Functional Analysis of the Osteoarthritis Susceptibility Loci Marked by the Polymorphisms rs10492367 and rs9350591

Katherine Johnson

A thesis submitted for the degree of Doctor of Philosophy

Institute of Cellular Medicine

March 2016

Abstract

Approximately 8.5 million people in the UK are affected by osteoarthritis (OA), a multifactorial, polygenic disease characterised by articular cartilage loss. In 2012, the arcOGEN Consortium reported on the largest OA genome-wide association scan (GWAS) to date, in which five regions of the genome were significantly associated with the disease in Europeans. I aimed to characterise two of these regions: rs10492367 is intergenic between *PTHLH* and *KLHL42*, while rs9350591 is intergenic between *FILIP1* and *SENP6*. *MYO6*, *TMEM30A*, *COX7A2* and *COL12A1* also surround rs9350591. There are no non-synonymous polymorphisms within either association region that could account for the signals. I first confirmed that the genes were expressed throughout *in vitro* chondrogenesis and osteoblastogenesis. Using quantitative real-time polymerase chain reaction, I identified the differential expression of *PTHLH*, *KLHL42*, *SENP6*, *MYO6*, *COX7A2* and *COL12A1* in articular cartilage stratified by disease state, joint and/or sex. There were no differences between the genotypic groups of either signal, corroborated by pyrosequencing which quantified allelic outputs. Expression quantitative trait loci, irrespective of rs9350591 genotype, acted upon *MYO6* and *COL12A1*. I used data from an Illumina BeadChip array to identify the hypermethylation of cg26466508 in rs9350591 risk allele carriers relative to non-risk allele homozygotes. In luciferase reporter assays, the alleles of polymorphisms in high linkage disequilibrium with rs10492367 displayed differential enhancer activity. Electrophoretic mobility shift assays were used to investigate protein binding to four of these polymorphisms, with RELA, SUB1 and TCF3 binding to rs10492367: chromatin immunoprecipitation confirmed these findings. Finally, I used silencing RNAs to knockdown the transcription factors in human articular chondrocytes, with SUB1 depletion resulting in a downregulation of *PTHLH*. Overall, I have highlighted the complexity of characterising GWAS signals. The data suggest functional roles for the regions, perhaps by mediating OA susceptibility during joint development rather than in end-stage diseased cartilage.

Dedication

To my darling parents.

Acknowledgements

I would like to thank my supervisors Prof. John Loughlin and Dr Louise Reynard for the support and advice they have given my throughout my Ph.D.

Dr Madhushika Ratnayake and Dr Emma Raine provided the raw quantitative real-time polymerase chain reaction data for the gene expression analyses of Chapter 3 and Chapter 4. I also thank Dr Madhushika Ratnayake, in addition to Maria Tselepi and Emma Rogers, for performing the *in vitro* chondrogenesis of Chapter 3.3.2 and Chapter 4.3.2.

Thank you to Dr Michael Rushton who provided the methylation microarray data, Dr Matthew Barter who provided the chondrogenesis microarray data and Dr Rodolfo Gomez who provided the osteoblastogenesis microarray data, all used in Chapter 3 and Chapter 4.

I would like to thank Dr Catherine McGillivray for experimental advice, particularly when I was performing the electrophoretic mobility shift assays in Chapter 5.

An undergraduate student under my supervision, Brooke Reed, performed the pyrosequencing experiments in Chapter 4.3.14.

Lucy Gentles collected and cultured human articular chondrocytes in preparation for the knockdown experiments of Chapter 5.

I would like to acknowledge the past and present members of the Osteoarthritis Genetics Group and the Musculoskeletal Research Group; particularly Prof. John Robinson, Dr Catherine McGillivray, Dr Madhushika Ratnayake, Lucy Gentles and Emma Rogers for their friendship and support during my Ph.D.

Finally, I would like to thank Arthritis Research UK for funding this research.

Contents

Contents of Figures, Tables and Equations

Figures

Tables

Equations

Abbreviations

3C	Chromosome conformation capture
3D	Three-dimensional
ACAN	Aggrecan
ACI	Autologous chondrocyte implantation
ADAM	A disintegrin and metalloproteinase
ADAMTS	A disintegrin and metalloproteinase with thrombospondin motif
AEI	Allelic expression imbalance
AGE	Advanced glycation end
ALDH1A2	Aldehyde dehydrogenase 1 family, member A2
ALPL	Alkaline phosphatase, liver/bone/kidney
AON	Antisense oligonucleotide
arcOGEN	Arthritis Research UK Osteoarthritis Genetics
ARNTL2	Aryl hydrocarbon receptor nuclear translocator-like 2
ARNTL2-AS1	ARNTL2 antisense RNA 1
ASTN2	Astrotactin 2
ASUN	Asunder spermatogenesis regulator
ATP	Adenosine triphosphate
ATP2A2a	ATPase Ca++ transporting cardiac muscle slow twitch 2a
ATPase	Adenosine triphosphatase
BACK	BTB and C-terminal kelch
BCAP29	B-cell receptor-associated protein 29
BLAT	BLAST-like alignment tool
BMI	Body mass index
BMP	Bone morphogenetic protein
BSA	Bovine serum albumin
BTB	Bric à brac, tamtrack and broad-complex
BTNL2	Butyrophilin-like 2
C12orf71	Chromosome 12 open reading frame 71
CCDC91	Coiled-coil domain containing 91
CDC50A	Cell division cycle 50A
CDMP	Cartilage-derived morphogenetic protein
cDNA	Complementary DNA
ChIA-PET	Chromatin interaction analysis with paired-end tag
ChIP	Chromatin immunoprecipitation

ChIP-Seq	Chromatin immunoprecipitation with sequencing
CHRD	Chordin
CHST11	Carbohydrate [chondroitin 4] sulfotransferase 11
CI	Confidence interval
COG5	Component of oligomeric Golgi complex 5
COL	Collagenous domain
COL11A1	Collagen, type XI, α1
COL12A1	Collagen, type XII, α1
COL2A1	Collagen, type II, α1
COMP	Cartilage oligomeric matrix protein
COX	Cyclooxygenase enzyme
COX7A2	Cytochrome c oxidase subunit VIIa polypeptide 2
CRISPR	Clustered regularly interspaced short palindromic repeats
CTGF	Connective tissue growth factor
DCN	Decorin
DEPC	Diethylpyrocarbonate
DLA	Disability living allowance
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferases
dNTP	Deoxynucleotide triphosphate
dsDNA	Double stranded DNA
DUS4L	Dihydrouridine synthase 4-like
ECM	Extracellular matrix
EMSA	Electrophoretic mobility shift assay
eQTL	Expression quantitative trait locus
ERG	V-ets avian erythroblastosis virus E26 oncogene homolog
F	Female
FACIT	Fibril-associated collagen with interrupted triple helices
FBS	Foetal bovine serum
FGF	Fibroblast growth factor
FGFR1	Fibroblast growth factor receptor 1
FGFR1OP2	FGFR1 oncogene partner 2
FILIP1	Filamin A interacting protein 1
FN3	Fibronectin type III
FP	Fat pad

GAG	Glycosaminoglycan
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GDF5	Growth differentiation factor 5
gDNA	Genomic DNA
GLT8D1	Glycosyltransferase 8 domain containing 1
GNL3	Guanine nucleotide binding protein-like 3
GPR22	G protein-coupled receptor 22
GWAS	Genome-wide association scan
HAC	Human articular chondrocyte
HBP1	HMG-box transcription factor 1
HGF	Hepatocyte growth factor
HLA-DQA1	Major histocompatibility complex, class II, DQ α1
HLA-DQB1	Major histocompatibility complex, class II, DQ β1
HLA-DRA	Major histocompatibility complex, class II, DR α
HLA-DRB1	Major histocompatibility complex, class II, DR β1
HLA-DRB5	Major histocompatibility complex, class II, DR β5
HMM	Humoral hypercalcaemia of malignancy
HPRT1	Hypoxanthine phosphoribosyltransferase 1
HRP	Horseradish peroxidase
IGF	Insulin-like growth factor
IHH	Indian hedgehog
IL	Interleukin
IMPG1	Interphotoreceptor matrix proteoglycan 1
IQ	Isoleucine [I] glutamine [Q]
ITS+L	Insulin, transferin, selenium, linoleic acid premix
IVL	Involucrin
JIA	Juvenile idiopathic arthritis
K-L	Kellgren-Lawrence
KLHDC5	Kelch domain containing 5
KLHL42	Kelch-like family member 42
LB	Lysogeny broth
LD	Linkage disequilibrium
M	Male
MAF	Minor allele frequency
MALDI	Matrix-assisted laser desorption/ionisation imaging

MANSC4	MANSC domain containing 4
MED21	Mediator complex subunit 21
MICAL3	Microtubule associated monooxygenase, calponin and LIM domain containing 3
min	Minute
miRNA	Micro ribonucleic acid
MITF	Microphthalmia-associated transcription factor
MMP	Matrix metalloproteinase
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
MRPS35	Mitochondrial ribosomal protein S35
MSC	Mesenchymal stem cell
MT-MMP	Membrane-type matrix metalloproteinase
MYO6	Myosin VI
NC	Non-collagenous domain
NCBI	National Center for Biotechnology Information
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NLS	Nuclear localisation signal
NOF	Neck of femur
NOG	Noggin
NSAID	Nonsteroidal anti-inflammatory drug
NT	Non-targeting
OA	Osteoarthritis
OR	Odds ratio
ORMDL3	ORMDL sphingolipid biosynthesis regulator 3
PAX3	Paired box 3
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PIP	Personal independence payment
PPFIBP1	PTPRF interacting protein, binding protein 1 [liprin β1]
PRKAR2B	Protein kinase, cAMP-dependent, regulatory type II β
PRM2	Protamine 2
PTGER4	Prostaglandin E receptor 4
PTH	Parathyroid hormone

PTH1R	Parathyroid hormone type 1 receptor
PTHLH	Parathyroid hormone-like hormone
PTHrP	Parathyroid hormone-related protein
PTPRF	Protein tyrosine phosphatase, receptor type F
qPCR	Quantitative real-time-PCR
RA	Rheumatoid arthritis
RELA	v-rel avian reticuloendotheliosis viral oncogene homolog A
REP15	RAB15 effector protein
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RNAi	RNA-mediated interference
rpm	Revolutions per minute
RT-PCR	Reverse transcriptase polymerase chain reaction
S100A4	S100 calcium binding protein A4
SAM	S-adenosylmethionine
sec	Second
SEM	Standard error of the mean
SENP6	SUMO1/sentrin-specific peptidase 6
siRNA	Small interfering RNA
SMCO2	Single-pass membrane protein with coiled-coil domains 2
SNP	Single nucleotide polymorphism
SOX9	SRY [sex determining region] Y-box 9
Sp1	Sp1 transcription factor
Sp3	Sp3 transcription factor
STK38L	Serine/threonine kinase 38 like
SUB1	SUB1 homolog [S. cerevisiae]
SUMO	Small ubiquitin-like modifier
sv	Snell's waltzer
Sy	Synovium
TBE	Tris/borate/EDTA
TBS	Tris-buffered saline
TCF3	Transcription factor 3
TGF	Transforming growth factor
THR	Total hip replacement
TIMP	Tissue inhibitor of metalloproteinase

TJR	Total joint replacement
TKR	Total knee replacement
TM7SF3	Transmembrane 7 superfamily member 3
TMCO1	Transmembrane and coiled-coil domains 1
TMEM30A	Transmembrane protein 30A
TNF	Tumour necrosis factor
TSPN	Thrombospondin N-terminal-like
TSS	Transcription start site
UTR	Untranslated region
UV	Ultraviolet
v/v	Volume/volume
VDR	Vitamin D receptor
VEGF	Vascular endothelial growth factor
vWA	von Willebrand factor A-like
w/v	Weight/volume
WOMAC	Western Ontario and McMaster Universities
XCL1	Chemokine [C motif] ligand 1
ΔCt	Delta Ct

Chapter 1. Introduction

1.1 The Musculoskeletal System

1.1.1 *Introduction to the musculoskeletal system*

The musculoskeletal system exists to allow the controlled movement of the human body by providing structure, support and protection. It comprises the skeleton and muscles in addition to connective tissues such as cartilage, ligaments and tendons. Perturbations of this fundamental system can lead to the development of musculoskeletal disorders, the leading cause of long-term pain and disability worldwide (Woolf and Pfleger, 2003). There are over 200 known musculoskeletal conditions (Arthritis Research UK, 2013) and some wide-ranging examples include muscular dystrophy, fibromyalgia, idiopathic lower back pain, Dupuytren's contractures, ankylosing spondylitis and arthritis.

1.1.2 *Economic impact of musculoskeletal health*

There is an ever increasing necessity to understand musculoskeletal disease mechanisms in order to provide more appropriate measures to prevent, treat and manage the associated conditions. In the 2012-2013 financial year, the budget assigned to the National Health Service (NHS) for the musculoskeletal health programme was £5.34 billion (NHS England, 2015) and it is forecast to rise in accordance with an ageing UK population. Each year, approximately one fifth of the UK population consult their general practitioner about a musculoskeletal condition (Arthritis Research UK Primary Care Centre Keele University, 2014). In addition, it is estimated that in 2013 alone, 30.60 million UK work days were lost as a result of musculoskeletal conditions (Office for National Statistics, 2014). A further economic burden is in the form of Disability Living Allowance (DLA; also known as Personal Independence Payment [PIP]), a tax-free, non-means-tested benefit system to aid those who have personal care or mobility needs as a result of disability. As detailed in Table 1.1, three of the top eight conditions with the highest number of claimants are directly related to musculoskeletal disorders (Department for Work and Pensions, 2015). Over half a million people in the UK claim DLA for arthritis, with an average weekly benefit of £88. Individuals with diseases of the muscles, bones or joints account for 173,530 claimants, receiving an average of £79 per week in benefits, while unspecified back pain accounts for 155,040 of claimants, each receiving an average of £87 per week. Other notable conditions not listed individually include spondylosis (80,600 claimants, £85/week), multiple sclerosis (64,080 claimants, £110/week), Parkinson's disease (17,530 claimants, £108/week) and motor

neurone disease (1,580 claimants, £113/week). These figures clearly highlight the financial need for the continued research into musculoskeletal disorders.

Disabling condition	Number of claimants	Average weekly amount of benefit (£)
Arthritis	532,920	88
Learning difficulties	448,300	78
Psychosis	256,000	71
Disease of the muscles, bones or joints	173,530	79
Psychoneurosis	172,150	59
Back pain – other/precise diagnosis not specified	155,040	87
Neurological diseases	134,340	97
Heart disease	106,260	88
Other conditions	1,199,710	91
Total	3,178,300	-

Table 1.1. Number of individuals in the UK who are in receipt of disability living allowance and the average benefit claimed per individual per week. The top eight conditions are listed individually, all of which have over 100,000 claimants each. Other conditions (total of 47 listed conditions each with less than 100,000 claimants). Data as of November 2014 (Department for Work and Pensions, 2015).

1.2 Defining Osteoarthritis

1.2.1 *Introduction to arthritis*

Arthritis is a disease of the joints and accounts for the greatest number of claimants of DLA for an individual disabling condition in the UK. The OANation 2012 report, commissioned by Arthritis Care, suggests that arthritis affects 10 million people in the UK alone (Arthritis Care, 2012). The two main forms of arthritis are osteoarthritis (OA) and rheumatoid arthritis (RA), both of which display the hallmark features of pain, stiffness and inflammation at an affected joint, but with different underlying causes and disease mechanisms.

1.2.2 *Rheumatoid arthritis and other arthritis-related conditions*

RA is the second most common form of arthritis, affecting over 400,000 people in the UK (Symmons *et al.*, 2002). The onset of RA is often seen between the ages of 40 and 50 years old, with women three times more likely to develop the disease than men. This chronic autoimmune disease is characterised by the inflammation of the outer membrane which surrounds a synovial joint. This causes irritation and swelling of the affected joint and consequently leads to changes in joint shape alongside bone and cartilage breakdown.

Other less common forms of arthritis include psoriatic arthritis (associated with the development of psoriasis), enteropathic arthritis (associated with inflammatory bowel disorders), reactive arthritis (short-term disorder following an infection), secondary arthritis (following joint damage or injury) and gout (episodic incidences caused by build-up of sodium urate crystals around the joint).

Around 15,000 children and young adults are also known to be affected by arthritis (Sacks *et al.*, 2007), specifically termed juvenile idiopathic arthritis (JIA). The two most common types of JIA are oligoarthritis and polyarthritis. Oligoarthritis accounts for around 60% of JIA, with the often mild symptoms affecting the knees. Polyarthritis is more widespread, affecting the joints of the hips, knees, hands, neck and jaw. Other forms of JIA include enthesitis-related JIA (at sites where bone and tendon connect), psoriatic arthritis (often affecting fingers and toes) and systemic-onset JIA (part of a more general illness that includes lethargy and weight loss). Generally, the symptoms of JIA improve with age.

1.2.3 *Osteoarthritis epidemiology and prevalence*

Affecting approximately 8.50 million people in the UK, OA is the most common form of arthritis, with this number predicted to rise to 17 million people by 2030 (Arthritis Care, 2012) in accordance with an ageing population (Office for National Statistics, 2015). Given the overlap between musculoskeletal disorders, the variations in disease definition, the wide range of affected joint sites and the extensive diversity between populations, the exact prevalence of OA is often difficult to calculate (Lawrence *et al.*, 2008). Estimates from the Global Burden of Disease Study 2010 (Cross *et al.*, 2014) suggest that the worldwide prevalence of radiographically-confirmed symptomatic knee OA (3.80%) is vastly greater than that of hip OA (0.85%). In both cases, the age-standardised prevalence was higher in females than in males (Table 1.2).

	Knee OA	**Hip OA**
Worldwide prevalence (male; female)	3.80% (2.80%; 4.80%)	0.85% (0.70%; 0.98%)
Prevalence peak	50 years of age	Consistent increase with age
Highest prevalence	Asia Pacific high income	North America high income
Lowest prevalence	South Asia, south east Asia	North Africa, Middle East, east Asia

Table 1.2. Estimates of the worldwide prevalence of osteoarthritis. The estimates were made as part of the Global Burden of Disease Study 2010 (Cross *et al.*, 2014); high income areas are deemed as such by the authors; OA (osteoarthritis).

1.2.4 *Osteoarthritis diagnosis*

OA is characterised by the progressive degradation of articular cartilage, the connective tissue that covers the joint surfaces and functions to allow smooth articulation. This causes pain and inflammation at the affected site and subsequently leads to diminished joint function and mobility. Other indications that may be present at an affected joint include effusion, tenderness, crepitus and locking. The disease primarily affects synovial joints that are subject to excessive or repeated mechanical loading such as the hands, knees, hips and spine (Felson, 2006).

OA can be classified into one of two subcategories, primary OA or secondary OA. This is dependent on the underlying cause of the disease. When a specific cause is unknown and the OA is most likely due to the sustained and repeated use of a joint that simply occurs with age, the disease is considered to be primary OA. OA can also develop after joint trauma or as a result of an underlying medical condition, in which case it is termed secondary OA (Altman *et al.*, 1986).

A diagnosis of OA is generally based on a physical examination by a general practitioner in order to identify any of the noted indications. The National Institute for Health and Care Excellence (NICE) guidelines published in 2014 recommend that OA should be diagnosed if a patient is i) 45 years old or over, ii) has activity-related joint pain and iii) experiences no morning joint-related stiffness or stiffness that lasts no longer than 30 minutes (National Institute for Health and Care Excellence, 2014). If further investigation is required, the clinical assessment can be performed in combination with X-rays. A radiographic procedure allows the detection of clinical features such as narrowing of the joint space, subchondral bone sclerosis, osteophyte formation and an overall change in joint structure (Arden and Nevitt, 2006). Although used less frequently, a further option is the use of magnetic resonance imaging (MRI) scans, which can detect early morphological changes during the development of OA (Agnesi *et al.*, 2008).

The Kellgren-Lawrence (K-L) grading system is a tool used to assess the severity of knee OA from radiographic images (Kellgren and Lawrence, 1957). This ranges from 0 (no radiographic evidence of OA) to 4 (osteophyte formations, joint space narrowing, subchondral bone sclerosis and deformed bone structure), whereby a score of 2 or greater lends itself to a definite classification of OA. A clinical study found that the K-L scores of patients with knee OA significantly and positively correlated with age and disease duration (Cubukcu *et al.*,

2012). The Western Ontario and McMaster Universities (WOMAC) Osteoarthritis Index is a validated set of questionnaires used to measure the clinical pain associated with hip and knee OA. This is a self-reporting questionnaire that considers the pain, stiffness and functional limitations of an individual. However, the correlation between radiographic disease severity and pain perception has been reported to be variable and oftentimes discordant (Hannan *et al.*, 2000; Felson, 2006). This may be due to the subjective nature of the pain scoring or perhaps due to the WOMAC OA Index capturing a more widespread suffering of pain and not specifically at the affected joints. Overall, a correct diagnosis of OA can be difficult to achieve due to the subjective nature of the condition and the variation in the presenting clinical symptoms.

1.2.5 *Management and treatment of osteoarthritis*

It is recommended that a holistic approach is taken when assessing the best course of action in treating an individual who has been diagnosed with OA (Figure 1.1), as living with such a disabling condition can lead to an individual's inability to work, reliance on others, isolation and depression (Lin, 2008). There are three main, non-surgical avenues discussed in the NICE guidelines (National Institute for Health and Care Excellence, 2014) that can be explored in the management of OA: pharmacological, non-pharmacological, and education and self-management.

Clinicians are advised to encourage positive behavioural changes and to offer ongoing verbal and written information on OA in order to help enhance a patient's understanding and management of the disease. Aerobic exercise and local muscle strengthening is central in the management of OA, regardless of the presentation or stage of disease, in order to achieve reduced pain and enhanced joint function. Similarly, weight loss of overweight or obese individuals is very strongly recommended. Assistive devices such as walking aids and appropriate footwear can also be offered.

Should these approaches be insufficient, pharmacological interventions can be considered but do carry potential off-target effects. Paracetamol and acetaminophen are usually the first analgesics to be offered, but where they prove to be ineffective, nonsteroidal anti-inflammatory drugs (NSAIDs) can be used as second-line agents. Non-selective NSAIDs target both isoforms of the cyclooxygenase enzyme (COX-1 and COX-2) and by inhibiting their catalysis of prostanoid biosynthesis (Simmons *et al.*, 2004), the inflammation and pain of arthritis can be alleviated. Specifically, COX-2 modulates inflammatory responses

(Willoughby *et al.*, 2000) and it is this isoform that more recent NSAIDs are targeted towards. COX-2-selective inhibitors such as celecoxib have the benefit of a reduced incidence of NSAID-associated gastrointestinal complications (Bensen *et al.*, 1999). Although previously thought to be suitable treatments for OA, glucosamine and chondroitin products are not recommended as therapeutics, as administration both alone and in combination did not result in a reduction of joint pain or of joint space narrowing compared to a placebo treatment (Sawitzke *et al.*, 2010). Moreover, intra-articular corticosteroid injections are similarly not recommended as only short-term benefits have been established (Arroll and Goodyear-Smith, 2004).

Surgical intervention will be considered if the management strategies fail to delay disease progression or to improve the patient's quality of life. Options include arthroscopic lavage to remove joint debris, osteotomy to realign the joint, and arthroplasty to replace the affected joint (Ronn *et al.*, 2011). Although minimally invasive, it has been shown that the outcome post-arthroscopic lavage was no different to a placebo surgery (Moseley *et al.*, 2002). In addition, the benefit of an osteotomy procedure is limited as it is reserved only for cases of unicompartmental OA. Total joint replacement is the only approach that can alter the progression of the disease. Although the management of OA can be largely tailored to suit the specific needs of an individual, there is currently no therapeutic intervention that can completely treat or prevent OA.

Current research into the treatment of OA focusses on cell-based therapies. One such approach is the use of autologous chondrocyte implantation (ACI), whereby chondrocytes are isolated from a cartilage biopsy and expanded *ex vivo* before transplantation back into the same individual (Ruano-Ravina and Jato Diaz, 2006). A more recent development is that of three dimensional (3D)-ACI, where extracted chondrocytes are embedded into a scaffold, such as the commercially available collagen type I matrix, before expansion and re-implantation (Kuroda *et al.*, 2011). However, as of yet this has not been translated into humans and approved for clinical use. In addition, the requirement of a substantial number of cells combined with the damage sustained by healthy cartilage means this technology is not considered suitable for routine OA treatment (Oldershaw, 2012). An alternative cell source can be found in the bone marrow of a patient where multipotent mesenchymal stem cells (MSCs) exist. MSCs can also be isolated from tissues such as synovium or muscle (De Bari *et al.*, 2001; Peng and Huard, 2004), making their isolation more accessible. In appropriate conditions, MSCs can differentiate into cells of a mesenchyme lineage, such as bone,

cartilage, tendon and ligament (Qi *et al.*, 2012), and it is this that can be exploited to generate *ex vivo* chondrocyte cultures. Akin to the procedure of ACI, the chondrocyte cultures can then be transplanted into the donor patient, with the aim of recapitulating healthy articular cartilage. Overall, despite the promising research, the success of cell therapies are disputed and as such are currently not recommended for treatment of OA in the UK.

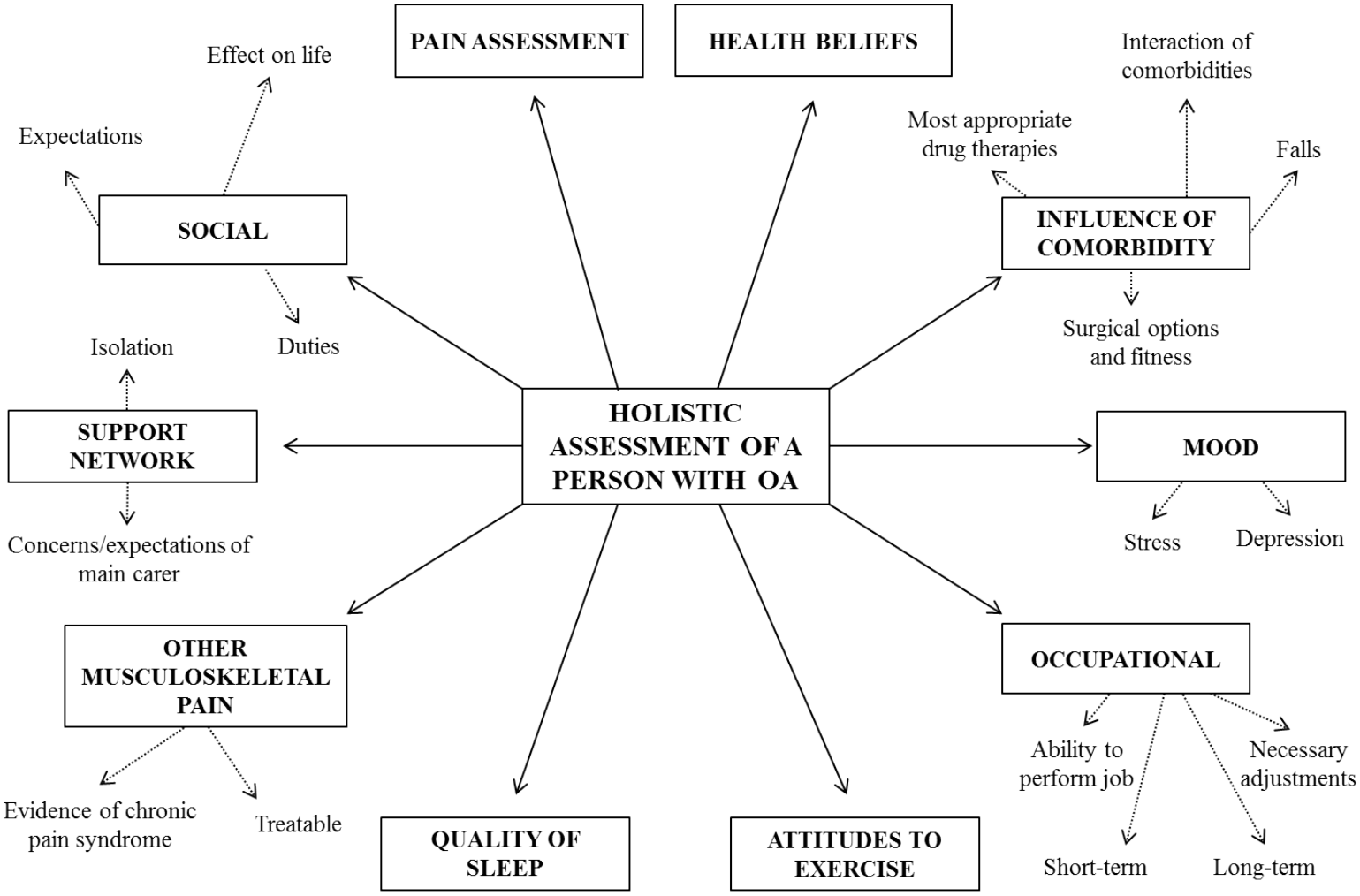

Figure 1.1. Considerations when assessing an individual with osteoarthritis. The key factors include the health, social and economic impact of the disease; comorbidities; and health beliefs including expectations and understanding of the disease. OA (osteoarthritis); adapted from the NICE guidelines (National Institute for Health and Care Excellence, 2014).

1.3 Synovial Joint Structure and Function

1.3.1 *Introduction to the synovial joint*

A synovial joint, or diarthrosis, allows free articulation of the skeletal structure and is defined by the presence of a surrounding fibrous capsule. Several tissues comprise a diarthrosis within the synovium and include subchondral bone, articular cartilage, synovial fluid, tendon and ligament. The maintenance and correct functioning of a synovial joint is entirely reliant upon both the individual contributions of each of these tissues and the interplay between them as an entire organ. It is when the integrity of the joint is impaired through the damage of one of the elements that the correct functioning is affected. Phenotypic changes can be observed in several of the joint tissues when an individual has OA, such as osteophyte formation, degradation of articular cartilage and damaged menisci (Figure 1.2). Because of this, OA is now more readily classified as a disease of the entire joint rather than only of the cartilage (Poole, 2012).

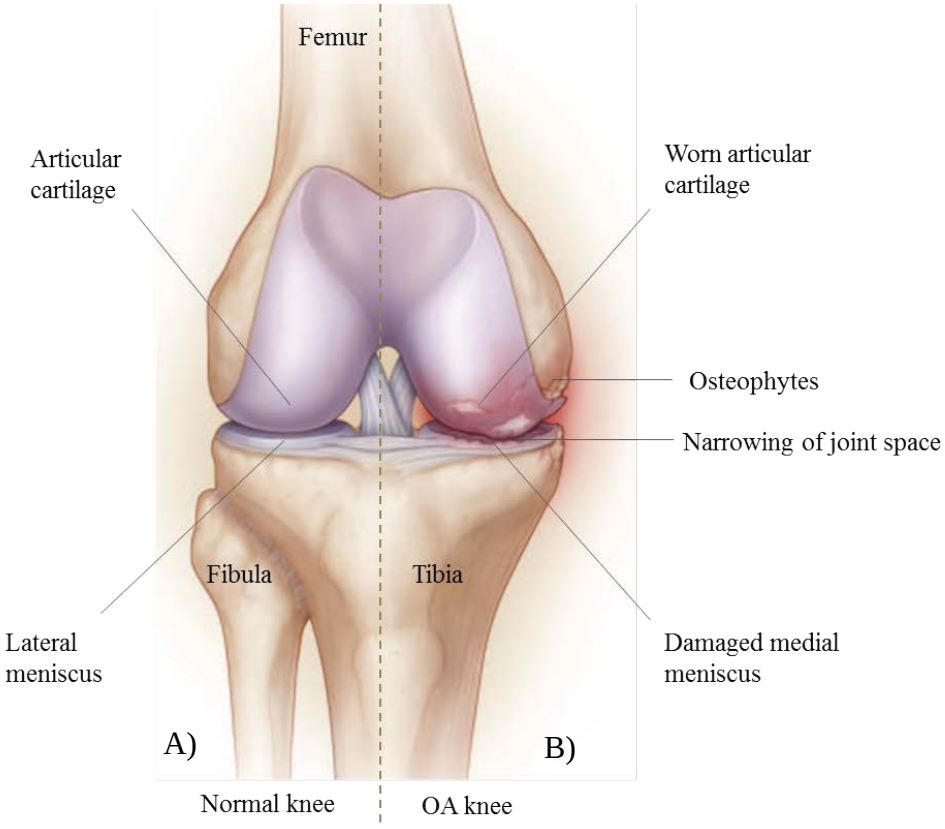

Figure 1.2. Comparison of the structure of a normal synovial joint structure with a joint displaying hallmark characteristics of osteoarthritis. A) Articular cartilage and synovial fluid provide protection and support to allow smooth movement of the joint. B) Damaged articular cartilage, osteophyte formation and narrowing of the joint space contribute to a painful and inflamed joint with restricted movement. OA (osteoarthritis); adapted from (Felson, 2006).

1.3.2 *Articular cartilage*

Within a synovial joint, articular cartilage is found as a 3-5 mm thick covering over the ends of bones. Its purpose is to provide smooth, frictionless movement of the joint, and to absorb and dissipate load in order to resist compressive stresses (Wieland *et al.*, 2005). Articular cartilage does not have a nervous, lymphatic or vascular system (Poole, 1997), and the only cell type to exist in this anaerobic tissue is the chondrocyte. MSC differentiation into chondroblasts is followed by a series of mitotic divisions that result in the production of extracellular matrix (ECM) components and the maturation of chondrocytes (Deng *et al.*, 2008). Chondrocytes are sparse in population, accounting for 2% of the volume of articular cartilage (Alford and Cole, 2005), and therefore do not form cell-cell interactions. The purpose of the metabolically-active chondrocytes is to maintain the surrounding ECM, and it is the composition and architectural arrangement of this that is responsible for the mechanical properties of the tissue. The principle components of the ECM, in order of contribution to the wet weight of articular cartilage, are water (65-85%), collagen (12-24%) and proteoglycans (3-6%) (Figure 1.3.A).

Type II collagen is the most abundant structural protein within the ECM, although other collagen types do contribute to the cartilage composition. Members of the collagen family are individually composed of three alpha-chain polypeptides that form a triple helix and then in turn, fibrils. The fibrils interact and form an overall mesh throughout the ECM, generating the characteristic tensile strength of articular cartilage (Responte *et al.*, 2007). This also supports the position of macromolecules within the matrix. To strengthen the overall fibril mesh, two additional collagens covalently interact with type II collagen. On the surface layers of the mesh, type IX collagen binds and, as a result of the outward COL3 and NC4 domain projections, permits interactions with other ECM proteins. Within the fibrillar structure, type XI collagen binds the core type II collagen (Buckwalter and Mankin, 1998). Other non-collagenous proteins, such as cartilage oligomeric matrix protein (COMP), have the ability to bind and stabilise the structure of the collagen fibrils (Rosenberg *et al.*, 1998).

A proteoglycan is formed by a central protein core to which linear glycosaminoglycans (GAGs) attach. Common GAGs present in articular cartilage include hyaluronan, chondroitin sulfate and keratan sulfate (Esko *et al.*, 2009). Aggrecan is both the most abundant and the largest proteoglycan in the articular cartilage ECM and is comprised of three globular domains (G1, G2 and G3). G1 binds hyaluronan and link proteins, while the interglobular domain between G2 and G3 covalently bind over 100 chondroitin sulfate and keratan sulfate

chains (Knudson and Knudson, 2001). Aggrecan aggregates are formed by non-covalently binding hyaluronan, stabilised by link proteins, meaning the large protein structures are retained in the collagenous mesh. Smaller, non-aggregating proteoglycans include decorin, biglycan and fibromodulin. It is postulated that decorin and fibromodulin interact with type II collagen, while biglycan has a tendency to interact with type VI collagen near the surface of chondrocytes (Roughley and Lee, 1994).

Nutrient transport is a vital role of the water contained in the articular cartilage ECM. In addition, the presence of water allows for a pressure within the ECM to be generated, which therefore creates the resistance necessary to withstand loading. The retention of water is largely due to the containment structures physically created by the collagenous mesh and the interacting macromolecules; but also due to the negative charges of chondroitin sulfate and keratan sulfate (Jerosch, 2011). Overall, the interactions of the main components of articular cartilage all contribute to the unique biomechanical properties of the tissue.

The morphology and organisation of the ECM components are not uniform throughout the layer of articular cartilage. The tissue is composed of non-calcified and calcified cartilage, discernible by the presence of a tidemark marking the boundary of calcification (Figure 1.3.B). The calcified cartilage represents a transition between the overlying articular cartilage and the underlying subchondral bone (Suri and Walsh, 2012), where type X collagen is the primary collagenous component (Madry *et al.*, 2010). Cells of this region are smaller with a presumed lower metabolic activity than the non-calcified tissue (Buckwalter *et al.*, 2005). The histological appearance of the tidemark suggests an accumulation of macromolecules, although the exact nature of the basophilic line is unclear (Buckwalter *et al.*, 2005; Suri and Walsh, 2012). Within the non-calcified cartilage, there are a further three sub-divisions determined by their structural appearance: the deep, middle and superficial zones. The deep zone represents approximately 30% of the total cartilage thickness. Here, the chondrocytes align in a columnar formation, parallel to the collagen fibrils which are perpendicular to the subchondral bone. The collagen fibrils anchor the non-calcified cartilage in position by extending through the tidemark into the calcified zone (Buckwalter *et al.*, 2005). The middle zone represents approximately 40-60% of the total cartilage thickness and has a more randomly organised collagen structure with sparsely distributed chondrocytes. The superficial zone is the region in closest physical proximity to the synovial space and represents 10-20% of the total cartilage thickness. This zone is responsible for the first line of protection against stresses on the joint, and as such collagen fibrils are densely packed in a formation parallel to

the plane of the articular cartilage surface (Responte *et al.*, 2007), while the chondrocytes are ellipsoid in shape (Buckwalter *et al.*, 2005). In addition, cells of the superficial zone synthesise and secrete the lubricant proteoglycan 4 (also known as lubricin) into the synovial space (Grogan *et al.*, 2013). Relative to the middle and deep zones, the content of both collagen and water is highest in the superficial zone, while the content of aggrecan is lowest (Buckwalter and Mankin, 1998; Buckwalter *et al.*, 2005).

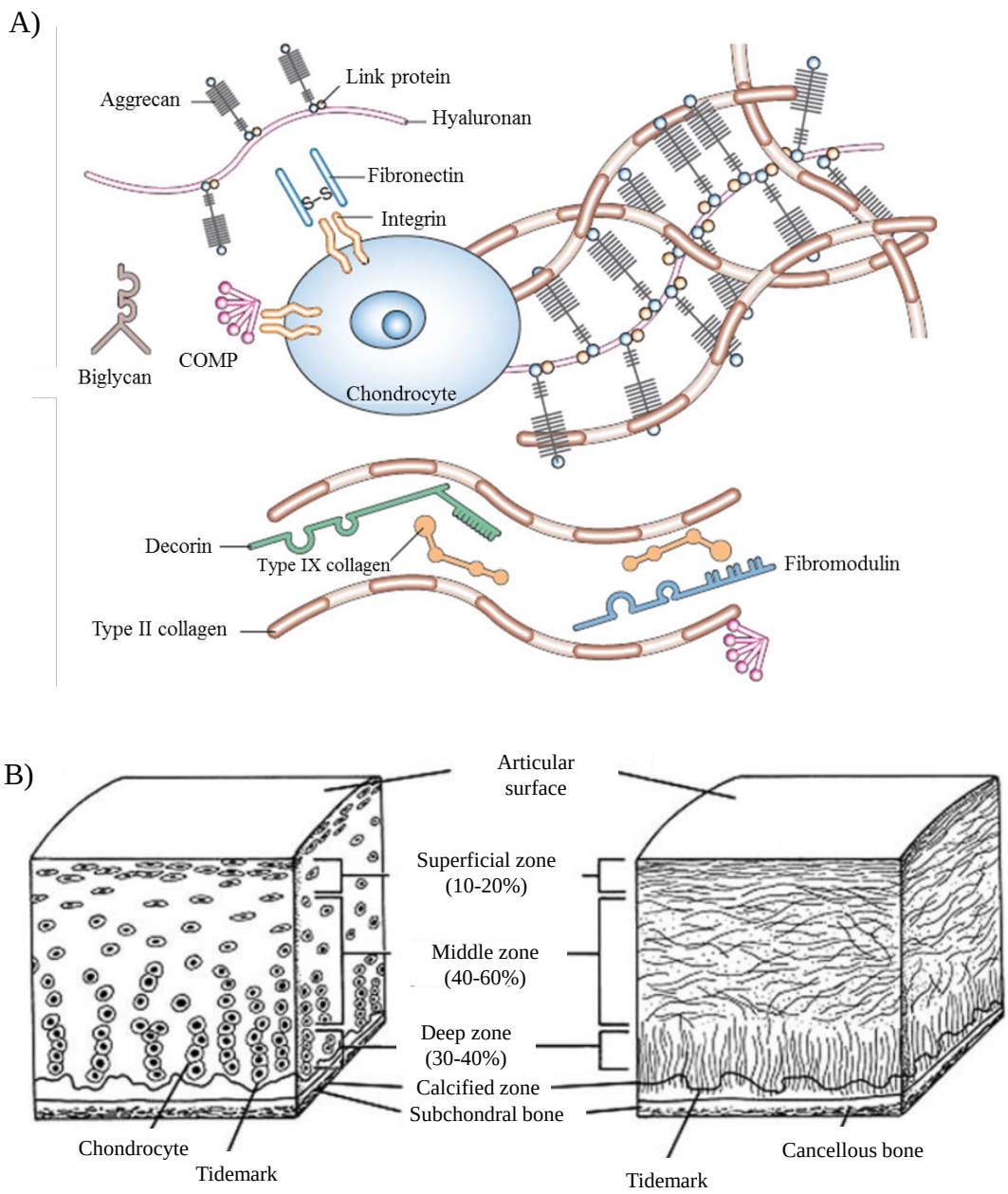

Figure 1.3. The structure of the articular cartilage of synovial joints. A) Within the extracellular matrix of articular cartilage, three classes of protein exist: collagens, proteoglycans and non-collagenous proteins. Tensile strength is created by the interaction between type II collagen and small, negatively charged proteoglycans. Image taken from (Chen *et al.*, 2006a). B) Non-calcified cartilage comprises three distinct zones, the superficial zone, the middle zone and the deep zone. Cells in the deep zone are aligned perpendicular to the subchondral bone plate; cells in the middle zone are sparse, retaining a similar shape to those of the deep zone; and cells of the superficial zone are elongated and lie parallel to the articular cartilage surface. The tidemark separates the hyaline cartilage from the calcified cartilage. Calcified cartilage is present at the interface between the non-calcified cartilage and the subchondral bone. Image taken from (Buckwalter *et al.*, 1994).

1.3.3 *Subchondral bone*

The subchondral bone lies directly beneath the articular cartilage, and can be classified into two distinct layers defined by their structure and function. The layer closest to the calcified zone of articular cartilage is termed the subchondral bone plate and beneath this is the cancellous bone. Through specialised channels, the subchondral bone serves as a route for the nervous and vascular systems to supply vital nutrients to the chondrocytes in the calcified cartilage (Madry *et al.*, 2010). A further primary function of the bone is to support the joint structure and dissipate stresses (Li *et al.*, 2013), which is achieved by the structure of the two layers. Type I collagen fibrils, the primary collagenous protein in the subchondral bone, is arranged into sheets within the subchondral bone plate (Madry *et al.*, 2010), while the cancellous bone is composed of precisely-orientated trabeculae (Li *et al.*, 2013). The subchondral bone matrix is mineralised by calcium hydroxyapatite crystals which intersperse the type I collagen fibrils (Wehrli, 2007). The dynamic nature of the subchondral bone is maintained by an equilibrium between the actions of osteoblasts that can deposit new bone, and osteoclasts that are involved in bone resorption (Wehrli, 2007).

1.3.4 *Joint capsule, synovial membrane and other tissues of the synovial joint*

The capsule surrounding a synovial joint is a fibrous tissue that physically connects the bones of a joint to add structure and stability. Within this, the synovial membrane is found, which is a lining made of two cell types that have discreet characteristics. Type A cells arise from a macrophage lineage and function to maintain an aseptic environment through the endocytosis of invading pathogens. Type B synoviocytes, on the other hand, are fibroblast-like and are responsible for the production of lubricin, an essential joint lubricant, and hyaluronan, a key ECM constituent that increases the viscosity of the synovial fluid (de Sousa *et al.*, 2014). The function of the synovial fluid is three-fold: to facilitate the transport of nutrients to the articular cartilage, to provide a lubricating medium through which the articulating components of a joint can move, and to supress the stresses and strains absorbed by the joint (Allan, 1998).

Tendons (connecting muscle to bone) and ligaments (connecting bone to bone) are both primarily composed of type I collagen and they act to stabilise the synovial joint structure (Benjamin and Ralphs, 1998). The presence of the infrapatellar fat pad additionally enhances stability as its flexibility allows for its adaptation to fit the contours of the joint, while its position allows for the absorption of excessive loads (Vahlensieck *et al.*, 2002). Finally, the functional unit of the menisci, composed mainly of type I collagen, acts to dissipate stress

evenly over the cartilage surface and to balance the loading of the joint (Messner and Gao, 1998).

1.4 Joint Homeostasis

1.4.1 *Introduction to joint homeostasis*

The maintenance of a healthy joint requires a fine balance of anabolic and catabolic factors that can be either mechanical or biochemical. Each of the different tissues within a synovial joint contribute to this overall equilibrium. The onset of OA is associated with a disruption of this dynamic balance to favour catabolism, observed by the resulting loss of articular cartilage integrity (Aigner *et al.*, 2006). Often, the mechanisms involved in anabolism and catabolism overlap and therefore do not have independent functions.

1.4.2 *Anabolic factors*

Anabolism refers to the promotion of tissue generation. Growth factors are particularly important in the synthesis and maintenance of joint tissue and include proteins such as transforming growth factor (TGF-β1, TGF-β2 and TGF-β3), bone morphogenetic protein (BMP), fibroblast growth factor (FGF-2, FGF-4 and FGF-8) and insulin-like growth factor (IGF-I). The anabolic function of the active forms of TGF-β is to regulate chondrocyte proliferation and stimulate the synthesis of type II collagen and aggrecan of the ECM (Finnson *et al.*, 2012). Additionally, TGF-β can promote the action of tissue inhibitors of metalloproteinases (TIMPs) to further inhibit catabolic proteases, while also inhibiting the catabolic interleukin-1 (IL-1) and tumour necrosis factor (TNF) (Hui *et al.*, 2001; Finnson *et al.*, 2012). Other members of the TGF-β superfamily include BMPs, which stimulate chondrocyte growth and bone formation (Martel-Pelletier *et al.*, 2008). FGFs and IGFs act to stimulate the synthesis of the ECM proteoglycans and type II collagen (Martel-Pelletier *et al.*, 2008). Other anabolic factors include cartilage-derived morphogenetic protein (CDMP), connective tissue growth factor (CTGF) and hepatocyte growth factor (HGF) (Martel-Pelletier *et al.*, 2008).

1.4.3 *Catabolic factors*

Catabolism refers to the promotion of tissue degeneration. Central to articular cartilage catabolism are active matrix metalloproteinases (MMPs), which are stimulated by the inflammatory cytokines IL-1, IL-17, IL-18 and TNF to target the cleavage of ECM proteins such as collagen and aggrecan (Sandell and Aigner, 2001). These cytokines, synthesised by articular cartilage and synovium, also inhibit the actions of TIMPs. Collagenases (MMP-1, -8

and -13), gelatinases (MMP-2 and -9), stromelysins (MMP-3, -10 and -11), membrane-type MMPs (MT-MMPs) and adamalysins (a disintegrin and metalloproteinase [ADAM] and a disintegrin and metalloproteinase with thrombospondin motifs [ADAMTS]) are all classified into the MMP family (Martel-Pelletier *et al.*, 2008).

1.5 Development of the Synovial Joint Structure

1.5.1 *Introduction to skeletogenesis*

The origin of a mammalian bone can be one of three lineages which determines the nature of the skeletal components that are created. The paraxial mesoderm gives rise to the axis skeleton, the lateral plate mesoderm forms bones of the appendicular skeleton and the neural crest is the origin of facial bones (Hojo *et al.*, 2010). To form the skeletal structures, there are two types of osteogenesis that can occur: intramembranous and endochondral. Intramembranous ossification is a one-step process whereby mesenchymal cells directly differentiate into the bone-forming osteoblast cells; in contrast, endochondral ossification is a two-step process that involves the intermediate formation of a cartilage template from which bone is then generated. Endochondral ossification is the typical process by which the axis and appendicular bones are formed (Hojo *et al.*, 2010).

1.5.2 *Synovial joint determination and formation*

During skeletogenesis, the individual bones of a synovial joint originate from one initial structure formed by a condensation of mesenchymal cells that assume the configuration of the bones (Yasuda and de Crombrugghe, 2009), which is mediated, in part, by cell adhesion molecules (Archer *et al.*, 2003). Joint site determination and bone formation occur concurrently. The site of an eventual joint is demarcated by a thin layer of flat mesenchymal cells that align perpendicular to the bone axis (Pacifici *et al.*, 2005), which forms across the plane of the cartilaginous template and is hence called an interzone (Decker *et al.*, 2014). This is specified by a number of factors including the activation of the expression of the TGF-β family member *GDF5* (growth differentiation factor 5); the downregulation of the chondrogenic marker *SOX9* (SRY [sex determining region] Y-box 9) and its target *COL2A1* (collagen, type II, α1); the expression of *FGF2*, *FGF4* and *FGF13*; the upregulation of the transcription factor *ERG* (v-ets avian erythroblastosis virus E26 oncogene homolog); and the increased activity of the Wnt/β-catenin signalling pathway (Pacifici *et al.*, 2005; Yasuda and de Crombrugghe, 2009). In addition, the expression of the TGF-β family members *BMP2* and *BMP4*, and the BMP antagonists *CHRD* (chordin) and *NOG* (noggin) are all upregulated at the interzone. In combination, this functions to determine a separate cell fate by preventing

the progenitor cells differentiating into chondrocytes in accordance with the surrounding mesenchymal cell condensation. The progression of joint formation involves the separation of the structure to form the synovial cavity followed by a morphogenic process that gives rise to the shape of the joint (Figure 1.4). The exact mechanisms of this are unclear, however it is likely to be a combination of GDF5-driven chondrogenic growth at the tip of the distal end and enhanced cell proliferation around the neck of the proximal end. The outcome is the protrusion of a convex distal end into a concave proximal end, which forms a complementary joint structure, the individual elements of which can then mature to complete joint formation (Pacifici *et al.*, 2005).

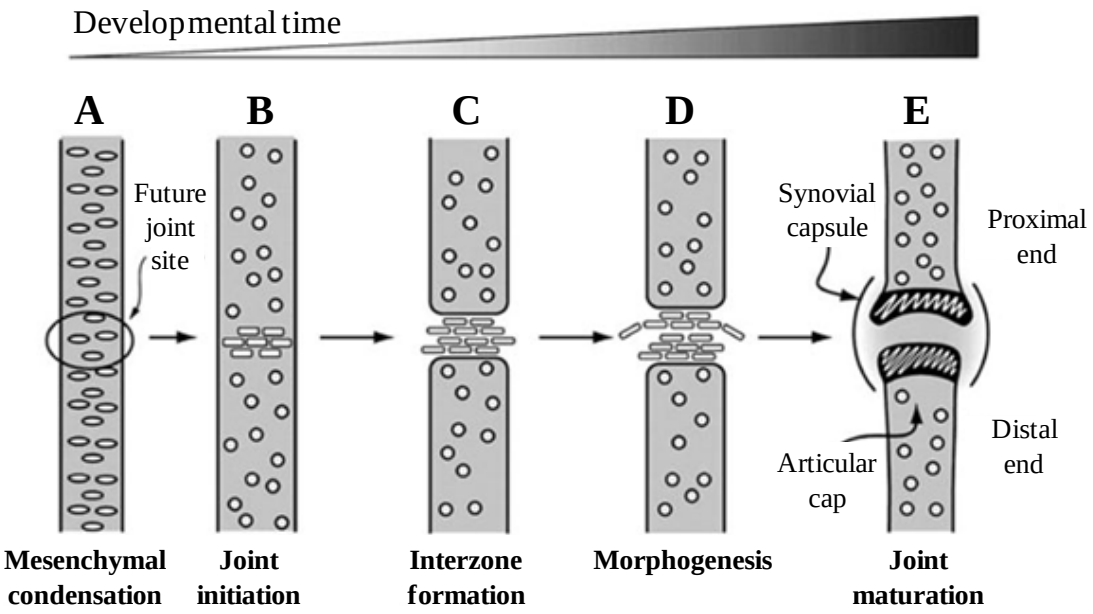

Figure 1.4. Schematic representation of the formation of a synovial joint. A) A mesenchymal cell condensation forms at the site of bone formation and B) the site of a joint is indicated by the alignment of flat mesenchymal cells. C) An interzone forms where the formation of a joint will occur, and a synovial cavity is also created here. D) A morphogenic process is initiated that leads to the generation of a definite joint structure before the E) maturation of the individual elements form a final joint. Adapted from (Pacifici *et al.*, 2006).

1.5.3 *Endochondral ossification*

Concurrent with the development of a joint from the mesenchymal cell condensation, the bones of a synovial joint are formed by endochondral ossification. Long bones can be divided into three zones: the diaphysis (the narrow central shaft), the metaphyses (the wider neck portions) and the epiphyses (the rounded ends). The growth plate is a layer of cartilage found within the metaphyses and acts to separate the diaphysis and epiphyses, such that the two

regions have separate centres of endochondral ossification (Mackie *et al.*, 2008). This allows the uninterrupted development of long bones through diaphysis elongation without affecting the formation of the joints.

Endochondral ossification begins by the differentiation of mesenchymal cells into chondrocytes, known as chondrogenesis, in order to generate a cartilage template (Figure 1.5). In the core, the characteristic markers of cartilage turnover are expressed, such as type II collagen and aggrecan (Hojo *et al.*, 2010), while the cells at the periphery differentiate into perichondrial cells before forming a bone collar of osteoblasts. The chondrocytes in the core, or primary ossification centre, cease proliferation and thereby become hypertrophic, synthesising type X collagen and mineralising the surrounding matrix. The invasion of a vasculature system is stimulated by the expression of 'markers' such as *VEGF* (vascular endothelial growth factor), as the chondrocytes undergo apoptosis and the cartilaginous matrix is degraded by chondroclasts. Bone matrix is deposited in place of the cartilage template, to form the precursor to trabecular bone, following the carryover of osteoblasts by the vasculature system from the bone collar. Overall, a similar process occurs at the secondary ossification centre within the epiphyses. The growth plate is essential in the longitudinal growth of long bones, remaining a place of transient chondrocyte proliferation, hypertrophy, mineralisation and bone deposition until early adulthood (Baron, 2000).

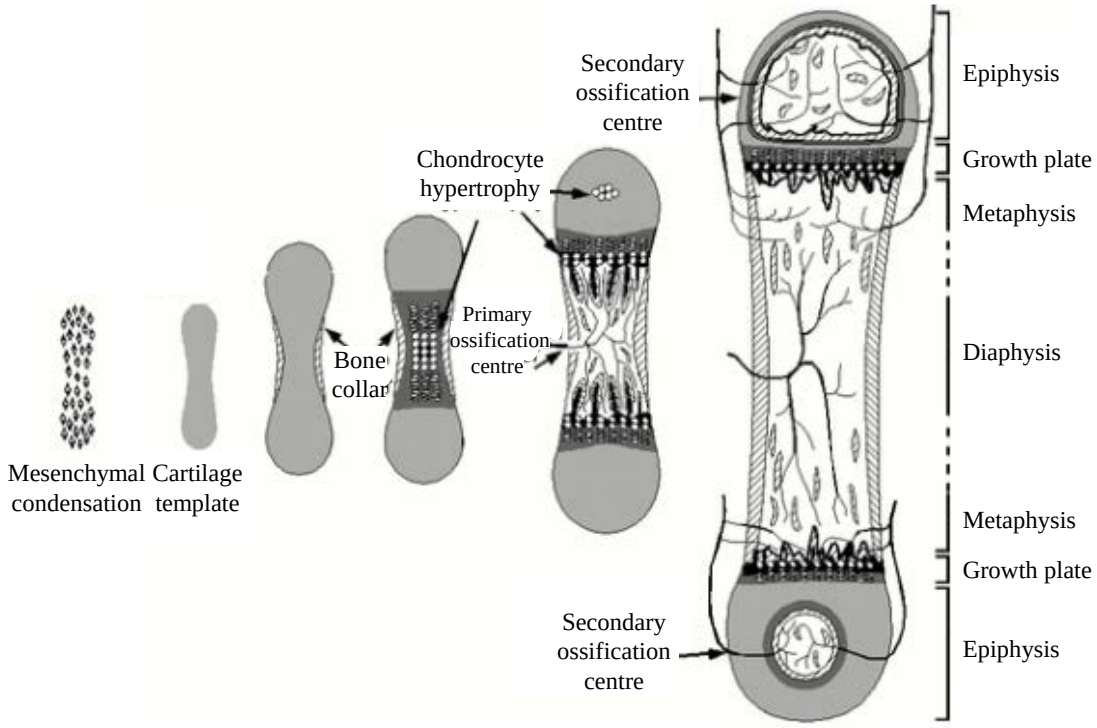

Figure 1.5. The stages of endochondral ossification. From a mesenchymal cell condensation, a cartilage template is formed. A bone collar is generated around hypertrophic chondrocytes at the primary ossification centre. A vasculature system is established and the cartilage template is mineralised before degradation and bone matrix deposition. A similar sequence occurs at the secondary ossification centre at the epiphyses. The bone continues to grow longitudinally at the growth plate until early adulthood. Adapted from (Baron, 2000).

1.6 Pathology of Osteoarthritis

1.6.1 *Introduction to the pathology of osteoarthritis*

Although the classic hallmark of OA is the loss of articular cartilage, it is in fact considered a disease of the entire joint as multiple tissues can be affected (Poole, 2012). The stress that initiates disease onset can be through one of two mechanisms: either there is abnormal loading on a normal joint structure, or there is normal loading on an abnormal joint structure. This causes cartilage degradation and subchondral bone sclerosis, however it is unclear as to which occurs first in an osteoarthritic joint (Bailey and Mansell, 1997). Subsequent biochemical and biomechanical changes then cause joint homeostasis to be unstable and thus affect other tissues (Samuels *et al.*, 2008). The pathological changes that occur are dependent upon the tissue that is affected and upon the stage of disease.

1.6.2 *Changes of articular cartilage and subchondral bone in osteoarthritis*

Healthy articular cartilage is pearly white and extremely smooth, however when damaged, macroscopic changes include cartilage softening, yellowing, surface irregularities, fibrillations and later, advanced fissures (Buckwalter *et al.*, 2005). In addition, the tidemark that normally

marks the calcification boundary appears to be multiplied, while vascular invasion is also common (Bonde *et al.*, 2005). Advancement of the tidemark indicates increased calcification and overall thinning of articular cartilage, although the mechanisms behind this are not fully understood (Goldring and Goldring, 2010). Eventually, cartilage degradation is so advanced that the underlying, exposed subchondral bone becomes sclerotic and eburnated.

It is the imbalance of anabolic and catabolic factors that cause the pathological changes of OA. Initial attempts are made to recapitulate the healthy homoeostatic balance, where the usually quiescent chondrocytes respond to cartilage damage by proliferating. This causes areas of cell clustering contrasted with regions of sparsely populated cells. Additionally, the synthesis of ECM proteins is increased in response to cartilage degradation. Ultimately, the anabolic activity cannot be maintained and so catabolism predominates, which leads to the further tissue destruction observed in advanced OA (Goldring and Goldring, 2010). Established catabolic degradation of the articular cartilage is associated with increased expression of aggrecanases and collagenases, such as *MMP1*, *MMP13*, *ADAMTS4* and *ADAMTS5* (Murphy and Nagase, 2008). The degradation and disorganisation of key ECM components cause a decrease in hydrostatic pressure and tensile strength, leading to the loss of cartilage integrity.

There are two aspects of subchondral bone changes during OA: the first is initiated prior to disease onset, and the second occurs in response to disease progression. Subchondral bone sclerosis occurs in response to abnormal stress placed on the joint, and can in some cases precede detectable changes in articular cartilage. Abnormal stress induces an attempt to remodel the cancellous bone structure resulting in an increase in trabeculae number and thickness, and therefore bone volume (Buckland-Wright, 2004). In addition, the subchondral bone plate thickens such that the subchondral bone loses its specific biomechanical integrity (Goldring and Goldring, 2010). Associated with disease progression is the alteration of underlying bone architecture. For example, osteophytes occur at joint margins, which are bony outgrowths covered with a fibrocartilage cap (Junker *et al.*, 2015). It has been reported that the presence of large osteophytes in knee OA confers no risk to disease progression (Felson *et al.*, 2005), with their purpose purported to be to stabilise the osteoarthritic joint (Goldring and Goldring, 2010). Finally, subchondral bone cysts are cavities that are found in the advanced stages of the disease. Cysts are found specifically in regions of uneven, excessive loading caused by damaged overlying articular cartilage (Ondrouch, 1963).

1.6.3 *Changes of other synovial joint tissues*

The products of cartilage breakdown, such as proteoglycan and type II collagen fragments, are able to diffuse into the synovial fluid. Their presence can elicit a mild immune response by synoviocytes, which is secondary to the disease. Catabolic factors such as MMPs and cytokines are secreted, including IL-1 and TNF (Scanzello and Goldring, 2012), which in turn promote the synthesis of further catabolic factors and cartilage destruction. Changes in the synovium include synovitis, thickening, increased vasculature and inflammatory cell (B cell and T cell) infiltration (Wenham and Conaghan, 2010). Structural changes may be apparent in the menisci, such as tears or maceration, and can precede or follow the development of knee OA (Englund *et al.*, 2009). Furthermore, the joint capsule may become calcified, shortened and less pliable, while the surrounding muscle becomes fibrotic (Lloyd-Roberts, 1953) and the ligaments thicken (Tan *et al.*, 2006).

1.7 Risk Factors for Osteoarthritis

1.7.1 *Introduction to factors that affect disease susceptibility, predisposition and progression*

OA is a multifactorial disorder with the relative significance of each element dependent on confounding factors, disease state and joint site (Zhang and Jordan, 2010). Risk factors can be systemic such as age, ethnicity, sex, nutrition, genetic and hormone; or local such as injury, deformity, obesity, occupation and muscle weakness (Felson *et al.*, 2000).

1.7.2 *Age*

The greatest risk factor for the development of OA is age, however, this is an association rather than a definitive inevitability. Mechanical changes that occur with age can affect disease predisposition, including altered proprioception and gait. However, biochemical changes within the ECM are the pivotal factors in the age-related tendency to develop OA. In accordance with increasing age, the structural organisation of the ECM is disrupted, which is associated with the number and size of aggrecan molecules diminishing. Moreover, advanced glycation end (AGE) products accumulate and are responsible for the increased abundance of cross-linked type II collagen fibrils, causing the articular cartilage to become stiffer (Martel-Pelletier *et al.*, 2008). Chondrocytes become senescent, have a diminished response to anabolic factors and ultimately lose their ability to maintain the ECM homeostasis (Loeser, 2009). Although age has a strong association with OA development, other factors are involved. Indeed, the observed association of disease prevalence with age may in part be a manifestation of the cumulative effects of other risk factors.

1.7.3 *Sex-specific differences*

In addition to an increased prevalence relative to males, post-menopausal females have a more severe OA phenotype (Srikanth *et al.*, 2005). An obvious assumption would be that this correlation can be explained by hormonal differences. However, the effect of oestrogen on OA development and the efficacy of post-menopausal oestrogen replacement therapy is widely debated (Wluka *et al.*, 2000). As viewed using MRI scans, females have consistently thinner articular cartilage than males (Maleki-Fischbach and Jordan, 2010), which could be the origin of the sex-specific differences although the mechanisms are not yet understood.

1.7.4 *Environmental factors*

Obesity is a considerable risk factor in the development of OA, highlighted by the NICE guidelines which strongly recommend weight loss for a first-line management of the disease. The implications of obesity are two-fold: an increased strain exerted on the load-bearing joints forces a responsive change in the joint architecture, while metabolic changes may account for enhanced ECM destruction. There is a strong correlation between obesity and knee OA (Sowers and Karvonen-Gutierrez, 2010), while there is moderate evidence for an association with hip OA (Lievense *et al.*, 2002). Increased body mass index (BMI) is also strongly correlated with the onset of knee OA (Blagojevic *et al.*, 2010), providing further evidence for an association between excessive load and OA onset. However, obesity has also been positively correlated with an increased risk of developing hand OA, which implies that a metabolic mechanism could be in effect (Carman *et al.*, 1994). Such an influence could be modulated by local obesity-associated gene expression, for example leptin. An increased expression of leptin is known to correlate with obesity, and the gene has also been shown to be expressed by cells of various synovial joint tissues including adipocytes and chondrocytes, and thus could trigger a cytokine signalling cascade that leads to the degradation of articular cartilage (Dumond *et al.*, 2003). Naturally, an unbalanced, unhealthy diet contributes to the progression of obesity and therefore indirectly increases OA susceptibility. To date, it remains unclear as to whether specific dietary factors, such as vitamin C, D, E and K, can influence disease progression more directly (Zhang and Jordan, 2010).

Certain occupations exert greater and more repetitive physical strains on the joints than that seen as a result of general joint use and movement. For example, male farmers after ten or more years of agricultural work were more likely to suffer from severe hip OA (classified as requiring hip replacement or with a joint space of less than or equal to 1.50 mm), as were individuals who stood for long periods of time or who routinely performed heavy lifting

(Croft *et al.*, 1992). In addition, local joint damage resulting from sporting injuries or accidents can predispose an individual to OA. Damage to the joint ligament, for example a common injury to the anterior cruciate ligament sustained by athletes (Beynnon *et al.*, 2005), has been shown to precede knee OA (Lohmander *et al.*, 2007).

1.8 Genetics of Osteoarthritis

1.8.1 *Introduction to the genetics of osteoarthritis*

OA is a multifactorial, heterogeneous disease with a complex, non-Mendelian pattern of inheritance. Many small-effect loci contribute to overall disease susceptibility, either individually or in combination with other variants (Figure 1.6). The heritability of OA is estimated to be at least 50% (Spector and MacGregor, 2004), meaning that around half of the variance in the occurrence of OA within the population can be attributed to genetics. There are several different approaches to understand the contribution of genetics to disease susceptibility including family studies, candidate gene approaches and genome-wide association scans (GWAS).

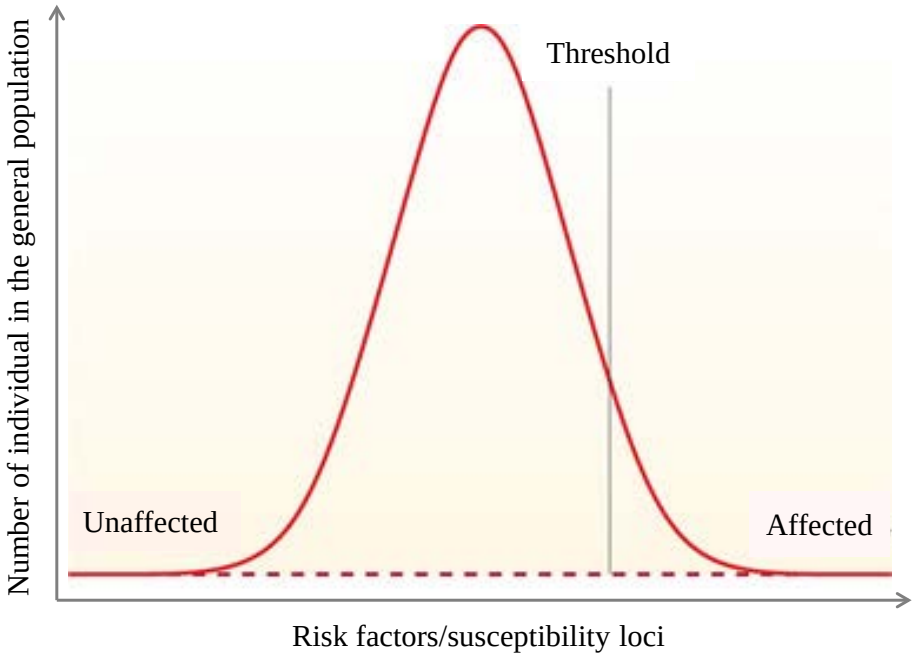

Figure 1.6. The liability threshold model. In the general population, the majority of individuals harbour an average number of susceptibility loci for osteoarthritis and have an average number of factors that confer osteoarthritis risk. If this number reaches a critical threshold, either solely susceptibility loci or in combination with risk factors, the individual will be affected. Adapted from (van Dongen *et al.*, 2012).

1.8.2 *Family and twin studies*

The first report to associate genetics with OA was from a family study published in 1941, in which it was found that Heberden's nodes, a common sign of hand OA, occurred more frequently in mothers and sisters of affected women compared to the general population (Stecher, 1941). Subsequent studies in the UK continued in the same vein (Kellgren *et al.*, 1963), confirming that OA of various synovial joints had a genetic component (Fernandez-Moreno *et al.*, 2008). Family studies are a good approach to investigate the contribution of genetics and environment to a disease, exploring if related individuals have disproportionate likelihoods of being affected by OA relative to the general population. However, this is sensitive to the natural differences between family members, including physical activity and age (Spector and MacGregor, 2004).

To circumvent such differences, twin studies are an alternative approach to estimate heritability, with a main advantage being that it corrects for age. Twin studies compare the disease concordance in monozygotic and dizygotic twins, where monozygotic twins share 100% of their genes and dizygotic twins share an average of 50% of their genes. If, for example, there was incomplete concordance in monozygotic twins, it could be concluded that environmental factors were also contributing to disease susceptibility. In a classic twin study, it was estimated that the heritability of female hand and knee OA was 39% – 65% (Spector *et al.*, 1996). Heritability of hip OA is somewhat more difficult to estimate given the lower prevalence of the disease. Nevertheless, joint space narrowing of female hip OA was shown to be influenced by genetics, with heritability estimated to be around 60% (Spector *et al.*, 1996). Finally, disc degeneration of the spine was assessed by MRI in 172 and 154 monozygotic and dizygotic twins, respectively, estimating a genetic influence of 73% – 74% (Sambrook *et al.*, 1999). A summary of the heritability estimates from various twin studies is detailed in Figure 1.7.

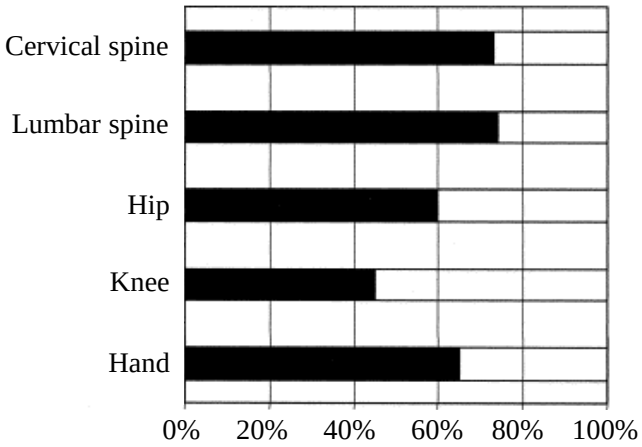

Figure 1.7. Estimates of osteoarthritis heritability at different joint sites from twin studies. Heritability of spine osteoarthritis is the highest at around 70%, while around half of the variation in knee osteoarthritis is due to genetics. Taken from (Spector and MacGregor, 2004).

1.8.3 *Candidate gene studies*

Candidate gene studies are a hypothesis-led approach aimed to identify the specific genes that account for the heritability of a trait. Genes are selected based on the expression profile, protein function or association with other similar traits. In the case of OA, candidate genes may encode structural components of the ECM, be involved in joint development, or have a role in joint homeostasis.

An obvious candidate gene is *COL2A1*, which encodes the α1 polypeptide chain of type II collagen. The *VDR* (vitamin D receptor) gene is similarly encoded in this region, 68 kb downstream of *COL2A1*, and so the genes are often interrogated together. Polymorphisms within *VDR* have been implicated in variations of bone mineral density (Lambrinoudaki *et al.*, 2011), while *COL2A1* mutations have been linked with skeletal phenotypes seen in disorders such as spondyloepiphyseal dysplasia congenita (Donahue *et al.*, 2003). However, various studies of this region have not yielded any conclusive association with OA (Loughlin *et al.*, 1994; Aerssens *et al.*, 1998).

The most promising candidate gene to date is *GDF5*. Mutations in this gene have been linked to skeletal disorders such as brachydactyly type C and chondrodysplasias (Farooq *et al.*, 2013). The protein encoded by the gene is a member of the TGF-β superfamily and is pivotal in the determination and differentiation of joint sites. Based on its role in skeletogenesis, its association with OA was examined. It was found that in an Asian population, the 5′ untranslated region (UTR) single nucleotide polymorphism (SNP) rs143383 was associated

with hip and knee OA (Miyamoto *et al.*, 2007). Subsequent luciferase reporter assays implied an influence of the SNP on the expression of *GDF5*, with the associated allele (T) corresponding to a reduced transcriptional activity (Miyamoto *et al.*, 2007). Furthermore, in a case-control study based on UK and Spanish populations, a reduction in the expression of the *GDF5* risk allele transcript relative to the minor (C) allele was observed (Southam *et al.*, 2007). Additional functional studies have continued to expand the understanding of the genetic basis of *GDF5* and its OA association (Egli *et al.*, 2009; Dodd *et al.*, 2013; Syddall *et al.*, 2013). Despite this success, candidate studies generally have limited power as they rely on current knowledge of functional associations and thus may omit more significant associations. As such, an alternative approach, in the form of a hypothesis-free methodology, is required.

1.8.4 *Linkage studies*

The aim of a linkage study is to interrogate the genomes of individuals within a family to identify if any regions co-segregate with the disease phenotype. Due to the age-association of OA, however, studying affected parents and their offspring, or affected individuals and their parents, is often impossible. Therefore, affected sibling pairs are more widely utilised (Loughlin *et al.*, 2002a).

The results of an early OA linkage study identified the female-specific association of a region on chromosome 11q with the disease (Chapman *et al.*, 1999), and linkage studies have since continued to similarly identify other associations with OA. In fact, the associations of twelve chromosomes (1, 2, 4, 6, 7, 9, 11, 12, 13, 16, 19 and X) with OA have been reported (Fernandez-Moreno *et al.*, 2008). Although this has not directly resulted in the identification of OA-associated genes, in-depth probing of the linkage regions have yielded promising results. For example, a genome-wide linkage scan followed by fine-mapping of the region revealed chromosome 6p12.3–q13 as being linked to female hip OA (Loughlin *et al.*, 1999; Loughlin *et al.*, 2002b). A subsequent analysis of the region presented evidence to suggest that the gene coding for BMP-5, a TGF-β superfamily member that is involved in modulating chondrocyte activity (Bailon-Plaza *et al.*, 1999), could be associated with female hip OA (Southam *et al.*, 2004).

1.8.5 *Genome-wide association scans*

A powerful advancement in the identification of genetic risk loci has been the development of GWAS, an unbiased, population-based methodology that uses large sample sizes and involves independent replication (Reynard and Loughlin, 2013). A GWAS allows for the examination

of up to 1 million polymorphisms across the genome of thousands of unrelated individuals (Bush and Moore, 2012), commonly using microarray platforms such as Affymetrix (Santa Clara, CA, USA) and Illumina (San Diego, CA, USA). The aim of a GWAS is to identify alleles of the genotyped SNPs that occur more frequently in a case group compared to a control group.

The human genome contains approximately 10 million variants with a minor allele frequency (MAF) of at least 1%, which is an average of 1 SNP per 300 bases (International HapMap Consortium, 2003). This means that physically, only a fraction of the SNPs are captured by a GWAS. However, the lawn of hybridised deoxyribonucleic acid (DNA) probes on the solid-surface microarray chip are specifically selected to target polymorphisms that allows for a much greater coverage, by exploiting the redundancy of SNPs across the genome (International HapMap Consortium, 2005). This so-called redundancy is caused by a co-inheritance of alleles, known as linkage disequilibrium (LD), between which there would be little or no genetic recombination. LD is a measure of the association between alleles at two given loci; there are two forms of this value that are generally reported. D' (ranges from 0.00 to 1.00) considers the recombination rate only, such that complete LD ($D' = 1.00$) denotes no recombination between the two loci. r^2 (ranges from 0.00 to 1.00) considers both the recombination rate and the allelic frequencies, such that perfect LD ($D' = 1.00$, $r^2 = 1.00$) denotes identical allelic frequencies of the two loci between which there is no recombination. Therefore, in the case of perfect LD, the genotype of additional SNPs can be inferred from the genotype of a SNP that is captured on the array (Figure 1.8). Generally, a positive identification of association between an allele and the disease is distinguished by a genome-wide significance threshold of $p \leq 5.00 \times 10^{-8}$, which accounts for the multiple comparisons performed, although this is not an absolute value in all studies (Panagiotou *et al.*, 2012). In addition, as GWAS signals can oftentimes be nominally significant but not so after accounting for the multiple tests performed, a replication phase is a necessary component in order to corroborate the initial findings of the discovery cohort. Following the identification of an association between an allele and the disease, the region must be interrogated to identify the causal SNP and the mechanisms of its action.

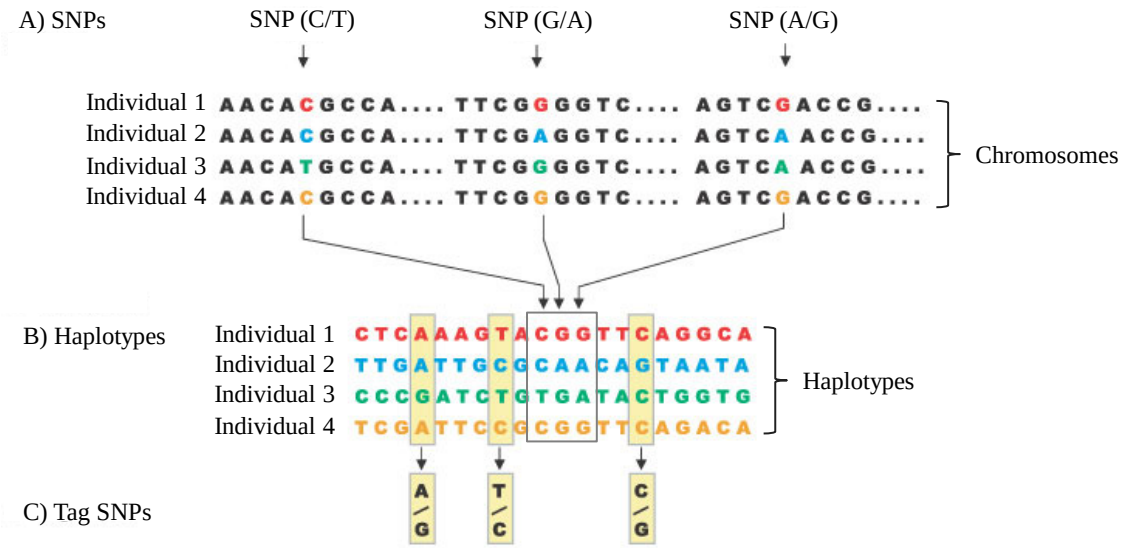

Figure 1.8. Example of the expansion of microarray coverage by exploiting the co-inheritance of alleles. A) The three SNPs of interest are highlighted on the chromosomes of four unrelated individuals. B) Between SNPs where there is no recombination, haplotypes are generated; that is, a particular set of polymorphisms that are always inherited together. C) Genotyping specific polymorphisms, known as 'tag SNPs', allows the genotypes of all the SNPs in this haplotype to be identified. For example, if the chromosome pattern of the tag SNP is G-T-C, this matches the haplotype of individual 3. Although humans are diploid organisms, for simplicity, only one chromosome is shown; SNP (single nucleotide polymorphism). Adapted from (International HapMap Consortium, 2003).

To date, four separate GWAS have been reported that aimed to identify OA susceptibility loci. The first genotyped over 500,000 SNPs in a total of 14,938 cases and 39,000 controls of a Dutch population (Kerkhof *et al.*, 2010). The study identified the C allele of rs3815148 as conferring a 1.14-fold increased risk of knee and/or hand OA ($p = 8.00 \times 10^{-8}$). The SNP is located within an intron of *COG5* (component of oligomeric Golgi complex 5) on chromosome 7q22, with *PRKAR2B* (protein kinase, cAMP-dependent, regulatory type II β), *HBP1* (HMG-box transcription factor 1), *GPR22* (G protein-coupled receptor 22), *DUS4L* (dihydrouridine synthase 4-like) and *BCAP29* (B-cell receptor-associated protein 29) also mapping to this locus. A subsequent meta-analysis recognised this region as being associated with knee OA (Evangelou *et al.*, 2011), although the predominating SNP was rs4730250 ($p = 9.20 \times 10^{-9}$). rs470250 resides within an intron of *DUS4L*, and is separated from rs3815148 by 269 kb with an r^2 of 0.82 and a D' of 1.00. Characterisation of this locus revealed that carriage of the OA-associated allele conferred a significant reduction of *HBP1* expression in cartilage and synovium, and of *DUS4L* in fat pad (Raine *et al.*, 2012).

The results of the second OA GWAS were published one month later, and this reported on the genotyping of approximately 450,000 SNPs in a total of 1,073 cases and 3,743 controls of a Japanese population (Nakajima *et al.*, 2010). The study found rs7775228 ($p = 2.43 \times 10^{-8}$) and rs10947262 ($p = 6.73 \times 10^{-8}$) to be associated with knee OA. The SNPs are located on chromosome 6p21, with rs7775228 intergenic upstream of *HLA-DQB1* (major histocompatibility complex, class II, DQ β1) and rs10947262 within an intron of *BTNL2* (butyrophilin-like 2). Also at this locus are *HLA-DRA* (major histocompatibility complex, class II, DR α), *HLA-DRB5* (major histocompatibility complex, class II, DR β5), *HLA-DRB1* (major histocompatibility complex, class II, DR β1), and *HLA-DQA1* (major histocompatibility complex, class II, DQ α1). The associations of the SNPs with knee OA were not replicated in an independent European cohort (Valdes *et al.*, 2011).

The largest OA GWAS was a two-stage study performed by the Arthritis Research UK Osteoarthritis Genetics (arcOGEN) Consortium. Each individual in the case groups had radiographic evidence of OA, and over 80% of those had also undergone hip or knee replacement surgery. The first stage of the study involved a total of 13,702 cases (3,177 of which were in the discovery cohort) and 53,286 controls (4,894 of which were in the discovery cohort) of a population of European descent (Panoutsopoulou *et al.*, 2011). The interim analysis reported that there were no SNPs associated with OA with genome-wide significance: the strongest signal for knee and/or hip OA was rs2277831 ($p = 2.30 \times 10^{-5}$), an intronic SNP within *MICAL3* (microtubule associated monooxygenase, calponin and LIM domain containing 3) on chromosome 22q11. When stratified by joint site, the strongest signal for knee OA was rs11280 within *Corf130* ($p = 3.20 \times 10^{-5}$) on chromosome 6p21, while the strongest for hip OA was rs2615977 in an intron of *COL11A1* (collagen, type XI, α1; $p = 1.10 \times 10^{-5}$) on chromosome 1p21. Similarly, a meta-analysis highlighted two independent signals within *COL11A1* that were associated with hip OA (Rodriguez-Fontenla *et al.*, 2014). Nevertheless, characterisation of the loci revealed that there was no correlation between gene expression at either the chromosome 22q11 locus and rs2277831 genotype (Ratnayake *et al.*, 2012), or the 1p21 locus and rs2615977 genotype (Raine *et al.*, 2013).

When searching for small-effect loci, false-negative results may be generated as a result of a lack of power. The second stage of the arcOGEN study with a much larger discovery cohort was implemented to overcome this (arcOGEN Consortium *et al.*, 2012). This involved a total of 14,883 cases (7,410 of which were in the discovery cohort) and 53,947 controls (11,009 of which were in the discovery cohort). Overall, the arcOGEN study analysed over 1.40 million

SNPs, which were either directly genotyped or imputed. Five regions were identified as being associated with OA with genome-wide significance, and a further three almost achieved genome-wide significance ($p \leq 5.00 \times 10^{-8}$; Table 1.3). rs11177, a missense variant within an exon of *GNL3* (guanine nucleotide binding protein-like 3), and rs6976, a 3′ UTR SNP within *GLT8D1* (glycosyltransferase 8 domain containing 1), were the strongest signals. Both SNPs are located at chromosome 3p21.1 and are in perfect LD with each other ($r^2 = 1.00$), meaning they represent the same OA-associated locus. The signal was associated with severe hip and knee OA, as defined by a requirement for total joint replacement surgery. The remaining genome-wide significant loci were associated with hip OA, with rs4836732 in an intron of *ASTN2* (astrotactin 2) at chromosome 9q33 being female-specific. The propensity of signals to emerge only following stratification by sex and joint site highlights the multifactorial nature of the disease.

SNP	Locus	Stratum	Nearest gene(s)	*p* value	OR (95% CI)
rs6976*	3p21.1	TJR	*GLT8D1*	7.24×10^{-11}	1.12 (1.06-1.12)
rs11177*	3p21.1	TJR	*GNL3*	1.25×10^{-10}	1.12 (1.08-1.16)
rs4836732	9q33.1	THR-F	*ASTN2*	6.11×10^{-10}	1.20 (1.13-1.27)
rs9350591	6q14.1	Hip	*FILIP1, SENP6*	2.42×10^{-9}	1.18 (1.12-1.25)
rs10492367	12p11.22	Hip	*PTHLH, KLHL42*	1.48×10^{-8}	1.14 (1.09-1.20)
rs835487	12q23.3	THR	*CHST11*	1.64×10^{-8}	1.13 (1.09-1.18)
rs12107036	3q28	TKR-F	*TP63*	6.71×10^{-8}	1.21 (1.13-1.29)
rs8044769	16q12.2	F	*FTO*	6.85×10^{-8}	1.11 (1.07-1.15)
rs10948172	6p21.1	M	*SUPT3H, CDC5L*	7.92×10^{-8}	1.14 (1.09-1.20)

Table 1.3. Loci significantly associated with osteoarthritis as identified by the arcOGEN study (arcOGEN Consortium *et al.*, 2012). Five regions were significantly associated with osteoarthritis and a further three were just below the threshold for genome-wide significance ($p \leq 5.00 \times 10^{-8}$). Data represent the combined discovery and replication analyses. * same signal, $r^2 = 1.00$. TJR (total joint replacement; severe end-stage of the disease), THR (total hip replacement; severe end-stage of the disease), TKR (total knee replacement; severe end-stage of the disease), F (female), M (male); hip (all cases of hip osteoarthritis; THR and radiographic evidence combined); OR (odds ratio); CI (confidence interval).

As a consequence of the arcOGEN Consortium findings, several studies have begun to functionally dissect the reported association signals. For example, it has been shown that carriage of the OA risk alleles of rs11177 and rs6976 correlate with a decrease in *GNL3* and *SPCS1* expression (Gee *et al.*, 2014). Another documented characterisation is that of rs835487, which is in an intron of *CHST11* (carbohydrate [chondroitin 4] sulfotransferase 11) on chromosome 12q23. *CHST11* catalyses the transfer of sulphate in the generation of chondroitin sulphate in the cartilage ECM. In relation to the OA signal, differential binding of

the transcription factors SP1 (Sp1 transcription factor), SP3 (Sp3 transcription factor) and SUB1 (SUB1 homolog [*S. cerevisiae*]) between the alleles of rs835487 has been shown (Reynard *et al.*, 2014), which could directly implicate the polymorphism in the OA susceptibility of the region.

The fourth and most recent OA GWAS was performed using an Icelandic discovery cohort of 623 cases of hand OA and 69,153 population-based controls (Styrkarsdottir *et al.*, 2014). Following replication in cases and controls of European ancestry, two SNPs emerged as being significantly associated with hand OA. rs4238326 is located in an intron of *ALDH1A2* (aldehyde dehydrogenase 1 family, member A2) on chromosome 15q21 ($p = 8.60 \times 10^{-11}$) and rs3204689 is a synonymous variant in an exon of the same gene ($p = 1.11 \times 10^{-11}$).

Overall, if large, unrelated cohorts can be obtained, GWAS provide an unbiased, hypothesis-free approach to identifying risk loci: to date, several loci have been associated with OA (Reynard and Loughlin, 2013). However, the polygenic nature of the disease remains a limiting factor in the subsequent functional analyses and characterisations of the regions. It is apparent that stratification of data by ethnicity, joint site, disease stage and sex are often required in identifying risk alleles, and this will no doubt be vital in further understanding the genetic association of OA in the future.

1.9 *Cis*-Acting Polymorphisms

1.9.1 *Expression quantitative trait loci (eQTL)*

A SNP that emerges as significantly associated with a disease phenotype from a GWAS study marks a region of association in which a number of other polymorphisms could reside. Typically, an arbitrary r^2 of 0.80 or greater is assigned as high LD between two SNPs (Dai *et al.*, 2013), and these polymorphisms are thus considered potential causal variants. Once the region has been narrowed down and the causal SNP identified, the mechanisms behind the association must be dissected. The polymorphism could be non-synonymous, causing an amino acid change that alters the encoded protein structure or function; the variant could be synonymous but leads to aberrant transcription or splicing; or the SNP could be non-coding, either intronic or intergenic, and act as a regulatory polymorphism to modulate gene expression. When the expression of a gene is regulated by the genomic locus, the SNP is referred to as an expression quantitative trait locus (eQTL), and if the genomic locus maps in close proximity to the regulated gene, it is known as a *cis*-eQTL (Franke and Jansen, 2009).

The understanding of how SNPs operate as regulators of nearby gene expression is ever expanding. The functionality of a specific polymorphism can traditionally be attributed to differential transcription factor binding between the alleles, which causes altered enhancer or promoter activity (Miyamoto *et al.*, 2007; Syddall *et al.*, 2013). Alternatively, the differential protein binding could instead control the conformation of the overall three-dimensional structure of the chromatin and in turn affect the binding of transcription factors to a nearby regulatory element (Gaffney *et al.*, 2012). Moreover, overall gene expression levels can be influenced through the regulation of transcript stability (Cheadle *et al.*, 2005). Ultimately, the mechanisms result in the regulation of target gene expression. A relatively recent development in the understanding of how *cis*-eQTLs function is the investigation into the role of methylation as an intermediate step between a *cis*-eQTL and gene expression. DNA methylation is mediated by DNA methyltransferases (DNMT) and it is commonly found within regulatory elements such as the transcription start sites of genes. Potentially, genetic variation at *cis*-eQTLs could regulate the binding and functioning of DNMT enzymes at the CpG sites thereby altering methylation. As such, the differential methylation may then regulate the target gene expression (Bell *et al.*, 2011). Overall, there are several known mechanisms by which disease association can be modulated, the dissection of which requires a thorough investigatory approach. For the purpose of my Ph.D, such mechanisms will be investigated for two of the arcOGEN OA association signals, rs9350591 and rs10492367.

1.10 Genes Surrounding rs9350591 on Chromosome 6q14.1

1.10.1 *Introduction to rs9350591*

The rs9350591 C to T SNP marks a region of the genome that is significantly associated with hip OA, as identified by the arcOGEN study (arcOGEN Consortium *et al.*, 2012) and detailed in Table 1.3. rs9350591 is an intergenic SNP 38 kb upstream of *FILIP1* (filamin A interacting protein 1) and 70 kb upstream of *SENP6* (SUMO1/sentrin-specific peptidase 6). A further five genes reside within 1 Mb upstream or downstream of the polymorphism (Table 1.4 and Figure 1.9). In addition, rs9350591 is upstream of two uncharacterised long non-coding ribonucleic acids (RNAs), *LOC100506804* and *LOC101928540*; and is within 103 kb of a predicted micro RNA (miRNA), *MIR4463*. As there are no amino acid changes that could account for the association signal, I postulate that rs9350591 marks a SNP in high LD with it that in some way regulates the expression of one of these genes on chromosome 6q.

Gene symbol	Gene name	Distance from rs9350591
FILIP1	Filamin A interacting protein 1	38 kb downstream
SENP6	SUMO1/sentrin-specific peptidase 6	70 kb downstream
MYO6	Myosin VI	217 kb downstream
TMEM30A	Transmembrane protein 30A	247 kb downstream
COX7A2	Cytochrome c oxidase subunit VIIa polypeptide 2	288 kb downstream
COL12A1	Collagen, type XII, α1	325 kb downstream
IMPG1	Interphotoreceptor matrix proteoglycan 1	390 kb upstream

Table 1.4. Genes within 1 Mb upstream and 1 Mb downstream of rs9350591. Seven genes reside in this region, with *FILIP1* and *SENP6* being the two genes in closest physical proximity to the SNP.

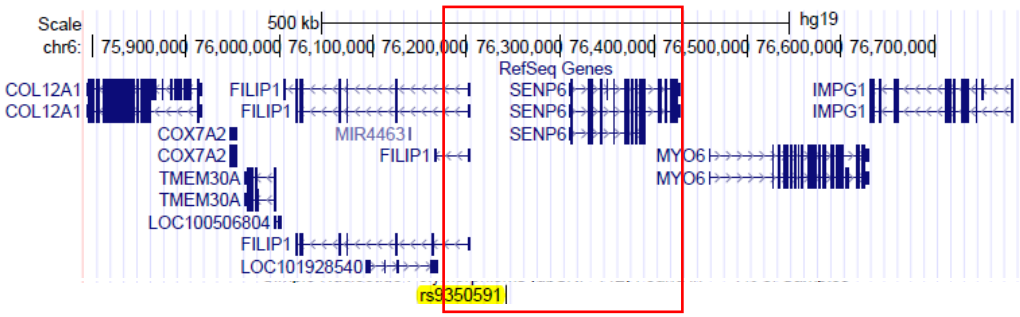

Figure 1.9. UCSC Genome Browser screenshot of the osteoarthritis association region marked by the polymorphism rs9350591 on chromosome 6q14.1. rs9350591 is an intergenic polymorphism upstream of *FILIP1*, *SENP6*, *MYO6*, *TMEM30A*, *COX7A2*, *COL12A1* and downstream of *IMPG1*. The red box marks the boundaries of the association interval; all SNPs with an $r^2 \geq 0.8$ relative to rs9350591 reside in this region.

1.10.2 *Filamin A interacting protein 1 (FILIP1)*

Filamin is an actin binding protein that is present in the cytosol of non-muscle tissue, with filamin A being one of three isoforms (Feng and Walsh, 2004). Filamins are primarily involved in bridging cortical actin into an overall three-dimensional cytoskeletal structure, aiding cell structure and motility. As the nomenclature suggests, the protein encoded by *FILIP1* interacts with filamin A, and could therefore be involved in its functions. Indeed, evidence supports this, with the encoded protein, FILIP1, functioning downstream of a master regulator of actin and cytoskeletal dynamics, RhoD (Nussinov, 2013).

To date, the known functions of FILIP1 are rather limited to the research initially performed by a laboratory in Japan. This group have linked FILIP1 to cell motility in the neocortex of Wistar rats, proposing that FILIP1 mediates the initiation of cell migration in the brain causing the degradation of filamin A and thus inhibiting its interaction with actin (Nagano *et*

al., 2002). In the fibroblast-like COS-7 cells, exogenous FILIP1 did suppress ventricular cell motility, and in addition, they reported an increase in *FILIP1* expression in *in situ* pre-migratory ventricular cells. Further to this, an established role of *FILIP1* and filamin A in corticogenesis has since been supported, showing their requirement in maintaining cell polarity during migration (Nagano *et al.*, 2004; Sato and Nagano, 2005). As a potential link to its corticogenesis function, deletions in the 6q14 region have been associated with intellectual disability (Becker *et al.*, 2012). This may prove to be a vital link to translate *FILIP1* expression into a role in disease pathology. However, the deletions are not uniform throughout affected individuals, and therefore *FILIP1* remains only one of several candidate genes in the region.

Although the literature primarily focusses on the role of *FILIP1* in cortical cell migration without any associated phenotypes, the known pathologies of filamin A are diverse and are not limited to brain tissue. It is therefore worthy of investigation in order to dissect the potential significance of *FILIP1*, however this of course is speculation only. Indeed, mutations within filamin A have been reported in a skeletal dysplasia called otopalatodigital syndrome (Clark *et al.*, 2009), where it seems that a gain-of-function mutation leads to an enhanced affinity for actin. However, it is currently unclear how this translates into the characteristic features of the disorder, which include limited joint flexibility, wide-set eyes and a small, flat nose. This is in contrast to neuronal disorders associated with filamin A loss-of-function mutations such as periventricular heterotopia. In this disorder, displaying the hallmark feature of seizures with an onset in young adults, a mutation in filamin A hinders neuronal cell migration (Fox *et al.*, 1998). Overall, there have been no reports on the function of *FILIP1* in skeletal development, although current knowledge is limited. In combination with a consideration of filamin A function, further investigations may one day confirm that *FILIP1* has a more diverse role than is currently understood.

1.10.3 *SUMO1/sentrin-specific peptidase 6 (SENP6)*

Small ubiquitin-like modifier (SUMO) proteins can covalently attach to target proteins in a fashion analogous to ubiquitination. In contrast to the often common degradative fate of ubiquitin-bound proteins however, sumoylation directs localisation and stability of the SUMO-bound protein (Geiss-Friedlander and Melchior, 2007). *SENP6* encodes a protein that is responsible for modulating the activity of sumoylation, and its function is two-fold. Firstly, SENP6 can mature SUMO proteins by cleaving the C terminal tail in order to process the precursor protein; and secondly, SENP6 can deconjugate SUMO proteins from their targets

by hydrolysing the isopetidic bond (Alegre and Reverter, 2011). Studies indicate that SENP6 has a preferred specificity for the poly- SUMO2 and SUMO3 isoform chains (Lima and Reverter, 2008). Moreover, there is evidence to suggest that SENP6 is involved in kinetochore dynamics. It has been reported that a depletion of the protein is associated with defective spindle morphology and metaphase progression (Mukhopadhyay *et al.*, 2010).

Sumoylation has been linked to a number of human diseases, although a role of SENP6 is yet to be associated with these. For example, sumoylation was shown to be required for Myc-dependent tumorigenesis (Kessler *et al.*, 2012), which is in contrast to its loss as part of a *MITF* (microphthalmia-associated transcription factor) loss-of-function mutation that is associated with renal cell carcinoma and sporadic melanoma (Bertolotto *et al.*, 2011). In addition, aberrant sumoylation of the ATP2A2a (ATPase Ca^{++} transporting cardiac muscle slow twitch 2a) protein is associated with heart failure (Kho *et al.*, 2011). With a link to skeletal development, the T/G polymorphism rs9360921 (r^2 = 0.81 relative to rs9350591), approximately 46 kb upstream of *SENP6,* has been associated with height (Lango Allen *et al.*, 2010). Interestingly, there is an emerging role of sumoylation in arthritis. *SENP3* overexpression in primary chondrocytes, for instance, has been reported to be associated with an increased expression of the catabolic factors *ADAMTS-4* and *MMP-13* (Yan *et al.*, 2010). However, the association with OA and cartilage has not extended specifically to *SENP6* and therefore remains speculative.

1.10.4 *Myosin VI (MYO6)*

MYO6 encodes the unconventional actin-based myosin VI motor protein, whereby the functional protein does not form filaments (Kalhammer and Bahler, 2000). The genetic structure is such that exons 1 to 20 encode the head domain, exons 21 to 23 encode the neck domain and exons 24 to 32 encode the tail domain (Ahituv *et al.*, 2000). Exon 30 can be alternatively spliced to generate two isoforms, both of which have been detected in brain tissue; the isoform without exon 30 is widely expressed throughout other tissues such as skeletal muscle, cardiac muscle and the cochlea (Avraham *et al.*, 1997). A putative casein kinase II phosphorylation site is within this exon, which could prove significant in the different functions of the isoforms. Within each of the three domains exist specific elements that allow myosin VI to perform its unique function (Figure 1.10). Firstly, the head domain contains an adenosine triphosphate (ATP) binding domain corresponding to exon 5, which is necessary for ATP hydrolysis, in addition to an actin binding domain corresponding to exon 19, which is needed for the cytoskeletal interactions. Also found here is a threonine residue

that has the potential to be phosphorylated. The neck region contains a converter domain and an isoleucine [I] glutamine [Q] (IQ) motif: the converter allows for the movement towards the minus end of the actin filament (Wells *et al.*, 1999), while the IQ motif presents a sequence for calmodulin binding. Finally, the tail region contains a coiled-coil domain to modulate dimerisation and a unique globular domain that is important for localisation (Frank *et al.*, 2004). The cellular function of myosin VI is primarily as a motor to carry vesicles and proteins along an actin track, and is reported to be involved in cell migration (Geisbrecht and Montell, 2002), endocytosis (Buss *et al.*, 2001) and the modulation of RNA polymerase II-mediated gene transcription (Vreugde *et al.*, 2006). For these purposes, it is localised to the Golgi complex, endocytic vesicles, the cytosol and membrane ruffles (Buss *et al.*, 1998).

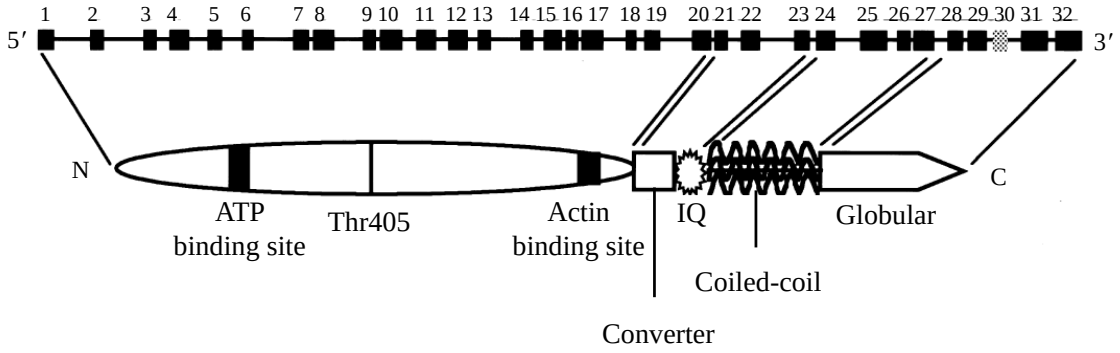

Figure 1.10. Gene and protein structure of *MYO6*. The gene comprises 32 exons, with exon 30 being alternatively spliced in brain tissue. The head domain is encoded by exons 1 to 20, the neck domain is encoded by exons 21 to 23, and the tail domain is encoded by exons 24 to 32. The functional MYO6 protein contains an ATP binding site, a threonine residue for potential phosphorylation, an actin binding site, a converter domain for movement towards the minus end of actin, an IQ (isoleucine [I] glutamine [Q]) domain for calmodulin binding, and a tail containing a coiled-coil and a globular domain. Adapted from (Ahituv *et al.*, 2000).

The expression of *MYO6* in the inner ear makes it an ideal candidate for a functional role in auditory disorders: indeed, mutations of *MYO6* are typically associated with deafness (Ahmed *et al.*, 2003; Sanggaard *et al.*, 2008). Moreover, the homozygous *Snell's waltzer* (*sv*) mouse displays deafness and spinning caused by a null mutation of *MYO6* (Avraham *et al.*, 1995). Given the nature of the protein function, *MYO6* may have wide-reaching implications in cellular physiology and disease pathology. This is exemplified by the diverse functionality of myosin VI in the *sv* mouse, where it has been shown that the protein is required for the correct tethering of the intestinal epithelial cell brush-border membrane (Hegan *et al.*, 2012). In addition, elevated expression of *MYO6* has been associated with cellular alterations observed in prostate cancer (Wei *et al.*, 2008), and its overexpression in high-grade ovarian carcinoma

has been shown to accelerate cell migration and subsequent tumour dissemination (Yoshida *et al.*, 2004). Such a fundamental role of *MYO6* in cellular physiology and its diverse disease pathologies mean it may prove to be relevant in OA development; however, as of yet this has not been established.

1.10.5 *Transmembrane protein 30A (TMEM30A)*

TMEM30A, also known as *CDC50A* (cell division cycle 50A), encodes the membrane-spanning β-subunit of the P4 family of adenosine triphosphatase (ATPase) flippase heterodimers (Katoh and Katoh, 2004; Folmer *et al.*, 2012). The purpose of the flippase complex, by coupling the process with ATP hydrolysis, is to maintain a bilayer asymmetry by translocating phospholipids from the exoplasmic domain to the cytosolic domain of the cell membrane (Paulusma and Oude Elferink, 2005). A knockdown of *TMEM30A* confirmed this role by showing a reduction in the level of phospholipid import (Chen *et al.*, 2011). TMEM30A is the most abundantly expressed out of the three family members (Folmer *et al.*, 2012; van der Mark *et al.*, 2013), and can interact with several of the fourteen members of the P4-ATPase catalytic domain family (van der Velden *et al.*, 2010), named from ATP8A1 through to ATP11C (Sebastian *et al.*, 2012).

The range of interactions between the ATPase and TMEM30 proteins therefore suggest a wider functional role of TMEM30A than solely the maintenance of bilayer asymmetry. For example, studies of the interaction between ATP8A1 and TMEM30A suggest that the complex is necessary for cell migration. A depletion of either subunit resulted in diminished cell motility, and furthermore, its translocation of phosphatidylethanolamine was necessary for correct ruffle formation – a factor required to promote cellular migration (Kato *et al.*, 2013). In addition, ATP8A2 has been shown to act synergistically with TMEM30A to regulate neurite outgrowth of rat hippocampal neurons (Xu *et al.*, 2012a).

The role of the transmembrane protein complex has been exploited by pharmaceutical companies. *TMEM30A* expression was upregulated in primary prostate cancer (Romanuik *et al.*, 2009) making this an ideal target as a route of entry for anti-cancer therapies. For example, the functional unit created by TMEM30A and an as of yet unidentified P4-ATPase aids the uptake of the alkyl-phospholipid anti-cancer drug perifosine (Munoz-Martinez *et al.*, 2010), which functions to inhibit the action of Akt. However, no overall benefit in survival was reported in patients with metastatic colorectal cancer (Bendell *et al.*, 2012); in addition, the drug did not significantly alter the course of multiple myeloma and so resulted in the

discontinuation of its phase III clinical trial (Aeterna Zentaris Inc., 2013). Edelfosine is a pro-apoptotic anti-cancer drug, and its uptake is similarly mediated by TMEM30A (van Blitterswijk and Verheij, 2008; Chen *et al.*, 2011).

There is limited knowledge pertaining to mutations within *TMEM30A*, however those of the ATPases could be suggestive of a functional role of TMEM30A. A mouse knockout of *ATP8A1* results in impaired hippocampus-dependent learning (Levano *et al.*, 2012), while mutations in *ATP8B1* are associated with cholestasis, characterised by a decrease in bile flow to the duodenum (Klomp *et al.*, 2004). It seems that *TMEM30A* expression has the potential to mediate a range of cellular functions, however, its physiological functions do not currently extend further than its use in cancer therapy.

1.10.6 *Cytochrome c oxidase subunit VIIa polypeptide 2 (COX7A2)*

Cytochrome c oxidase, an integral membrane protein, is the terminal enzyme of the electron transport chain of cellular respiration (Van Beeumen *et al.*, 1990). Through the coupling of the electron transport chain to proton transfer within the mitochondria, energy in the form of ATP is generated. Cytochrome c oxidase is composed of nuclear- and mitochondrial- encoded subunits, which are thought to function as structural and catalytic components, respectively. COX7A2 is a nuclear-encoded subunit of the cytochrome c oxidase complex (Little *et al.*, 2010). This isoform is expressed in all tissue types, while polypeptide 1 is specific to cardiac and skeletal muscle (Arnaudo *et al.*, 1992).

A number of studies have implicated *COX7A2* in various physiological presentations. For example, as testosterone production in aged men was downregulated, *COX7A2* expression was found to be upregulated (Xin *et al.*, 2003). Further investigations have revealed that COX7A2 negatively affects steroidogenesis in murine Leydig cells, potentially acting through an increase in reactive oxygen species production to inhibit testosterone production (Chen *et al.*, 2006b). Using data obtained from a Diabetes Genetic Initiative GWAS, the G allele of rs1323070 (r^2 = 0.00 relative to rs9350591), approximately 8 kb upstream of *COX7A2*, was identified as being associated with a decrease in glucose-stimulated insulin secretion (Olsson *et al.*, 2011). Given the importance of ATP production in the regulation of this process, it is reasonable to suggest *COX7A2* is involved in the mechanism through which this association occurs, and therefore could be implicated in type II diabetes. Finally, using matrix-assisted laser desorption/ionisation imaging (MALDI) mass spectrometry, a decrease in *COX7A2*

expression has been reported to be a prognostic marker in Barrett's adenocarcinoma (Elsner *et al.*, 2012).

Perhaps the requirement of ATP in the chondrocyte-mediated maintenance of the ECM (Wolff *et al.*, 2013) and the production of ATP in the electron transport chain may implicate *COX7A2* in OA susceptibility (Martin *et al.*, 2012).

1.10.7 *Collagen, type XII, α1 (COL12A1)*

COL12A1 encodes a homotrimeric protein of α1-chain polypeptides that is classified as a fibril-associated collagen with interrupted triple helices (FACIT) collagen. As illustrated in Figure 1.11, the structure is such that it has two triple-helical collagenous (COL1 and COL2) domains interspaced by non-collagenous, non-triple helical (NC1, NC2 and NC3) domains and a thrombospondin N-terminal-like (TSPN) region. The NC3 domain is globular and has repeating von Willebrand factor A-like (vWA) and fibronectin type III (FN3) domains (Bader *et al.*, 2009). It is the NC3 domain that differs by approximately 100 kDa between the two splice variants of collagen type XII, long (XIIA) and short (XIIB), which are known to exist in mammals (Bohme *et al.*, 1995). Type XII collagen has been identified in cartilage (Watt *et al.*, 1992), tendon (Dublet *et al.*, 1989), bone and ligament (Oh *et al.*, 1993). It is understood to localise with type I collagen (Walchli *et al.*, 1994), perhaps to act as a bridge or anchor for the collagen fibrils, however its exact function remains unclear (Gerecke *et al.*, 1997).

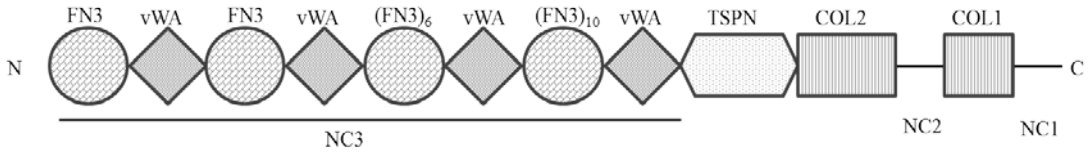

Figure 1.11. Structure of the type XII collagen α1 polypeptide. Two collagenous domains (COL1 and COL2) are interrupted by three non-collagenous domains (NC1, NC2 and NC3). A TSPN domain separates COL2 and NC3. The globular NC3 is composed of repeating vWA domains and FN3 repeats. Adapted from (Ricard-Blum and Ruggiero, 2005).

Type XII collagen could be a marker for cancer progression as an upregulation of *COL12A1* was observed in colon cancer cells (Karagiannis *et al.*, 2012). However, given the expression of various collagens in the bone and ECM of synovial joints, it is reasonable to predict a more relevant role for type XII collagen somewhere during skeletal development or function. In wild type mice, type XII collagen is secreted by osteoblasts in regions of bone formation; knockout mice, on the other hand, have disorganised osteoblasts which generate lower levels

of bone matrix deposition and thus display a short stature phenotype with skeletal abnormalities (Izu *et al.*, 2011). Moreover, an upregulation of *COL12A1* in a murine osteoblast cell line was identified in response to stretch stress, however, this was not replicated by the group in other murine cell line cultures (Arai *et al.*, 2008). In humans, a loss-of-function mutation within *COL12A1* presents symptoms including joint hyperlaxity, which is reminiscent of connective tissue disorders such as Bethlem myopathy and Ehlers-Danlos syndrome (Zou *et al.*, 2014). Finally, the A/G SNP rs970547 (r^2 = 0.00 relative to rs9350591) within *COL12A1* was significantly associated with anterior cruciate ligament ruptures in females (Posthumus *et al.*, 2010). These studies imply that *COL12A1* is involved in skeletogenesis, trauma response and the overall correct functioning of synovial joints, which could therefore perhaps extend to a role in OA.

1.10.8 *Interphotoreceptor matrix proteoglycan 1 (IMPG1)*

IMPG1 encodes a glycoprotein component of the interphotoreceptor matrix of retinal rod and cone cells (Felbor *et al.*, 1998; Lee *et al.*, 2000). The protein helps to maintain the homeostasis of the extracellular matrix surrounding photoreceptors, and perhaps functions as a scaffold component to retain the environmental integrity (Acharya *et al.*, 1998a). Additionally, evidence suggests that IMPG1 could be integral in the adhesion of the neural retina to the retinal pigmented epithelium (Kuehn and Hageman, 1999). In accordance with its expression in the interphotoreceptor matrix, *IMPG1* has been linked to a role in human retinopathies. For example, some forms of vitelliform macular dystrophy, characterised by deteriorating central vision related to an accumulation of lipofuscin, have been reported to be associated with mutations within *IMPG1* (Manes *et al.*, 2013). Furthermore, genetic linkage analyses have mapped retinal dystrophies, such as North Carolina macular dystrophy, to this region (Small *et al.*, 1992); however it has since been shown that there is no apparent correlation (Gehrig *et al.*, 1998).

Naturally, the fact that *IMPG1* encodes an ECM proteoglycan and that it possesses a functional hyaluronan binding domain (Acharya *et al.*, 1998b) advocates its potential role in articular cartilage ECM homeostasis. However, the interphotoreceptor matrix is unique from the well-documented cartilage ECM (Ishikawa *et al.*, 2015), and thus *IMPG1* is unlikely to emerge as a functional proteoglycan here. In addition, data from Prof. Loughlin's group (Institute of Cellular Medicine, Newcastle University) have shown that gene expression was not detectable in articular cartilage, fat pad, meniscus, synovium, tendon, ligament or osteophyte (arcOGEN Consortium *et al.*, 2012).

1.11 Genes Surrounding rs10492367 on Chromosome 12p11.22

1.11.1 *Introduction to rs10492367*

The rs1049267 G to T SNP marks a region of the genome that is significantly associated with hip OA, as identified by the arcOGEN study (arcOGEN Consortium *et al.*, 2012) and detailed in Table 1.3. rs10492367 is an intergenic SNP 59 kb downstream of *KLHL42* (kelch-like family member 42) and 96 kb downstream of *PTHLH* (parathyroid hormone-like hormone). A further thirteen genes reside within a 1 Mb span upstream or downstream of the polymorphism (Table 1.5 and Figure 1.12). In addition, rs10492367 is 415 kb upstream of the long non-coding RNA *ARNTL2-AS1* (ARNTL2 antisense RNA 1). As there are no amino acid changes that could account for the association signal, I postulate that rs10492367 marks a SNP in high LD with it that in some way regulates the expression of one of these genes on chromosome 12p.

Gene symbol	Gene name	Distance from rs9350591
KLHL42	Kelch-like family member 42	59 kb downstream
PTHLH	Parathyroid hormone-like hormone	96 kb downstream
ASUN	Asunder spermatogenesis regulator	924 kb downstream
FGFR1OP2	FGFR1 oncogene partner 2	901 kb upstream
TM7SF3	Transmembrane 7 superfamily member 3	848 kb downstream
MED21	Mediator complex subunit 21	831 kb upstream
C12orf71	Chromosome 12 open reading frame 71	780 kb downstream
STK38L	Serine/threonine kinase 38 like	536 kb upstream
ARNTL2	Aryl hydrocarbon receptor nuclear translocator-like 2	436 kb upstream
SMCO2	Single-pass membrane protein with coiled-coil domains 2	360 kb upstream
PPFIBP1	PTPRF interacting protein, binding protein 1 [liprin β1]	166 kb upstream
REP15	RAB15 effector protein	164 kb upstream
MRPS35	Mitochondrial ribosomal protein S35	106 kb upstream
MANSC4	MANSC domain containing 4	91 kb downstream
CCDC91	Coiled-coil domain containing 91	40 kb downstream

Table 1.5. Genes within 1 Mb upstream and 1 Mb downstream of rs10492367. Fifteen genes resided at this locus, with *PTHLH* and *KLHL42* being the two genes in closest physical proximity to the SNP.

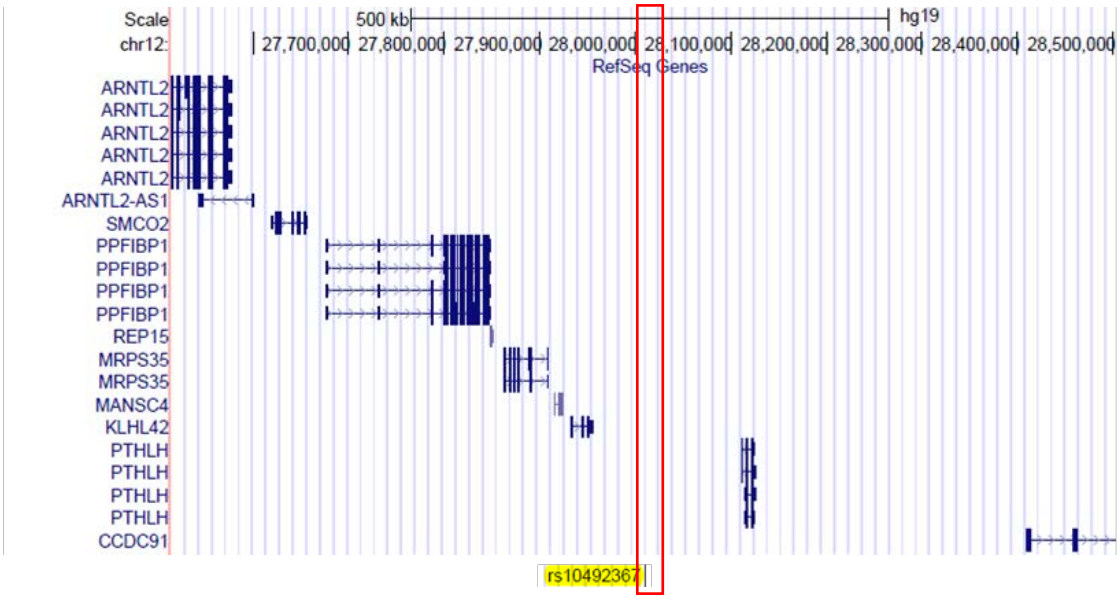

Figure 1.12. UCSC Genome Browser screenshot of the OA association region marked by the polymorphism rs10492367 on chromosome 12p11.22. rs10492367 is an intergenic polymorphism downstream of *PTHLH* and *KLHL42*. The red box marks the boundaries of the association interval; all SNPs with an $r^2 \geq 0.8$ relative to rs10492367 reside in this region.

1.11.2 *Parathyroid hormone-like hormone (PTHLH)*

PTHLH encodes the protein parathyroid hormone-related protein (PTHrP), which was first discovered nearly three decades ago (Burtis *et al.*, 1987). The protein was purified from a tumour of a patient with breast carcinoma and humoral hypercalcaemia of malignancy (HMM), and it was shown to be characteristically comparable to parathyroid hormone (PTH). Increased levels of secreted PTHrP by malignant tumours cause excessive bone resorption and therefore hypercalcaemia, mimicking the innate effect of PTH, which is a key regulator of calcium homeostasis (Nakajima *et al.*, 2013). In fact, eight of the first 13 amino acids of the N-terminal domain are identical between PTH and PTHrP (Wysolmerski, 2012), meaning their shared structural homology allows for a shared G protein-coupled receptor, parathyroid hormone type 1 receptor (PTH1R).

The *PTHLH* gene is orientated on the reverse strand of genomic DNA (gDNA) and comprises nine exons with three different promoters. It is the variation in promoter use, in addition to alternative splicing, which generates three distinct isoforms of the prohormone: PTHrP 139, PTHrP 141 and PTHrP 173 (Figure 1.13). The intervening intron between exons 5 and 6 is spliced in all variants, and this is the basis for PTHrP 139. Alternatively, additional spicing events can draw together this region to the coding sequence of exon 8 for PTHrP 173, or the coding sequence of exon 9 to create PTHrP 141. Consequently, these resulting isoforms have

largely similar sequences, and differ mainly in the 3′-most region of the molecule. The significance of possessing three different isoforms is not clearly understood, although it is known that the expression of PTHrP 173 is human-specific (Sellers *et al.*, 2004).

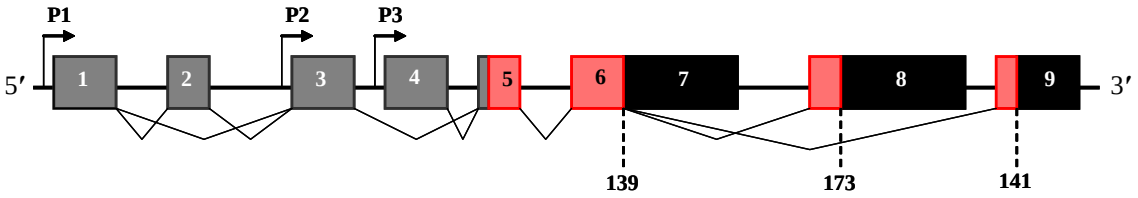

Figure 1.13. Genetic structure of *PTHLH*. The gene comprises nine exons, of which there are four coding regions. The use of three different promoters (P1, P2 and P3) and alternative splicing of the mRNA transcript result in three initial translation products of 139, 141 or 173 amino acids (dotted lines). Grey boxes represent the 5′ UTR and black boxes represent the 3′ UTR. Red boxes denote the coding exons. Adapted from (Bouizar *et al.*, 1999).

Post-translational modifications mature the secretory prohormones into various proteins (Figure 1.14), which have distinct physiological roles mainly in paracrine signalling pathways. The primary protein forms are PTHrP (1-36), which shares sequence homology with the N-terminal of PTH (Mundy and Edwards, 2008); PTHrP (38-94), which has a known role in the activation of intracellular calcium transport, particularly through the placenta from the maternal to the foetal circulation (Kovacs *et al.*, 1996); and PTHrP (107-139), which has been shown to inhibit bone resorption and stimulate osteoblast function (Cornish *et al.*, 1997; Cornish *et al.*, 1999). In general, however, the individual biological relevance of the separate regions are not fully understood.

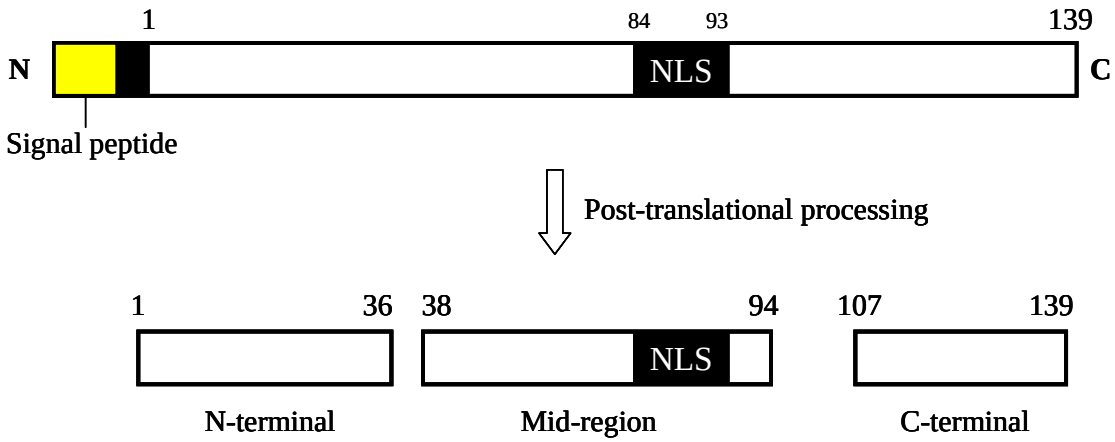

Figure 1.14. Post-translational proteolytic processing of the PTHrP 139 prohormone. NLS (nuclear localisation signal). Adapted from (Atlas of Genetics and Cytogenetics in Oncology and Haematology, 2015).

In addition to the secretory forms, PTHrP also has an intracrine role. In some cases, translation may be initiated downstream to the usual start site which causes the signal peptide sequence to be truncated (Wysolmerski, 2012). This means the prohormone remains in the cytosol, localises to the cell nucleus on the strength of its nuclear localisation signal (NLS), and therefore bypasses direction into endoplasmic reticulum for secretion (Nguyen *et al.*, 2001). The exact position of the NLS is debated, although definitions are generally within the mid-region peptide between amino acids 88 and 106 (Fiaschi-Taesch and Stewart, 2003). The function of nuclear PTHrP is unclear, however it has been suggested that it acts to regulate the transcription of ribosomal genes (Nguyen *et al.*, 2001).

The secreted form of PTHrP functions during endochondral ossification to promote long bone growth before skeletal maturity and the closure of the growth plate (Jiang *et al.*, 2008). Prehypertrophic chondrocytes secrete Indian hedgehog (IHH) which signals to undifferentiated chondrocytes at the end of long bones to stimulate PTHrP secretion (St-Jacques *et al.*, 1999). This then signals to prehypertrophic chondrocytes (Figure 1.15), which express the cognate receptor PTH1R in order to inhibit their hypertrophic differentiation (Hirai *et al.*, 2011). By way of a negative feedback loop, IHH secretion is consequently reduced given the maintenance of a proliferative state, causing PTHrP secretion to decrease and thus stimulate hypertrophy (Vortkamp *et al.*, 1996; Deng *et al.*, 2008). Indeed, PTHrP-deficient mice have reduced chondrocyte proliferation and die at birth (Miao *et al.*, 2002).

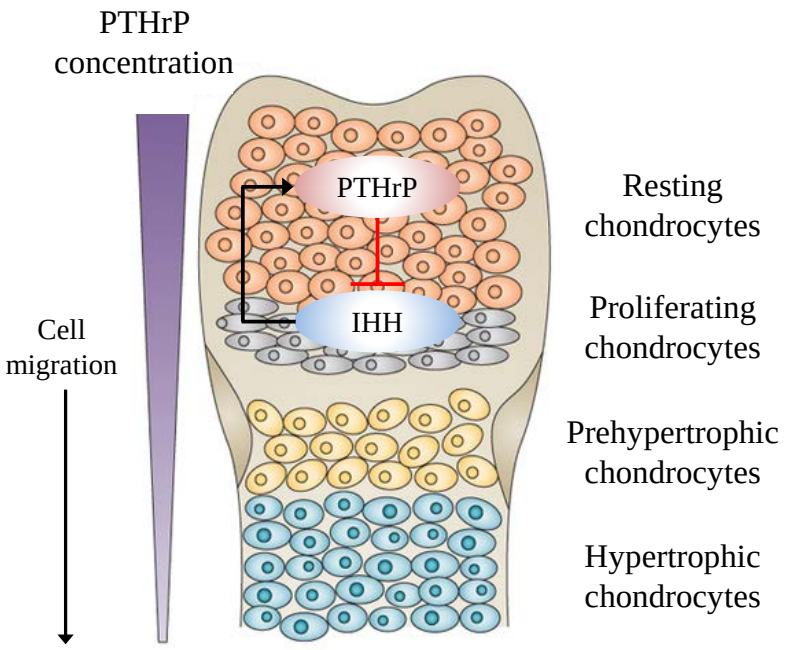

Figure 1.15. The negative feedback loop between PTHrP and IHH in the regulation of growth plate chondrocyte differentiation. The role of PTHrP is to keep chondrocytes proliferating. PTHrP is secreted by cells at the end of long bones which acts to inhibit IHH, a key driver of chondrocyte hypertrophy. IHH inhibition signals to decrease PTHrP secretion and therefore lead to hypertrophy. Adapted from (Alman, 2015) .

Within articular cartilage, reports of the expression of *PTHLH* are often conflicting. For example, it has been demonstrated that cellular interactions, mediated by PTHrP, between superficial zone and deep zone cells regulate chondrocyte mineralisation in a co-culture (Jiang *et al.*, 2008). Perhaps this control is lost following articular cartilage degradation, further perpetuating the development of OA. In addition, the authors describe low *PTHLH* expression in the superficial zone cell culture. This is largely in contrast to an observed increase in *PTHLH* expression in articular chondrocytes relative to the osteophytic chondrocytes used to recapitulate the transient phenotype of those destined to be mineralised during endochondral ossification (Gelse *et al.*, 2012). In terms of OA, an upregulation of PTHrP has been reported in OA knee cartilage relative to the non-OA controls (Terkeltaub *et al.*, 1998). In addition to *PTHLH* expression in skeletal tissues, the gene is also known to be expressed in a variety of cell types including endothelial cells, smooth muscle cells and those of the cardiovascular system (Liu *et al.*, 2011).

The administration of exogenous PTH and PTHrP induce entirely different biological effects (Esbrit and Alcaraz, 2013). Continuous administration of PTH leads to bone resorption, while

continuous administration of PTHrP favours bone formation (Lippuner, 2012). The biochemical function of PTHrP has been exploited by the biopharmaceutical industry, with subcutaneous injections of a synthetic analogue of the N-terminal domain (PTHrP [1-34]) being trialled for the treatment of osteoporosis. Osteoporosis is characterised by weak and fragile bones caused by an imbalance of bone formation and bone resorption. The therapy, known as Abaloparatide or BA058 (Radius Health Inc., Cambridge, MA, USA), has been reported to increase the bone mineral density of the lumbar spine, femoral neck and the total hip of post-menopausal women in a randomised placebo-controlled trial (Leder *et al.*, 2015).

To support the evidence of a non-redundant role of PTHrP signalling in the correct development of a skeletal phenotype, GWAS have identified polymorphisms within or nearby *PTHLH* to be associated with bone mineral density (Estrada *et al.*, 2012) and height (Lango Allen *et al.*, 2010). Moreover, Blomstrand chondrodysplasia, caused by a loss-of-function mutation of PTH1R, is characterised by advanced endochondral ossification resulting in a short-limbed stature, premature birth and neonatal death (Jobert *et al.*, 1998). Jansen's metaphyseal chondrodysplasia, on the other hand, is caused by a ligand-independent activating mutation of PTH1R and presents with short stature, joint swelling and hypercalcaemia (Mannstadt *et al.*, 1999). The opposing nature of the mutations causing similar phenotypes suggests that a fine balance of PTHrP signalling is required for correct long bone development. Finally, an association of Brachydactyly Type E, a disorder displaying characteristic short digits, with the balanced translocation t(8;12)(q13;p11.22) has been reported and correlates with a decreased *PTHLH* expression (Maass *et al.*, 2010).

The diverse nature of the gene structure, expression and the regulatory roles of its encoded protein make understanding *PTHLH* extremely complex. Overall, however, the role of PTHrP in endochondral ossification and the associated skeletal phenotypes makes *PTHLH* an ideal candidate gene for OA susceptibility.

1.11.3 *Kelch-like family member 42 (KLHL42)*

Very little is currently known about *KLHL42* or the function of its encoded protein. The member of the kelch superfamily has recently been reclassified from *KLHDC5* (kelch domain containing 5) to *KLHL42* based on emerging information with regards to its structure (Figure 1.16). Standard *KLHL* family members have a bric à brac, tamtrack and broad-complex (BTB) domain, a BTB and C-terminal kelch (BACK) domain, and five to six kelch domain

repeats (Dhanoa *et al.*, 2013). This is in contrast to *KLHDC* family members that have no BTB or BACK domains.

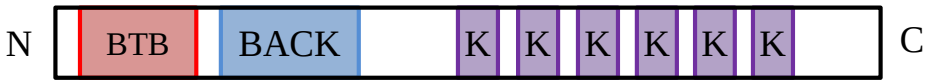

Figure 1.16. Re-annotation of *KLHDC5* has shown the typical domains of a *KLHL* family member. *KLHL42* contains one BTB, one BACK and six kelch domains. Adapted from (Dhanoa *et al.*, 2013).

The domains of *KLHL42* appear to offer unique functionality to the protein. For example, other BTB domain-containing proteins facilitate ubiquitination by forming cullin-based E3 ligase complexes (Furukawa *et al.*, 2003; Pintard *et al.*, 2004). While there is no definitive function for the BACK domain, it is speculated to be necessary for the correct cullin-based E3 ligase complex formation with BTB (Stogios and Prive, 2004). Finally, kelch domains are known to be required for cellular morphology and cytoskeleton organisation through direct and indirect actin binding (von Bulow *et al.*, 1995; Kim *et al.*, 1999; Adams *et al.*, 2000). Each kelch repeat forms a 4-stranded β-sheet, which then, in combination with the surrounding kelch repeats, folds into a conformation that creates a β-propeller structure around a central axis (Adams *et al.*, 2000), as depicted in Figure 1.17. This structure certainly fits the proposed function of KLHL42 based on the individual domain functions: the central core might be a key factor in the binding of the protein to regulatory or cytoskeletal proteins.

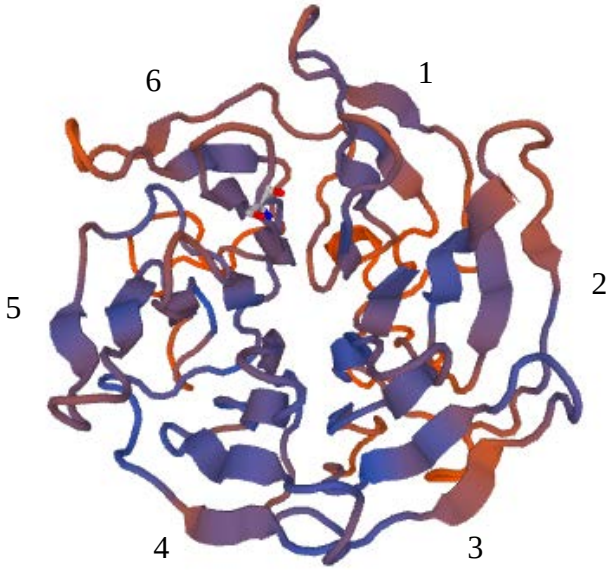

Figure 1.17. Structure of KLHL42. A bird's eye view of the protein clearly shows the six β-sheets created by the individual kelch repeats. The resulting structure is a cone that has a central core, potentially to aid protein binding and interactions. Numbers 1 – 6 highlight the individual 4-stranded β-sheets. Image created using the SWISS-PDB Viewer 4.0.4 (Guex and Peitsch, 1997) available at http://www.expasy.org/spdbv/.

One of the very few publications relating to *KLHL42* was only six years ago, in which through a two-hybrid screen, KLHL42 was shown to interact with cullin-3, a core component of the E3 ligase complex that is involved in target protein ubiquitination and degradation (Cummings *et al.*, 2009). This process occurs during mitosis to sequester and remove the microtubule-severing complex p60/katanin; naturally, overexpression of *KLHL42* resulted in an increased density of microtubules, while protein knockdown led to a loss of the microtubule framework.

Aside from this, there is sparse published data pertaining specifically to *KLHL42* or the function of its protein. In fact, a detailed literature search for "*KLHL42*" using the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/), which includes PubMed and PubMed Central, yielded only three publications. One of these was an update on kelch-like gene family members as previously referenced (Dhanoa *et al.*, 2013), while the remaining two publications did not report on the structure or function of *KLHL42* (Silveira *et al.*, 2014; Ross *et al.*, 2015). A similar database search for "*KLHDC5*" yielded a modest 57 publications, however apart from the research discussed here, none of the papers were primary research publications that advanced the current knowledge of *KLHL42* acting as an adapter for cullin-based E3 ligases. The standout publications include GWAS that identify disease susceptibility loci nearby, but not specifically at, *KLHL42*; for example, type

2 diabetes (Morris *et al.*, 2012), bone mineral density (Kemp *et al.*, 2014), OA (arcOGEN Consortium *et al.*, 2012), and discussions thereof. Overall, there appears to be no evidence to date that may imply a relevant functional role for *KLHL42* in OA susceptibility.

1.11.4 *Other genes within 1Mb upstream or downstream of rs10492367*

Conventionally, *cis*-eQTLs are classified as such if they reside within 1 Mb of the transcription start site of the regulated gene (Nica and Dermitzakis, 2013). Aside from *PTHLH* and *KLHL42*, as already noted a further thirteen genes plus a long non-coding RNA can be found within the 2 Mb window of rs10492367. There is limited published data with regards to these genes.

According to data derived from RNA sequencing (Xu *et al.*, 2012b), the following genes from amongst the thirteen are not, or are very lowly, expressed in OA and non-OA hip cartilage: *ARNTL2*, *SMCO2*, *REP15*, *MANSC4*, *CCDC91* and *C12orf71*. These genes also have limited published knowledge with regards to their function, with no obviously relevant reference to joint structure or function. For example, *ARNTL2* is a core component of the circadian clock, a 24 hour system of oscillations in gene expression that is translated into a metabolic rhythm (Ciarleglio *et al.*, 2008); *REP15* co-localises to the endocytic recycling compartment with RAB15-GTP to facilitate transferrin recycling (Strick and Elferink, 2005); and *CCDC91* is involved in the *trans*-Golgi network of membrane trafficking, and has been implicated in the maintenance of the microstructure of the white matter in the brain (Sprooten *et al.*, 2014).

Amongst the genes that are known to be expressed in cartilage (Xu *et al.*, 2012b), *PPFIBP1* is a liprin (LAR protein-tyrosine phosphatase-interacting protein) family member, also known as liprin β1. Liprins are known to interact with transmembrane PTPRF (protein tyrosine phosphatase, receptor type F; also known as LAR) proteins, which mediate axon guidance (Ensslen-Craig and Brady-Kalnay, 2004), however the involvement of liprin β1 in this process is not well studied. It has been reported that the metastasis-associated protein S100A4 (S100 calcium binding protein A4) binds liprin β1, which could potentially affect cell adhesion through interactions with PTPRF (Kriajevska *et al.*, 2002). Liprin β1 has also been implicated in the regulation of lymphatic vasculature integrity in *Xenopus* tadpoles (Norrmen *et al.*, 2010).

MRPS35 encodes a mitochondrial ribosome protein subunit that aids protein synthesis. Studies have suggested that an overexpression of *MRPS35* is involved in testicular germ cell

tumours (Rodriguez *et al.*, 2003), while a mutation in the gene in *S. cerevisiae* is associated with increased lipid droplet size (Fei *et al.*, 2011) and perhaps, therefore, could be implicated in metabolic syndromes that present the hallmark accumulation of lipid droplets.

FGFR1OP2 has been implicated in the 8p11 myeloproliferative syndrome through its fusion to the *FGFR1* (fibroblast growth factor receptor 1) gene, resulting in the activation of the FGFR1 tyrosine kinase, unregulated haemopoiesis and a subsequent enhanced disease pathogenesis (Grand *et al.*, 2004). In addition, it has been postulated that the gene may promote wound closure through cytoskeletal interactions and enhanced cell migration (Lin *et al.*, 2010).

There is even less known about *ASUN*, *TM7SF3*, *MED21* and *STK38L*. *ASUN* modulates the correct progression through the mitotic cell cycle, particularly prophase, by recruiting dynein to the nuclear surface as the nucleus couples to the centrosome (Jodoin *et al.*, 2012). *TM7SF3*, meanwhile, is a seven-span transmembrane protein (Akashi *et al.*, 2000) that is thought to act as an inhibitor of cytokine-induced pancreatic β cell death, resulting in the rescue of associated diminished insulin secretion (Beck *et al.*, 2011). *MED21* forms a co-activating complex to regulate the expression of RNA polymerase II-controlled genes (Baumli *et al.*, 2005), and implications of its expression have been linked to keratinocyte proliferation and differentiation control (Oda *et al.*, 2010). Finally, *STK38L* is a protein kinase that possesses a diverse range of functions. For example, it is involved in cell cycle progression (Cornils *et al.*, 2011), hippocampal neuron polarisation (Yang *et al.*, 2014) and tumour cell invasion and survival (Suzuki *et al.*, 2006).

Overall, *PTHLH* appears to be the strongest candidate gene for the OA susceptibility of this region. However, the lack of knowledge with regards to the functionality of the remaining genes is not reason enough to discount them entirely.

1.12 Summary

OA is the most common form of arthritis, characterised by articular cartilage degradation at the end of long bones within synovial joints. This presents as joint pain, stiffness and a lack of mobility, with effective treatments currently limited to pain management and total joint replacement.

The disease is multifactorial and polygenic: the primary causes of OA are age, lifestyle, environment, sex and genetics. Various techniques have been implemented to ascertain the extent to which the disorder is caused by genetics. Through the arcOGEN GWAS, several regions of the genome have been associated with OA including chromosome 12p11.22 marked by rs10492367, and chromosome 6q14.1 marked by rs9350591.

PTHLH and *KLHL42* are the genes in closest physical proximity to rs10492367, although there are several others within a 2 Mb span of the SNP. *PTHLH* is the strongest candidate gene for OA association at this locus, and it could transpire that rs10492367 marks a regulatory polymorphism that modulates gene expression. *SENP6* and *FILIP1* are the two genes in closest physical proximity to rs9350591, although again, several other genes reside at this locus, including *COL12A1*, any of which could mediate the OA association of this region.

Functionally characterising the two loci will help dissect the OA association signals, providing an understanding of disease mechanisms and potentially furthering the diagnostic and treatment technologies currently available.

1.13 Overall Aims

The overall aim of this research is to characterise the two OA association signals marked by the polymorphisms rs10492367 and rs9350591.

This will be achieved by investigating:

- gene expression
- DNA methylation
- transcription factor binding and function

Chapter 2. Materials and Methods

2.1 Database Searches to Characterise the Association Signals

Two databases were utilised for the initial characterisations of the association signals: the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway; (Kent *et al.*, 2002)) and RegulomeDB (http://regulomedb.org/; (Boyle *et al.*, 2012)). RegulomeDB annotates intergenic polymorphisms that have known or predicted regulatory activity, while the UCSC Genome Browser acts as a portal through which a vast array of data can be accessed. This includes the NCBI RefSeq collection (Pruitt *et al.*, 2005) of annotated gene transcripts, and data generated by the ENCODE Consortium (Encode Project Consortium, 2012). The ENCODE Project utilises data from 147 cell types, with nine common cell lines capturing eight different tissue types (blood, embryonic stem cell, liver, breast, muscle, blood vessel, skin and lung) and therefore the mesoderm, inner cell mass, endoderm and ectoderm lineages.

2.2 Human Mesenchymal Stem Cell (MSC) Differentiation Down a Chondrogenic Lineage

In vitro chondrogenesis was performed by three members of Prof. Loughlin's research group (Institute of Cellular Medicine, Newcastle University) – Dr Madhushika Ratnayake, Maria Tselepi and Emma Rogers – and by Dr Mathew Barter (Institute of Cellular Medicine, Newcastle University), following a well-established differentiation model for Transwell cultures (Tew *et al.*, 2008). Briefly, MSCs were cultured at 37^{o}C in Dulbecco's Modified Eagle Medium supplemented with 100 IU/ml penicillin, 100 µg/ml streptomycin and 2 mM of L-glutamine (Sigma-Aldrich, UK). The medium also contained: 40 µg/ml proline, 10 ng/ml TGF-β3 (PeproTech, UK), 100 mM dexamethasone (Sigma-Aldrich, UK), 50 µg/ml ascorbic acid-2-phosphate (Sigma-Aldrich, UK), 1 x insulin, transferrin, selenium, linoleic acid premix (ITS+L; BD Biosciences, Oxford, UK) and 4.50 g/l glucose (Lonza, UK). Cells were cultured in Corning Costar 24-well cell culture plates (Sigma-Aldrich, UK) using Millicell 0.40 µm hanging polyethylene terephthalate cell culture inserts (Merck Millipore, UK). RNA was extracted at various time points using TRIzol® Reagent (Life Technologies, UK) following the manufacturer's protocol: the Loughlin group extracted RNA at days 3, 7 and 14, and Dr Barter extracted RNA at days 0, 1, 3, 6, 10 and 14. Dr Barter subsequently used an Illumina Human HT-12 V4 expression array to profile a range of gene expressions during chondrogenesis (Barter *et al.*, 2015).

2.3 Human Mesenchymal Stem Cell (MSC) Differentiation Down an Osteoblastogenic Lineage

Osteoblastogenesis was performed by Dr Rodolfo Gomez (Institute of Cellular Medicine, Newcastle University). Briefly, MSCs at a density of 17,500 cells/cm^2 were cultured at 37^oC for 48 hours in Mesenchymal Stem Cell Growth Medium (Lonza, UK) supplemented with 5 ng/ml FGF-2 (R&D Systems, UK). The medium was replaced with Dulbecco's Modified Eagle Medium supplemented with 10% volume/volume (v/v) foetal bovine serum (FBS), 10 nM dexamethasone (Sigma-Aldrich, UK), 5 mM β-glycerol phosphate and 50 mg/ml ascorbic acid-2-phosphate. Cells were cultured in 60 mm Corning tissue-culture treated culture dishes (Sigma-Aldrich, UK) for 21 days to achieve full mineralisation. RNA was extracted at days 0 and 21 using TRIzol® Reagent (Life Technologies, UK) following the manufacturer's protocol. The RNA was used by Dr Gomez to profile a range of gene expressions during osteoblastogenesis on an Illumina Human HT-12 V4 expression array.

2.4 Tissue Sample Collection

Informed written consent for the use of joint tissue was provided by each patient who had undergone elective total joint replacement for OA of the knee or hip, and by patients who had total hip replacements due to neck of femur (NOF) fractures. Surgeries were performed at the Freeman Hospital and the Royal Victoria Infirmary, Newcastle-upon-Tyne. Ethical approval was granted by the Newcastle and North Tyneside Research Ethics Committee (REC reference number 09/H0906/72). Post-surgery, samples were stored in Hank's Balanced Salt Solution supplemented with penicillin, streptomycin and nystatin at 4^oC. Articular cartilage, infrapatellar fat pad and synovium were excised from the joints and were snap frozen at -80^oC on the day of surgery. The OA cartilage samples had visible lesions and were screened to exclude other pathologies, while cartilage from the NOF control joints did not show macroscopic damage or visible signs of OA.

2.5 Nucleic Acid Extraction from Joint Tissue

Under liquid nitrogen, 1 g – 3 g of frozen tissue was ground in a Retsch Mixer Mill MM 200 (Retsch, Leeds, UK). To 250 mg of ground tissue, 1 ml of TRIzol Reagent (Ambion, Life Technologies, UK) was added before thorough homogenisation by vortexing. Samples were incubated for 15 minutes (min) at room temperature followed by a 3 min centrifugation at 13,000 revolutions per minute (rpm) at 4^oC. The supernatant was transferred to a fresh microcentrifuge tube containing 200 µl chloroform, shaken vigorously for 15 seconds (sec) and incubated for 3 min at room temperature. The gDNA, total RNA and protein phases were

separated by centrifugation for 15 min at 13,000 rpm and 4°C. The top aqueous phase containing total RNA was transferred to a fresh microcentrifuge tube and the total RNA extracted using an RNeasy kit (QIAGEN, Crawley, UK) according to the manufacturer's instructions. The remaining phases were discarded. gDNA was extracted from 250 mg of ground tissue using an E.Z.N.A.® DNA/RNA Isolation Kit (Omega Bio-Tek, Georgia, USA). The manufacturer's protocol was followed with only slight modifications. Briefly: the ground tissue was vortexed with 700 µl GTC Lysis Buffer and centrifuged for 5 min at 11,700 rpm. The supernatant was transferred to a HiBind® column and centrifuged at 10,300 rpm for 1 min. The centrifugation was repeated following the addition of HB Buffer, and again upon the addition of DNA Wash Buffer. The membrane in the column was dried by centrifugation at 13,000 rpm for 2 min and the DNA subsequently eluted in 100 µl Elution Buffer.

2.6 Polymerase Chain Reaction (PCR) Optimisation

The polymerase chain reaction (PCR) annealing conditions of all primer pairs were optimised before use (Appendix A: Table A.1 and Table A.2). For the basic PCR reactions of genotyping, pyrosequencing and cloning, AmpliTaq Gold® *Taq* Polymerase (Applied Biosystems, Life Technologies, USA) was used, and for bisulfite converted DNA samples, Titanium® *Taq* DNA Polymerase (Clontech Laboratories, Inc., France) was used.

2.6.1 *Polymerase chain reaction (PCR) optimisation using AmpliTaq Gold® Taq Polymerase*

The optimal $MgCl_2$ concentration was tested (1 mM, 1.50 mM or 2 mM) over a range of annealing temperatures from 55°C to 70°C. Each 15 µl PCR reaction contained 1 x PCR Buffer II (Applied Biosystems, Life Technologies, USA), 0.50 µM forward primer (Sigma-Aldrich, UK), 0.50 µM reverse primer (Sigma-Aldrich, UK), 200 µM deoxynucleotide triphosphates (dNTPs; Bioline, UK), 0.40 U AmpliTaq Gold® *Taq* Polymerase (Applied Biosystems, Life Technologies, USA), $MgCl_2$ (Applied Biosystems, Life Technologies, USA) and 50 ng gDNA. PCR conditions were: initialisation for 14 min at 94°C, denaturation for 30 sec at 94°C, annealing for 30 sec, and extension for 1 min per 1 kb of template at 72°C. Denaturation, annealing and extension steps were repeated for a further 35 cycles before a final elongation step for 5 min at 72°C. Correct amplification was confirmed by gel electrophoresis, loading 4 µl of the PCR products onto a 2% weight/volume (w/v) agarose Tris/borate/EDTA (TBE) gel containing 20 µg ethidium bromide per 100 ml total volume. Amplimers were visualised under ultraviolet (UV) light using a G:BOX gel visualisation system (Syngene, UK).

2.6.2 *Polymerase chain reaction (PCR) optimisation using Titanium® Taq DNA Polymerase*

The annealing temperature of each primer pair was optimised over a temperature gradient from 55°C to 70°C. Each 20 µl reaction mix contained: 1 x Titanium® *Taq* PCR Buffer (Clontech Laboratories, Inc., France), 0.20 µM forward primer (Sigma-Aldrich, UK), 0.20 µM reverse primer (Sigma-Aldrich, UK), 200 µM dNTPs (Bioline, UK), 1 x Titanium® *Taq* DNA Polymerase (Clontech Laboratories, Inc., France) and 50 ng bisulfite converted DNA (Chapter 2.25). Thermal cycling conditions were: initialisation for 5 min at 95°C, denaturation for 30 sec at 95°C, annealing for 30 sec, and extension for 1 min at 68°C. Denaturation, annealing and extension steps were repeated for a further 35 cycles before a final elongation step for 5 min at 68°C. Correct amplification was confirmed by gel electrophoresis, loading 4 µl of the PCR products onto a 2% (w/v) agarose TBE gel containing 20 µg ethidium bromide per 100 ml total volume. Amplimers were visualised under ultraviolet UV light using a G:BOX gel visualisation system (Syngene, UK).

2.7 Complementary DNA (cDNA) Synthesis

2.7.1 *Reverse Transcriptase Polymerase Chain Reaction (RT-PCR)*

A final reaction volume of 20 µl was used to reverse transcribe 1 µg of total RNA. To remove contaminating DNA, total RNA was incubated at 37°C for 30 min with 1 U TURBO™ DNase (Invitrogen, Life Technologies, UK) and 1 x TURBO™ DNase Buffer (Invitrogen, Life Technologies, USA), followed by DNase inactivation at 75°C with 100 mM EDTA for 10 min. Total RNA was incubated at 65°C for 5 min with 1 µg random primers (Invitrogen, Life Technologies, USA), 10 µM dNTP mix and made up to 8 µl with diethylpyrocarbonate (DEPC)-treated water (Invitrogen, Life Technologies, USA). Each reaction was incubated for 1 min at 25°C with 1 x First Strand Buffer (Invitrogen, Life Technologies, USA), 5 mM $MgCl_2$ (Applied Biosystems, Life Technologies, USA), 10 mM DTT (Invitrogen, Life Technologies, USA) and 40 U RNaseOUT™ (Invitrogen, Life Technologies, USA). Total RNA was reverse transcribed into complementary DNA (cDNA) following the addition of 200 U SuperScript™ II Reverse Transcriptase (Invitrogen, Life Technologies, USA) for 10 min at 25°C, 50 min at 42°C and 10 min at 70°C. To degrade any complementary RNA, 2 U *E. coli* RNase H (Invitrogen, Life Technologies, USA) was incubated with each reverse transcriptase polymerase chain reaction (RT-PCR) sample for 20 min at 37°C. The cDNA was stored at -20°C.

2.7.2 *Polymerase Chain Reaction (PCR) to assess complementary DNA (cDNA) integrity*

Following cDNA synthesis, the integrity of the cDNA was assessed via PCR amplification of *HBP1*. The primers specific for this region (Appendix A: Table A.1) are located within two different exons of *HBP1*, meaning that cDNA and residual gDNA contamination can be distinguished. PCR was performed as previously described (Chapter 2.6.1), with 0.50 µl of cDNA added to a 14.50 µl master mix containing 2 mM $MgCl_2$. The annealing temperature was 60^{o}C. Only cDNA that had no contaminating gDNA was carried forward for downstream applications.

2.8 Quantitative Real-Time Polymerase Chain Reaction (qPCR)

In MicroAmp Fast Optical 96-well Reaction Plates (Applied Biosystems, Life Technologies, USA), PrimeTime Quantitative Real-Time-PCR (qPCR) Assays (Appendix A: Table A.3; Integrated DNA Technologies [IDT], Iowa, USA) were used in 10 µl total reactions with TaqMan Fast Universal 1 x PCR Master Mix (Applied Biosystems, Life Technologies, USA) and 2.50 µl of cDNA diluted 1:20. Gene expression was analysed in real-time for three replicates per sample per gene using the ABI PRISM 7900HT Sequence Detection System (Applied Biosystems, Life Technologies, USA). Thermal cycling conditions were: 95°C for 20 sec followed by 40 cycles of 95°C for 1 sec and 60°C for 20 sec. The housekeeper genes *18S*, *GAPDH* (glyceraldehyde 3-phosphate dehydrogenase) and *HPRT1* (hypoxanthine phosphoribosyltransferase 1) were also analysed following this protocol. For each of the three replicates for each patient for each gene, delta C_t (ΔC_t) was calculated using the formula: ΔC_t = C_t (test gene) – C_t (mean of control genes). Relative gene expression of the test gene compared to the housekeeper genes was analysed using the $2^{-\Delta Ct}$ method. Significance was assessed using a Mann-Whitney *U* test (two groups) or a Kruskal-Wallis one-way analysis of variance (three or more groups).

2.9 Online Database Search for Transcript Single Nucleotide Polymorphisms (SNPs)

Transcript SNPs of each gene were identified using the online software from the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway; (Kent *et al.*, 2002)). The Broad Institute (http://www.broadinstitute.org/mpg/snap/; (Johnson *et al.*, 2008)) online software was used to conduct pairwise searches to identify the LD between the transcript SNPs and the association SNPs.

2.10 Restriction Fragment Length Polymorphism (RFLP) Primer Design

The DNA sequences flanking approximately 200 bp either side of the SNPs were obtained from the UCSC Genome Browser. Forward and reverse primers specific for each SNP were designed using the Primer3 Input online software (http://primer3.wi.mit.edu/; (Rozen and Skaletsky, 2000)). A BLAST-like alignment tool (BLAT) search of the primers was performed using the UCSC Genome Browser BLAT software (Kent, 2002) to ensure specificity for the region of interest. The NEBcutter online software (http://tools.neb.com/NEBcutter2/; (Vincze *et al.*, 2003)) was used to identify restriction enzymes which cut at the SNP sites. A restriction fragment length polymorphism (RFLP) assay could not be designed for rs10492367 and so rs11049204, which is in perfect LD with rs10492367, was used as a proxy SNP.

2.11 Restriction Fragment Length Polymorphism (RFLP) Assay

DNA was amplified as per primer optimisation, replacing the temperature gradient and $MgCl_2$ concentration with the optimised conditions (Appendix A: Table A.1). PCR controls contained no template DNA. In a 15 µl final volume, 7.50 µl of amplified DNA was digested with 5 U of the restriction enzyme (New England BioLabs, UK) appropriate to the SNP of interest and 1 x reaction buffer (New England BioLabs, UK) as recommended by the manufacturer. After 3 hours at the appropriate temperature, the resulting DNA fragments were electrophoresed through a 3% (w/v) agarose TBE gel containing 20 µg ethidium bromide per 100 ml total volume. Restriction fragments were visualised under UV light in a G:BOX gel visualisation system (Syngene, UK). Patient details and genotypes are listed in Appendix D: Table D.1.

2.12 Allelic Quantification by Pyrosequencing

Pyrosequencing primers (Appendix A: Table A.1) were designed using PyroMark Assay Design Software 2.0 (QIAGEN, Crawley, UK), and the PCR conditions optimised for the primers (Sigma-Aldrich, UK) as detailed previously (Chapter 2.6.1). Cartilage DNA and cDNA samples were PCR amplified, 12 µl of which was added to separate wells of a 0.20 ml 24-well PCR plate (STARLAB, Milton Keynes, UK). Each 15 µl PCR sample was agitated using a 96-well plate shaker with a mix of 40 µl PyroMark binding buffer (QIAGEN, Crawley, UK), 2 µl streptavidin-coated sepharose beads (GE Healthcare, UK) and 28 µl deionised water for 10 min. Using a PyroMark Q24 Vacuum Workstation (QIAGEN, Crawley, UK), the sepharose beads were captured with the filter probes before 5 sec aspirations of 70% ethanol, 0.20 M sodium hydroxide and PyroMark wash buffer (QIAGEN,

Crawley, UK). The beads were released into a PyroMark Q24 24-well plate (QIAGEN, Crawley, UK) containing 24.75 µl PyroMark annealing buffer (QIAGEN, Crawley, UK) and 0.75 µl sequencing primer (Sigma-Aldrich, UK) per well. Denaturation proceeded for 2 min at 80^{o}C followed by primer annealing for 5 min at room temperature. A PyroMark Cassette (QIAGEN, Crawley, UK) was loaded with the appropriate volumes of PyroMark Gold Q24 Reagents (QIAGEN, Crawley, UK), followed by pyrosequencing in a PyroMark Q24 Pyrosequencing machine (QIAGEN, Crawley, UK). The capacity of all assays to differentiate between allelic quantities was validated in duplicate prior to use by comparing known synthetic allelic ratios with the ratios detected by the Pyrosequencing machine (Appendix B: Figure B.1). Each allelic expression imbalance (AEI) reaction was assayed in triplicate for DNA and cDNA. Significance was assessed using a Mann-Whitney *U* test.

2.13 Cloning of DNA into pGL3-Promoter Luciferase Reporter Vectors

2.13.1 *Amplification of DNA fragments containing alleles of the polymorphisms of interest*

gDNA of known major allele and minor allele homozygotes for each SNP of interest was used as starting material. Using the optimised PCR conditions, DNA fragments were PCR amplified with SNP-specific primers (Appendix A: Table A.2) containing an enforced restriction site for either *Mlu*I or *Bgl*II to allow ligation into the multiple cloning site of the pGL3-promoter vector (Appendix C: Figure C.1; Promega, UK). The reactions were purified using a QIAquick Gel Extraction kit (QIAGEN, Crawley, UK) according to the manufacturer's instructions.

2.13.2 *Digestion of DNA fragments with Mlu*I *and Bgl*II

The purified PCR products were digested in a master mix containing: 1 x NEBuffer 3.1 (New England BioLabs, UK), 10 U *Mlu*I (New England BioLabs, UK), 10 U *Bgl*II (New England BioLabs, UK), 1 µg DNA, and made up to 40 µl total volume with DEPC-treated water (Invitrogen, Life Technologies, UK). The reactions were incubated overnight at 37^{o}C. The samples were electrophoresed through a 1% (w/v) agarose TBE gel containing 20 µg ethidium bromide per 100 ml total volume. Amplimers were visualised under UV light, excised from the gel and purified using a QIAquick Gel Extraction kit (QIAGEN, Crawley, UK) according to the manufacturer's instructions. The pGL3-promoter vector was also digested and purified in this manner.

2.13.3 *Ligation of purified DNA fragments into pGL3-promoter vectors*

The purified DNA was ligated into the pGL3-promoter vector in a 3:1 DNA to plasmid molar ratio. This was performed overnight at 16^{o}C with 1 x T4 DNA Ligase Reaction Buffer (New England BioLabs, UK) and 400 U T4 Ligase (New England BioLabs, UK), and made up to a total volume of 20 µl.

2.13.4 *Transformation of pGL3-promoter vector constructs into chemically competent cells*

Two microlitres of the pGL3-promoter vector constructs were transformed into One Shot Mach1 T1 Phage-Resistant Chemically Competent *E.coli* cells (Invitrogen, Life Technologies, USA) according to the manufacturer's instructions. Transformed cells were spread onto agar plates supplemented with 100 µg/ml ampicillin (Sigma-Aldrich, UK) and incubated overnight at 37^{o}C. Three millilitres of lysogeny broth (LB) medium supplemented with 100 µg/ml ampicillin (Sigma-Aldrich, UK) was inoculated with unique clones that were picked from distinct colonies, and incubated at 37^{o}C with shaking overnight. Glycerol stocks were made by combining 300 µl bacterial culture with 300 µl glycerol, and stored at -80^{o}C. Plasmid cultures were purified using a PureYield^{TM} Plasmid Miniprep System (Promega, UK) according to the manufacturer's guidelines. The identity of the DNA insert was confirmed using sequencing performed by Source BioScience, UK.

2.14 Site-Directed Mutagenesis

Site-directed mutagenesis was performed using the QuikChange II Site-Directed Mutagenesis kit (Agilent Technologies, UK) according to the manufacturer's guidelines, with minor changes. Briefly, a 50 µl reaction contained: 1 x Reaction Buffer, 125 ng forward primer and 125 ng reverse primer (Appendix A: Table A.2), 0.01 mM dNTP mix, 2.50 U *PfuUltra* High-Fidelity DNA polymerase, 3 µl QuikSolution and 50 ng plasmid DNA. The PCR steps were an initial 95^{o}C for 1 min followed by 18 cycles of denaturation at 95^{o}C for 50 sec, annealing at 60^{o}C for 50 sec and extension at 68^{o}C for 7 min. Samples were treated with 10 U of *Dpn*I to digest the parental DNA and were incubated at 37^{o}C for 1 hour. Plasmid DNA was transformed into XL10-Gold Ultracompetent Cells (Agilent Technologies, UK), spread onto agar plates supplemented with 100 µg/ml ampicillin (Sigma-Aldrich, UK) and incubated overnight at 37^{o}C according to the manufacturer's instructions. Positive colonies were picked, cultured, purified and sequenced as previously described (Chapter 2.13.4).

2.15 Cell Line Culture

All cell culture was performed under sterile conditions. SW1353 cells are an immortalised chondrosarcoma cell line taken from the humerus of a 72 year old female Caucasian. Cells were cultured in Dulbecco's Modified Eagle Medium with a 1:1 ratio of Ham's F-12 nutrient mixture (GIBCO, Life Technologies, UK) supplemented with 10% (v/v) FBS, 100 IU/ml penicillin, 100 µg/ml streptomycin and 2 mM of L-glutamine (Sigma-Aldrich, UK). U2OS cells are an immortalised osteosarcoma cell line taken from the tibia of a 15 year old female Caucasian. Cells were cultured in Dulbecco's Modified Eagle Medium (GIBCO, Life Technologies, UK) supplemented with 10% (v/v) FBS, 100 IU/ml penicillin, 100 µg/ml streptomycin and 2 mM of L-glutamine (Sigma-Aldrich, UK). For each cell line, a vial of one million cells was resurrected from liquid nitrogen and warmed in a 37^{o}C water bath until defrosted. The cells were transferred to Corning vented 75 cm^{2} cell culture flasks (Sigma-Aldrich, UK) containing 9 ml of the respective media and cultured at 37^{o}C with 5% CO_2. After 4 hours, the medium was changed and the cells then cultured to 80% confluency. Upon reaching 80% confluency, the medium for each cell line was aspirated and the cells washed in PBS before incubation with 0.05% trypsin-EDTA solution (Sigma-Aldrich, UK) for 5 min at 37^{o}C with 5% CO_2. Once fully detached, the cells were passaged 1 in 3 and cultured as before.

2.16 Transfection of Cell Lines with pGL3-Promoter Luciferase Reporter Vectors

In Corning Costar 96-well cell culture plates (Sigma-Aldrich, UK), SW1353 cells were seeded at a density of 6,000 cells per well and U2OS cells at a density of 10,000 cells per well, and incubated for 24 hours at 37^{o}C with 5% CO_2. Cells were transfected with 1.50 ng pRL-TK *Renilla* Luciferase Reporter Vector (Appendix C: Figure C.2; Promega, UK) and 50 ng plasmid DNA using FuGENE HD Transfection Reagent (Promega, UK) in a total volume of 100 µl. Six wells were transfected as technical replicates per condition of an overall individual replication. Following a 24 hour incubation at 37^{o}C with 5% CO_2, cells were washed in 1 x phosphate-buffered saline (PBS) and lysed in 30 µl 1 x passive lysis buffer (Promega, UK) for 20 min with constant rocking. Twenty microlitres of the lysate were taken to quantify the luciferase and *Renilla* luciferase activities using the Dual Luciferase Reporter Assay System (Promega, UK) with a MicroLumat Plus LB96V luminometer (Berthold Technologies, UK). All luciferase values were normalised to the corresponding *Renilla* luciferase values. These were then normalised to the mean of the corresponding empty vector controls, which were given an arbitrary value of 1. Significance was assessed using a Mann-Whitney *U* test (two groups).

2.17 Human Articular Chondrocyte (HAC) Cell Culture

Excised cartilage samples were sliced and digested with three enzymes in order to extract the chondrocytes, with intervening washes with PBS: i) to digest hyaluronan, the tissue was incubated for 15 min at 37^{o}C with 1 mg/ml hyaluronidase in PBS (5 ml/g of cartilage); ii) to digest aggrecan, the sample was incubated for 30 min at 37^{o}C with 2.50 mg/ml trypsin in PBS (5 ml/g of cartilage); and iii) to digest collagen, the sample was incubated overnight at 35.5^{o}C with 2.00 mg/ml collagenase in Dulbecco's Modified Eagle Medium (GIBCO, Life Technologies, UK) containing 10% (v/v) FBS (3 ml/g of cartilage). The digested sample was then filtered through a 100 µm cell strainer (BD Biosciences, Oxford, UK) and the cells pelleted at 1,200 rpm for 7 min. The cell pellet was washed in PBS and centrifuged once more. Chondrocytes were seeded at a density of 40,000 cells/cm^2 in Corning vented 75 cm^2 cell culture flasks (Sigma-Aldrich, UK) containing Dulbecco's Modified Eagle Medium (GIBCO, Life Technologies, UK) supplemented with 10% (v/v) FBS, 2mM L-glutamine, 200 IU/ml penicillin, 200µg/ml streptomycin and 40 IU/ml nystatin until 90% confluent.

2.18 Nuclear Protein Extraction

2.18.1 *Preparation of buffers*

Hypotonic buffer (10 mM HEPES pH 7.60, 1.50 mM $MgCl_2$, 10 mM KCl, 1 mM DTT, 10 mM NaF, 1 mM Na_3VO_4, 0.10% [v/v] NP-40, 1 x complete protease inhibitor cocktail tablet per 50 ml of buffer [Roche, UK]) and high salt buffer (20 mM HEPES pH 7.90, 420 mM NaCl, 20% [v/v] glycerol, 1 mM DTT, 10 mM NaF, 1 mM Na_3VO_4, 1 x complete protease inhibitor cocktail tablet per 50 ml of buffer [Roche, UK]) were prepared prior to protein extraction.

2.18.2 *Preparation of cells*

SW1353 cells were seeded onto Corning 500 cm^2 square cell culture dishes (Sigma-Aldrich, UK) at a density of 12 x 10^6 cells per plate. U2OS cells were seeded onto Corning 500 cm^2 square cell culture dishes (Sigma-Aldrich, UK) at a density of 20 x 10^6 cells per plate. Human articular chondrocytes (HACs) were seeded at a density of 3 x 10^6 cells per Corning vented 75 cm^2 cell culture flask (Sigma-Aldrich, UK).

2.18.3 *Protein extraction from SW1353 and U2OS cell lines*

When 80% confluency was reached, cells were washed in 5 ml ice cold PBS. Cells were detached from the plate using a Corning cell scraper (Sigma-Aldrich, UK), collected in 5 ml fresh ice cold PBS and centrifuged at 10,000 rpm for 30 sec at 4^{o}C. The cell pellet was

resuspended in 1 ml hypotonic buffer and incubated on ice for 15 min with regular vortexing. Cells were centrifuged at 10,000 rpm for 30 sec at 4^{o}C and the supernatant containing cytosolic protein was snap-frozen on dry ice before storage at -80^{o}C. In order to fractionate the nuclei, the remaining pellet was resuspended in 1 ml hypotonic buffer supplemented with 0.25 M sucrose. Cells were centrifuged as before and the pellet resuspended in 0.80 µl high salt buffer per cm^{2} originally seeded. The sample was incubated on ice for 30 min with regular vortexing, followed by a final centrifugation at 10,000 rpm for 2 min at 4^{o}C. The supernatant containing nuclear protein was snap-frozen on dry ice before storage at -80^{o}C. After quantification of the protein (Chapter 2.19), the purity was assessed by western blot analysis (Chapter 2.20) using antibodies against lamin A/C (1:4,000) to detect the nuclear fractions and GAPDH (1:40,000) to detect the cytosolic fractions.

2.18.4 *Protein extraction from human articular chondrocytes (HACs)*

Protein was extracted from HACs in an identical manner, with only a slight modification. Rather than directly scraping the cells into 5 ml PBS upon reaching 80% confluency, HACs were instead detached by incubation for 5 min at 37^{o}C with 0.05% trypsin-EDTA solution (Sigma-Aldrich, UK).

2.19 Bradford Assay to Quantify Protein

A bovine serum albumin (BSA) standard curve was set up from 0.00 mg/ml to 1.50 mg/ml, using 2 mg/ml BSA stock solution (ThermoScientific, Surrey, UK) diluted in the lysis buffer used for protein extraction. Extracted protein was diluted in the respective lysis buffer. To each sample, 300 µl of Bradford Ultra (Expedeon, UK) was added and incubated for 5 min at room temperature. Absorbances were read at 595 nm using a Tecan Sunrise Microplate Absorbance Reader (Tecan, Reading, UK).

2.20 Western Blot for Protein Detection

Ten micrograms of protein was heated to 95^{o}C for 5 min with 1 x Laemmli sample buffer (0.10 M Tris-HCl, 0.35 M SDS, 20% [v/v] glycerol, 0.01% [v/v] bromophenol blue and 5% [v/v] β-mercaptoethanol) and then resolved on an 8% (w/v) SDS polyacrylamide gel. By electroblotting in a Scie-Plas V20-SDB 20 cm x 20 cm semi-dry blotter (Scie-Plas, Cambridge, UK), proteins were transferred to an Immobilon-P polyvinylidene fluoride membrane. The membrane was blocked for non-specific protein binding for 30 min at room temperature in 1 x PBS containing 5% (w/v) Marvel milk and 0.02% (v/v) Tween-20. Following a 5 min wash with 1 x PBS containing 0.02% (v/v) Tween-20, the membrane was

incubated at 4^{o}C overnight with a primary antibody diluted in 1 x PBS containing 5% (w/v) Marvel milk and 0.02% (v/v) Tween-20. All primary antibodies were used at a 1:2,000 dilution unless otherwise stated. The membrane was washed as before, followed by a 1 hour incubation at room temperature with the required secondary antibody: polyclonal goat anti-rabbit (1:2,000) or polyclonal goat anti-mouse (1:10,000) horseradish peroxidase (HRP)-conjugated antibodies (Dako, Denmark). Protein was detected using either Immobilon Western Chemiluminescent HRP Substrate (Millipore, UK), or ECL or ECL Select Western Blotting Substrate (GE Healthcare, Lille Chalfont, UK) in a G:BOX gel visualisation system (Syngene, UK). The antibodies used are listed in Appendix E: Table E.1 and Table E.2.

2.21 Prediction of Protein Binding Sites

Three publicly available online search tools were used to predict protein binding to each allele of the SNPs of interest: PROMO version 3.0 (http://alggen.lsi.upc.es; (Messeguer *et al.*, 2002; Farre *et al.*, 2003)), JASPAR version 5.0 (http://jaspar.binf.ku.dk; (Sandelin *et al.*, 2004)) and TFSEARCH version 1.3 (http://www.cbrc.jp/research/db/TFSEARCH; (Heinemeyer *et al.*, 1998)). In addition, the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway; (Kent *et al.*, 2002)) was used as a predictor of protein binding based on experimental cell line chromatin immunoprecipitation (ChIP) assays with sequencing (ChIP-Seq) data from ENCODE (Encode Project Consortium, 2012).

2.22 Selection of Transcription Factors for Functional Analysis

To assess whether a transcription factor should be included in electrophoretic mobility shift assay (EMSA) investigations, the gene had to be expressed in cartilage (Xu *et al.*, 2012b): of those that were expressed, any transcription factors that were predicted to bind only one allele of the SNP were selected. From the remaining transcription factors, the ones carried forward were selected based on the experimental evidence of binding and on the number of databases in which the protein was predicted to bind (Figure 2.1). Consensus sequences were found through a combination of literature searches and database searches, as detailed previously in Chapter 2.21, in addition to utilising data from the GeneCards website (www.genecards.org; (Safran *et al.*, 2010)).

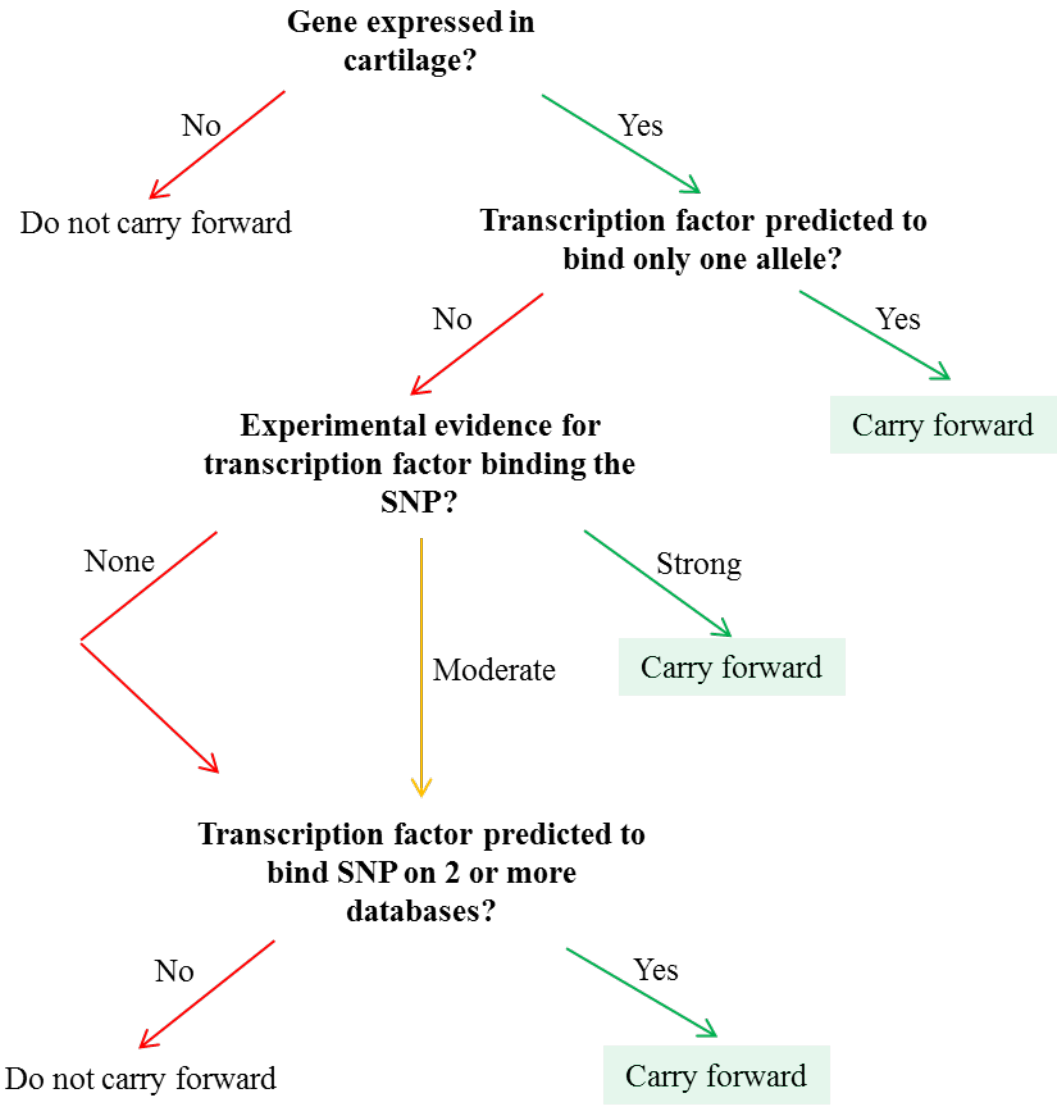

Figure 2.1. Flow diagram of the selection process for transcription factors known or predicted to bind the SNP of interest. The transcription factor must be expressed in cartilage to be appropriate for the selection process. Subsequently, the considerations are the number of alleles it is predicted to bind, the existing experimental evidence for binding, and the number of databases on which binding is predicted.

2.23 Electrophoretic Mobility Shift Assay (EMSA)

2.23.1 *Preparation of probes and buffers*

Fluorescently-labelled (5′DY682) oligonucleotides (Eurofins MWG Operon, Ebersberg, Germany) spanning 15 bp upstream and 15 bp downstream of each allele, for both the + and – strands, of the SNPs of interest were resuspended in water to a concentration of 100 pmol/µl. The complementary oligonucleotides were heated to 95^{o}C for 5 min with annealing buffer (100 mM Tris-HCl, 500 mM NaCl, 10 mM EDTA) and allowed to cool to room temperature for 2 hours in order for the + and – strands to anneal. A stock 5 x TBE buffer (445 mM Tris, 445 mM boric acid, 10 mM EDTA pH 8) and a native 5% (v/v) acrylamide gel (1 x TBE,

1:1,000 TMED and 0.07% [v/v] ammonium persulfate) were prepared in advance, and the gel allowed to set overnight at 4^{o}C. The gel was pre-run for 30 min at 4^{o}C in 0.50 x TBE buffer before samples were loaded: this was to guarantee a constant gel temperature, to remove traces of ammonium persulfate and to equilibrate ions in the running buffer.

2.23.2 *Binding reaction and electrophoresis*

The binding reactions were prepared using the components of an Odyssey Infrared EMSA kit (LiCor Biosciences, Cambridge, UK) and incubated at room temperature for 20 min in the dark. All 20 µl reactions contained 200 fmol of labelled probe. For competition EMSAs, unlabelled competitors were at 5 x (1 pmol), 10 x (2 pmol), 25 x (5 pmol), and 50 x (10 pmol) that of the labelled probe concentration. Supershift EMSAs contained either 3 µg or 6 µg of antibody (Appendix E: Table E.1 and Table E.2). Where appropriate, control reactions contained either i) no competitor or ii) a species-matched IgG antibody. Orange G loading dye was added to a final 1 x concentration and the samples then loaded onto the gel. Electrophoresis was at 100 V for 3 hours at 4^{o}C in the dark, and the protein binding was visualised using a LiCor Odyssey Infrared Imager (LiCor Biosciences, Cambridge, UK).

2.24 Chromatin Immunoprecipitation (ChIP)

ChIP was repeated six times in total. This constituted two independent rounds of cross-linking, cell harvesting and sonication (Chapter 2.24.2), followed by three immunoprecipitation repeats for each sonication (Chapter 2.24.3 and Chapter 2.24.4).

2.24.1 *Buffers*

Prior to ChIP, the following buffers were prepared: FA lysis buffer (50 mM HEPES-NaOH pH 7.50, 140 mM NaCl, 1 mM EDTA pH 8, 1% Triton X-100, 0.10% sodium deoxycholate, 0.10% SDS, 1 x complete protease inhibitor cocktail tablet per 100 ml of buffer [Roche, UK]), RIPA buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 2 mM EDTA pH 8, 1% NP-40, 0.50% sodium deoxycholate, 0.10% SDS, 1 x complete protease inhibitor cocktail tablet per 100 ml of buffer [Roche, UK]), wash buffer (0.10% SDS, 1% Triton X-100, 2 mM EDTA pH 8, 150 mM NaCl, 20 mM Tris-HCl pH 8), final wash buffer (0.10% SDS, 1% Triton X-100, 2 mM EDTA pH 8, 500 mM NaCl, 20 mM Tris-HCl pH 8) and elution buffer (1% SDS, 100 mM $NaHCO_3$).

2.24.2 *Cross-linking, cell harvesting and sonication*

SW1353 cells were seeded at a density of 2 x 10^6 cells per 10 cm circular Corning cell culture dish (Sigma-Aldrich, UK), and incubated for 24 hours at 37^{o}C with 5% CO_2. Proteins were cross-linked to the DNA by the addition of formaldehyde (Sigma-Aldrich, UK) to a final concentration of 0.75% and the plates rotated for 10 min at room temperature. To quench the formaldehyde, glycine was then added to the medium to a final concentration of 125 mM and incubated for 5 min at room temperature with shaking. Following two washes in ice cold PBS, the cells were scraped into fresh PBS and centrifuged for 5 min at 2,000 rpm at 4^{o}C. The cell pellet was resuspended in 750 µl FA lysis buffer per 1 x 10^7 cells. To shear the DNA to an average fragment size of 100-200 bp, the cell lysates were sonicated for 20 cycles of 30 sec intervals and 20 sec rest periods using a Bioruptor Standard (Diagenode, Belgium). The sonicated lysates were centrifuged at 4^{o}C for 30 sec at 8,000 rpm and the supernatant removed to a fresh tube. Fifty microlitres of the supernatant was retained as the input sample.

2.24.3 *Immunoprecipitation, elution and reverse cross-linking*

The concentration of protein in the supernatant was quantified using a Bradford assay (Chapter 2.19). For each immunoprecipitation reaction, 50 µg of protein was diluted 1:10 in RIPA buffer and 10 µg of the required antibody added. Acetyl-histone H3 was used as a positive control, while IgG was used as a species-matched control. The samples were incubated overnight with rotation at 4^{o}C with 20 µl protein A beads (Santa Cruz Biotechnology, USA). The beads were pelleted for 1 min at 5,000 rpm and washed in 3 x 1 ml wash buffer followed by 3 x 1 ml final wash buffer. The beads were incubated at 30^{o}C for 15 min in 120 µl elution buffer and then pelleted as above. At this stage, 400 µl of Tris-buffered saline (TBS) was added to the input sample and was hereafter treated in the same manner as the immunoprecipitated samples.

2.24.4 *Phenol-chloroform extraction of DNA*

To initiate DNA extraction by degrading the protein, each sample was incubated overnight at 65^{o}C with 20 mg/ml proteinase K. Equal volumes of the organic phase of acid-phenol: chloroform (Ambion, Life Technologies, UK) was added to the eluates, vortexed for 20 sec and centrifuged for 5 min at 13,000 rpm. The upper aqueous phase was vortexed for 20 sec with 2 x volume of 100% ethanol and incubated for 30 min at -20^{o}C. The DNA was pelleted at 13,000 rpm for 30 min at 4^{o}C, washed in 70% ethanol and centrifuged for a further 5 min at 13,000 rpm. The DNA pellet was resuspended in 50 µl water and stored at -20^{o}C.

2.24.5 *Quantification of DNA by quantitative real-time polymerase chain reaction (qPCR)*

Primers were designed to the DNA spanning no more than 100 bp (Appendix A: Table A.3). To quantify the DNA pulled-down with the protein of interest, a 7.50 µl master mix of KiCqStart® SYBR® Green qPCR ReadyMix™ (Sigma-Aldrich, UK), 0.20 µM forward primer, 0.20 µM reverse primer and DEPC-treated water (Invitrogen, Life Technologies, UK) was added to 2.50 µl DNA. Each reaction was performed in triplicate using the ABI PRISM 7900HT Sequence Detection System (Applied Biosystems, Life Technologies, USA). Thermal cycling conditions were: 95^{o}C for 10 min followed by 40 cycles of 95^{o}C for 3 sec and 60^{o}C for 30 sec, before a final sequence of 95^{o}C, 60^{o}C and 95^{o}C each for 15 sec. For each of the six repeats, the mean C_t values of the three technical replicates were normalised to the mean of the IgG control replicates. The normalised values were combined and significance was assessed using the Student's *t* test.

2.25 Bisulfite Conversion of DNA

Prior to bisulfite conversion, DNA was extracted from hip cartilage using the E.Z.N.A.® DNA/RNA Isolation Kit (Omega Bio-Tek, Georgia, USA) as detailed in Chapter 2.5. DNA was bisulfite converted using an EZ DNA Methylation™ Kit (Zymo Research, California, USA) according to the manufacturer's protocol. To 500 ng of DNA, 5 µl of M-Dilution Buffer was added and the volume adjusted to 50 µl with water. Following a 15 min incubation at 37^{o}C, 100 µl CT Conversion Reagent was mixed with the sample and incubated in the dark for 16 cycles of: 95^{o}C for 30 sec and 50^{o}C for 60 min. Four hundred microlitres of M-Binding Buffer was added to a Zymo-Spin™ IC Column containing the sample, and the column centrifuged for 30 sec at 13,000 rpm. Centrifugation was repeated following the addition of 100 µl M-Wash Buffer, and repeated again following a 20 min incubation with 200 µl M-Desulphonation Buffer. The column was washed twice with 200 µl M-Wash Buffer and the DNA eluted in 25 µl M-Elution Buffer.

2.26 Quantification of DNA Methylation by Pyrosequencing

Bisulfite converted DNA was amplified as per primer optimisation (Chapter 2.6.2), replacing the temperature gradient with the optimised annealing temperatures. Each reaction was performed in duplicate and pyrosequencing was carried out as previously described (Chapter 2.12). The mean of the duplicates was calculated and significance was assessed using a Mann-Whitney *U* test (two groups) or a Kruskal-Wallis one-way analysis of variance (three or more groups).

2.27 Cloning of DNA into pCpGL-Basic/EF_1 Luciferase Reporter Vectors

2.27.1 *Amplification of DNA fragments containing alleles of the polymorphisms of interest*

DNA fragments were PCR amplified from the previously-generated pGL3-promoter vector constructs. This used the optimised PCR conditions for the SNP-specific primers that contained enforced restriction site for either *Pst*I or *Spe*I (Appendix A: Table A.4), to allow for ligation into the multiple cloning site of the pCpGL-basic luciferase reporter vector that contained an EF_1 promoter (Appendix C: Figure C.3; (Klug and Rehli, 2006)). The reactions were purified using a QIAquick Gel Extraction kit (QIAGEN, Crawley, UK) according to the manufacturer's instructions.

2.27.2 *Digestion of DNA fragments with Pst*I *and Spe*I

The purified PCR products were digested in a master mix containing: 1 x NEBuffer 2 (New England BioLabs, UK), 4 µg BSA, 10 U *Pst*I (New England BioLabs, UK), 10 U *Spe*I (New England BioLabs, UK), 1 µg DNA, and made up to 40 µl total volume with DEPC-treated water (Invitrogen, Life Technologies, UK). The reactions were incubated overnight at 37°C. The samples were electrophoresed through a 1% (w/v) agarose TBE gel containing 20 µg ethidium bromide per 100 ml total volume. Amplimers were visualised under UV light, excised from the gel and purified using a QIAquick Gel Extraction kit (QIAGEN, Crawley, UK) according to the manufacturer's instructions. The pCpGL-basic/EF_1 vector was also digested and purified in this manner.

2.27.3 *Ligation of purified DNA fragments into pCpGL-basic/EF_1 vectors*

The purified DNA was ligated into the pCpGL-basic/EF_1 vector in a 3:1 DNA to plasmid molar ratio. This was performed for 1 hour at 25°C with 1 x T4 DNA Ligase Reaction Buffer (Invitrogen, Life Technologies, UK), 0.10 U T4 Ligase (Invitrogen, Life Technologies, UK), and made up to a total volume of 20 µl.

2.27.4 *Transformation of pCpGL-basic/EF_1 vector constructs into chemically competent cells*

Two microlitres of the pGL3-basic/EF_1 vector constructs were transformed into ChemiComp GT115 *E. coli* Cells (Invivogen, France) according to the manufacturer's instructions. Transformed cells were spread onto agar plates supplemented with 30 µg/ml zeocin (Invitrogen, Life Technologies, UK) and incubated overnight at 37°C. Three millilitres of LB medium supplemented with 30 µg/ml zeocin (Invitrogen, Life Technologies, UK) was inoculated with unique clones that were picked from distinct colonies, and incubated at 37°C

with shaking overnight. Glycerol stocks were made by combining 300 µl bacterial culture and 300 µl glycerol, and stored at -80°C. Plasmid cultures were purified using a PureYield™ Plasmid Miniprep System (Promega, UK) according to the manufacturer's guidelines. The identity of the DNA insert was confirmed using sequencing performed by Source BioScience, UK. Two hundred and fifty millilitres of LB medium supplemented with 30 µg/ml zeocin (Invitrogen, Life Technologies, UK) was inoculated with 5 µl of the glycerol stocks from the positive clones and incubated at 37°C with shaking overnight. Plasmid cultures were purified using a PureYield™ Plasmid Maxiprep System (Promega, UK) according to the manufacturer's guidelines.

2.28 Methylation of pCpGL-Basic/EF_1 Luciferase Reporter Vectors

Prior to transfection, all pCpGL-basic/EF_1 luciferase reporter vector constructs were methylated and mock-methylated. For both treatments, 5 µg of plasmid DNA was mixed with 1 x NEBuffer 2 (New England BioLabs, UK) and 1600 µM S-adenosylmethionine (SAM; New England BioLabs, UK). The constructs were methylated by the addition of 4 U *M.Sss*I (New England BioLabs, UK), or were mock-treated with DEPC-treated water (Invitrogen, Life Technologies, USA). The 30 µl reactions were incubated for 4 hours at 37°C followed by 20 min at 65°C, and the DNA precipitated. Briefly, the reactions were incubated at -80°C for 20 min with 100 µl 100% ethanol followed by centrifugation for 10 min at 13,000 rpm. After the supernatants were removed, the pellets were resuspended in 500 µl 70% ethanol followed by centrifugation for 5 min at 13,000 rpm. The supernatants were again removed. The pellets were then left to air dry for 5 min before resuspension in 40 µl DEPC-treated water (Invitrogen, Life Technologies, USA). Methylation was confirmed by digesting 250 ng of plasmid DNA with a relevant methylation-sensitive enzyme. DNA was diluted to 50 ng/µl and stored at -20°C.

2.29 Transfection of Cell Lines with pCpGL-Basic/EF_1 Luciferase Reporter Vectors

Transfection of the pCpGL-basic/EF_1 luciferase reporter vectors was performed as previously described for the pGL3-promoter vectors (Chapter 2.16), but instead using 100 ng of plasmid DNA per well rather than 50 ng per well. Luciferase/*Renilla* values were normalised to the corresponding empty vectors.

2.30 RNA-Mediated Interference (RNAi)

2.30.1 *Gene knockdown using small interfering RNA (siRNA)*

HACs were cultured as detailed in Chapter 2.17. After reaching 90% confluency, cells were seeded in Corning Costar 6-well cell culture plates (Sigma-Aldrich, UK) at a density of 120,000 cells per well and incubated for 24 hours at 37°C with 5% CO_2. In serum-free medium, 3.25 µl per well of DharmaFECT 1 Transfection Reagent (Dharmacon, GE Healthcare, UK) was diluted. Separately, SMARTpool ON-TARGET*plus* small interfering RNA (siRNA; Dharmacon, GE Healthcare, UK) was diluted in serum-free medium in an intermediate step to provide a final concentration of 100 nM. After a 5 min incubation at room temperature, the mixes were combined and the cells transfected for 48 hours at 37°C with 5% CO_2. The siRNAs are listed in Appendix E: Table E.3.

2.30.2 *Total RNA and protein extraction and quantification*

Total RNA and protein were extracted simultaneously using a NucleoSpin® RNA/Protein Extraction Kit (Macherey-Nagel, Germany) following the manufacturer's protocol. The extracted protein was quantified using a Bradford assay (Chapter 2.19), and a western blot (Chapter 2.20) was carried out to confirm protein depletion. One microgram of the RNA was reverse transcribed into cDNA (Chapter 2.7) before qPCR was used to assess gene knockdown (Chapter 2.8) using the assays predesigned by Integrated DNA Technologies [IDT], Iowa, USA, detailed in Appendix A: Table A.3. ΔC_t was calculated using the formula: $\Delta C_t = C_t$ (test gene) – C_t (mean of control genes). Relative gene expression of the test gene compared to the housekeeper genes was analysed using the $2^{-\Delta Ct}$ method. The mean of the $2^{-\Delta Ct}$ values for each knockdown was normalised to the non-targeting control. The normalised values for the three biological repeats were combined and significance was assessed using the Student's *t* test.

2.31 Co-Transfection of Small Interfering RNA (siRNA) and pGL3-Promoter Luciferase Reporter Vectors

For each condition, six wells of a Corning Costar 96-well cell culture plate (Sigma-Aldrich, UK) were transfected for quantification of the luciferase and *Renilla* reporter genes, and one well of a Corning Costar 6-well cell culture plate (Sigma-Aldrich, UK) was transfected for gene expression quantification. SW1353 cells were cultured and seeded as previously described (Chapter 2.15), with 6,000 cells per well of a 96-well plate and 120,000 cells per well of a 6-well plate. Transfections with the pGL3-promoter vector constructs were as detailed previously (Chapter 2.16), additionally using 90 ng of pRL-TK *Renilla* Luciferase

Reporter Vector (Promega, UK) and 3 µg of plasmid DNA per well of a six-well plate. After a 5 hour incubation at 37^{o}C with 5% CO_2, the cells were transfected with siRNAs as previously detailed (Chapter 2.30.1), with 0.26 µl of DharmaFECT 1 Transfection Reagent (Dharmacon, GE Healthcare, UK) used per well of a 96-well plate. Subsequently, the luciferase and *Renilla* expressions were quantified and the data analysed as before. Similarly, the total RNA and protein were extracted, quantified and analysed as detailed previously (Chapter 2.30.2).

Chapter 3. Characterisation of the 6q14.1 Locus Marked by the Polymorphism rs9350591

3.1 Introduction

One of the five genome-wide significant loci identified as being associated with OA in the arcOGEN GWAS was marked by the C/T polymorphism rs9350591 on chromosome 6q14.1 (arcOGEN Consortium *et al.*, 2012). The minor allele (T; MAF = 0.14) was significantly associated with hip OA in patients of European descent compared to the population-based controls, with an odds ratio (OR) of 1.18 and a *p* value of 2.42×10^{-09} (Table 3.1). There was a weaker significance in the association for all cases of OA, with an OR of 1.09 and a *p* value of 2.78×10^{-04}.

Discovery		**Replication**		**Discovery and replication**	
OR (95% CI)	***p* value**	**OR (95% CI)**	***p* value**	**OR (95% CI)**	***p* value**
1.20 (1.11-1.30)	2.49×10^{-06}	1.16 (1.07-1.25)	1.64×10^{-04}	1.18 (1.12-1.25)	2.42×10^{-09}

Table 3.1. Association statistics from the arcOGEN GWAS for rs9350591 in the hip stratum only. Odds ratio (OR), confidence interval (CI), genome-wide significance *p* value = 5×10^{-08}. Adapted from (arcOGEN Consortium *et al.*, 2012).

The association interval, defined as the region in which all SNPs in an LD of $\geq$ 0.80 with rs9350591 reside, spans 240 kb and encompasses *SENP6*. There are no non-synonymous transcript polymorphisms that are in high LD ($\geq$ 0.80) with rs9350591 and as such, it is unlikely that the OA association is mediated by a change in the coding sequence of any of the nearby genes. Instead, the causal SNP could be mediating its effect by regulating gene transcription. Gene expressions were quantified by Dr Madhushika Ratnayake and Dr Emma Raine of Prof. Loughlin's group (Institute of Cellular Medicine, Newcastle University) prior to the commencement of this project. The results confirmed that all of the genes at this locus, excluding *IMPG1*, were expressed in the following human joint tissues: articular cartilage, fat pad, synovium, meniscus, tendon and ligament (arcOGEN Consortium *et al.*, 2012). This chapter will therefore characterise the OA-associated region marked by rs9350591 by investigating the expression of *FILIP1*, *SENP6*, *MYO6*, *TMEM30A*, *COX7A2* and *COL12A1*.

3.2 Aim

The aim of this chapter was to characterise the OA association signal marked by the polymorphism rs9350591. This was achieved by:

- performing database searches to characterise the association region
- examining the expression profiles of *FILIP1*, *SENP6*, *MYO6*, *TMEM30A*, *COX7A2* and *COL12A1* during chondrogenesis using RNA generated by Dr Madhushika Ratnayake, Maria Tselepi and Emma Rogers
- analysing *FILIP1*, *SENP6*, *MYO6*, *TMEM30A*, *COX7A2* and *COL12A1* expression throughout chondrogenesis as assayed on a microarray performed by Dr Matthew Barter
- analysing *FILIP1*, *SENP6*, *MYO6*, *TMEM30A*, *COX7A2* and *COL12A1* expression throughout osteoblastogenesis as assayed on a microarray performed by Dr Rodolfo Gomez
- quantitatively analysing the expression of *FILIP1*, *SENP6*, *MYO6*, *TMEM30A*, *COX7A2* and *COL12A1* in synovial joint tissues
- characterising the expression profiles of *FILIP1*, *SENP6*, *MYO6*, *TMEM30A*, *COX7A2* and *COL12A1* in OA hip, OA knee and NOF cartilage, using data previously generated by Dr Madhushika Ratnayake and Dr Emma Raine
- replicating the gene expression quantification experiments in an independent group of OA hip cartilage samples
- characterising the expression profiles of *FILIP1*, *SENP6*, *MYO6*, *TMEM30A*, *COX7A2* and *COL12A1* in fat pad and synovium samples
- investigating if rs9350591 marks a *cis*-eQTL by producing an allelic expression imbalance of the transcription of *FILIP1*, *SENP6*, *MYO6*, *TMEM30A*, *COX7A2* and/or *COL12A1*
- analysing CpG methylation levels in hip and knee cartilage 1 Mb upstream and 1 Mb downstream of rs9350591 as assayed on a microarray performed by Dr Michael Rushton

3.3 Results

3.3.1 *Initial database searches to characterise the rs9350591 locus*

The UCSC Genome Browser (Kent *et al.*, 2002) collates information regarding the reference sequence of the human genome (Chapter 2.1); Figure 3.1 is a screenshot from the website covering the 6q14.1 locus. The first track displays RefSeq genes (Pruitt *et al.*, 2005), showing rs9350591 as an intergenic SNP 38 kb upstream of *FILIP1* and 70 kb upstream of *SENP6*. A further five genes reside within 1 Mb upstream or downstream of the polymorphism: *MYO6* (217 kb downstream), *TMEM30A* (247 kb downstream), *COX7A2* (288 kb downstream), *COL12A1* (325 kb downstream) and *IMPG1* (390 kb upstream). In addition, rs9350591 is upstream of two uncharacterised long non-coding RNAs, LOC100506804 and LOC101928540; and is within 103 kb of a provisional (that is, not independently reviewed) miRNA, MIR4463. Using ChIP-Seq data, the chromatin state of the region was modelled using a multivariate Hidden Markov Model (Ernst and Kellis, 2010; Ernst *et al.*, 2011). It is apparent, denoted by a grey shaded area, that the SNP resides in a region of heterochromatin. Furthermore, there are no transcription factors known to bind the polymorphism (Gerstein *et al.*, 2012; Wang *et al.*, 2012; Wang *et al.*, 2013), and the SNP is not in a DNase I hypersensitive region. This implies that the polymorphism is not mediating the OA association by affecting direct transcription factor binding; nor is it a regulatory element, which are generally identified as DNase I-sensitive loci (Song and Crawford, 2010). Finally, in order to characterise the three-dimensional structure of the genome, chromatin interaction analysis with paired-end tag (ChIA-PET) sequencing was undertaken as part of the ENCODE project (Fullwood *et al.*, 2010). Loci that are in close physical proximity with each other are cross-linked and ligated together, the graphical representation of which is denoted by blocks connected by horizontal lines. Based on this dataset it seems rs9350591 does not interact with other genomic loci. This is the same for all SNPs in high LD with rs9350591, but due to the large size of the association region (231 kb), only rs9350591 is depicted in Figure 3.1.B.

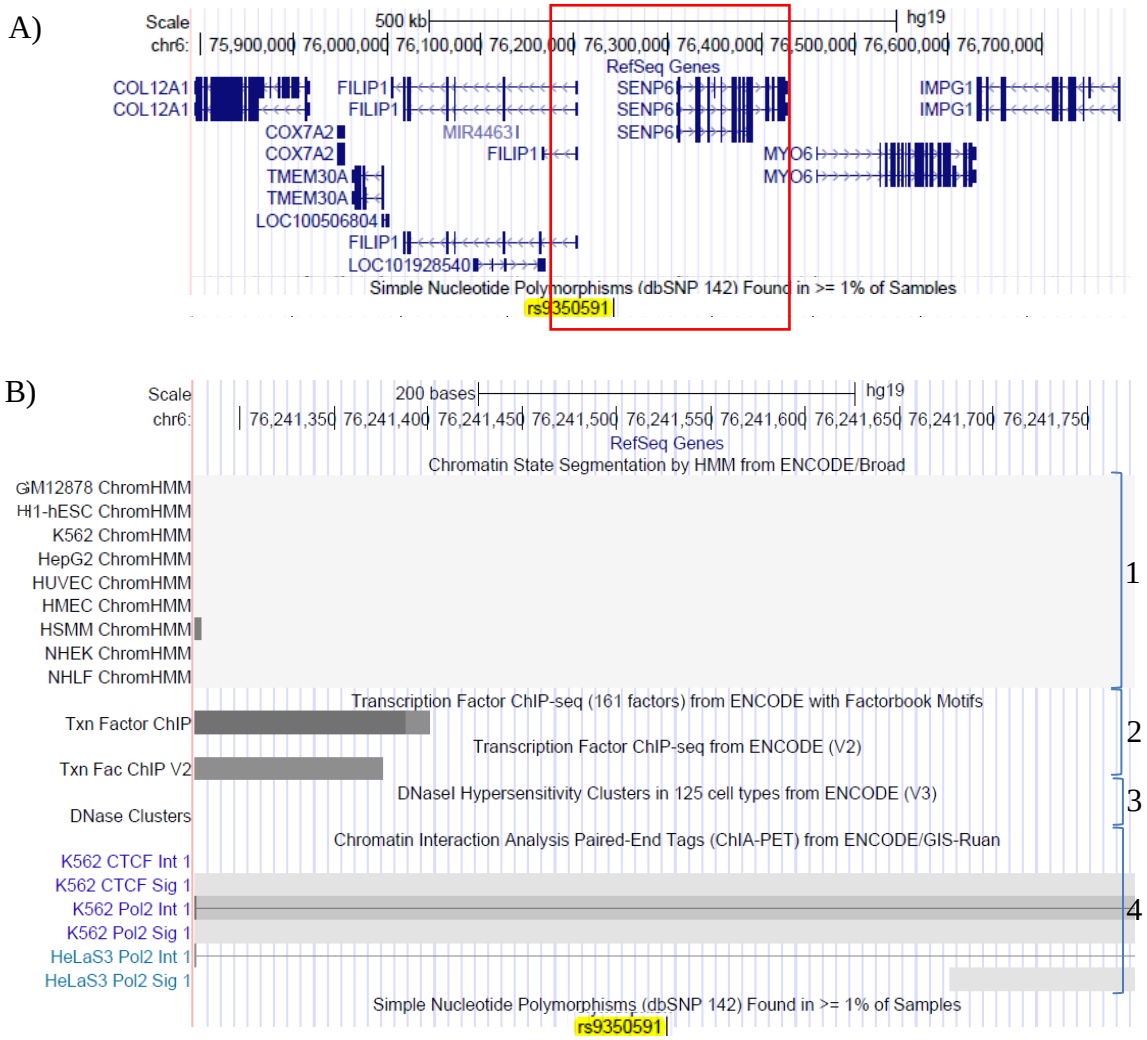

Figure 3.1. UCSC Genome Browser screenshots of the OA association region marked by the polymorphism rs9350591 on chromosome 6q14.1. A) rs9350591 is an intergenic polymorphism upstream of *FILIP1*, *SENP6*, *MYO6*, *TMEM30A*, *COX7A2*, *COL12A1* and downstream of *IMPG1*. The red box marks the boundaries of the association interval; all SNPs with an $r^2 \geq 0.80$ relative to rs9350591 reside in this region. B) The SNP does not reside in a predicted regulatory region, denoted by grey boxes within the Chromatin State Segmentation track [1]. No transcription factors have been identified as binding over the polymorphism within the Transcription Factor ChIP-Seq tracks [2], nor is it within a DNase I-hypersensitive region of the DNase I Hypersensitivity Clusters track [3]. The SNP has not been identified as interacting with other genomic loci in the Chromatin Interaction Analysis Paired-End Tags track [4]. The images were obtained using the hg19 reference genome.

As rs9350591 is not necessarily causal in mediating the OA association of this region, any SNPs in high LD with the polymorphism could instead account for the signal. In total, there are 39 SNPs with an r^2 of ≥ 0.80 relative to the association SNP, nine of which are in perfect LD with rs9350591. A search of the RegulomeDB online database (Boyle *et al.*, 2012) shows that some SNPs have transcription factors known to bind or reside in functional regions in relevant cell lines (Table 3.2). One of the 39 SNPs, rs9360921, was discussed in Chapter 1: following a meta-analysis of GWAS data and *in silico* replication, the polymorphism was

significantly associated with adult height in individuals of European descent (Lango Allen *et al.*, 2010). Such signals highlight the association of this region with skeletal development and perhaps could ultimately prove to be of relevance to OA susceptibility.

SNP	Distance from rs9350591 (bp)	r^2 relative to rs9350591	D' relative to rs9350591	Genotyped on arcOGEN array	SNP location	Transcription factor binding	Chromatin State		
							Bone marrow-derived cultured MSCs	MSC-derived chondrocyte cultured cells	Osteoblast primary cells
rs12211255	53,197	0.810	1.000	No	Intronic	No data	No data	No data	No data
rs10943249	42,308	1.000	1.000	No	Intronic	No data	No data	No data	No data
rs35985089	39,845	1.000	1.000	No	Intronic	No data	Weak transcription	Weak repressed polycomb	Weak repressed polycomb
rs9343292	38,871	1.000	1.000	No	Intronic	ATF2, NFKB1, POLR2A, TBP, NR3C1, USF1	Weak transcription	Flanking active TSS	Repressed polycomb
rs11964634	21,727	1.000	1.000	No	Intergenic	No data	No data	No data	No data
rs9360913	17,208	0.935	1.000	No	Intergenic	No data	Quiescent/low	Quiescent/low	Quiescent/low
rs9359125	10,223	0.872	1.000	No	Intergenic	No data	Quiescent/low	Quiescent/low	Quiescent/low
rs12190734	9,543	0.935	1.000	No	Intergenic	No data	No data	No data	No data
rs9343297	6,830	0.935	1.000	No	Intergenic	EP300, TCF7L2, TRIM28, ZNF263, GATA3, REST	Enhancer	Enhancer	Enhancer
rs9341526	5,138	0.935	1.000	No	Intergenic	ARID3A, CTCF, CEBPB, FOXM1, JUND, SMC3, YY1, RAD21, RUNX3	Quiescent/low	Quiescent/low	Weak repressed polycomb
rs12200169	4,406	1.000	1.000	No	Intergenic	No data	Quiescent/low	Quiescent/low	Weak repressed polycomb
rs12207675	3,786	0.935	1.000	No	Intergenic	No data	Quiescent/low	Quiescent/low	Weak repressed polycomb
rs12202443	2,913	0.935	1.000	No	Intergenic	CTCF	Quiescent/low	Quiescent/low	Weak repressed polycomb
rs13192994	2,265	0.935	1.000	No	Intergenic	No data	No data	No data	No data
rs9343299	1,682	0.935	1.000	No	Intergenic	No data	No data	No data	No data
rs9352215	1,646	0.935	1.000	No	Intergenic	No data	No data	No data	No data
rs9350591	0	1.000	1.000	Yes	Intergenic	No data	Weak repressed polycomb	Quiescent/low	Weak repressed polycomb

Table 3.2. All SNPs in high linkage disequilibrium ($r^2 \geq 0.80$) with rs9350591. Out of the thirty nine polymorphisms, nine were in perfect linkage disequilibrium with this association SNP. Transcription start site (TSS). Continued overleaf.

SNP	Distance from rs9350591 (bp)	r^2 relative to rs9350591	D' relative to rs9350591	Genotyped on arcOGEN array	SNP location	Transcription factor binding	Chromatin State		
							Bone marrow-derived cultured MSCs	MSC-derived chondrocyte cultured cells	Osteoblast primary cells
rs117337795	3,443	0.935	1.000	No	Intergenic	No data	No data	No data	No data
rs7756065	5,985	1.000	1.000	Yes	Intergenic	No data	Weak repressed polycomb	Weak repressed polycomb	Weak repressed polycomb
rs9359127	6,118	1.000	1.000	No	Intergenic	No data	Weak repressed polycomb	Repressed polycomb	Weak repressed polycomb
rs9359128	8,929	1.000	1.000	No	Intergenic	No data	Weak repressed polycomb	Weak repressed polycomb	Weak repressed polycomb
rs11963619	11,538	1.000	1.000	No	Intergenic	No data	No data	No data	No data
rs9352217	17,598	0.810	1.000	No	Intergenic	No data	Quiescent/low	Quiescent/low	Weak repressed polycomb
rs9360921	24,115	0.810	1.000	No	Intergenic	USF1	Quiescent/low	Weak repressed polycomb	Weak repressed polycomb
rs12201305	42,095	0.810	1.000	No	Intergenic	No data	No data	No data	No data
rs12213476	49,846	0.810	1.000	No	Intergenic	No data	Quiescent/low	Quiescent/low	Quiescent/low
rs9360926	58,467	0.810	1.000	No	Intergenic	No data	No data	No data	No data
rs9359133	84,148	0.810	1.000	No	Intronic	No data	Quiescent/low	Quiescent/low	Quiescent/low
rs67016585	88,824	0.810	1.000	No	Intronic	GATA1	Flanking active TSS	Enhancer	Enhancer
rs33997653	91,407	0.810	1.000	No	Intronic	No data	No data	No data	No data
rs12214738	102,256	0.810	1.000	No	Intronic	No data	Strong transcription	Strong transcription	Strong transcription
rs9360930	139,150	0.810	1.000	No	Intronic	No data	Strong transcription	Strong transcription	Strong transcription

Table 3.2. All SNPs in high linkage disequilibrium ($r^2 \geq 0.80$) with rs9350591. Out of the thirty nine polymorphisms, nine were in perfect linkage disequilibrium with this association SNP. Transcription start site (TSS). Continued overleaf.

SNP	Distance from rs9350591 (bp)	r^2 relative to rs9350591	D' relative to rs9350591	Genotyped on arcOGEN array	SNP location	Transcription factor binding	Chromatin State		
							Bone marrow-derived cultured MSCs	MSC-derived chondrocyte cultured cells	Osteoblast primary cells
rs9350596	139,519	0.810	1.000	No	Intronic	No data	No data	No data	No data
rs9341531	140,538	0.810	1.000	No	Intronic	No data	Strong transcription	Strong transcription	Weak transcription
rs9360932	146,410	0.810	1.000	No	Intronic	No data	No data	No data	No data
rs12192223	156,682	0.810	1.000	No	Intronic	No data	No data	No data	No data
rs9343320	174,156	0.810	1.000	No	Intronic	No data	Weak transcription	Strong transcription	Weak transcription
rs12208368	176,259	0.810	1.000	No	Intronic	No data	Strong transcription	Strong transcription	Weak transcription
rs12212171	178,209	0.810	1.000	No	Intronic	No data	Strong transcription	Enhancer	Weak transcription
rs17792773	186,968	0.810	1.000	No	Intergenic	No data	Strong transcription	Strong transcription	Quiescent/low

Table 3.2. All SNPs in high linkage disequilibrium ($r^2 \geq 0.80$) with rs9350591. Out of the thirty nine polymorphisms, nine were in perfect linkage disequilibrium with this association SNP. Transcription start site (TSS).

3.3.2 *Examination of the expression profiles of FILIP1, SENP6, MYO6, TMEM30A, COX7A2 and COL12A1 during chondrogenesis*

As the association signal resides in a region suggested to be functional, I first sought to investigate the expression of the surrounding genes throughout the development of cartilage. I hypothesised that the functional effects of the association signal could be exerted at early stages of joint development, resulting in altered joint structure and thus predisposing an individual to developing OA later in life. I tracked overall gene expression throughout chondrogenesis, irrespective of rs9350951 genotype, to clarify whether the genes were expressed and had the potential to contribute to joint development and OA susceptibility. Dr Madhushika Ratnayake, Maria Tselepi and Emma Rogers of Prof. Loughlin's group (Institute of Cellular Medicine, Newcastle University) performed *in vitro* chondrogenesis to replicate cartilage development (Chapter 2.2): I performed the qPCR (Chapter 2.8), data analysis and genotyping (Chapter 2.11). A summary of the donor information is detailed in Table 3.3. qPCR primer and probe sequences can be found in Appendix A: Table A.3, and the assay positions are detailed in Figure 3.2 (housekeeping genes) and Figure 3.3 (target genes). The probes and primers were designed to the cDNA sequences, and thus excluded intronic regions. For all of the genes, every known protein-coding transcript was covered by the assays.

Donor ID	Sex	Age at donation (years)	Joint	rs9350591 genotype
52	F	61	Hip	CC
93	F	51	Hip	No data
225	M	55	Hip	CC
276*	F	41	Iliac crest	CC
277*	F	24	Iliac crest	CC
278*	M	25	Iliac crest	CC

Table 3.3. Characteristics and genotype at rs9350591 for donors used in chondrogenesis. The donors marked with an asterisk (*) were donors with an unknown OA status purchased from Lonza, UK. There was no DNA available to genotype donor 93.

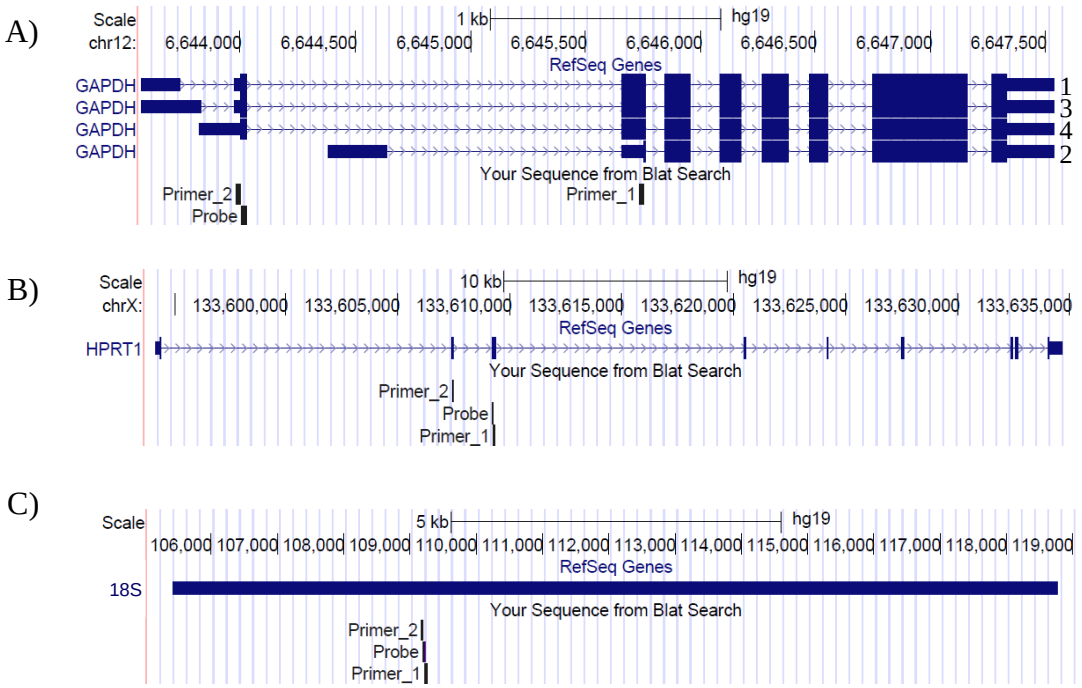

Figure 3.2. Location of qPCR primers and probes used for quantitative gene expression analysis. Assays were designed to the exons of A) *GAPDH*, B) *HPRT1* and C) *18S*. Every transcript (numbered 1 to 4) of *GAPDH* was covered by the assays. The images were obtained using the hg19 reference genome.

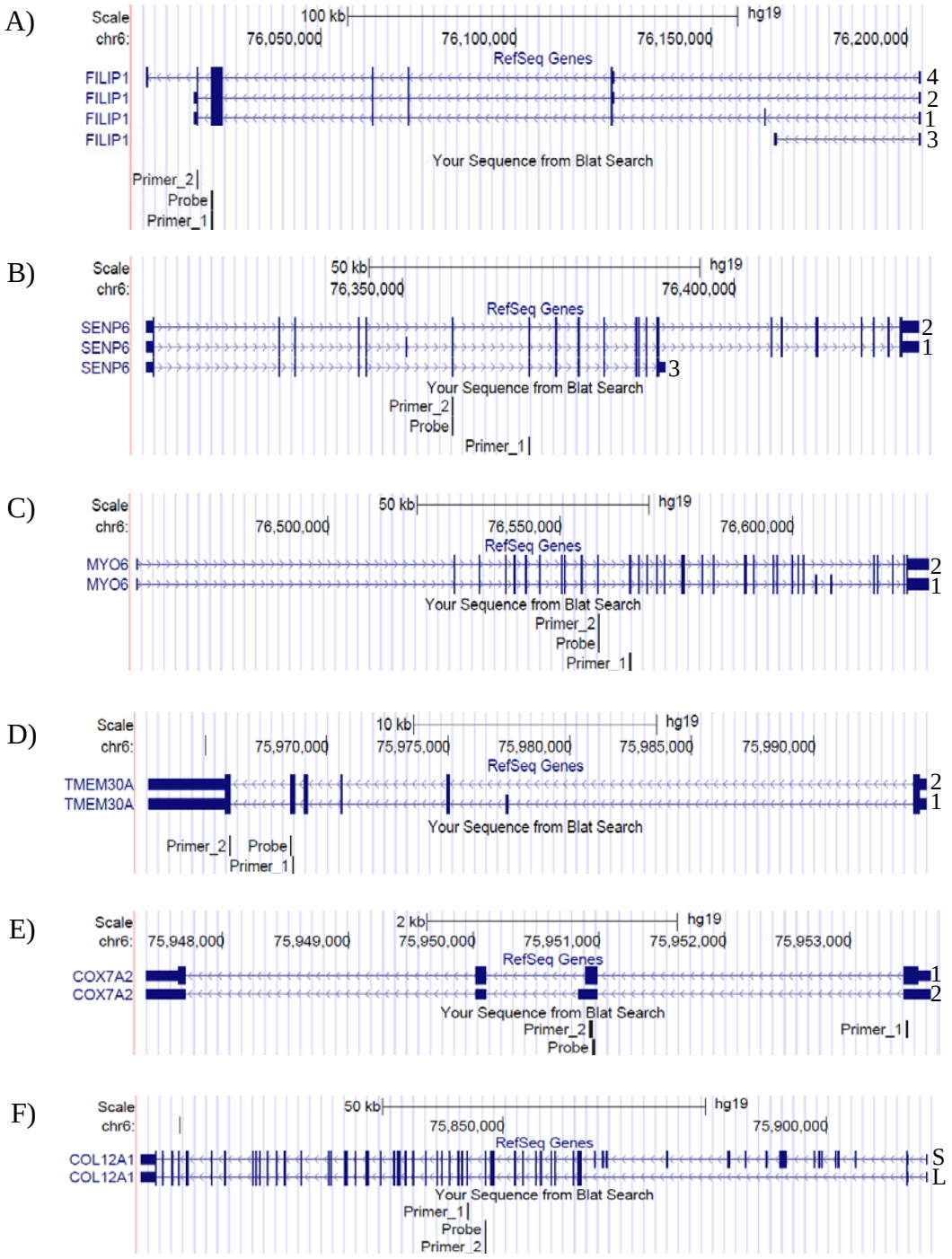

Figure 3.3. Location of qPCR primers and probes used for quantitative gene expression analysis. Assays were designed to the exons of A) *FILIP1*, B) *SENP6*, C) *MYO6*, D) *TMEM30A*, E) *COX7A2* and F) *COL12A1*. Every transcript of each gene was covered by the assays, excluding transcript variant 3 of *FILIP1*, which is annotated as non-coding. Transcript isoforms are numbered for each gene; *COL12A1* isoforms are annotated short (S) and long (L). The images were obtained using the hg19 reference genome.

Gene expressions were quantified at three time points throughout chondrogenesis: day 3, day 7 and day 14. At day 0, the MSCs would be undifferentiated; by day 7 the cells would have undergone rapid proliferation and differentiation; and by day 14 there would be evidence of an established collagen fibril network (Murdoch *et al.*, 2007). To confirm that chondrogenesis

had progressed, the expressions of the chondrogenic markers *ACAN* (aggrecan), *COL2A1* and *SOX9* (Bhang *et al.*, 2011) were confirmed at each time point by Dr Madhushika Ratnayake, Maria Tselepi and Emma Rogers (*data not shown*). All six genes under study were expressed throughout the chondrogenesis time course (Figure 3.4). There were no defined patterns of expression associated with the stage of cartilage development, nor were there distinct differences in expression between cells from Lonza donors and OA patients. The inter-individual variability and dynamic patterns of gene expression suggest that any of the genes could function at any of the stages of chondrogenesis to contribute to OA susceptibility.

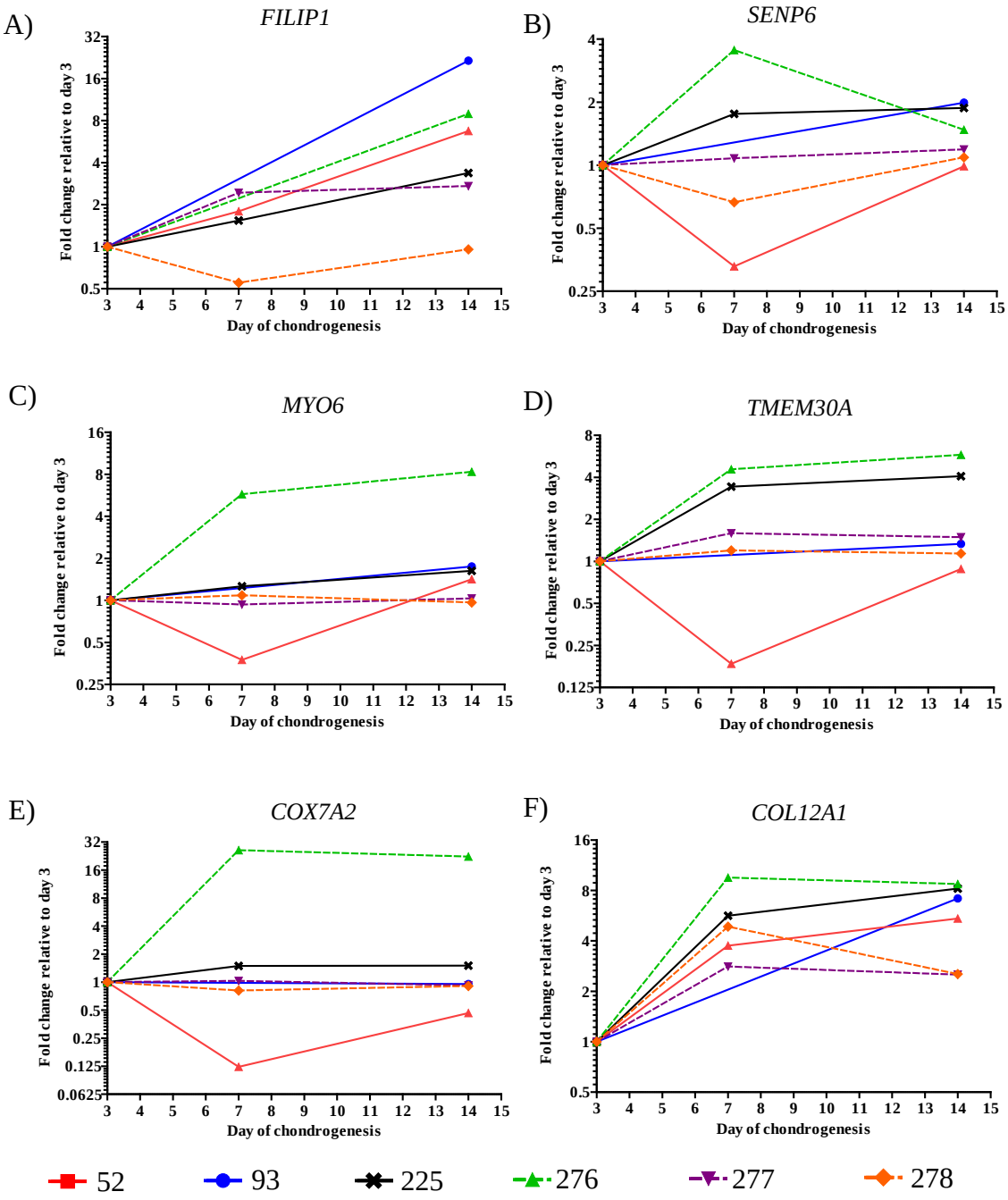

Figure 3.4. Expression of *FILIP1*, *SENP6*, *MYO6*, *TMEM30A*, *COX7A2* and *COL12A1* during chondrogenesis. MSCs were differentiated into chondrocytes and gene expression was quantified at various time points throughout chondrogenesis. A) *FILIP1*, B) *SENP6*, C) *MYO6*, D) *TMEM30A*, E) *COX7A2* and F) *COL12A1* were all expressed throughout the time course of chondrogenesis. The inter-individual variation and the dynamic expression patterns imply that any of the genes have the potential to function during cartilage development.

3.3.3 *Analysis of FILIP1, SENP6, MYO6, TMEM30A, COX7A2 and COL12A1 expression throughout chondrogenesis as assayed on a microarray*

All six genes were shown to be expressed during chondrogenesis in the preceding investigation (Chapter 3.3.2). As an additional approach to validate these results, I acquired normalised data from a microarray performed and analysed by Dr Matthew Barter (*personal communication*). MSCs from a 22 year old female (Lonza, UK) were differentiated into chondrocytes and, using RNA extracted at days 0, 1, 3, 6, 10 and 14, an Illumina Human HT-12 V4 expression array was used to profile a range of gene expressions (Barter *et al.*, 2015). Table 3.4 details the day 14 relative to day 0 expression fold change of the genes surrounding rs9350591 and of the chondrogenic markers *ACAN*, *COL2A1* and *SOX9* (Bhang *et al.*, 2011). Confirming that chondrogenesis had progressed, the chondrogenic markers were all significantly upregulated by day 14 relative to day 0. Similarly, *TMEM30A*, *COX7A2* and *COL12A1* had significantly increased levels of expression by day 14. Conversely, *MYO6* and *SENP6* were downregulated relative to day 0; however the difference was not significant for *SENP6* expression. *FILIP1* was not included in the data set due to failure to meet the quality control standards. The microarray output generates a detection *p* value, a measure of how confident the detection of a signal is. If a particular probe was anomalous or the gene was not expressed, the *p* value would be ≥ 0.01 and therefore considered not detected. In addition to day 14 relative to day 0, I also considered the gene expression at each individual time point for this donor (Figure 3.5) in order to independently corroborate the findings presented in Figure 3.4. All six genes were expressed throughout chondrogenesis, often at comparable levels to the chondrogenic markers.

Probe	Gene	Day 14 vs day 0 log_2 fold change	Average expression	Adjusted *p* value
5080215	*SENP6*	-0.026	8.566	0.816
1430497	*MYO6*	-0.554	7.033	1.148×10^{-4}
1240039	*TMEM30A*	0.811	8.458	1.516×10^{-6}
540491	*COX7A2*	0.247	13.017	0.015
3060095	*COL12A1* (short variant)	1.767	11.520	7.449×10^{-12}
4850129	*COL12A1* (long variant)	0.225	6.680	0.031
4480747	*ACAN* (variant 1)	2.237	7.445	7.573×10^{-13}
6770470	*ACAN* (variant 2)	3.851	8.518	2.217×10^{-16}
4010136	*COL2A1* (variant 1)	8.044	11.288	7.762×10^{-20}
650113	*COL2A1* (variant 1)	3.937	8.690	7.969×10^{-14}
4230475	*SOX9*	3.276	10.895	1.574×10^{-14}

Table 3.4. Comparison of day 14 and day 0 normalised gene expression during chondrogenesis at the rs9350591 locus. *MYO6* was significantly downregulated by day 14 relative to day 0, while *TMEM30A*, *COX7A2* and *COL12A1* were significantly upregulated. There was no significant difference in *SENP6* expression between day 14 and day 0. The chondrogenic markers were all significantly upregulated throughout the chondrogenesis time course. Both isoforms of *COL12A1* and *ACAN* were captured by the array, while two probes captured the same variant of *COL2A1*.

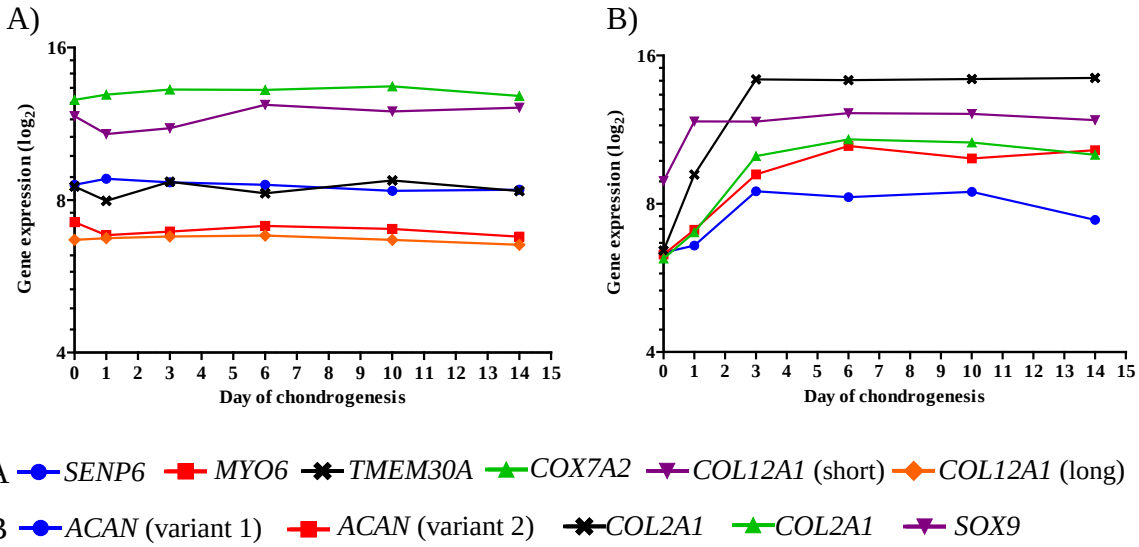

Figure 3.5. Expression of *SENP6, MYO6, TMEM30A, COX7A2* and *COL12A1* during chondrogenesis. MSCs were differentiated into chondrocytes and gene expression was quantified at various time points throughout chondrogenesis on an Illumina Human HT-12 V4 expression array. A) *SENP6*, *MYO6*, *TMEM30A*, *COX7A2* and *COL12A1* and B) the chondrogenic markers *ACAN*, *COL2A1* (covered by two probes [4010136, black x; and 650113, green triangle]) and *SOX9* were all expressed throughout the time course of chondrogenesis. The levels of expression imply that any of the genes at the 6q locus have the potential to function during cartilage development.

3.3.4 *Analysis of FILIP1, SENP6, MYO6, TMEM30A, COX7A2 and COL12A1 expression throughout osteoblastogenesis as assayed on a microarray*

As OA is increasingly becoming recognised as a disease of the entire synovial joint and not specifically of cartilage, I sought to investigate the expression of the genes during osteoblastogenesis. I hypothesised that, akin to the diverse expression patterns observed throughout chondrogenesis, any of the genes could in fact contribute to disease susceptibility by affecting joint development and structure by acting during osteoblastogenesis. I acquired normalised data from a microarray performed and analysed by Dr Rodolfo Gomez (*personal communication*). MSCs from a 19 year old female and a 24 year old female (Lonza, UK) were differentiated into osteoblasts and, using RNA extracted at days 0 and 21, an Illumina Human HT-12 V4 expression array was used to profile a range of gene expressions (Chapter 2.3). By day 21, mineralisation should be established, oftentimes having begun by day 17 (Matthews *et al.*, 2014). Table 3.5 details the day 21 relative to day 0 expression fold change of the genes surrounding rs9350591 and of the osteogenic markers *ALPL* (alkaline phosphatase, liver/bone/kidney) and *DCN* (decorin) (Waddington *et al.*, 2003; Graneli *et al.*, 2014). The osteogenic markers were both significantly upregulated by day 21 relative to day 0, confirming that osteoblastogenesis had progressed. Similarly, *FILIP1*, *SENP6*, *MYO6* and *TMEM30A* were upregulated by day 21, although this difference was only significant for *MYO6*. Conversely, *COX7A2* and *COL12A1* were downregulated relative to day 0, with the difference only significant for *COL12A1* (short variant). This is in contrast to the differences observed during chondrogenesis (Chapter 3.3.3). *COX7A2* and *COL12A1* (short variant) had the highest overall levels of expression, whereas *COL12A1* (long variant) and *FILIP1* had the lowest expression levels. *SENP6*, *MYO6* and *TMEM30A* had very similar levels of intermediate expression. Overall, all genes were expressed throughout osteoblastogenesis, often at comparable levels to the osteogenic markers. The upregulation of *MYO6* and the downregulation of *COL12A1* during osteoblastogenesis imply that either of these genes could contribute to OA susceptibility by affecting joint structure.

Probe	Gene	Day 21 vs day 0 log_2 fold change	Average expression	Adjusted *p* value
6200600	*FILIP1*	0.174	7.120	0.517
5080215	*SENP6*	0.130	8.748	0.654
1430497	*MYO6*	0.734	8.321	0.011
1240039	*TMEM30A*	0.290	9.037	0.383
540491	*COX7A2*	-0.241	12.936	0.347
3060095	*COL12A1* (short)	-1.706	12.711	0.001
4850129	*COL12A1* (long)	-0.183	7.191	0.507
6100356	*ALPL*	2.071	11.788	0.007
50368	*DCN* (variant A1)	0.786	11.061	0.033
5550719	*DCN* (variant A2)	1.495	11.857	5.75×10^{-06}
7650296	*DCN* (variant A2)	0.909	14.067	0.003

Table 3.5. Comparison of day 21 and day 0 normalised gene expression during osteoblastogenesis at the rs9350591 locus. *MYO6* was significantly upregulated by day 21 relative to day 0, while *COL12A1* was significantly downregulated. There were no significant differences in the expression of the remaining genes between day 21 and day 0. The osteogenic markers were both significantly upregulated throughout the osteoblastogenesis time course. Both isoforms of *COL12A1* and *DCN* were captured by the array.

3.3.5 *Quantitative expression analysis of FILIP1, SENP6, MYO6, TMEM30A, COX7A2 and COL12A1 in synovial joint tissues*

Since Chapter 3.3.1 indicated that the association signal resides in a region that is functional, and Chapters 3.3.2 – 3.3.4 confirmed that all of the six genes are expressed during chondrogenesis and osteoblastogenesis, I next aimed to characterise the expression of the genes in primary joint tissues. Following the publication of the arcOGEN GWAS in 2012 (arcOGEN Consortium *et al.*, 2012) and prior to the start of my Ph.D, expression datasets pertaining to the genes surrounding the 6q14.1 locus were generated by Dr Madhushika Ratnayake and Dr Emma Raine, two members of Prof. Loughlin's group (Institute of Cellular Medicine, Newcastle University). qPCR was performed using OA hip, OA knee and non-OA hip (NOF) cartilage samples (Chapter 2.4 – Chapter 2.8). Similarly, I performed qPCR using fat pad and synovium tissue originating from the knees of OA patients: all reactions were performed in triplicate and each reaction was normalised to the mean of the housekeeping genes *18S*, *GAPDH* and *HPRT1*.

In all of the joint tissues tested, expression of the six genes was detected (Figure 3.6): for a more in-depth analysis, the gene expression datasets for cartilage were separated into OA hip, OA knee and NOF cartilage. *FILIP1* and *COL12A1* were most highly expressed in fat pad relative to the other tissues tested. *SENP6*, *MYO6*, *TMEM30A* and *COX7A2* were most highly

expressed in OA knee cartilage. In addition, *SENP6* was also very highly expressed in NOF cartilage. The expression of *COX7A2* was lowest in fat pad, while the expressions of the remaining genes were all lowest in synovium. Importantly, the expression of each of the six genes was detected in the synovial joint tissues tested. This strengthens the hypothesis that any of the genes could contribute to joint structure and development, and potentially OA susceptibility.

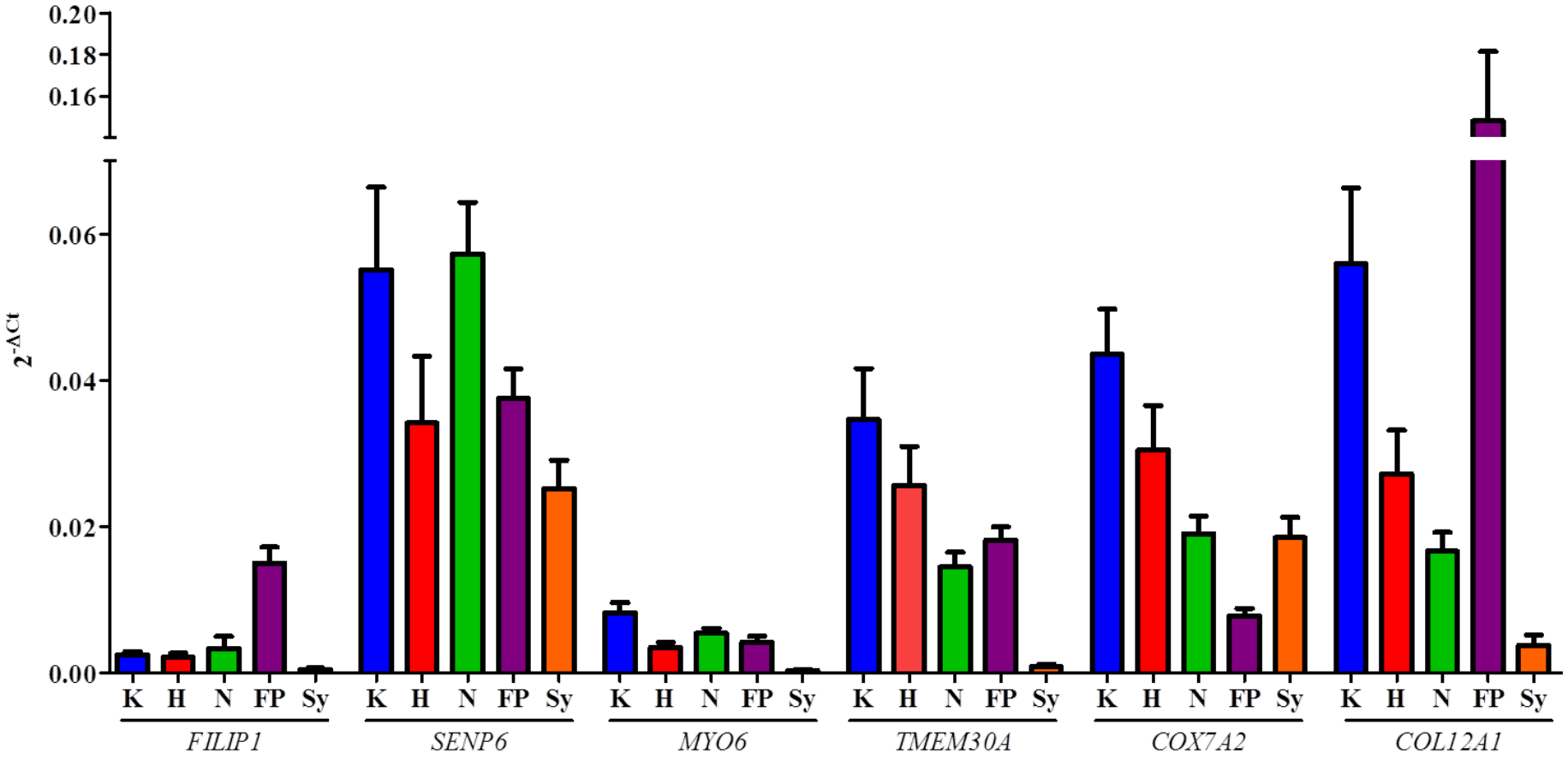

Figure 3.6. Average gene expression in the joint tissues assayed. *FILIP1* had the lowest overall expression of all the genes in OA hip, OA knee and NOF cartilage. In fat pad and synovium, *MYO6* had the lowest overall expression of the genes. *SENP6* was the most highly expressed gene in OA hip cartilage, NOF cartilage and synovium, while *COL12A1* was the most highly expressed gene in OA knee cartilage and fat pad. Tissues tested: OA knee cartilage (K; $n \leq 53$), OA hip cartilage (H; $n \leq 21$), NOF cartilage (N; $n \leq 19$), fat pad (FP; $n \leq 25$) and synovium (Sy; $n \leq 21$). Error bars represent the standard error of the mean (SEM).

3.3.6 *Characterising the expression profiles of FILIP1, SENP6, MYO6, TMEM30A, COX7A2 and COL12A1 in cartilage: comparisons of disease state, sex, skeletal site and age*

For each donor I compiled the following information: genotype at rs9350591, age (in years) at joint replacement, sex and joint replaced (Appendix D: Table D.1). I subsequently analysed the expression profiles of *FILIP1*, *SENP6*, *MYO6*, *TMEM30A*, *COX7A2* and *COL12A1* in cartilage using various stratifications (Table 3.6 – Table 3.11). As differences in gene expression depending on disease state have previously been shown in the context of OA-associated regions (Raine *et al.*, 2012), I included a comparison of OA and NOF. In addition, as discussed in Chapter 1, there are sex-specific differences in the prevalence of OA. Finally, given that OA is an age-associated disease, a comparison of gene expression relative to age was performed. For these analyses, solely due to the availability of tissue, all NOF donors were female. Only the statistically significant comparisons are included in the graphical representations (Figure 3.7 – Figure 3.10).

Stratification for *FILIP1* qPCR data	***p* value**
All OA (*n* = 68) vs NOF (*n* = 14)	0.635
OA hip (*n* = 19) vs NOF (*n* = 14)	0.600
OA female (*n* = 38) vs NOF female (*n* = 14)	0.557
OA female (*n* = 38) vs OA male (*n* = 30)	0.377
OA knee (*n* = 49) vs OA hip (*n* = 19)	0.891
OA female knee (*n* = 24) vs OA female hip (*n* = 14)	0.728
OA male knee (*n* = 25) vs OA male hip (*n* = 5)	0.404
All OA age: 50 (*n* = 13) vs 60 (*n* = 22) vs 70 (*n* = 24) vs 80 (*n* = 9)	0.314

Table 3.6. Analysis of *FILIP1* expression in OA hip, OA knee and NOF cartilage. There were no significant differences in gene expression for any of the stratifications. Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons and a one-way analysis of variance for a comparison of more than two groups; *n* represents the number of individuals in the comparison group.

Stratification for *SENP6* qPCR data	*p* value
All OA (*n* = 74) vs NOF (*n* = 19)	0.001
OA hip (*n* = 21) vs NOF (*n* = 19)	0.005
OA female (*n* = 40) vs NOF female (*n* = 19)	0.007
OA female (*n* = 40) vs OA male (*n* = 34)	0.736
OA knee (*n* = 53) vs OA hip (*n* = 21)	0.502
OA female knee (*n* = 25) vs OA female hip (*n* = 15)	0.468
OA male knee (*n* = 28) vs OA male hip (*n* = 6)	0.946
All OA age: 50 (*n* = 13) vs 60 (*n* = 23) vs 70 (*n* = 28) vs 80 (*n* = 10)	0.651

Table 3.7. Analysis of *SENP6* expression in OA hip, OA knee and NOF cartilage. Expression of *SENP6* was significantly lower in OA cartilage relative to NOF cartilage when comparing all OA vs NOF, OA hip vs NOF and OA female vs NOF. Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons and a one-way analysis of variance for a comparison of more than two groups; *n* represents the number of individuals in the comparison group.

Stratification for *MYO6* qPCR data	*p* value
All OA (*n* = 74) vs NOF (*n* = 19)	0.214
OA hip (*n* = 21) vs NOF (*n* = 19)	0.026
OA female (*n* = 40) vs NOF female (*n* = 19)	0.303
OA female (*n* = 40) vs OA male (*n* = 34)	0.948
OA knee (*n* = 53) vs OA hip (*n* = 21)	0.047
OA female knee (*n* = 25) vs OA female hip (*n* = 15)	0.288
OA male knee (*n* = 28) vs OA male hip (*n* = 6)	0.061
All OA age: 50 (*n* = 13) vs 60 (*n* = 23) vs 70 (*n* = 28) vs 80 (*n* = 10)	0.779

Table 3.8. Analysis of *MYO6* expression in OA hip, OA knee and NOF cartilage. Expression of *MYO6* was significantly lower in OA hip cartilage relative to both NOF cartilage and OA knee cartilage. Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons and a one-way analysis of variance for a comparison of more than two groups; *n* represents the number of individuals in the comparison group.

Stratification for *TMEM30A* qPCR data	*p* value
All OA (*n* = 74) vs NOF (*n* = 19)	0.128
OA hip (*n* = 21) vs NOF (*n* = 19)	0.176
OA female (*n* = 40) vs NOF female (*n* = 19)	0.107
OA female (*n* = 40) vs OA male (*n* = 34)	0.676
OA knee (*n* = 53) vs OA hip (*n* = 21)	0.774
OA female knee (*n* = 25) vs OA female hip (*n* = 15)	0.867
OA male knee (*n* = 28) vs OA male hip (*n* = 6)	0.288
All OA age: 50 (*n* = 13) vs 60 (*n* = 23) vs 70 (*n* = 28) vs 80 (*n* = 10)	0.697

Table 3.9. Analysis of *TMEM30A* expression in OA hip, OA knee and NOF cartilage. There were no significant differences in gene expression for any of the stratifications. Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons and a one-way analysis of variance for a comparison of more than two groups; *n* represents the number of individuals in the comparison group.

Stratification for *COX7A2* qPCR data	*p* value
All OA (*n* = 74) vs NOF (*n* = 19)	0.030
OA hip (*n* = 21) vs NOF (*n* = 19)	0.390
OA female (*n* = 40) vs NOF female (*n* = 19)	0.091
OA female (*n* = 40) vs OA male (*n* = 34)	0.725
OA knee (*n* = 53) vs OA hip (*n* = 21)	0.291
OA female knee (*n* = 25) vs OA female hip (*n* = 15)	0.534
OA male knee (*n* = 28) vs OA male hip (*n* = 6)	0.456
All OA age: 50 (*n* = 13) vs 60 (*n* = 23) vs 70 (*n* = 28) vs 80 (*n* = 10)	0.939

Table 3.10. Analysis of *COX7A2* expression in OA hip, OA knee and NOF cartilage. *COX7A2* expression was significantly increased in all OA cartilage relative to NOF cartilage. Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons and a one-way analysis of variance for a comparison of more than two groups; *n* represents the number of individuals in the comparison group.

Stratification for *COL12A1* qPCR data	*p* value
All OA (*n* = 74) vs NOF (*n* = 19)	0.010
OA hip (*n* = 21) vs NOF (*n* = 19)	0.144
OA female (*n* = 40) vs NOF female (*n* = 19)	0.021
OA female (*n* = 40) vs OA male (*n* = 34)	0.733
OA knee (*n* = 53) vs OA hip (*n* = 21)	0.140
OA female knee (*n* = 25) vs OA female hip (*n* = 15)	0.118
OA male knee (*n* = 28) vs OA male hip (*n* = 6)	0.946
All OA age: 50 (*n* = 13) vs 60 (*n* = 23) vs 70 (*n* = 28) vs 80 (*n* = 10)	0.315

Table 3.11. Analysis of *COL12A1* expression in OA hip, OA knee and NOF cartilage. *COL12A1* expression was significantly increased in all OA cartilage relative to NOF cartilage. Expression was also significantly increased in OA female cartilage relative to NOF cartilage. Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons and a one-way analysis of variance for a comparison of more than two groups; *n* represents the number of individuals in the comparison group.

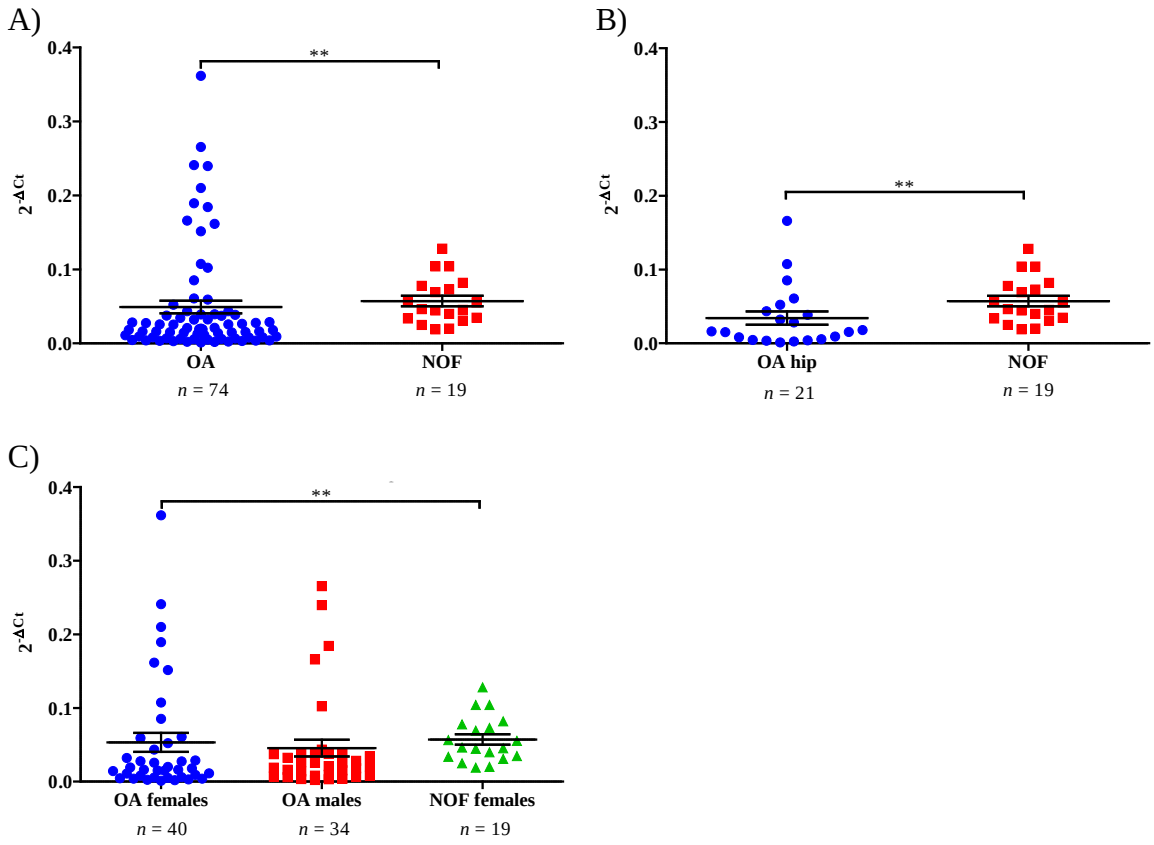

Figure 3.7. Statistically significant differences in the expression of *SENP6*. Gene expression was significantly decreased in A) all OA cartilage relative to NOF cartilage, B) OA hip cartilage relative to NOF cartilage, and C) OA female cartilage relative to NOF female cartilage. Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group; * $p \leq 0.05$; ** $p \leq 0.01$; error bars represent the mean $\pm$ SEM.

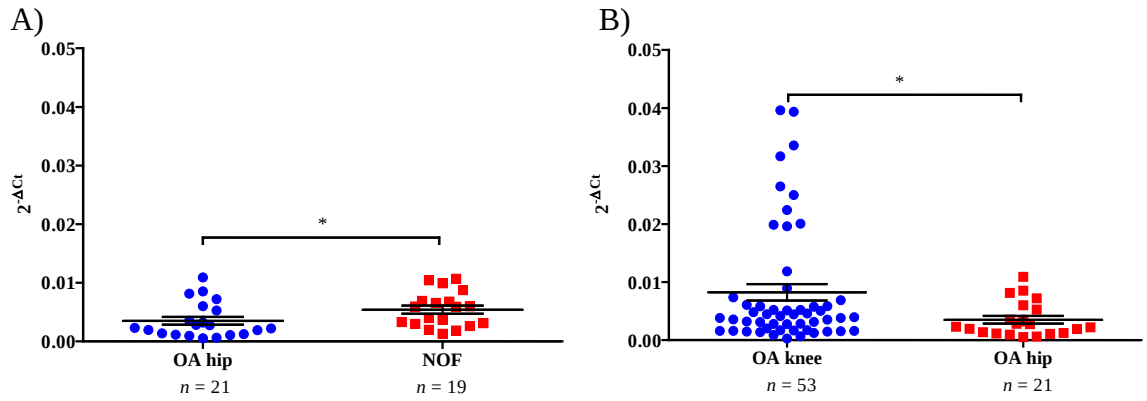

Figure 3.8. Statistically significant differences in the expression of *MYO6*. Gene expression was significantly decreased in A) OA hip cartilage relative to NOF cartilage and B) OA hip cartilage relative to OA knee cartilage. Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group; * $p \leq 0.05$; error bars represent the mean ± SEM.

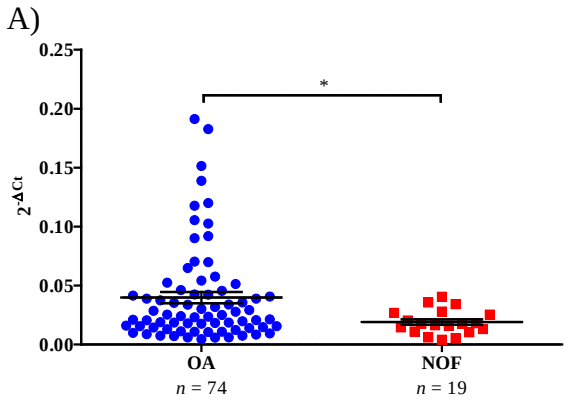

Figure 3.9. Statistically significant difference in the expression of *COX7A2*. Gene expression was significantly increased in A) all OA cartilage relative to NOF cartilage. Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group; * $p \leq 0.05$; error bars represent the mean ± SEM.

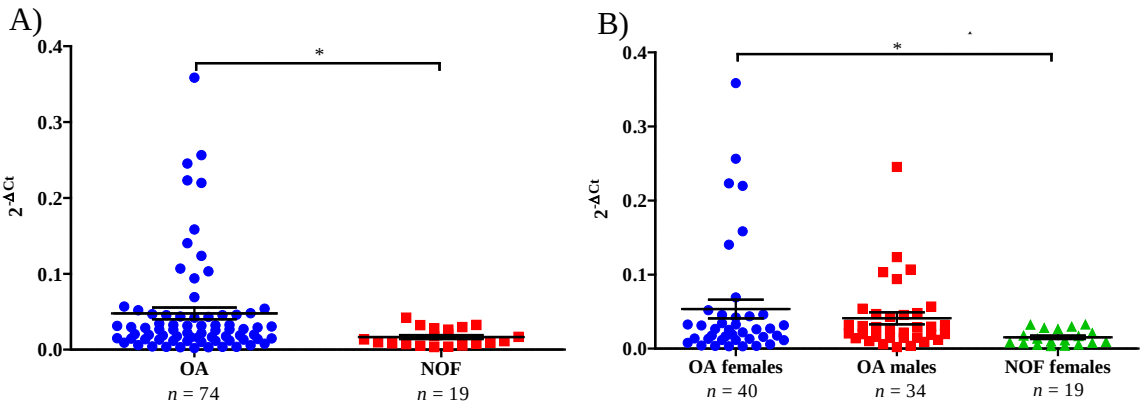

Figure 3.10. Statistically significant differences in the expression of *COL12A1*. Gene expression was significantly increased in A) all OA cartilage relative to NOF cartilage and B) OA female cartilage relative to NOF cartilage. Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group; * $p \leq 0.05$; error bars represent the mean ± SEM.

From the analysis of these data, I identified differential expression of *SENP6*, *MYO6*, *COX7A2* and *COL12A1*. Expression of *SENP6* was decreased in cartilage from all OA, OA hip and OA female relative to NOF cartilage. *MYO6* was similarly decreased in cartilage from OA hip relative to both NOF and OA knee cartilage. Conversely, the expression of *COX7A2* was increased in all OA cartilage relative to NOF cartilage, while *COL12A1* expression was increased in OA female cartilage relative to NOF female cartilage.

3.3.7 *Characterising the expression profiles of FILIP1, SENP6, MYO6, TMEM30A, COX7A2 and COL12A1 in cartilage: comparisons of rs9350591 genotype*

The stratification necessary to investigate if an eQTL is operating at this locus is the comparison of gene expression between the genotypic groups of rs9350591; that is CC, CT and TT. However, due to the low MAF (0.14) of rs9350591, it was not feasible in the timeframe of my Ph.D to acquire enough individuals that were TT at the association signal. As such, T allele carriers (CT and TT genotypes) were grouped together and compared to CC homozygotes. From these data, it was not possible to identify the actions of an eQTL influencing the expression of any of the genes tested (Figure 3.11 – Figure 3.16).

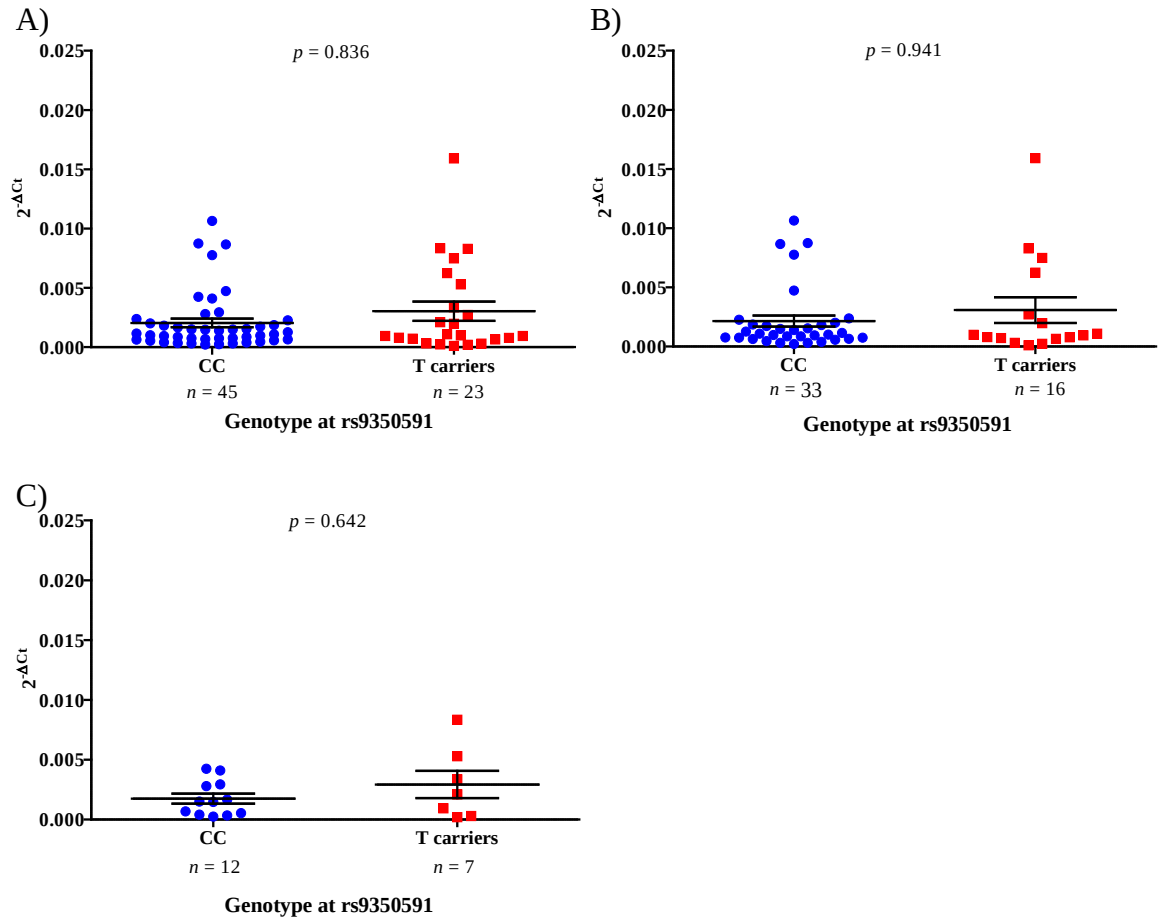

Figure 3.11. Analysis of *FILIP1* expression in OA hip and OA knee cartilage. There were no significant differences in gene expression between the genotypic groups of rs9350591 in A) all OA cartilage (hip and knee combined), B) OA knee cartilage, or C) OA hip cartilage. Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group; error bars represent the mean ± SEM.

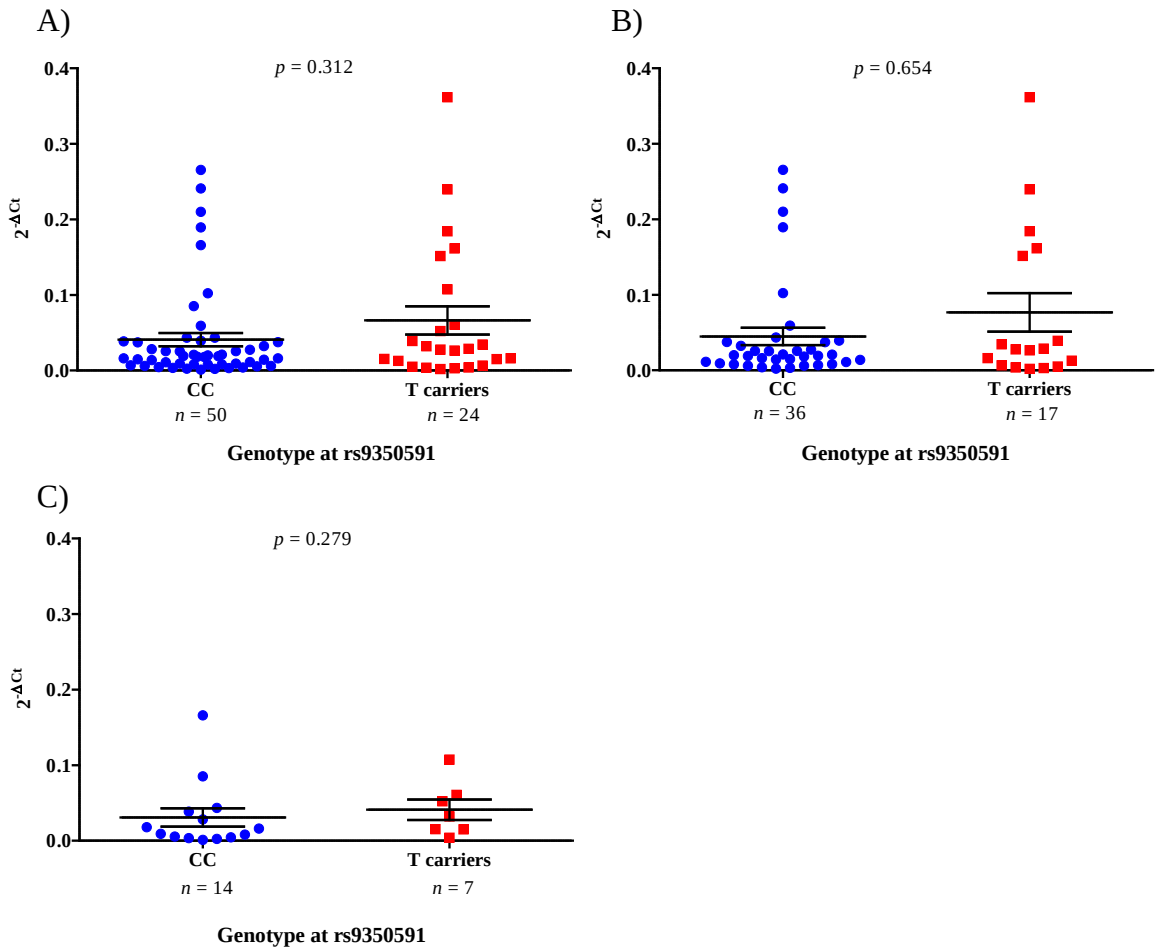

Figure 3.12. Analysis of *SENP6* expression in OA hip and OA knee cartilage. There were no significant differences in gene expression between the genotypic groups of rs9350591 in A) all OA cartilage (hip and knee combined), B) OA knee cartilage, or C) OA hip cartilage. Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group; error bars represent the mean ± SEM.

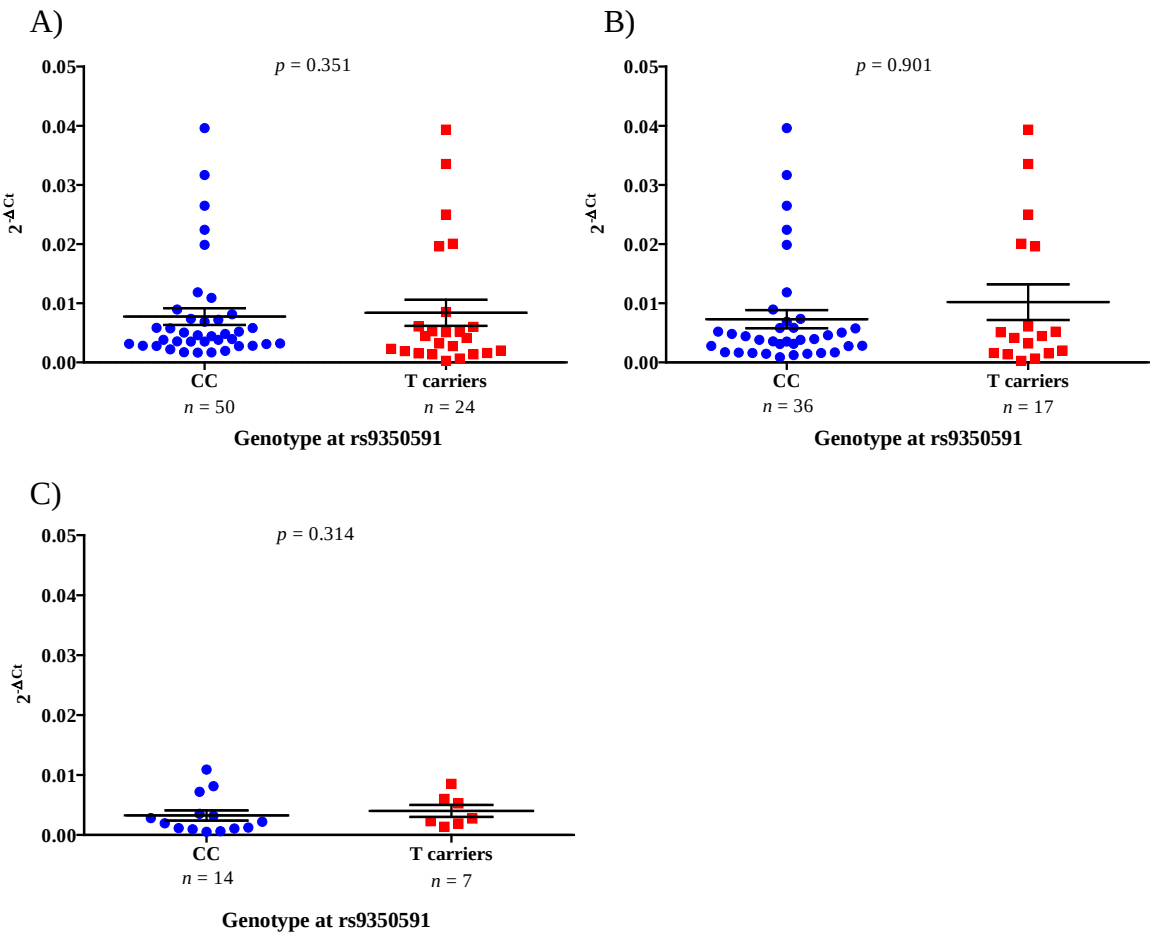

Figure 3.13. Analysis of *MYO6* expression in OA hip and OA knee cartilage. There were no significant differences in gene expression between the genotypic groups of rs9350591 in A) all OA cartilage (hip and knee combined), B) OA knee cartilage, or C) OA hip cartilage. Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group; error bars represent the mean ± SEM.

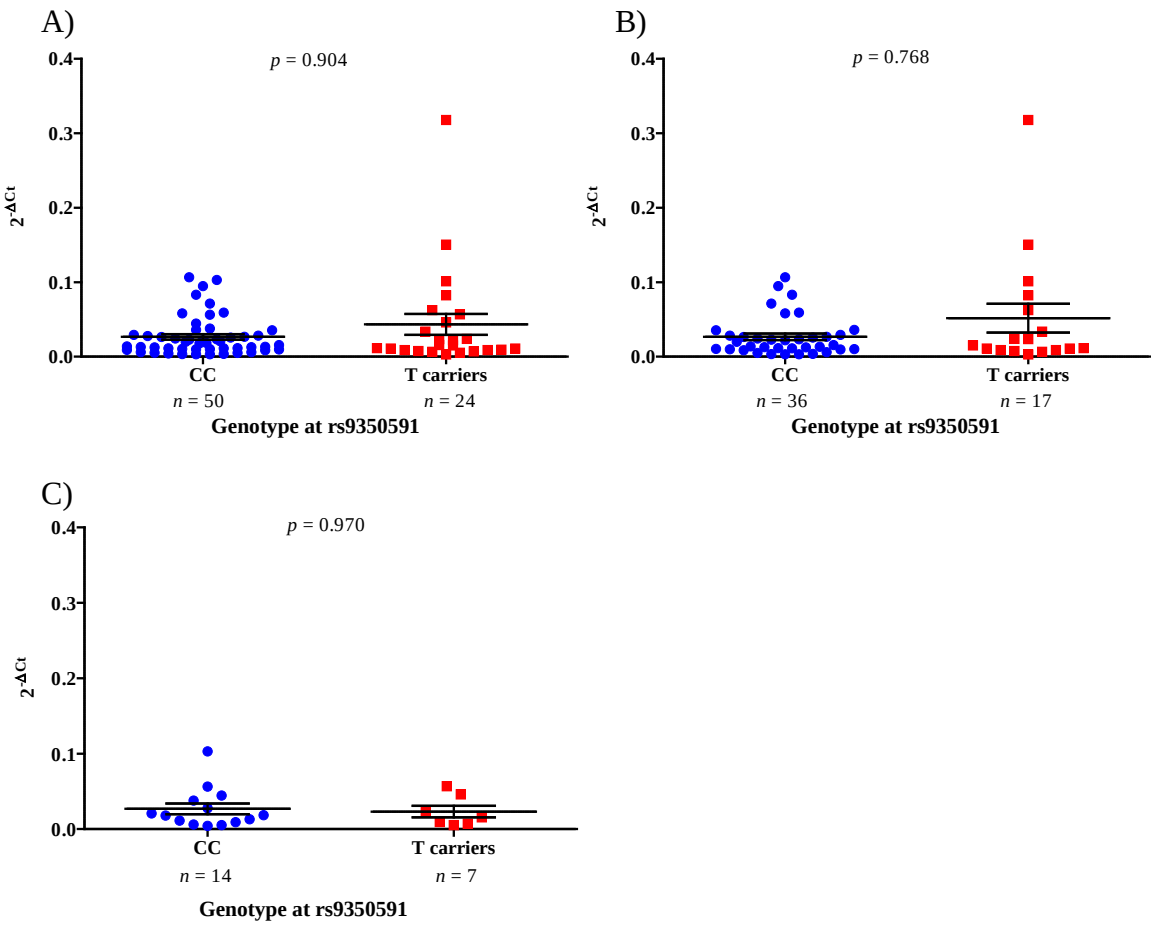

Figure 3.14. Analysis of *TMEM30A* expression in OA hip and OA knee cartilage. There were no significant differences in gene expression between the genotypic groups of rs9350591 in A) all OA cartilage (hip and knee combined), B) OA knee cartilage, or C) OA hip cartilage. Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group; error bars represent the mean ± SEM.

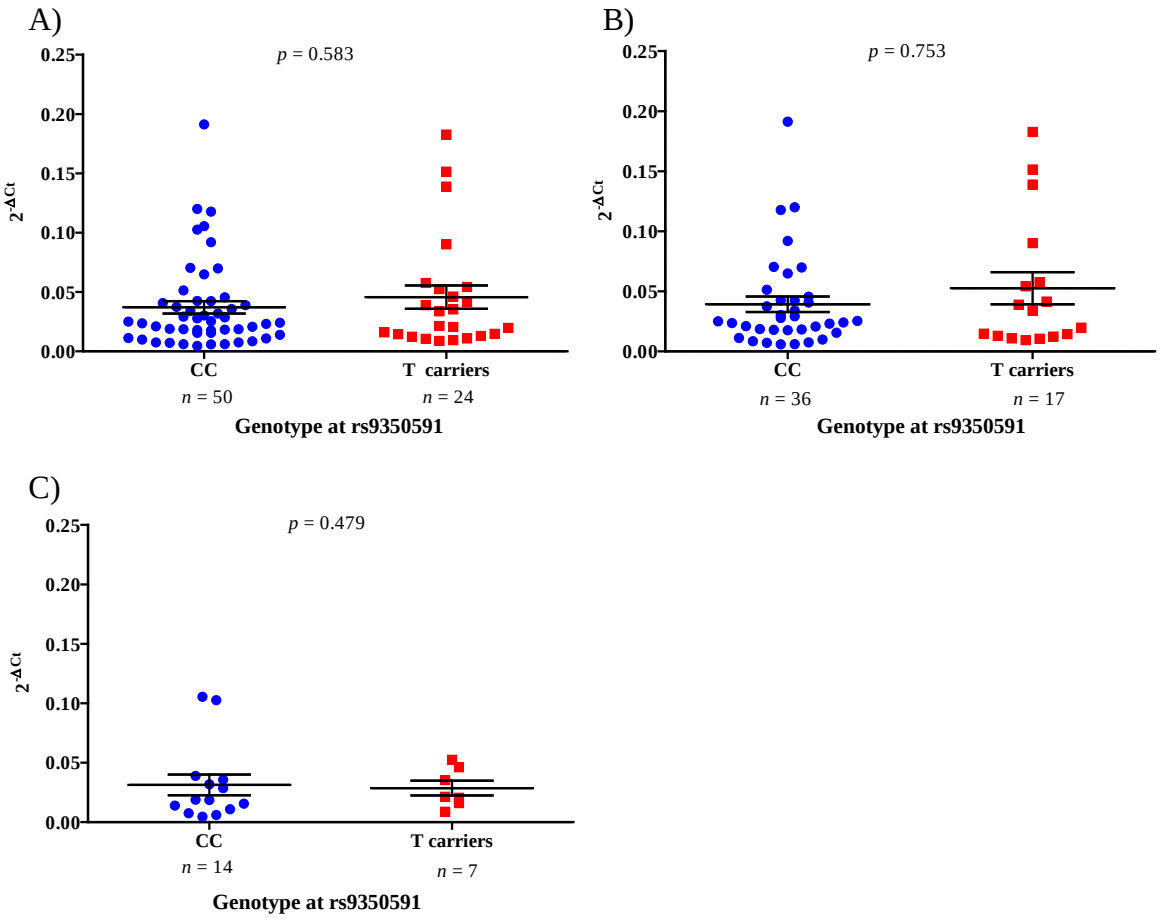

Figure 3.15. Analysis of *COX7A2* expression in OA hip and OA knee cartilage. There were no significant differences in gene expression between the genotypic groups of rs9350591 in A) all OA cartilage (hip and knee combined), B) OA knee cartilage, or C) OA hip cartilage. Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group; error bars represent the mean ± SEM.

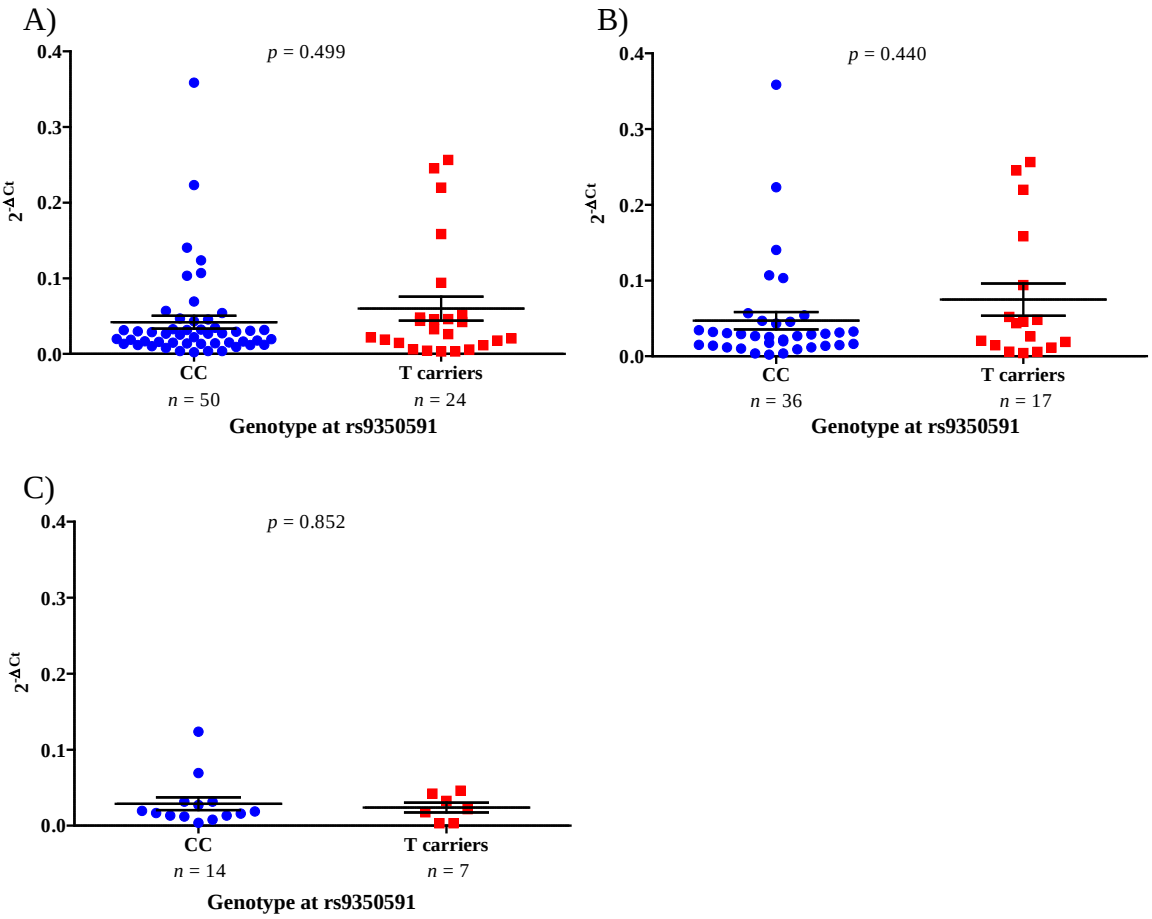

Figure 3.16. Analysis of *COL12A1* expression in OA hip and OA knee cartilage. There were no significant differences in gene expression between the genotypic groups of rs9350591 in A) all OA cartilage (hip and knee combined), B) OA knee cartilage, or C) OA hip cartilage. Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group; error bars represent the mean ± SEM.

3.3.8 *Replication of the gene expression quantification experiments in an independent group of OA hip cartilage samples*

As rs9350591 marks a region of association in hip OA, I replicated the gene expression qPCR in an independent group of OA hip cartilage donors to confirm the findings presented in Chapter 3.3.7. It was not appropriate to combine these data with the corresponding data of Chapter 3.3.7 as there were no common samples in the cohorts to which the data could be normalised. As observed previously, none of the differences were statistically significant (Figure 3.17).

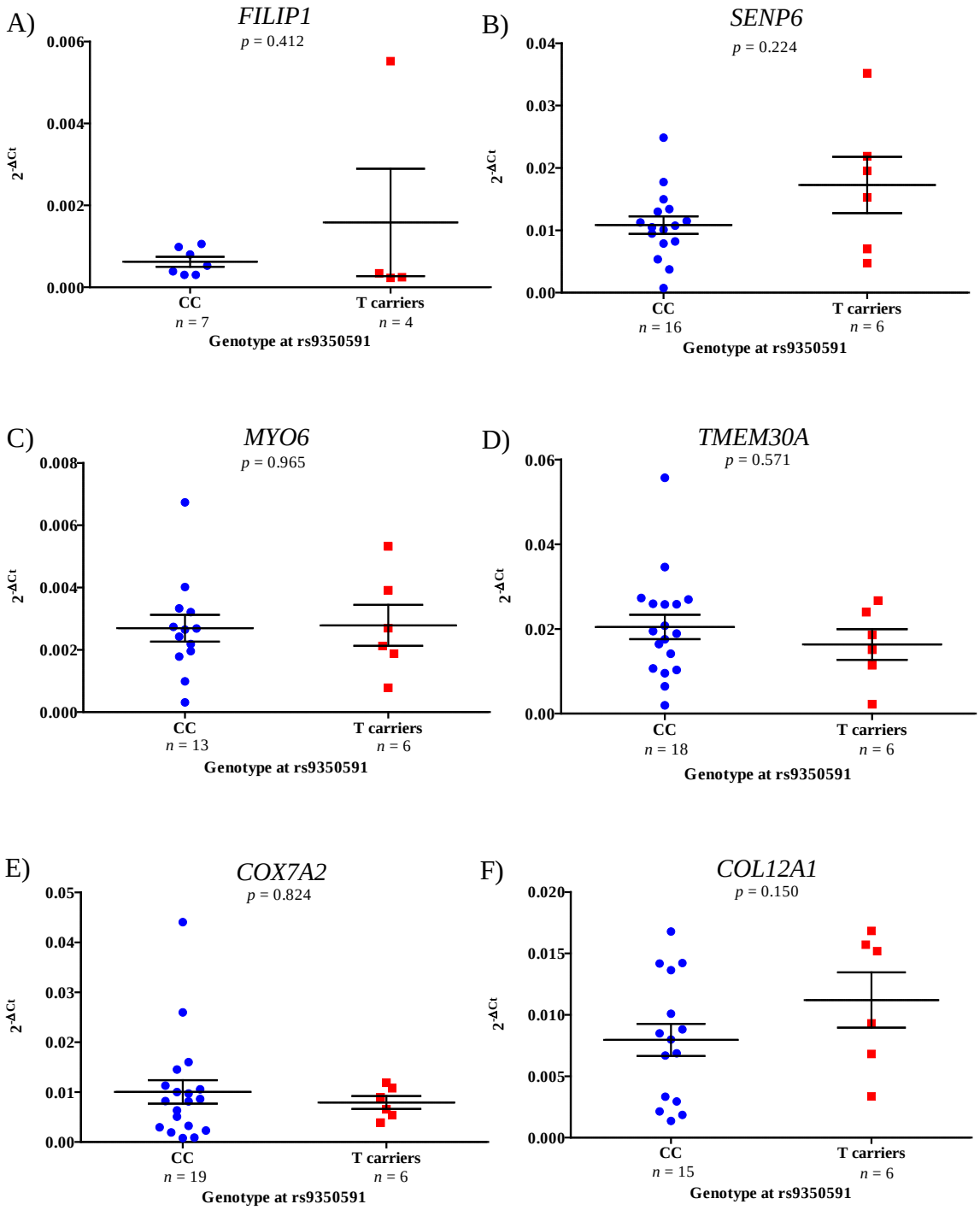

Figure 3.17. Independent replication of gene expression in OA hip cartilage. Expressions of A) *FILIP1*, B) *SENP6*, C) *MYO6*, D) *TMEM30A*, E) *COX7A2* and F) *COL12A1* were quantified in OA hip cartilage and normalised to the housekeeping genes *18S*, *GAPDH* and *HPRT1*. Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group; error bars represent the mean ± SEM.

3.3.9 *Characterising the expression profiles of FILIP1, SENP6, MYO6, TMEM30A, COX7A2 and COL12A1 in OA fat pad and OA synovium: comparisons of sex and age*

From the hip and knee tissues supplied after joint replacement surgery, our laboratory was able to collect cartilage, fat pad and synovium from knee samples and cartilage from hip

samples. Although originating from patients with OA of the knee, gene expressions in fat pad and synovium were investigated for this hip OA-associated locus. These tissues are also present around the hip joint and contribute to normal joint function (Okada *et al.*, 1989; Manaster, 2000). By quantifying such gene expressions, it might be possible to detect differences that could be translated into hip OA susceptibility. I first compared levels of gene expression between female and male donors and then between ages (Table 3.12 – Table 3.17). Through these analyses, there were no significant differences of expression in fat pad or synovium.

Tissue	Stratification for *FILIP1* qPCR data	*p* value
FP	Female OA (*n* = 15) vs male OA (*n* = 10)	0.390
FP	All OA age: 50 (*n* = 3) vs 60 (*n* = 11) vs 70 (*n* = 9) vs 80 (*n* = 2)	0.165
Sy	Female OA (*n* = 9) vs male OA (*n* = 10)	0.905
Sy	All OA age: 50 (*n* = 6) vs 60 (*n* = 3) vs 70 (*n* = 8) vs 80 (*n* = 2)	0.956

Table 3.12. Analysis of *FILIP1* expression in OA fat pad and synovium. There were no significant differences in gene expression for either of the stratifications in fat pad (FP) or synovium (Sy). Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons and a one-way analysis of variance for a comparison of more than two groups; *n* represents the number of individuals in the comparison group.

Tissue	Stratification for *SENP6* qPCR data	*p* value
FP	Female OA (*n* = 15) vs male OA (*n* = 7)	0.833
FP	All OA age: 50 (*n* = 2) vs 60 (*n* = 9) vs 70 (*n* = 9) vs 80 (*n* = 2)	0.279
Sy	Female OA (*n* = 11) vs male OA (*n* = 10)	0.916
Sy	All OA age: 50 (*n* = 6) vs 60 (*n* = 5) vs 70 (*n* = 9) vs 80 (*n* = 1)	0.636

Table 3.13. Analysis of *SENP6* expression in OA fat pad and synovium. There were no significant differences in gene expression for either of the stratifications in fat pad (FP) or synovium (Sy). Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons and a one-way analysis of variance for a comparison of more than two groups; *n* represents the number of individuals in the comparison group.

Tissue	Stratification for *MYO6* qPCR data	*p* value
FP	Female OA (*n* = 15) vs male OA (*n* = 7)	0.778
FP	All OA age: 50 (*n* = 2) vs 60 (*n* = 9) vs 70 (*n* = 9) vs 80 (*n* = 2)	0.996
Sy	Female OA (*n* = 10) vs male OA (*n* = 10)	0.436
Sy	All OA age: 50 (*n* = 5) vs 60 (*n* = 4) vs 70 (*n* = 9) vs 80 (*n* = 2)	0.454

Table 3.14. Analysis of *MYO6* expression in OA fat pad and synovium. There were no significant differences in gene expression for either of the stratifications in fat pad (FP) or synovium (Sy). Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons and a one-way analysis of variance for a comparison of more than two groups; *n* represents the number of individuals in the comparison group.

Tissue	Stratification for *TMEM30A* qPCR data	*p* value
FP	Female OA (*n* = 14) vs male OA (*n* = 10)	0.573
FP	All OA age: 50 (*n* = 3) vs 60 (*n* = 11) vs 70 (*n* = 8) vs 80 (*n* = 2)	0.165
Sy	Female OA (*n* = 11) vs male OA (*n* = 10)	0.751
Sy	All OA age: 50 (*n* = 6) vs 60 (*n* = 4) vs 70 (*n* = 9) vs 80 (*n* = 2)	0.659

Table 3.15. Analysis of *TMEM30A* expression in OA fat pad and synovium. There were no significant differences in gene expression for either of the stratifications in fat pad (FP) or synovium (Sy). Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons and a one-way analysis of variance for a comparison of more than two groups; *n* represents the number of individuals in the comparison group.

Tissue	Stratification for *COX7A2* qPCR data	*p* value
FP	Female OA (*n* = 14) vs male OA (*n* = 5)	0.105
FP	All OA age: 50 (*n* = 2) vs 60 (*n* = 8) vs 70 (*n* = 7) vs 80 (*n* = 2)	0.245
Sy	Female OA (*n* = 11) vs male OA (*n* = 10)	0.597
Sy	All OA age: 50 (*n* = 6) vs 60 (*n* = 5) vs 70 (*n* = 9) vs 80 (*n* = 1)	0.829

Table 3.16. Analysis of *COX7A2* expression in OA fat pad and synovium. There were no significant differences in gene expression for either of the stratifications in fat pad (FP) or synovium (Sy). Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons and a one-way analysis of variance for a comparison of more than two groups; *n* represents the number of individuals in the comparison group.

Tissue	Stratification for *COL12A1* qPCR data	*p* value
FP	Female OA (*n* = 14) vs male OA (*n* = 10)	0.977
FP	All OA age: 50 (*n* = 3) vs 60 (*n* = 11) vs 70 (*n* = 8) vs 80 (*n* = 2)	0.400
Sy	Female OA (*n* = 10) vs male OA (*n* = 10)	0.190
Sy	All OA age: 50 (*n* = 6) vs 60 (*n* = 4) vs 70 (*n* = 8) vs 80 (*n* = 2)	0.655

Table 3.17. Analysis of *COL12A1* expression in OA fat pad and synovium. There were no significant differences in gene expression for either of the stratifications in fat pad (FP) or synovium (Sy). Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons and a one-way analysis of variance for a comparison of more than two groups; *n* represents the number of individuals in the comparison group.

3.3.10 *Characterising the expression profiles of FILIP1, SENP6, MYO6, TMEM30A, COX7A2 and COL12A1 in OA fat pad and OA synovium: comparisons of rs9350591 genotype*

As for the analyses in cartilage, the stratification necessary to investigate if an eQTL is operating at this locus is the comparison of gene expression between the genotypic groups of rs9350591. Again, T allele carriers (TT and CT genotypes) were grouped together and compared to CC homozygotes. From these data, it was not possible to identify the actions of an eQTL influencing the expression of any of the genes tested. However, this interpretation must be with caution as i) the *n* numbers are low compared to previous qPCR analyses, such that statistical tests were often not possible, and ii) rs9350591 is specifically an OA-associated locus in the hip stratum of the arcOGEN study, whereas these tissues originate from the joints of patients with knee OA.

Tissue	Stratification for *FILIP1* qPCR data	*p* value
FP	All OA CC (*n* = 19) vs all OA T carriers (*n* = 6)	0.824
FP	OA female CC (*n* = 12) vs OA female T carriers (*n* = 2)	-
FP	OA male CC (*n* = 6) vs OA male T carriers (*n* = 4)	1.000
Sy	All OA CC (*n* = 16) vs all OA T carriers (*n* = 3)	0.955
Sy	OA female CC (*n* = 7) vs OA female T carriers (*n* = 2)	-
Sy	OA male CC (*n* = 9) vs OA male T carriers (*n* = 1)	-

Table 3.18. Analysis of *FILIP1* expression in OA fat pad and synovium. There were no significant differences in gene expression for any of the stratifications in fat pad (FP) or synovium (Sy). Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group. Statistical tests were not performed on datasets with $\leq$2 individuals.

Tissue	Stratification for *SENP6* qPCR data	*p* value
FP	All OA CC (*n* = 17) vs all OA T carriers (*n* = 5)	0.060
FP	OA female CC (*n* = 12) vs OA female T carriers (*n* = 2)	-
FP	OA male CC (*n* = 4) vs OA male T carriers (*n* = 3)	0.400
Sy	All OA CC (*n* = 18) vs all OA T carriers (*n* = 3)	0.880
Sy	OA female CC (*n* = 9) vs OA female T carriers (*n* = 2)	-
Sy	OA male CC (*n* = 9) vs OA male T carriers (*n* = 1)	-

Table 3.19. Analysis of *SENP6* expression in OA fat pad and synovium. There were no significant differences in gene expression for any of the stratifications in fat pad (FP) or synovium (Sy). Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group. Statistical tests were not performed on datasets with ≤2 individuals.

Tissue	Stratification for *MYO6* qPCR data	*p* value
FP	All OA CC (*n* = 17) vs all OA T carriers (*n* = 5)	0.754
FP	OA female CC (*n* = 13) vs OA female T carriers (*n* = 2)	-
FP	OA male CC (*n* = 4) vs OA male T carriers (*n* = 3)	0.400
Sy	All OA CC (*n* = 17) vs all OA T carriers (*n* = 3)	0.459
Sy	OA female CC (*n* = 8) vs OA female T carriers (*n* = 2)	-
Sy	OA male CC (*n* = 9) vs OA male T carriers (*n* = 1)	-

Table 3.20. Analysis of *MYO6* expression in OA fat pad and synovium. There were no significant differences in gene expression for any of the stratifications in fat pad (FP) or synovium (Sy). Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group. Statistical tests were not performed on datasets with ≤2 individuals.

Tissue	Stratification for *TMEM30A* qPCR data	*p* value
FP	All OA CC (*n* = 18) vs all OA T carriers (*n* = 6)	0.583
FP	OA female CC (*n* = 12) vs OA female T carriers (*n* = 2)	-
FP	OA male CC (*n* = 6) vs OA male T carriers (*n* = 4)	0.629
Sy	All OA CC (*n* = 18) vs all OA T carriers (*n* = 3)	0.514
Sy	OA female CC (*n* = 9) vs OA female T carriers (*n* = 2)	-
Sy	OA male CC (*n* = 9) vs OA male T carriers (*n* = 1)	-

Table 3.21. Analysis of *TMEM30A* expression in OA fat pad and synovium. There were no significant differences in gene expression for any of the stratifications in fat pad (FP) or synovium (Sy). Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group. Statistical tests were not performed on datasets with ≤2 individuals.

Tissue	Stratification for *COX7A2* qPCR data	*p* value
FP	All OA CC (*n* = 14) vs all OA T carriers (*n* = 5)	0.247
FP	OA female CC (*n* = 12) vs OA female T carriers (*n* = 2)	-
FP	OA male CC (*n* = 2) vs OA male T carriers (*n* = 3)	-
Sy	All OA CC (*n* = 18) vs all OA T carriers (*n* = 3)	0.725
Sy	OA female CC (*n* = 9) vs OA female T carriers (*n* = 2)	-
Sy	OA male CC (*n* = 9) vs OA male T carriers (*n* = 1)	-

Table 3.22. Analysis of *COX7A2* expression in OA fat pad and synovium. There were no significant differences in gene expression for any of the stratifications in fat pad (FP) or synovium (Sy). Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group. Statistical tests were not performed on datasets with ≤2 individuals.

Tissue	Stratification for *COL12A1* qPCR data	*p* value
FP	All OA CC (*n* = 18) vs all OA T carriers (*n* = 6)	0.665
FP	OA female CC (*n* = 12) vs OA female T carriers (*n* = 2)	-
FP	OA male CC (*n* = 6) vs OA male T carriers (*n* = 4)	0.914
Sy	All OA CC (*n* = 17) vs all OA T carriers (*n* = 3)	0.597
Sy	OA female CC (*n* = 8) vs OA female T carriers (*n* = 2)	-
Sy	OA male CC (*n* = 9) vs OA male T carriers (*n* = 1)	-

Table 3.23. Analysis of *COL12A1* expression in OA fat pad and synovium. There were no significant differences in gene expression for any of the stratifications in fat pad (FP) or synovium (Sy). Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group. Statistical tests were not performed on datasets with ≤2 individuals.

3.3.11 *Investigating the effect of the rs9350591 association signal on the allelic output of the transcripts of FILIP1, SENP6, MYO6, TMEM30A, COX7A2 and COL12A1*

In the overall gene expression analysis discussed in the preceding sections of this chapter, I have observed no evidence for rs9350591 marking a *cis*-eQTL at this locus in any of the end-stage OA tissue samples tested. However, inter-individual variability could be masking the effect of the polymorphism, with differences in the output of the gene transcript in fact being more subtle. As this would be detected with a more specific investigatory approach, I used pyrosequencing to quantify the allelic output of the gene transcripts. In order to do this for the intergenic association signal, transcript SNPs for each of the genes were required as markers of messenger RNA (mRNA) output (Figure 3.18).

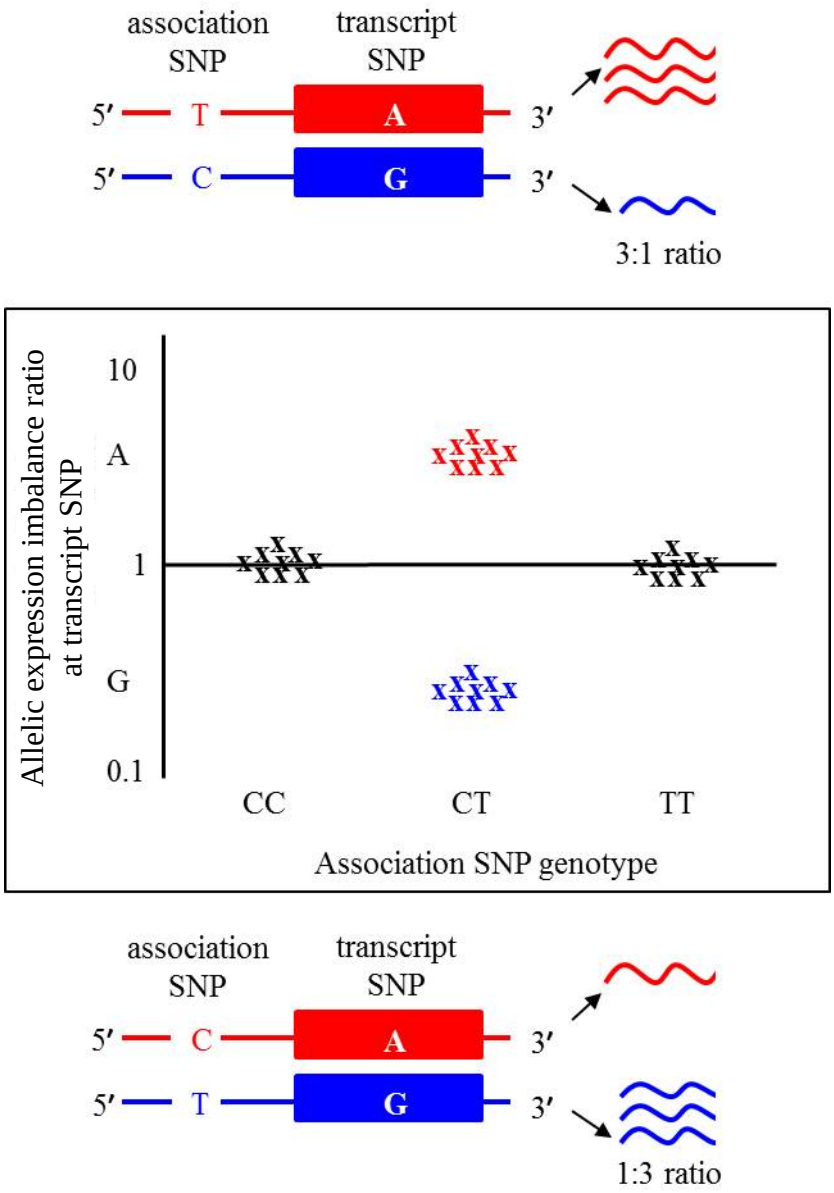

Figure 3.18. Schematic diagram of the expected results if there is a correlation between association SNP genotype and the allelic expression imbalance of a gene when the polymorphisms are not in high linkage disequilibrium. As the frequency of recombination between the two polymorphisms is high, there are two haplotypes that can occur in compound heterozygotes: in this case, TA/CG or TG/CA. If the intergenic polymorphism is acting as an eQTL, a resulting imbalance of allelic transcripts would be detected. Whether this would be observed as a greater transcription of the A allele or G allele of the transcript SNP would be dependent on the haplotype of each individual. Clustering would become apparent for compound heterozygotes, which would indicate the actions of an eQTL.

Each gene was searched for transcript SNPs using the UCSC Genome Browser (Kent *et al.*, 2002) and a SNAP Pairwise LD (Johnson *et al.*, 2008) search was performed for all of the polymorphisms to assess the degree of correlation with rs9350591 (Chapter 2.9). The heterozygote frequencies were obtained from the dbSNP online database (Sherry *et al.*, 2001):

AEI analysis requires transcript SNP heterozygotes in order to investigate differential allelic outputs. *FILIP1* had two known transcript SNPs, however the heterozygote frequencies were < 5%, meaning that acquiring a sufficient number of heterozygotes to allow for a robust analysis was not feasible (Table 3.24). In addition, there were no known transcript polymorphisms within *COX7A2*. There were no transcript SNPs in high LD ($r^2 \geq 0.80$) with rs9350591 in the remaining four genes. Therefore, in order to maximise the sample sizes, two transcript SNPs for each of the four genes were selected (Table 3.25 – Table 3.28). All donors were genotyped at the association SNP using an RFLP assay (Chapter 2.10 and Chapter 2.11), and transcript SNPs were genotyped using pyrosequencing (Chapter 2.12).

SNP	r^2 relative to rs9350591	D' relative to rs9350591	Heterozygote frequency (%)	Genetic location
rs112839775	0.005	1.000	4.88	Intron/3′ UTR
rs62415695	0.003	1.000	3.34	Exon

Table 3.24. Transcript polymorphisms within *FILIP1*. Both SNPs have heterozygote frequencies < 5% and so were not used for allelic expression imbalance analysis. The multiple genetic loci for rs112839775 reflects the position on different transcript isoforms.

SNP	r^2 relative to rs9350591	D' relative to rs9350591	Heterozygote Frequency (%)	Genetic location
rs17414086	0.085	1.000	45.01	Exon
rs17414687	0.085	1.000	45.01	Exon
rs9250	0.085	1.000	44.42	Exon
rs7385	0.225	0.724	42.47	3′ UTR
rs276683	0.225	0.724	41.35	3′ UTR
rs276684	0.035	1.000	27.82	3′ UTR
rs71561434	0.746	0.925	23.10	5′ UTR
rs507662	0.145	1.000	4.30	Exon
rs16886792	0.048	1.000	1.59	Exon

Table 3.25. Transcript polymorphisms within *SENP6*. rs17414687 and rs71561434 were selected as markers for mRNA output in AEI analysis.

SNP	r^2 relative to rs9350591	D' relative to rs9350591	Heterozygote Frequency (%)	Genetic location
rs699186	0.079	1.000	45.91	Exon
rs12606	0.076	1.000	42.08	Intron/3′ UTR
rs7746476	0.036	0.237	32.95	3′ UTR
rs1045758	0.426	0.715	28.90	3′ UTR
rs11756446	0.02	1.000	18.00	Exon
rs9360957	0.024	0.226	13.88	3′ UTR
rs7741414	0.005	1.000	4.88	3′ UTR
rs2273857	0.048	1.000	3.34	Exon

Table 3.26. Transcript polymorphisms within *MYO6*. rs699186 and rs1045758 were selected as markers for mRNA output in AEI analysis. The multiple genetic loci for rs12606 reflects the position on different transcript isoforms.

SNP	r^2 relative to rs9350591	D' relative to rs9350591	Heterozygote Frequency (%)	Genetic location
rs240375	0.012	0.202	43.08	5′ UTR
rs240374	0.041	0.255	15.39	Exon
rs41269315	0.110	0.522	12.50	3′ UTR
rs414624	0.001	0.428	11.46	3′ UTR
rs117512	0.001	0.428	10.22	3′ UTR
rs654428	0.001	0.428	9.50	3′ UTR
rs397039	0.001	0.428	8.41	3′ UTR
rs381141	0.001	0.428	8.41	3′ UTR
rs638590	0.001	0.428	8.41	3′ UTR
rs45596238	0.006	0.872	8.05	3′ UTR
rs15616	0.001	0.428	6.38	3′ UTR

Table 3.27. Transcript polymorphisms within *TMEM30A*. rs41269315 and rs240375 were selected as markers for mRNA output in AEI analysis.

SNP	r^2 relative to rs9350591	D' relative to rs9350591	Heterozygote Frequency (%)	Genetic location
rs240736	0.000	0.020	39.88	Exon
rs970547	0.000	0.009	33.98	Exon
rs594012	0.018	1.000	16.70	Exon
rs35523808	0.006	1.000	6.40	Exon
rs560250	No data	No data	0.70	3′ UTR

Table 3.28. Transcript polymorphisms within *COL12A1*. rs240736 and rs594012 were selected as markers for mRNA output in AEI analysis.

Pyrosequencing uses a third primer in addition to the forward and reverse primers of a standard PCR. The sequencing primer binds nearby the polymorphism and is extended over the SNP toward the direction of the captured biotinylated primer. It is necessary to validate the ability of sequencing primers to distinguish between allelic ratios at specified polymorphisms. This is achieved by combining the DNA of major and minor allele homozygotes in order to generate known allelic ratios before comparing the values detected experimentally to the expected outcome (Appendix B: Figure B.1). The low MAFs for rs594012 (*COL12A1*), rs41269315 (*TMEM30A*) and rs71561434 (*SENP6*) meant that the generation of allelic ratios was limited to heterozygotes and major allele homozygotes, that is, ratios containing $\leq$ 50% of the minor allele. All of the validations had a positive correlation between observed and expected ratios, each with a goodness of fit $r^2 \geq 0.90$, and therefore were considered suitable for AEI analysis. Reactions were performed in triplicate and the mean of the cDNA allelic ratios was normalised to the corresponding gDNA ratios. There was excessive variability in the cDNA technical replicates for rs240375 (*TMEM30A*) in addition to failing quality control checks, and as a result this assay was removed from the investigation. Primer sequences for the remaining assays can be found in Appendix A: Table A.1, and the assay positions detailed in Figure 3.19.

A GWAS performed in an Icelandic population found that a region within *ALDH1A2* was significantly associated with hand OA (Styrkarsdottir *et al.*, 2014). Despite examining meta-analysis data to investigate the association of the locus with hip and knee OA, the association remained unique to hand OA. However, in hip and knee cartilage, the risk allele of the association SNP rs3204689 was transcribed at lower levels relative to the non-risk allele. Thus, despite being specifically associated with hand OA, AEI was observed in knee and hip cartilage. Accordingly, I analysed hip OA, knee OA and NOF samples collectively for AEI correlations with rs9350591 genotype. The low MAF of the association SNP meant that minor allele homozygotes were scarce, and as such, the analyses were restricted to rs9350591 CC and CT individuals.

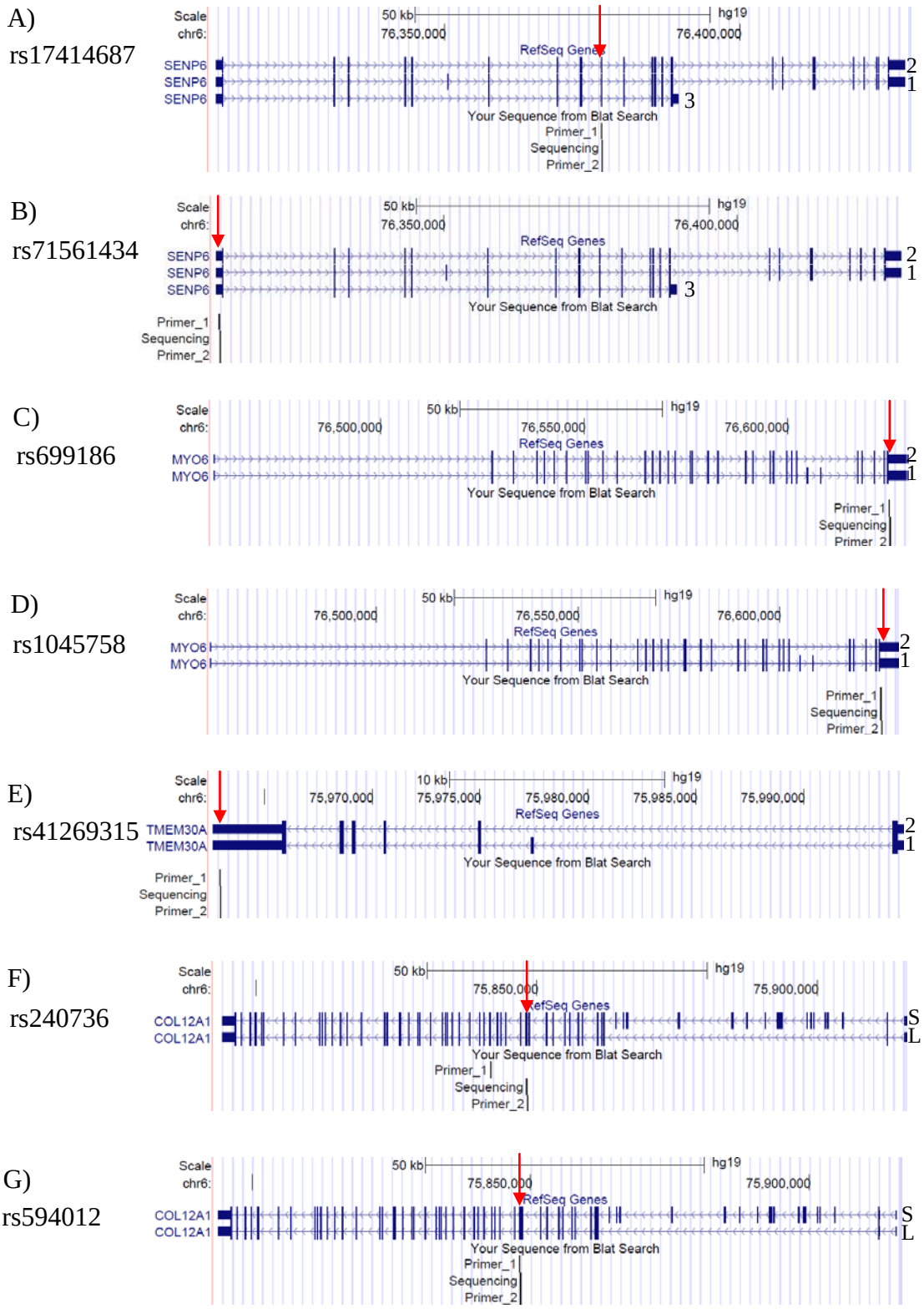

Figure 3.19. Location of pyrosequencing primers used for allelic expression analysis. Assays were designed to A) rs17414687 of *SENP6*, B) rs71561434 of *SENP6*, C) rs699186 of *MYO6*, D) rs1045758 of *MYO6*, E) rs41269315 of *TMEM30A*, F) rs240736 of *COL12A1* and G) rs594012 of *COL12A1*. Every transcript of each gene was covered by the assays. Red arrows (→) indicate transcript SNP positions. Transcript isoforms are numbered for each gene; *COL12A1* isoforms are annotated short (S) and long (L). The images were obtained using the hg19 reference genome.

There was no evidence to indicate rs9350591 correlates with a *cis*-eQTL acting on *SENP6* (Figure 3.20.A), *MYO6* (Figure 3.20.B), *TMEM30A* (Figure 3.20.C) and *COL12A1* (Figure 3.20.D) in cartilage. However, an eQTL was operating on the *COL12A1* transcript, with more of the minor allele transcribed relative to the major allele. An eQTL was also operating on the *MYO6* transcript causing more of the major allele to be transcribed relative to the minor allele: this significance is likely to be caused by the large spread of data observed in the CC homozygotes. As neither of the imbalances fit the model detailed in Figure 3.18, it is likely that they are not relevant to the OA association signal at this locus, but instead are mediated by another eQTL operating here.

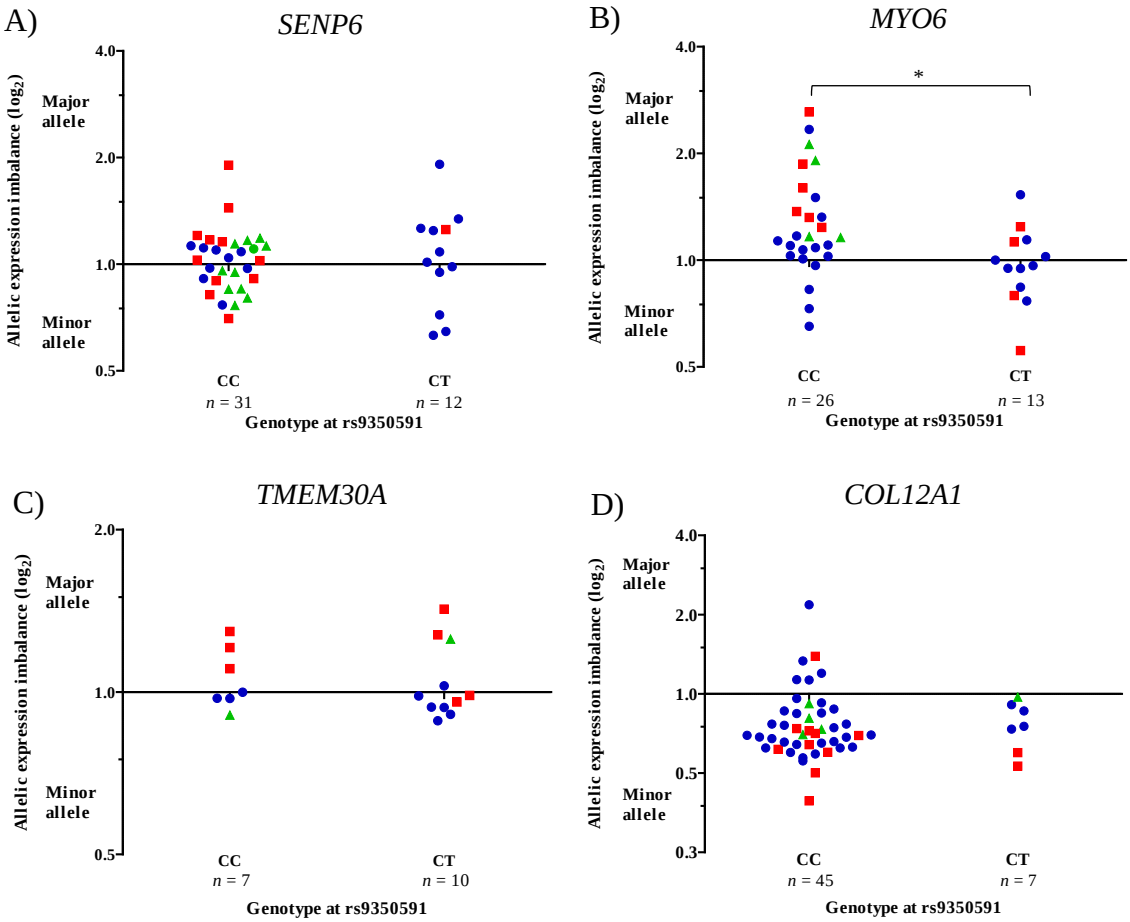

Figure 3.20. Allelic expression imbalance of *SENP6*, *MYO6*, *TMEM30A* and *COL12A1* in hip and knee cartilage stratified by rs9350591 genotype. There was no distinct clustering of compound heterozygote donors that would imply a correlation between rs9350591 genotype and an imbalance of mRNA output of A) *SENP6*, B) *MYO6*, C) *TMEM30A* or D) *COL12A1*. An eQTL was operating on *MYO6* causing a greater output of the major allele transcript, while an eQTL was causing a greater output of the minor allele transcript of *COL12A1*. These did not correlate with association SNP genotype, as the imbalances occurred in both rs9350591 genotype groups. OA knee (blue circles), OA hip (red squares) and NOF (green triangles) cartilage. Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group; * $p \leq 0.05$.

3.3.12 *Analysis of methylation levels at CpG sites surrounding the 6q14.1 locus in hip and knee cartilage*

An underlying cause of the OA association marked by rs9350591 could be differential methylation of the surrounding genomic region. Epigenetic regulation has been shown to be tissue-specific (Davies *et al.*, 2012) and variations in genomic sequences are known to contribute to differences in methylation profiles (Bell *et al.*, 2011). It is therefore possible that the association signal could impact upon the methylation levels of the surrounding region, perhaps by affecting the binding of DNA methyltransferases. In turn, differential methylation may either directly impede or strengthen binding of transcription factors, or it could lead to changes in chromatin formation through the recruitment of chromatin remodelling proteins. The result of such changes would be observed as differences in target gene expression, potentially as a 'signature' of a previous effect earlier in joint development. As I have shown thus far, however, there is no correlation between rs9350591 genotype and nearby gene expression. Nevertheless, an analysis of the epigenetic profile of the region surrounding rs9350591 would allow for a more detailed characterisation of the OA-associated locus. Particularly, if methylation is affecting gene expression, this may have been overlooked in the allelic expression analysis of Chapter 3.3.11, for example, as *FILIP1* and *COX7A2* were omitted from the investigation. As such, I acquired data generated by Dr Michael Rushton, a member of Prof. Loughlin's research group (Institute of Cellular Medicine, Newcastle University), whereby genome-wide methylation levels were assessed on an Illumina Infinium HumanMethylation450 BeadChip array (Rushton *et al.*, 2014). I extracted the data for all 164 CpG sites that were annotated within 2 Mb of rs9350591: ranging from cg20428196 (1,064,579 bp downstream of rs9350591) to cg15162000 (1,023,383 bp upstream of rs9350591). For every CpG site, the cartilage methylation profiles of 17 hip OA (14 CC, 3 T carriers), 63 knee OA (50 CC, 13 T carriers) and 35 NOF (29 CC, 6 T carriers) donors were included in the various comparisons that I performed. To correct for multiple comparisons, a new significance threshold was calculated (Equation 3.1).

$$\text{significance threshold} = \frac{p \text{ value}}{\text{number of CpG sites}} = \frac{0.05}{164} = \mathbf{0.000305}$$

Equation 3.1. Bonferroni correction used to counteract the multiple tests performed for the methylation microarray analysis. A new significance threshold of 0.000305 was calculated: *p* values must be less than this to be considered statistically significant.

Specifically, as rs9350591 is a hip OA locus, it is these samples that are of particular relevance to the characterisation of the 6q14.1 locus. When comparing the average levels of methylation in OA hip donors with NOF donors, 13 CpG sites remained significantly different between the two groups after Bonferroni correction, eight of which had a difference in methylation $\geq$ 5% between the two groups (Figure 3.21.A). Subsequently, the average levels of methylation for the 164 CpG sites were compared between hip OA CC and OA hip T carriers (Figure 3.21.B). In this case, cg26466508 was the only CpG site to remain significantly different between the two groups following Bonferroni correction and, as in the hip OA versus NOF comparison, had a difference in methylation $\geq$ 5% between the two groups.

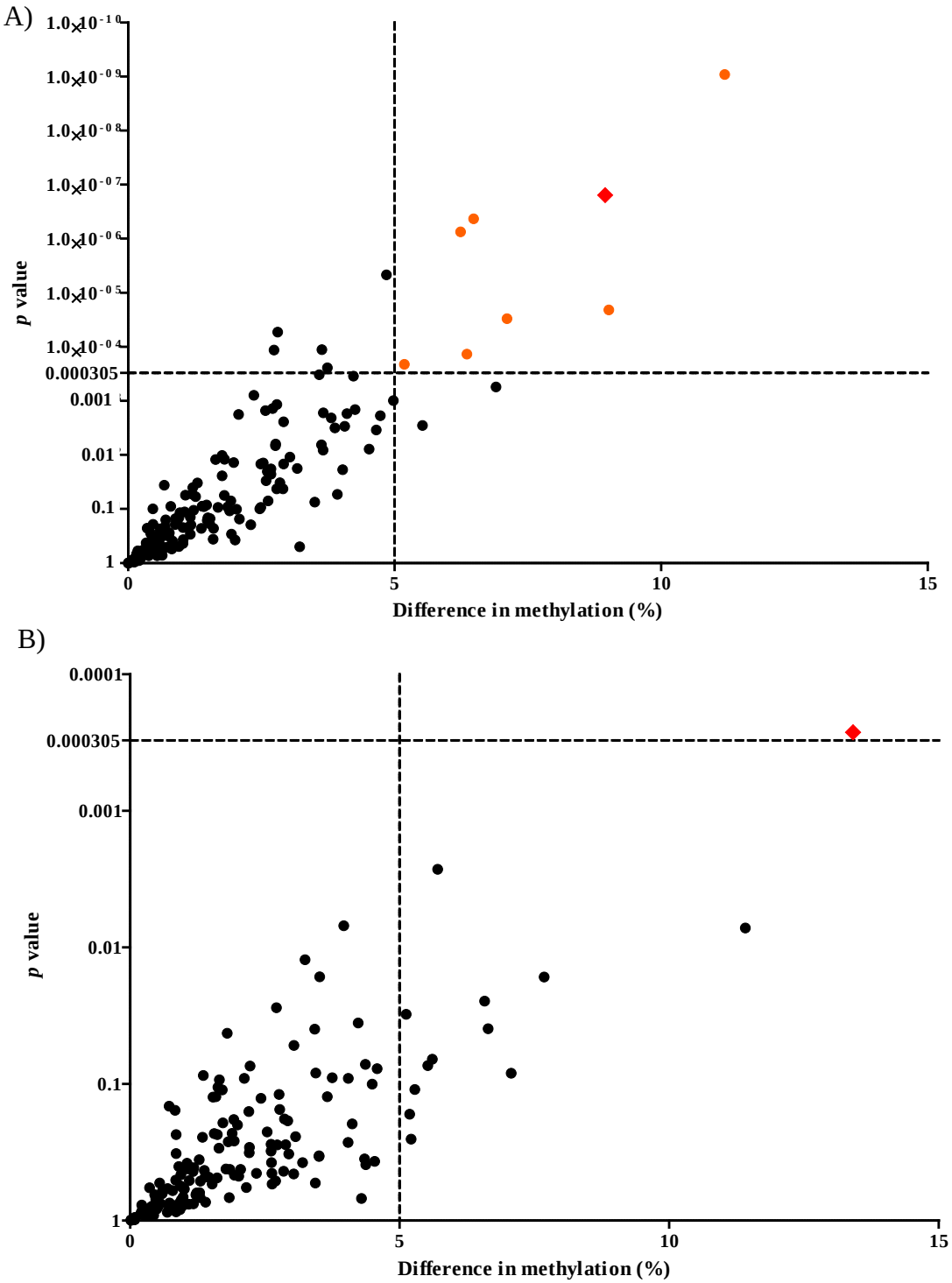

Figure 3.21. Scatter plots to compare the levels of methylation in OA hip and NOF donors. One hundred and sixty four CpG sites, all within 2 Mb of rs9350591, were analysed for differential methylation in A) hip OA vs NOF donors and B) hip OA CC vs hip OA T carriers. Eight of the thirteen CpG sites that remained significant after Bonferroni correction also differed by $\geq$ 5%, marked by orange circles. One of these CpG sites was cg26466508, marked by a red diamond. The methylation levels of cg26466508 were also significantly different between hip OA CC and hip OA T carriers. Statistical significance was assessed using the Student's *t* test. Horizontal dotted lines represent the significance threshold after Bonferroni correction.

Due to a lack of overlapping data, it was not possible to investigate any correlation between methylation at cg26466508 and the expressions of the genes at this locus. However, a database search as performed in Figure 3.1 indicated that cg26466508 resides in a functional region. In fact, the CpG site is 53 bp upstream of *FILIP1* in a region predicted to have enhancer activity with transcription factor binding. In addition, cg26466508 is 942 bp upstream of the *FILIP1* intronic SNP rs9343292 and 16,202 bp downstream of the intergenic SNP rs11964634: both SNPs are in perfect LD with rs9350591.

In OA hip cartilage, cg26466508 was significantly ($p \leq 0.000305$) hypermethylated relative to NOF cartilage (Table 3.29). Statistical significance did not endure following a Bonferroni correction for methylation levels at this CpG site in OA hip cartilage relative to OA knee cartilage, nor for NOF cartilage relative to OA knee cartilage. When stratified by rs9350591 genotype, cg26466508 was hypermethylated in the cartilage of T carriers relative to the respective CC individuals for OA hip cartilage: this was not statistically significant in OA knee or NOF cartilage (Table 3.30). The statistically significant differences for which methylation levels differ by 5% or greater are shown in the graphical representations of Figure 3.22. Such differences imply that methylation may be key in regulating gene expression at this OA-associated locus: perhaps it is acting to regulate *FILIP1* expression, a result that could have been overlooked in the previous characterisation sections of this chapter.

Stratum	β value	Difference (%)	*p* value
OA knee (*n* = 63)	0.255	5.55	0.0004
OA hip (*n* = 17)	0.310		
OA knee (*n* = 63)	0.255	3.40	0.0008
NOF (*n* = 35)	0.221		
OA hip (*n* = 17)	0.310	**8.95**	**1.552×10^{-07}**
NOF (*n* = 35)	0.221		

Table 3.29. Analysis of the methylation profile of cg26466508 in OA hip, OA knee and NOF cartilage. The only analysis with differences in methylation greater than 5% that endured after Bonferroni correction ($p = 0.000305$) was OA hip cartilage relative to NOF cartilage (highlighted in **bold**). Statistical significance was assessed using the Student's *t* test.

Stratum	β value	Difference (%)	*p* value
OA hip, OA knee and NOF CC (*n* = 93)	0.245	3.90	0.004
OA hip, OA knee and NOF T carriers (*n* = 22)	0.284		
OA hip and NOF CC (*n* = 43)	0.239	6.22	0.008
OA hip and NOF T carriers (*n* = 9)	0.302		
OA hip CC (*n* = 14)	0.287	**13.41**	**2.656 x 10^{-04}**
OA hip T carriers (*n* = 3)	0.421		
OA knee CC (*n* = 50)	0.250	2.19	0.165
OA knee T carriers (*n* = 13)	0.272		
NOF CC (*n* = 29)	0.217	2.54	0.138
NOF T carriers (*n* = 6)	0.242		

Table 3.30. Analysis of the methylation profile of cg26466508 in OA hip, OA knee and NOF cartilage. The only analysis with differences in methylation greater than 5% that endured after Bonferroni correction ($p = 0.000305$) was OA hip cartilage CC relative to OA hip cartilage T carriers (highlighted in **bold**). Statistical significance was assessed using the Student's *t* test.

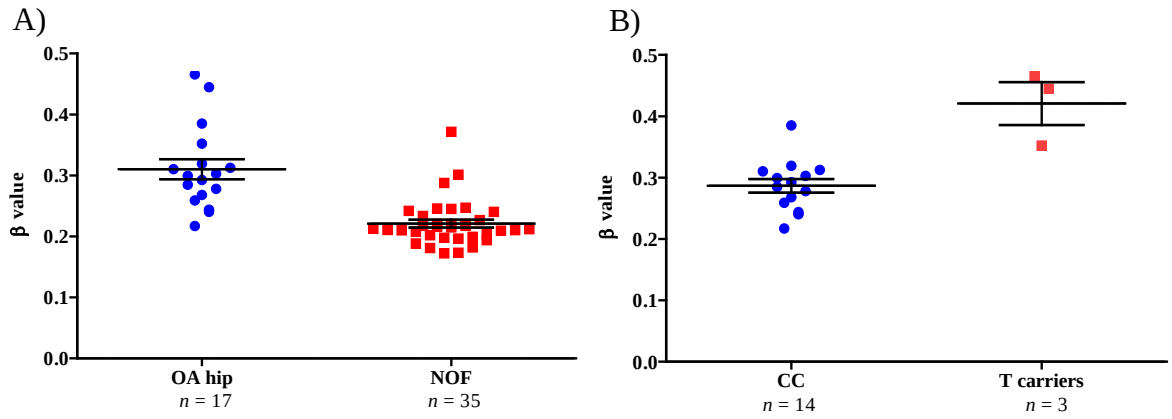

Figure 3.22. Statistically significant differences in the methylation profile of cg26466508 in OA hip, OA knee and NOF cartilage. Methylation at cg26466508 was significantly increased in A) OA hip cartilage relative to NOF cartilage and B) carriers of the risk allele of rs9350591 in OA hip cartilage only. Statistical significance was assessed using the Student's *t* test; *n* represents the number of individuals in the comparison group; error bars represent the mean ± SEM.

3.4 Discussion

The aim of the research in this chapter was to characterise the OA association locus marked by the polymorphism rs9350591. It is widely known that GWAS signals often mark the actions of other polymorphisms, commonly affecting the process of transcription or the stability of the resulting transcript (Montgomery and Dermitzakis, 2011). I hypothesised that rs9350591 marks a functional polymorphism that acts to regulate the expression of at least one of the genes at this locus. This investigation was to be achieved by interrogating the expression profiles of the six genes that reside within 1 Mb upstream and 1 Mb downstream of the SNP in order to identify if rs9350591 genotype correlates with differences in gene expression in synovial joints. It was not necessary to genotype any other polymorphisms. If the true causal SNP was in high enough LD with rs9350591 to cause its effect to be detected on the arcOGEN GWAS, then any correlations should also be detected through rs9350591 in these investigations (Styrkarsdottir *et al.*, 2014).

I first postulated that the association signal could be mediating its effects during joint development. I utilised microarray data that quantified gene expression during MSC differentiation down a chondrogenic lineage and an osteoblastic lineage, showing that all the genes were expressed throughout the time courses. Chondrogenesis was independently investigated using qPCR, and again confirmed the dynamic gene expression patterns, overall suggesting that any one of the genes have the potential to modulate OA susceptibility.

I then analysed data that were generated prior to the beginning of my Ph.D to confirm that the genes were expressed in primary articular cartilage. Additionally, I quantified the expressions in fat pad and synovium. Taken as a whole, *FILIP1* and *MYO6* had the lowest expression levels, while the remaining genes were all highly expressed in the joint tissues. The expression patterns bear no relevance to the known functions of any of the genes as discussed in Chapter 1, yet the results do support the evidence that the gene expressions are not isolated to specific tissue types. Once I had confirmed that all genes were expressed in the synovial joint tissues tested, I was able to begin to investigate the OA association signal.

The two main ways to investigate gene expression and genotype correlations are overall gene expression quantification and allelic expression quantification. The quantification of overall gene expression allows all samples, irrespective of genotype, to be used. It is, however, vulnerable to the natural fluctuation in gene expression between individuals, making it liable to false negatives. Since investigating the allelic expression is internally controlled (that is, the

allelic output of one allele is quantified relative to the other) and comparisons are within an individual, it is not affected by the variation between different donors. This approach, however, is limited by the heterozygote frequency of the transcript SNP under investigation, as only transcript SNP heterozygotes can be studied.

Previous overall gene expression studies have considered OA as a disease of the entire joint and have therefore investigated tissues other than solely cartilage (Raine *et al.*, 2012). Based on this, fat pad, synovium and cartilage excised from the knee of OA donors were included alongside hip cartilage analyses. In all tissue types, there were no significant differences in any of the overall gene expressions relative to the OA association SNP genotype. Nevertheless, *SENP6*, *MYO6*, *COX7A2* and *COL12A1* were all differentially expressed in cartilage depending on the disease state and joint site. This suggests that the genes could contribute to joint-specific OA development (Karlsson *et al.*, 2010) but are acting independently of the OA association signal detected by the arcOGEN study. For example, a downregulation of *SENP6* in OA hip cartilage relative to NOF cartilage could be associated with aberrant sumoylation (Chapter 1.10.3) and therefore affect ECM homeostasis. Alternatively, the changes could simply be a consequence of OA, rather than a cause, and so the exact mechanisms of action would need to be established.

The aim of AEI analysis is to identify if there are differences in the mRNA outputs that correspond to the different alleles of a heterozygote individual (Wang and Elbein, 2007). In this case, for example, even though the overall abundance of a gene may be comparable between individuals, an allelic imbalance within this may be apparent. Similar routes of investigation have previously been followed and have yielded positive results in the identification of *cis*-eQTLs (Bos *et al.*, 2012; Raine *et al.*, 2012; Gee *et al.*, 2014). Despite rs9350591 being associated with only hip OA, it is possible to detect AEI in more than one joint tissue (Egli *et al.*, 2009; Styrkarsdottir *et al.*, 2014; Gee *et al.*, 2015), and so knee cartilage was combined with hip cartilage is this study. However, following this in-depth approach of investigating AEI, there were still no significant differences in gene expression that could be attributed to rs9350591 genotype. The most notable outcome of this investigation was the significant difference between rs9350591 major allele homozygotes and minor allele carriers for the *MYO6* allelic outputs. This, however, can be accounted for by the large allelic ratio spread in the major allele homozygotes. In addition, an eQTL operating on the *COL12A1* gene transcript was also identified, causing more of the minor allele transcript to be produced relative to the major allele. Again, this did not correlate with rs9350591

genotype, and so it was concluded that there was no OA association with these genes in the cartilage samples tested. Importantly, it was not possible to interrogate *FILIP1* and *COX7A2* and so a correlation between their expressions and the OA association of this region cannot be ruled out.

DNA methylation is now an established mechanism through which the genome can be regulated and has been reported to correlate with DNA sequence variations (Bell *et al.*, 2011; Smith *et al.*, 2014). It would therefore be remiss not to explore this as a part of the characterisation of this locus. Indeed, analysis of data from a microarray that included 164 CpG sites across a 2 Mb span of the region in OA hip, OA knee and NOF cartilage yielded promising results. cg26466508 was hypermethylated in both OA hip relative to NOF and in OA hip rs9350591 T carriers relative to OA hip rs9350591 CC individuals. Moreover, the difference in methylation was greater than 5%, a threshold value considered to strengthen the biological relevance of detected signals (Hall *et al.*, 2014). Although the *n* number for T carriers was rather low, the significance survived Bonferroni correction and so can be considered robust. Perhaps the T allele of rs9350591 permits the recruitment of DNA methyltransferase enzymes that methylate cg26466508. In turn, the aberrant methylation could prevent transcription factor binding and thus regulate gene expression. This may indeed be the mechanism by which the *cis*-eQTL is modulating OA association, and it is because *FILIP1* and *COX7A2* could not be interrogated through AEI that it has been overlooked.

Despite a number of the genes at this locus being differentially expressed in cartilage, it is clear that the association signal does not modulate its effects in the end-stage OA synovial joint tissues tested. Nevertheless, the genes are all dynamically expressed throughout chondrogenesis and osteoblastogenesis, which imply a potential role for any of the genes at earlier stages of joint development. Moreover, differential methylation was observed between the rs9350591 genotype groups in OA hip cartilage at a CpG site that resides 53 bp upstream of *FILIP1*. Overall, this chapter highlights the huge amount of diversity at this OA association locus, and shows that there is scope to further investigate the functional region and its association to OA.

Chapter 4. Characterisation of the 12p11.22 Locus Marked by the Polymorphism rs10492367

4.1 Introduction

Another of the five genome-wide significant loci identified as being associated with OA in the arcOGEN GWAS was marked by the G/T polymorphism rs10492367 on chromosome 12p11.22 (arcOGEN Consortium *et al.*, 2012). The minor allele (T; MAF = 0.21) was significantly associated with hip OA in patients of European descent compared to the population-based controls, with an OR of 1.14 and a *p* value of 1.48×10^{-08} (Table 4.1). There was a weaker significance in the association for all cases of OA, with an OR of 1.06 and a *p* value of 9.02×10^{-04}.

Discovery		Replication		Discovery and replication	
OR (95% CI)	***p* value**	**OR (95% CI)**	***p* value**	**OR (95% CI)**	***p* value**
1.18 (1.11-1.27)	1.20×10^{-06}	1.11 (1.04-1.18)	1.18×10^{-03}	1.14 (1.09-1.20)	1.48×10^{-08}

Table 4.1. Association statistics from the arcOGEN GWAS for rs10492367 in the hip stratum only. Odds ratio (OR), confidence interval (CI), genome-wide significance *p* value = 5×10^{-08}. Adapted from (arcOGEN Consortium *et al.*, 2012).

The association interval, defined as the region in which all SNPs in an LD of ≥ 0.80 with rs10492367 reside, spans an intergenic region of 21 kb. There are no non-synonymous transcript polymorphisms that are in high LD ($r^2 \geq 0.80$) with rs10492367 and as such, it is unlikely that the OA association is mediated by a change in the coding sequence of any of the nearby genes. Instead, the causal SNP could be mediating its effect by regulating gene transcription. *PTHLH* and *KLHL42* expressions were quantified by Dr Madhushika Ratnayake and Dr Emma Raine of Prof. Loughlin's group (Institute of Cellular Medicine, Newcastle University) prior to the commencement of this project. The results confirmed that the genes were expressed in the following human tissues: articular cartilage, fat pad, synovium, meniscus, tendon and ligament (arcOGEN Consortium *et al.*, 2012). This chapter will therefore characterise the OA-associated region marked by rs10492367 by investigating the expression of *PTHLH* and *KLHL42*.

4.2 Aim

The aim of this chapter was to characterise the OA association signal marked by the polymorphism rs10492367. This was achieved by:

- performing database searches to characterise the association region
- examining the expression profiles of *PTHLH* and *KLHL42* during chondrogenesis using RNA generated by Dr Madhushika Ratnayake, Maria Tselepi and Emma Rogers
- analysing *PTHLH* and *KLHL42* expression throughout chondrogenesis as assayed on a microarray performed by Dr Matthew Barter
- analysing *PTHLH* and *KLHL42* expression throughout osteoblastogenesis as assayed on a microarray performed by Dr Rodolfo Gomez
- quantitatively analysing the expression of *PTHLH* and *KLHL42* in synovial joint tissues
- characterising the expression profiles of *PTHLH* and *KLHL42* in OA hip, OA knee and NOF cartilage, using data previously generated by Dr Madhushika Ratnayake and Dr Emma Raine
- replicating the gene expression quantification experiments in an independent group of OA hip cartilage samples
- characterising the expression profiles of *PTHLH* and *KLHL42* in fat pad and synovium samples
- investigating if rs10492367 marks a *cis*-eQTL by producing an allelic expression imbalance of the transcription of *PTHLH* and/or *KLHL42*
- analysing CpG methylation levels in hip and knee cartilage 1 Mb upstream and 1 Mb downstream of rs10492367 as assayed on a microarray performed by Dr Michael Rushton
- identifying if differential methylation of CpG sites within 20 bp upstream or downstream of rs10492367, or any SNP in high LD with it, correlates with rs10492367 genotype
- identifying if differential methylation of CpG sites within 20 bp upstream or downstream of rs10492367, or any SNP in high LD it, affects the enhancer activity of the regions

4.3 Results

4.3.1 *Initial database searches to characterise the rs10492367 locus*

Figure 4.1 is a screenshot from the UCSC Genome Browser website (Kent *et al.*, 2002) covering the 12p11.22 locus. The first track displays RefSeq genes (Pruitt *et al.*, 2005), showing rs10492367 as an intergenic SNP 59 kb downstream of *KLHL42* and 96 kb downstream of *PTHLH*. A further thirteen genes reside within a 1 Mb span upstream or downstream of the polymorphism: *ASUN* (924 kb downstream), *FGFR1OP2* (901 kb upstream), *TM7SF3* (848 kb downstream), *MED21* (831 kb upstream), *C12orf71* (780 kb downstream), *STK38L* (536 kb upstream), *ARNTL2* (436 kb upstream), *SMCO2* (360 kb upstream), *PPFIBP1* (166 kb upstream), *REP15* (164 kb upstream), *MRPS35* (106 kb upstream), *MANSC4* (91 kb downstream) and *CCDC91* (40 kb downstream). In addition, rs10492367 is 415 kb upstream of the long non-coding RNA *ARNTL2-AS1*. Due to the comparatively small size of the association region, only a 1 Mb region is shown in Figure 4.1, rather than 2 Mb. Using ChIP-Seq data, the chromatin state of the region was modelled using a multivariate Hidden Markov Model (Ernst and Kellis, 2010; Ernst *et al.*, 2011). It is apparent that the functionality of this region is dependent on the cell line under investigation, with the green boxes of the track representing an area of active transcription, the grey boxes denoting a region of heterochromatin and the orange boxes denoting a region with strong enhancer activity. Furthermore, there are several transcription factors known to bind rs10492367 (Gerstein *et al.*, 2012; Wang *et al.*, 2012; Wang *et al.*, 2013) – including RELA (v-rel avian reticuloendotheliosis viral oncogene homolog A), TCF3 (transcription factor 3) and TCF12 – and the signal is also in a DNase I hypersensitivity region (Song and Crawford, 2010). The sensitivity to DNase I digestion implies that the region is in an open conformation and could therefore be exposed to transcription factor binding that could regulate nearby gene expression. Finally, based on the ChIA-PET dataset (Fullwood *et al.*, 2010), it seems rs10492367 does not interact with genomic loci outside this region. This is the same for all SNPs in high LD with rs10492367, but due to the large size of the association region (21 kb), only rs10492367 is depicted in Figure 4.1.B.

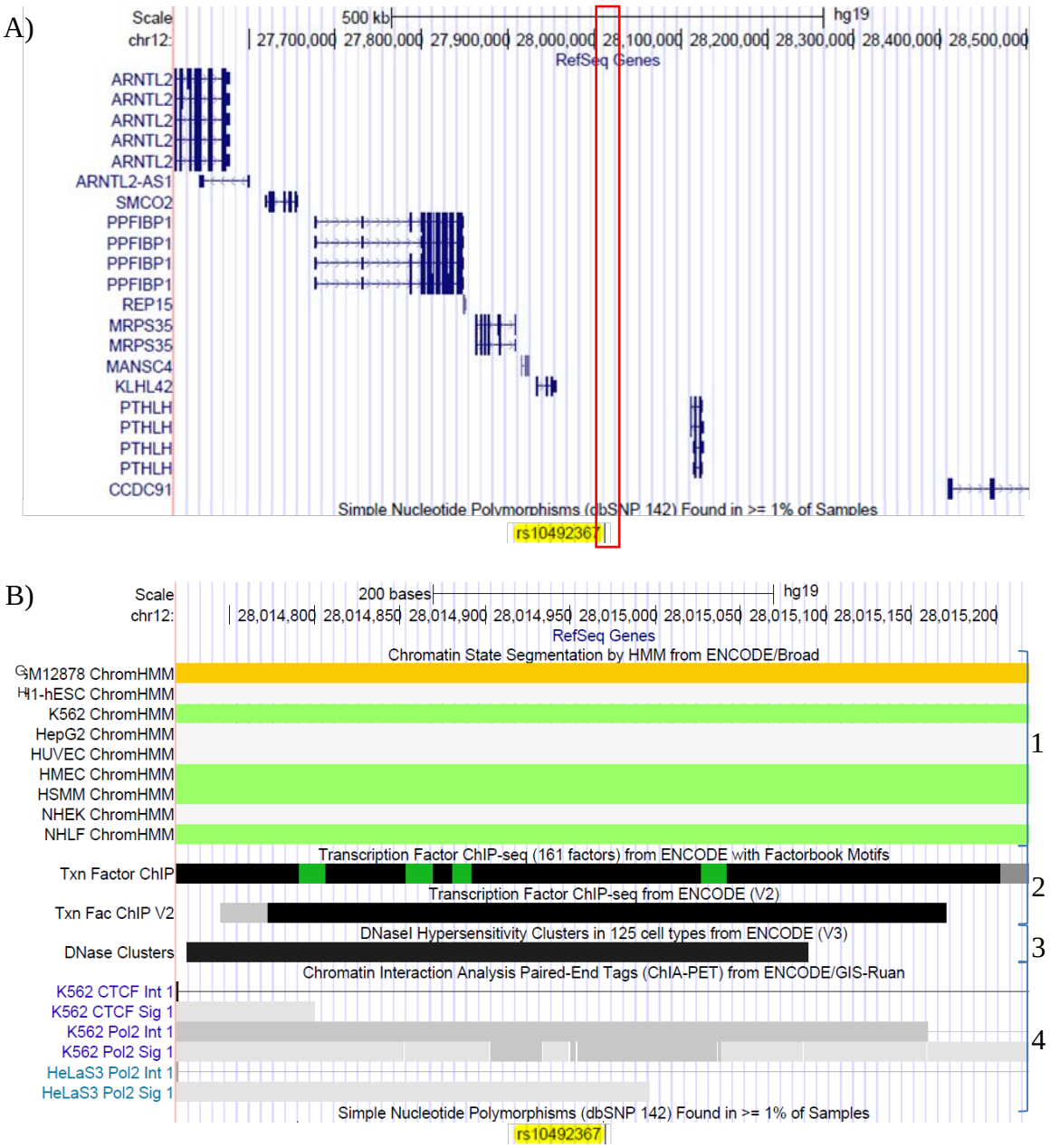

Figure 4.1. UCSC Genome Browser screenshot of the OA association region marked by the polymorphism rs10492367 on chromosome 12p11.22. A) rs10492367 is an intergenic polymorphism downstream of *PTHLH* and *KLHL42*. The red box marks the boundaries of the association interval; all SNPs with an $r^2 \geq 0.80$ relative to rs10492367 reside in this region. B) The SNP resides in a predicted enhancer in the GM12878 cell line, denoted by an orange box within the Chromatin State Segmentation track [1]. Transcription factors have been identified as binding the polymorphism within the Transcription Factor ChIP-Seq tracks [2], and it is within a DNase I-hypersensitive region of the DNase I Hypersensitivity Clusters track [3]. The SNP has not been identified as interacting with other genomic loci in the Chromatin Interaction Analysis Paired-End Tags track [4]. The images were obtained using the hg19 reference genome.

As rs10492367 is not necessarily causal in mediating the OA association of this region, any SNPs in high LD with the polymorphism could instead account for the signal. In total, there

are nine SNPs with an r^2 of $\geq$ 0.80 relative to the association SNP, three of which are in perfect LD with rs10492367. A search of the RegulomeDB online database (Boyle *et al.*, 2012) shows that some SNPs, including rs10492367, have transcription factors known to bind or reside in functional regions in relevant cell lines (Table 4.2). Currently, aside from rs10492367, there is no published literature relating any of the SNPs to an OA or musculoskeletal phenotype.

SNP	Distance from rs10492367 (bp)	r^2 relative to rs10492367	*D′* relative to rs10492367	Genotyped on arcOGEN array	SNP location	Transcription factor binding	Chromatin State		
							Bone marrow-derived cultured MSCs	MSC-derived chondrocyte cultured cells	Osteoblast primary cells
rs58649696	10,617	0.853	1.000	No	Intergenic	IRF1, HNF4A, EP300, GATA2	Quiescent/low	Quiescent/low	Quiescent/low
rs57380671	4,563	0.806	1.000	No	Intergenic	No data	Quiescent/low	Quiescent/low	Quiescent/low
rs61916489	1,682	0.853	1.000	No	Intergenic	No data	Quiescent/low	Quiescent/low	Quiescent/low
rs11049204	1,407	1.000	1.000	No	Intergenic	No data	Quiescent/low	Quiescent/low	Quiescent/low
rs10492367	0	1.000	1.000	Yes	Intergenic	FOXM1, ATF2, IKZF1, EP300, NFATC1, NFKB1, NFIC, GATA2, MEF2A, MEF2C, TCF3, BCL11A, PAX5, TCF12, ZNF143	Quiescent/low	Enhancer	Quiescent/low
rs10743612	421	0.865	1.000	No	Intergenic	No data	Quiescent/low	Enhancer	Quiescent/low
rs11049206	531	1.000	1.000	No	Intergenic	No data	Enhancer	Enhancer	Quiescent/low
rs11049207	3,683	0.853	1.000	No	Intergenic	CEBPB, FOS, MYC, STAT3	Enhancer	Enhancer	Enhancer
rs79881709	4,054	0.853	1.000	No	Intergenic	No data	No data	No data	No data
rs10843013	10,226	1.000	1.000	Yes	Intergenic	No data	Quiescent/low	Quiescent/low	Quiescent/low

Table 4.2. All SNPs in high linkage disequilibrium ($r^2 \geq 0.80$) with rs10492367. Out of the nine polymorphisms, three were in perfect linkage disequilibrium with this association SNP, one of which was also genotyped on the arcOGEN microarray.

4.3.2 *Examination of the expression profiles of PTHLH and KLHL42 during chondrogenesis*

As the association signal resides in a region suggested to be functional, I first sought to investigate the expression of *PTHLH* and *KLHL42* throughout the development of cartilage. As for Chapter 3, I hypothesised that the functional effects of the association signal could be exerted at early stages of joint development, resulting in altered joint structure and thus predisposing an individual to developing OA later in life. I therefore sought to replicate the overall gene expression quantification experiments throughout chondrogenesis irrespective of rs10492367 genotype. This would clarify whether the genes were expressed and whether they had the potential to contribute to joint development and OA susceptibility. Dr Madhushika Ratnayake, Maria Tselepi and Emma Rogers of Prof. Loughlin's group (Institute of Cellular Medicine, Newcastle University) performed *in vitro* chondrogenesis to replicate cartilage development: I performed the qPCR, data analysis and genotyping (Table 4.3). Primer and probe sequences can be found in Appendix A: Table A.3, and the assay positions for *PTHLH* and *KLHL42* are detailed in Figure 4.2. The assays for *18S*, *GAPDH* and *HPRT1* were the same as those used for the qPCR of Chapter 3 (Figure 3.2). The probes and primers were designed to the cDNA sequences, and thus excluded intronic regions. For all of the genes, every known protein-coding transcript was covered by the assays.

Donor ID	Sex	Age at donation (years)	Joint	rs10492367 genotype
52	F	61	Hip	GG
93	F	51	Hip	GG
225	M	55	Hip	GG
276*	F	41	Iliac crest	GG
277*	F	24	Iliac crest	GG
278*	M	25	Iliac crest	GG

Table 4.3. Characteristics and genotype at rs10492367 for donors used in chondrogenesis. The donors marked with an asterisk (*) were donors with an unknown OA status purchased from Lonza, UK.

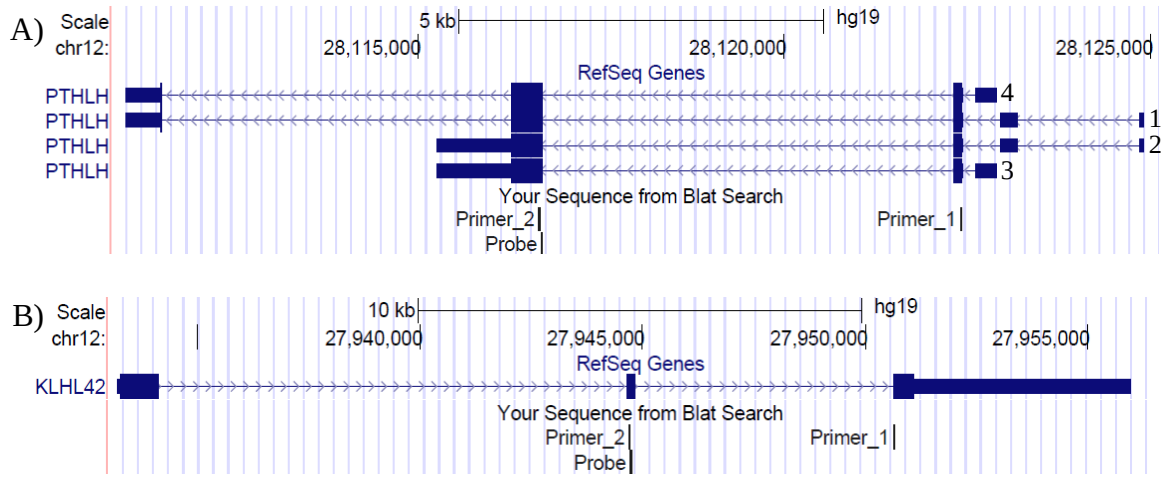

Figure 4.2. Location of qPCR primers and probes used for quantitative gene expression analysis. Assays were designed to the exons of A) *PTHLH* and B) *KLHL42*. Every transcript of *PTHLH* was covered by the assays. Variants 1 to 4 are numbered as such for *PTHLH*. The images were obtained using the hg19 reference genome.

Gene expressions were quantified at three time points throughout chondrogenesis: day 3, day 7 and day 14. Both genes were expressed throughout the chondrogenesis time course (Figure 4.3), although *PTHLH* was variable with low levels of expression. To confirm that chondrogenesis had progressed, the expressions of the chondrogenic markers *ACAN*, *COL2A1* and *SOX9* (Bhang *et al.*, 2011) were confirmed at each time point by Dr Madhushika Ratnayake, Maria Tselepi and Emma Rogers (*data not shown*). Table 4.4 details the mean $2^{-\Delta Ct}$ values for *PTHLH*, highlighting the difficulty in calculating the fold change of gene expression relative to day 3. There were no distinct differences in expression depending on the disease state, nor were there defined patterns of expression associated with the stage of cartilage development. The inter-individual variability and dynamic patterns of gene expression suggest that either of the genes could function at any of the stages of chondrogenesis to contribute to OA susceptibility.

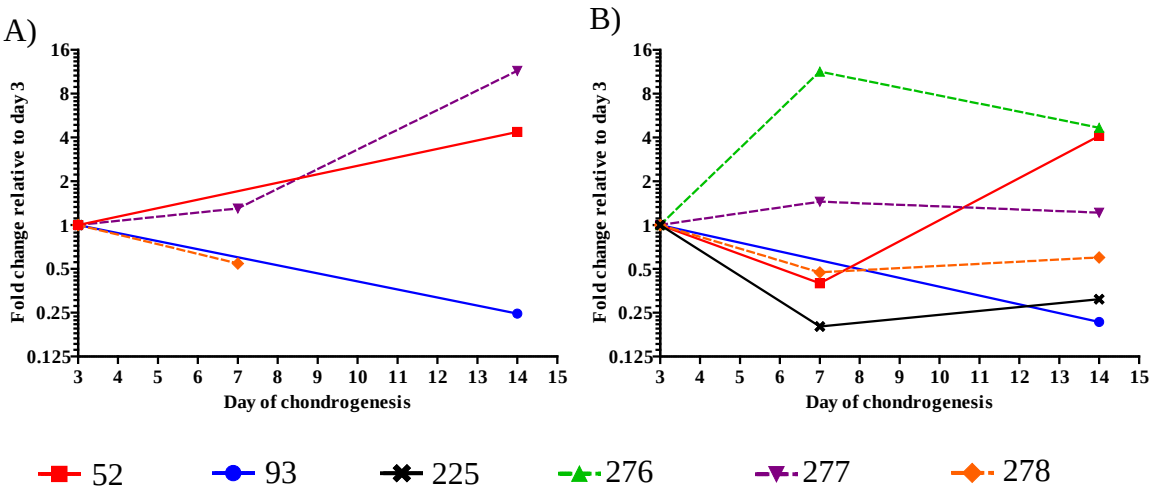

Figure 4.3. Expression of *PTHLH* and *KLHL42* during chondrogenesis. MSCs were differentiated into chondrocytes and gene expression was quantified at various time points throughout chondrogenesis. A) *PTHLH* expression was very low throughout chondrogenesis and quantification was not possible at several time points. B) *KLHL42* was expressed throughout the time course of chondrogenesis in all donors.

	Donor					
Day	**52**	**93**	**225**	**276**	**277**	**278**
3	9.287×10^{-5}	2.280×10^{-3}	8.196×10^{-5}	Undetermined	2.645×10^{-6}	1.845×10^{-5}
7	Undetermined	Undetermined	Undetermined	Undetermined	3.436×10^{-6}	1.005×10^{-5}
14	4.057×10^{-4}	5.625×10^{-4}	Undetermined	2.936×10^{-5}	3.038×10^{-5}	Undetermined

Table 4.4. Calculated mean $2^{-\Delta Ct}$ values for *PTHLH* for each donor at the three time points that gene expression was quantified. At many time points, the gene expression was undetected.

4.3.3 *Analysis of PTHLH and KLHL42 expression throughout chondrogenesis as assayed on a microarray*

Both genes were shown to be expressed during chondrogenesis in the preceding investigation (Chapter 4.3.2), albeit at low levels particularly for *PTHLH*. As an additional approach to validate these results, I acquired normalised data from a microarray performed and analysed by Dr Matthew Barter (*personal communication*). MSCs from a 22 year old female (Lonza, UK) were differentiated into chondrocytes and, using RNA extracted at days 0, 1, 3, 6, 10 and 14, an Illumina Human HT-12 V4 expression array was used to profile a range of gene expressions (Barter *et al.*, 2015). Table 4.5 details the day 14 relative to day 0 expression fold change of the genes surrounding rs10492367 and of the chondrogenic markers *ACAN*, *COL2A1* and *SOX9* (Bhang *et al.*, 2011). The analysis included all genes within 1 Mb upstream or downstream of the association SNP, and the genes that failed to meet the quality

control standards were not included. The chondrogenic markers were all significantly upregulated by day 14 relative to day 0. Similarly, *STK38L*, *REP15* and variant 1 of *PPFIBP1* were all upregulated by day 14. Conversely, *PTHLH*, *MRPS35* and *TM7SF3* were downregulated relative to day 0. In addition, I also considered the overall gene expression at each individual time point (Figure 4.4) in order to independently corroborate the findings presented in Figure 4.3. In this donor, akin to the pattern observed previously, *KLHL42* has a higher overall expression compared to *PTHLH*. The remaining genes were also similarly expressed, and so overall, it can be concluded that the genes at this locus were expressed throughout chondrogenesis.

Probe	Gene	Day 14 vs day 0 $\log_2$ fold change	Average expression	Adjusted *p* value
6900414	*PTHLH* (variant 1)	-0.248	6.764	0.014
1980593	*PTHLH* (variant 3)	-0.244	6.442	0.015
4490273	*KLHL42*	-0.146	9.116	0.171
1340681	*FGFR1OP2*	-0.018	7.864	0.866
130672	*MED21*	0.101	6.884	0.313
2760653	*STK38L*	2.980	8.500	2.82×10^{-15}
3390364	*PPFIBP1* (variant 1)	0.665	7.224	1.23×10^{-05}
4010414	*PPFIBP1* (variant 1)	0.355	9.148	0.002
1940450	*PPFIBP1* (variant 2)	-0.079	6.752	0.515
2680131	*PPFIBP1* (variant 1)	-0.041	7.369	0.740
20753	*REP15*	0.292	6.901	0.005
7380270	*MRPS35*	-0.263	9.144	0.012
2340594	*CCDC91*	-0.190	8.388	0.081
2070474	*TM7SF3*	-0.338	8.292	0.002
4480747	*ACAN* (variant 1)	2.237	7.445	7.57×10^{-13}
6770470	*ACAN* (variant 2)	3.851	8.518	2.22×10^{-16}
4010136	*COL2A1* (variant 1)	8.044	11.288	7.76×10^{-20}
650113	*COL2A1* (variant 1)	3.937	8.69	7.97×10^{-14}
4230475	*SOX9*	3.276	10.895	1.57×10^{-14}

Table 4.5. Comparison of day 14 and day 0 normalised gene expression during chondrogenesis at the rs10492367 locus. *PTHLH*, *MRPS35* and *TM7SF3* were significantly downregulated by day 14 relative to day 0, while *STK38L*, *REP15* and *PPFIBP1* variant 1 were significantly upregulated. There were no significant differences in *KLHL42*, *FGFR1OP2*, *MED21* or *CCDC91* expressions between day 14 and day 0. The chondrogenic markers were all significantly upregulated throughout the chondrogenesis time course. Both isoforms of *PPFIBP1* and *ACAN* were captured by the array, while two probes captured the same variant of *COL2A1*. Two of the four *PTHLH* isoforms were captured by the microarray.

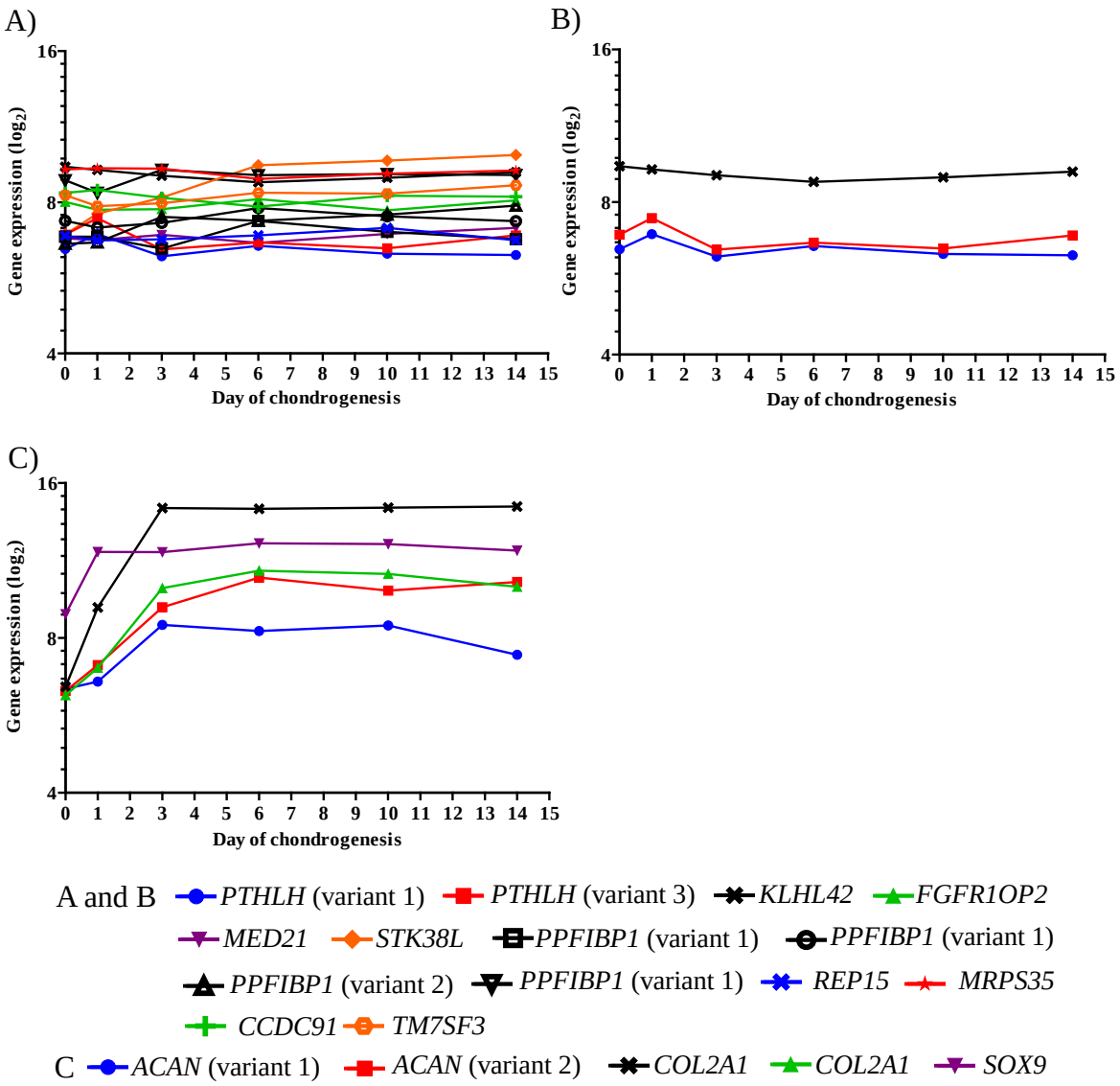

Figure 4.4. Expression of genes at the rs10492367 locus during chondrogenesis. MSCs were differentiated into chondrocytes and gene expression was quantified at various time points throughout chondrogenesis on an Illumina Human HT-12 V4 expression array. A) all genes within 2 Mb of rs10492367, B) *PTHLH* and *KLHL42* only and C) the chondrogenic markers *ACAN*, *COL2A1* (covered by two probes [4010136, black x; and 650113, green triangle]) and *SOX9* were all expressed throughout the time course of chondrogenesis. The levels of expression imply that any of the genes at the 12p11.22 locus have the potential to function during cartilage development.

4.3.4 *Analysis of PTHLH and KLHL42 expression throughout osteoblastogenesis as assayed on a microarray*

As OA is increasingly becoming recognised as a disease of the entire synovial joint and not specifically of cartilage, I sought to investigate the expression of the genes during osteoblastogenesis. In addition, *PTHLH* has an established role in endochondral ossification and so I hypothesised that this gene, or perhaps *KLHL42*, could in fact contribute to disease susceptibility by affecting joint development and structure by acting during

osteoblastogenesis. I acquired normalised data from a microarray performed and analysed by Dr Rodolfo Gomez (*personal communication*). MSCs from a 19 year old female and a 24 year old female (Lonza, UK) were differentiated into osteoblasts and, using RNA extracted at days 0 and 21, an Illumina Human HT-12 V4 expression array was used to profile a range of gene expressions. Table 4.6 details the day 21 relative to day 0 expression fold change of the genes surrounding rs10492367 and of the osteogenic markers *ALPL* and *DCN* (Waddington *et al.*, 2003; Graneli *et al.*, 2014). The osteogenic markers were both significantly upregulated by day 21 relative to day 0, confirming that osteoblastogenesis had progressed. Out of the genes of interest, only *TM7SF3* significantly differed by day 21, with its expression upregulated relative to day 0. Overall, all genes were expressed throughout osteoblastogenesis, which implies that any of the genes, particularly *TM7SF3*, could contribute to OA susceptibility by affecting joint structure.

Probe	Gene	Day 14 vs day 0 log_2 fold change	Average expression	Adjusted *p* value
6900414	*PTHLH*	0.133	7.205	0.639
4490273	*KLHL42*	0.074	9.834	0.802
5270451	*ASUN*	-0.074	7.210	0.808
6400082	*ASUN*	-0.075	8.570	0.801
1340681	*FGFR1OP2*	0.183	7.875	0.499
2760653	*STK38L*	-0.081	7.615	0.781
4010414	*PPFIBP1*	0.427	9.301	0.106
2680131	*PPFIBP1*	0.275	8.449	0.392
20753	*REP15*	0.069	7.276	0.813
7380270	*MRPS35*	0.286	8.386	0.267
2340594	*CCDC91*	0.355	7.855	0.157
2070474	*TM7SF3*	0.583	9.365	0.028
6100356	*ALPL*	2.071	11.788	0.007
50368	*DCN* (variant A1)	0.786	11.061	0.033
5550719	*DCN* (variant A2)	1.495	11.857	5.75×10^{-06}
7650296	*DCN* (variant A2)	0.909	14.067	0.003

Table 4.6. Comparison of day 21 and day 0 normalised gene expression during osteoblastogenesis at the rs10492367 locus. *TM7SF3* was the only gene to be significantly upregulated by day 21 relative to day 0. There were no significant differences in the expression of the remaining genes between day 21 and day 0. The osteogenic markers were both significantly upregulated throughout the osteoblastogenesis time course. Both isoforms of *DCN* were captured by the array.

4.3.5 *Quantitative expression analysis of PTHLH and KLHL42 in synovial joint tissues*

Since Chapter 4.3.1 indicated that the association signal resides in a region that is functional, and Chapters 4.3.2 – 4.3.4 confirmed that the genes are expressed during chondrogenesis and osteoblastogenesis, I next aimed to characterise the expression of the genes in primary joint tissues. Following the publication of the arcOGEN GWAS in 2012 (arcOGEN Consortium *et al.*, 2012) and prior to the start of my Ph.D, expression datasets pertaining to *PTHLH* and *KLHL42* were generated by Dr Madhushika Ratnayake and Dr Emma Raine, two members of Prof. Loughlin's group (Institute of Cellular Medicine, Newcastle University). Given that *PTHLH* is such a strong candidate gene and *KLHL42* is the closest gene physically to rs10492367, characterisation efforts focussed on these two genes. qPCR was performed using OA hip, OA knee and NOF cartilage samples. Similarly, I performed qPCR using fat pad and synovium tissue originating from the knees of OA patients: all reactions were performed in triplicate and each reaction was normalised to the mean of the housekeeping genes *18S*, *GAPDH* and *HPRT1*.

In all of the joint tissues tested, expression of the genes was detected (Figure 4.5): for this analysis, the gene expression datasets for cartilage were separated into OA hip, OA knee and NOF. The expression of both genes was lowest in the synovium samples. *KLHL42* was most highly expressed in fat pad, while *PTHLH* was most highly expressed in OA knee cartilage.

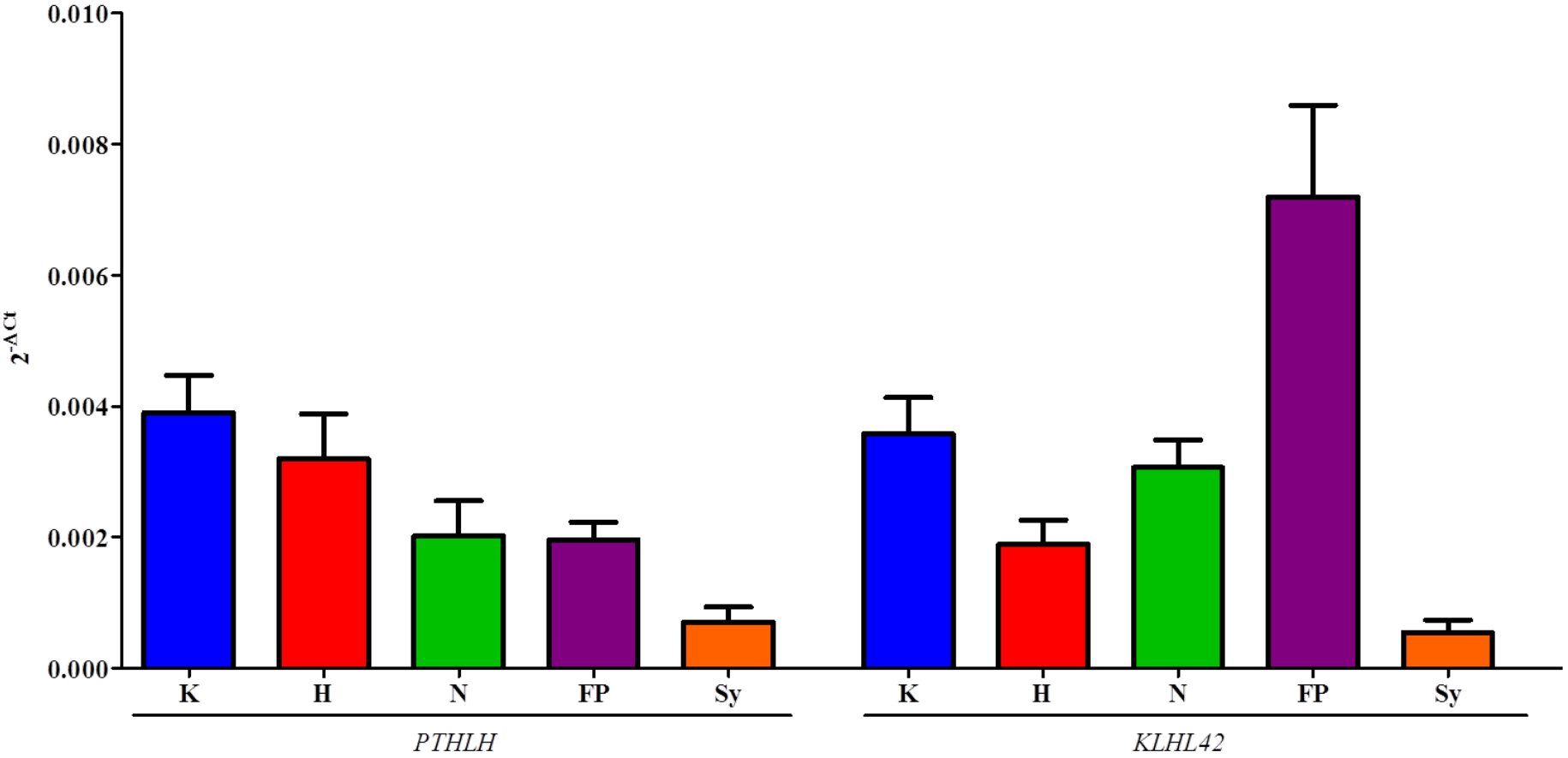

Figure 4.5. Average gene expression at the12p11.22 locus in all joint tissues assayed. Both genes were expressed in all tissues tested. The highest expression of *PTHLH* was in OA knee cartilage, and the lowest expression was in OA synovium. The highest expression of *KLHL42* was in OA fat pad, and the lowest expression was in OA synovium. Tissues tested: OA knee cartilage (K; $n \leq 52$), OA hip cartilage (H; $n \leq 21$), NOF cartilage (N; $n \leq 18$), fat pad (FP; $n \leq 26$) and synovium (Sy; $n \leq 22$). Error bars represent the SEM.

4.3.6 *Characterising the expression profiles of PTHLH and KLHL42 in cartilage: comparisons of disease state, sex, skeletal site and age*

For each donor I compiled the following information: genotype at rs10492367, age (in years) at joint replacement, sex and joint replaced (Appendix D: Table D.1). I subsequently analysed the expression profiles of *PTHLH* and *KLHL42* in cartilage using various routes of stratification (Table 4.7 and Table 4.8, respectively). As differences in gene expression depending on disease state have previously been shown in the context of OA-associated regions (Raine *et al.*, 2012), I included a comparison of OA and NOF. In addition, as discussed in Chapter 1, there are sex-specific differences in the prevalence of OA. Finally, given that OA is an age-associated disease, a comparison of gene expression relative to age was performed. For these analyses, solely due to the availability of tissue, all NOF donors were female. Only the statistically significant comparisons are included in the graphical representations (Figure 4.6 and Figure 4.7).

Stratification for *PTHLH* qPCR data	***p* value**
OA hip and OA knee (n = 72) vs NOF (n = 18)	0.075
OA hip (n = 21) vs NOF (n = 18)	0.330
OA female (n = 40) vs NOF female (n = 18)	0.021
OA female (n = 40) vs OA male (n = 32)	0.080
OA knee (n = 51) vs OA hip (n = 21)	0.552
OA female knee (n = 25) vs OA female hip (n = 15)	0.889
OA male knee (n = 26) vs OA male hip (n = 6)	0.201
All OA age: 50 (n = 13) vs 60 (n = 22) vs 70 (n = 27) vs 80 (n = 10)	0.988

Table 4.7. Analysis of *PTHLH* expression in OA hip, OA knee and NOF cartilage. Expression of *PTHLH* was significantly upregulated in female OA cartilage relative to female NOF cartilage. Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons and a one-way analysis of variance for a comparison of more than two groups; *n* represents the number of individuals in the comparison group.

Stratification for *KLHL42* qPCR data	*p* value
OA hip and OA knee (*n* = 72) vs NOF (*n* = 18)	0.265
OA hip (*n* = 20) vs NOF (*n* = 18)	0.027
OA female (*n* = 39) vs NOF female (*n* = 18)	0.350
OA female (*n* = 39) vs OA male (*n* = 33)	0.964
OA knee (*n* = 52) vs OA hip (*n* = 20)	0.040
OA female knee (*n* = 25) vs OA female hip (*n* = 15)	0.356
OA male knee (*n* = 26) vs OA male hip (*n* = 6)	0.018
All OA age: 50 (*n* = 13) vs 60 (*n* = 22) vs 70 (*n* = 27) vs 80 (*n* = 10)	0.256

Table 4.8. Analysis of *KLHL42* expression in OA hip, OA knee and NOF cartilage. Expression of *KLHL42* was significantly downregulated in OA hip cartilage relative to NOF cartilage and OA knee cartilage. Expression in male OA hip cartilage was significantly decreased relative to male OA knee cartilage. Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons and a one-way analysis of variance for a comparison of more than two groups; *n* represents the number of individuals in the comparison group.

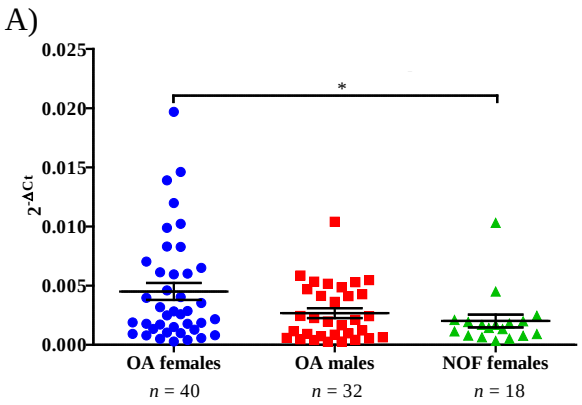

Figure 4.6. Statistically significant difference in the expression of *PTHLH*. Gene expression was significantly increased in A) female OA cartilage relative to female NOF cartilage. Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group; * $p \leq 0.05$; error bars represent the mean $\pm$ SEM.

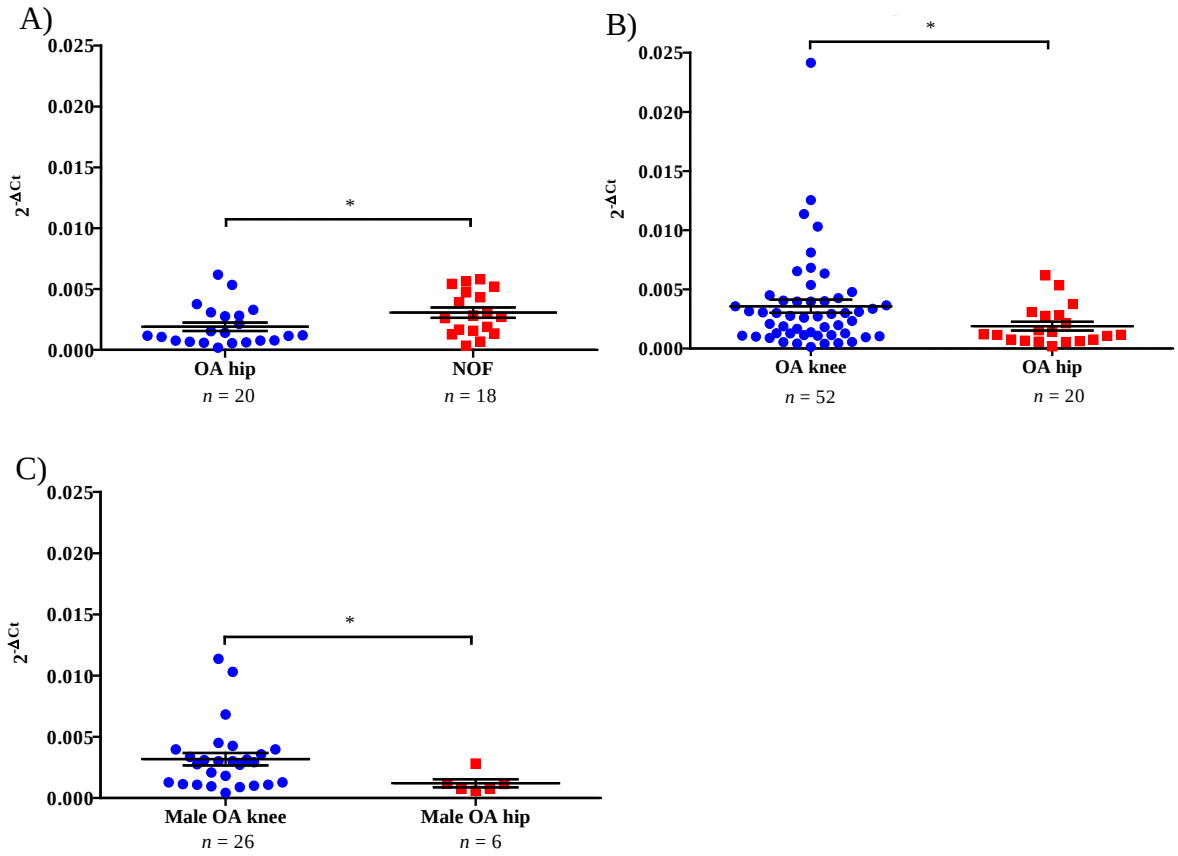

Figure 4.7. Statistically significant differences in the expression of *KLHL42*. Gene expression was significantly decreased in A) OA hip cartilage relative to NOF cartilage, B) OA hip cartilage relative to OA knee cartilage, and C) male OA hip cartilage relative to male OA knee cartilage. Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group; * $p \leq 0.05$; error bars represent the mean ± SEM.

From the analysis of these data, I identified differential expression of both genes. Expression of *PTHLH* was significantly increased in cartilage from all OA females relative to female NOF cartilage. *KLHL42* was significantly decreased in cartilage from OA hip relative to both NOF and OA knee cartilage. In addition, the expression of *KLHL42* was increased in the cartilage from the knees of OA males relative to male hip OA cartilage. Overall, the genes had relatively low levels of expression, but expression was confirmed in synovial joint cartilage.

4.3.7 *Characterising the expression profiles of PTHLH and KLHL42 in cartilage: comparisons of rs10492367 genotype*

The stratification necessary to investigate if an eQTL is operating at this locus is the comparison of gene expression between the genotypic groups of rs10492367; that is GG, GT

and TT. However, as for rs9350591, due to the low MAF (0.21) of rs10492367, it was not feasible in the timeframe of my Ph.D to acquire enough individuals that were TT at the association signal for hip OA and NOF donors. As such, T allele carriers (GT and TT) were grouped together and compared to GG homozygotes in all cases where TT $n \leq 3$. From these data, it was not possible to identify the actions of an eQTL influencing the expression of any of the genes tested (Figure 4.8 and Figure 4.9).

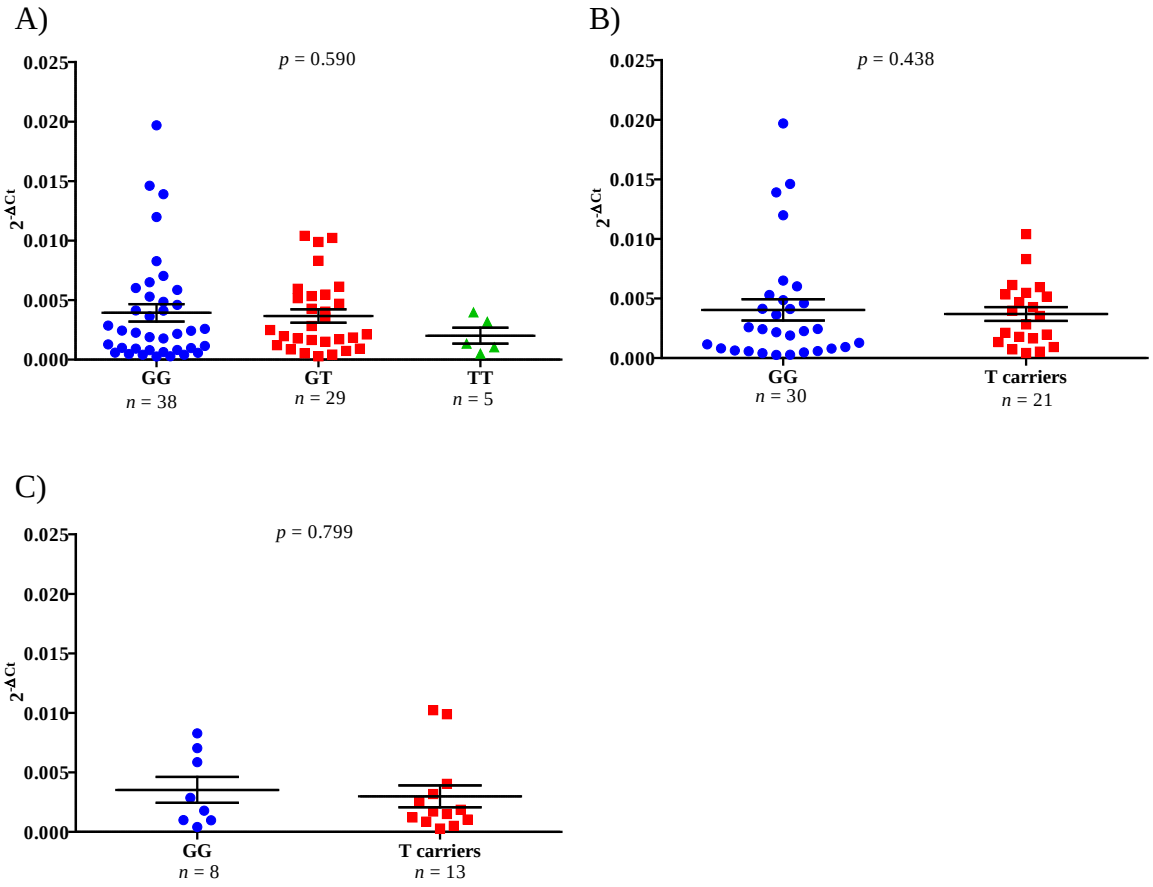

Figure 4.8. Analysis of *PTHLH* expression in OA hip and OA knee cartilage. There were no significant differences in gene expression between the genotypic groups of rs10492367 in A) all OA cartilage (hip and knee combined), B) OA knee cartilage, or C) OA hip cartilage. Statistical significance was assessed using a Mann-Whitney *U* test (two groups) or a Kruskal-Wallis one-way analysis of variance (three groups); *n* represents the number of individuals in the comparison group; error bars represent the mean ± SEM.

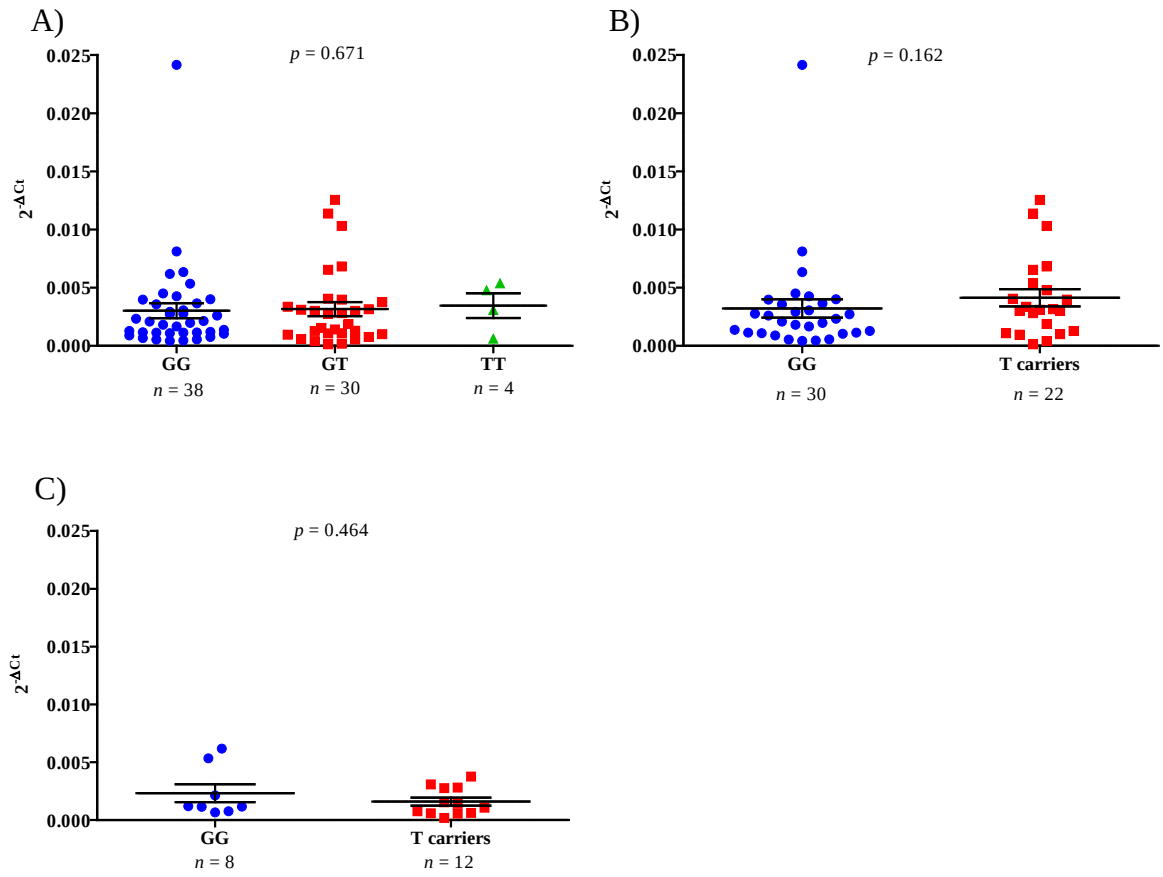

Figure 4.9. Analysis of *KLHL42* expression in OA hip and OA knee cartilage. There were no significant differences in gene expression between the genotypic groups of rs10492367 in A) all OA cartilage (hip and knee combined), B) OA knee cartilage, or C) OA hip cartilage. Statistical significance was assessed using a Mann-Whitney *U* test (two groups) or a Kruskal-Wallis one-way analysis of variance (three groups); *n* represents the number of individuals in the comparison group; error bars represent the mean ± SEM.

4.3.8 *Replication of the gene expression quantification experiments in an independent group of OA hip cartilage samples*

As rs10492367 marks a region of association in hip OA, I replicated the gene expression qPCR in an independent group of OA hip cartilage donors to confirm the findings presented in Chapter 4.3.7. Again, the genes had low levels of expression, and the carriage of the risk allele of rs10492367 did not cause a significant difference in the expression of either of the genes (Figure 4.10).

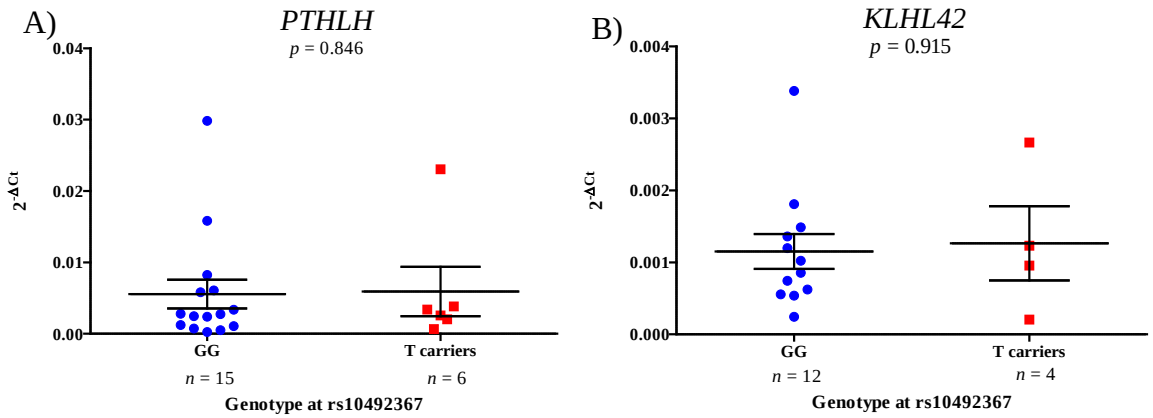

Figure 4.10. Independent replication of gene expression in OA hip cartilage. Expressions of A) *PTHLH* and B) *KLHL42* were quantified in OA hip cartilage and normalised to the housekeeping genes *18S*, *GAPDH* and *HPRT1*. Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group; error bars represent the mean ± SEM.

4.3.9 *Characterising the expression profiles of PTHLH and KLHL42 in OA fat pad and OA synovium: comparisons of sex and age*

As for the hip locus in Chapter 3, knee fat pad and synovium were considered in this investigation. I first compared levels of gene expression between female and male donors and then between ages (Table 4.9 and Table 4.10). Through these analyses, there were no significant differences of expression in fat pad or synovium.

Tissue	Stratification for *PTHLH* qPCR data	*p* value
FP	Female OA (*n* = 15) vs male OA (*n* = 11)	0.756
FP	All OA age: 50 (*n* = 4) vs 60 (*n* = 11) vs 70 (*n* = 9) vs 80 (*n* = 2)	0.533
Sy	Female OA (*n* = 11) vs male OA (*n* = 11)	0.860
Sy	All OA age: 50 (*n* = 6) vs 60 (*n* = 3) vs 70 (*n* = 9) vs 80 (*n* = 3)	0.630

Table 4.9. Analysis of *PTHLH* expression in OA fat pad and synovium. There were no significant differences in gene expression for either of the stratifications in fat pad (FP) or synovium (Sy). Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons and a one-way analysis of variance for a comparison of more than two groups; *n* represents the number of individuals in the comparison group.

Tissue	Stratification for *KLHL42* qPCR data	*p* value
FP	Female OA (*n* = 15) vs male OA (*n* = 10)	0.718
FP	All OA age: 50 (*n* = 3) vs 60 (*n* = 11) vs 70 (*n* = 9) vs 80 (*n* = 2)	0.267
Sy	Female OA (*n* = 10) vs male OA (*n* = 10)	0.739
Sy	All OA age: 50 (*n* = 5) vs 60 (*n* = 4) vs 70 (*n* = 9) vs 80 (*n* = 2)	0.763

Table 4.10. Analysis of *KLHL42* expression in OA fat pad and synovium. There were no significant differences in gene expression for either of the stratifications in fat pad (FP) or synovium (Sy). Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons and a one-way analysis of variance for a comparison of more than two groups; *n* represents the number of individuals in the comparison group.

4.3.10 *Characterising the expression profiles of PTHLH and KLHL42 in OA fat pad and OA synovium: comparisons of rs10492367 genotype*

As for the analyses in cartilage, the stratification necessary to investigate if an eQTL is operating at this locus is the comparison of gene expression between the genotypic groups of rs10492367. Again, T allele carriers (GT and TT genotypes) were grouped together and compared to GG homozygotes (Table 4.11 and Table 4.12). From these data, it was not possible to identify the actions of an eQTL influencing the expression of either of the genes tested, however, this interpretation must be with caution as rs10492367 is specifically an OA-associated locus in the hip stratum of the arcOGEN study, whereas these tissues originate from the joints of patients with knee OA.

Tissue	Stratification for *PTHLH* qPCR data	*p* value
FP	All OA GG (*n* = 16) vs all OA T carriers (*n* = 10)	0.200
FP	OA female GG (*n* = 9) vs OA female T carriers (*n* = 6)	0.224
FP	OA male CC (*n* = 7) vs OA male T carriers (*n* = 4)	0.649
Sy	All OA GG (*n* = 12) vs all OA T carriers (*n* = 9)	0.749
Sy	OA female GG (*n* = 6) vs OA female T carriers (*n* = 5)	0.931
Sy	OA male GG (*n* = 6) vs OA male T carriers (*n* = 4)	0.914

Table 4.11. Analysis of *PTHLH* expression in OA fat pad and synovium. There were no significant differences in gene expression for any of the stratifications in fat pad (FP) or synovium (Sy). Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group.

Tissue	Stratification for *KLHL42* qPCR data	*p* value
FP	All OA GG (*n* = 16) vs all OA T carriers (*n* = 9)	0.843
FP	OA female GG (*n* = 9) vs OA female T carriers (*n* = 6)	0.776
FP	OA male CC (*n* = 7) vs OA male T carriers (*n* = 3)	0.517
Sy	All OA GG (*n* = 12) vs all OA T carriers (*n* = 8)	0.671
Sy	OA female GG (*n* = 6) vs OA female T carriers (*n* = 4)	1.000
Sy	OA male GG (*n* = 6) vs OA male T carriers (*n* = 4)	0.610

Table 4.12. Analysis of *KLHL42* expression in OA fat pad and synovium. There were no significant differences in gene expression for any of the stratifications in fat pad (FP) or synovium (Sy). Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group.

4.3.11 *Investigating the effect of the rs10492367 association signal on the allelic output of the transcripts of PTHLH and KLHL42*

In the overall gene expression analysis discussed in the preceding sections of this chapter, I have observed no evidence for rs10492367 marking a *cis*-eQTL at this locus in any of the end-stage OA tissue samples tested. As an additional approach to investigate this, the allelic outputs of the gene transcripts were quantified. Transcript SNPs for each of the genes were required as markers of mRNA output, the premise of which is explained in Figure 3.18. Each gene was examined for transcripts SNPs (Table 4.13 and Table 4.14) using the UCSC Genome Browser (Kent *et al.*, 2002), and a SNAP Pairwise LD (Johnson *et al.*, 2008) search was performed for all of the polymorphisms to assess the degree of correlation with rs10492367. The heterozygote frequencies were obtained from the dbSNP online database (Sherry *et al.*, 2001) and one transcript SNP per gene was selected.

SNP	r^2 relative to rs10492367	D' relative to rs10492367	Heterozygote Frequency (%)	Genetic location
rs6246	0.013	0.146	42.00	Intron/5′ UTR
rs6253	0.003	0.068	42.00	Intron/3′ UTR
rs6244	0.008	0.160	13.88	Intron/3′ UTR
rs6252	0.008	0.160	12.33	3′ UTR
rs6245	No data	No data	12.33	Intron/3′ UTR
rs2796	0.008	0.160	12.33	Intron/3′ UTR

Table 4.13. Transcript polymorphisms within *PTHLH*. rs6253 was selected as a marker for mRNA output in AEI analysis. The multiple genetic loci for some of the SNPs reflect the position on different transcript isoforms.

SNP	r^2 relative to rs10492367	D' relative to rs10492367	Heterozygote Frequency (%)	Genetic location
rs17224065	0.045	0.436	49.88	3′ UTR
rs17801400	0.039	0.410	49.78	3′ UTR
rs1050287	0.011	0.228	49.50	3′ UTR
rs1050288	0.021	0.170	39.14	3′ UTR
rs9029	0.054	0.769	38.30	3′ UTR
rs7971518	0.046	0.240	35.42	3′ UTR
rs11613049	0.004	1.000	3.34	3′ UTR
rs11609108	0.004	1.000	3.34	3′ UTR
rs61244584	0.004	1.000	3.34	3′ UTR
rs12301204	0.004	1.000	3.34	3′ UTR
rs11614346	0.004	1.000	3.34	3′ UTR

Table 4.14. Transcript polymorphisms within *KLHL42*. rs9029 was selected as a marker for mRNA output in AEI analysis.

It was necessary to validate the ability of the sequencing primers to distinguish between allelic ratios at the polymorphisms by combining the DNA of major and minor allele homozygotes in order to generate known allelic ratios before comparing the values detected experimentally to the expected outcome (Appendix B: Figure B.1). There were no donors available who were homozygous for the minor allele at rs9029, and so this meant that the generation of allelic ratios was limited to heterozygotes and major allele homozygotes. Both of the validations had a positive correlation between observed and expected ratios, each with a goodness of fit $r^2 \geq 0.90$, and therefore were considered suitable for AEI analysis. Reactions were performed in triplicate and the mean of the cDNA allelic ratios was normalised to the corresponding gDNA ratios. Primer sequences for the assays can be found in Appendix A: Table A.1, and the assay positions detailed in Figure 4.11. The low MAF of the association SNP meant that minor allele homozygotes were scarce, and as such, the analyses were restricted to rs10492367 GG and GT individuals. As for the AEI analyses of Chapter 3, hip OA, knee OA and NOF cartilage were analysed collectively.

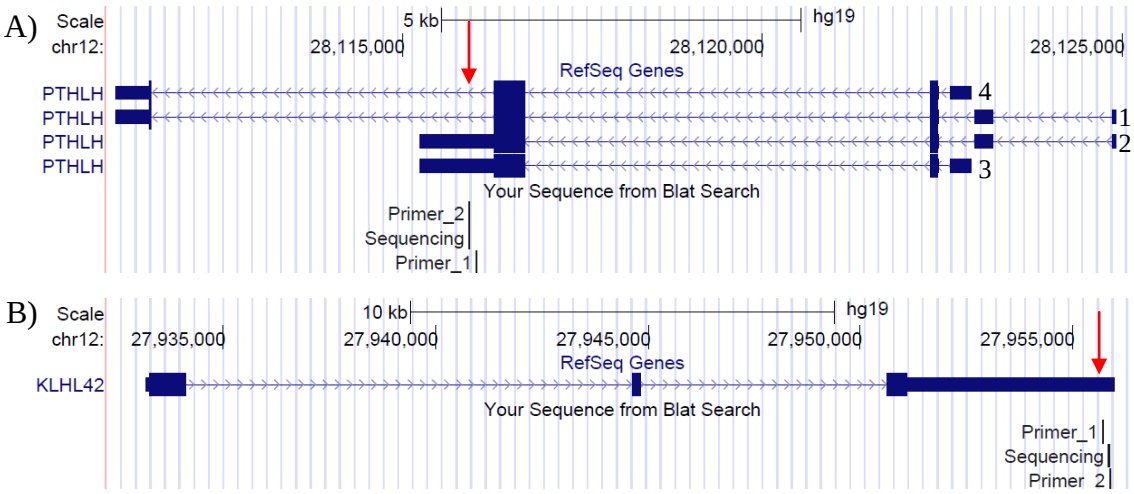

Figure 4.11. Location of pyrosequencing primers used for allelic expression analysis. Assays were designed to the exons of A) rs6253 of *PTHLH* and B) rs9029 of *KLHL42*. There were no polymorphisms that could be used to cover every transcript variant of *PTHLH*. Red arrows (→) indicate transcript SNP positions. Variants 1 to 4 are numbered as such for *PTHLH*. The images were obtained using the hg19 reference genome.

A caveat of the AEI performed for *PTHLH* is that only two (variants 2 and 3) of the four RefSeq transcripts were captured by the assay. This was impossible to avoid as no other SNPs were present in all of the transcripts. The addition of another transcript SNP into the investigation was not achievable, the first reason being due to the physical position of the SNPs. rs6246, the only other SNP with a high heterozygote frequency, was present in variants 3 and 4, meaning that this combined with rs6253 would still leave variant 1 without any quantification. Secondly, the low heterozygote frequencies of the transcript SNPs rs6244, rs6252, rs6245 and rs2796 meant that the compound heterozygote frequencies with rs10492367 would not allow for a large enough *n* number to be gathered.

There was no evidence to indicate rs10492367 correlates with a *cis*-eQTL acting on *PTHLH* (Figure 4.12.A) or *KLHL42* (Figure 4.12.B) in cartilage. Despite some isolated allelic imbalances, on the whole, unlike the trends observed for *COL12A1* and *MYO6* (Figure 3.20.D and Figure 3.20.B, respectively), there were no *cis*-eQTLs operating on the gene transcripts that were unrelated to the OA association signal: both genotype groups clustered around a 1:1 output ratio ($y = 1$). Neither of the differences between the genotype groups were statistically significant.

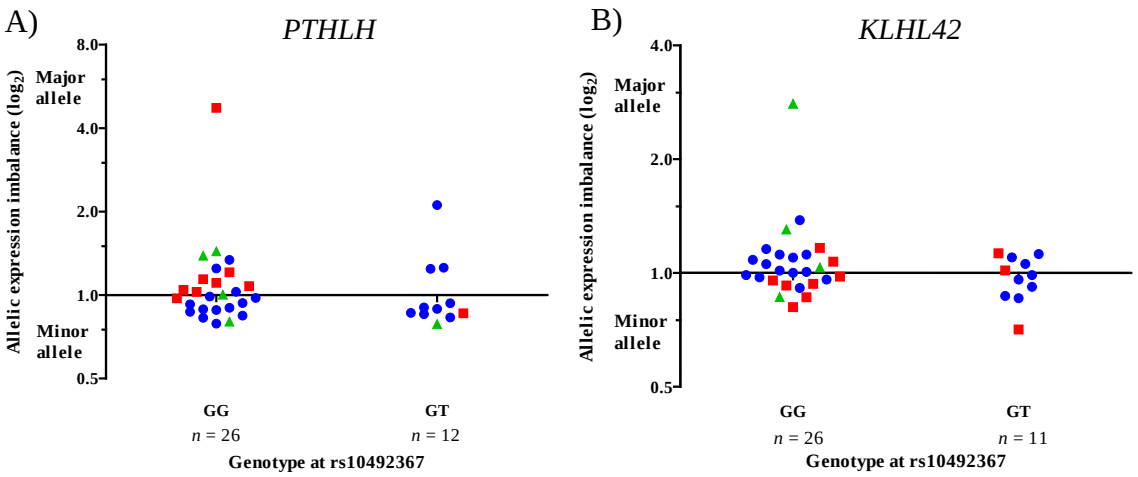

Figure 4.12. Allelic expression imbalance of *PTHLH* and *KLHL42* in hip and knee cartilage stratified by rs10492367 genotype. There was no distinct clustering of compound heterozygote donors that would imply a correlation between rs10492367 genotype and an imbalance of mRNA output of A) *PTHLH* or B) *KLHL42*. OA knee (blue circles), OA hip (red squares) and NOF (green triangles) cartilage. Statistical significance was assessed using a Mann-Whitney *U* test; *n* represents the number of individuals in the comparison group.

4.3.12 *Analysis of methylation levels at CpG sites surrounding the 12p11.22 locus in hip and knee cartilage*

As discussed in Chapter 3.3.12, an underlying cause of the OA association signals could be differential methylation of the surrounding genomic regions. It is possible that rs10492367 could impact upon the methylation levels of the surrounding region, perhaps by affecting the binding of DNA methyltransferases. Although I have shown thus far that there is no correlation between rs10492367 genotype and nearby gene expression, a similar analysis of the epigenetic profile of this region to that of Chapter 3.3.12 would allow for a more detailed characterisation. Particularly, if methylation is affecting gene expression, this may have been overlooked so far in this chapter as genes that have been omitted from the investigation also reside at this locus. As such, I acquired data generated by Dr Michael Rushton (*personal communication*), a member of Prof. Loughlin's research group (Institute of Cellular Medicine, Newcastle University), whereby genome-wide methylation levels were assessed on an Illumina Infinium HumanMethylation450 BeadChip array (Rushton *et al.*, 2014). I extracted the data for all 212 CpG sites that were annotated within 2 Mb of rs10492367: ranging from cg09778963 (982,321 bp downstream of rs10492367) to cg27198040 (897,375 bp upstream of rs10492367). For every CpG site, the cartilage methylation profile of 17 hip OA (13 GG, 4 T carriers), 63 knee OA (46 GG, 17 T carriers) and 35 NOF (29 GG, 6 T carriers) donors

were included in the various comparisons that I performed. To correct for multiple comparisons, a new significance threshold of 0.0002 was calculated using Equation 3.1.

When comparing the average levels of methylation in hip OA donors with NOF donors, 15 CpG sites remained significant after Bonferroni correction, ten of which had a difference in methylation $\geq$ 5% between the two groups (Figure 4.13.A). Subsequently, the average levels of methylation for the 212 CpG sites were compared between hip OA GG and hip OA T carriers (Figure 4.13.B). In this case, no significant differences in methylation at any of the CpG sites endured following Bonferroni correction, and so there is no evidence to suggest methylation is a mechanism by which rs10492367 modulates OA susceptibility.

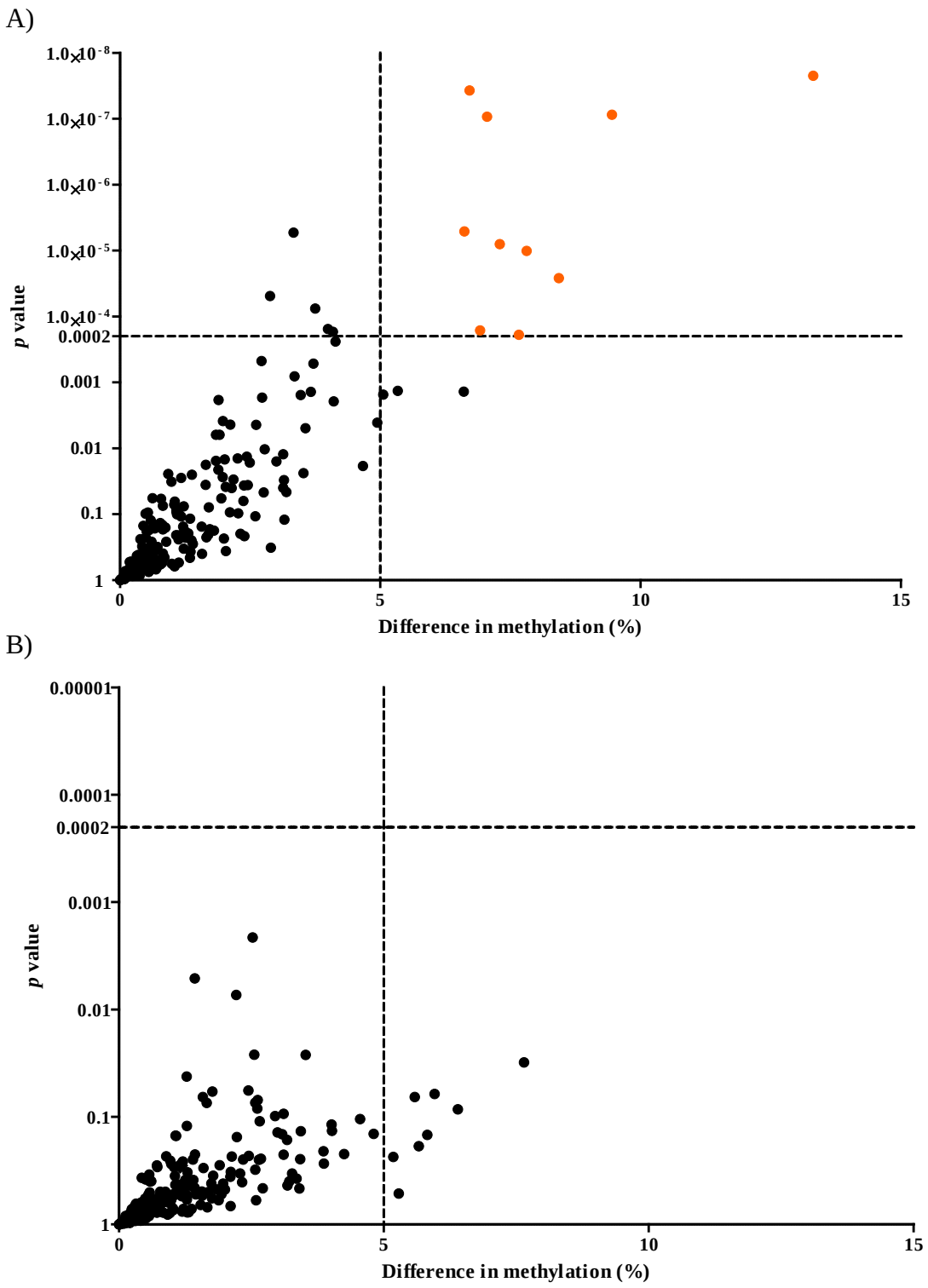

Figure 4.13. Scatter plots to compare the levels of methylation in hip OA and NOF donors. Two hundred and twelve CpG sites, all within 2 Mb of rs10492367, were analysed for differential methylation in A) hip OA vs NOF donors and B) hip OA GG vs hip OA T carriers. Ten of the fifteen CpG sites that remained significant after Bonferroni correction also differed by ≥ 5%, marked by orange circles. None of these CpG sites significantly differed between hip OA GG and hip OA T carriers. Statistical significance was assessed using the Student's *t* test. Horizontal dotted lines represent the significance threshold after Bonferroni correction.

4.3.13 *Identification of CpG sites within 20 bp upstream or downstream of rs10492367 or any polymorphism in high linkage disequilibrium with it*

Although the Illumina Infinium HumanMethylation450 BeadChip array covers 99% of RefSeq genes, with 96% coverage of CpG islands, the average number of CpG sites per gene region included on the array is only 17. This means that when performing an epigenome-wide association scan, a considerable proportion of the 28 million CpG sites of the human genome will be missed. Therefore, rather than acting through a CpG site captured on the array, it may be that this region modulates OA susceptibility through another CpG site in close proximity to the association SNP or a SNP in high LD with it. During the final year of my Ph.D, I designed a research project for an undergraduate student, Brooke Reed, to investigate this. The work detailed in the following sub-section (Chapter 4.3.14) was performed by Brooke under my supervision.

Firstly, a 40 bp region (20 bp upstream and 20 bp downstream) surrounding rs10492367 and each of the SNPs in an LD of $\geq$ 0.80 with it were scanned for CpG sites (Table 4.15). rs10743612 had a CpG site at -2 bp relative to the polymorphism, while rs11049207 created a polymorphic CpG site. These CpG sites from the two regions were selected for subsequent analyses.

SNP	40 bp sequence surrounding SNP (5′-3′)
rs58649696	GCATTCATCTGCCTCTTTCA[C/T]TTTCCTAATGGGACTTTGTA
rs57380671	ACAACTGCTTTTGGCATTTT[C/T]CTCATGAATTTTTTGCCCAT
rs61916489	ACAGCTGTCAATTATGTAAA[A/G]TGTAAATGATATGAAACTGG
rs11049204	CATTTTAAAAAATGAAACTG[A/G]ATAGAAAGATCATAGGCAAA
rs10492367	GTTCTACTTATTATTAGACC[C/A]AGAGTGCTAGAGAGAAAGTG
rs10743612	CTGGGCCTTTCTGCCATGCG[G/A]CACCATAAAAATAAGAGATG
rs11049206	GTTTGTTTTGCTGGTGTTTG[G/C]TTTATACATAAACATGAGTG
rs11049207	TTGGTGCTTGTGTGTGTGGC[G/A]GTTCTATCTTCTAGGAGGAA
rs79881709	ATGGTTTCCTGAACTGGAAG[G/A]ATCTCACTCTCCAACTTGGT
rs10843013	GATAGAGCAGTACCAGTTTT[A/C]ACTAGCACAGAAGTACTTGA

Table 4.15. DNA sequences surrounding the SNPs in high LD ($\geq$ 0.80) with rs10492367. Sequences were analysed for CpG sites (underlined in red text) within 40 bp of the SNPs (in [square] brackets).

4.3.14 *Characterisation of the methylation profiles of CpG sites within 20 bp upstream or downstream of rs10743612 and rs11049207*

The first aim of this investigation was to assess if the CpG sites surrounding rs104743612 and at rs11049207 were differentially methylated in the cartilage of hip OA, knee OA and NOF donors. Secondly, the data were to be stratified by rs10492367 genotype to assess if methylation of the CpG sites correlated with the association SNP genotype. gDNA extracted from the cartilage samples was bisulfite converted, a treatment that converts unmethylated cytosine bases to thymine bases: methylated cytosine bases are protected from this process (Chapter 2.25). Therefore, pyrosequencing can be used to sequence the region and detect levels of methylation at these sites, detailing the proportion of protected cytosine bases relative to converted thymine bases (Chapter 2.26). Primers were designed to the bisulfite converted sequences of the three regions (Appendix A: Table A.4).

Methylation at the CpG site -2 bp of rs10743612 was significantly lower in NOF cartilage relative to hip OA and knee OA cartilage (Figure 4.14.A), however this did not correlate with rs10492367 genotype in the cartilage samples analysed (Figure 4.14.B) or when only hip OA cartilage was studied (Figure 4.14.C). Similarly, methylation at the polymorphic rs11049207 CpG site was significantly lower in NOF cartilage relative to knee OA cartilage (Figure 4.15.A), however, again this did not correlate with rs10492367 genotype (Figure 4.15.B and Figure 4.15.C). Both CpG sites had greater than 50% methylation.

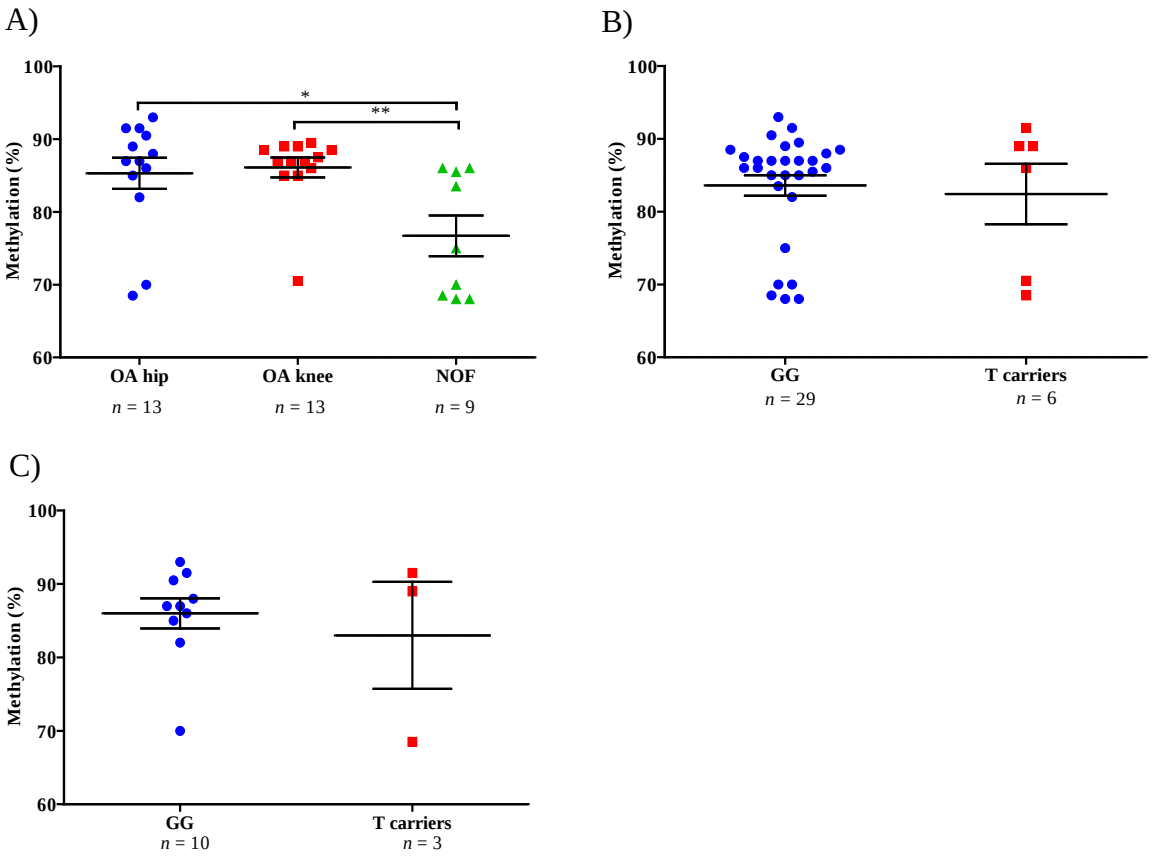

Figure 4.14. Analysis of the levels of methylation in OA hip, OA knee and NOF cartilage at the CpG site 2 bp upstream of rs10743612. Methylation at the CpG site in cartilage was quantified using pyrosequencing and compared between A) hip OA, knee OA and NOF. Levels were significantly decreased in NOF relative to both hip OA and knee OA. The data were stratified by genotype at rs10492367 in B) all donors and C) hip OA donors only. There were no significant correlations between rs10492367 genotype and methylation. Error bars represent the SEM. Statistical significance was assessed using the Mann-Whitney *U* test. * $p < 0.05$; ** $p < 0.01$.

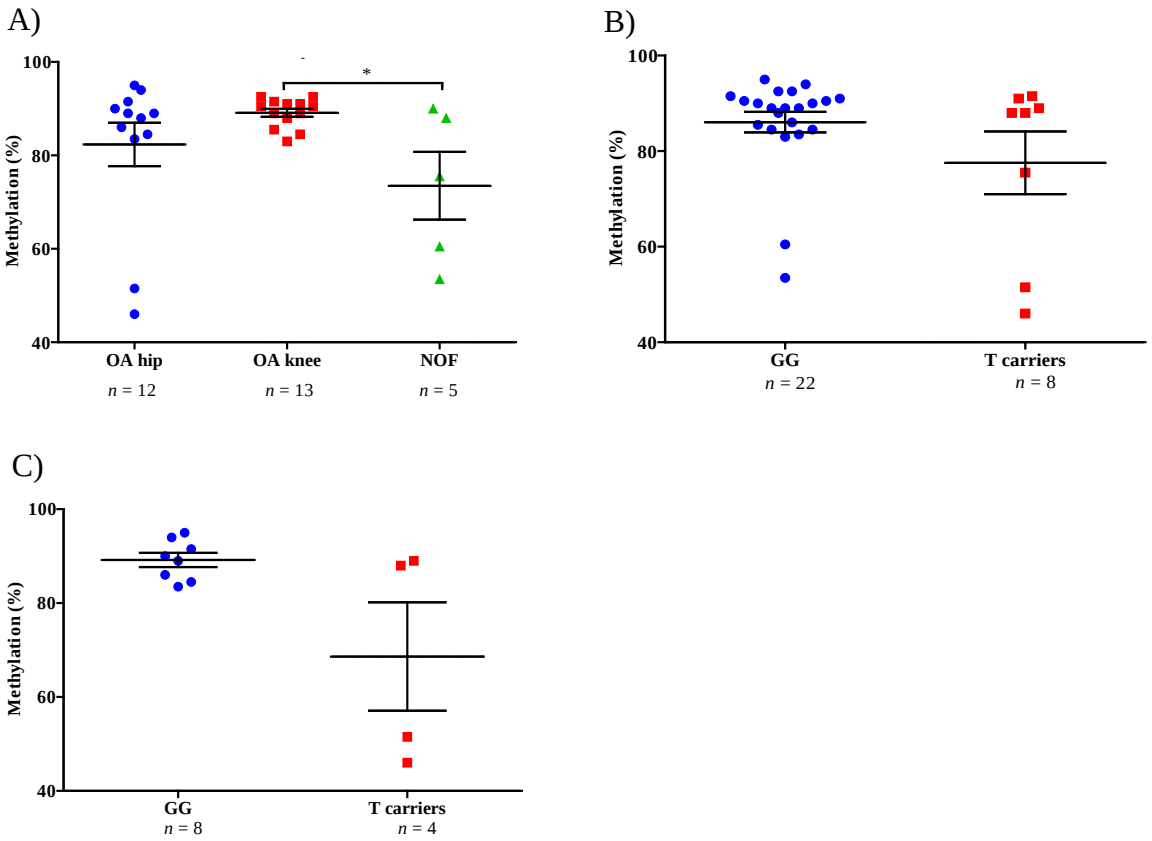

Figure 4.15. Analysis of the levels of methylation in hip OA, knee OA and NOF cartilage at the CpG site at rs11049207. Methylation at the CpG site in cartilage was quantified using pyrosequencing and compared between A) hip OA, knee OA and NOF. Levels were significantly decreased in NOF relative to knee OA. The data were stratified by genotype at rs10492367 in B) all donors and C) hip OA donors only. There were no significant correlations between rs10492367 genotype and methylation. Error bars represent the SEM. Statistical significance was assessed using the Mann-Whitney *U* test. * $p < 0.05$.

4.3.15 *Investigation into the effects on enhancer activity of the CpG sites surrounding rs10743612 and rs11049207*

To further characterise the methylation profile of the two regions, I sought to investigate the effect methylation has on enhancer activity. Regions surrounding the CpG sites were cloned into pCpGL-basic/EF_1 luciferase reporter vectors (Chapter 2.27). To achieve this, the fragments that were cloned into pGL3-promoter vector constructs (discussed in Chapter 5.3.3) were ligated into the *Pst*I and *Spe*I restriction sites of the pCpGL-basic/EF_1 vector (Appendix C: Figure C.3), using the primers listed in Appendix A: Table A.4. Sequencing confirmed the correct ligation of the fragments into the plasmids, and revealed the presence of an additional two CpG sites in each of the allelic constructs compared to the smaller fragments used in Chapter 4.3.14. Treatment with the methyltransferase enzyme *M.Sss*I consequently only methylated the CpG sites of the inserted fragments, and not the CpG-free vector backbone,

meaning a direct comparison of this methylation relative to the corresponding mock-methylated vector was possible (Chapter 2.28). The methylation treatments were confirmed for all rs10743612 and rs11049207 constructs by digesting the vectors with methylation-sensitive enzymes. A representative example is shown in Figure 4.16, where methylation at the restriction site protected the vector from digestion.

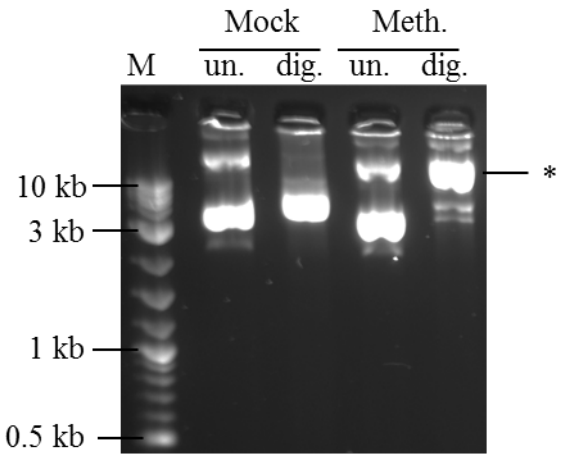

Figure 4.16. Gel electrophoresis of the digested products of the pCpGL rs10743612 A allele construct after *in vitro* methylation with *M.Sss*I and mock methylation. The pCpGL constructs were digested with *Bce*AI. Undigested and digested samples were electrophoresed through a 1% agarose TBE gel. Mock-methylation does not protect the product marked by an asterisk (*) from digestion. The banding pattern for the undigested methylated vector mirrors that of the mock-methylated construct. When the methylated vector is treated with *Bce*AI, the monomer is protected from digestion. M (marker), mock (mock-methylated), meth. (*in vitro* methylated with *M.Sss*I), un. (undigested product), dig. (digested product).

All vectors were transfected into the SW1353 chondrosarcoma and U2OS osteosarcoma cell lines followed by quantification of the luciferase reporter gene expression after 24 hours (Chapter 2.29). The luciferase/*Renilla* absorbances of the *M.Sss*I-treated and mock-methylated vectors were normalised to the correspondingly treated empty vectors (Figure 4.17). One of the most notable activities of the constructs was the overall heightened enhancer activity of rs11049207 for all constructs relative to the empty vector controls in the SW1353 cell line. In terms of investigating the effects of methylation, there were no significant differences in the enhancer activity of the methylated form of either allele construct of rs10743612 relative to the non-methylated controls. Similarly, there were no differences in either cell line for the G allele constructs of rs11049207. When methylated, the A allele construct rs11049207 had a significantly lower enhancer activity relative to the non-methylated control in both cell lines. This is particularly interesting as the A allele of rs11049207 replaces the G of the polymorphic CpG site, and thus there is no CpG site to be

methylated. Instead, it may be the case that the genotype of rs11049207 correlates with the methylation of a distal CpG site that was captured in the cloned fragment. Overall, therefore, by methylating the A allele construct of rs11049207, the A allele could influence the binding of DNA methyltransferases at the distal CpG site, and in turn cause a decrease in enhancer activity: the G allele may not have such an influence.

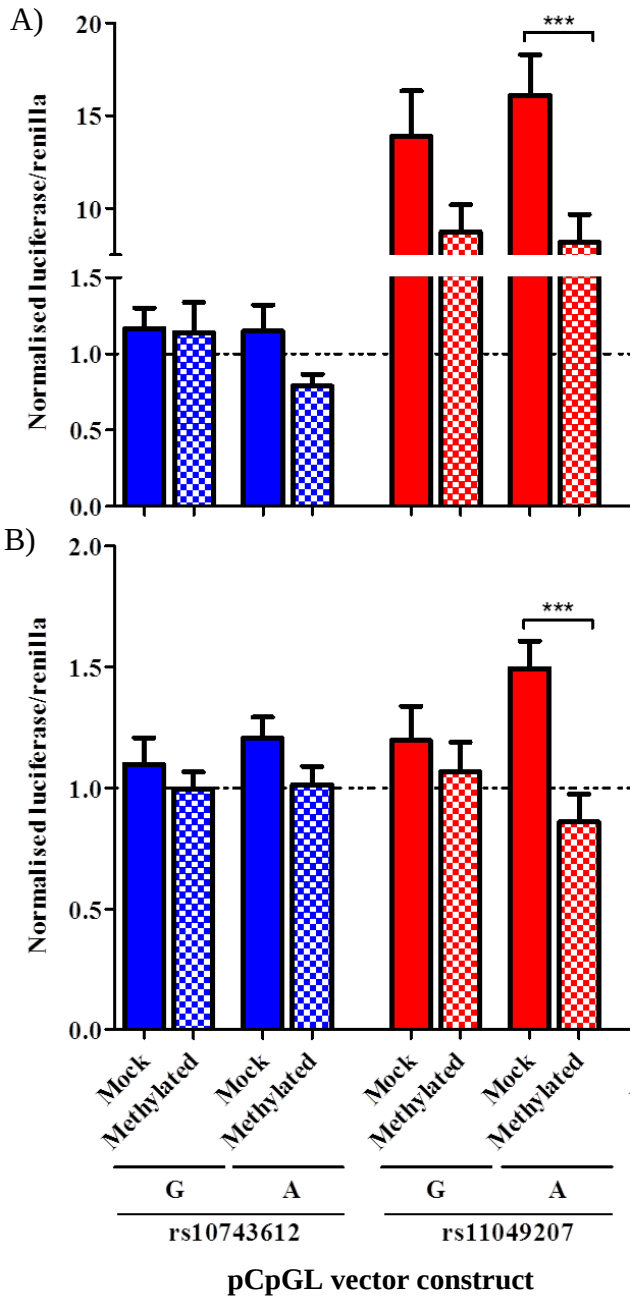

Figure 4.17. Investigation into the effects of methylation on the enhancer activity of rs10743612 and rs11049207. Fragments surrounding the alleles of the polymorphisms were excised from the previously created pGL3-promoter vector constructs and ligated into pCpGL-basic/EF_1 luciferase reporter vectors. The constructs were either methylated with *M.Sss*I or mock-methylated before transfection into A) SW1353 and B) U2OS cell cultures. The activity of the luciferase gene downstream of the inserts was quantified after 24 hours. All absorbances were read at 595 nm and the luciferase readings normalised to the internal control (*Renilla*). Luciferase/*Renilla* values were normalised to the basal levels of activity measured in the empty vector controls with the corresponding treatments ($y = 1$). Six technical replicates were performed per three independent replicates for both cell line transfections. Error bars represent the SEM. Statistical significance was assessed using the Mann-Whitney *U* test. *** $p < 0.001$.

4.4 Discussion

The aim of the research in this chapter was to characterise the OA association locus marked by the polymorphism rs10492367. This was achieved by mirroring the rationale and practical investigations of the 6p14.1 locus (Chapter 3). I hypothesised that the association SNP marks a functional polymorphism that acts to regulate the expression of at least one of the genes at this locus. Prior to the start of my Ph.D, the expressions of *PTHLH* and *KLHL42* were quantified by Dr Madhushika Ratnayake and Dr Emma Raine, with *PTHLH* being an excellent candidate gene for OA susceptibility. As such, I continued with the subsequent characterisation of this locus, meaning that that genes distal to *PTHLH* and *KLHL42* were excluded. As for rs9350591, it was not necessary to genotype any polymorphism other than the association signal, as the nature of LD meant that functional correlations of other SNPs should be detected through rs10492367 (Styrkarsdottir *et al.*, 2014).

I first postulated that the association signal could be mediating its effects during joint development, particularly as *PTHLH* has a well-established role in endochondral ossification. I utilised microarray data that quantified gene expression during MSC differentiation down a chondrogenic lineage and an osteoblastic lineage, showing that both genes were expressed throughout the time courses. I additionally considered the other genes within 1 Mb upstream and 1 Mb downstream of rs10492367 for both time courses. For chondrogenesis, the expression of *STK38L*, *REP15* and *PPFIBP1* were increased by day 14 of differentiation, while the expression of *PTHLH*, *MRPS35* and *TM7SF3* were decreased. Particularly for *PTHLH*, this could explain why its expression was so low in synovial joint tissues. Only *TM7SF3* was significantly upregulated during osteoblastogenesis, which is in contrast to its chondrogenic profile and is not substantiated by any known roles of the gene. Chondrogenesis was independently investigated using qPCR for *PTHLH* and *KLHL42* only, and again confirmed the dynamic gene expression patterns, overall suggesting that either of the genes have the potential to modulate OA susceptibility.

I then analysed data that were generated prior to the beginning of my Ph.D to confirm that the genes were expressed in articular cartilage; additionally, I confirmed the expressions in fat pad and synovium. Both *PTHLH* and *KLHL42* were expressed at relatively low levels, however their diverse expression profiles support the evidence that the genes are not tissue-specific. Once I had confirmed that both genes were expressed in the synovial joint tissues tested, I was able to begin to investigate the OA association SNP.

The two main ways to do this, as discussed in Chapter 3, are overall gene expression quantification and allelic expression quantification. The former technique considers the overall total abundance of the gene transcripts, while the latter method interrogates the genes in a much more in-depth manner. Investigating the allelic expression allows the quantification of mRNA transcripts and thus focusses on whether the association signal specifically modulates this output.

Fat pad, synovium and cartilage excised from the knee of OA donors were included alongside hip cartilage analyses for overall gene expression quantification. In all tissue types, there were no significant differences in either of the gene expressions relative to the OA association SNP genotype. Nevertheless, *PTHLH* and *KLHL42* expressions were differentially expressed in cartilage depending on the disease state and joint site. This suggests that the genes might contribute to joint-specific OA development (Karlsson *et al.*, 2010) but are acting independently of the OA association signal detected by the arcOGEN study. For example, *PTHLH* expression was increased in OA female cartilage relative to non-OA control cartilage, an observation similarly reported in OA knee cartilage (Terkeltaub *et al.*, 1998).

The aim of AEI analysis is to identify if there are differences in the mRNA outputs that correspond to the different alleles of a heterozygote individual (Wang and Elbein, 2007). Although the AEI investigation of Chapter 3 yielded no correlations with association SNP genotype, other studies have previously shown positive results in the identification of *cis*-eQTLs (Bos *et al.*, 2012; Raine *et al.*, 2012; Gee *et al.*, 2014). Despite rs10492367 being associated with only hip OA, it is possible to detect AEI in more than one joint tissue (Egli *et al.*, 2009; Bos *et al.*, 2012; Styrkarsdottir *et al.*, 2014), and so knee cartilage was combined with hip cartilage is this study. However, following this in-depth approach of investigating AEI, there were still no significant differences in gene expression that could be attributed to rs10492367 genotype. An explanation for this could be that the association marked by rs10492367 may be regulating the expression of a gene other than those characterised, as *cis*-eQTLs can act at up to megabase distances (Nica and Dermitzakis, 2013).

DNA methylation is now an established mechanism through which the genome can be regulated and has been reported to correlate with DNA sequence variations (Bell *et al.*, 2011; Smith *et al.*, 2014). Unlike the 6q14.1 locus, however, analysis of the CpG sites across a 2 Mb span of the region did not implicate methylation in the modulation of OA association. Following this, I designed an undergraduate project to further probe the region, as the

microarray used for this investigation could have excluded functional CpG sites. The project interrogated CpG sites surrounding the association SNP and those in high LD with it, and had three aims: i) to characterise methylation in hip OA, knee OA and NOF cartilage, ii) to identify any correlation with rs10492367 genotype, and iii) to investigate the effect of CpG site methylation on enhancer activity. Stratification by disease state often correlated with differential methylation of the CpG sites suggesting, as for overall gene expression, that these loci could be modulating OA susceptibility but are unrelated to the association signal. However, the assays need to be validated to ensure these results are robust, as this was not attainable in the timeframe of the undergraduate project.

The final aim of investigating the methylation of this region was to identify if differential methylation of the CpG sites affected the enhancer activity of the region. pGL3-promoter vector constructs were already generated (Chapter 5) that contained the CpG sites of interest, and so the inserts were taken from the constructs and ligated into CpG-free pCpGL-basic/EF_1 luciferase reporter vectors. A caveat of this approach is that the fragments were not identical to those used in the pyrosequencing, meaning an additional two CpG sites were captured in each of the inserts. It is impossible, therefore, to rule out the effects of additional CpG sites on the differential enhancer activities observed. This is exemplified by rs11049207, where methylation of the A allele construct caused a significantly decreased enhancer activity relative to the non-methylated control, even though the presence of the A allele replaces the G allele of the CpG site. Overall, these results reinforce the functional effects that methylation could have on the OA association of this region.

It is clear that, like the signal studied in Chapter 3, the association signal marked by rs10492367 does not modulate its effects in the end-stage OA cartilage, fat pad or synovium tissues tested. Nevertheless, the genes at this locus are all dynamically expressed throughout chondrogenesis and osteoblastogenesis, which implies a potential role for any of the genes to act at earlier stages of joint development. Moreover, differential methylation at CpG sites at or near SNPs in high LD with rs10492367 was observed between joint site and disease state, while methylation of the A allele construct of rs11049207 appeared to regulate enhancer activity of the 12p11.22 locus. Overall, this chapter highlights the huge amount of diversity at this OA association locus, and shows that there is scope to further investigate the functional region and its association to OA.

Chapter 5. Functional Studies of the 12p11.22 Locus Marked by the Polymorphism rs10492367

5.1 Introduction

Although there was no correlation observed between the carriage of the risk allele of rs10492367 and expression of either *PTHLH* or *KLHL42* in the end-stage OA cartilage tested, it by no means implies that the region is not functional in OA susceptibility. The association signal may exert its effects in other tissue types, at a different stage of development or by modulating gene expression outside the region of LD. Indeed, Table 4.2 suggests several of the SNPs are in a functionally active state in bone marrow-derived cultured MSCs, MSC-derived cultured cells and osteoblast primary cells. It was therefore necessary to functionally dissect the association signal, by investigating the potential of the region to modulate enhancer activity in relevant cell lines. Based on this, I hypothesised that rs10492367, or a polymorphism in high LD with it, is functional in regulating the enhancer activity of the region. Should differential activity be observed, it would be necessary to further investigate the identity and function of *trans*-acting factor binding that could account for such differences.

5.2 Aim

The aim of this chapter was to characterise the functionality of the association signal marked by rs10492367. This was achieved by:

- cloning DNA fragments of each allele of rs10492367 and of the polymorphisms in high LD with it into pGL3-promoter vectors
- investigating if there is allele-specific differential enhancer activity of the regions in chondrosarcoma and osteosarcoma cell lines
- creating a set of criteria that allows a ranking system to identify the SNP(s) that should be carried forward for further functional analysis
- characterising protein binding to the selected SNP(s) using EMSAs
- confirming protein binding *in vitro* in a chondrosarcoma cell line using ChIP
- investigating the effect of protein binding on target gene expression by knocking-down the transcription factors in a primary cell line
- assessing if knocking-down transcription factors in a chondrosarcoma cell line has differential effects depending on which allele of rs10492367 is present

5.3 Results

5.3.1 *Selection of polymorphisms for functional characterisation of the region*

Following the database searches detailed in Figure 4.1 and Table 4.2, it was clear that the OA association SNP resides in a predicted enhancer region with transcription factors known to bind in various cell lines, including MSC-derived chondrocytes. In addition to this, there are nine polymorphisms that are known to be in an LD of $\geq$ 0.80 with rs10492367, all of which are intergenic (Table 5.1). I therefore hypothesised that rs10492367, or a SNP in high LD ($r^2 \geq 0.80$) with it, acts on target gene expression in an allele-specific manner by regulating the enhancer activity of the region.

SNP	bp from rs10492367	r^2 relative to rs10492367	D' relative to rs10492367	Major/minor allele	MAF
rs58649696	10,617	0.853	1.000	C/T	0.183
rs57380671	4,563	0.806	1.000	C/T	0.233
rs61916489	1,682	0.853	1.000	G/A	0.183
rs11049204	1,407	1.000	1.000	A/G	0.208
rs10492367	0	1.000	1.000	G/T	0.208
rs10743612	421	0.865	1.000	G/A	0.233
rs11049206	531	1.000	1.000	G/C	0.208
rs11049207	3,683	0.853	1.000	G/A	0.183
rs79881709	4,054	0.853	1.000	G/A	0.183
rs10843013	10,226	1.000	1.000	A/C	0.208

Table 5.1. SNPs in high LD ($r^2 \geq 0.80$) with rs10492367. All of the SNPs are intergenic between *PTHLH* and *KLHL42*, three of which are in perfect LD with the association SNP. The association signal is highlighted by a dotted box.

Before commencing on any functional analyses of the region, it was important to identify whether or not the SNPs in Table 5.1 were themselves genotyped on the arcOGEN array, or if they were in perfect LD with other SNPs that were captured by the array. The cut-off r^2 used in this selection was $\geq$ 0.80 relative to rs10492367, the premise being that any of the SNPs could be causal in the OA association of the region, but with the effects detected via rs10492367 as this was the SNP genotyped on the microarray. Therefore, if any of the polymorphisms are in high LD with another SNP genotyped on the array, and the r^2 is greater than the r^2 relative to the association SNP, it is likely that the OA association would have additionally been detected by that SNP. As such, I performed a SNAP Proxy Search for each of the SNPs (Table 5.2), and identified that none of the r^2 values were greater than those relative to rs10492367. It was therefore reasonable to carry forward all the SNPs for characterisation of the region.

SNP in LD ≥ 0.80 with rs10492367	SNP genotyped on the arcOGEN array	Distance from SNP of interest (bp)	r^2 relative to SNP of interest	D' relative to SNP of interest
rs58649696	rs258394	533	0.853	1.000
	rs258396	1,301	0.853	1.000
	rs10492367	**10,617**	**0.853**	**1.000**
	rs10843013	20,843	0.853	1.000
rs57380671	**rs10492367**	**4,563**	**0.806**	**1.000**
	rs10843013	14,789	0.806	1.000
rs61916489	**rs10492367**	**1,682**	**0.853**	**1.000**
	rs258396	7,634	0.853	1.000
	rs258394	8,402	0.853	1.000
	rs10843013	11,908	0.853	1.000
rs11049204	**rs10492367**	**1,407**	**1.000**	**1.000**
	rs10843013	11,633	1.000	1.000
rs10492367	**rs10492367**	**0**	**1.000**	**1.000**
	rs10843013	10,226	1.000	1.000
rs10743612	**rs10492367**	**421**	**0.865**	**1.000**
	rs258396	9,737	0.865	1.000
	rs10843013	9,805	0.865	1.000
	rs258394	10,505	0.865	1.000
rs11049206	**rs10492367**	**531**	**1.000**	**1.000**
	rs10843013	9,695	1.000	1.000
rs11049207	**rs10492367**	**3,683**	**0.853**	**1.000**
	rs10843013	6,543	0.853	1.000
	rs258396	12,999	0.853	1.000
	rs258394	13,767	0.853	1.000
rs79881709	**rs10492367**	**4,054**	**0.853**	**1.000**
	rs10843013	6,172	0.853	1.000
	rs258396	13,370	0.853	1.000
	rs258394	14,138	0.853	1.000
rs10843013	rs10843013	0	1.000	1.000
	rs10492367	**10,226**	**1.000**	**1.000**

Table 5.2. List of SNPs that are in high LD (r^2 ≥ 0.80) with rs10492367, detailing the polymorphisms that are both in high LD with them and that were genotyped by the arcOGEN study (arcOGEN Consortium *et al.*, 2012). None of the SNPs had a higher LD with the other polymorphisms than that relative to rs10492367 (in **bold**).

5.3.2 *Selection of cell lines and characterisation of their expressions of PTHLH and KLHL42*

To functionally characterise the region, I decided to perform luciferase reporter assays. These assays involve the transfection of reporter vectors into a relevant cell line in order to quantify

the resulting expression of the encoded luciferase gene. Given that OA is thought to be primarily a disease of cartilage, the SW1353 chondrosarcoma cell line was selected for use in the luciferase reporter assays. In addition, as *PTHLH* has a widely known role in endochondral ossification (Wysolmerski, 2012) the U2OS osteosarcoma cell line was also selected. As transcription factors are a likely modulator of gene expression, which will be investigated should differential enhancer activity be observed, it was important to confirm that *PTHLH* and *KLHL42* were expressed. Moreover, as future experiments could be directed more toward primary chondrocytes, I also investigated the gene expressions in five HAC donors (Figure 5.1). Cell culture conditions are detailed in Chapter 2.15 and Chapter 2.17. Gene expression was detected in all cell lines. It was therefore inferred that the transcription factors potentially modulating the expressions were also expressed in the cell lines. Genotyping of DNA extracted from SW1353 and U2OS cell cultures revealed both cell lines were homozygous for the major allele of rs10492367.

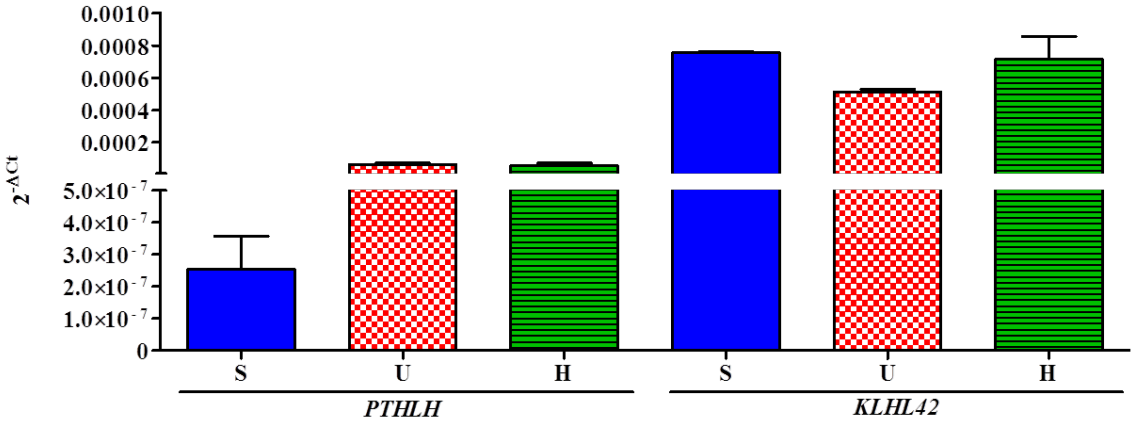

Figure 5.1. Expression of *PTHLH* and *KLHL42* in SW1353, U2OS and HAC cells. Both genes were expressed in SW1353 (S), U2OS (U) and HAC (H) cells. RNA was extracted from cultures of SW1353 and U2OS cells, cDNA synthesised and three technical repeats used for qPCR to assess gene expression. RNA was extracted from five HAC donors, cDNA synthesised and three technical repeats performed for qPCR per donor. Error bars represent the SEM.

5.3.3 *Investigating the allele-specific effects on the enhancer activity of rs10492367 and the polymorphisms with an $r^2 \geq 0.80$ relative to it*

To assess if the region is functional in regulating enhancer activity, I performed luciferase reporter assays. As the total distance covered by the polymorphisms was almost 21 kb from rs58649696 to rs10843013, it was not possible to investigate the functionality of the region as a whole. As such, I designed primers (Appendix A: Table A.2) covering regions of no more than 650 bp surrounding each individual polymorphism and used genomic DNA from major

and minor allele homozygotes for the allele templates. At this stage, the repetitive nature of the surrounding DNA meant that a single, specific amplimer for rs57380671 could not be produced and was therefore excluded from the experiment. Due to technical difficulties experienced in identifying minor allele homozygotes, site-directed mutagenesis was used to generate minor allele amplimers for rs11049204 and rs79881709 (Chapter 2.14). The regions were ligated into pGL3-promoter vectors (Appendix C: Figure C.1) that contained an SV40 promoter and a multiple cloning site upstream of the luciferase gene, and crucially did not possess an enhancer (Chapter 2.13). All vectors were co-transfected with a *Renilla* control vector into chondrosarcoma and osteosarcoma cell lines, followed by quantification of the luciferase and *Renilla* reporter gene expressions after 24 hours (Chapter 2.16). The empty pGL3 promoter vector was used as a control, and the luciferase divided by *Renilla* value of this control plasmid gives an arbitrary value of 1.

Taken as a whole, it was apparent that the regions assayed had functional activities particularly in the chondrosarcoma cell line (Figure 5.2). rs61916489, rs11049204, rs10743612 and rs11049207 had normalised *y*-axis values > 1, suggesting an increase in levels of enhancer activity relative to the basal levels of the empty vector. Conversely, rs58649696 and rs10843013 both had apparent repressive activities relative to the basal empty vector levels. In the osteosarcoma cell line, the values more closely resembled the activity of the empty vector. There were no significant differences in either cell line between the alleles of rs61916489, rs11049204 and rs10743612. rs58649696 and rs11049206 were the only polymorphisms that were significantly different in both cell lines. There was a decrease in enhancer activity in the presence of the minor allele of rs58649696 ($p = 0.0009$ [SW1353] and $p < 0.0001$ [U2OS]), and this was mirrored by rs11049206 ($p = 0.002$) in the osteosarcoma cell line. Conversely, the presence of the minor allele of rs11049206 ($p = 0.0003$) in the chondrosarcoma cell line caused an increase in enhancer activity. The remaining polymorphisms – rs10492367 ($p < 0.0001$), rs11049207 ($p = 0.019$), rs79881709 ($p < 0.0001$) and rs10843013 ($p = 0.037$) – all differed in the chondrosarcoma cell line only. The minor alleles of rs11049207 and rs79881709 both caused a decrease in enhancer activity relative to the major alleles, whereas the minor alleles of the association SNP and rs10843013 caused an increase in enhancer activity.

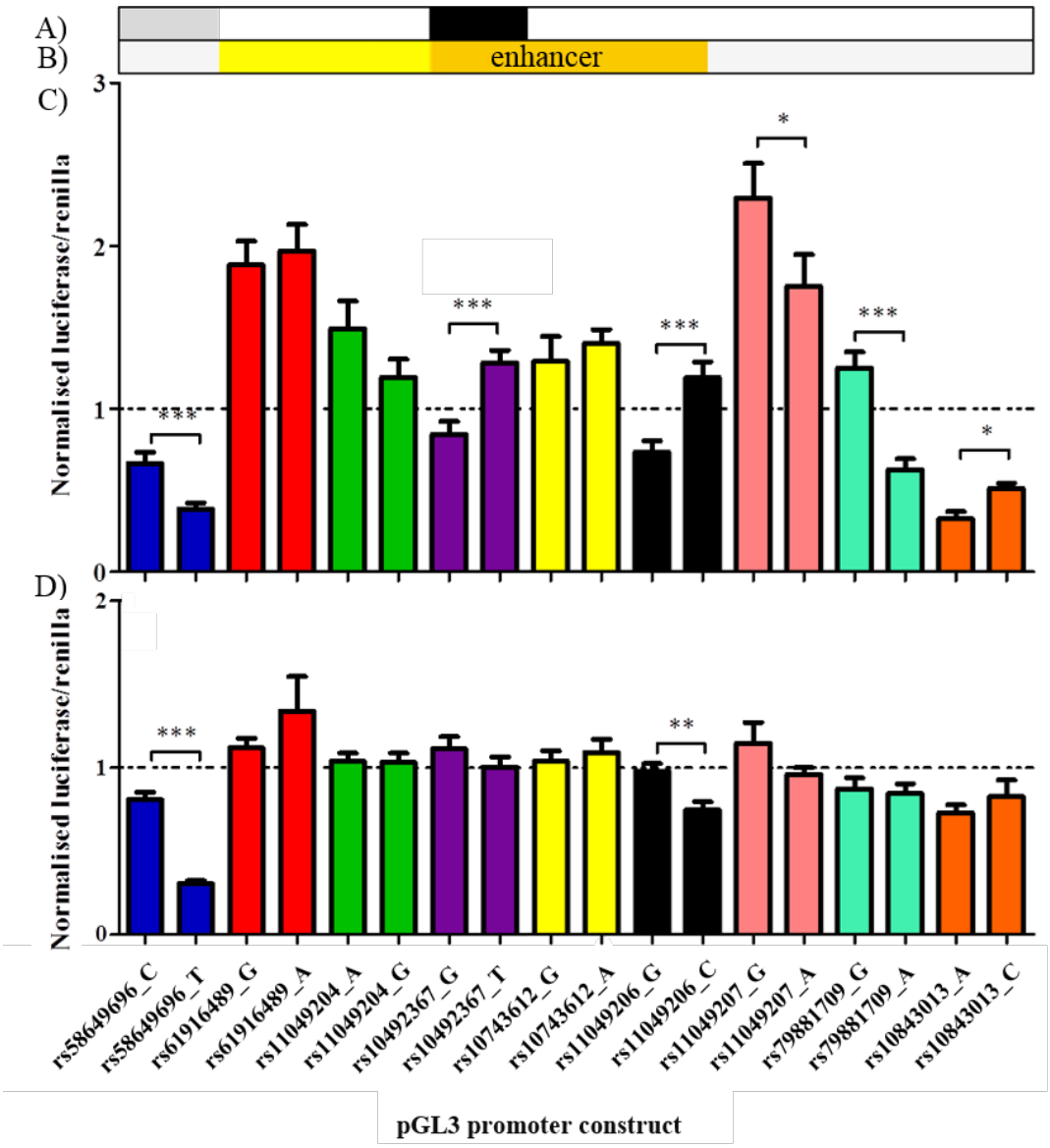

Figure 5.2. Investigation into the allelic effects of rs10492367 and the SNPs in high LD with it. Fragments surrounding the alleles of the association SNP and SNPs in high LD with it were ligated into pGL3-promoter vectors. The constructs were transfected into C) the SW1353 chondrosarcoma cell line and D) the U2OS osteosarcoma cell line, and after 24 hours the activity of the luciferase gene downstream of the inserts was quantified. All absorbances were read at 595 nm and the luciferase readings normalised to the internal control (*Renilla*). Luciferase/*Renilla* values were normalised to the basal levels of activity measured in the empty vector control (*y*-axis = 1). A) Transcription Factor ChIP-Seq track and B) Chromatin State Segmentation track both taken from the UCSC Genome Browser: a black or grey box in A) indicates transcription factor binding and an orange or yellow box in B) indicates the region is predicted to have enhancer activity. Six technical replicates were performed per six independent replicates for SW1353 transfections and four independent replicates for U2OS transfections. Error bars represent the SEM. Statistical significance was assessed for the C) SW1353 and D) U2OS cell line luciferase assay results using the Mann-Whitney *U* test. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

5.3.4 *Selection of polymorphisms for further functional analysis*

The dynamic functionality of the region certainly warranted further investigation, with the aim of the next step being to identify transcription factors that bind the polymorphisms. It would not be realistic to investigate all the polymorphisms, or even the six SNPs that showed differential allelic enhancer activity, and so it was necessary to narrow the field of investigation. To do this, I designed a set of criteria that would allow the selection of only the most promising polymorphisms. The factors that I deemed most important in determining whether a polymorphism could be contributing to the association signal were:

- the r^2 relative to rs10492367: an ideal situation would be an r^2 of 1.000
- chromatin state: based on the ENCODE dataset, if the region is predicted to have enhancer activity (as detailed in Figure 5.2), it could be more likely to be functional *in vivo* in regulating gene expression
- transcription factor binding: any relevant indications of protein binding from the ENCODE dataset would be a necessary consideration
- significant differences in enhancer activity between alleles in the SW1353 cell line (Figure 5.2.C): significant differences in allelic activity could translate into an *in vivo* disease susceptibility locus
- significant differences in enhancer activity between alleles in the SW1353 cell line in the same direction as rs10492367: that is, increased enhancer activity with the risk allele (Figure 5.2.C)
- significant differences in enhancer activity between alleles in the U2OS cell line (Figure 5.2.D)

Each polymorphism was scored on these criteria, and those that had a total score of ≥ 3 out of a possible 6 were selected for further analysis (Table 5.3). Based on these criteria, rs58649696, rs10492367, rs11049206 and rs10843013 were identified as the most promising candidates.

SNP	r^2 relative to association SNP	Perfect LD (r^2 = 1) with rs10492367	Resides in predicted enhancer	Transcription factors known to bind	Significant difference in SW1353	Same direction as association SNP in SW1353	Significant difference in U2OS	Overall score (/6)
rs58649696	0.853	×	×	✓	✓	×	✓	3
rs57380671	0.806	×	×	×	No data	No data	No data	0
rs61916489	0.853	×	✓	×	×	×	×	1
rs11049204	1.000	✓	✓	×	×	×	×	2
rs10492367	1.000	✓	✓	✓	✓	✓	×	5
rs10743612	0.850	×	✓	×	×	×	×	1
rs11049206	1.000	✓	✓	×	✓	✓	✓	5
rs11049207	0.853	×	×	×	✓	×	×	1
rs79881709	0.853	×	×	×	✓	×	×	1
rs10843013	1.000	✓	×	×	✓	✓	×	3

Table 5.3. Criteria used for selecting SNPs to carry forward for investigating protein:DNA binding using EMSAs. One point was given for each match of the following criteria: i) an LD relative to the association SNP of 1.00, ii) located within a predicted enhancer region, iii) transcription factors known to bind; iv) significant allelic differences in luciferase activity in the SW1353 cell line, v) significant differences in luciferase activity in SW1353 cells in the same direction as rs10492367 (that is, increased enhancer activity with the risk allele), and vi) significant allelic differences in luciferase activity in the U2OS cell line. All polymorphisms with a score of 3 or above were selected for investigation using EMSAs: rs58649696, rs10492367, rs11049206 and rs10843013. The association signal is highlighted by a dotted box.

5.3.5 *Assessing the purity of nuclear protein extracts from SW1353 cells, U2OS cells and human articular chondrocytes (HACs)*

Having identified differential enhancer activity between the alleles of polymorphisms within the association interval, I next aimed to investigate the transcription factors that could be responsible for the functional effects. EMSAs are a technique used to investigate protein binding to specific DNA sequences, and requires the extraction of nuclear protein from the cell lines of interest. Nuclear protein from SW1353, U2OS and five HAC donor cell cultures was extracted and the purity of the extracts assessed using anti-lamin A/C and anti-GAPDH antibodies on a western blot (Chapter 2.18 – Chapter 2.20). Lamins are structural components of the nuclear membrane whereas GAPDH is involved in glycolysis in the cell cytosol. Lamin A/C was detected in all three nuclear extracts and GAPDH was abundant in the cytosolic extracts (Figure 5.3).

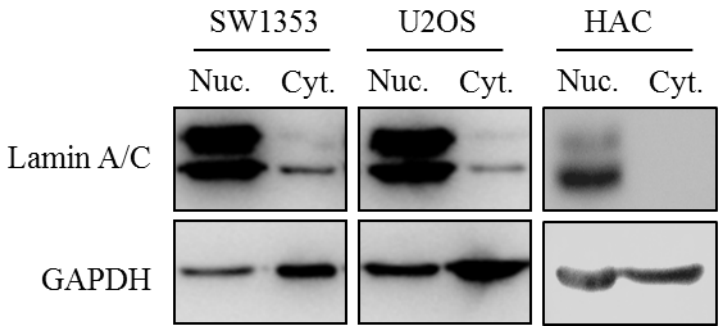

Figure 5.3. Assessment of the purity of the nuclear protein extract from SW1353, U2OS and HAC cell cultures using immunoblotting. The presence of lamin A/C is an indicator of nuclear (Nuc.) protein, while GAPDH indicates cytosolic (Cyt.) protein.

5.3.6 *Investigating transcription factor binding to rs10492367 using chondrosarcoma and osteosarcoma cell line nuclear protein*

I have shown that the risk (T) allele of rs10492367 caused an increase in enhancer activity relative to the non-risk (G) allele. EMSA conditions were first optimised for 5 µg of protein binding to 200 fmol fluorescently-labelled G and T allele probes (Chapter 2.23), using NP-40 and glycerol as optional additional components that have the potential to stabilise and strengthen the protein:DNA complexes (Appendix F: Figure F.1). Protein binding to the G and T allele probes produced seven distinct complexes, highlighted by the numbered arrows, with little difference between the two conditions. Therefore, NP-40 was selected as the only additional component in the EMSA reaction mixes for rs10492367.

Following optimisation of the binding conditions, the specificity and affinity of the protein complexes to the G and T allele probes were tested. This required the use of increasing concentrations of unlabelled competitors identical to the labelled probes. These competition assays were performed with a 5-, 10-, 25- or 50- fold excess of unlabelled competitor relative to the labelled probe (Figure 5.4). Most strikingly, protein binding to both allele probes in complexes 1, 2, 4 and 5 were outcompeted at lower concentrations of the G allele competitor compared to the T allele competitors. This suggests the protein in these complexes bind more avidly to the G allele than the T allele of rs10492367. Moreover, complex 7 appeared to bind the T allele probe specifically, with only a shadowing of a band with the G allele probe. In addition, increasing competition with the T allele competitor with both allele probes caused an increase in the band intensity of complex 4, which could be caused by an increasing availability of the probe from an outcompeted protein:DNA complex.

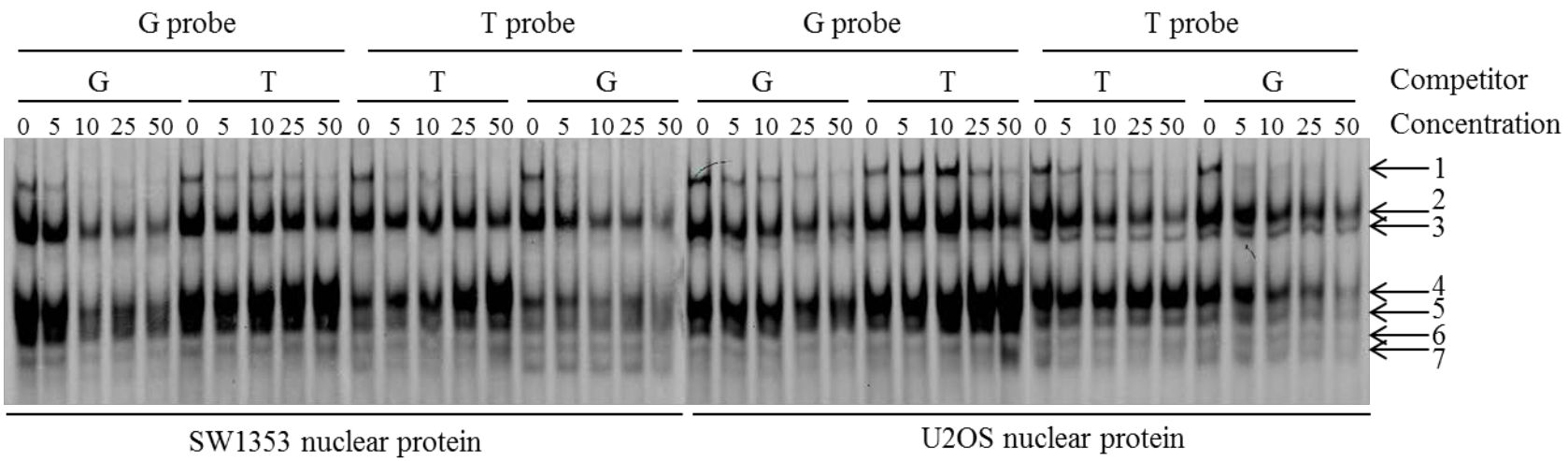

Figure 5.4. Competition EMSAs to investigate allele-specific binding of SW1353 and U2OS nuclear extract to the G and T alleles of rs10492367. The concentrations of unlabelled competitors were increased from 0 to 50 times that of the labelled probes. The protein:probe complexes marked by arrows 3, 6 and 7 were not noticeably disrupted upon increasing competitor concentrations. Protein binding to both allele probes in complexes 1 and 2 were outcompeted by the G allele competitor, and to less of a degree with the T allele competitor. Complexes 4 and 5 were similarly disrupted over the concentration gradient of the G allele competitor, whereas increasing competition with the T allele competitor in both allele probes caused an increase in the band intensity. Concentration = 0, 5, 10, 25, 50 x probe concentration.

The region of the labelled probe that was involved in binding the protein was investigated by using competitor sequences that consisted of one region identical to the labelled probe and the remaining sequence entirely random (Table 5.4). The random sequence had no effect on protein binding to either allele probe, indicating that the protein:DNA complexes are specific for the probe sequences (Figure 5.5). Complexes 2 and 4 binding to the G allele probe were outcompeted by competitor 3, indicating that it is likely to be the 3'-most sequence of the G allele probe that is required for protein binding. The increased protein binding in complexes 6 and 7 with competitor 1 with the T allele probe may have been caused by a new protein binding site being generated by the combination of sequences.

Sequence name	Probe sequence (5'-3')
Random primer full length	ATGGGGCGTGCGATCGTACTGCCTACGGTGG
G allele probe and competitor	TCTCTCTAGCACTCTGGGTCTAATAATAAGT
T allele probe and competitor	TCTCTCTAGCACTCTTGGTCTAATAATAAGT
Competitor 1: G allele	TCTCTCTAGCACTCTGGACTGCCTACGGTGG
Competitor 1: T allele	TCTCTCTAGCACTCTTGACTGCCTACGGTGG
Competitor 2: G allele	ATGGGGCGTCACTCTGGGTCTACTACGGTGG
Competitor 2: T allele	ATGGGGCGTCACTCTTGGTCTACTACGGTGG
Competitor 3: G allele	ATGGGGCGTGCGATTGGGTCTAATAATAAGT
Competitor 3: T allele	ATGGGGCGTGCGATTTGGTCTAATAATAAGT

Table 5.4. Primer sequences used for competition EMSAs to investigate the regions of the G and T allele probes of rs10492367 to which the SW1353 and U2OS nuclear extracts bind. Each primer was annealed to its reverse complement, creating double stranded DNA (dsDNA), prior to use in EMSAs. The original competitor sequences are underlined in red; the random sequences are in black text.

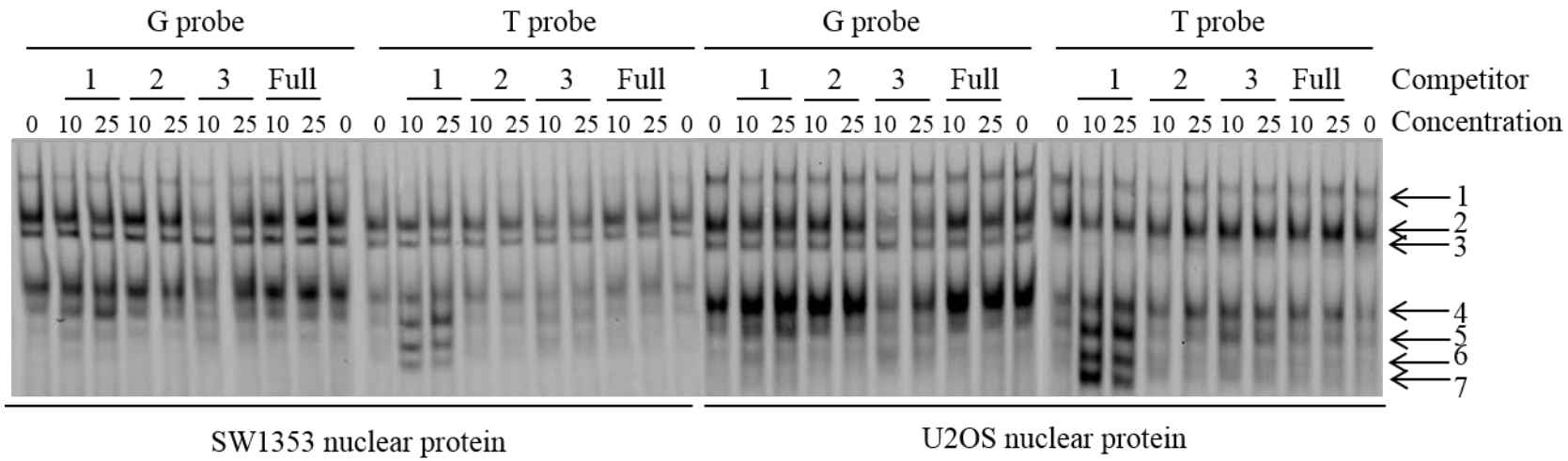

Figure 5.5. Competition EMSAs to investigate the regions of the G and T allele probes of rs10492367 to which the SW1353 and U2OS nuclear extracts bind. The protein:probe mixes were incubated with unlabelled competitors that had random sequences replacing the original competitor sequence. An entirely random competitor (full) had no effect on any of the protein:probe complexes. The G allele probe binding to the protein in complexes 2 and 4 were outcompeted by competitor 3. Competitor 1 caused the band intensity to increase for complexes 5, 6 and 7 with the T allele probe. Concentration = 0, 10, 25 x probe concentration. Competitor = 31 bp random competitor (full); 3′-most region of the competitor replaced by a 14 bp random sequence (1); the central 13 bp region flanked either side by a 9 bp random sequence (2); 5′-most region of the competitor replaced by a 14 bp random sequence (3).

Searches of four databases – UCSC Genome Browser, PROMO, JASPAR and TFSEARCH – yielded an extensive list of transcription factors that were either known or predicted to bind rs10492367 (Chapter 2.21 – Chapter 2.22). Predictions were based solely on the probe sequences, whereas known binding was based on experimental evidence from the Transcription Factor ChIP-Seq from ENCODE track available on the UCSC Genome Browser. In order to prioritise those that were most likely to bind *in vivo*, it was necessary to implement the selection criteria detailed in Figure 2.1. As a result, nine transcription factors were selected for competition EMSAs to investigate the identity of the protein complexes binding to rs10492367 (Table 5.5). For each *trans*-acting factor, a competitor sequence was designed that included the consensus sequence at the position predicted to bind, flanked by a random sequence. The majority of the predicted binding sites were 5' to the SNP or in the central region, which is contradictory to the region predicted to be necessary for protein binding to the G allele probe (Figure 5.5).

Factor	Database	Competitor sequence (5'-3')	Key
C/EBPβ	PROMO	CCACCGTAGGCAGGCAAATCGCACGCCCCAT	1
MEF2A	UCSC; JASPAR	CCATATTTTTGGCTACGATCGCACGCCCCAT	1
MEF2C	UCSC; JASPAR	CCCTAAAAATAGGTACGATCGCACGCCCCAT	2
NFIC	UCSC; PROMO	CCACCGTAGGCAGCCAAACGCCACGCCCCAT	1
NFκB	UCSC	CCACCGGGAAAGTCCCGATCGCACGCCCCAT	3
RELA	UCSC	CCACCGGAGTTTCCCCTATCGCACGCCCCAT	1
RXRα	PROMO	CCACCGTAGTGAACCCGATCGCACGCCCCAT	1
TCF3	UCSC	CCACCGTAGGCAGTACAGCTGCACGCCCCAT	4
TCF12	UCSC	CCACCGTAGGCAGTACAGCTGCACGCCCCAT	4

Table 5.5. Nine transcription factors selected for competition EMSAs to investigate the identity of the protein complexes binding to rs10492367. After selecting the most promising transcription factors, competitor sequences were designed to include the protein consensus sequence at the site predicted to bind, flanked by a random sequence. Transcription factor consensus sequences are underlined in red, the random sequences are in black text. Origin of consensus sequences are numbered in the key: 1 (PROMO), 2 (Milligan and Jolly, 2012), 3 (Pierce *et al.*, 1988) and 4 (GeneCards).

Initially, the basic reaction mixture was incubated with competitors at 10 x and 25 x the concentration of the fluorescently-labelled probe (Appendix F: Figure F.2) in order to select those necessitating a more detailed competition EMSA (Figure 5.6 and Figure 5.7). From both the intermediate and the full competition EMSAs, the most compelling results were: i) the

TCF3/12 consensus sequence outcompeting complexes 2 and 4 with both allele probes with SW1353 nuclear protein, and ii) complex 4 binding to both alleles being outcompeted by the MEF2A consensus sequence. The remaining consensus sequences resulted in subtle, less defined competitions, including RxRα which appeared to have no effect on protein binding.

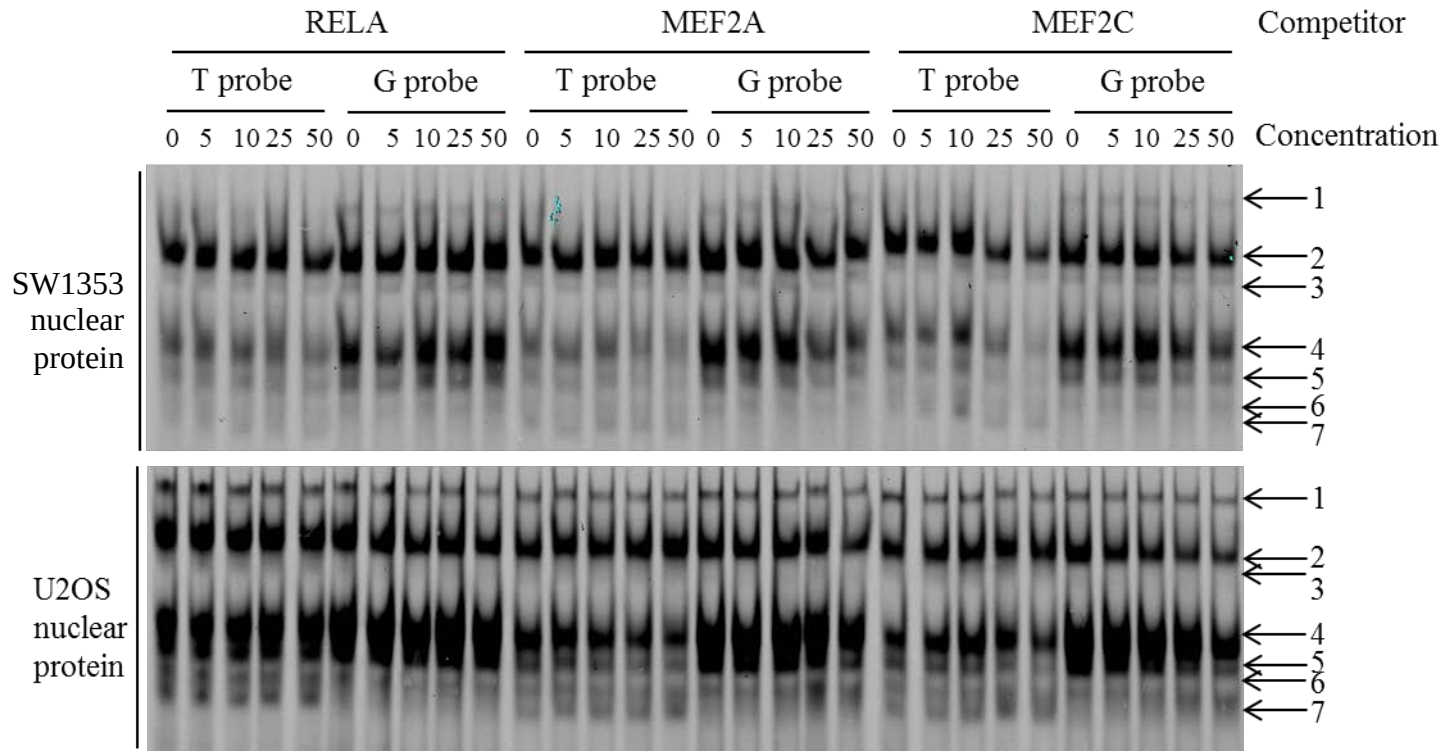

Figure 5.6. Competition EMSAs to investigate the consensus sequences necessary for SW1353 and U2OS nuclear extract binding to the G and T alleles of rs10492367. The protein:probe mixes were incubated with a full concentration range of unlabelled competitors that contained the consensus sequences for selected transcription factors predicted to bind the G or T allele probes. The results confirmed the preliminary indications that competition with the MEF2A and MEF2C consensus sequences caused changes in band intensity. Competition of the T allele probe with the RELA consensus sequence appeared to increase the band intensity of complex 4 with the U2OS nuclear extract. Concentration = 0, 5, 10, 25, 50 x probe concentration.

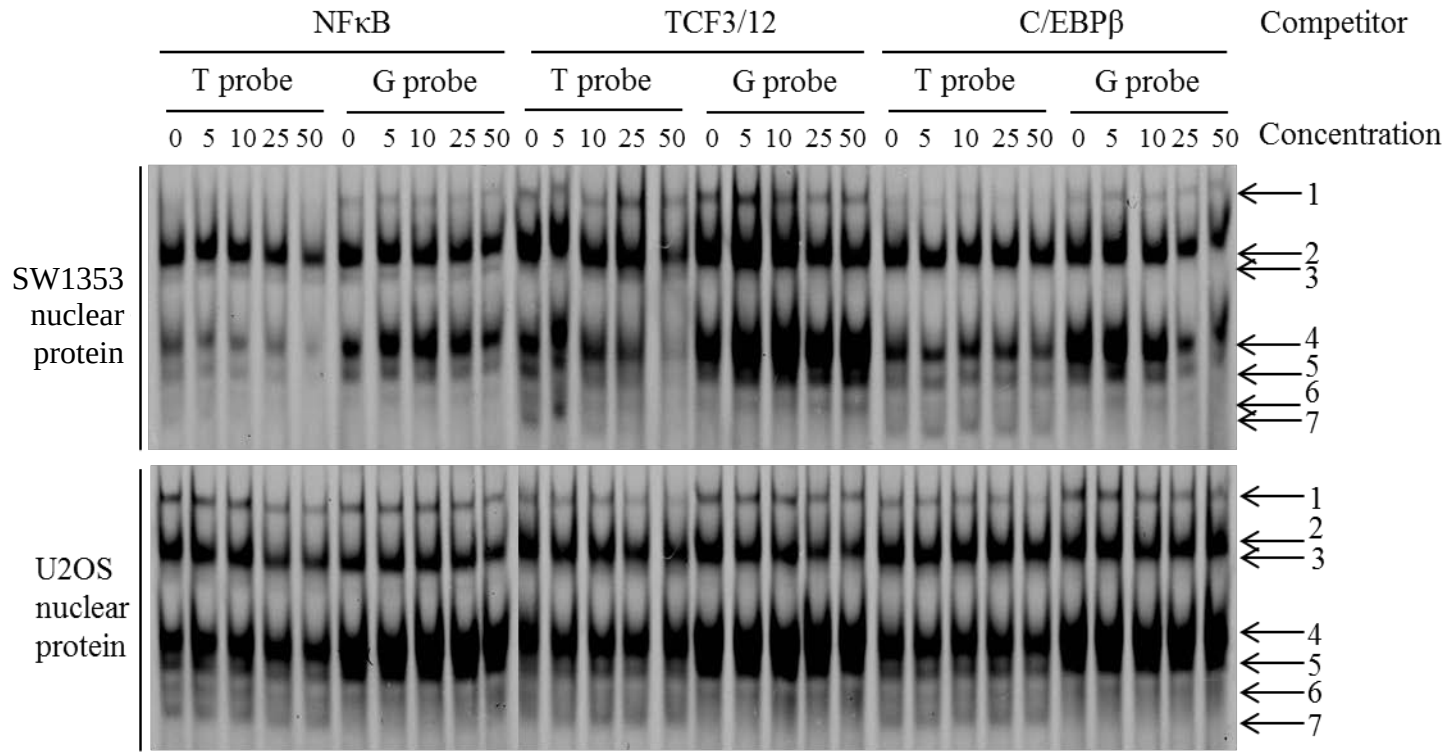

Figure 5.7. Competition EMSAs to investigate the consensus sequences necessary for SW1353 and U2OS nuclear extract binding to the G and T alleles of rs10492367. The protein:probe mixes were incubated with a full concentration range of unlabelled competitors that contained the consensus sequences for selected transcription factors predicted to bind the G or T allele probes. The results confirmed the preliminary indications that competition with TCF3/12 and C/EBPβ consensus sequences caused changes in band intensity. Competition with NFκB caused disruption of the protein binding to the T probe in complexes 2 and 4. Concentration = 0, 5, 10, 25, 50 x probe concentration.

Based on the competition EMSAs, the binding of RELA, MEF2A, MEF2C, TCF3, TCF12 and C/EBPβ were further investigated by means of supershift EMSAs. This involved the addition of antibodies raised against the specific transcription factors, with the premise being that the antibody would bind the *trans*-acting factor:DNA complex, causing the complex to migrate more slowly through the gel which would be observed as a 'supershift'. The binding of each transcription factor was investigated in this manner, in addition to a panel of *trans*-acting factors that are known to be expressed in cartilage (Appendix E: Table E.1). To test the concentration of antibody required for the supershift EMSAs, the basic reaction mixes were incubated with either 2 or 6 µg of antibody.

A supershift was observed for TCF3, marked by an asterisk in Figure 5.8, for both alleles with both nuclear extracts. The supershift can be observed more clearly with 6 µg of antibody in the SW1353 nuclear protein. In addition, complex 4 binding to the T allele probe with the SW1353 nuclear extract became smeared upon the addition of the RELA antibody. There were no other supershifts observed for the remaining predicted transcription factors (Figure 5.8 and Figure 5.9). Of the panel of *trans*-acting factors known to be expressed in cartilage that were tested, SUB1 (also known as PC4) caused complex 1 to disappear (Appendix F: Figure F.3 – Figure F.7). Other transcription factors that appeared to cause supershifts were AR, Deaf1 and Ets1/2, although these could not be replicated. For TCF3, RELA and SUB1, the supershifts were repeated and replicated with both SW1353 and U2OS nuclear protein, before confirmation using HAC nuclear extract (Figure 5.10). The supershift for RELA was rather ambiguous, and subsequent repeats of this experiment produced equally unclear results. It is unclear which band supershifted to produce the banding pattern observed for TCF3; however competition EMSAs did suggest the protein in complex 2 was TCF3. For RELA, competition EMSAs caused complex 4 to become more intense, and it was this complex that was affected by incubation with the RELA antibody. SUB1, RELA and TCF3 were selected as candidates for further functional analysis (Chapter 5.3.10 – 5.3.13). The competition and supershift EMSA results of transcription factors binding to rs10492367 are summarised in Table 5.6.

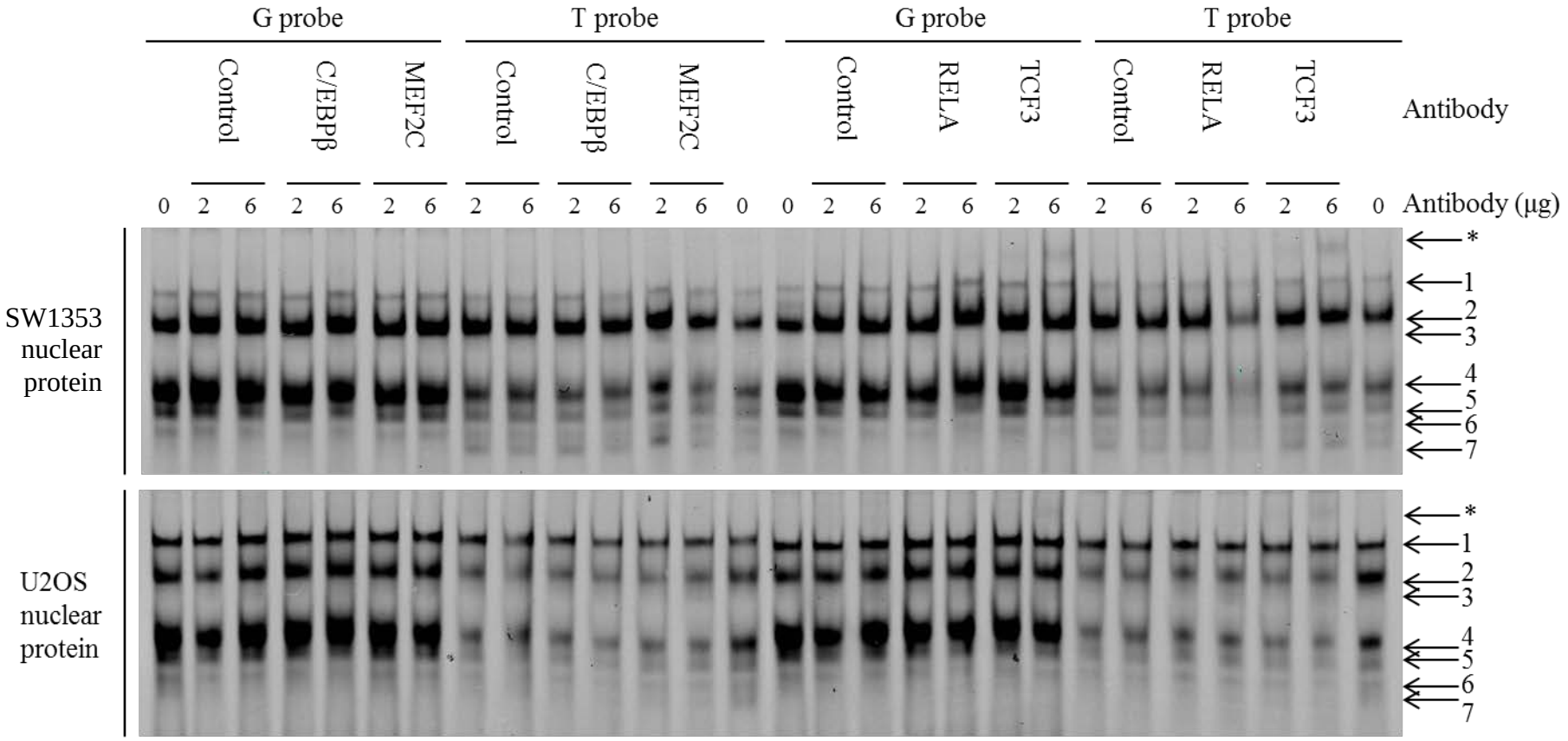

Figure 5.8. Supershift EMSAs to investigate the transcription factors of SW1353 and U2OS nuclear extracts binding to the G and T alleles of rs10492367. The protein:probe mixes were incubated with either 2 µg or 6 µg of antibody. The results confirmed the preliminary indications that RELA and TCF3 interacted with the fluorescently labelled DNA. Bands 4 – 7 of the G allele probe were fainter when incubated with 6 µg RELA antibody relative to the species-matched IgG control antibody. A supershift, marked by an asterisk (*), was observed for both alleles after incubation with the TCF3 antibody. No changes in the banding patterns were observed after incubation with C/EBPβ or MEF2C antibodies. Control (IgG species-matched antibody).

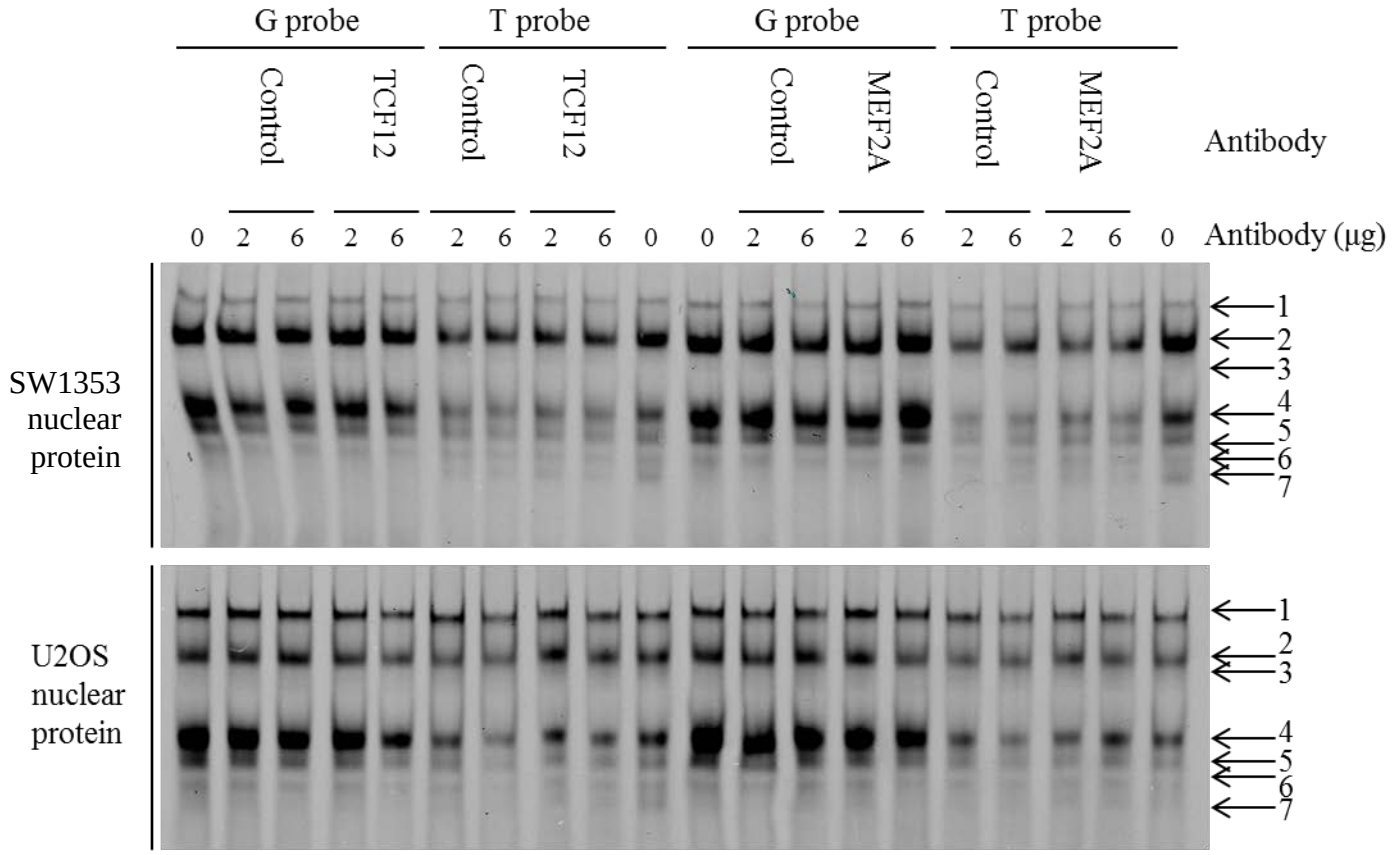

Figure 5.9. Supershift EMSAs to investigate the transcription factors of SW1353 and U2OS nuclear extracts binding to the G and T alleles of rs10492367. The protein:probe mixes were incubated with either 2 µg or 6 µg of antibody. No changes in the banding patterns were observed after incubation with TCF12 or MEF2A antibodies relative to the control antibody. Control (IgG species-matched antibody).

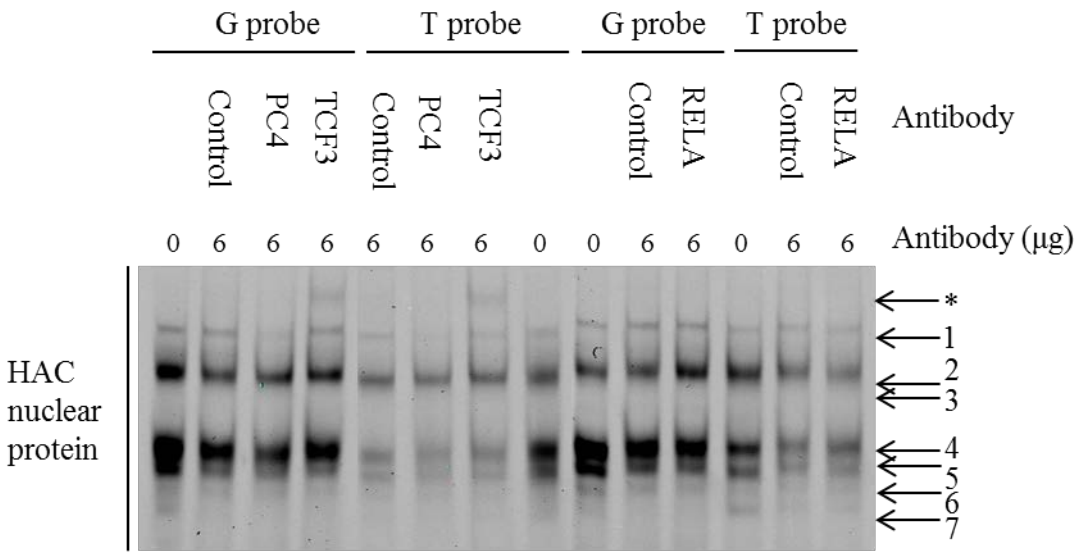

Figure 5.10. Supershift EMSAs to investigate the transcription factors of HAC nuclear extract binding to the G and T alleles of rs10492367. The protein:probe mixes were incubated with 6 µg of antibody. The results confirmed the SW1353 and U2OS indications that PC4 and TCF3 interacted with the fluorescently labelled DNA. Band 1 becomes fainter for both alleles after incubation with the PC4 antibody relative to the species-matched IgG control. A supershift, marked by an asterisk (*), was observed for both alleles after incubation with the TCF3 antibody. The results were less clear for RELA, where previously bands 4 – 7 were seen to become fainter relative to the IgG control antibody. Control (IgG species-matched antibody). Lanes irrelevant to this analysis have been removed.

	Competition EMSA				Supershift		
Transcription factor	**Nuclear protein**	**Allele**	**Effect**	**Complex**	**Nuclear protein**	**Allele**	**Complex**
C/EBPβ	SW1353 and U2OS	G and T	Outcompeted	4	No effect	No effect	No effect
MEF2A	SW1353 and U2OS	G and T	Outcompeted	4	No effect	No effect	No effect
MEF2C	SW1353 and U2OS	G and T	Outcompeted	4	No effect	No effect	No effect
NFIC	No effect	No effect	No effect	No effect	No effect	No effect	No effect
NFκB	SW1353 and U2OS	T only	Outcompeted	2 and 4	No effect	No effect	No effect
RELA	U2OS	T only	Stronger binding	4	SW1353 and HAC	T only	4
RXRα	No effect	No effect	No effect	No effect	No effect	No effect	No effect
TCF3	SW1353	G and T	Outcompeted	2 and 4	SW1353, U2OS and HAC	G and T	2
TCF12	SW1353	G and T	Outcompeted	2 and 4	No effect	No effect	No effect
SUB1 (PC4)	No data	No data	No data	No data	SW1353, U2OS and HAC	G and T	1

Table 5.6. Summary of competition and supershift EMSAs to investigate transcription factor binding to rs10492367. All transcription factors, apart from SUB1 (PC4), were predicted to bind the SNP. RELA and TCF3 were positively identified as binding rs10492367 through supershift EMSAs, in addition to SUB1 (PC4). RELA, TCF3 and SUB1 (PC4) were selected for further functional analysis. The nuclear protein extracts for which effects were observed are listed in the 'Nuclear protein' column.

5.3.7 *Investigating transcription factor binding to rs58649696 using chondrosarcoma and osteosarcoma cell line nuclear protein*

The luciferase reporter assays of Chapter 5.3.3 have demonstrated that the risk (T) allele of rs58649696 had a decreased enhancer activity relative to the non-risk (C) allele. EMSA conditions were first optimised using NP-40 and glycerol as optional additional components (Appendix F: Figure F.1). Protein binding to the C and T allele probes produced six distinct complexes, highlighted by the numbered arrows, with little difference between the two conditions. Therefore, NP-40 was selected as the only additional component in the EMSA reaction mixes for rs58649696.

Following optimisation of the binding conditions, the specificity and affinity of the protein complexes to the C and T allele probes were tested using unlabelled competitors (Figure 5.11). The resulting banding patterns were markedly different both between the alleles and the nuclear extract used. For example, complex 1 was present with both allele probes, but only with the U2OS cell line nuclear extract; and complex 3 was T allele-specific. Protein binding to the C allele probe in complex 2 appeared stronger than to the T allele probe, and was disrupted more readily by the C allele competitor. The protein binding to complexes 4 and 6 were stronger to the T allele probe than the C allele probe. In addition, the T allele competitor strongly outcompeted these complexes, suggesting that the proteins have an increased binding specificity for the T allele rather than the C allele. An additional protein complex, marked by an asterisk in Figure 5.11, appeared upon increasing C allele competitor concentration for both alleles, which could be a result of the probe from a different complex being freed to bind another protein and make a new complex.

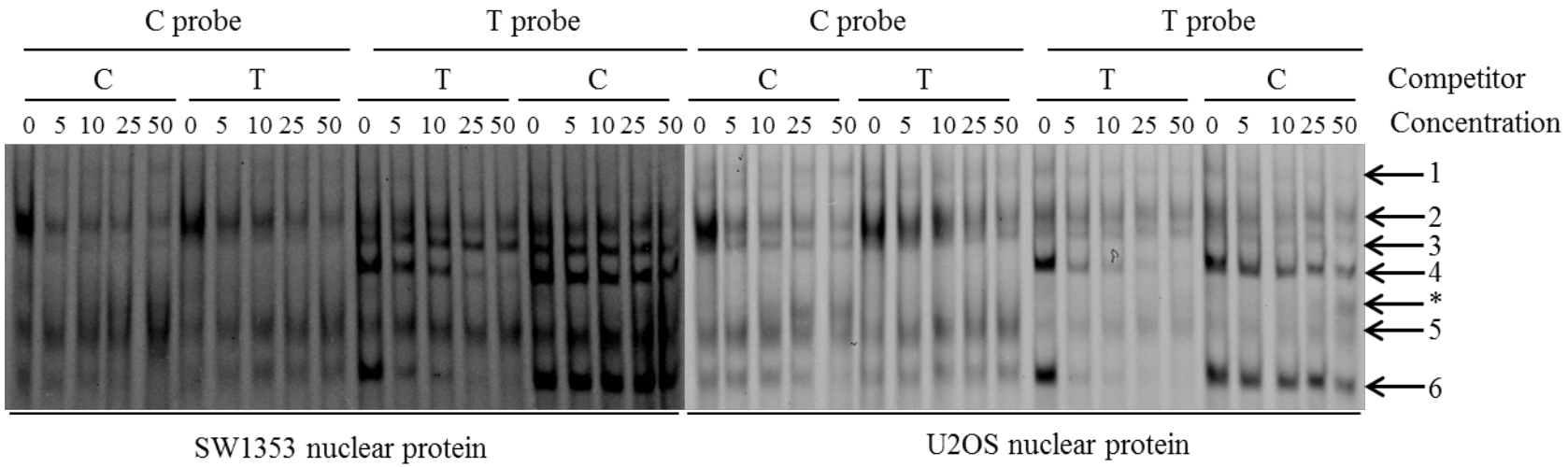

Figure 5.11. Competition EMSAs to investigate allele-specific binding of SW1353 and U2OS nuclear extract to the C and T alleles of rs58649696. The concentrations of unlabelled competitors were increased from 0 to 50 times that of the labelled probes. The protein:probe complexes marked by arrows 1, 3 and 5 were not noticeably disrupted upon increasing competitor concentrations. Protein binding to the C allele probe in complex 2 was disrupted by the C allele competitors and to less of a degree with the T allele competitor. Complexes 4 and 6 binding to the T allele probe were outcompeted by the T allele competitor only. A protein complex, marked by an asterisk (*), appeared upon increasing C allele competitor concentration for both alleles. Concentration = 0, 5, 10, 25, 50 x probe concentration.

The region of the labelled probe that was involved in binding the protein was next investigated using competitors containing random sequences adjacent to probe sequences (Table 5.7). The random sequence had no effect on protein binding to either allele probe, indicating that the protein:DNA complexes are specific for the probe sequences (Figure 5.12). Complex 5 binding to the C allele probe was outcompeted by competitor 2, which contained a central region that was homologous to the probe sequence, while complex 6 binding to the T allele probe was outcompeted by competitor 1. This suggests that the middle section and the 5'-most sequence of the probes are required for protein binding to the C and T allele probes, respectively. Lack of competition for the other complexes suggests that the entire probe sequence is important for binding.

Sequence name	Probe sequence (5'-3')
Random primer full length	CCACCGTAGGCAGTACGATCGCACGCCCCAT
C allele probe and competitor	CATCTGCCTCTTTCACTTTCCTAATGGGACT
T allele probe and competitor	CATCTGCCTCTTTCATTTTCCTAATGGGACT
Competitor 1: C allele	CATCTGCCTCTTTCACTATCGCACGCCCCAT
Competitor 1: T allele	CATCTGCCTCTTTCATTATCGCACGCCCCAT
Competitor 2: C allele	CCACCGTAGCTTTCACTTTCCTACGCCCCAT
Competitor 2: T allele	CCACCGTAGCTTTCATTTTCCTACGCCCCAT
Competitor 3: C allele	CCACCGTAGGCAGTACTTTCCTAATGGGACT
Competitor 3: T allele	CCACCGTAGGCAGTATTTTCCTAATGGGACT

Table 5.7. Primer sequences used for competition EMSAs to investigate the regions of the C and T allele probes of rs58649696 to which the SW1353 and U2OS nuclear extracts bind. Each primer was annealed to its reverse complement, creating dsDNA, prior to use in EMSAs. The original competitor sequences are underlined in red; the random sequences are in black text.

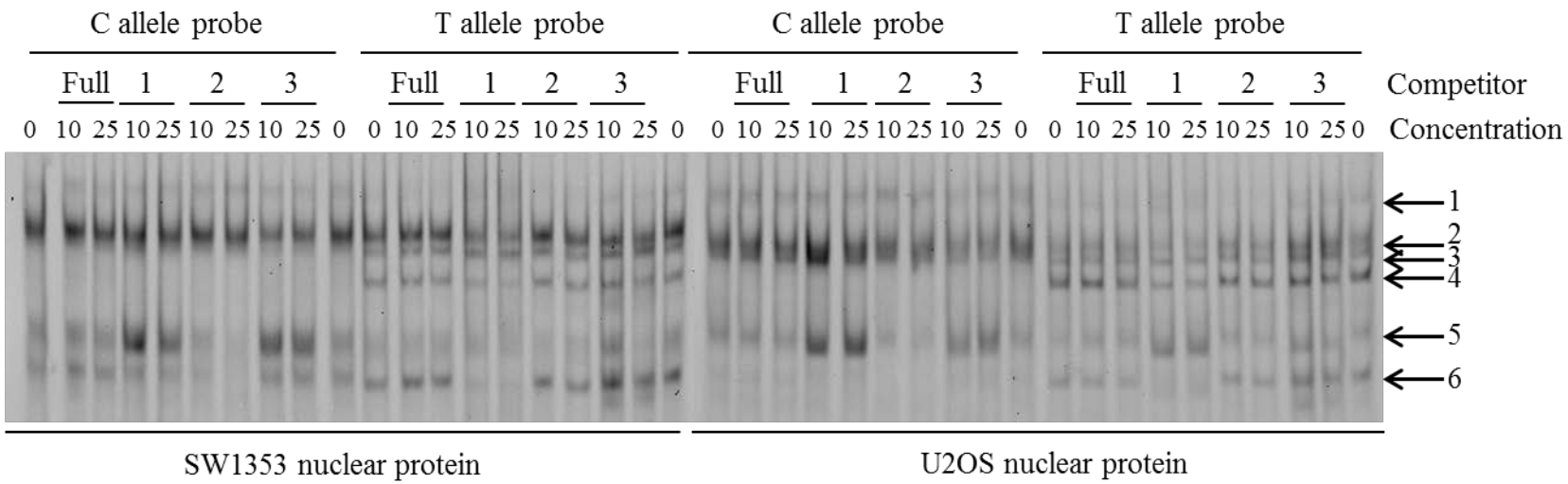

Figure 5.12. Competition EMSAs to investigate the regions of the C and T allele probes of rs58649696 to which the SW1353 and U2OS nuclear extracts bind. The protein:probe mixes were incubated with unlabelled competitors that had random sequences replacing the original competitor sequence. An entirely random competitor (full) had no effect on any of the protein:probe complexes. The aberrant strength of complex 2 with the C allele probe was likely due to a technical error. The C allele probe binding to the protein in complexes 5 and 6 were outcompeted by competitor 2. Competitors 1 and 3 caused the band intensity to increase for complex 5. Concentration = 0, 10, 25 x probe concentration. Competitor = 31 bp random competitor (full); 3′-most region of the competitor replaced by a 14 bp random sequence (1); the central 13 bp region flanked either side by a 9 bp random sequence (2); 5′-most region of the competitor replaced by a 14 bp random sequence (3).

Database searches identified five transcription factors that had predicted, or evidence to suggest, binding over rs58649696. These proteins were selected for competition EMSAs, using probes that had binding sites in the central region of the probe (Table 5.8), as was shown to be important for protein binding to the C allele in Figure 5.12.

Factor	Database	Competitor sequence (5'-3')	Key
IRF1	UCSC; TFSEARCH; JASPAR	CCACCGTAGGTTTCCCTTTCGCACGCCCCAT	1
IRF2	PROMO; TFSEARCH; JASPAR	CCACCGTAGGCATCACTTTCGCACGCCCCAT	1
POU2F2	JASPAR	CCACCGTAGGGCGGATTTGCATATTCCCCAT	1
TCF4	JASPAR	CCACCGTACCTTTGATGATCGCACGCCCCAT	2
XBP1	PROMO	CCACCGTAGGCGTCATGATCGCACGCCCCAT	1

Table 5.8. Five transcription factors selected for competition EMSAs to investigate the identity of the protein complexes binding to rs58649696. After selecting the most promising transcription factors, competitor sequences were designed to include the protein consensus sequence at the site predicted to bind, flanked by a random sequence. Transcription factor consensus sequences are underlined in red, the random sequences are in black text. Origin of consensus sequences are numbered in the key: 1 (PROMO) and 2 (Hatzis *et al.*, 2008).

Initially, the basic reaction mixture was incubated with competitors at 10 x and 25 x the concentration of the fluorescently-labelled probe (Appendix F: Figure F.8) in order to select those necessitating a more detailed competition EMSA (Figure 5.13). From both the intermediate and the full competition EMSAs, the most compelling results were for IRF2 and XBP1, where complexes 2, 4 and 6 binding to the T allele probe were outcompeted by the unlabelled competitors. These complexes were shown to be allele-specific in Figure 5.11. Competition of the C allele probe with the TCF4 consensus sequence appeared to slightly decrease the band intensity of complex 6 with the SW1353 nuclear extract. There was no observed competition with the consensus sequence for IRF1.

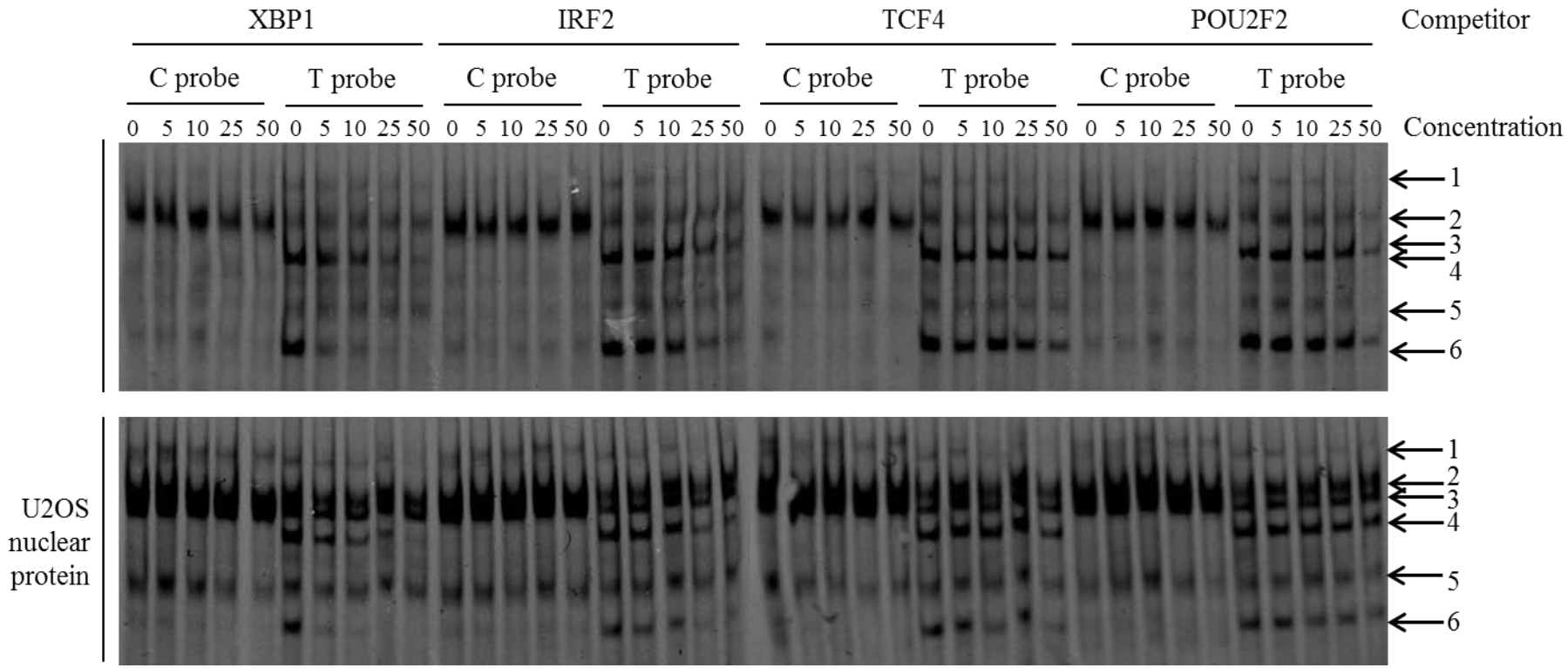

Figure 5.13. Competition EMSAs to investigate the consensus sequences necessary for SW1353 and U2OS nuclear extract binding to the C and T alleles of rs58649696. The protein:probe mixes were incubated with a full concentration range of unlabelled competitors that contained the consensus sequences for selected transcription factors predicted to bind the C or T allele probes. The results confirmed the preliminary indications that competition with the XBP1, IRF2, TCF4 and POU2F2 consensus sequences caused changes in band intensity. Concentration = 0, 5, 10, 25, 50 x probe concentration.

Based on the competition EMSAs, the binding of IRF2, POU2F2, TCF4 and XBP1 were further investigated by means of supershift EMSAs. The binding of each transcription factor was investigated in this manner, in addition to a panel of *trans*-acting factors that are known to be expressed in cartilage (Appendix E: Table E.1). The basic reaction mixes for the transcription factors predicted to bind were incubated with either 2 or 6 µg of antibody.

There were no supershifts observed with either cell line nuclear extract for any of the transcription factors that were predicted to bind the polymorphism (Figure 5.14). However, incubation of the T allele probe with antibodies for PAX9 and SUB1 (PC4), neither of which were predicted to bind, caused a band to appear with both cell line nuclear extracts (Appendix F: Figure F.9 – Figure F.11). The supershift for SUB1 (PC4) was repeated and replicated with both SW1353 and U2OS nuclear extracts before confirmation using HAC nuclear extract, however that of PAX9 was not replicated (Figure 5.15). It is unclear which band supershifted to produce the banding pattern observed for SUB1 (PC4), and moreover, it cannot be deduced from the competition EMSAs as it was not predicted to bind the SNP. Nevertheless, this is a positive identification of protein binding to rs58649696. The competition and supershift EMSA results of transcription factors binding to rs58649696 are summarised in Table 5.9.

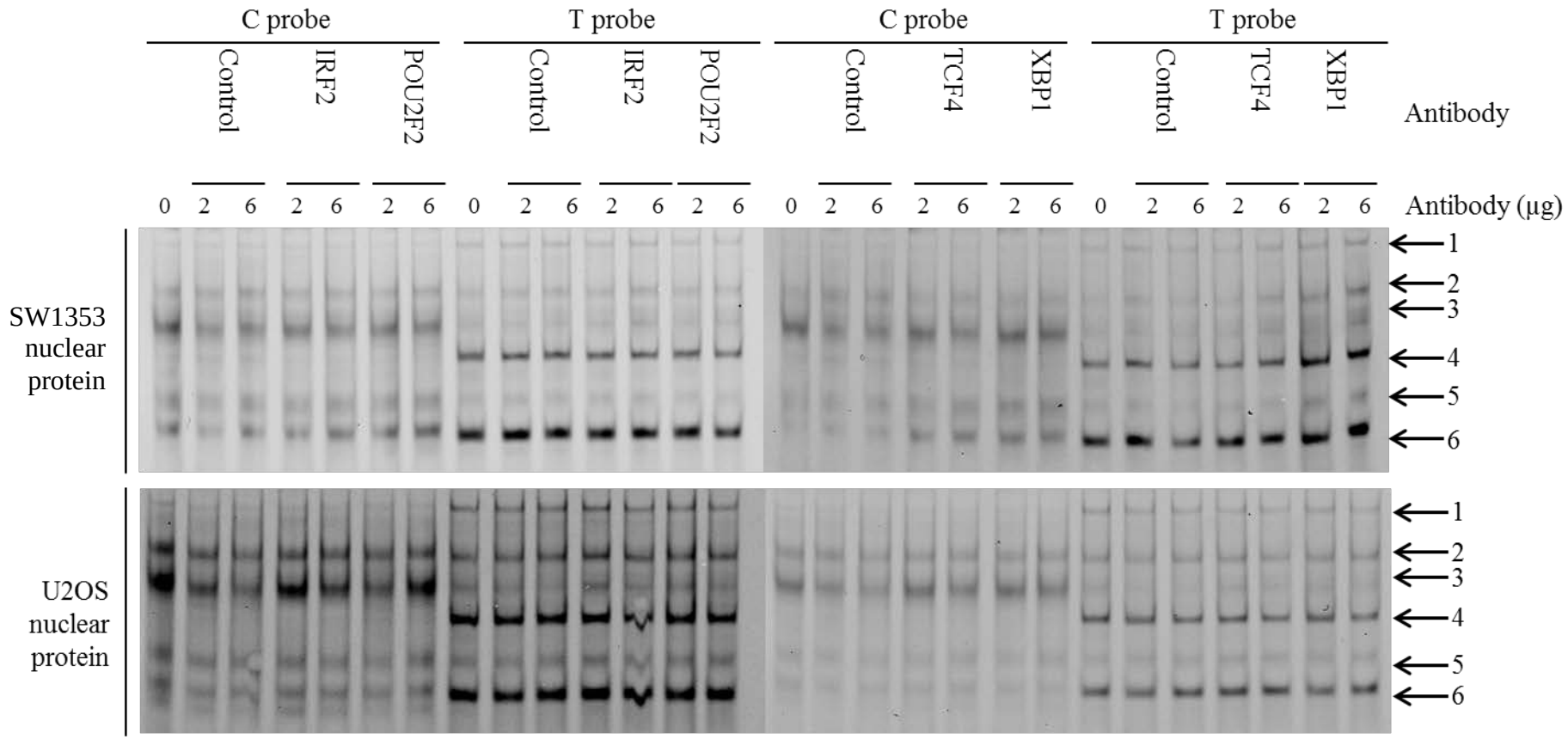

Figure 5.14. Supershift EMSAs to investigate the transcription factors of SW1353 and U2OS nuclear extracts binding to the C and T alleles of rs58649696. The protein:probe mixes were incubated with either 2 µg or 6 µg of antibody. No changes in the banding patterns were observed after incubation with any of the antibodies. Control (IgG species-matched antibody).

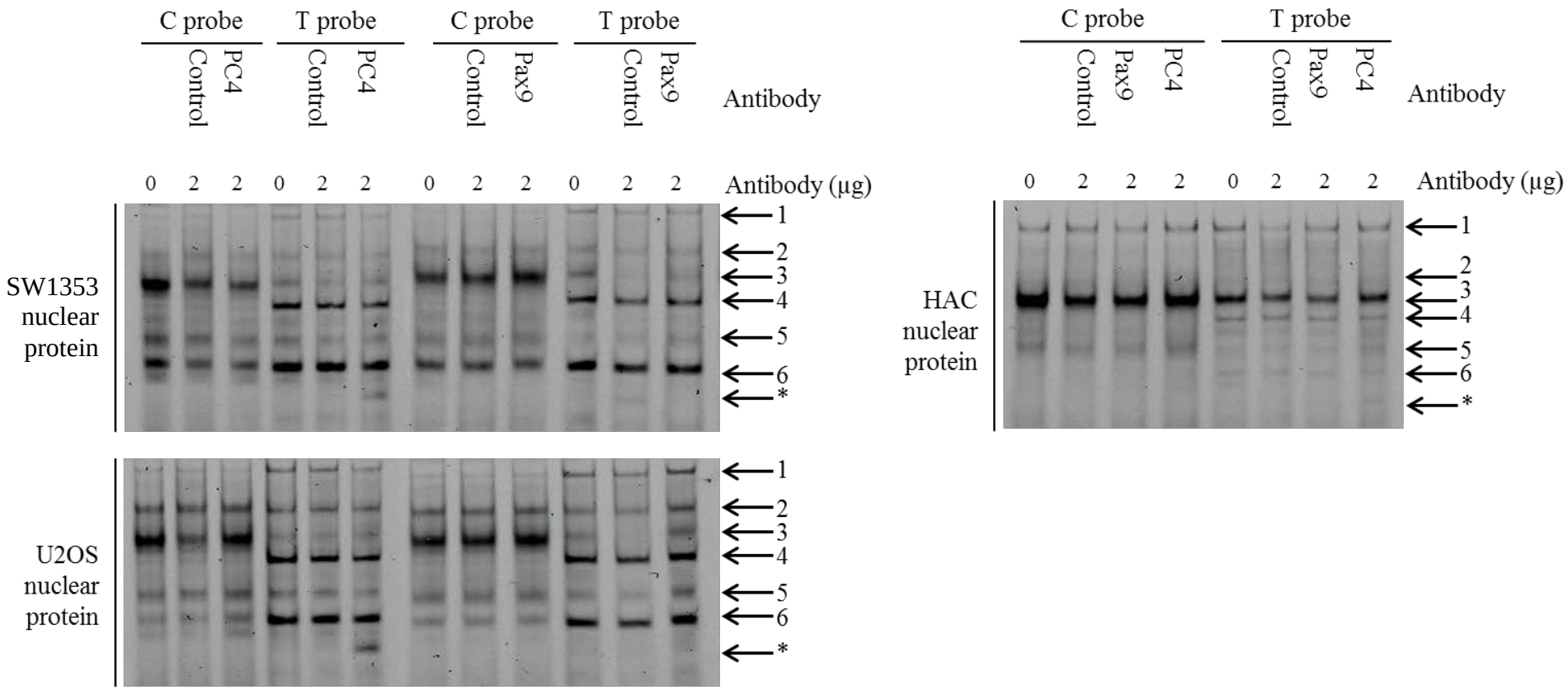

Figure 5.15. Supershift EMSAs to confirm the transcription factors of SW1353, U2OS and HAC nuclear extract binding to the C and T alleles of rs58649696. The protein:probe mixes were incubated with 2 µg of antibody. The results confirmed the SW1353 and U2OS indications that PC4 interacted with the fluorescently labelled T allele probe in all cell line nuclear extracts, causing a band to appear as marked by an asterisk (*). A repeat of the PAX9 supershifts did not result in any observable supershifts in with SW1353, U2OS or HAC nuclear protein. Control (IgG species-matched antibody). Lanes irrelevant to this analysis have been removed.

	Competition EMSA				Supershift		
Transcription factor	**Nuclear protein**	**Allele**	**Effect**	**Complex**	**Nuclear protein**	**Allele**	**Complex**
IRF1	No effect	No effect	No effect	No effect	No effect	No effect	No effect
IRF2	SW1353 and U2OS	T only	Outcompeted	2, 4 and 6	No effect	No effect	No effect
POU2F2	SW1353	T only	Outcompeted	2, 4 and 6	No effect	No effect	No effect
TCF4	SW1353	C only	Outcompeted	6	No effect	No effect	No effect
XBP1	SW1353 and U2OS	T only	Outcompeted	2, 4 and 6	No effect	No effect	No effect
PAX9	No data	No data	No data	No data	U2OS	T only	Below 5
SUB1 (PC4)	No data	No data	No data	No data	SW1353, U2OS and HAC	T only	Below 6

Table 5.9. Summary of competition and supershift EMSAs to investigate transcription factor binding to rs58649696. All transcription factors, apart from PAX9 and SUB1 (PC4), were predicted to bind the SNP. PAX9 and SUB1 (PC4) were positively identified as binding rs58649696 through supershift EMSAs. The nuclear protein extracts for which effects were observed are listed in the 'Nuclear protein' column.

5.3.8 *Investigating transcription factor binding to rs11049206 using chondrosarcoma and osteosarcoma cell line nuclear protein*

In luciferase assays performed in the SW1353 cell line, the risk (C) allele of rs11049206 caused an increase in enhancer activity relative to the non-risk (G) allele, while the converse was observed for the U2OS cell culture. Optimisation of the EMSA binding conditions for this SNP using NP-40 with or without glycerol produced seven distinct complexes, highlighted by the numbered arrows, with little difference between the two conditions (Appendix F: Figure F.1). Therefore, NP-40 was selected as the only additional component in the EMSA reaction mixes for rs11049206. Due to technical difficulties in extracting nuclear protein, extracts at a higher passage were used for this optimisation. As used for the experiments in Chapter 5.3.6, Chapter 5.3.7 and Chapter 5.3.9, extracts from a lower passage were used for the remaining EMSAs.

Following optimisation of the binding conditions, the specificity and affinity of the protein complexes to the C and G allele probes were tested (Figure 5.16). The protein of complex 1 binding to both allele probes was outcompeted by both unlabelled competitors, although this band is comparatively fainter with the SW1353 extract. In addition, the binding of the protein in complex 4 was outcompeted by both the C and G allele competitors. The protein of complex 5 binding to the G allele probe was outcompeted primarily by the G allele competitor, with some competition at the higher concentrations of the C allele competitor with the U2OS nuclear extract. Since returning to use nuclear extracts at a lower passage, an additional band appeared beneath complex 3. Although it could be forming a doublet with complex 3, this protein seemed to bind more strongly to the C allele with SW1353 nuclear extracts, with only faint shadowing in the presence of the G allele. Similarly, it was non-specific with the U2OS nuclear extract, however the strength of binding between the alleles was more comparable.

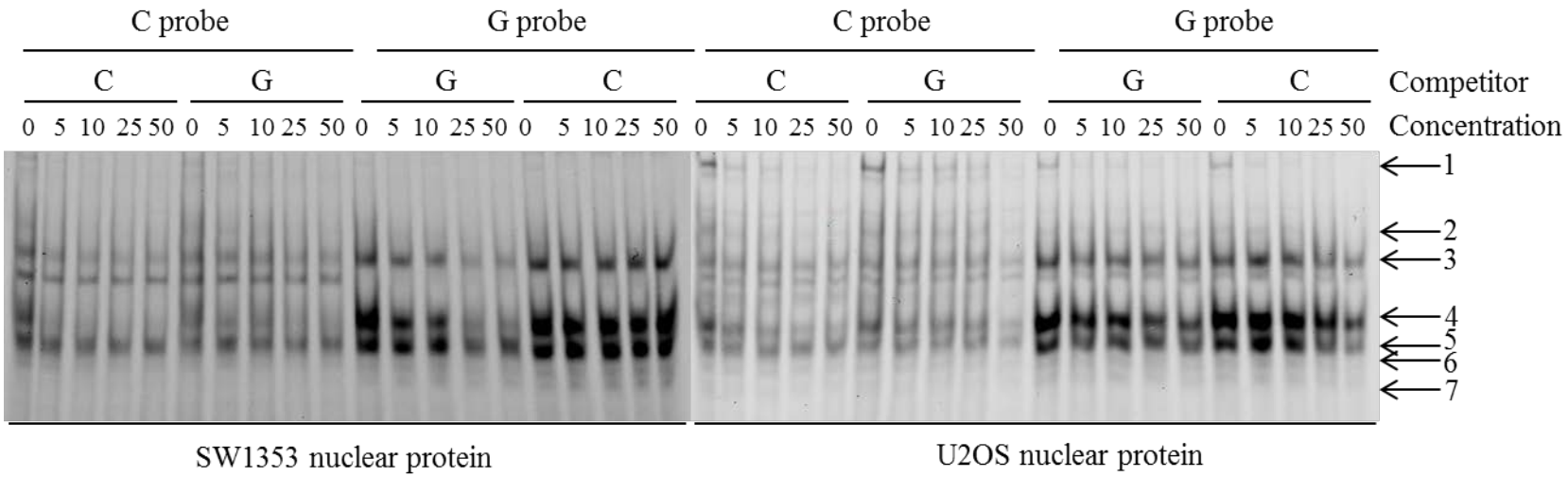

Figure 5.16. Competition EMSAs to investigate allele-specific binding of SW1353 and U2OS nuclear extract to the C and G alleles of rs11049206. The concentrations of unlabelled competitors were increased from 0 to 50 times that of the labelled probes. The protein:probe complexes marked by arrows 2, 6 and 7 were not noticeably disrupted upon increasing competitor concentrations. Protein binding to both allele probes in complex 1 was outcompeted by both allele competitors in the U2OS nuclear extract. Complex 4 was disrupted over the concentration gradient of the respective competitors for both allele probes, whereas complex 5 was outcompeted by the G allele competitor for the G allele probe only. A band appeared beneath complex 3, which seemed to bind more strongly to the C allele than the G allele with SW1353 nuclear extract. Concentration = 0, 5, 10, 25, 50 x probe concentration.

The region of the labelled probe that was involved in binding the protein was next investigated using competitors containing random sequences adjacent to probe sequences (Table 5.10). The random sequence had no effect on protein binding to either allele probe, indicating that the protein:DNA complexes are specific for the probe sequences (Figure 5.17). Complex 1 binding to the C allele probe was outcompeted by competitors 1 and 2, meaning that it is likely to be the 5′-most and central region of the DNA probe that is required for protein binding in this complex. Additionally, the intensity of the protein binding to complexes 5, 6 and 7 for both allele probes was increased upon the addition of competitor 3. This could be due to the generation of a new binding site with a greater sequence similarity to the protein consensus sequence than that of the original probe sequence.

Sequence name	Probe sequence (5'-3')
Random primer full length	CCACCGTAGGCAGTACGATCGCACGCCCCAT
C allele probe and competitor	ATGTTTATGTATAAACCAAACACCAGCAAAA
G allele probe and competitor	ATGTTTATGTATAAAGCAAACACCAGCAAAA
Competitor 1: C allele	ATGTTTATGTATAAACCATCGCACGCCCCAT
Competitor 1: G allele	ATGTTTATGTATAAAGCATCGCACGCCCCAT
Competitor 2: C allele	CCACCGTAGTATAAACCAAACAACGCCCCAT
Competitor 2: G allele	CCACCGTAGTATAAAGCAAACAACGCCCCAT
Competitor 3: C allele	CCACCGTAGGCAGTACCAAACACCAGCAAAA
Competitor 3: G allele	CCACCGTAGGCAGTAGCAAACACCAGCAAAA

Table 5.10. Primer sequences used for competition EMSAs to investigate the regions of the C and G allele probes of rs11049206 to which the SW1353 and U2OS nuclear extracts bind. Each primer was annealed to its reverse complement, creating dsDNA, prior to use in EMSAs. The original competitor sequences are underlined in red; the random sequences are in black text.

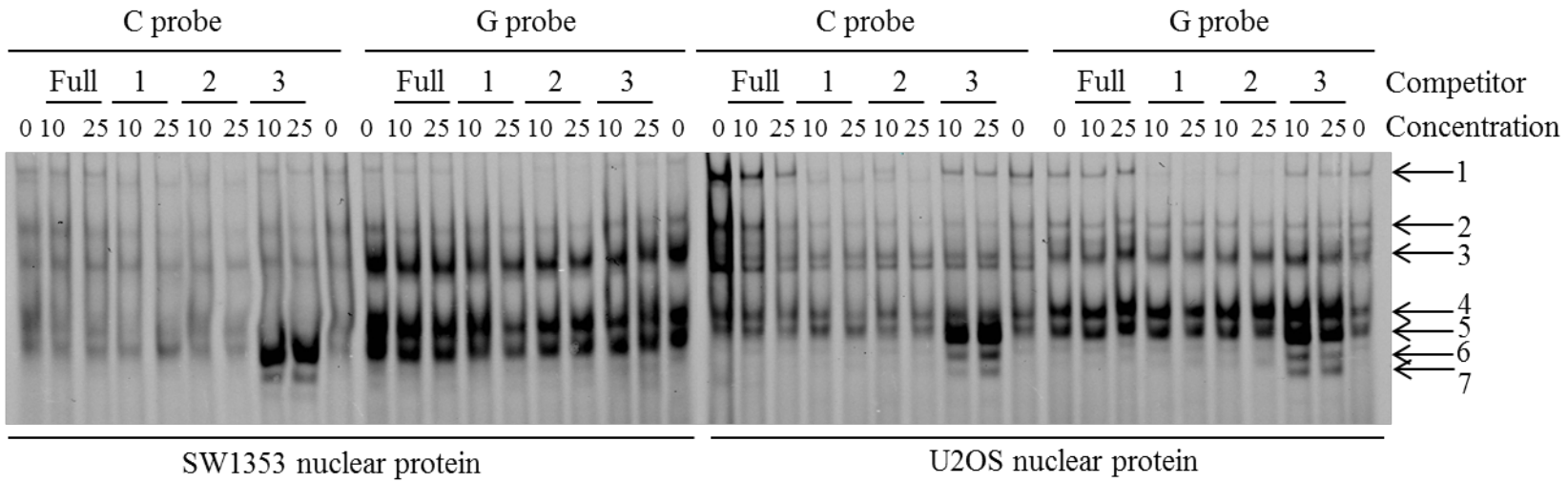

Figure 5.17. Competition EMSAs to investigate the regions of the C and G allele probes of rs11049206 to which the SW1353 and U2OS nuclear extracts bind. The protein:probe mixes were incubated with unlabelled competitors that had random sequences replacing the original competitor sequence. An entirely random competitor (full) had no effect on any of the protein:probe complexes. The C allele probe binding to the protein in complex 1 was outcompeted by competitors 1 and 2. Competitor 3 caused the band intensity to increase for complexes 5, 6 and 7. Concentration = 0, 10, 25 x probe concentration. Competitor = 31 bp random competitor (full); 3′-most region of the competitor replaced by a 14 bp random sequence (1); the central 13 bp region flanked either side by a 9 bp random sequence (2); 5′-most region of the competitor replaced by a 14 bp random sequence (3).

Database searches identified seven transcription factors that had predicted, or evidence to suggest, binding over rs11049206. These proteins were selected for competition EMSAs, using probes that included the consensus sequence at the position predicted to bind, flanked by a random sequence (Table 5.11).

Factor	Database	Competitor sequence (5'-3')	Key
FOXP1	JASPAR	CCACCGTAGGCAGTAATAAATAACGCCCCAT	1
FOXA2	TFSEARCH	CCACCGTAGGCAGAAGCAAACAATTCCCCAT	2
MEF2A	JASPAR	CTATTTTTGGCTGTACGATCGCACGCCCCAT	2
NFIC	PROMO	CCACCGTAGGCAGTACCAAACGCCGCCCCAT	2
NFYA	JASPAR	CCACCGTAGGCAGTACCAAACACCGCCCCAT	2
NFYB	JASPAR	CCACCGTAGGCAGTACCAATGCACGCCCCAT	2
RUNX2	JASPAR	CCACCGTAGGCAGTACCACAGCACGCCCCAT	3

Table 5.11. Six transcription factors selected for competition EMSAs to investigate the identity of the protein complexes binding to rs11049206. After selecting the most promising transcription factors, competitor sequences were designed to include the protein consensus sequence at the site predicted to bind, flanked by a random sequence. Transcription factor consensus sequences are underlined in red, the random sequences are in black text. Origin of consensus sequences are numbered in the key: 1 (Wang *et al.*, 2003); 2 (PROMO); 3 (Little *et al.*, 2012).

Initially, the basic reaction mixture was incubated with competitors at 10 x and 25 x the concentration of the fluorescently-labelled probe (Appendix F: Figure F.12) in order to select those necessitating a more detailed competition EMSA (Figure 5.18 and Figure 5.19). From both the intermediate and the full competition EMSAs, the most compelling results were for FOXP1, MEF2A and NFYA, while NFIC had a minimal effect and FOXA2 caused no observed competition. Briefly, complexes 1, 3, 4 and 6 were outcompeted with the FOXP1 consensus sequence, protein binding to the C allele probe of complexes 4 and 5 appeared to be decreased upon competition with MEF2A, complexes 4 and 6 were outcompeted by NFYA and complex 2 binding to both allele probes was outcompeted by RUNX2.

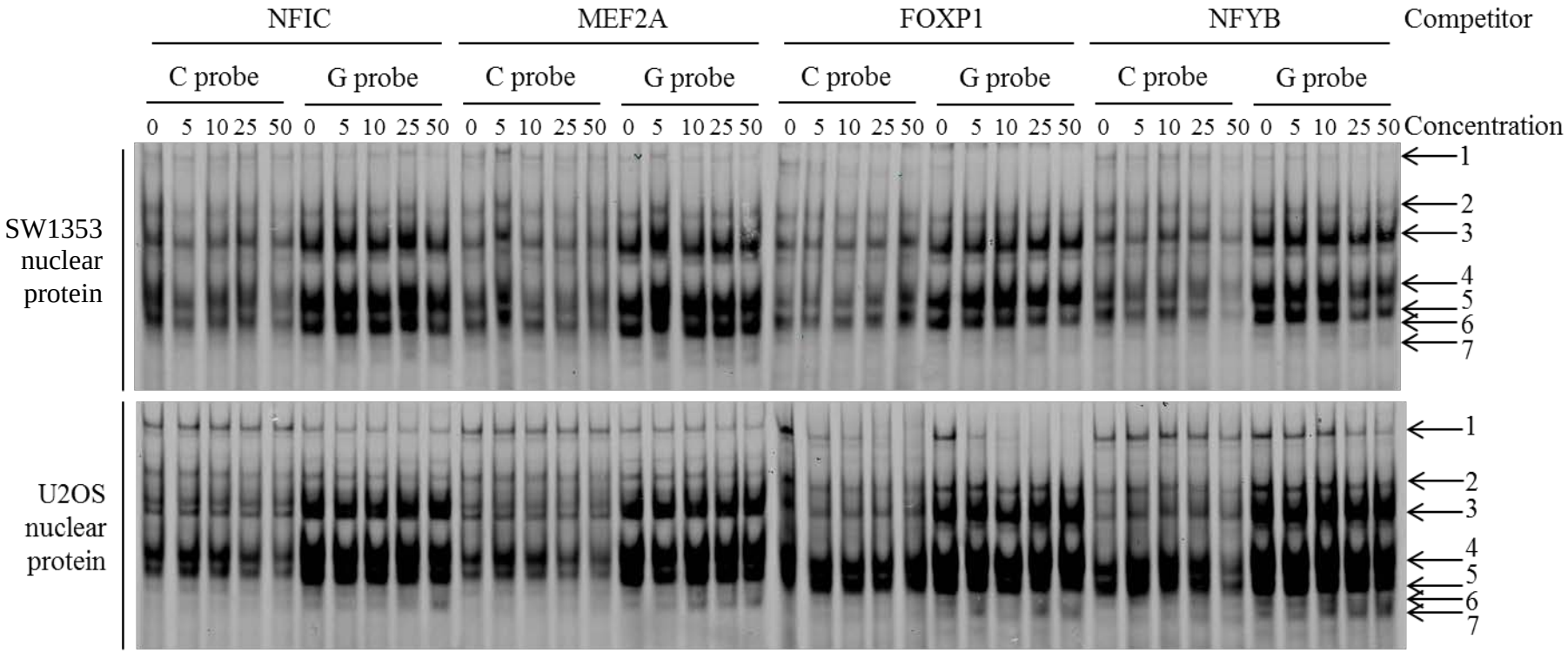

Figure 5.18. Competition EMSAs to investigate the consensus sequences necessary for SW1353 and U2OS nuclear extract binding to the C and G alleles of rs11049206. The protein:probe mixes were incubated with a full concentration range of unlabelled competitors that contained the consensus sequences for selected transcription factors predicted to bind the C or G allele probes. The results confirmed the preliminary indications that competition with NFIC, MEF2A, FOXP1 and NFYB caused changes in band intensity. Additionally, the competitions for NFIC, MEF2A and NFYB were replicated with the U2OS nuclear protein. Concentration = 0, 5, 10, 25, 50 x probe concentration.

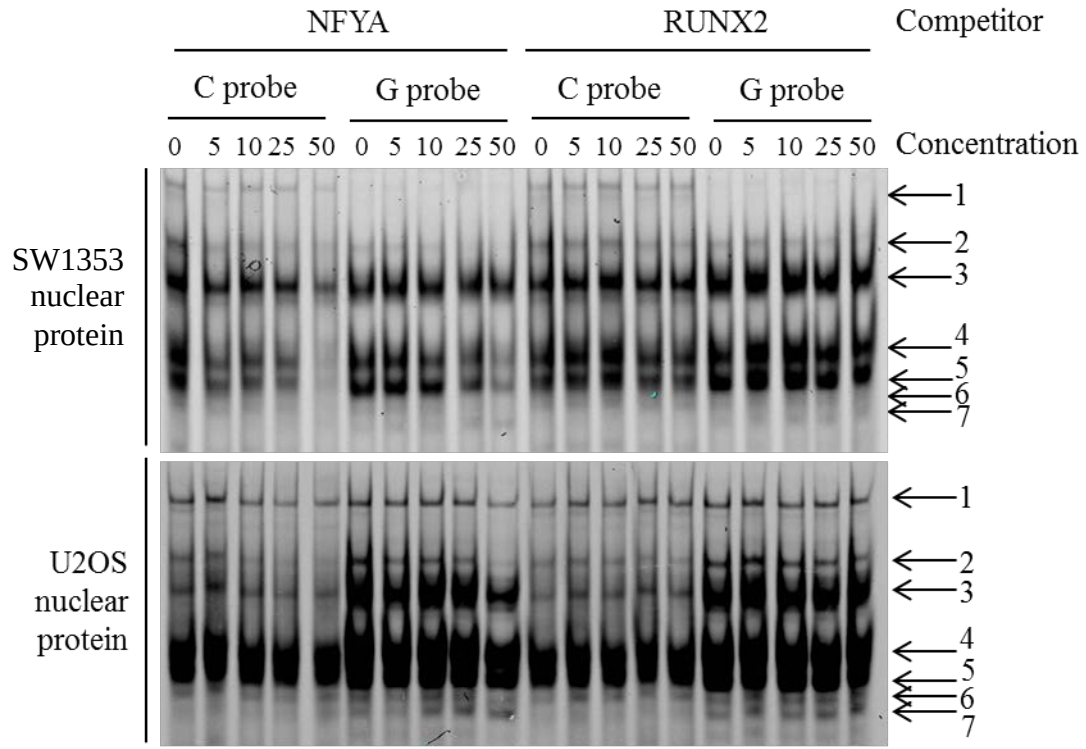

Figure 5.19. Competition EMSAs to investigate the consensus sequences necessary for SW1353 and U2OS nuclear extract binding to the C and G alleles of rs11049206. The protein:probe mixes were incubated with a full concentration range of unlabelled competitors that contained the consensus sequences for selected transcription factors predicted to bind the C or G allele probes. The results confirmed the preliminary indications that competition with NFYA and RUNX2 caused changes in band intensity. Additionally, competition with NFYA caused complexes 4 and 5 to be outcompeted in both allele probes and SW1353 nuclear extract. Concentration = 0, 5, 10, 25, 50 x probe concentration.

Based on the competition EMSAs, the binding of FOXP1, MEF2A and NFYA were further investigated by means of supershift EMSAs. The binding of each transcription factor was investigated in this manner, in addition to a panel of *trans*-acting factors that are known to be expressed in cartilage (Appendix E: Table E.1). The basic reaction mixes for the transcription factors predicted to bind were incubated with either 2 or 6 µg of antibody. There were no supershifts observed with either cell line nuclear extract for any of the transcription factors that were predicted to bind the polymorphism (Figure 5.20). However, incubation of the G allele probe with the antibody for RELA (NFκβ p65) caused bands 4 and 5 to disappear with the U2OS cell line nuclear extracts (Appendix F: Figure F.13 and Figure F.14). This supershift was not replicated using the SW1353 or HAC cell line nuclear extracts (Appendix F: Figure F.15). Additionally, the appearance of the protein beneath complex 3 was not supershifted by the antibodies, meaning its identity could not be revealed. The competition and supershift EMSA results of transcription factors binding to rs11049206 are summarised in Table 5.12.

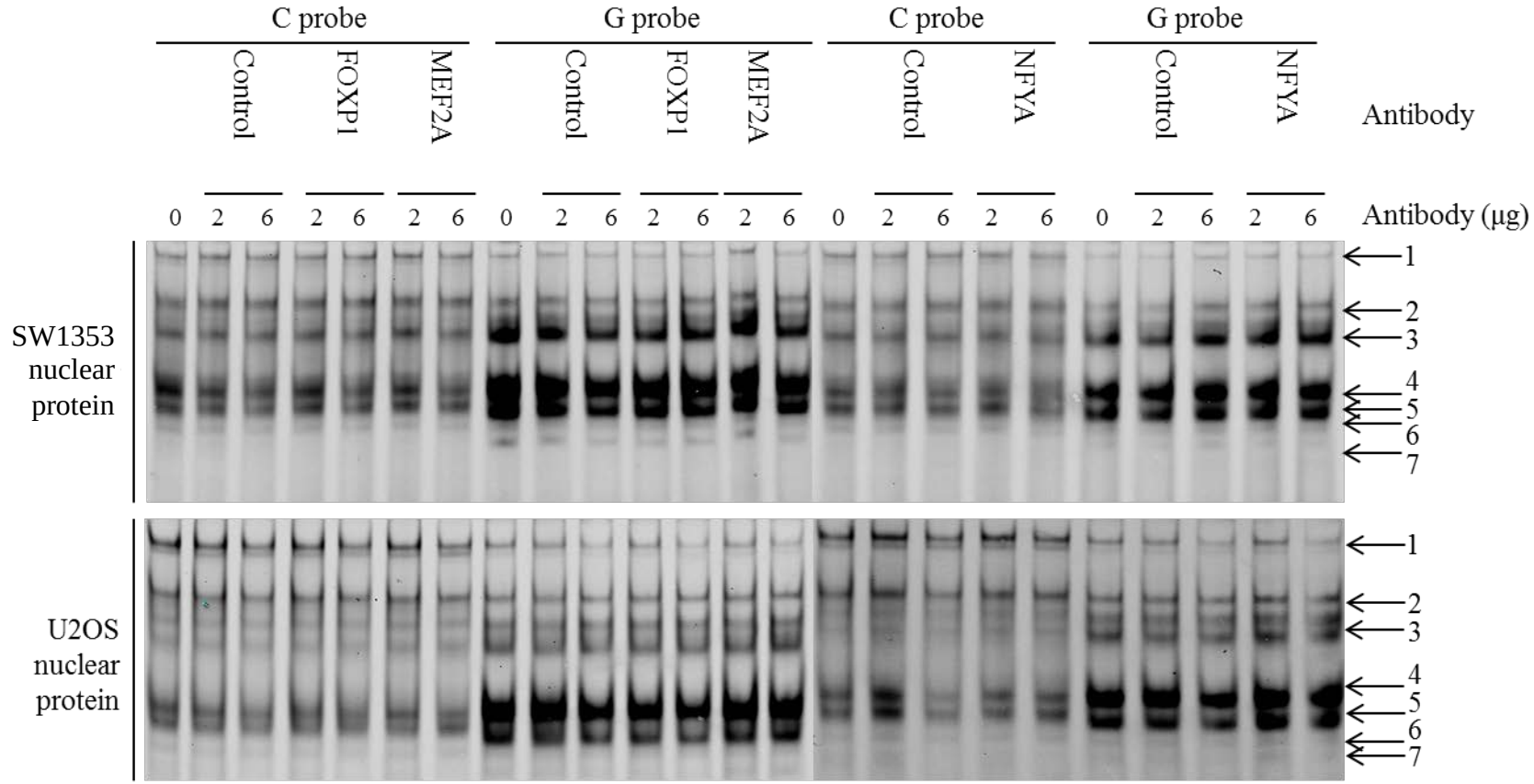

Figure 5.20. Supershift EMSAs to investigate the transcription factors of SW1353 and U2OS nuclear extracts binding to the C and G alleles of rs11049206. The protein:probe mixes were incubated with either 2 µg or 6 µg of antibody. No changes in the banding patterns were observed after incubation with any of the antibodies. Control (IgG species-matched antibody).

	Competition EMSA				Supershift		
Transcription factor	**Nuclear protein**	**Allele**	**Effect**	**Complex**	**Nuclear protein**	**Allele**	**Complex**
FOXP1	SW1353 and U2OS	C and G	Outcompeted	1, 3, 4 and 6	No effect	No effect	No effect
FOXA2	No effect	No effect	No effect	No effect	No effect	No effect	No effect
MEF2A	SW1353 and U2OS	C only	Outcompeted	4 and 5	No effect	No effect	No effect
NFIC	SW1353 and U2OS	C only	Outcompeted	4 and 5	No effect	No effect	No effect
NFYA	SW1353	C and G	Outcompeted	4 and 5	No effect	No effect	No effect
NFYB	SW1353 and U2OS	C and G	Outcompeted	4 and 5	No effect	No effect	No effect
RUNX2	SW1353	C and G	Outcompeted	2	No effect	No effect	No effect
RELA	No data	No data	No data	No data	U2OS	G only	4 and 5

Table 5.12. Summary of competition and supershift EMSAs to investigate transcription factor binding to rs11049206. All transcription factors, apart from RELA, were predicted to bind the SNP. RELA was positively identified as binding rs11049206 through supershift EMSAs. The nuclear protein extracts for which effects were observed are listed in the 'Nuclear protein' column.

5.3.9 *Investigating transcription factor binding to rs10843013 using chondrosarcoma and osteosarcoma cell line nuclear protein*

I have shown that the risk (C) allele caused an increased enhancer activity relative to the non-risk (A) allele of rs10843013 in the SW1353 cell line. Importantly, this regulation mirrored that observed for rs10492367. EMSA conditions were first optimised using glycerol, NP-40, KCl, $MgCl_2$ and EDTA as optional additional components (Appendix F: Figure F.1). Protein binding to the A and C allele probes produced seven distinct complexes, highlighted by the numbered arrows, with little difference between the conditions. Therefore, NP-40 was selected as the only additional component in the EMSA reaction mixes for rs10843013.

Following optimisation of the binding conditions, the specificity and affinity of the protein complexes to the A and C allele probes were tested (Figure 5.21). Even without competitors, the banding pattern was more intense for the A allele probe than for the C allele probe, suggesting that the proteins bind more avidly to the A allele. This was confirmed for complex 3, which required a higher concentration of its competitor to outcompete binding than the C allele probe. Moreover, protein binding to the A allele probe of complexes 1 and 6 was outcompeted by the A allele competitor, while complex 2 was specific to the U2OS nuclear extract.

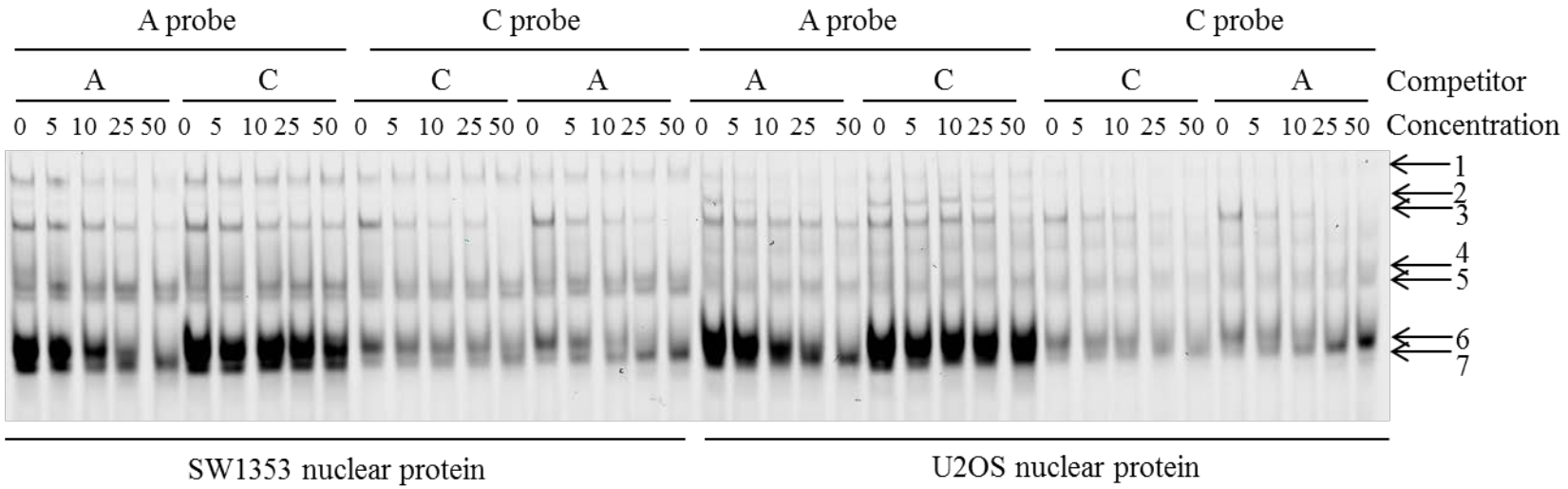

Figure 5.21. Competition EMSAs to investigate allele-specific binding of SW1353 and U2OS nuclear extract to the A and C alleles of rs10843013. The concentrations of unlabelled competitors were increased from 0 to 50 times that of the labelled probes. Protein binding to the A allele probe in complex 1 was outcompeted by the A allele competitor. Complex 2 is only present with the U2OS nuclear extract and was outcompeted at lower concentrations of the A allele competitor relative to the C allele competitor. Protein in complex 3 was binding more strongly to the A allele probe as a greater concentration of competitor was required to outcompete binding. Binding of complex 6 to both allele probes was disrupted upon the addition of the A allele competitor with both nuclear extracts. Concentration = 0, 5, 10, 25, 50 x probe concentration.

The region of the labelled probe that was involved in binding the protein was investigated by incubating the basic reaction mix with competitors that consisted of one region identical to the labelled probe and the remaining sequence entirely random (Table 5.13). The random sequence had no effect on protein binding to either allele probe, indicating that the protein:DNA complexes are specific for the probe sequences (Figure 5.22). Competitors 1 and 3 both caused the intensity of complex 7 to increase in addition to the appearance of bands beneath this complex for the C allele probe only. Competitor 2 outcompeted complex 2 with the U2OS nuclear extract only, meaning the central portion of the probe is required for protein binding in this complex.

Sequence name	Probe sequence (5'-3')
Random primer full length	CCACCGTAGGCAGGACGAACAGTTGCCCCAT
A allele probe and competitor	AGCAGTACCAGTTTTAACTAGCACAGAAGTA
C allele probe and competitor	AGCAGTACCAGTTTTCACTAGCACAGAAGTA
Competitor 1: A allele	AGCAGTACCAGTTTTAAAACAGTTGCCCCAT
Competitor 1: C allele	AGCAGTACCAGTTTTCAAACAGTTGCCCCAT
Competitor 2: A allele	CCACCGTAGAGTTTTAACTAGCTTGCCCCAT
Competitor 2: C allele	CCACCGTAGAGTTTTCACTAGCTTGCCCCAT
Competitor 3: A allele	CCACCGTAGGCAGGTAACTAGCACAGAAGTA
Competitor 3: C allele	CCACCGTAGGCAGGTCACTAGCACAGAAGTA

Table 5.13. Primer sequences used for competition EMSAs to investigate the regions of the A and C allele probes of rs10843013 to which the SW1353 and U2OS nuclear extracts bind. Each primer was annealed to its reverse complement, creating dsDNA, prior to use in EMSAs. The original competitor sequences are underlined in red; the random sequences are in black text.

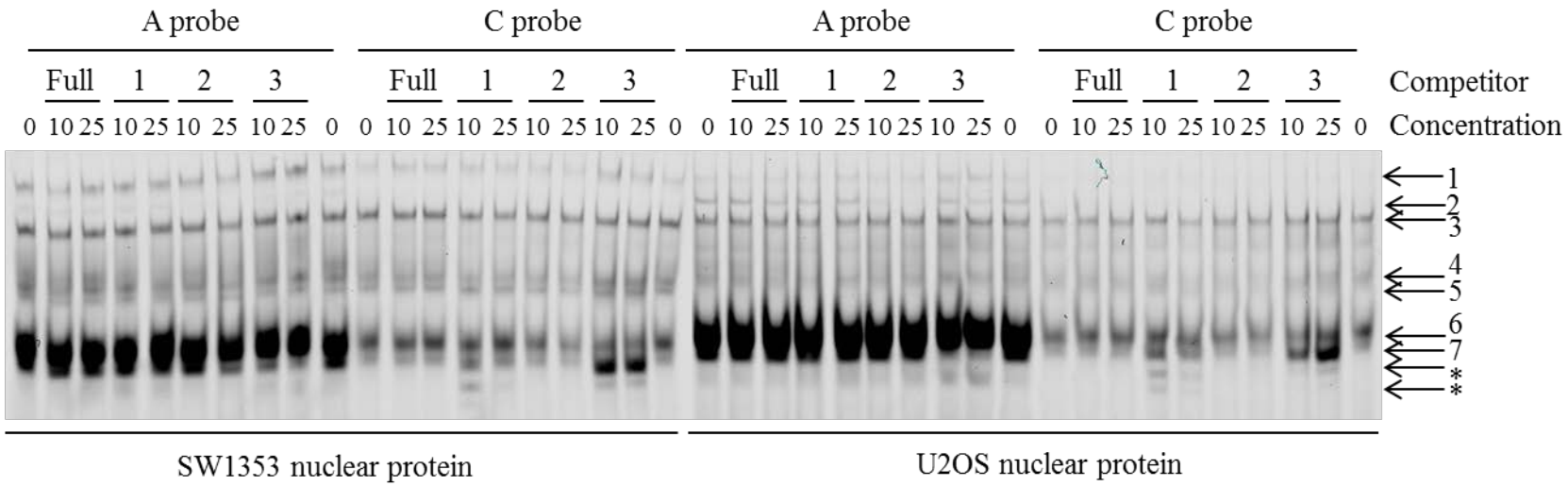

Figure 5.22. Competition EMSAs to investigate the regions of the A and C allele probes of rs10843013 to which the SW1353 and U2OS nuclear extracts bind. The protein:probe mixes were incubated with unlabelled competitors that had random sequences replacing the original competitor sequence. An entirely random competitor (full) had no effect on any of the protein:probe complexes. Competitor 2 caused the band intensity of complex 2 to decrease with the U2OS nuclear extract. Competitors 1 and 3 caused the band intensity of complex 7 with the C allele probe to increase and additional bands to appear, marked by asterisks (*). Concentration = 0, 10, 25 x probe concentration. Competitor = 31 bp random competitor (full); 5′-most region of the competitor replaced by a 14 bp random sequence (1); the central 13 bp region flanked either side by a 9 bp random sequence (2); 3′-most region of the competitor replaced by a 14 bp random sequence (3).

There were no transcription factors that were known or predicted to bind rs10843013 following database searches. As a result, only the panel of *trans*-acting factors that were known to be expressed in cartilage (Appendix E: Table E.1) were tested by means of supershift EMSAs (Appendix F: Figure F.16, Figure F.17 and Figure F.18). The basic reaction mixes were incubated with 2 μg of antibody. Incubation of the C allele probe with an antibody for RELA (NFκβ p65) appeared to cause complexes 6 and 7 to become fainter with the SW1353 nuclear extract. Additionally, complex 2 became fainter and complex 1 more intense upon the incubation of both allele probes with an antibody for XBP1, which is suggestive of a positive supershift. The EMSAs could not be replicated for the RELA (NFκβ p65) incubation, while the XBP1 supershifts were only replicated with the U2OS nuclear extract (Figure 5.23). The supershift EMSA results of transcription factors binding to rs10843013 are summarised in Table 5.14.

Overall, the EMSA investigations in this and the preceding sections (Chapter 5.3.6, Chapter 5.3.7 and Chapter 5.3.8) have identified a number of transcription factors that bind the most promising functional polymorphisms within the association interval. Again, it would not be realistic to further investigate all four of these polymorphisms, and so the focus of this investigation must be narrowed further. Of the EMSA experiments, the most compelling results were for rs10492367, as typical supershifts were observed and replicable. Therefore, this SNP and the proteins identified as binding to it, were selected for the next stage of this investigation.

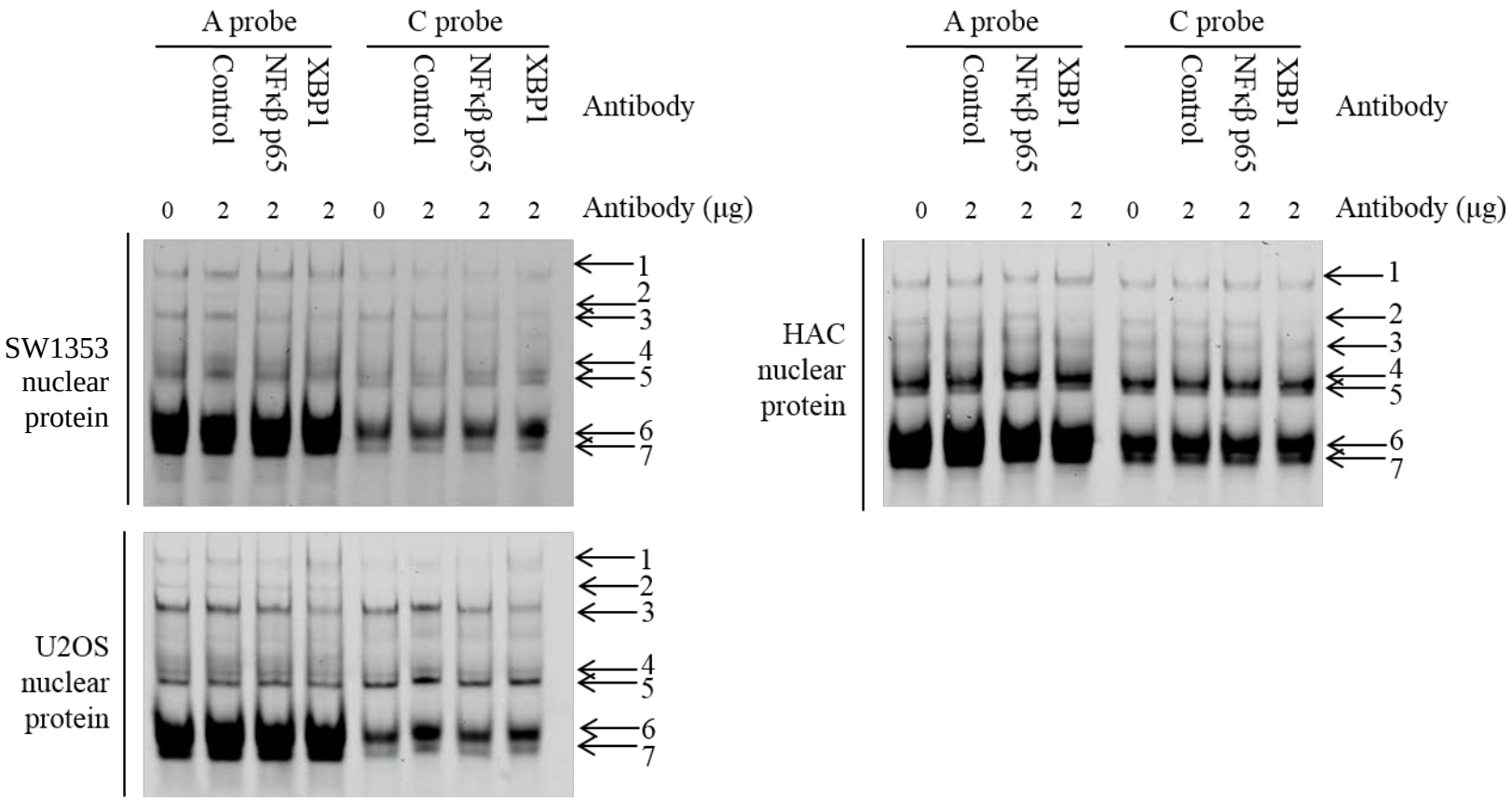

Figure 5.23. Supershift EMSAs to confirm the transcription factors of SW1353, U2OS and HAC nuclear extract binding to the A and C alleles of rs10843013. The protein:probe mixes were incubated with 2 μg of antibody. The results showed that RELA (NFκβ p65**)** of both nuclear extracts does not interact with the fluorescently labelled DNA. An extremely slight supershift of band 2, causing band 1 to become more intense with the XBP1 antibody, was seen with the U2OS nuclear extract only. Control (IgG species-matched antibody). Lanes irrelevant to this analysis have been removed.

	Competition EMSA				Supershift		
Transcription factor	**Nuclear protein**	**Allele**	**Effect**	**Complex**	**Nuclear protein**	**Allele**	**Complex**
RELA	No data	No data	No data	No data	SW1353	C only	6 and 7
XBP1	No data	No data	No data	No data	U2OS	A and C	1 and 2

Table 5.14. Summary of supershift EMSAs to investigate transcription factor binding to rs10843013. No transcription factors were predicted to bind the SNP. RELA and XBP1 were positively identified as binding to rs10843013 in a cell line-specific manner. The nuclear protein extracts for which effects were observed are listed in the 'Nuclear protein' column.

5.3.10 *Characterisation of trans-acting factor expression profiles in SW1353, U2OS and human articular chondrocyte (HAC) cell cultures*

Since SUB1, RELA and TCF3 were identified as binding to rs10492367 and were selected as the most promising candidates for further functional characterisation, it was necessary to confirm the presence of the *trans*-acting factors in the cell lines that were used in the EMSA experiments. Firstly, the expressions of the genes were confirmed in SW1353 and U2OS cell lines, and additionally in five HAC donors using qPCR assays detailed in Figure 5.24 (sequences are listed in Appendix A: Table A.3). All transcription factors were expressed: SUB1 was the most highly expressed (Figure 5.25). The variability of the HAC gene expression was presumably caused by inter-individual differences in gene expression – this is not seen in the SW1353 cell line or the U2OS cell line, as the replicates originate from the same cell culture.

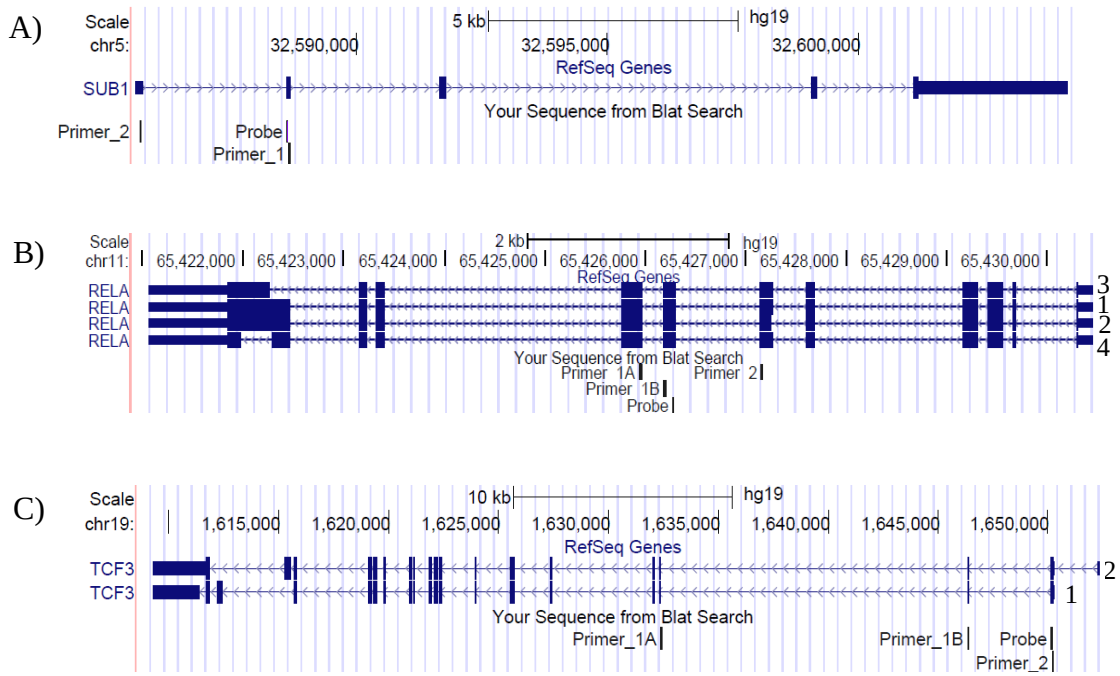

Figure 5.24. Location of qPCR primers and probes used for quantitative gene expression analysis. Assays were predesigned to the exons of A) *SUB1*, B) *RELA* and C) *TCF3*. Primer 1 for *RELA* and *TCF3* spanned an intron in each gene, targeting only the coding exons. Transcript isoforms are numbered for *RELA* and *TCF3*. The images were obtained using the hg19 reference genome.

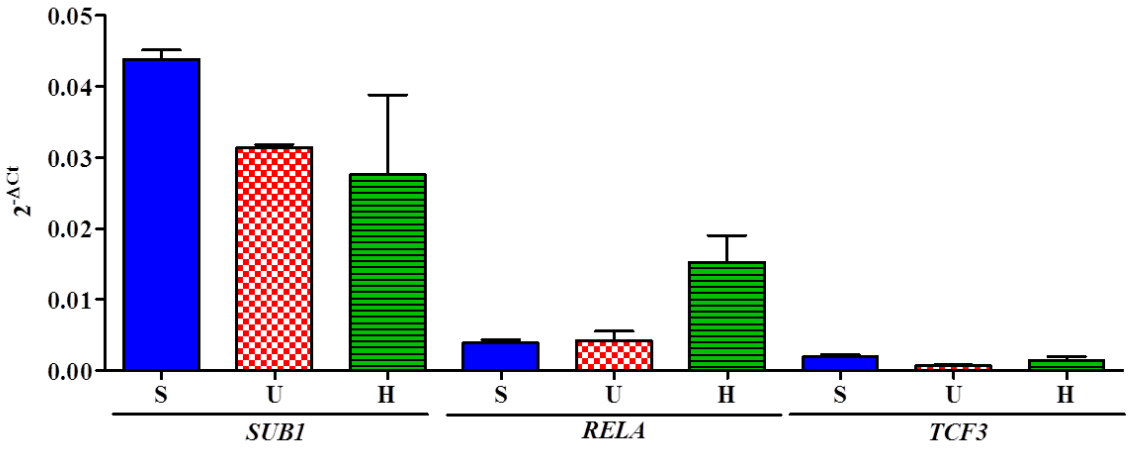

Figure 5.25. Expression of *SUB1*, *RELA* and *TCF3* in SW1353, U2OS and HAC cells. All transcription factors were expressed, with *SUB1* the most highly expressed, while *RELA* and *TCF3* were more lowly expressed. RNA was extracted, cDNA synthesised and three technical repeats used for qPCR to assess gene expression for the SW1353 and U2OS cell lines. RNA was extracted from five HAC donors, cDNA synthesised and three technical repeats performed for qPCR per donor. S (SW1353), U (U2OS) and H (HAC). Error bars represent the SEM.

To corroborate these findings, the protein levels were also investigated. All three transcription factors were present in each of the cell line extracts (Figure 5.26). RELA was most abundant in the HAC nuclear protein, while SUB1 was most abundant in the SW1353 extract. These observations confirmed the gene expression data. TCF3 was most abundant in the U2OS nuclear protein and was only detected in the HAC extract when a stronger exposure was used, which was not in keeping with the expression levels detected. A cause of the differences may be that firstly, the HAC donors used for gene expression quantification were not the same as those used for protein detection; secondly, the protein and total RNA were not extracted simultaneously from the SW1353 and U2OS cell cultures; and thirdly, the different cell types may have differing transcript stabilities, translation efficiencies or protein stabilities. Finally, as this was a qualitative assessment of protein levels, a loading control was not used, and so the discrepancies could be caused by different amounts of protein in each lane. Nevertheless, the purpose of the western blot was achieved, simply confirming the presence of each of the transcription factors in the protein extracts.

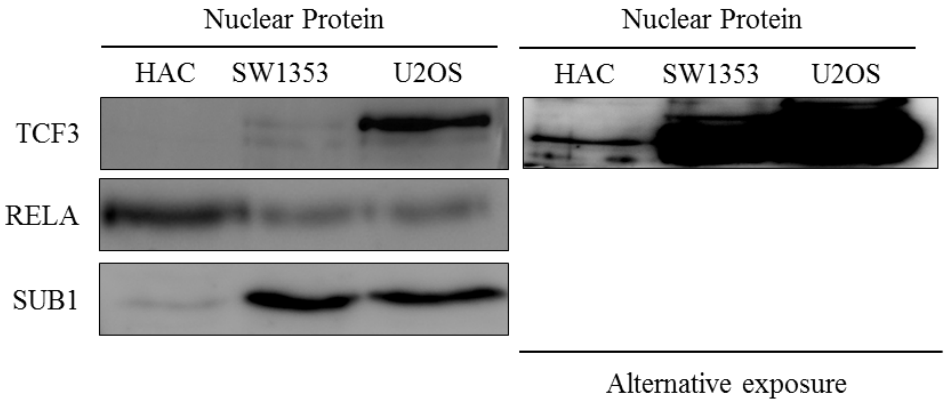

Figure 5.26. Detection of protein levels of TCF3, RELA and SUB1 in HAC, SW1353 and U2OS nuclear protein. The *trans*-acting factors were detected in all nuclear extracts. A stronger exposure of the TCF3 blot allowed the detection of the protein in HAC nuclear extracts.

5.3.11 *Validation of SUB1, RELA and TCF3 binding to rs10492367 using chromatin immunoprecipitation (ChIP)*

Having putatively shown that SUB1, RELA and TCF3 bind the association SNP in *in vitro* EMSAs, I next aimed to corroborate these findings *in vivo* using ChIP (Chapter 2.24). This technique involves cross-linking the protein to the DNA *in situ* in the cell, allowing the detection of transcription factor binding to the endogenous genomic loci *in vivo*. As differential activity between the alleles of rs10492367 was observed in SW1353 cells, and the transcription factors were identified as binding the SNP in the nuclear extract from this cell line, the chondrosarcoma cell line was selected for ChIP. The cross-linked chromatin was first sonicated to fragments of 100 – 300 bp in size (Figure 5.27).

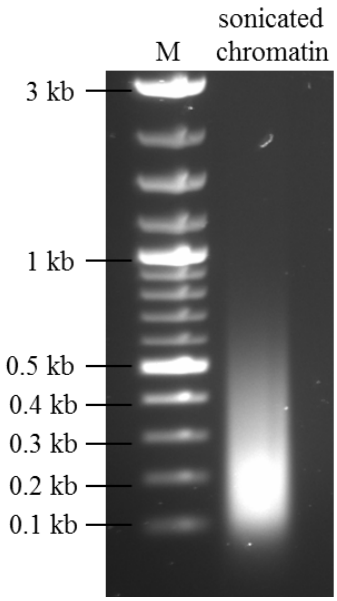

Figure 5.27. Gel electrophoresis of the sonicated chromatin extracted from an SW1353 cell culture. The chromatin was sonicated to fragment sizes of 100 – 300 bp prior to ChIP. The sample was electrophoresed through a 2% agarose TBE gel. M (DNA marker).

In order to pull-down the proteins of interest, the sonicated chromatin was incubated with antibodies specific for SUB1, TCF3, RELA and acetyl-histone H3. Acetyl-histone H3 is a marker of active gene transcription, so combined with a constitutively active housekeeping gene, in this case *GAPDH*, this acted as positive control. In addition, IgG was used as a species-matched control to measure the basal levels of non-specific antibody binding to the chromatin. Following the pull-down, the cross-links between the protein and DNA were reversed and the DNA purified. The DNA was quantified by qPCR, investigating if the rs10492367 locus was enriched in the DNA immunoprecipitated by the different antibodies (Figure 5.28.A). Additionally, the same quantification was carried out for the *GAPDH* locus (Figure 5.28.B) and a testis-specific locus (protamine 2 [*PRM2*]; Figure 5.28.C) as positive and negative controls, respectively. The qPCR primer and probe sequences are listed in Appendix A: Table A.3. All data were normalised to the non-specific binding of the species-matched IgG and repeated three times for each of the two rounds of sonication.

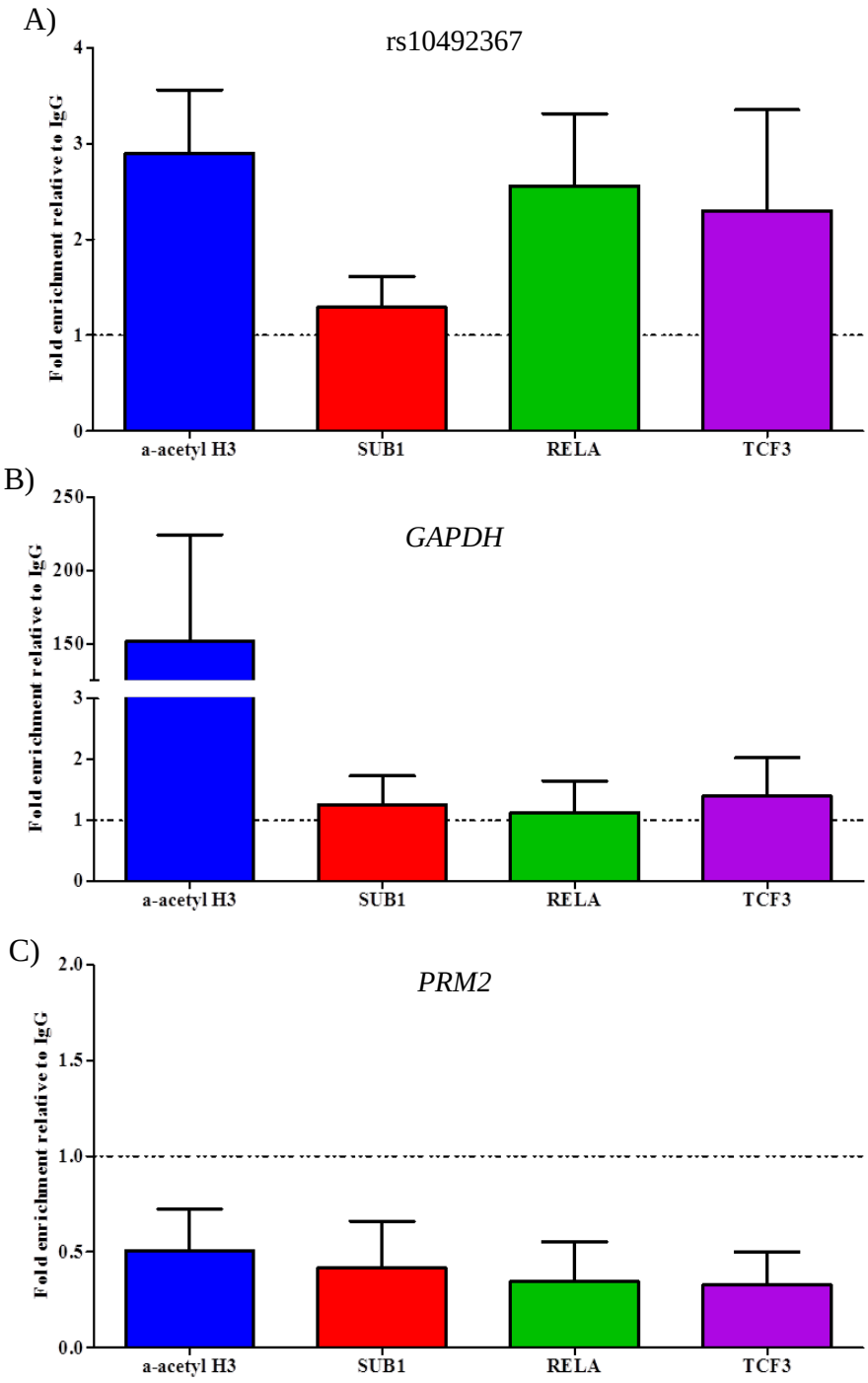

Figure 5.28. Fold enrichment of the *trans*-acting factors SUB1, RELA and TCF3 at rs10492367 and the *GAPDH* and *PRM2* loci. Protein:DNA complexes from sonicated SW1353 chromatin were captured using antibodies specific for the transcription factors SUB1, RELA and TCF3, and the controls acetylated histone H3 and IgG. qPCR was used to assess if the proteins were enriched at A) rs10492367, B) the *GAPDH* active promoter, and C) the testis-specific *PRM2* gene. SUB1, RELA and TCF3 were enriched at rs10492367 although these did not reach statistical significance relative to IgG. The positive control of acetyl-histone H3 was enriched at the *GAPDH* locus, but was not significant. None of the proteins were enriched at the *PRM2* locus. Three independent replicates were performed for each of the two independent cell culture and sonications. Error bars represent the SEM. Statistical significance was assessed using the Student's *t* test.

None of the transcription factors were significantly enriched at the rs10492367 locus, although there was a trend of enrichment relative to the non-specific binding of IgG. This could be due to the SW1353 cell line being homozygous for the major allele of rs10492367. Perhaps the *trans*-acting factors bind more strongly to the T allele *in vivo*, and as such, a significant enrichment was not observed. Acetyl-histone H3 was considerably enriched at the *GAPDH* locus, however variation in the biological replicates meant that this was not significant. Crucially, the transcription factors were depleted at the *PRM2* locus. Overall, the ChIP investigations showed promising, but not definite, trends of protein binding.

5.3.12 *Knockdown of SUB1, RELA and TCF3 in human articular chondrocytes (HACs)*

Since protein binding to rs10492367 was identified using EMSAs and ChIP, the next stage of investigation was to characterise if the *trans*-acting factors function to modulate *PTHLH* and/or *KLHL42* expression. As all transcription factors were shown to be expressed in HACs, I aimed to deplete the expression of each gene and assess the resulting influence on *PTHLH* and *KLHL42*. Chondrocytes from the knee cartilage of three OA donors (Table 5.15) were extracted, cultured and transfected with siRNAs targeted to *SUB1*, *RELA* or *TCF3* (Chapter 2.30). A non-targeting siRNA was also included as a transfection control.

Donor	Sex	Age at joint replacement (years)	Joint	rs10492367
226	F	58	Knee	GT
227	M	67	Knee	GG
228	F	74	Knee	GT

Table 5.15. Characteristics and genotype at rs10492367 for donors used in the HAC knockdowns of SUB1, RELA and TCF3.

Following total RNA and protein extraction from each condition, gene expressions were quantified for *SUB1*, *RELA* and *TCF3* in the corresponding knockdowns and in the non-targeting control condition. All genes were significantly downregulated ($p < 0.0001$ [*SUB1*], $p = 0.001$ [*RELA*], $p = 0.010$ [*TCF3*]) after targeted siRNA knockdown relative to their expressions in the non-targeting siRNA control (Figure 5.29.A). Moreover, the resulting levels of protein after the siRNA knockdowns were assessed by a western blot. As previously shown in Figure 5.25 and Figure 5.26, the protein level of TCF3 and SUB1 were low in HACs. Accordingly, it was with difficulty that protein levels of the knockdowns were detected. Figure 5.29.B is a representation of the knockdowns, using the clearest images for

this figure. Nevertheless, a depletion of each transcription factor was observed, with the loading controls confirming this was not an artefact of unequal quantities of protein between conditions.

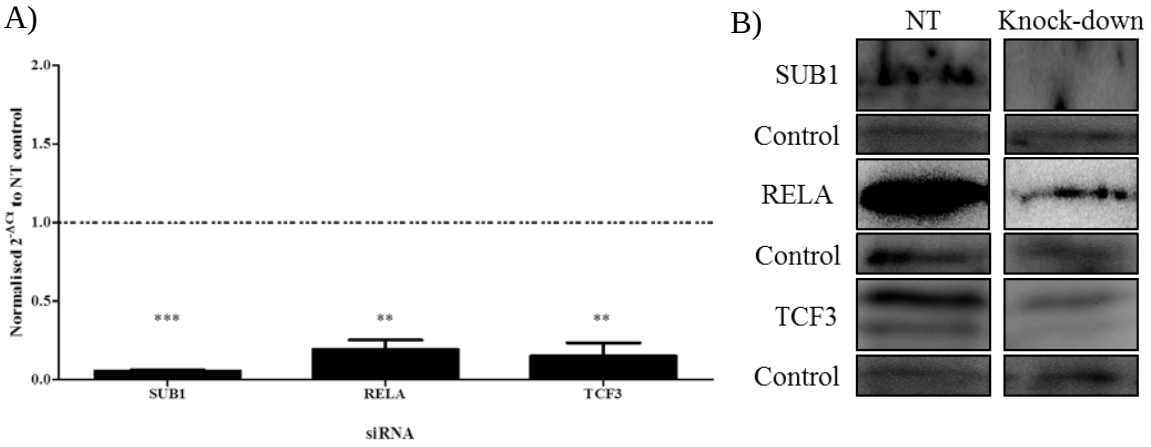

Figure 5.29. Knockdown of the *trans*-acting factors *SUB1*, *RELA* and *TCF3* in human articular chondrocytes (HACs). Chondrocytes, after isolation from the cartilage of OA donors who had undergone knee replacement surgery, were transfected with non-targeting siRNA, or siRNA specifically targeting *SUB1*, *RELA* or *TCF3*. A) Gene expression for each condition was quantified using qPCR and normalised to the corresponding expression in the non-targeting siRNA control (*y*-axis = 1). Gene expression was significantly downregulated relative to the expression in the non-targeting control for all conditions. *n* = knockdown of three independent donors (biological replicates) with one technical replicate each. qPCR was performed as standard, including three technical replicates. Error bars represent the SEM. Statistical significance was assessed using a paired Student's *t*-test. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. B) Western blot to confirm protein depletion. Each of the three knockdowns resulted in a decrease at the protein level. The loading controls are shown directly beneath the corresponding knockdowns. GAPDH was the loading control used for the SUB1 and TCF3 knockdowns. The loading control for the RELA knockdown was TCF3. SUB1 and TCF3 are from patient 277, and RELA is from patient 226. NT (non-targeting siRNA control).

As the siRNAs successfully decreased the expressions of the *trans*-acting factors, which was also confirmed at the protein level, the expressions of *PTHLH* and *KLHL42* were next assessed. In every condition, the gene expression levels were quantified by qPCR and normalised to the corresponding values in the non-targeting siRNA controls: the non-targeting control represents an arbitrary value of 1. After normalisation, there was considerable variation in *PTHLH* and *KLHL42* expression between the biological replicates for the *RELA* and *TCF3* knockdowns. However, knockdown of *SUB1* resulted in a consistent, significant ($p = 0.018$) downregulation of *PTHLH* (Figure 5.30).

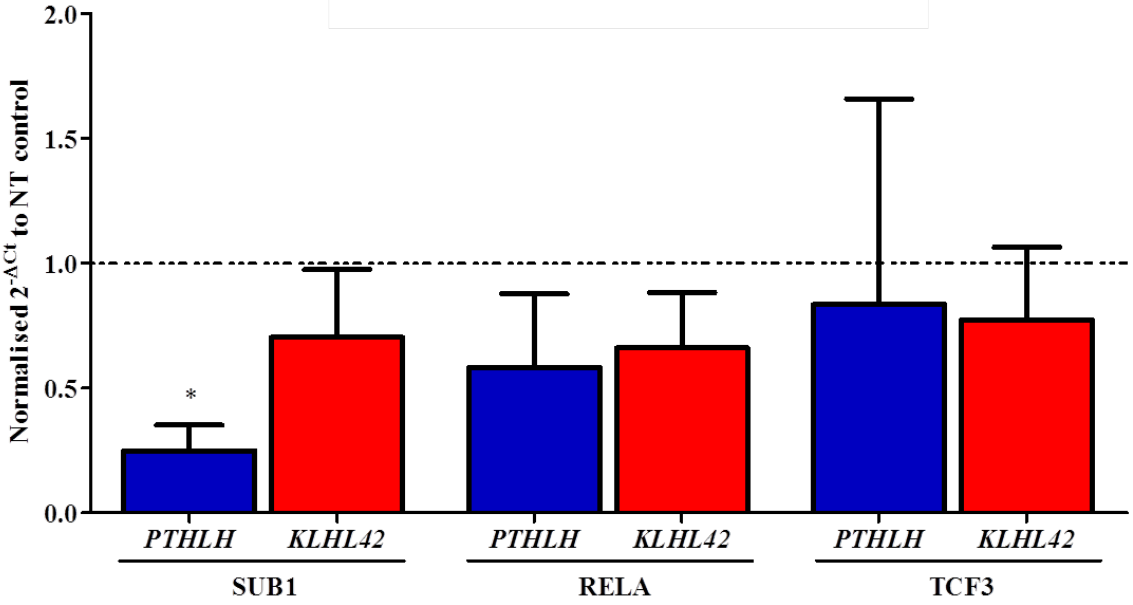

Figure 5.30. *PTHLH* and *KLHL42* expression after knockdown of the *trans*-acting factors SUB1, RELA and TCF3 in human articular chondrocytes (HACs). Chondrocytes, after isolation from the cartilage of OA donors who had undergone knee replacement surgery, were cultured until confluent and then transfected with non-targeting siRNA, or siRNA specifically targeting *SUB1*, *RELA* or *TCF3*. *PTHLH* and *KLHL42* expression for each condition was quantified using qPCR and normalised to the corresponding expression in the non-targeting siRNA control (*y*-axis = 1). *PTHLH* was significantly downregulated following knockdown of SUB1 relative to the expression in the non-targeting control. *n* = three independent biological replicates. Error bars represent the SEM. Statistical significance was assessed using a paired Student's *t*-test. * $p < 0.05$.

5.3.13 *Assessing if the knockdown of SUB1, RELA or TCF3 has differential effects on the enhancer activity of pGL3-promoter vectors containing the alleles of rs10492367*

The final route of my investigation to characterise the OA association region was to assess the combined effects of transcription factor knockdown on the differential enhancer activity of rs10492367. The previous luciferase studies (Chapter 5.3.3) demonstrated that the allelic differences in rs10492367 activity were functional only in the SW1353 cell line, and so for the following characterisation, this chondrosarcoma cell line was utilised. The pGL3-promoter vector constructs containing the alleles of rs10492367 (previously generated in Chapter 5.3.3) were similarly used for this investigation. Five hours post-transfection with each allele construct, the cell populations were transfected with the *SUB1*, *RELA*, *TCF3* siRNAs or the non-targeting siRNA control (Chapter 2.31). The luciferase activities were quantified after 24 hours and for each vector construct, the values were normalised to those of the corresponding vectors that had been transfected with the non-targeting siRNA control.

Firstly, the expressions of *SUB1*, *RELA* and *TCF3* were quantified using qPCR to ensure the knockdowns were successful (Figure 5.31). All siRNA transfections resulted in the significant downregulation of the targeted gene relative to the corresponding non-targeting control. Due to insufficient yields of protein extracted from the cells, it was not possible to confirm knockdown at the protein level.

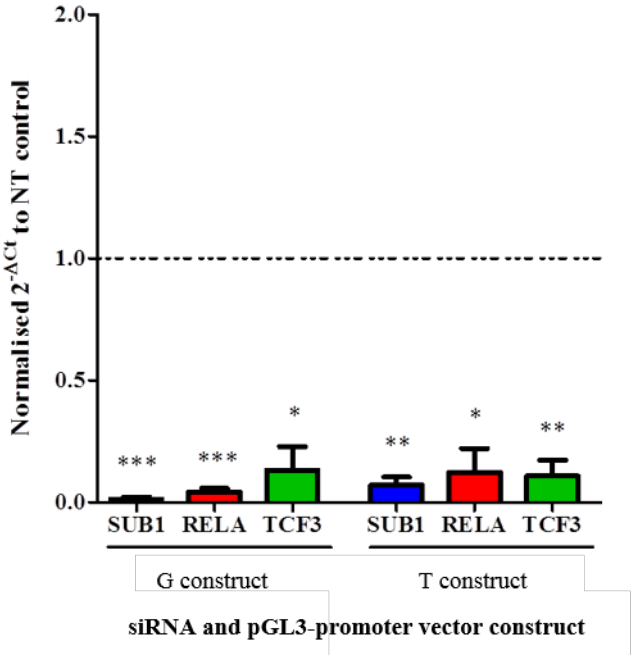

Figure 5.31. Knockdown of the *trans*-acting factors *SUB1*, *RELA* and *TCF3* in the SW1353 cell line. pGL3-promoter vectors contained either the G or the T allele of rs10492367. After transfection for 5 hours with the luciferase reporter vectors, cells were transfected for 24 hours with *SUB1*, *RELA*, *TCF3* or non-targeting (NT) siRNA. Gene expression for each condition was quantified using qPCR and normalised to the corresponding expression in the non-targeting siRNA control (*y*-axis = 1). Gene expression was downregulated relative to the expression in the non-targeting control for all conditions. *n* = three independent biological replicates. Error bars represent the SEM. Statistical significance was assessed using a paired Student's *t*-test. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Since the knockdowns were confirmed at the gene level, I quantified the luciferase activities of the constructs. Knockdown of *SUB1*, *RELA* and *TCF3* had no significant effect on enhancer activity for either of the constructs relative to the corresponding non-targeting control. In addition, there were no differences in enhancer activity between the allelic constructs following *SUB1* and *RELA* knockdown. However, the construct containing the G allele of rs10492367 had significantly lower levels of enhancer activity relative to the T allele construct ($p = 0.004$) after *TCF3* knockdown. Overall, this investigation revealed that *in vitro*,

a lack of TCF3 correlated with an increase in enhancer activity of rs10492367 only in the presence of the risk (T) allele.

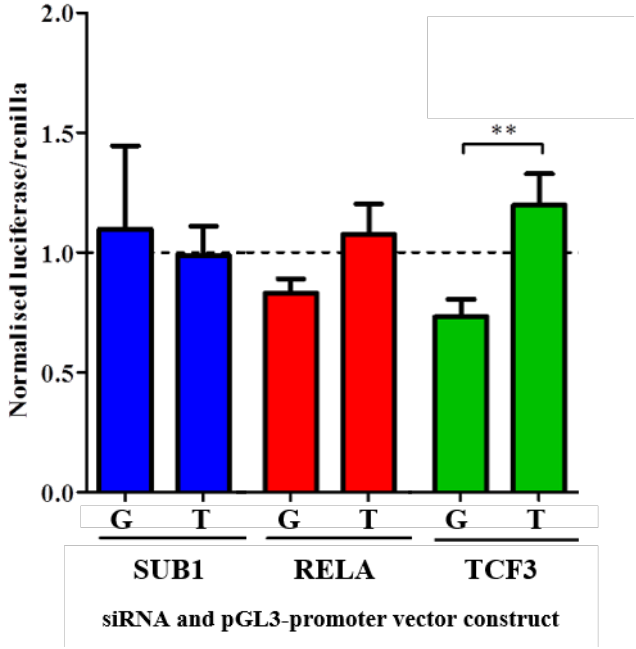

Figure 5.32. Co-transfection of pGL3-promoter vector constructs for rs10492367 and siRNA targeting *SUB1*, *RELA* and *TCF3* in the SW1353 cell line. pGL3-promoter vectors contained either the G or the T allele of rs10492367. After transfection for 5 hours with the luciferase reporter vectors, cells were transfected for 24 hours with *SUB1*, *RELA*, *TCF3* or non-targeting siRNA before the activity of the luciferase gene downstream of the inserts was quantified. All absorbances were read at 595 nm and the luciferase readings normalised to the internal control (*Renilla*). Luciferase/*Renilla* values were normalised to the basal levels of activity measured with the corresponding vector co-transfected with a non-targeting control (*y*-axis = 1). When transfected with TCF3 siRNA, the activity of the G allele construct was significantly lower than the T allele construct. Error bars represent the SEM. Statistical significance was assessed using the Mann-Whitney *U* test. ** $p < 0.01$.

5.4 Discussion

In Chapter 5, I aimed to functionally dissect the OA association of the 12p11.22 signal marked by the polymorphism rs10492367. I performed luciferase reporter assays to assess differential allelic activity of the SNPs in high LD with the association signal, EMSAs to identify differential transcription factor binding, and siRNA knockdowns to characterise the effect of the identified transcription factors on *PTHLH* and *KLHL42* expression.

I began with a panel of ten intergenic polymorphisms that could be responsible for the OA association of the region, one of which was rs10492367 itself: the remaining nine SNPs all had an LD of at least 0.80 with rs10492367. Luciferase reporter assays are an established technique used to identify differential allelic activities (Egli *et al.*, 2009), and so I similarly used this to investigate the enhancer activities of the SNPs in this region. A limitation that could not be avoided was the exclusion of rs57380671 from the investigation, as specific constructs representing this region could not be generated. This was despite different *Taq* polymerases, PCR conditions and annealing temperatures being used during the amplification stage. Nevertheless, the remaining nine SNPs were investigated for differential enhancer activity, the diverse results of which required the design of a set of criteria to shortlist the most promising functional candidates. As a result, rs10492367 and three other polymorphisms were carried forward to investigate protein binding.

I hypothesised that differential transcription factor binding to the SNPs could regulate gene expression in an allele-specific manner. As is commonly performed to characterise protein:DNA interactions, I used EMSAs as the basis to investigate this (Syddall *et al.*, 2013). To summarise: RELA (NFκβ p65) in the U2OS nuclear extract was identified as binding to the G allele of rs11049206; XBP1 in the U2OS nuclear extract was identified as binding to both alleles of rs10843013; SUB1 (PC4) in all cell line nuclear extracts was identified as binding the T allele of rs58649696; and SUB1 (PC4), TCF3 and RELA (NFκβ p65) in all cell line nuclear extracts were identified as binding to rs10492367. Although the banding patterns for each SNP were similar between cell line extracts, the fact that some of the supershifts were cell line-specific suggests that the complexes are not composed of the same proteins. This could also explain the differences observed between cell lines for the luciferase reporter assays. Despite numerous replications, some of the EMSA results were inconsistent, a possible result of the artificial conditions of the experiment. The labelled probe DNA is linear and is thus not representative of the native three-dimensional conformation of the chromatin, which could influence the avidity of transcription factor binding. For example, therefore,

SUB1 (PC4) may in fact only bind one allele of rs10492367 *in vivo* but appears to bind both alleles in the EMSA conditions. Additionally, using nuclear protein extracted at different cell passages appeared to affect the EMSA banding patterns. While this did not influence the outcome of these investigations, it highlights that transcription factors can have differential activity with different population doublings. Perhaps the activities can change or different complements of *trans*-acting factors are present depending on the passage number of the cells. Overall, I have identified that the SNPs under study were functional, with a range of transcription factors binding in the assays used.

To further assess the protein:DNA binding, ChIP was used because it offers a native representation of the cellular interactions, allowing the study of transcription factor binding to the endogenous loci rather than the artificial conditions produced in the EMSA experiments. For both the ChIP and the subsequent knockdown experiments, the four SNPs under investigation were further reduced to only one. This would allow for a more in-depth analysis in the timeframe available. It was rs10492367 that was selected as differential enhancer activity was observed in addition to the identification of three potential transcription factors binding. Although none of the results reached significance, ChIP did indicate that the transcription factors were enriched at rs10492367 in the SW1353 cell line, with enrichment beginning to approach similar levels observed for other loci and proteins (Verzi *et al.*, 2010; Ouma *et al.*, 2015). As this was using only SW1353 cell cultures, further experiments could utilise other cells such as HACs, which may have different abundances of protein present or different regulatory mechanisms. Additionally, it would be prudent to investigate cell lines that are heterozygous or minor allele homozygous for the association SNP, which could affect the avidity of protein binding and could be reflected in the ChIP enrichment.

The results yielded in the experiments thus far indicated that SUB1 (PC4), TCF3 and RELA had the potential to bind rs10492367. As it was unclear whether this was directly relevant to the association signal, I sought next to investigate whether the proteins had an effect on *PTHLH* or *KLHL42* expression. I successfully knocked-down the transcription factors at the gene and protein levels, and indeed, observed a consistent downregulation of *PTHLH* following SUB1 (PC4) depletion. Although this did not implicate a particular genotype of rs10492367 as controlling gene expression through the action of differential SUB1 (PC4) binding, it did show that the transcription factor modulates *PTHLH* expression *in vivo* in HAC cell cultures. An MSC chondrogenesis or osteoblastogenesis model could be used to investigate if the effects of SUB1 (PC4) depletion on *PTHLH* act throughout MSC

differentiation and also whether this has a direct effect on joint development. By using cell cultures each with a different rs10492367 genotype, the effect of the polymorphism on the SUB1 (PC4) regulation of *PTHLH* could be established.

To investigate if the genotype of rs10492367 affected protein activity, I performed co-transfections with the constructs used for the luciferase reporter assays of Chapter 5.3.3. Again, this was using artificial conditions and so is only suggestive of possible correlations. A significant difference was observed between the alleles of rs10492367 following the knockdown of TCF3. This suggested that in the presence of the risk (T) allele of rs10492367, knocking-down TCF3 caused an increase in downstream luciferase expression relative to the non-risk (G) allele. There were no significant differences for SUB1 (PC4) and RELA.

SUB1 (PC4) mediates interactions between proteins and DNA thereby acting as a transcriptional coactivator (Conesa and Acker, 2010). The binding of SUB1 (PC4) to another OA-associated SNP has recently been investigated, and it was found to bind both alleles of the *GDF5* 5′ UTR polymorphism rs143383 (Syddall *et al.*, 2013). As for rs10492367, protein binding could not be linked to association SNP genotype. Crucially, the protein did not exert any significant effects in isolation, and so it was hypothesised that it instead acts as a linker in a larger protein complex. This could be explained by the lack of a definitive consensus sequence, meaning that the protein was not predicted to bind the DNA directly, yet supershift EMSAs revealed its presence at the SNP site. In addition, as it is a coactivator, it may well bind the alleles of rs10492367 with equal affinity and avidity, but it could have differential effects on binding other proteins. This could achieve overall differential transcription factor binding to the alleles of rs10492367, yet it would not be reflected in the binding of SUB1 (PC4).

TCF transcription factors are involved in the downstream cascade of the Wnt/β-catenin signalling pathway. This pathway has been implicated in OA progression (Corr, 2008), and so to this end, TCF proteins are of interest in OA aetiology. Should the protein be associated with OA progression, an upregulation of MMPs upon protein overexpression, resulting in the degradation of the cartilage ECM, would likely be observed. Conversely, however, a potential link between TCF3 and the mechanisms of OA was shown following the overexpression of the protein in HACs, where this resulted in a reduced MMP-1, -3 and -13 mRNA expression (Ma *et al.*, 2013). Overall, having shown that TCF3 binds rs10492367 yet does not correlate with genotype, the role of TCF3 in OA remains unclear.

RELA is the p65 subunit of the essential, widely-expressed transcription factor NFκβ, and represents the most obvious candidate protein of the three identified. Several studies have implicated RELA in cartilage homeostasis and the development of OA. For example, it has been reported that RELA has the capacity to activate anabolic factors such as SOX9, forming part of a network that can induce chondrogenic differentiation (Ushita *et al.*, 2009). Moreover, the protein is a reported activator of HIF-2α, which itself targets genes, such as *COL10A1*, that are involved in endochondral ossification (Saito *et al.*, 2010). RELA has also been shown to co-localise with MMP-3 in macrophages and smooth muscle cells (Souslova *et al.*, 2010), suggesting that perhaps it could regulate MMP-3 induction during cartilage degradation. Further evidence for the involvement of RELA in cartilage homeostasis is its activation of the ADAMTS5 aggrecanase, most likely as part of a larger molecular network (Kobayashi *et al.*, 2013). The authors consequently hypothesised that the protein induces aggrecanase activity in chondrocytes throughout OA development, highlighting its relevance in the dissection of the rs10492367 association signal.

In summary, the involvement of additional proteins cannot be excluded, given that SUB1 (PC4) was itself not predicted to bind rs10492367, yet is displaying the most promising results to date. In order to investigate whether the proteins interact with each other to enhance the effects on gene expression, an additional approach could be to perform co-immunoprecipitations. In addition, it may prove that the *in vivo* analyses proposed earlier in the discussion could reveal such enhanced effects as well as allele-specific differences in protein binding.

Chapter 6. General Discussion

6.1 Perspective

OA is an age-associated, multifactorial arthritis characterised by the progressive loss of articular cartilage in synovial joints. Through candidate gene studies, linkage studies and GWAS approaches, several regions of the genome have been identified as being associated with the polygenic disorder. The challenge that faces researchers, however, is to dissect the signals to such a degree that the OA association is fully understood and characterised. One mechanism through which polymorphisms can modulate disease susceptibility is by causing a change in the coding amino acid sequence and thus, altering protein function. Alternatively, SNPs can act as eQTLs to modulate gene expression, a strong example of which is the *GDF5* 5′ UTR polymorphism rs143383. This is the most robust OA signal that has been functionally studied to date, with the SNP mediating gene expression through the binding of Sp1, Sp3, and DEAF-1 (Syddall *et al.*, 2013).

The most powerful OA GWAS yet is the 2012 study by the arcOGEN Consortium (arcOGEN Consortium *et al.*, 2012). Five regions of the human genome were identified as being associated with the disease, two of which are the primary focus of the research discussed in this thesis. Despite likely candidate genes residing at the two loci, both rs10492367 and rs9350591 are novel in their association with OA. Other group members have interrogated the remaining three signals. Firstly, *GNL3* and *SPCS1* were shown to be subject to the actions of a *cis*-eQTL that correlated with the association SNP genotype (Gee *et al.*, 2014); the signal at *CHST11* is postulated to mediate OA susceptibility through differential protein binding (Reynard *et al.*, 2014); and finally, while no eQTL has been identified at the *ASTN2* locus, the signal is speculated to regulate *PAPPA* expression during joint development (*unpublished data; personal communication*).

Both rs10492367 and rs9350591 have so far remained largely unstudied, and have not previously been annotated in the context of musculoskeletal disorders. Indeed, of the nine polymorphisms that are in high LD with rs10492367, none have been associated with a musculoskeletal disease phenotype. For the SNPs in high LD with rs9350591, rs9360921 was reported to be associated with height in individuals of European descent (Lango Allen *et al.*, 2010). Not in LD with the association signal, but still at the locus of interest, the A allele of rs7953528, 2 kb upstream of rs10492367 and 94 kb downstream of *PTHLH*, is associated with variations in bone mineral density (Estrada *et al.*, 2012). This association is not

surprising given the known function of *PTHLH* in endochondral ossification and its presence as a strong candidate gene in this thesis. Similarly, the *COL12A1* exonic polymorphism rs970547 is not in LD with rs9350591 but has been reported to be associated with a predisposition to anterior cruciate ligament ruptures in a female cohort (Posthumus *et al.*, 2010; O'Connell *et al.*, 2015). Emerging evidence for the involvement of the two regions with skeletal development, together with the results presented in this thesis, suggest that the OA signals could ultimately prove to be linked with OA susceptibility during joint development rather than in end-stage diseased tissue.

6.2 Key Results

The overall aim of my Ph.D was to functionally dissect the two hip OA-associated loci that emerged from the arcOGEN GWAS (arcOGEN Consortium *et al.*, 2012). The key findings of this research are:

- despite the discovery of several significant differences between gene expressions stratified by disease state or joint site, in the end-stage disease tissue tested there were no correlations between the association SNP genotypes and the respective gene expressions at the two loci
- all genes were expressed during chondrogenesis and osteoblastogenesis
- the CpG dinucleotide cg26466508, 53 bp upstream of *FILIP1*, was hypermethylated in OA hip cartilage T (risk) allele carriers relative to major allele homozygotes
- the region of association marked by rs10492367 is functional in regulating enhancer activity, can significantly differ between alleles of the same SNP, and can differ depending on the tissue of origin of the cell
- transcription factors were found to bind all four SNPs that were carried forward for EMSA investigations into protein:DNA interactions, however this was often dependent on the origin of the nuclear extract
- ChIP confirmed an enrichment of TCF3, SUB1 (PC4) and RELA at rs10492367
- knockdown of SUB1 (PC4) in HACs resulted in a consistent and significant downregulation of *PTHLH*

The data presented here highlight the diverse nature both of signals that are generated from GWAS and of OA susceptibility. The particular results of this thesis will allow for subsequent research to be focussed on the effect of the OA association regions on joint development and

cell differentiation rather than end-stage diseased tissue. In addition, continued functional work to investigate the nature of protein binding to rs10492367 will be key in fully understanding this locus.

6.3 Assessing the Correlation of Gene Expressions and the Respective Association Single Nucleotide Polymorphism (SNP) Genotypes

In Chapter 3 and Chapter 4, I confirmed that there were no non-synonymous polymorphisms that could account for the association signals. I therefore hypothesised that rs10492367 and rs9350591 instead mark the actions of *cis*-eQTLs that modulate nearby gene expression, and thereby contribute to OA susceptibility. As *cis*-eQTLs can act on genes from up to 1 Mb from the transcription start site (Nica and Dermitzakis, 2013), the identification of the specific gene upon which the association signal acts often requires the analysis of several genes within and surrounding the disease-associated region. For example, in a study investigating susceptibility to childhood asthma, the expressions of a total of 14 genes within or near the association region at chromosome 17q21 were first considered before a positive correlation was identified between the transcript levels of *ORMDL3* (ORMDL sphingolipid biosynthesis regulator 3) and the disease-associated markers (Moffatt *et al.*, 2007). In addition, susceptibility variants within a 1.25 Mb region associated with Crohn's disease on chromosome 5p13.1 were found to contribute to the variance in expression of *PTGER4* (prostaglandin E receptor 4), a gene 270 kb away from the associated region (Libioulle *et al.*, 2007).

Through overall gene expression and AEI analysis, several functional studies have reported OA-associated signals to be marking eQTLs similarly acting on nearby genes (Reynard and Loughlin, 2013). This is exemplified at the 7q22 locus, a region of association in Dutch Caucasians marked by the polymorphisms rs3815148 (Kerkhof *et al.*, 2010) and rs4730250 (Evangelou *et al.*, 2011), whereby the OA-associated alleles correlate with a reduced expression of *HBP1* and *DUS4L* (Raine *et al.*, 2012). Moreover, at the 14q31.1 locus, AEI indicated that the OA-associated allele of rs225014 correlates with an increased transcript output of *DIO2* (Bos *et al.*, 2012). Finally, as previously discussed in Chapter 3, the risk allele of the hand OA-associated SNP within *ALDH1A2* was significantly associated with knee and hip OA (Styrkarsdottir *et al.*, 2014). To investigate my hypothesis, I similarly quantified the gene expressions in diseased and non-diseased hip cartilage; however there were no such correlations observed with the association SNP genotypes. This typifies the complexity of dissecting regions of OA association where investigations, such as that of the OA-associated

SNP rs2277831 and *MICAL3* expression (Ratnayake *et al.*, 2012), oftentimes do not yield positive results.

Interestingly, the expressions of *SENP6* and *MYO6* from the rs9350591 locus and *KLHL42* from the rs10492367 locus were all decreased in OA hip cartilage relative to NOF cartilage. This suggests that these genes might be required for normal cartilage homeostasis, irrespective of the association SNP genotypes: simply, a lack of the encoded proteins could contribute to the onset or progression of OA. While there is little published literature about *KLHL42*, and pathologies associated with *MYO6* include autosomal dominant hearing loss (Sanggaard *et al.*, 2008), *SENP6* is a more likely candidate gene. Firstly, the gene resides within the region of high LD with rs9350591, and secondly, its protein function has a potential link with OA susceptibility. Through sumoylation, the covalent attachment of SUMO proteins is a reversible regulator of target protein function. The protein encoded by *SENP6* is able to regulate sumoylation by deconjugating the SUMO proteins from their targets (Lima and Reverter, 2008). There is emerging evidence to suggest that sumoylation plays a role in arthritis (Yan *et al.*, 2010), and consequently, a decrease in *SENP6* expression might be a mechanism through which OA progresses.

A caveat to studying overall gene expression is that an eQTL might be tissue-specific, such as that detected at the 7q22 locus (Raine *et al.*, 2012). To overcome this, I included an analyses of fat pad and synovium samples to ensure an eQTL acting in these tissues was not overlooked. Similarly, I analysed OA knee cartilage to safeguard against false negative results. Again, this yielded no correlation with association SNP genotype. A further limitation of studying overall gene expression is that inter-individual differences could be masking true correlations. Indeed, AEI was observed at the Type 2 diabetes-associated locus at chromosome 1q21-24 for *IVL* (involucrin), *XCL1* (chemokine [C motif] ligand 1), and *TMCO1* (transmembrane and coiled-coil domains 1), yet overall gene expression did not correlate with association signal genotype (Mondal *et al.*, 2013). By similarly implementing AEI in Chapter 3 and Chapter 4, a more detailed assessment of gene expression can be made; however, for this study, AEI was not without its own limitations. Firstly, *FILIP1* and *COX7A2* were omitted from the analyses due to the lack of transcript polymorphisms; and secondly, the lack of informative OA hip patient samples meant a substantial *n* number was not achievable. Therefore, particularly for the rs9350591 6q14.1 locus, an eQTL operating in end-stage diseased tissues could have gone undetected.

With the knowledge that all genes are expressed during chondrogenesis and osteoblastogenesis, the most likely reason for why the actions of an eQTL were not detected is that the effects could instead be exerted at earlier stages of development. Perhaps the causal SNP is functional in OA susceptibility and progression rather than at the end-stage of the disease where a total joint replacement was necessary. In particular, *PTHLH* has a well-established role in endochondral ossification (Wysolmerski, 2012), and could therefore be regulated by an eQTL that predisposes an individual to developing OA later in life. Overall, both gene expression analyses and AEI failed to detect any correlation with the association SNP genotypes. It can be concluded that the OA associations marked by the polymorphisms rs10492367 and rs9350591 do not contribute to disease susceptibility by modulating the tested genes in end-stage diseased synovial joint tissue. Alternatively, the OA-associated signals could mark *trans*-eQTLs, modulating gene expression beyond the regions currently investigated (Fehrmann *et al.*, 2011), or eQTLs that act during joint development.

6.4 Differential Methylation as a Mechanism to Mediate Osteoarthritis Susceptibility

DNA methylation is a widely characterised mechanism through which epigenetics can regulate the human genome, and has been studied in a variety of diseases such as cancer (Esteller *et al.*, 2000), metabolic disorders (Nilsson *et al.*, 2014) and Alzheimer's disease (De Jager *et al.*, 2014). Until relatively recently, however, DNA methylation has been largely overlooked in the context of genome-wide OA susceptibility. From approximately 27,000 CpG sites, the first OA-specific genome-wide investigation into methylation identified 91 differentially methylated loci between the cartilage of donors with knee OA and healthy controls (Fernandez-Tajes *et al.*, 2014). A matter of months later, approximately 480,000 CpG sites were interrogated by Dr Michael Rushton, a member of Prof. Loughlin's research group (Institute of Cellular Medicine, Newcastle University), in the cartilage of OA hip, OA knee and healthy controls, identifying differentially methylated loci depending on both the disease state and the joint site (Rushton *et al.*, 2014). Prior to the publication of the latter paper, I accessed the raw data in order to conduct my own investigations into the methylation levels of a 2 Mb span at each of the two loci. Despite not detecting an eQTL in the end-stage diseased tissues tested, it was important that methylation as a regulatory mechanism in OA should still be considered. Firstly, *FILIP1* and *COX7A2* were excluded from the in-depth AEI investigations, meaning that although not detected, an eQTL could still be functional here. Moreover, differential methylation of the region may mark a historical event that is no longer functional, with the activity of regulatory regions known to act at specific stages of cellular differentiation (West *et al.*, 2014). For example, differential DNA methylation between young

adult rhesus macaques was associated with a response to events earlier in life (Provencal *et al.*, 2012). In addition, the methylation mark could be present and detected in cartilage, yet it might not be functional: perhaps the functionality is only apparent in the presence of tissue-specific transcription factors.

The identification of differentially methylated CpG sites dependent on the disease state and the joint site at both loci suggests, akin to the gene expression analyses, that methylation could be key in modulating OA susceptibility (den Hollander *et al.*, 2012; Bomer *et al.*, 2015). Crucially for the dissection of the OA signals, however, cg26466508 was hypermethylated in OA hip cartilage risk allele (T) carriers of rs9350951 relative to the non-risk CC homozygotes. As this CpG site is 53 bp upstream of *FILIP1* in a region predicted to have enhancer activity, its function in OA cartilage may have been missed since AEI was not studied for this gene. A mechanism of action of cg26466508 might be that relative to the non-risk allele, the T allele of rs9350591 creates a more open chromatin conformation such that DNA methyltransferase enzymes can more readily bind cg26466508 (Robertson *et al.*, 2004). The increased methylation of the CpG site could then affect transcription factor binding to the region, causing differential regulation of target gene expression. One mode of action is through the direct inhibition of transcription factor binding due to the altered binding site (Tate and Bird, 1993). An alternative mode of action is through the recruitment of histone deacetylase complexes which cause a change in chromatin structure, and thus modulates the accessibility of the binding site (Jones *et al.*, 1998). Of course, the causal SNP could be one in high LD with the association signal: in fact rs9343292 is in perfect LD with rs9350591 and resides less than 1 kb downstream of cg26466508. In the context of myogenic tumours, a CpG island surrounding the transcription start site of *PAX3* (paired box 3) was found to be hypermethylated in embryonal rhabdomyosarcomas (common in young children) relative to alveolar rhabdomyosarcomas (common in older children), while an inverse correlation existed between a particular site within this island and *PAX3* expression in both forms (Kurmasheva *et al.*, 2005). The authors postulated that the decreased gene expression is caused by inhibited transcription factor binding by the hypermethylation, which could thus result in abnormal development and tumorigenesis. It would be essential for subsequent investigations of the 6q14.1 region to similarly quantify *FILIP1* expression and correlate this with methylation levels at cg26466508.

For the methylation analysis of the CpG sites spanning the 12p11.22 locus, there were no significant correlations with rs10492367 genotype. As differential methylation can be tissue-

specific (Lokk *et al.*, 2014), performing a similar microarray investigation using alternative synovial joint tissues might reveal differentially methylated sites. As an additional interrogation of this region, I designed an undergraduate project to quantify methylation at CpG sites in close physical proximity to SNPs in high LD with rs10492367. Although methylation did not correlate with rs10492367 genotype, the CpG site at rs11049207 was hypermethylated in OA knee cartilage relative to NOF cartilage, while the CpG site at rs10743612 was hypomethylated in NOF cartilage relative to both OA hip and OA knee cartilage. These were intriguing significant findings that could be important in modulating the development of OA, irrespective of the association signal at this locus. Overall, however, this requires further investigation to both validate the assay and increase the *n* number.

6.5 Differential Transcription Factor Binding as a Mechanism to Mediate Osteoarthritis Susceptibility

The aim of Chapter 5 was to characterise the region of association marked by rs10492367 primarily by investigating its functionality. Although regulation of gene expression was not identified, it does not imply that the region is not functional. Differential transcription factor binding has previously been identified in the context of OA (Syddall *et al.*, 2013), and so I hypothesised that rs10492367 could function in a similar manner. It was necessary to shortlist the functional polymorphisms prior to performing EMSAs in order to focus on only the most likely causal variants. While the EMSAs indicated protein binding to the SNPs, none of the supershifts were allele-specific. An explanation for this could be that the EMSA probes are linear and do not provide a true representation of the natural chromatin conformation. For functionality, a particular haplotype or three-dimensional conformation might be required, which is lacking in this artificial environment (Knight, 2005). Moreover, the difference between the allelic probes of each SNP is one base pair – it is conceivable that, regardless of any *in vivo* allelic specificity of the transcription factors, the linear presentation of the probes are so similar that the protein could in fact bind both. Combined with ChIP, the EMSA results for rs10492367 are suggestive of transcription factor binding. Although the EMSA results imply the proteins do not differentially bind the alleles of the SNP, this might again be an artefact of the linear nature of the EMSA probes. The most compelling result was SUB1 (PC4), which when knocked-down in HACs caused a significant downregulation of *PTHLH*.

Currently, the connections between SUB1 (PC4) and OA, synovial joint development and joint homeostasis are limited. Rather, it has been more widely studied as a general transcriptional coactivator (Conesa and Acker, 2010). SUB1 (PC4) has been reported to

activate non-homologous end joining of double-stranded DNA breaks (Batta *et al.*, 2009), damage that is commonly caused by ionising radiation during cancer therapy. Knockdown of SUB1 (PC4) in oesophageal squamous cell carcinoma cells increased the sensitivity of the cells to radiation, meaning that the presence of the protein might be associated with a poorer response of the patient to radiotherapy (Qian *et al.*, 2014). The protein is also essential for embryogenesis, with SUB1 (PC4) knockout mice being lethal despite normal development until E5.5 (Li, 2010). More relevant to OA, SUB1 (PC4) has been reported to be functional in chromatin organisation. The transcription factor interacts with histones H3 and H2B to condense the chromatin, while its knockdown leads to cell cycle arrest (Das *et al.*, 2006). As discussed in Chapter 5, SUB1 (PC4) could be central in the formation of a larger complex, with other unidentified proteins mediating allele-specific differences at rs10492367. Similarly, the protein was identified as a component of a multi-protein complex that binds the OA-associated SNP rs143383 (Syddall *et al.*, 2013). Together with its regulation of *PTHLH*, the wide-ranging function and expression of SUB1 (PC4) means that this has the potential to be equally important in the mediation of OA susceptibility.

Understanding a link between transcription factor binding and the association signal is vital if this is to be exploited through therapeutic interventions. I propose that SUB1 (PC4), in combination with other unidentified transcription factors, binds to the non-risk (G) allele of rs10492367 *in vivo*, increasing *PTHLH* expression and thus stimulating chondrocyte proliferation and inhibiting hypertrophy. This regulation could be lacking with the risk (T) allele, and so accelerated terminal differentiation of chondrocytes, a hallmark of OA (van der Kraan and van den Berg, 2012), would be induced. The fact that SUB1 (PC4) is widely-expressed means that targeting this in the treatment or prevention of OA would be difficult, particularly as off-target effects must be considered. Increasing the amount of SUB1 (PC4) within a patient that is susceptible to developing OA could be achieved by either administering exogenous protein (Gil-Bernabe *et al.*, 2012) or by stimulating its endogenous expression by targeting upstream effectors in its metabolic pathway. Should it be the case that the actions of SUB1 (PC4) need to be inhibited, several established mechanisms could be explored (Gambari, 2011). For example, the expression of the gene encoding the protein could be targeted using an antisense oligonucleotide (AON). The purpose of an AON is to induce exon skipping within the gene transcript, which overall achieves a downregulation of the full length protein (Kemaladewi *et al.*, 2014). Alternatively, exogenous mutated protein with a dominant negative effect can be introduced to effectively override the native protein function (Attardi *et al.*, 1993). Moreover, double-stranded oligonucleotides can be used as

decoys by mimicking the target sequence and therefore competing directly for binding to the transcription factor (Mann, 2005). Overall, the functional findings of this chapter illustrates how diverse the region is and highlights the huge potential for its modulation in targeted therapies.

6.6 Future Work

By tailoring the experiments to be more specific for the loci under investigation, the existing data could be made considerably more robust. Firstly, a basic enhancement in Chapter 3 and Chapter 4 would be to increase the number of hip cartilage samples for the gene expression, AEI and methylation analyses. Of course, validating the assays used in the undergraduate project to investigate differential methylation is essential in order to draw any reliable conclusions from the data. Additionally for the 6q14.1 locus, it would be prudent to explore functionality by following the same route of investigation as was performed for rs10492367 in Chapter 5. The associated volume of work, however, meant that this was not possible in the timeframe of my Ph.D. For both the loci, it would then be worthwhile performing additional luciferase reporter assays using HAC cell cultures, although through personal communication with colleagues, I was advised that such transfections would be difficult to achieve (Syddall, 2013). Likewise, the EMSA and ChIP experiments of Chapter 5 could be improved by using nuclear protein extracted specifically from hip cartilage. The probes that were used for the EMSAs were a standard length and were considered a suitable size to capture transcription factors binding directly over the polymorphism. However, this could limit the experiments and exclude proteins that bind over larger distances. Thus increasing the length of the probes, although perhaps capturing non-specific binding, might allow additional proteins to be identified. Finally, a limitation of using SW1353 and U2OS cell lines for the luciferase reporter assays, EMSAs and ChIP is that both cell lines are homozygous for the major allele of rs10492367. It is possible that transcription factors that bind the risk allele are less abundant, or even absent, in these cell cultures and so a true representation of functionality is not provided. A way to correct this would be to use a heterozygous chondrosarcoma or osteosarcoma cell line, however there were no such cell lines available in our research group. The clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system is a technique that uses customised nucleases to target and edit specific regions of the genome and could be utilised to generate the necessary genotype of the association SNP (Sander and Joung, 2014).

The findings discussed in this thesis provide a basis for further dissecting the association loci. Several additional experiments would be vital in achieving this, particularly for the functional analysis of the rs10492367 locus in Chapter 5. Firstly, despite the luciferase reporter assays allowing a shortlist of functional variants to be identified, this differed depending on the cell line under investigation. Therefore, a thorough investigation into protein binding to all of the functional SNPs, by performing EMSAs and ChIP, would ensure that no variants were missed. Moreover, EMSAs and ChIP are two different techniques that ultimately investigate the same principle, with mass spectrometry being another complementary method that could be employed (Glish and Vachet, 2003). Here, oligonucleotide pull-down assays – whereby allele-specific regions surrounding rs10492367 would be amplified with biotinylated primers, incubated with nuclear protein and separated by electrophoresis – could precede mass spectrometry which would then identify the proteins binding the DNA (Syddall, 2013; Syddall *et al.*, 2013).

If functionality of the association signal could be identified, the next step of investigation would be to identify at what stage of disease susceptibility this occurs. It is likely that in end-stage diseased cartilage, the region is not exerting its maximum functionality given that the disease is so far advanced. Instead, the region could be functional during joint development, and so a strong approach to investigate this would be to quantify gene expression and AEI during chondrogenesis and osteoblastogenesis. Furthermore, to develop the identification of SUB1 (PC4) binding to rs10492367, knocking-down this protein during MSC differentiation would allow its function during joint development, specifically on *PTHLH* expression, to be investigated. In addition, combination knockdowns to include RELA and TCF3 would show whether the proteins act in isolation or as a larger protein complex. Despite showing that the transcription factors were expressed in the cell lines of interest, the levels of expression were often low. As such, a similar set of experiments to the knockdowns could be performed, but instead the transcription factors could be overexpressed in HACs and MSCs. It is imperative that the focus should be on understanding the link between SUB1 (PC4) binding to rs10492367, the knockdown of SUB1 (PC4) causing a downregulation of *PTHLH*, and the role of *PTHLH* in OA susceptibility.

The vital limitation of the experiments discussed in this thesis is that the entire premise is based on the hypothesis that the association signals mark *cis*-eQTLs. Therefore, the present investigations have been limited to only the regions surrounding the signals. An alternative methodology could be to utilise a gene expression array, where gene-specific probes are

hybridised to a solid support to capture nucleic acids in a sample preparation, giving a snapshot of the transcriptome composition (Macgregor and Squire, 2002). Additionally, rather than being targeted only to known genes, RNA-Seq involves the ligation of adaptors to all fragments of a cDNA library before high-throughput sequencing is performed (Wang *et al.*, 2009). Both technologies could enhance the present study, and performing subsequent eQTL investigations could be useful in further investigations. Moreover, employing chromosome conformation capture (3C) in relevant cell types such as HACs would elucidate the *in vivo* physical conformation of the chromatin and identify whether the regions interact with other genomic loci (Gavrilov *et al.*, 2009). From this, the investigation could be directed towards more appropriate genes. Finally, if evidence implies that the functional polymorphism acts in combination with other SNPs, it would be important to assess the effects of haplotypes rather than the SNPs in isolation as has so far been done.

6.7 Summary

In summary, I have fulfilled the aim of my Ph.D, which was to functionally analyse the OA association signals marked by the polymorphisms rs10492367 and rs9350591. Although the functionality of the regions could not be correlated with the respective OA association signal genotypes, the research presented here provides a foundation for the understanding of these loci. The polymorphisms do not mark the actions of *cis*-eQTLs in end-stage diseased tissue, but more likely, the OA associations are modulated at an earlier stage of development or in different synovial joint tissues.

Appendix A. Primer Sequences

SNP	Forward primer (5′ - 3′)	Reverse primer (5′ - 3′)	Sequencing primer (5′ - 3′)	$MgCl_2$ (mM)	Anneal (^oC)
rs6253	[Btn]CGTCACCCAACATCAATCC	GGTGTGAGAGTAAGGGGAAGAA	TGAGAGTAAGGGGAAGAA	2.0	55
rs9029	[Btn]AGCTGTCCCTGCAATGTTTTC	GGGTATTTCATAGGCATTCTTGTA	ATGTTTACACTTTTAAGCTA	2.0	64
rs594012	ACGCGATATTGCAGCACA	[Btn]ACACCTAACAGCCTCGATGTTC	CAGCACAGGTCCTGG	2.0	55
rs240736	[Btn]CTTGTGTTGTTAAGATAGGCA	TGAACCCCAACACCATCTAT	AGTCAGAAAGTGATGACCTG	2.0	55
rs41269315	[Btn]CAGGAGAGGTAGCAGGCTGAAC	ACTAGTGAGGAGGAGCAGCTGAG	GAGAGATAGGGTCAGTGAA	2.0	64
rs1045758	CCGCTGTAATTCCCAAAACTC	[Btn]ACTGCAAAGGAGGAAAAAAATAGA	CACTAATGTTTTGGTCTGAA	2.0	55
rs699186	[Btn]GGCATATTCTGATGTTTCTCATCC	AGTGCCTTGATCATTTTAAGTGGT	GTACTGTGCCATCCTTAA	2.0	55
rs71561434	CAAAGGCCTCCGCTGATG	[Btn]GCCCCTAGCTCGTTCTGCA	CACGCCTGGGCGGGGT	1.0	62
rs17414687	[Btn]CACCATCCACTGGAAAAGTAGAA	TTCTGGAATGCTTCGTAGTTCA	TCTGCAAGTATTTTCATTTA	1.0	60

SNP	Forward primer (5′ - 3′)	Reverse primer (5′ - 3′)	Restriction enzyme	$MgCl_2$ (mM)	Anneal (^oC)
rs9350591	CATAAGAAAGGCATGTTGC	CAGCTTTCATTGTATAACAAC	*Msp*I	2.0	55
rs11049204	GCTCTTAACCATCTTTCAACC	CATATAGATTTACAAAAGGC	*Bsr*I	2.0	55

Gene	Forward primer (5′ - 3′)	Reverse primer (5′ - 3′)	cDNA/gDNA fragment size (bp)	$MgCl_2$ (mM)	Anneal (^oC)
HBP1	TCGAAGAGTGAACCAGCCTT	GAAGGCCAGGAATTGCACCATCC	152/570	2.0	55

Table A.1. Primer sequences and conditions used for PCR, restriction digest and pyrosequencing. Top panel: primer sequences used for genotyping and allelic expression imbalance analysis by pyrosequencing. Middle panel: primer sequences and restriction enzymes used for genotyping by restriction fragment length polymorphism analysis. Bottom panel: primer sequences used for validation of cDNA synthesis.

SNP	Forward primer (5′ - 3′)	Reverse primer (5′ - 3′)	$MgCl_2$ (mM)	Anneal (^oC)
rs10492367	GGGGACGCGTATTTTAAACTGCTGGTTCCCACT	GGGGAGATCTTTCCATTGACTTCTTAACCCAAA	1.5	55
rs11049206	GGGGACGCGTACTAAAATGCCAGTTTAACCACC	GGGGAGATCTCAAATCCTCATGAGCTGTCCTG	2.0	70
rs11049204	GGGGACGCGTTGTTCCATTTCTGTTGTTACCCC	GGGGAGATCTAGGGAAGCTGTCATAAGGAAGAT	2.0	66
rs10843013	GGGGACGCGTGTGTTGTTGGATAGAGCAGTACC	AACCAAAAGCTAGATCTGCCTTT	2.0	66
rs10743612	GGGGACGCGTACTGAGTTAGGTGATCAGAGTGG	GGGGAGATCTACTGGCATTTTAGTTATGAGCCA	1.5	70
rs61916489	GGGGACGCGTGAATCACCTCCCCAGTAAACAAC	GGGGAGATCTCAAGTCACTATGGGAAGGAACAC	2.0	66
rs11049207	GGGGACGCGTCCTTGAACTACCCAAAGATGGTA	GGGGAGATCTGAATCAAAACCTTCTTCCCTGTT	1.5	55
rs79881709	GGGGACGCGTTAAAGTGCAAACAGGGAAGAAGG	GGGGAGATCTGCCTGTGGTTCCATCAAATCTTA	2.0	66
rs58649696	GGGGACGCGTGCAGACATTCTTACTCATCCAGC	GGGGAGATCTCTCCCCTCACATTACCTTTGTTC	2.0	66

SNP	Forward primer (5′ - 3′)	Reverse primer (5′ - 3′)
rs58649696	CATCTGCCTCTTTCATTTTCCTAATGGGAC	GTCCCATTAGGAAAATGAAAGAGGCAGATG
rs11049204	CATTTTAAAAAATGAAACTGGATAGAAAGATCATAG	CTATGATCTTTCTATCCAGTTTCATTTTTTAAAATG
rs79881709	GGTTTCCTGAACTGGAAGAATCTCACTCTCCAAC	GTTGGAGAGTGAGATTCTTCCAGTTCAGGAAACC

Table A.2. Primer sequences used for the cloning of fragments into pGL3 promoter vector constructs. Top panel: PCR primers and conditions used for standard cloning into pGL3 vectors between the restriction sites *Mlu*I and *Bgl*II; the restriction sites are underlined in red text preceded by a 5′ GGGG. The rs10843013 fragment contained a naturally-occurring *Bgl*II restriction site. Bottom panel: primer sequences used for site-directed mutagenesis.

Gene	Probe sequence (5′-3′)	Primer 1 (5′-3′)	Primer 2 (5′-3′)
18S	56-FAM/TCCTTTGGTCGCTCGCTCCTCTCCC/TAMRA	TATTAGCTCTAGAATTACCACAGTTATCC	CGAATGGCTCATTAAATCAGTTATGG
COL12A1	56-FAM/TCCCCTGTG/ZEN/GAAGGCTGATAAGTGA/3IABkFQ	CGGTGATAGTGTAAGGAGTGTC	TGGTCGTGTGCAGAAATATAGG
COX7A2	56-FAM/AGATTGGGC/ZEN/AGAGGACGATAAGCAC/3IABkFQ	TGGTCAGTAACAGCCAAGATG	TTTTAAAATGCCTGCGGGAAG
FILIP1	56-FAM/ACAACGTCA/ZEN/TCTGCTCGAGGAACC/3IABkFQ	AGCACTATCACCATAACACCG	CTTTTGACATAGGAATGCGGG
GAPDH	56-FAM/AAGGTCGGAGTCAACGGATTTGGTC/IABkFQ/36-TAMSp	TGTAGTTGAGGTCAATGAAGGG	ACATCGCTCAGACACCATG
HPRT1	56-FAM/AGGACTGAACGTCTTGCTCGAGATG/36-TAMSp	ACAGAGGGCTACAATGTGATG	TGCTGAGGATTTGGAAAGGG
KLHL42	56-FAM/CGCCACAAA/ZEN/ATTCCACGCATCCT/3IABkFQ	TGCAAAATGTTCATGTTCCGG	TTGAGTGTTACAACCCCGAG
MYO6	56-FAM/TGGCGTCCT/ZEN/GCACCTTGGAAATA/3IABkFQ	CCAAACCCAGTAATTCAGCAC	AAAGCTTGATCTCTTCCGGG
PTHLH	56-FAM/CCGCCTCAA/ZEN/AAGAGCTGTGTCTGAA/3IABkFQ	CAGCGGAGACTGGTTCAG	ATGGACTTCCCCTTGTCATG
RELA	56-FAM/CCCAACACT/ZEN/GCCGAGCTCAAGAT/3IABkFQ	ACCTCAATGTCCTCTTTCTGC	CCTGTCCTTTCTCATCCCATC
SENP6	56-FAM/CTGTAAGGT/ZEN/TAAGTCGGCTCCAAGGT/3IABkFQ	TCCTCTTAATTTCAGGCTCCAC	AAAGAATACCCACCTCATGTCC
SUB1	56-FAM/CGAAGCGAT/ZEN/GCCTAAATCAAAGGAACTT/3IABkFQ	GTCAGAATCACTGCCAGAAGAG	CGAGCGAACGACCAAGAG
TCF3	56-FAM/TCCTGGACT/ZEN/TCAGCATGATGTTCCC/3IABkFQ	GTCCTCAAGACCTGAACCTC	GCACAGACAAGGAGCTCAG
TMEM30A	56-FAM/TGTGACATT/ZEN/CAAAGAGTATCGGCCAGC/3IABkFQ	CAGCATTACCTACTTTTCGCAAG	TCATCCGTTTTCGTCCATCAA

Locus	Probe sequence (5′-3′)	Primer 1 (5′-3′)
rs10492367	AGCTCAAACTGATGAAACCATGT	CAGCTGTTGGTTGTCACACT
GAPDH	CGTAGCTCAGGCCTCAAGAC	GCTGCGGGCTCAATTTATAG
PRM2	GGCGCAGACACTGCTCTC	CCTTCTGCAGGAGCGATG

Table A.3. Primer and probe sequences used for qPCR. Top panel: primer and probe sequences used for quantification of gene expression. Bottom panel: primer sequences used for DNA quantification in chromatin immunoprecipitation.

SNP	Forward primer (5′ - 3′)	Reverse primer (5′ - 3′)	Sequencing primer (5′ - 3′)	Anneal (^{o}C)
rs11049207	TGTGTGTATATGTTTGGTGTT	[Btn]TCCAACAAAAATACACTATAA	GTGTTTGTGTGTGTG	56
rs10743612	GGTATTTGGATATTTATTTG	[Btn]AACATTTTAATTATAAACCA	GGGTTTTTTGTTAT	56

SNP	Forward primer (5′ - 3′)	Reverse primer (5′ - 3′)	$MgCl_2$ (mM)	Anneal (^{o}C)
rs11049207	GGGGACTAGTCCTTGAACTACCCAAAGATGGTA	GGGGCTGCAGGAATCAAAACCTTCTTCCCTGTT	1.5	55
rs10743612	GGGGACTAGTACTGAGTTAGGTGATCAGAGTGG	GGGGCTGCAGACTGGCATTTTAGTTATGAGCCA	1.5	55

Table A.4. Primer sequences and conditions used for PCR, methylation quantification and the cloning of fragments into pCpGL-basic/EF_1 vectors. Top panel: primer sequences used for methylation quantification by pyrosequencing; primers target bisulfite converted DNA. Bottom panel: PCR primers and conditions used for standard cloning into pCpGL-basic/EF_1 vectors between the restriction sites *Pst*I and *Spe*I; the restriction sites are underlined in red text preceded by a 5′ GGGG.

Appendix B. Pyrosequencing Validations

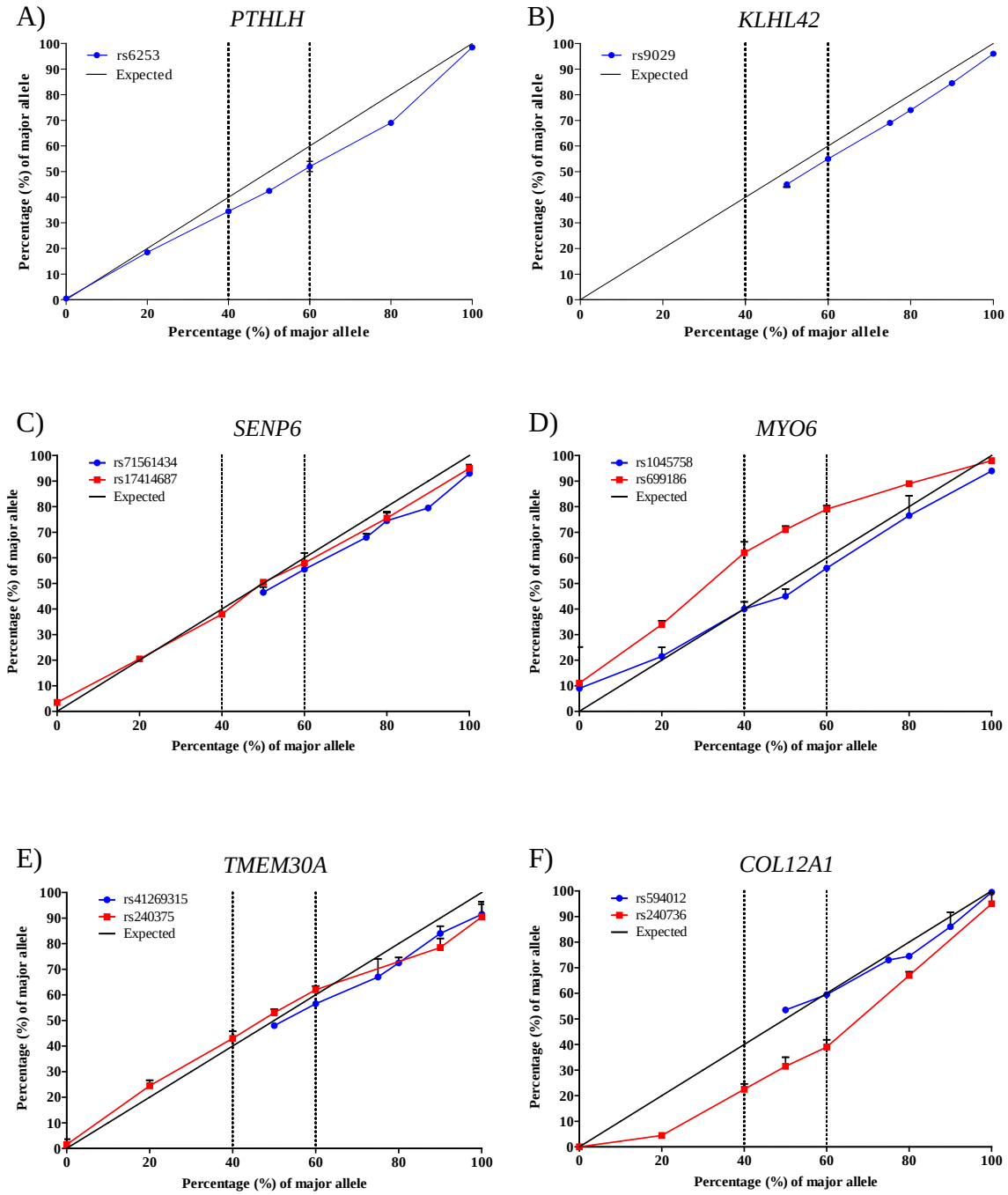

Figure B.1. Validation of pyrosequencing assays used for genotyping and allelic expression imbalance. Known artificial allelic ratios were generated by mixing the PCR products of major allele homozygotes with minor allele homozygotes (or heterozygotes in the absence of minor allele homozygotes) in duplicate for A) *PTHLH*, B) *KLHL42*, C) *SENP6*, D) *MYO6*, E) *TMEM30A* and F) *COL12A1*. For validations that used heterozygotes, the plots begin at 50%. The allelic ratios, as determined by the pyrosequencer, were compared to the expected ratios. All assays were considered suitable for use. The detected ratios closely mirrored the expected outcome, with a positive correlation particularly between 40% and 60% of the major allele (dotted lines) where allelic imbalance is likely to be observed. Error bars represent the SEM.

Appendix C. Reporter Vector Constructs

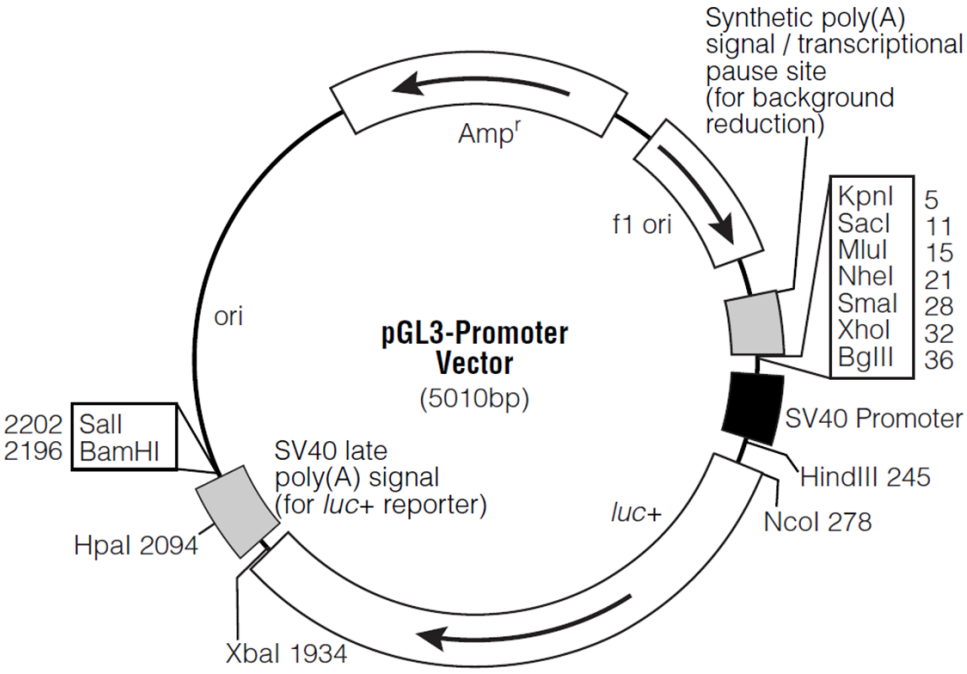

Figure C.1. Map of the pGL3-promoter vector. The vector contains an ampicillin resistance gene for selection, a multiple cloning site for fragment ligation, an SV40 promoter to initiate transcription and a luciferase gene for use as a reporter of enhancer activity. Image taken from the Promega Corporation website (http://www.promega.co.uk/).

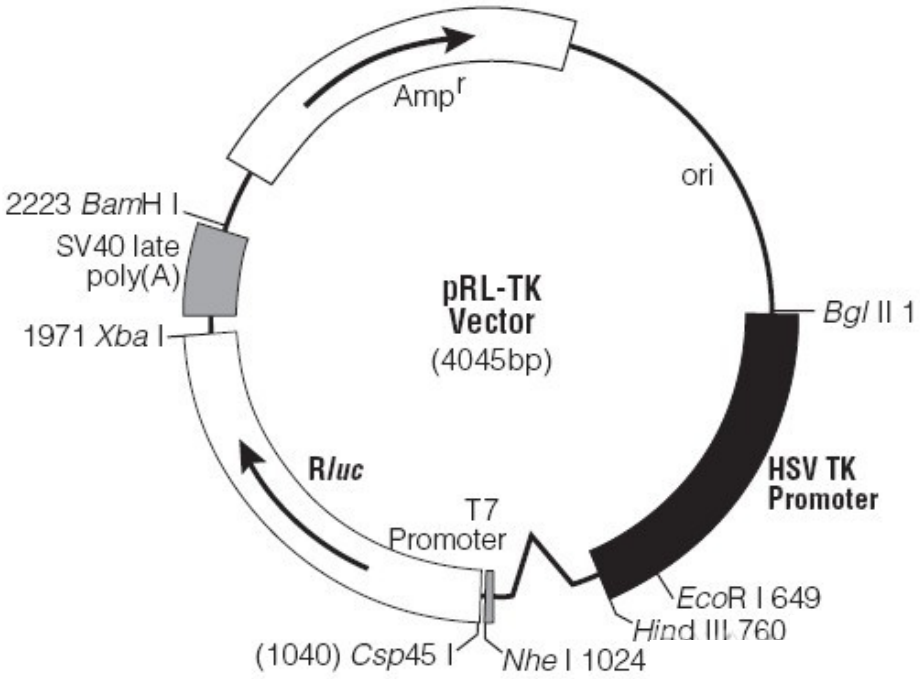

Figure C.2. Map of the pRL-TK vector. The vector contains an ampicillin resistance gene for selection, an HSV TK promoter to initiate transcription and a *Renilla* luciferase gene for use as a reporter of enhancer activity. Image taken from the Promega Corporation website (http://www.promega.co.uk/).

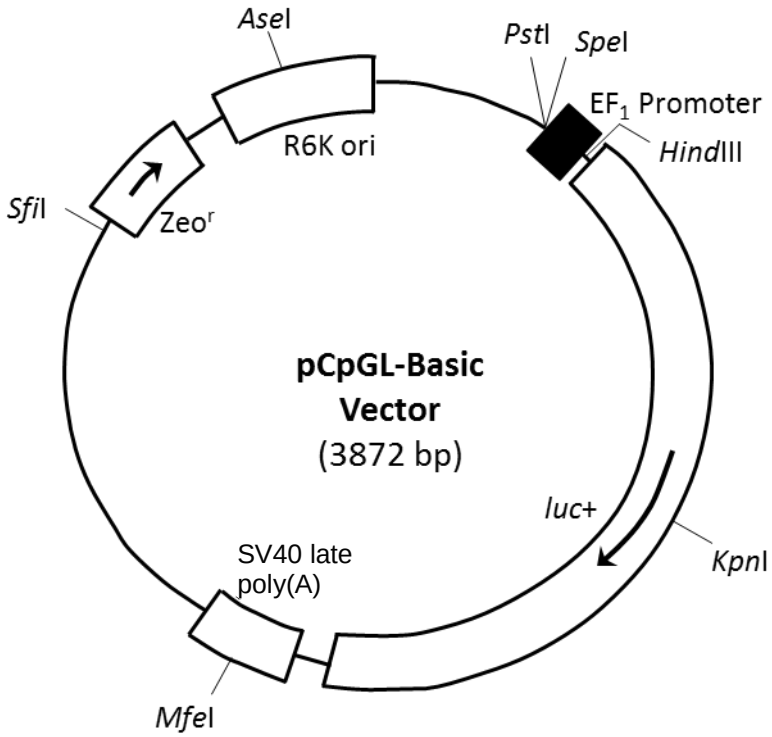

Figure C.3. Map of the pCpGL-basic/EF_1 vector. The vector backbone is CpG-free. The construct contains a zeocin resistance gene for selection, a cloning site for fragment ligation, an EF_1 promoter to initiate transcription and a luciferase gene for use as a reporter of enhancer activity. Image adapted from (Klug and Rehli, 2006).

Appendix D. Patient Details and Genotypes

Donor	Sex	Age (years)	Joint	rs9350591	rs10492367	S1	S2	M1	M2	T1	T2	C1	C2	P1	K1
1	F	75	H	CC	GG										
2	F	67	K	CT	GT	CT		AG		GA		TT		CT	GC
3	M	69	K	CC	GT										
4	F	77	H	CT	GT										
5	F	73	K	CC	GT	CC		AA		GG		TA		CC	GC
6	M	71	K	CC	GG										
7	F	73	K	CC	GG										
8	F	51	H	CT	GT										
9	M	71	H	CC	GT										
10	F	60	K	CC	GG	CC		AA		GG		TT		CC	GC
11	F	76	K	CT	GT										
12	F	78	H	CT	GG	CC		AA		GG		TT		TT	GG
13	M	74	K	CC	GT	CC		AG		GG		TT		CC	GC
14	F	83	H	CT	GG	CC		AG		GG		TT		CC	GC
15	M	50	K	CC	TT	CC		AA		GG		TT		CC	GC
16	F	76	H	TT	TT										
17	F	71	K	CT	GG										
18	F	67	H	CC	GT										
19	F	70	H	CC	GT										

Table D.1.Characteristics and genotypes of all donors whose tissues have been used. Transcript SNPs are named S1 (rs71561434), S2 (rs17414687), M1 (rs1045758), M2 (rs699186), T1 (rs41269315), T2 (rs240375), C1 (rs594012), C2 (rs240736), P1 (rs6253) and K1 (rs9029). Donors 1-230 OA, donors 231-275 NOF. Age (age at joint replacement); rs9350591and rs10492367 (association SNPs). The donors marked with an asterisk (*) were young, non-OA donors purchased from Lonza, UK. Continued overleaf.

Donor	Sex	Age (years)	Joint	rs9350591	rs10492367	S1	S2	M1	M2	T1	T2	C1	C2	P1	K1
20	F	70	H	CC	GG										
21	F	60	H	CC	TT										
22	F	78	H	CT	GG										
23	F	81	K	CT	GG										
24	M	76	K	CC	GG	CC	GA	AA	TC	GG	GG	TA		TT	GG
25	M	82	K	CT	GT	CT		AG		GG		TT		CC	GG
26	M	63	K	CT	GT										
27	M	57	K	CC	GG	CC	GG	AG		GG	AA	TT	TT	CC	GC
28	M	57	H	CC	GG										
29	M	69	K	CC	GT	CC	AA	AA	TT	GG	GA	TT	TT	CC	GC
30	M	74	K	CC	GT	CC		AA		GG		TT		CT	GC
31	F	67	K	CT	GG	CC	GA	AG		GA		TT	TT	CC	GG
32	F	69	K	CC	GG										
33	M	57	K	CT	GG										
34	F	61	H	CC	TT	CC		AA		GG		TA		CT	GG
35	F	60	K	CC	GG	CC		AA		GG		TT	TT	CC	GC
36	F	66	K	CT	GG	CC		AA		GG		TT		CT	GG
37	M	63	K	CC	GG										
38	M	77	K	CT	GG	CT		AG		GG		TT	TC	CT	CC
39	M	82	K	CT	GT	CC		AA		GG		TT		TT	GG
40	F	78	K	CC	GG								TT		

Table D.1. Characteristics and genotypes of all donors whose tissues have been used. Transcript SNPs are named S1 (rs71561434), S2 (rs17414687), M1 (rs1045758), M2 (rs699186), T1 (rs41269315), T2 (rs240375), C1 (rs594012), C2 (rs240736), P1 (rs6253) and K1 (rs9029). Donors 1-230 OA, donors 231-275 NOF. Age (age at joint replacement); rs9350591and rs10492367 (association SNPs). The donors marked with an asterisk (*) were young, non-OA donors purchased from Lonza, UK. Continued overleaf.

Donor	Sex	Age (years)	Joint	rs9350591	rs10492367	S1	S2	M1	M2	T1	T2	C1	C2	P1	K1
41	M	82	K	CC	GG										
42	M	56	K	CC	GG					GG			TT		
43	M	56	K	CC	GG										
44	F	54	K	CC	GT	CC		AA		GG		TA		CC	GG
45	M	71	K	CC	GT								TC		
46	F	58	H	CC	GT										
47	F	69	H	CC	GG										
48	M	63	K	CT	GT	CC		AG		GA		TT		CT	GG
49	F	71	H	CT	GT										
50	M	70	K	CC	GG	CC		AA		GG		TT		CC	GC
51	M	67	K	CC	GG	CC		AA		GG		TT		CT	GC
52	F	61	H	CC	GG										
53	M	86	K	CC	GT	CC	GG	AA	TC	GG	GG	TT	TC	CT	GC
54	F	80	K	TT	GG	CC		AA		GG		TT	TT	CT	GG
55	F	67	K	CC	GG	CC		AA		GG		TT		CT	GG
56	M	71	K	CT	GG										
57	F	46	K	CC	GT										
58	F	62	K	CC	GT										
59	F	67	K	CC	GT	CC		AA		GG		TA		CT	GG
60	F	58	K	CC	GG										

Table D.1. Characteristics and genotypes of all donors whose tissues have been used. Transcript SNPs are named S1 (rs71561434), S2 (rs17414687), M1 (rs1045758), M2 (rs699186), T1 (rs41269315), T2 (rs240375), C1 (rs594012), C2 (rs240736), P1 (rs6253) and K1 (rs9029). Donors 1-230 OA, donors 231-275 NOF. Age (age at joint replacement); rs9350591and rs10492367 (association SNPs). The donors marked with an asterisk (*) were young, non-OA donors purchased from Lonza, UK. Continued overleaf.

Donor	Sex	Age (years)	Joint	rs9350591	rs10492367	S1	S2	M1	M2	T1	T2	C1	C2	P1	K1
61	M	59	K	CT	GG										
62	M	64	K	CT	GG								CC		
63	F	81	K	CT	GG	CT		AG		GG		TA		TT	GC
64	F	80	K	CC	GG	CC		AG		GG		TT	TT	CC	GG
65	F	55	K	CT	GG	CC		AG		GG		TT		CT	GG
66	F	64	K	CC	GG	CC		AA		GA		AA	TC	CT	GC
67	F	78	K	CC	GG										
68	F	61	K	CC	GT	CC		AA		GG		TT		CC	GG
69	F	80	K	CT	GT	CC		AA		GA		TT		CC	GC
70	F	80	K	CT	TT										
71	F	59	K	CC	GT	CC		AA		GG		TT	TC	CC	GC
72	F	71	H	CT	GT					GG					
73	M	74	K	CC	GT	CC		AA		GG		TA		CC	GC
74	F	74	H	CT	GG										
75	M	72	K	CC	GG	CC		AA		GG		TT		CC	GG
76	M	72	K	CC	GT								TT		
77	M	70	H	CC	GT					GG					
78	F	72	H	CT	GG										
79	F	63	H	CC	GG	CC		AA		GG		TT		CC	GC
80	M	77	K	CC	GT	CC		AA		GG		TT		CC	GG

Table D.1. Characteristics and genotypes of all donors whose tissues have been used. Transcript SNPs are named S1 (rs71561434), S2 (rs17414687), M1 (rs1045758), M2 (rs699186), T1 (rs41269315), T2 (rs240375), C1 (rs594012), C2 (rs240736), P1 (rs6253) and K1 (rs9029). Donors 1-230 OA, donors 231-275 NOF. Age (age at joint replacement); rs9350591and rs10492367 (association SNPs). The donors marked with an asterisk (*) were young, non-OA donors purchased from Lonza, UK. Continued overleaf.

Donor	Sex	Age (years)	Joint	rs9350591	rs10492367	S1	S2	M1	M2	T1	T2	C1	C2	P1	K1
81	M	71	H	CC	GG	CC		AA		GG		TA		CT	GC
82	M	71	H	CT	GT	CT		AG		GG		TA		CC	GG
83	F	73	K	CC	GT	CC		AA		GG		TA		CT	GC
84	F	82	K	CT	GG								TT		
85	M	71	K	CC	GG	CC		AA		GG		TT		CC	GG
86	M	80	K	CC	GG	CC		AG		GG		TT		CT	GC
87	M	66	H	CC	GG	CC		AA		GG		TT	TT	CT	GG
88	M	69	K	CC	GG	CC		AA		GG		TT		CT	GC
89	F	58	K	CC	GG	CC		AA		GG		TT		TT	GC
90	F	60	K	CC	GG	CC		AA		GG		TT	TT	CC	GC
91	F	62	K	CC	TT	CC		AA		GG		TA		CT	GG
92	F	67	H	CT	GT	CC		AA		GG		TT		CC	GC
93	F	51	H		GG										
94	M	53	K	CC	GG	CC		AA		GG		TT	TC	CT	GG
95	M	65	K	CC	GG	CC		AG		GG		TT		CC	CC
96	M	85	H	CC	GG	CC	GA	AA	TC	GG	GA	TT	TC	TT	GC
97	F	79	K	CC	GT	CC		AA		GG		TT		CC	GG
98	F	88	K	CC	GG	CC	GG	AA	CC	GG	AA	TT	TT	TT	GC
99	M	65	K	CC	GT	CC		AA		GG		TT		CT	GG
100	F	76	H	CC	GG	CC		AA		GG		TT		CT	GC

Table D.1. Characteristics and genotypes of all donors whose tissues have been used. Transcript SNPs are named S1 (rs71561434), S2 (rs17414687), M1 (rs1045758), M2 (rs699186), T1 (rs41269315), T2 (rs240375), C1 (rs594012), C2 (rs240736), P1 (rs6253) and K1 (rs9029). Donors 1-230 OA, donors 231-275 NOF. Age (age at joint replacement); rs9350591and rs10492367 (association SNPs). The donors marked with an asterisk (*) were young, non-OA donors purchased from Lonza, UK. Continued overleaf.

Donor	Sex	Age (years)	Joint	rs9350591	rs10492367	S1	S2	M1	M2	T1	T2	C1	C2	P1	K1
101	F	68	H	CC	GG	CC	GG	GG	TT	GG	GA	TT	TT	CT	GG
102	M	69	K	CC	GG	CC		AA		GG		TT		CC	GC
103	M	63	K	CT	GG	CC		AA		GG		TT	CC	CC	CC
104	F	76	H	CC	GG										
105	F	71	K	CC	GG	CC		AA		GA		TT		CT	CC
106	M	49	K	CT	GG	CT		GG	TT	GG	GA	TT	TT	CC	GG
107	F	61	K	CC	GG	CC		AA		GG		TT	TT	TT	GG
108	F	64	K	CC	GG										
109	F	62	H	CC	GT	CC	GG	AA	TC	GG	GG	TT	TC	CT	GG
110	M	83	K	CC	GG										
111	M	67	K	CC	GG	CC	AA	AA	TC	GG	GG	TT	TT	CT	GG
112	F	60	K	CC	GG										
113	F	64	K	CC	GG	CC	GA	AA	TC	GG	GG	TT	TT	CT	GG
114	M	59	H	CC	GT	CC		AA		GG		TT	TT	CC	GG
115	M	65	K	CC	GG	CC		AA		GA		TT	TT	CT	GC
116	M	68	K	CC	GG	CC	GG	AA	CC	GG	AA	TT	TT	CT	GG
117	F	72	K	CT	GT	CT		AG		GG		TT		TT	GG
118	F	70	K	CC	GG	CC		AA		GG		TT		CC	GG
119	M	85	K	CC	GG	CC		AA		GG		AA	TT	TT	GC
120	F	54	K	CT	GT	CT		AA		GG		TT		CC	GG

Table D.1. Characteristics and genotypes of all donors whose tissues have been used. Transcript SNPs are named S1 (rs71561434), S2 (rs17414687), M1 (rs1045758), M2 (rs699186), T1 (rs41269315), T2 (rs240375), C1 (rs594012), C2 (rs240736), P1 (rs6253) and K1 (rs9029). Donors 1-230 OA, donors 231-275 NOF. Age (age at joint replacement); rs9350591and rs10492367 (association SNPs). The donors marked with an asterisk (*) were young, non-OA donors purchased from Lonza, UK. Continued overleaf.

Donor	Sex	Age (years)	Joint	rs9350591	rs10492367	S1	S2	M1	M2	T1	T2	C1	C2	P1	K1
121	F	76	K	CC	GG	CC		AA		GG		TT		CC	GG
122	F	72	K	CT	GG	CT		AG		GG	GA	TT	TC	CC	GG
123	F	76	H	CC	GG	CC	GA	AA	TC	GG	AA	TA		CC	GG
124	F	71	K	CC	GG	CC	GA	AA	TT	GG	GG	TT	TC	CC	GC
125	F	72	K	CC	GT	CC		AA		GG		TT		CC	GG
126	M	48	H	CT	TT	CT		AG		GG		TT		CC	CC
127	F	59	H	CC	GG	CC	GA	AA	TC	GA		TA		CT	CC
128	M	90	H	CC	GG	CC		AA		GG		TT		TT	GC
129	F	52	H	CT	GG		GG		TC	GG	AA		TT	CC	GC
130	F	73	K	CC	GG	CC	GG	AA	TC	GG	GG	TT	TT	CT	GC
131	F	68	K	CC	GG	CC		AA		GG		TT		CT	GG
132	F	63	K	CT	GG	CC		AA		GG		TT	TT	TT	GG
133	F	64	K	CC	GG	CC	GA	AG		GG	GA	AA	TC	CT	GC
134	M	81	K	CC	GT	CC	AA	AA	TT	GG	GA	TT	CC	CT	GC
135	F	82	K	CC	GG	CC		AA		GG		TT		CT	GG
136	M	69	K	CC	GT	CC	GA	AA	CC	GG	GG	TT	TT	CT	GC
137	F	68	K	CT	GT	CT		AG		GA		TT	CC	CT	GG
138	F	86	K	CC	GT	CC		AA		GG		TT		CT	GG
139	M	76	K	CC	GG	CC		AA		GG		TT		CT	GG
140	F	34	K	CC	GT	CC		AA		GG		TT	TT	CT	GG

Table D.1. Characteristics and genotypes of all donors whose tissues have been used. Transcript SNPs are named S1 (rs71561434), S2 (rs17414687), M1 (rs1045758), M2 (rs699186), T1 (rs41269315), T2 (rs240375), C1 (rs594012), C2 (rs240736), P1 (rs6253) and K1 (rs9029). Donors 1-230 OA, donors 231-275 NOF. Age (age at joint replacement); rs9350591and rs10492367 (association SNPs). The donors marked with an asterisk (*) were young, non-OA donors purchased from Lonza, UK. Continued overleaf.

Donor	Sex	Age (years)	Joint	rs9350591	rs10492367	S1	S2	M1	M2	T1	T2	C1	C2	P1	K1
141	F	73	H	CC	GG	CC		AA		GG		TT		CT	GG
142	M	78	H	CT	GG	CT		GG		GA		TT		CT	GG
143	F	88	K	CT	GT	CT		AG		GG		TT	TC	CT	GG
144	F	69	K	CC	GG	CC		AA		GG		TT	TT	CT	GG
145	F	61	K	CC	GT		GA		CC	GG	GA		TT		GC
146	F	70	H	CT	GG		GG		TT	GA	GA		TC	CC	GG
147	F	76	K	CC	GG		AA		TC	GG	GG		TC		GC
148	M	58	H	CC	GG		GA		TT	GG	GA		TT	CT	GC
149	M	54	K	CT	GG		GG		TC	GG	AA		TT		GG
150	M	54	K	CT	GG		GG		TC	GG	AA		TT		GG
151	F	62	K	CT	GT										
152	F	66	K	CC						GG			TC		
153	F	78	K	CC	GG		GA		TC	GG	GA		TT		GG
154	M	73	K	CC	GG		AA		TT	GG	GG		TC		CC
155	F	67	H	CC	GG		AA		CC	GG	GG		TC	CT	GG
156	M	58	K	CC	GG		GG		TC	GG	GA		TC		GC
157	F	45	H	CC	GG		GA		TT	GG	GA		TC	CT	GG
158	F	70	K	CC	GG		GG		CC	GG	GG		TT		GC
159	F	67	H	CC	GT		GG		TC	GG	GG		TC	CC	GC
160	F	45	H	CC	GG		GA		TT	GG	GG		TC	TT	GG

Table D.1. Characteristics and genotypes of all donors whose tissues have been used. Transcript SNPs are named S1 (rs71561434), S2 (rs17414687), M1 (rs1045758), M2 (rs699186), T1 (rs41269315), T2 (rs240375), C1 (rs594012), C2 (rs240736), P1 (rs6253) and K1 (rs9029). Donors 1-230 OA, donors 231-275 NOF. Age (age at joint replacement); rs9350591and rs10492367 (association SNPs). The donors marked with an asterisk (*) were young, non-OA donors purchased from Lonza, UK. Continued overleaf.

Donor	Sex	Age (years)	Joint	rs9350591	rs10492367	S1	S2	M1	M2	T1	T2	C1	C2	P1	K1
161	F	68	K	CC	GG		AA		TC	GG	AA		TC		GG
162	M	82	K	CC	GT		GG		TC	GG	AA		TC		GG
163	F	60	K	CT	GT		GA		TT	GG	AA		TT		GG
164	M	79	K	CT	GG		GA		TT	GA	GA		TT		GC
165	M	67	K	CC	GT		GA		TC	GG	GA		TT		GG
166	F	68	H	CC	GG		GG		TC	GG	GA		TC	CC	GC
167	F	63	K	CC	GT					GG	GA		TC		GG
168	F	87	K		GT		GA		TT	GG	GG		TT		GG
169	F	57	K	CC	GG					GG	GG		TC		
170	M	59	K	CC						GG			TT		
171	F	69	K	CC	GG						GA		CC		
172	M	68	K	CC	GG					GG	GG		TC		
173	M	75	K	CC	GG					GG	GA		TC		
174	F	57	K	CC	GG					GG			TC		
175	M	55	K	CC	GT					GG	GG		TC		GG
176	M	62	K	CC						GG			TC		
177	F	61	K	CT						GA			TT		
178	M	54	K	CC						GG			TT		
179	F	74	K	CT	GT		GA		TT	GA			TT		GG
180	M	64	K	CT	GG					GG	GG		TT		

Table D.1. Characteristics and genotypes of all donors whose tissues have been used. Transcript SNPs are named S1 (rs71561434), S2 (rs17414687), M1 (rs1045758), M2 (rs699186), T1 (rs41269315), T2 (rs240375), C1 (rs594012), C2 (rs240736), P1 (rs6253) and K1 (rs9029). Donors 1-230 OA, donors 231-275 NOF. Age (age at joint replacement); rs9350591and rs10492367 (association SNPs). The donors marked with an asterisk (*) were young, non-OA donors purchased from Lonza, UK. Continued overleaf.

Donor	Sex	Age (years)	Joint	rs9350591	rs10492367	S1	S2	M1	M2	T1	T2	C1	C2	P1	K1
181	F	54	K	CC	GG					GG	GG		TC		
182	M	55	K	CC	GG					GG	GA		TC		
183	M	50	K	CC	GG		AA		TT	GG	GA		TC		GG
184	F	56	K	CC	GT					GG	GG		TC		GG
185	F	73	K	CC	GG				CC	GG			TC		
186	F	61	K	CC	GG					GG	GA		TT		
187	M	82	K								AA				
188	F	73	K	CC	GG					GG	GA		TC		
189	F	56	K	CT	GG					GG	GA		TC		
190	F	77	K	CC	GT					GG	AA		TT		GC
191	F	66	H	CC	GG		GA		TC	GG	GA		CC	TT	GC
192	F	76	K	CC	GG					GG			TC		
193	M	55	K	CC	GT					GG			TT		
194	F	89	H	CC	GG		GA		TT	GG			TT	CC	GG
195	M	81	K	CC	GG					GG			TT		
196	M	65	H	CC	GG		GA		TT	GG			TC	CT	GG
197	F	72	H	CC	GG		GG		CC	GG			TC	CT	GG
198	F	64	K	CC	GT					GG			TC		
199	F	59	H		GG										
200	F	59	H	CC	GT		GA		TT	GG			CC	CT	GG

Table D.1. Characteristics and genotypes of all donors whose tissues have been used. Transcript SNPs are named S1 (rs71561434), S2 (rs17414687), M1 (rs1045758), M2 (rs699186), T1 (rs41269315), T2 (rs240375), C1 (rs594012), C2 (rs240736), P1 (rs6253) and K1 (rs9029). Donors 1-230 OA, donors 231-275 NOF. Age (age at joint replacement); rs9350591and rs10492367 (association SNPs). The donors marked with an asterisk (*) were young, non-OA donors purchased from Lonza, UK. Continued overleaf.

Donor	Sex	Age (years)	Joint	rs9350591	rs10492367	S1	S2	M1	M2	T1	T2	C1	C2	P1	K1
201	M	74	H	CT	GT				TC	GA			TC	CC	GC
202	F	72	H	CC	GT		GG		TT	GG			TT	CT	GG
203	F	67	K	CC	GT					GG			CC		
204	F	76	H	CC	GG		GA		TC	GG			TC	CC	GC
205	F	68	K	CC	GG					GG			TC		
206	F	77	H	CC	GG		GA		TT	GG	AA		TT	TT	GC
207	F	62	H	CC	GG		AA		TC	GG	GA		TT	CT	GC
208	F	79	H	CT	GG		GG		TC	GG			TT	CT	GC
209	F	51	H	CC	GT		GA		TC	GG			TT	CC	CC
210	F	36	K							GG				CT	
211	M	63	H	CT	GG		GA		TC	GG			TC	TT	GC
212	M	73	H	CC	GG		AA		TT	GG	GG		CC	CT	GC
213	F	51	H	CC	GG		AA		TT	GG			TC	CC	GC
214	M	64	H	CC	GG		AA		TC	GG			TT	CT	GC
215	F	31	H	CC	GG		GG		TT	GA			TC	CT	GC
216	F	64	H	CT	GG		GG		TC	GA			CC	CC	GG
217	F	58	H	CC	GG		GG		TT	GA			TT	TT	GC
218	F	58	H	CC	GG										
219	F	79	H	CC	GT		AA		TT	GG			CC	CT	GG
220	F	76	H	CC	GT		GA		TC	GG			CC	CC	GG

Table D.1. Characteristics and genotypes of all donors whose tissues have been used. Transcript SNPs are named S1 (rs71561434), S2 (rs17414687), M1 (rs1045758), M2 (rs699186), T1 (rs41269315), T2 (rs240375), C1 (rs594012), C2 (rs240736), P1 (rs6253) and K1 (rs9029). Donors 1-230 OA, donors 231-275 NOF. Age (age at joint replacement); rs9350591and rs10492367 (association SNPs). The donors marked with an asterisk (*) were young, non-OA donors purchased from Lonza, UK. Continued overleaf.

Donor	Sex	Age (years)	Joint	rs9350591	rs10492367	S1	S2	M1	M2	T1	T2	C1	C2	P1	K1
221	F	50	H	CC	GT		AA		TC	GG			TC	CC	GG
222	F	62	H	CC	GT										
223	M	67	H	CC	GG										
224	M	84	H	CC	GT										
225	M	55	H	CC	GG										
226	F	58	K	CC	GT										
227	M	67	K	CT	GG										
228	F	74	K	CC	GT										
229	M	68	K	CT	GT								TC		GC
230	M	75	K	CT	GG								TT		
231	F	69	H	CT	GT	CC		AG		GG		TT		CC	GG
232	F	71	H	CC	GG	CC		AG		GG		TT		CC	GC
233	F	81	H	CC	GG	CC		AA		GG		TT		CT	GC
234	F	72	H	CC	GG	CC		AA		GG		TT		CC	GC
235	F	84	H	CT	GG	CC		AG		GA		TT		CC	GC
236	F	79	H	CC	GG	CC		AA		GG		TT		CT	GC
237	F	94	H	CC	GG	CC		AA		GA		TT		CT	GC
238	F	84	H	CC	GG	CC		AG		GG		TT		CT	GG
239	F	84	H	CC	GG	CC		AA		GG		TA		CT	GG
240	F	52	H	CC	GG	CC		AA		GG		TT		CC	CC

Table D.1. Characteristics and genotypes of all donors whose tissues have been used. Transcript SNPs are named S1 (rs71561434), S2 (rs17414687), M1 (rs1045758), M2 (rs699186), T1 (rs41269315), T2 (rs240375), C1 (rs594012), C2 (rs240736), P1 (rs6253) and K1 (rs9029). Donors 1-230 OA, donors 231-275 NOF. Age (age at joint replacement); rs9350591and rs10492367 (association SNPs). The donors marked with an asterisk (*) were young, non-OA donors purchased from Lonza, UK. Continued overleaf.

Donor	Sex	Age (years)	Joint	rs9350591	rs10492367	S1	S2	M1	M2	T1	T2	C1	C2	P1	K1
241	F	80	H	CC	GG	CC		AA		GG		TT		TT	GC
242	F	86	H	CC	GG	CC		AA		GG		TT		CT	GC
243	F	89	H	CC	GT	CC		AG		GG		TT		CC	GC
244	F	91	H	CC	GT	CC		AG		GG		TA		CC	GG
245	F	82	H	CT	GG	CT		AG		GG		TT		CT	GC
246	F	80	H	CC	GG	CC		AG		GG		TT		TT	GC
247	F	83	H	CC	GG	CC		AA		GG		TT		TT	GG
248	F	80	H	CC	GG	CC		AA		GA		TA		CT	GG
249	M	84	H	CC	GG	CC		AG		GG		TT		CT	GC
250	F	82	H	CC	GT	CC		AG		GG		TA		CC	GC
251	F	91	H	CC	GT	CC		AA		GG		TT		CT	GG
252	M	85	H	CC	GG		GA		TT	GG			TC	CC	GG
253	M	85	H	CC	GG		GA		CC	GG			TT	CC	GC
254	M	78	H	CC	GG		GG		CC	GG	AA		TT	CC	CC
255	F	80	H	CC	GG		GA		TT	GG	GG		TT	CC	CC
256	M	62	H	CT	TT		GG		TT	GG	GG		TT	CC	GG
257	F	95	H	CC	GG		GA		TT	GG	AA		TT	CC	CC
258	M	75	H	CC	GG		GA		TC	GG	AA		TC	CC	GC
259	M	79	H	CC	GG		GA		TT	GG	GG		TT	CT	GC
260	F	81	H	CT	TT		GG		TT	GA	GA		TC	TT	GG

Table D.1. Characteristics and genotypes of all donors whose tissues have been used. Transcript SNPs are named S1 (rs71561434), S2 (rs17414687), M1 (rs1045758), M2 (rs699186), T1 (rs41269315), T2 (rs240375), C1 (rs594012), C2 (rs240736), P1 (rs6253) and K1 (rs9029). Donors 1-230 OA, donors 231-275 NOF. Age (age at joint replacement); rs9350591and rs10492367 (association SNPs). The donors marked with an asterisk (*) were young, non-OA donors purchased from Lonza, UK. Continued overleaf.

Donor	Sex	Age (years)	Joint	rs9350591	rs10492367	S1	S2	M1	M2	T1	T2	C1	C2	P1	K1
261	F	77	H	CC	GG		GA		TT	GG			TT	CT	GC
262	M	86	H	CC	GG		AA		TC	GG			TC	CT	GC
263	F	72	H	CC	GT				CC	GG	AA		TT	CT	GG
264	F	62	H	CC	GG		GA		TC	GG			TT	CC	GC
265	F	81	H	CC	GG		GG		TT	GG			TT	CC	GC
266	F	62	H	CC	GG		AA		TC	GG	GG		CC	CT	GG
267	F	87	H	CC	GG		GA		TC	GG	AA		TC	CT	GG
268	F	71	H	CC	GG		GG		CC	GG	AA		TT	CT	CC
269	F	86	H	CC	GG		GA		TT	GG	GG		TT	CC	GG
270	F	80	H	CC	GG		AA		TC	GG			TC	TT	GC
271	F	83	H	CC	TT		GG		TT	GA			TT	CC	GC
272	F	92	H	CC	GG		GA		TT	GG			TC	TT	GC
273	F	73	H	CC	GG		GA		TC	GG			TT	CC	GG
274	M	69	H	CT	GG										
275	F	66	H	CC	GG		GA		TT	GG	GA		CC	CT	GC
276*	F	41	H	CC	GG										
277*	F	24	H	CC	GG										
278*	M	25	H	CC	GG										

Table D.1. Characteristics and genotypes of all donors whose tissues have been used. Transcript SNPs are named S1 (rs71561434), S2 (rs17414687), M1 (rs1045758), M2 (rs699186), T1 (rs41269315), T2 (rs240375), C1 (rs594012), C2 (rs240736), P1 (rs6253) and K1 (rs9029). Donors 1-230 OA, donors 231-275 NOF. Age (age at joint replacement); rs9350591and rs10492367 (association SNPs). The donors marked with an asterisk (*) were young, non-OA donors purchased from Lonza, UK.

Appendix E. Antibodies and siRNAs

Antibody	Cat. No.	rs10492367	rs58649696	rs11049206	rs10843013
apCREB	06-519	Stock	-	-	-
AR	sc-816	Stock	Stock	Stock	Stock
C/EBPβ	sc-150	Predicted	Stock	Stock	Stock
DABP	sc-98411	-	-	-	Stock
Deaf1	AP2711b	Stock	Stock	Stock	Stock
E2A.E12 (TCF3)	sc-349	Predicted	Stock	-	Stock
E2F1	sc-193	Stock	-	-	-
Egr1	sc-20689	Stock	Stock	-	-
Ets1/2	sc-112	Stock	-	-	-
FOXP1	sc-66896	-	-	Predicted	Stock
GR	sc-1003	Stock	Stock	Stock	Stock
HEB (TCF12)	sc-357	Predicted	Stock	Stock	Stock
HLF	sc-367607	-	-	-	Stock
HNF4a	sc-8987	Stock	Stock	Stock	Stock
IRF2	sc-498	-	Predicted	-	Stock
KLF16	sc-131168	Stock	Stock	Stock	Stock
LEF1	sc-28687	Stock	Stock	Stock	Stock
MEF2A	sc-10794	Predicted	Stock	Predicted	Stock
MEF2C	sc-13268	Predicted	Stock	Stock	Stock
NFATC3	sc-1152	Stock	Stock	Stock	Stock
NFkB p65 (RELA)	sc-372	Predicted	Stock	Stock	Stock
NFYA	sc-10779	-	-	Predicted	Stock
Nkx3.2	sc-25066	Stock	Stock	Stock	Stock
Oct2 (POU2F2)	sc-25400	-	Predicted	-	Stock
Pax5	sc-1975	Stock	Stock	Stock	Stock
Pax9	sc-7746	Stock	Stock	-	Stock
PC4	sc-48778	Stock	Stock	Stock	Stock
PR	sc-538	Stock	Stock	Stock	Stock
Sox9	sc-20095	Stock	Stock	Stock	Stock
Sp1	sc-59	Stock	Stock	-	Stock
Sp3	sc-644	Stock	Stock	Stock	Stock
TCF4	sc-13027	Stock	Predicted	Stock	Stock
XBP1	sc-7160	-	Predicted	-	Stock

Table E.1. Antibodies used for supershift EMSAs to investigate protein binding to rs10492367, rs58649696, rs11049206 and rs10843013. Stock = transcription factor not predicted to bind the polymorphism; predicted = transcription factor predicted to bind to the polymorphism; - = binding of transcription factor not investigated. All antibodies with catalogue numbers (Cat. No.) “sc-” were purchased from Santa Cruz Biotechnology (USA), apCREB from Millipore (UK) and Deaf1 from Abgent (USA).

Antibody	Cat. No.	Manufacturer	Use
α-acetyl-histone H3	06-599	Millipore, UK	ChIP: positive control
α-GAPDH	MAB374	Millipore, UK	Western blot: cytosolic protein
α-lamin A/C	2032	CST, USA	Western blot: nuclear protein
Goat IgG	I5256	Sigma-Aldrich, UK	EMSA: IgG control
Rabbit IgG	I5006	Sigma-Aldrich, UK	EMSA: IgG control

Table E.2. Additional antibodies used in ChIP, western blot and EMSA experiments.

siRNA	Catalogue number
SMARTpool: ON-TARGETplus *RELA* siRNA	L-003533-00
SMARTpool: ON-TARGETplus *SUB1* siRNA	L-009815-00
SMARTpool: ON-TARGETplus *TCF3* siRNA	L-009384-00
ON-TARGETplus Non-targeting Pool	D-001810-10

Table E.3. siRNA (Dharmacon, GE Healthcare, UK) used for the knockdown of *RELA*, *SUB1*, and *TCF3*.

Appendix F. Electrophoretic Mobility Shift Assays

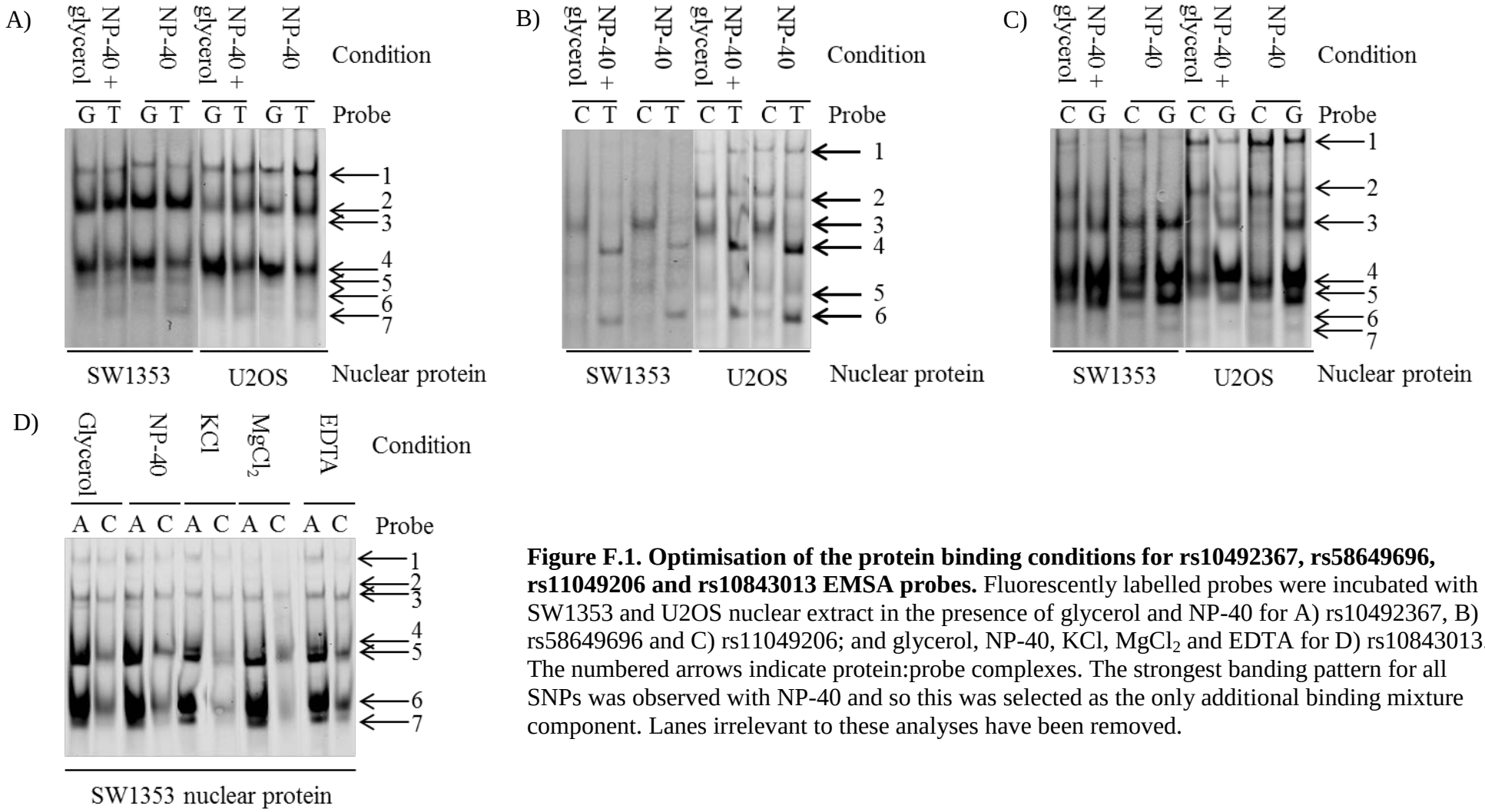

Figure F.1. Optimisation of the protein binding conditions for rs10492367, rs58649696, rs11049206 and rs10843013 EMSA probes. Fluorescently labelled probes were incubated with SW1353 and U2OS nuclear extract in the presence of glycerol and NP-40 for A) rs10492367, B) rs58649696 and C) rs11049206; and glycerol, NP-40, KCl, $MgCl_2$ and EDTA for D) rs10843013. The numbered arrows indicate protein:probe complexes. The strongest banding pattern for all SNPs was observed with NP-40 and so this was selected as the only additional binding mixture component. Lanes irrelevant to these analyses have been removed.

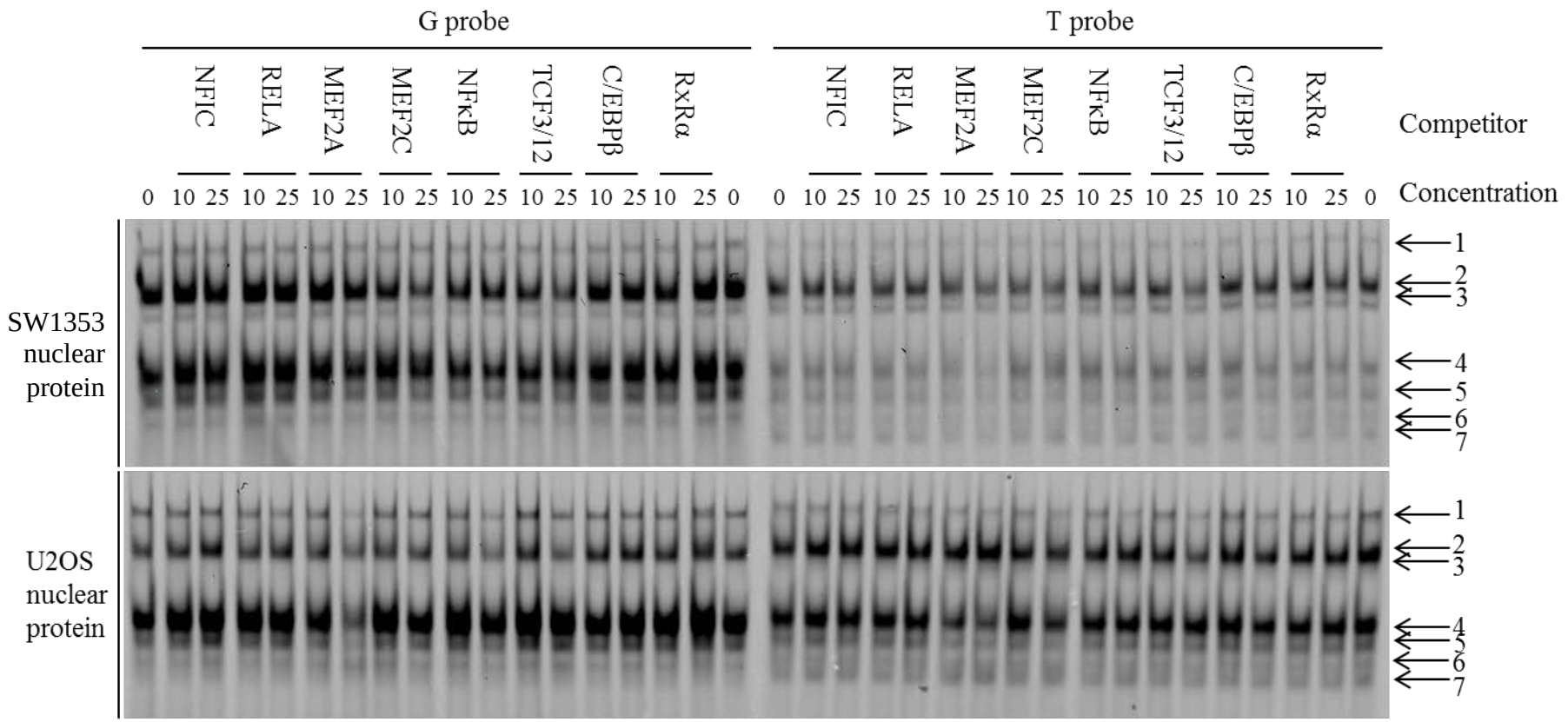

Figure F.2. Competition EMSAs to investigate the consensus sequences necessary for SW1353 and U2OS nuclear extract binding to the G and T alleles of rs10492367. The protein:probe mixes were incubated with unlabelled competitors that contained the consensus sequences for transcription factors predicted to bind the G or T allele probes. With both nuclear extracts, the G and T allele probe binding to the protein in complex 4 was disrupted by the MEF2A competitor, whereas the T allele probe binding to the protein in complex 2 was disrupted by the MEF2C competitor with the U2OS nuclear extract. The G allele probe binding to the U2OS nuclear extract in complex 2 was disrupted by competition with the NFκB consensus sequence. Competition with TCF3/12 caused disruption of complex 2 for both alleles in both cell lines. C/EBPβ appeared to cause a slight decrease in band intensity of complex 1 of the T allele probe with both nuclear extracts. Concentration = 0, 10, 25 x probe concentration.

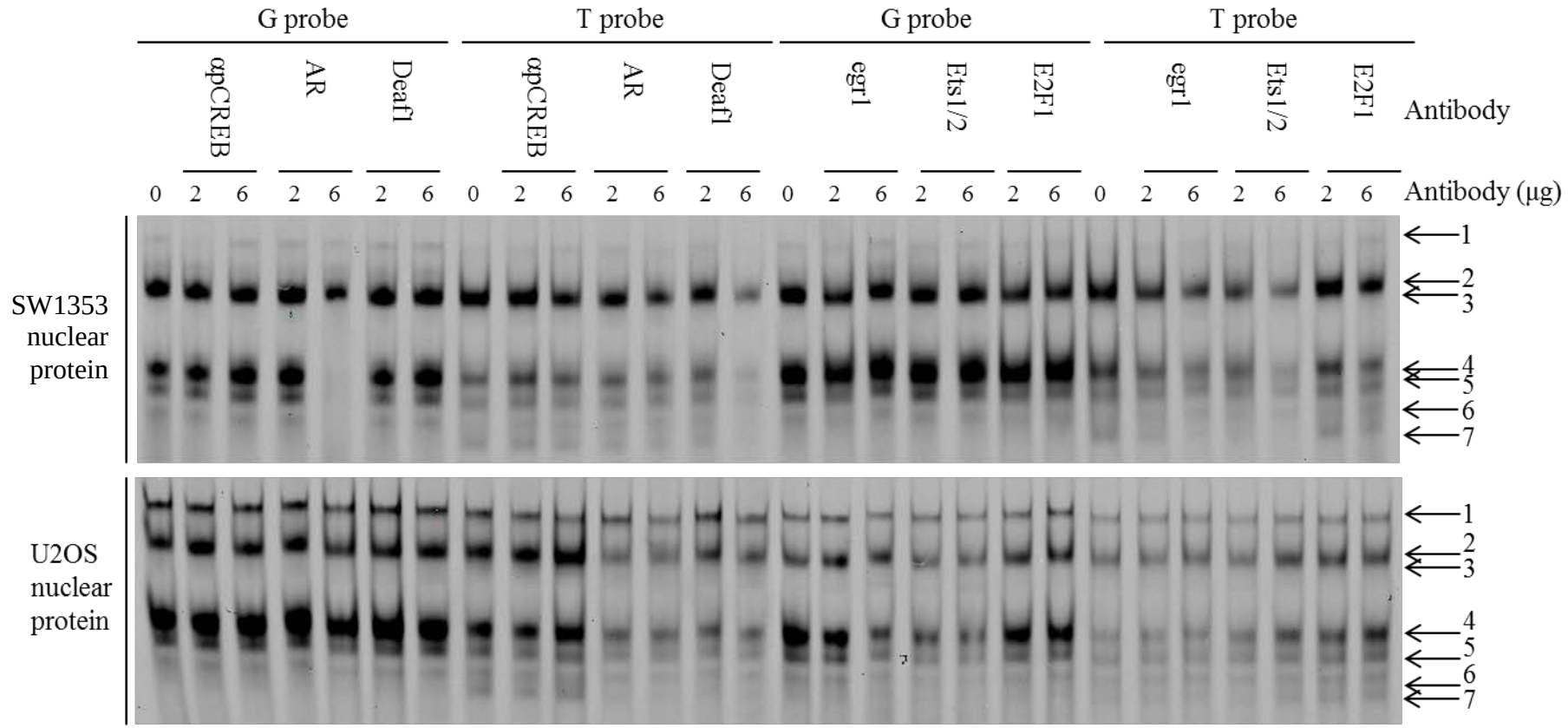

Figure F.3. Supershift EMSAs to investigate the transcription factors of SW1353 and U2OS nuclear extracts binding to the G and T alleles of rs10492367. The protein:probe mixes were incubated with either 2 µg or 6 µg of antibody. Band 4 binding to the G allele probe with SW1353 nuclear protein was outcompeted when incubated with AR. Band 4 binding to the T allele probe with SW1353 nuclear protein was outcompeted when incubated with Deaf1 and Ets1/2. No changes in the banding patterns were observed after incubation with the remaining antibodies.

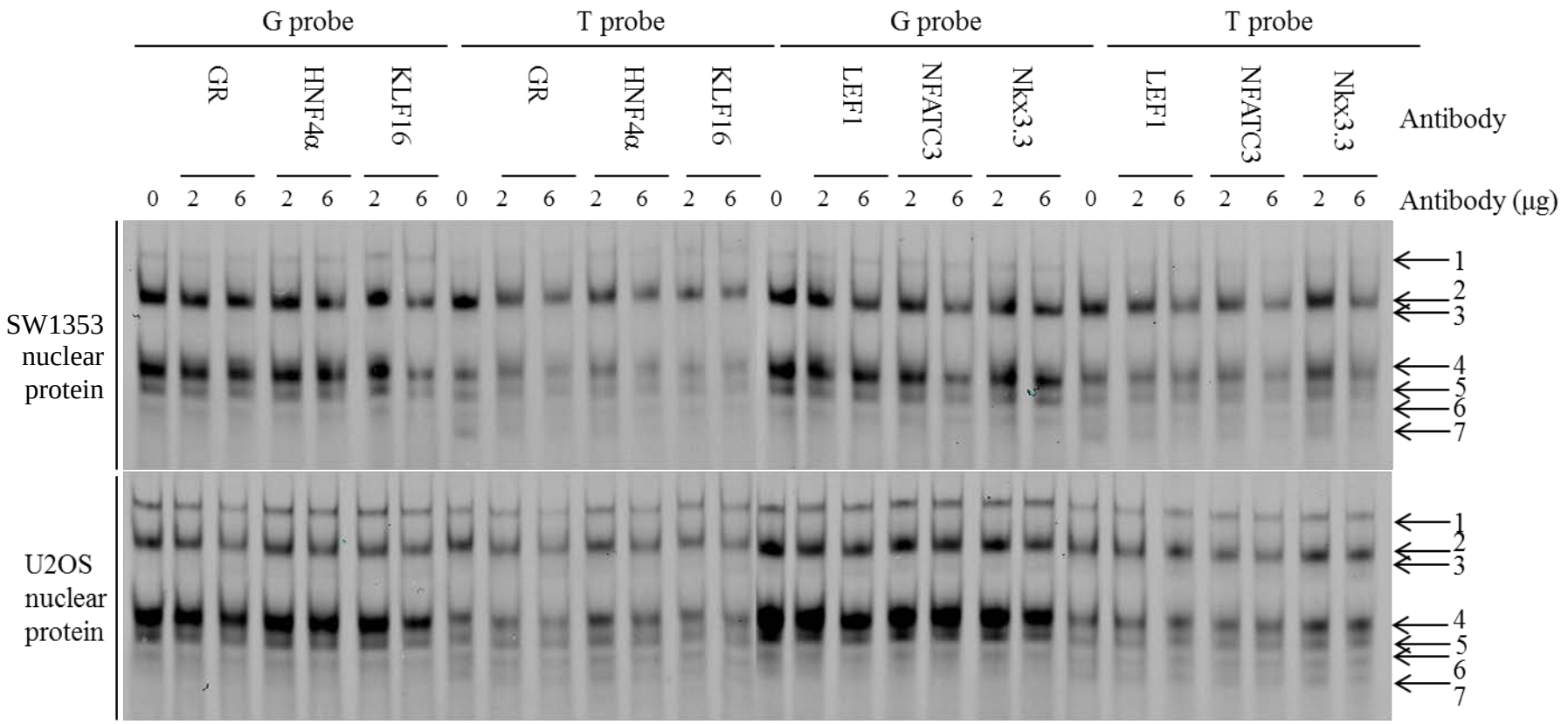

Figure F.4. Supershift EMSAs to investigate the transcription factors of SW1353 and U2OS nuclear extracts binding to the G and T alleles of rs10492367. The protein:probe mixes were incubated with either 2 µg or 6 µg of antibody. No changes in the banding patterns were observed after incubation with any of the antibodies.

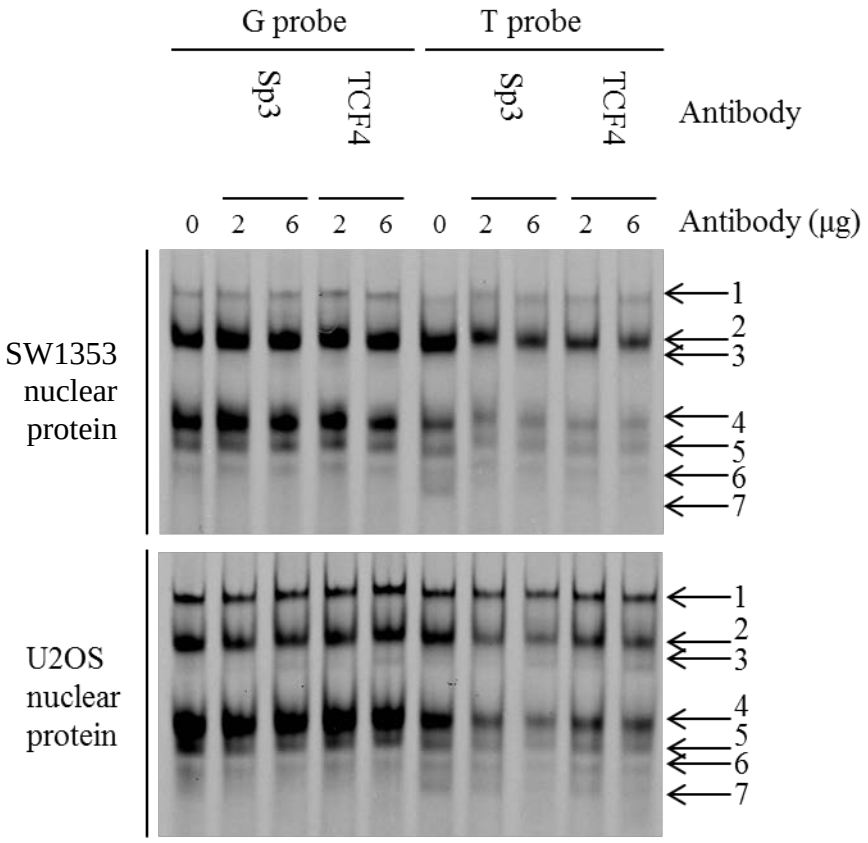

Figure F.5. Supershift EMSAs to investigate the transcription factors of SW1353 and U2OS nuclear extracts binding to the G and T alleles of rs10492367. The protein:probe mixes were incubated with either 2 µg or 6 µg of antibody. No changes in the banding patterns were observed after incubation with any of the antibodies.

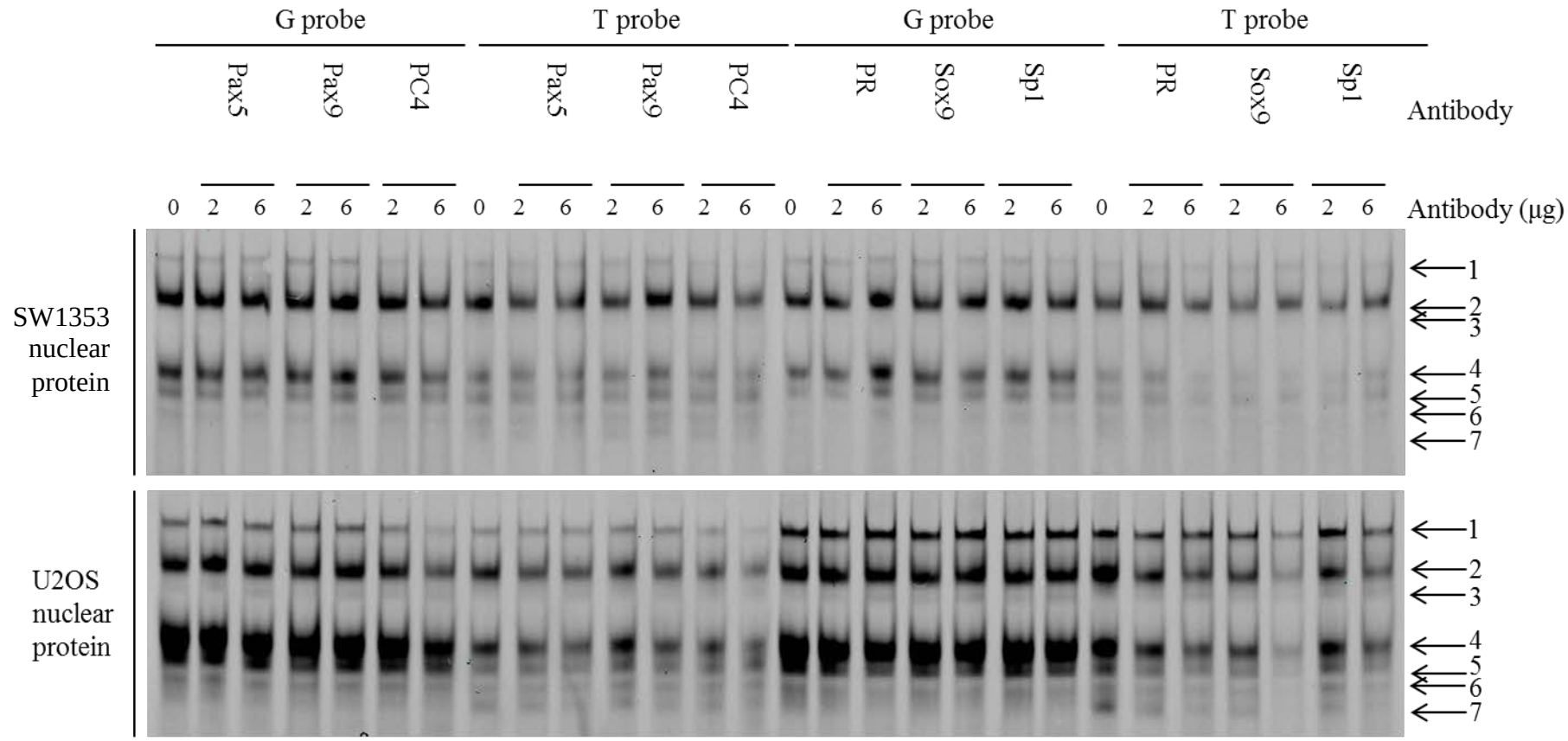

Figure F.6. Supershift EMSAs to investigate the transcription factors of SW1353 and U2OS nuclear extracts binding to the G and T alleles of rs10492367. The protein:probe mixes were incubated with either 2 µg or 6 µg of antibody. Nuclear protein binding to both allele probes in complex 1 was outcompeted by PC4. Incubation with Sox9 caused bands 4, 5, 6 and 7 of the T allele probe to become fainter with the U2OS nuclear extract. No changes in the banding patterns were observed after incubation with the remaining antibodies.

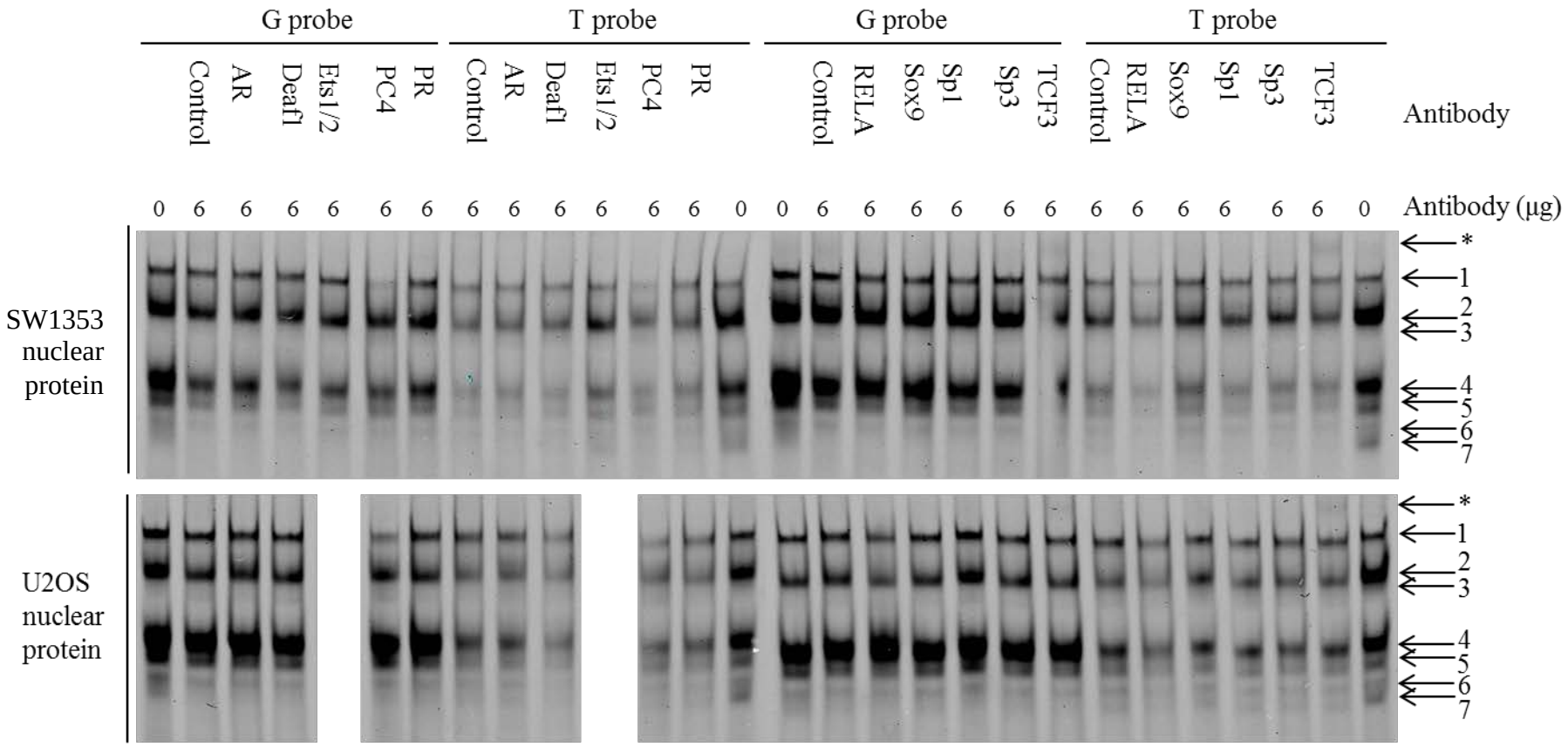

Figure F.7. Replication supershift EMSAs to investigate the transcription factors of SW1353 and U2OS nuclear extracts binding to the G and T alleles of rs10492367. The protein:probe mixes were incubated with either 2 µg or 6 µg of antibody that showed supershifts or were ambiguous previously. PC4, RELA and TCF3 supershifts were replicated. A supershift, marked by an asterisk (*), was observed for both alleles after incubation with the TCF3 antibody. No changes in the banding patterns were observed after incubation with the remaining antibodies.

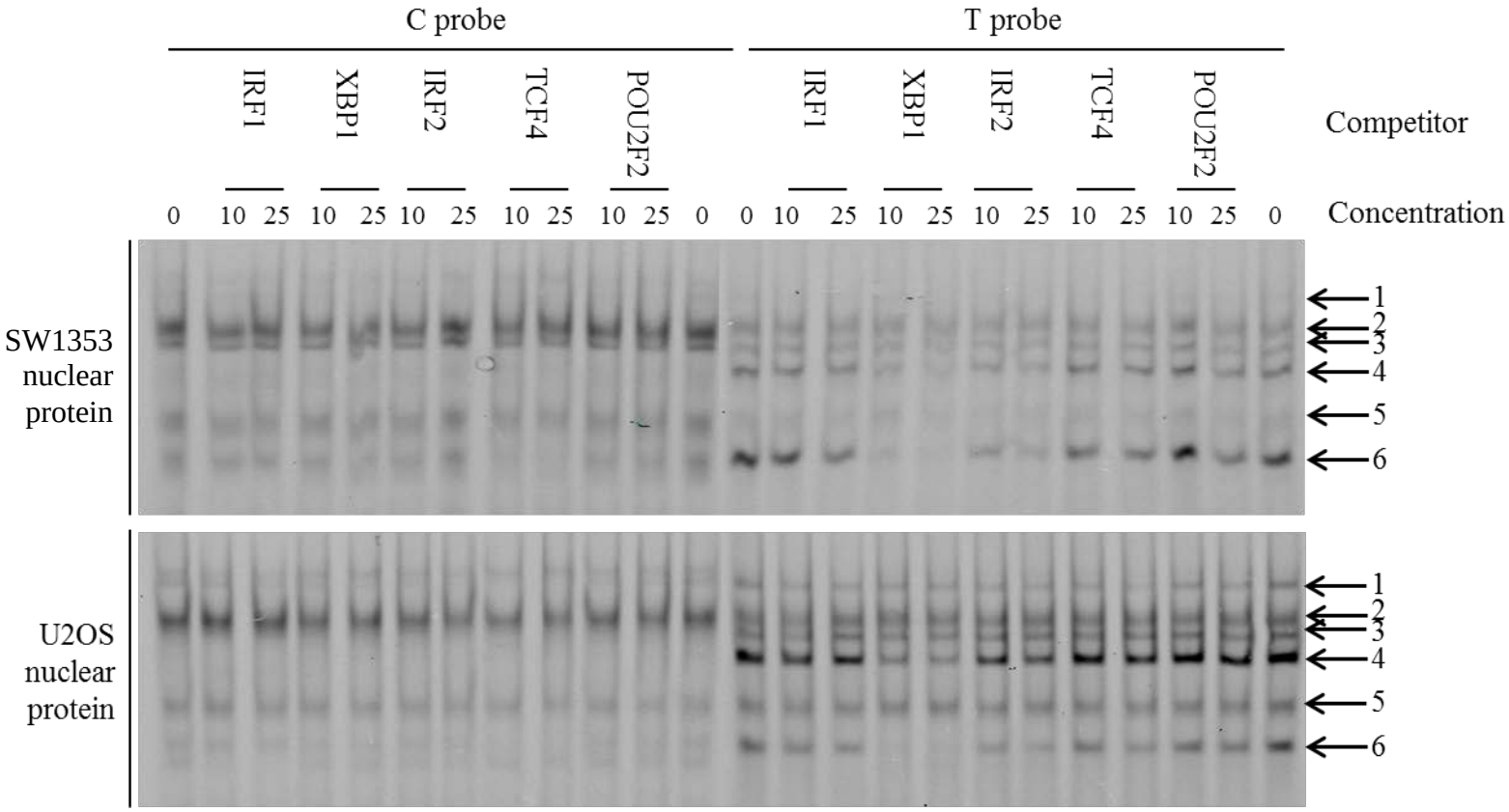

Figure F.8. Competition EMSAs to investigate the consensus sequences necessary for SW1353 and U2OS nuclear extract binding to the C and T alleles of rs58649696. The protein:probe mixes were incubated with unlabelled competitors that contained the consensus sequences for transcription factors predicted to bind the C or T allele probes. With both nuclear extracts, the T allele probe binding to the protein in complexes 4 and 6 was disrupted by the XBP1, IRF2 and POU2F2 competitors. Competition of the C allele probe with the TCF4 consensus sequence appeared to slightly decrease the band intensity of complex 6 with the SW1353 nuclear extract. There was no observed competition with the consensus sequence for IRF1. Concentration = 0, 10, 25 x probe concentration.

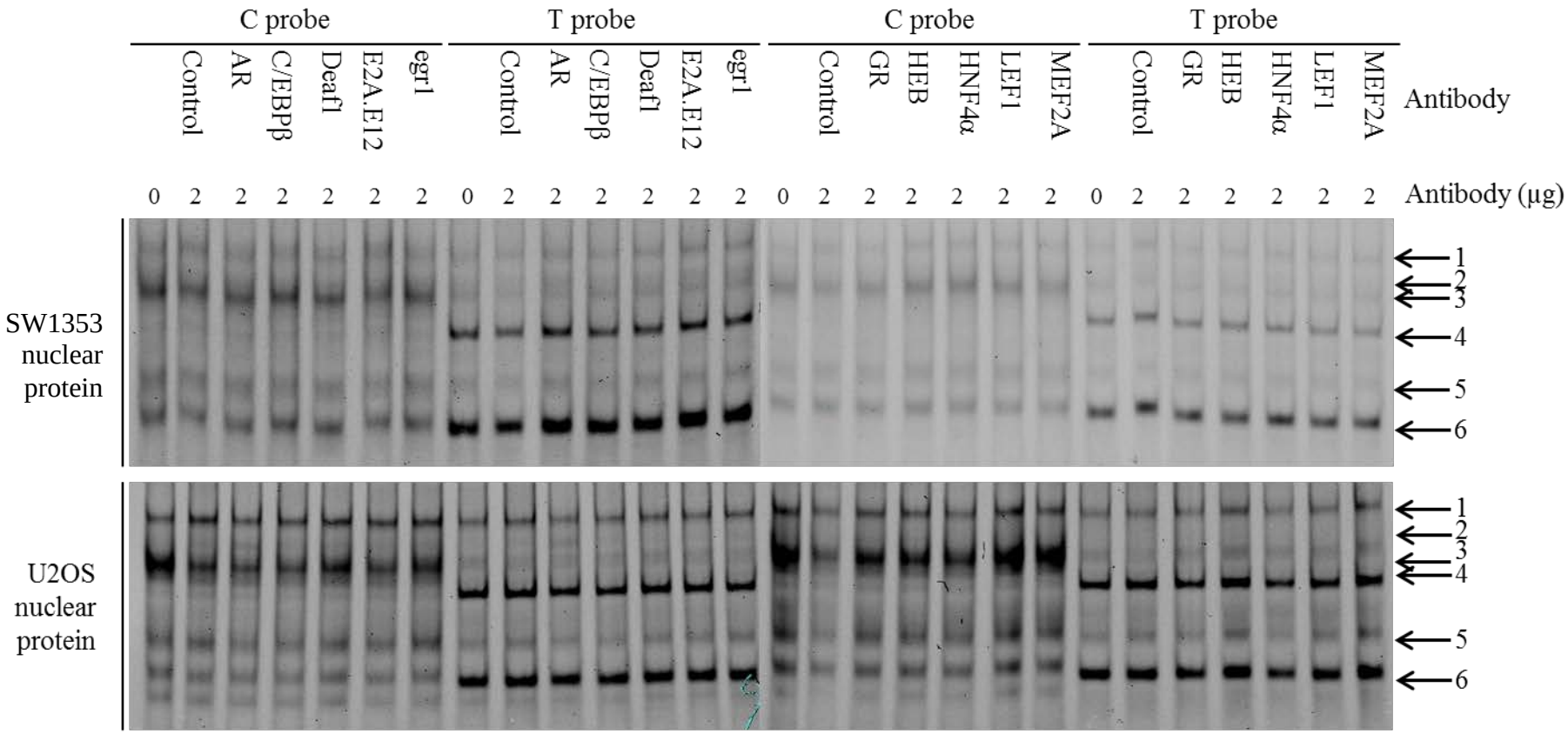

Figure F.9. Supershift EMSAs to investigate the transcription factors of SW1353 and U2OS nuclear extracts binding to the C and T alleles of rs58649696. The protein:probe mixes were incubated with 2 µg of antibody. No changes in the banding patterns were observed after incubation with any of the antibodies.

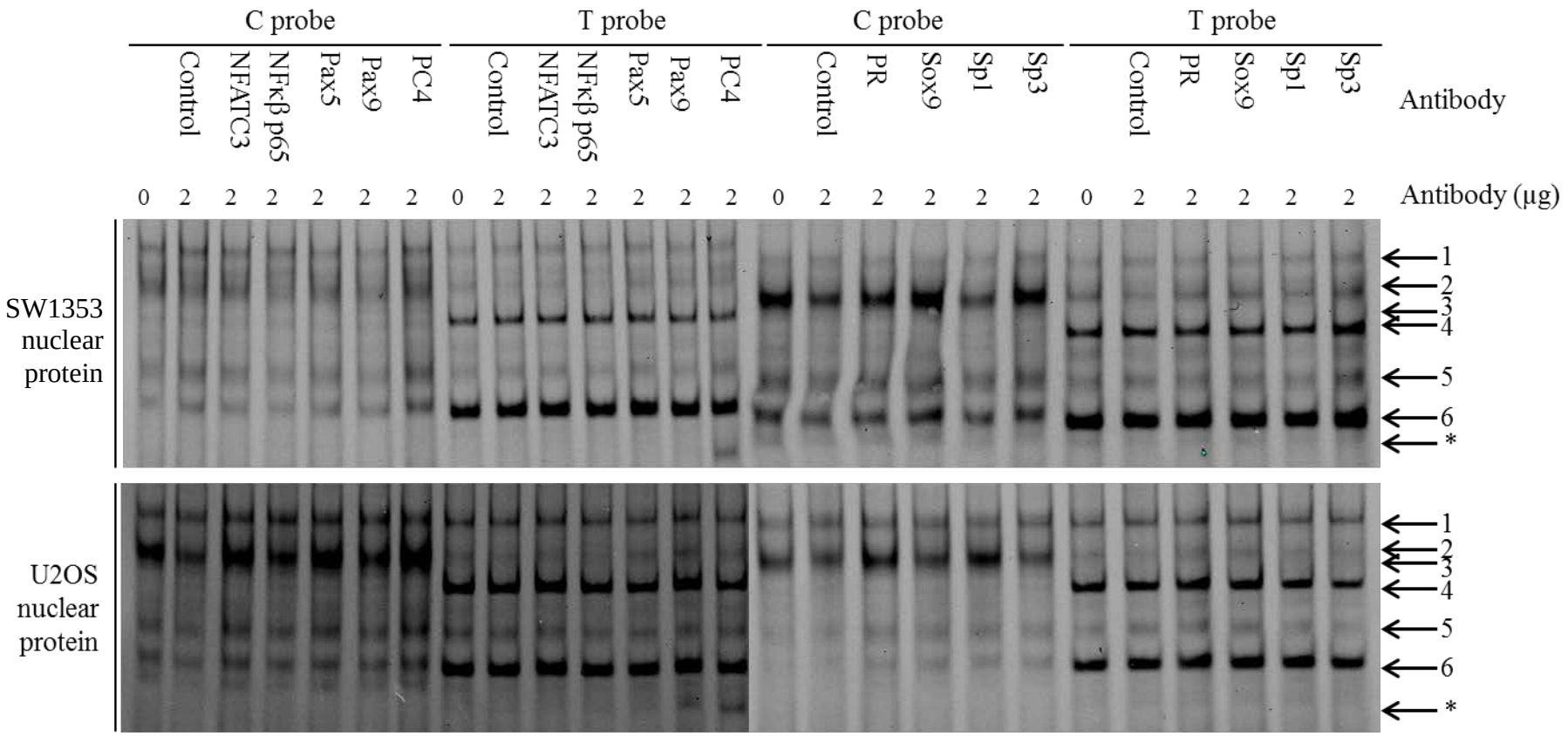

Figure F.10. Supershift EMSAs to investigate the transcription factors of SW1353 and U2OS nuclear extracts binding to the C and T alleles of rs58649696. The protein:probe mixes were incubated with 2 µg of antibody. A supershift, marked by an asterisk (*), with PC4 and PAX9 was observed for the T allele probe with both cell line nuclear extracts. No changes in the banding patterns were observed after incubation with the remaining antibodies.

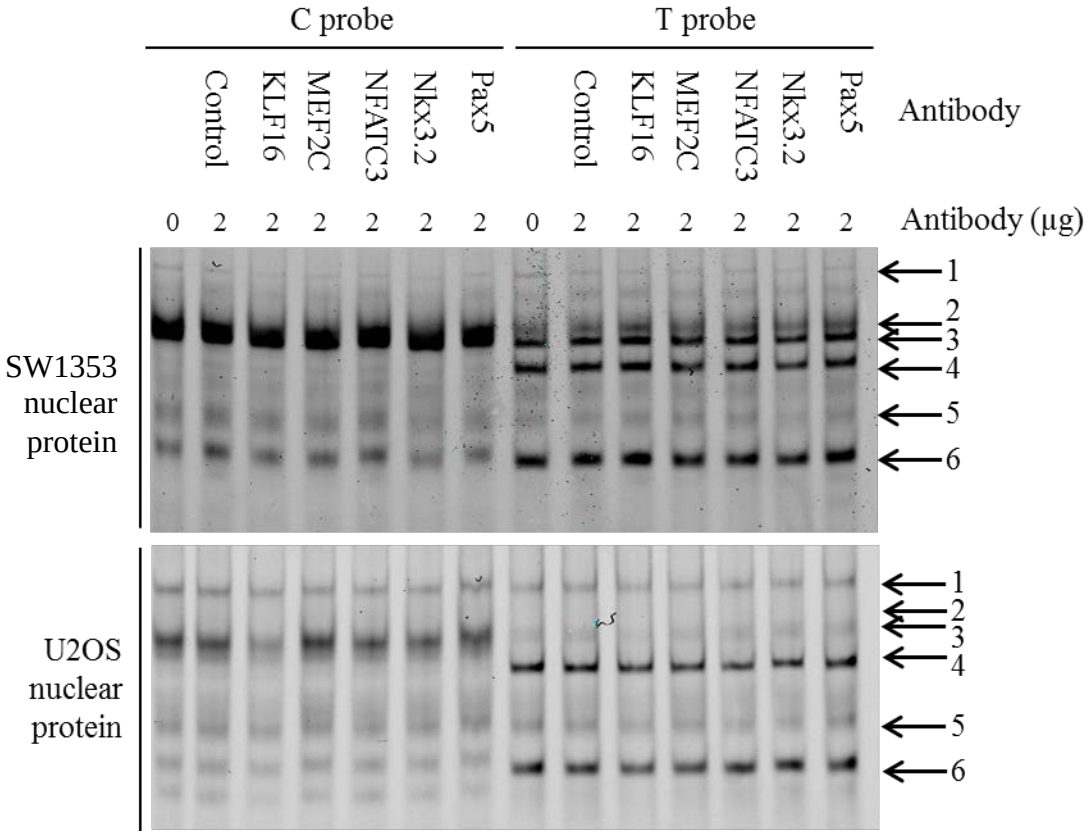

Figure F.11. Supershift EMSAs to investigate the transcription factors of SW1353 and U2OS nuclear extracts binding to the C and T alleles of rs58649696. The protein:probe mixes were incubated with 2 µg of antibody. No changes in the banding patterns were observed after incubation with any of the antibodies.

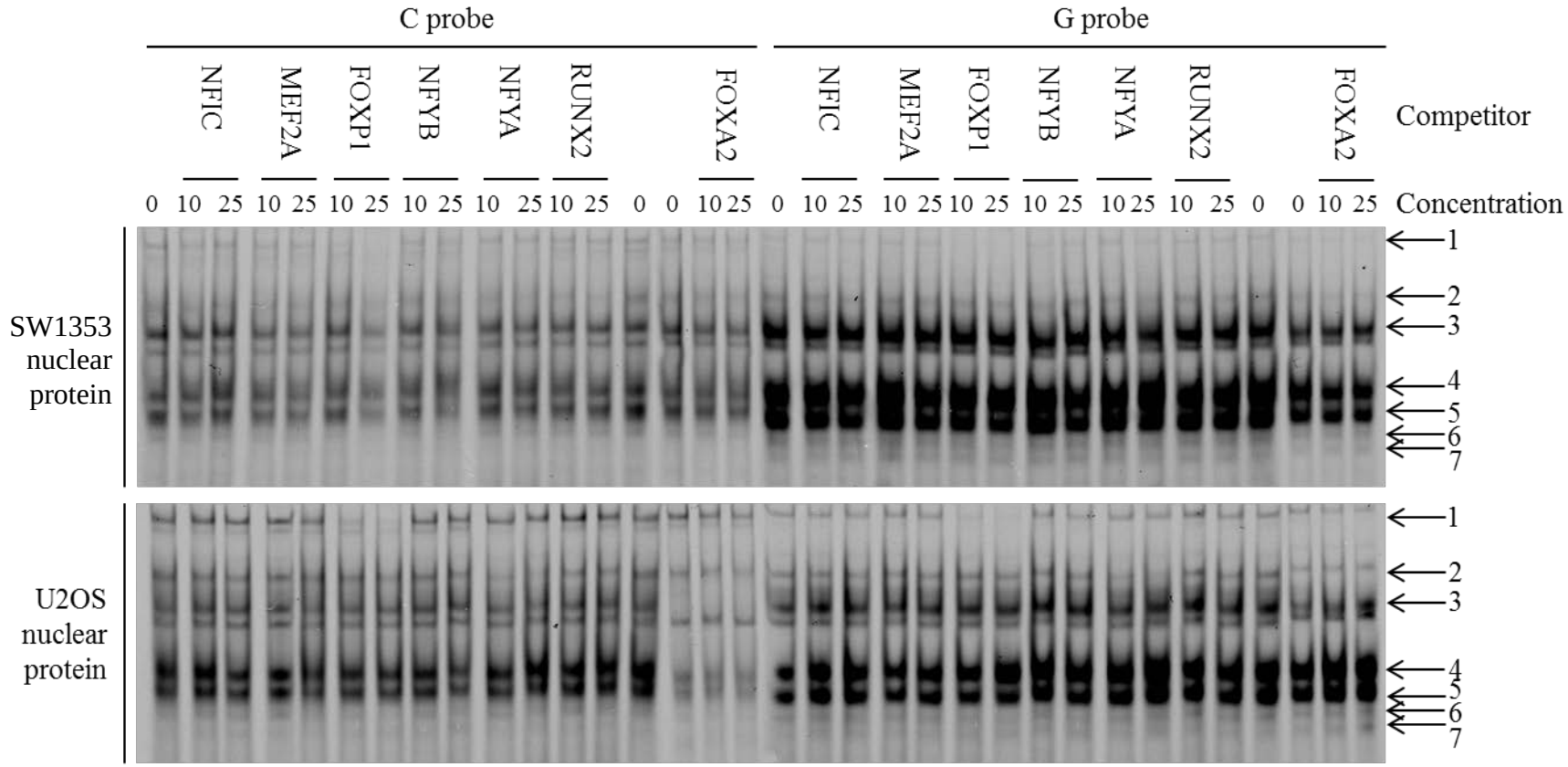

Figure F.12. Competition EMSAs to investigate the consensus sequences necessary for SW1353 and U2OS nuclear extract binding to the C and G alleles of rs11049206. The protein:probe mixes were incubated with unlabelled competitors that contained the consensus sequences for transcription factors predicted to bind the C or G allele probes. Complex 1 was outcompeted with both nuclear extracts and in both alleles with the FOXP1 consensus sequence. Protein binding to the C allele probe appeared to be decreased upon competition with MEF2A (complexes 4 and 5), FOXP1 (complexes 3, 4 and 6), NFYA (complex 2) and RUNX2 (complex 2). Competition of the C allele probe with the NFYB competitor caused a slight decrease in band intensity of complexes 4 and 5 in both cell lines. Complexes 2, 4 and 5 were slightly outcompeted by the NFIC competitor with the U2OS nuclear protein in the C allele probe. There was no observed competition with the consensus sequence for FOXA2. Concentration = 0, 10, 25 x probe concentration.

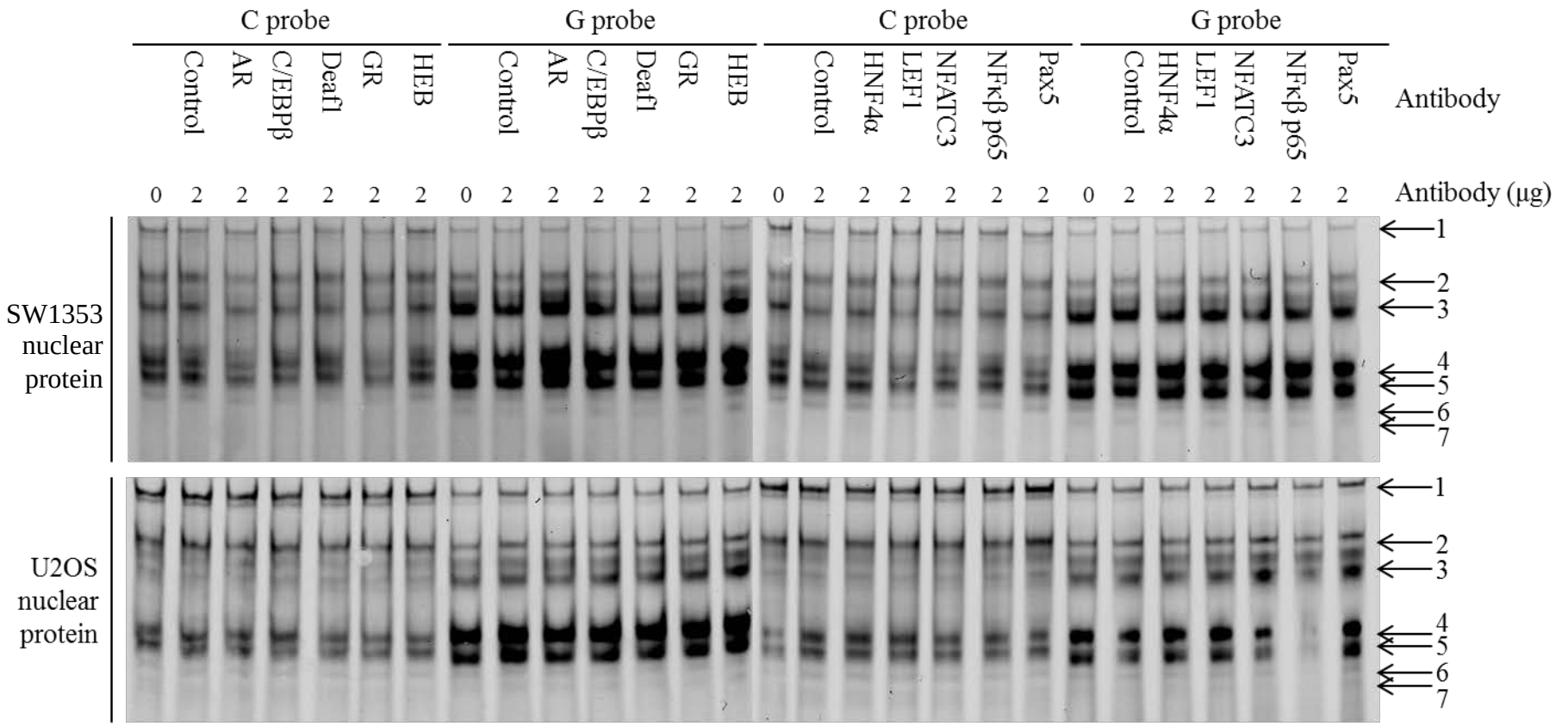

Figure F.13. Supershift EMSAs to investigate the transcription factors of SW1353 and U2OS nuclear extracts binding to the C and G alleles of rs11049206. The protein:probe mixes were incubated with 2 µg of antibody. Incubation with the NFκβ p65 antibody caused bands 4 and 5 to become fainter in the G allele probe with the U2OS nuclear extract. No changes in the banding patterns were observed after incubation with the remaining antibodies.

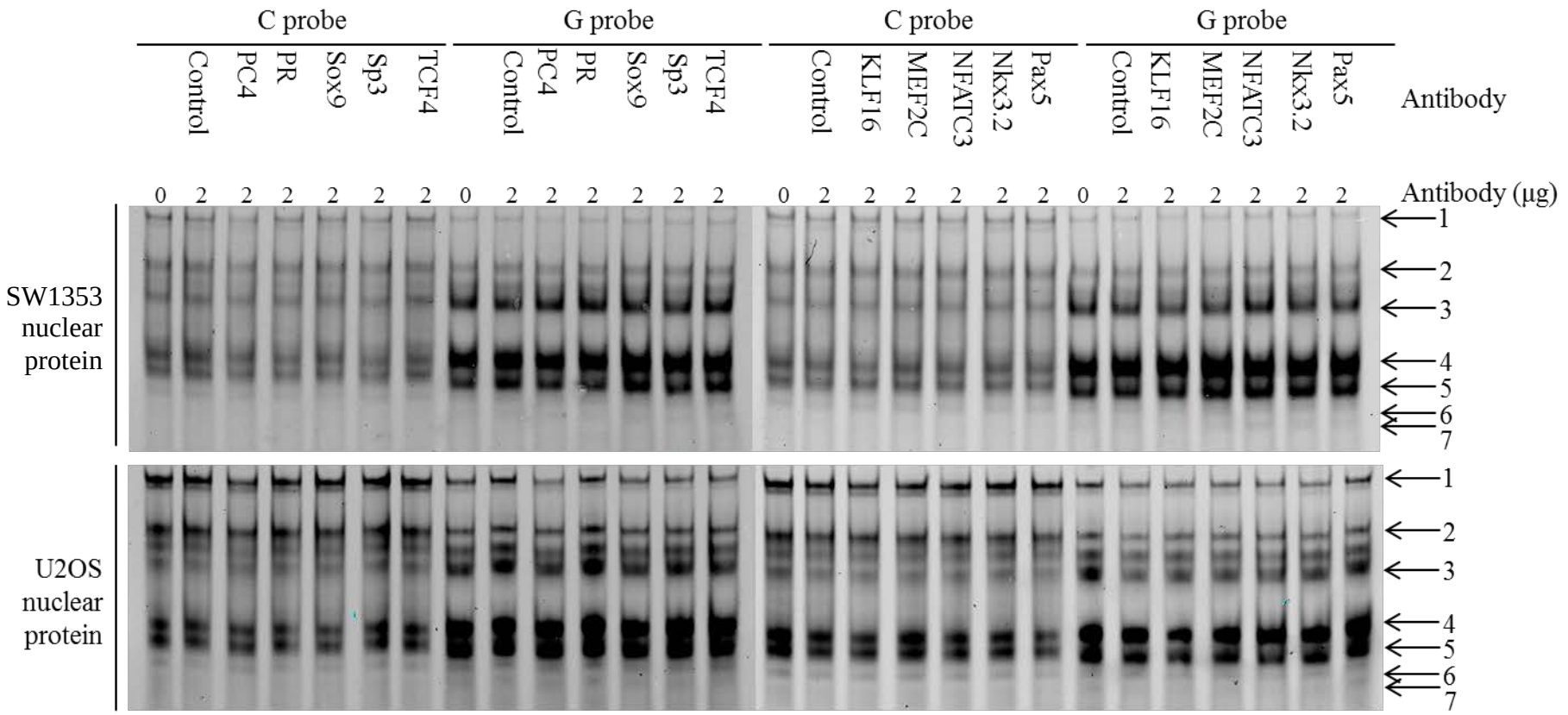

Figure F.14. Supershift EMSAs to investigate the transcription factors of SW1353 and U2OS nuclear extracts binding to the C and G alleles of rs11049206. The protein:probe mixes were incubated with 2 µg of antibody. No changes in the banding patterns were observed after incubation with any of the antibodies.

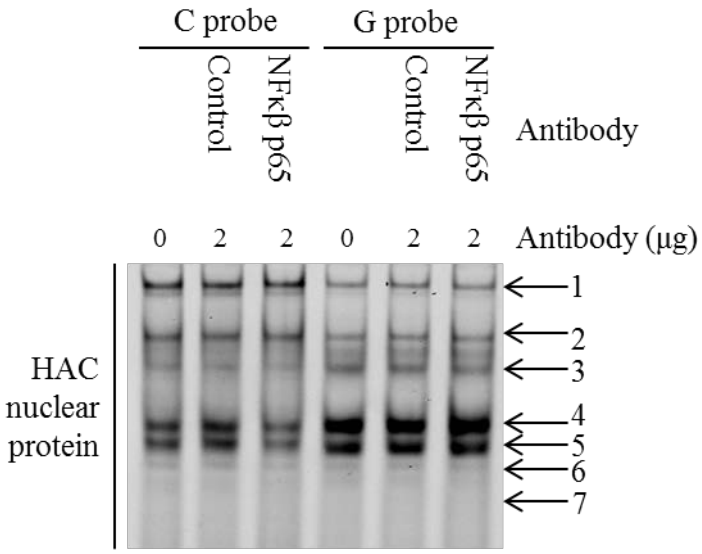

Figure F.15. Supershift EMSAs to investigate the transcription factors of human articular chondrocyte (HAC) nuclear extract binding to the C and G alleles of rs11049206. The protein:probe mixes were incubated with 2 µg of antibody. The results confirmed that, unlike with U2OS nuclear protein, NFκβ p65 of HAC nuclear protein does not interact with the fluorescently labelled DNA. Control (IgG species-matched antibody). Lanes irrelevant to this analysis have been removed.

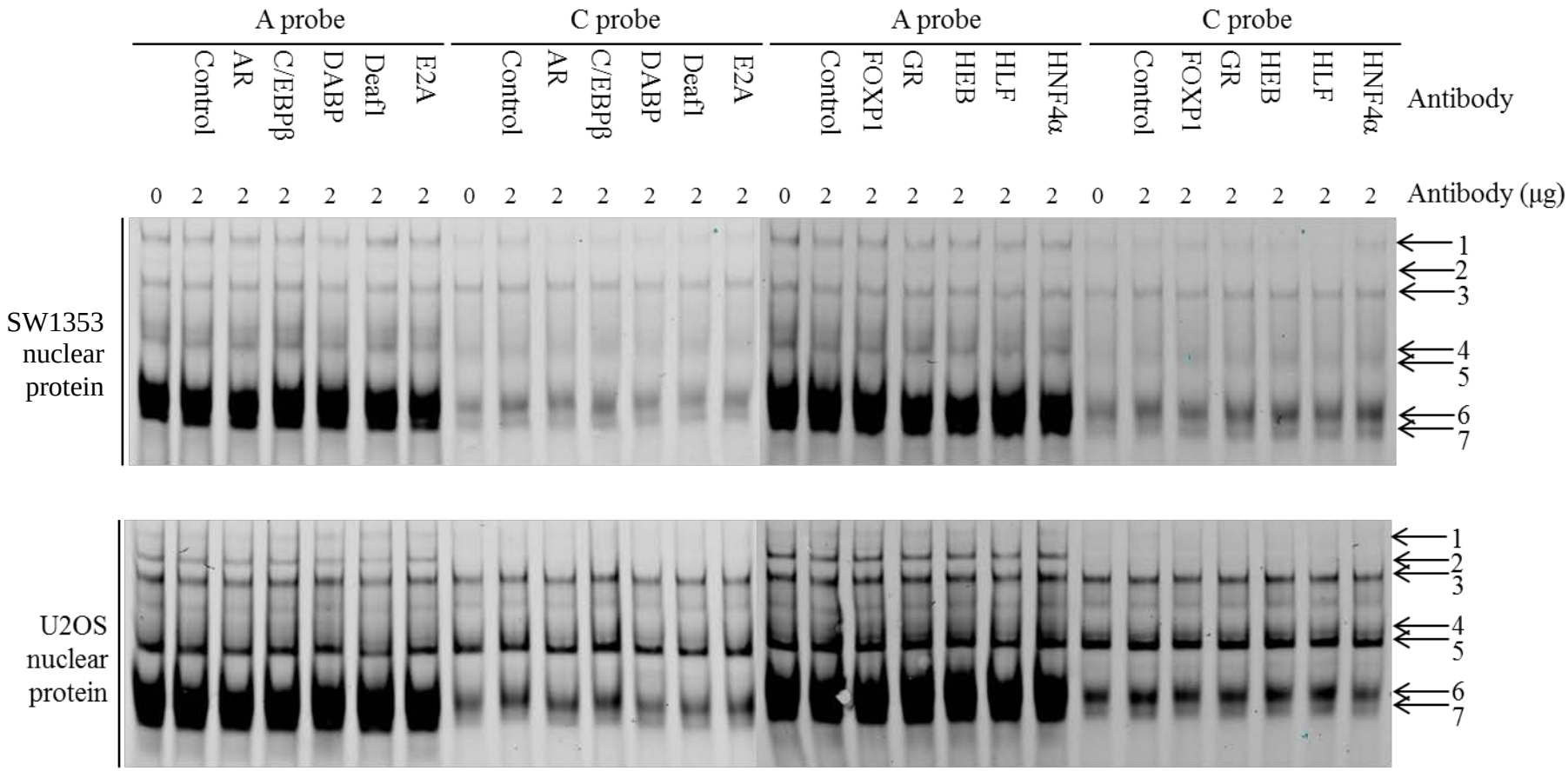

Figure F.16. Supershift EMSAs to investigate the transcription factors of SW1353 and U2OS nuclear extracts binding to the A and C alleles of rs10843013. The protein:probe mixes were incubated with 2 µg of antibody. No changes in the banding patterns were observed after incubation with any of the antibodies.

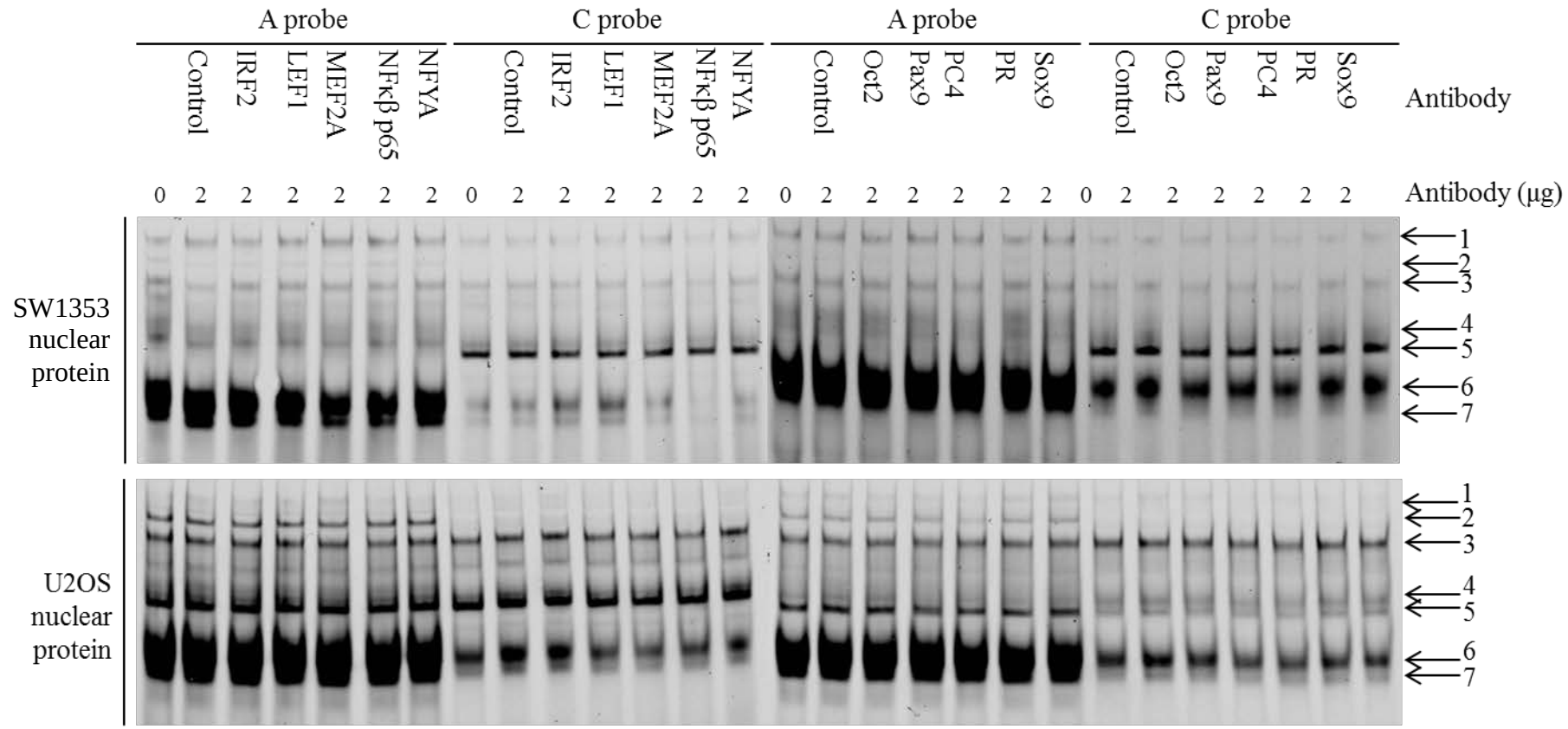

Figure F.17. Supershift EMSAs to investigate the transcription factors of SW1353 and U2OS nuclear extracts binding to the A and C alleles of rs10843013. The protein:probe mixes were incubated with 2 µg of antibody. Complexes 6 and 7 were outcompeted following incubation with the NFκβ p65 antibody with the SW1353 nuclear protein and the C allele probe. No changes in the banding patterns were observed after incubation with the remaining antibodies.

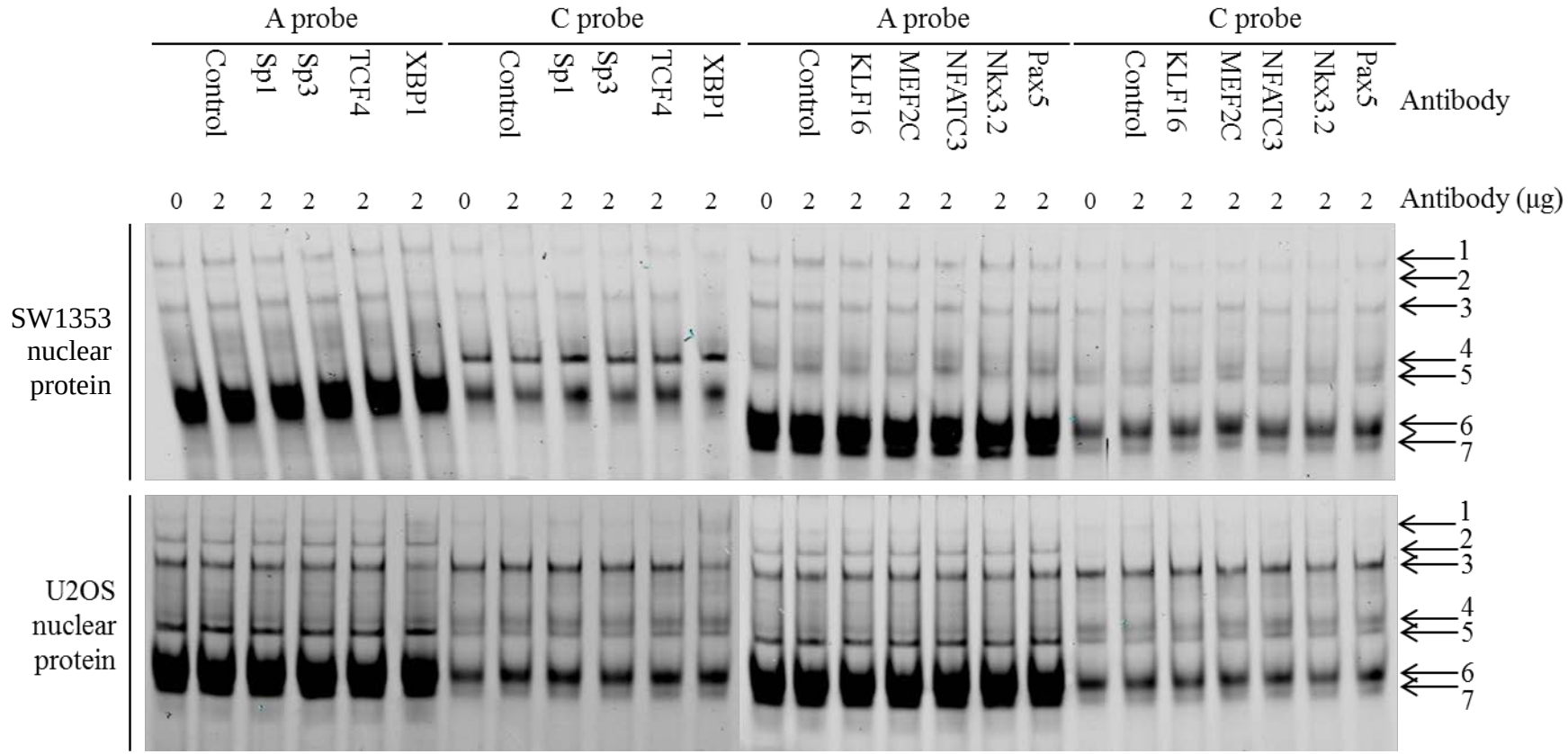

Figure F.18. Supershift EMSAs to investigate the transcription factors of SW1353 and U2OS nuclear extracts binding to the A and C alleles of rs10843013. The protein:probe mixes were incubated with 2 µg of antibody. Band 2 appeared to become fainter and band 1 more intense after incubation with the XBP1 antibody for both nuclear extracts and both allele probes. No changes in the banding patterns were observed after incubation with the remaining antibodies.

Presentations and Publications

Presentations

- March 2014: Arthritis Research UK Annual Fellows' Meeting, Loughborough, UK (*poster presentation*)
- April 2014: World Congress on Osteoarthritis, Paris, France (*poster presentation*)
- June 2014: Institute of Cellular Medicine Research Day, Newcastle University, UK (*poster presentation*)
- September 2014: UK-German Connective Tissue Meeting, Allendale, UK (*oral presentation*)
- October 2014: North East Postgraduate Conference, Newcastle, UK (*oral presentation*)
- March 2015: Arthritis Research UK Annual Fellows' Meeting, Loughborough, UK (*oral presentation*)
- October 2015: American Society of Human Genetics Annual Meeting, Baltimore, Maryland, USA (*poster presentation*)

Publications

- Johnson K., Reynard L. N. and Loughlin J. (2014) Functional analysis of the osteoarthritis susceptibility locus marked by the polymorphism rs10492367. *Osteoarthritis Cartilage*. **22**: S236-S237
- Johnson K., Reynard L. N. and Loughlin J. (2015) Functional characterisation of the osteoarthritis susceptibility locus at chromosome 6q14.1 marked by the polymorphism rs9350591. *BMC Med Genet.* **16**: 81
- Johnson K., Reynard L. N. and Loughlin J. (2015) Functional characterisation of the osteoarthritis susceptibility locus marked by the polymorphism rs10492367 at chromosome 12p11.22; programme number 853. Presented at the 65th Annual Meeting of The American Society of Human Genetics, 9th October 2015, Baltimore, MD
- Johnson K., Reynard L. N. and Loughlin J. The identification of *trans*-acting factors that regulate the expression of *PTHLH* via the osteoarthritis susceptibility SNP rs10492367 (*manuscript in preparation*)

References

Acharya, S., Rayborn, M. E. and Hollyfield, J. G. (1998a) Characterization of SPACR, a sialoprotein associated with cones and rods present in the interphotoreceptor matrix of the human retina: immunological and lectin binding analysis. *Glycobiology*. **8**: 997-1006

Acharya, S., Rodriguez, I. R., Moreira, E. F., Midura, R. J., Misono, K., Todres, E. and Hollyfield, J. G. (1998b) SPACR, a novel interphotoreceptor matrix glycoprotein in human retina that interacts with hyaluronan. *J Biol Chem*. **273**: 31599-606

Adams, J., Kelso, R. and Cooley, L. (2000) The kelch repeat superfamily of proteins: propellers of cell function. *Trends Cell Biol*. **10**: 17-24

Aerssens, J., Dequeker, J., Peeters, J., Breemans, S. and Boonen, S. (1998) Lack of association between osteoarthritis of the hip and gene polymorphisms of VDR, COL1A1, and COL2A1 in postmenopausal women. *Arthritis Rheum*. **41**: 1946-50

Aeterna Zentaris Inc. (2013) Aeterna Zentaris to Discontinue Phase 3 Trial in Multiple Myeloma with Perifosine Following Data Safety Monitoring Board Recommendation [Web Page]. Available at: http://www.aezsinc.com/en/page.php?p=60&q=550 (Accessed: 20th July 2015).

Agnesi, F., Amrami, K. K., Frigo, C. A. and Kaufman, K. R. (2008) Comparison of cartilage thickness with radiologic grade of knee osteoarthritis. *Skeletal Radiol*. **37**: 639-43

Ahituv, N., Sobe, T., Robertson, N. G., Morton, C. C., Taggart, R. T. and Avraham, K. B. (2000) Genomic structure of the human unconventional myosin VI gene. *Gene*. **261**: 269-75

Ahmed, Z. M., Morell, R. J., Riazuddin, S., Gropman, A., Shaukat, S., Ahmad, M. M., Mohiddin, S. A., *et al.* (2003) Mutations of MYO6 are associated with recessive deafness, DFNB37. *Am J Hum Genet*. **72**: 1315-22

Aigner, T., Soeder, S. and Haag, J. (2006) IL-1beta and BMPs--interactive players of cartilage matrix degradation and regeneration. *Eur Cell Mater*. **12**: 49-56; discussion 56

Akashi, H., Han, H. J., Iizaka, M., Nakajima, Y., Furukawa, Y., Sugano, S., Imai, K., *et al.* (2000) Isolation and characterization of a novel gene encoding a putative seven-span transmembrane protein, TM7SF3. *Cytogenet Cell Genet*. **88**: 305-9

Alegre, K. O. and Reverter, D. (2011) Swapping small ubiquitin-like modifier (SUMO) isoform specificity of SUMO proteases SENP6 and SENP7. *J Biol Chem*. **286**: 36142-51

Alford, J. W. and Cole, B. J. (2005) Cartilage restoration, part 1: basic science, historical perspective, patient evaluation, and treatment options. *Am J Sports Med*. **33**: 295-306

Allan, D. A. (1998) Structure and physiology of joints and their relationship to repetitive strain injuries. *Clin Orthop Relat Res*. 32-8

Alman, B. A. (2015) The role of hedgehog signalling in skeletal health and disease. *Nat Rev Rheumatol*. **11**: 552-60

Altman, R., Asch, E., Bloch, D., Bole, G., Borenstein, D., Brandt, K., Christy, W., *et al.* (1986) Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis Rheum*. **29**: 1039-49

Arai, K., Nagashima, Y., Takemoto, T. and Nishiyama, T. (2008) Mechanical strain increases expression of type XII collagen in murine osteoblastic MC3T3-E1 cells. *Cell Struct Funct*. **33**: 203-10

Archer, C. W., Dowthwaite, G. P. and Francis-West, P. (2003) Development of synovial joints. *Birth Defects Res C Embryo Today*. **69**: 144-55

arcOGEN Consortium, arcOGEN Collaborators, Zeggini, E., Panoutsopoulou, K., Southam, L., Rayner, N. W., Day-Williams, A. G., *et al.* (2012) Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet*. **380**: 815-23

Arden, N. and Nevitt, M. C. (2006) Osteoarthritis: epidemiology. *Best Pract Res Clin Rheumatol*. **20**: 3-25

Arnaudo, E., Hirano, M., Seelan, R. S., Milatovich, A., Hsieh, C. L., Fabrizi, G. M., Grossman, L. I., *et al.* (1992) Tissue-specific expression and chromosome assignment of genes specifying two isoforms of subunit VIIa of human cytochrome c oxidase. *Gene*. **119**: 299-305

Arroll, B. and Goodyear-Smith, F. (2004) Corticosteroid injections for osteoarthritis of the knee: meta-analysis. *BMJ*. **328**: 869

Arthritis Care (2012) OANation 2012 [Web Page]. Available at: http://www.arthritiscare.org.uk/LivingwithArthritis/oanation-2012 (Accessed: 25th June 2015).

Arthritis Research UK (2013) What is arthritis? [Web Page]. Available at: http://www.arthritisresearchuk.org/arthritis-information/conditions/arthritis/what-is-arthritis.aspx (Accessed: 24th June 2015).

Arthritis Research UK Primary Care Centre Keele University (2014) Musculoskeletal Matters Bulletins [Web Page]. Available at: http://www.keele.ac.uk/pchs/disseminatingourresearch/newslettersandresources/bulletins/ (Accessed: 24th June 2015).

Atlas of Genetics and Cytogenetics in Oncology and Haematology (2015) *PTHLH* (parathyroid hormone-like hormone) [Web Page]. Available at: http://atlasgeneticsoncology.org/Genes/PTHLHID41897ch12p11.html (Accessed: 13th July 2015).

Attardi, L. D., Von Seggern, D. and Tjian, R. (1993) Ectopic expression of wild-type or a dominant-negative mutant of transcription factor NTF-1 disrupts normal Drosophila development. *Proc Natl Acad Sci U S A*. **90**: 10563-7

Avraham, K. B., Hasson, T., Sobe, T., Balsara, B., Testa, J. R., Skvorak, A. B., Morton, C. C., *et al.* (1997) Characterization of unconventional MYO6, the human homologue of the gene responsible for deafness in Snell's waltzer mice. *Hum Mol Genet*. **6**: 1225-31

Avraham, K. B., Hasson, T., Steel, K. P., Kingsley, D. M., Russell, L. B., Mooseker, M. S., Copeland, N. G., *et al.* (1995) The mouse Snell's waltzer deafness gene encodes an unconventional myosin required for structural integrity of inner ear hair cells. *Nat Genet*. **11**: 369-75

Bader, H. L., Keene, D. R., Charvet, B., Veit, G., Driever, W., Koch, M. and Ruggiero, F. (2009) Zebrafish collagen XII is present in embryonic connective tissue sheaths (fascia) and basement membranes. *Matrix Biol*. **28**: 32-43

Bailey, A. J. and Mansell, J. P. (1997) Do subchondral bone changes exacerbate or precede articular cartilage destruction in osteoarthritis of the elderly? *Gerontology*. **43**: 296-304

Bailon-Plaza, A., Lee, A. O., Veson, E. C., Farnum, C. E. and van der Meulen, M. C. (1999) BMP-5 deficiency alters chondrocytic activity in the mouse proximal tibial growth plate. *Bone*. **24**: 211-6

Baron, R. (2000) 'Anatomy and Ultrastructure of Bone - Histogenesis, Growth and Remodeling', in L. J. De Groot, P. Beck-Peccoz, G. Chrousos, K. Dungan, A. Grossman, J. M. Hershman, C. Koch, R. McLachlan, M. New, R. Rebar, F. Singer, A. Vinik and M. O. Weickert (eds.) Endotext. South Dartmouth (MA).

Barter, M. J., Tselepi, M., Gomez, R., Woods, S., Hui, W., Smith, G. R., Shanley, D. P., *et al.* (2015) Genome-wide microRNA and gene analysis of mesenchymal stem cell chondrogenesis identifies an essential role and multiple targets for miR-140-5p. *Stem Cells*.

Batta, K., Yokokawa, M., Takeyasu, K. and Kundu, T. K. (2009) Human transcriptional coactivator PC4 stimulates DNA end joining and activates DSB repair activity. *J Mol Biol*. **385**: 788-99

Baumli, S., Hoeppner, S. and Cramer, P. (2005) A conserved mediator hinge revealed in the structure of the MED7.MED21 (Med7.Srb7) heterodimer. *J Biol Chem*. **280**: 18171-8

Beck, A., Isaac, R., Lavelin, I., Hart, Y., Volberg, T., Shatz-Azoulay, H., Geiger, B., *et al.* (2011) An siRNA screen identifies transmembrane 7 superfamily member 3 (TM7SF3), a seven transmembrane orphan receptor, as an inhibitor of cytokine-induced death of pancreatic beta cells. *Diabetologia*. **54**: 2845-55

Becker, K., Di Donato, N., Holder-Espinasse, M., Andrieux, J., Cuisset, J. M., Vallee, L., Plessis, G., *et al.* (2012) De novo microdeletions of chromosome 6q14.1-q14.3 and 6q12.1-q14.1 in two patients with intellectual disability - further delineation of the 6q14 microdeletion syndrome and review of the literature. *Eur J Med Genet*. **55**: 490-7

Bell, J. T., Pai, A. A., Pickrell, J. K., Gaffney, D. J., Pique-Regi, R., Degner, J. F., Gilad, Y., *et al.* (2011) DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol*. **12**: R10

Bendell, J. C., Ervin, T. J., Senzer, N. N., Richards, D. A., Firdaus, I., Lockhart, A. C., Cohn, A. L., *et al.* (2012) Results of the X-PECT study: A phase III randomized double-blind, placebo-controlled study of perifosine plus capecitabine (P-CAP) versus placebo plus capecitabine (CAP) in patients (pts) with refractory metastatic colorectal cancer (mCRC). *J Clin Oncol*. **30 suppl**: LBA3501

Benjamin, M. and Ralphs, J. R. (1998) Fibrocartilage in tendons and ligaments--an adaptation to compressive load. *J Anat*. **193 (Pt 4)**: 481-94

Bensen, W. G., Fiechtner, J. J., McMillen, J. I., Zhao, W. W., Yu, S. S., Woods, E. M., Hubbard, R. C., *et al.* (1999) Treatment of osteoarthritis with celecoxib, a cyclooxygenase-2 inhibitor: a randomized controlled trial. *Mayo Clin Proc*. **74**: 1095-105

Bertolotto, C., Lesueur, F., Giuliano, S., Strub, T., de Lichy, M., Bille, K., Dessen, P., *et al.* (2011) A SUMOylation-defective MITF germline mutation predisposes to melanoma and renal carcinoma. *Nature*. **480**: 94-8

Beynnon, B. D., Johnson, R. J., Abate, J. A., Fleming, B. C. and Nichols, C. E. (2005) Treatment of anterior cruciate ligament injuries, part I. *Am J Sports Med*. **33**: 1579-602

Bhang, S. H., Jeon, J. Y., La, W. G., Seong, J. Y., Hwang, J. W., Ryu, S. E. and Kim, B. S. (2011) Enhanced chondrogenic marker expression of human mesenchymal stem cells by interaction with both TGF-beta3 and hyaluronic acid. *Biotechnol Appl Biochem*. **58**: 271-6

Blagojevic, M., Jinks, C., Jeffery, A. and Jordan, K. P. (2010) Risk factors for onset of osteoarthritis of the knee in older adults: a systematic review and meta-analysis. *Osteoarthritis Cartilage*. **18**: 24-33

Bohme, K., Li, Y., Oh, P. S. and Olsen, B. R. (1995) Primary structure of the long and short splice variants of mouse collagen XII and their tissue-specific expression during embryonic development. *Dev Dyn*. **204**: 432-45

Bomer, N., den Hollander, W., Ramos, Y. F., Bos, S. D., van der Breggen, R., Lakenberg, N., Pepers, B. A., *et al.* (2015) Underlying molecular mechanisms of DIO2 susceptibility in symptomatic osteoarthritis. *Ann Rheum Dis*. **74**: 1571-9

Bonde, H. V., Talman, M. L. and Kofoed, H. (2005) The area of the tidemark in osteoarthritis--a three-dimensional stereological study in 21 patients. *APMIS*. **113**: 349-52

Bos, S. D., Bovee, J. V., Duijnisveld, B. J., Raine, E. V., van Dalen, W. J., Ramos, Y. F., van der Breggen, R., *et al.* (2012) Increased type II deiodinase protein in OA-affected cartilage and allelic imbalance of OA risk polymorphism rs225014 at DIO2 in human OA joint tissues. *Annals of the Rheumatic Diseases*. **71**: 1254-8

Bouizar, Z., Spyratos, F. and De vernejoul, M. C. (1999) The parathyroid hormone-related protein (PTHrP) gene: use of downstream TATA promotor and PTHrP 1-139 coding

pathways in primary breast cancers vary with the occurrence of bone metastasis. *J Bone Miner Res*. **14**: 406-14

Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., Karczewski, K. J., *et al.* (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. **22**: 1790-7

Buckland-Wright, C. (2004) Subchondral bone changes in hand and knee osteoarthritis detected by radiography. *Osteoarthritis Cartilage*. **12 Suppl A**: S10-9

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., Mankin, H. J. and Grodzinsky, A. J. (2005) Articular cartilage and osteoarthritis. *Instr Course Lect*. **54**: 465-80

Buckwalter, J. A., Mow, V. C. and Ratcliffe, A. (1994) Restoration of Injured or Degenerated Articular Cartilage. *J Am Acad Orthop Surg*. **2**: 192-201

Burtis, W. J., Wu, T., Bunch, C., Wysolmerski, J. J., Insogna, K. L., Weir, E. C., Broadus, A. E., *et al.* (1987) Identification of a novel 17,000-dalton parathyroid hormone-like adenylate cyclase-stimulating protein from a tumor associated with humoral hypercalcemia of malignancy. *J Biol Chem*. **262**: 7151-6

Bush, W. S. and Moore, J. H. (2012) Chapter 11: Genome-wide association studies. *PLoS Comput Biol*. **8**: e1002822

Buss, F., Arden, S. D., Lindsay, M., Luzio, J. P. and Kendrick-Jones, J. (2001) Myosin VI isoform localized to clathrin-coated vesicles with a role in clathrin-mediated endocytosis. *EMBO J*. **20**: 3676-84

Buss, F., Kendrick-Jones, J., Lionne, C., Knight, A. E., Cote, G. P. and Paul Luzio, J. (1998) The localization of myosin VI at the golgi complex and leading edge of fibroblasts and its phosphorylation and recruitment into membrane ruffles of A431 cells after growth factor stimulation. *J Cell Biol*. **143**: 1535-45

Carman, W. J., Sowers, M., Hawthorne, V. M. and Weissfeld, L. A. (1994) Obesity as a risk factor for osteoarthritis of the hand and wrist: a prospective study. *Am J Epidemiol*. **139**: 119-29

Chapman, K., Mustafa, Z., Irven, C., Carr, A. J., Clipsham, K., Smith, A., Chitnavis, J., *et al.* (1999) Osteoarthritis-susceptibility locus on chromosome 11q, detected by linkage. *Am J Hum Genet*. **65**: 167-74

Cheadle, C., Fan, J., Cho-Chung, Y. S., Werner, T., Ray, J., Do, L., Gorospe, M., *et al.* (2005) Stability regulation of mRNA and the control of gene expression. *Ann N Y Acad Sci*. **1058**: 196-204

Chen, F. H., Rousche, K. T. and Tuan, R. S. (2006a) Technology Insight: adult stem cells in cartilage regeneration and tissue engineering. *Nat Clin Pract Rheumatol*. **2**: 373-82

Chen, L., Xin, Z. C., Li, X., Tian, L., Yuan, Y. M., Liu, G., Jiang, X. J., *et al.* (2006b) Cox7a2 mediates steroidogenesis in TM3 mouse Leydig cells. *Asian J Androl*. **8**: 589-94

Chen, R., Brady, E. and McIntyre, T. M. (2011) Human TMEM30a promotes uptake of antitumor and bioactive choline phospholipids into mammalian cells. *J Immunol*. **186**: 3215-25

Ciarleglio, C. M., Ryckman, K. K., Servick, S. V., Hida, A., Robbins, S., Wells, N., Hicks, J., *et al.* (2008) Genetic differences in human circadian clock genes among worldwide populations. *J Biol Rhythms*. **23**: 330-40

Clark, A. R., Sawyer, G. M., Robertson, S. P. and Sutherland-Smith, A. J. (2009) Skeletal dysplasias due to filamin A mutations result from a gain-of-function mechanism distinct from allelic neurological disorders. *Hum Mol Genet*. **18**: 4791-800

Conesa, C. and Acker, J. (2010) Sub1/PC4 a chromatin associated protein with multiple functions in transcription. *RNA Biol*. **7**: 287-90

Cornils, H., Kohler, R. S., Hergovich, A. and Hemmings, B. A. (2011) Human NDR kinases control G(1)/S cell cycle transition by directly regulating p21 stability. *Mol Cell Biol*. **31**: 1382-95

Cornish, J., Callon, K. E., Lin, C., Xiao, C., Moseley, J. M. and Reid, I. R. (1999) Stimulation of osteoblast proliferation by C-terminal fragments of parathyroid hormone-related protein. *J Bone Miner Res*. **14**: 915-22

Cornish, J., Callon, K. E., Nicholson, G. C. and Reid, I. R. (1997) Parathyroid hormone-related protein-(107-139) inhibits bone resorption in vivo. *Endocrinology*. **138**: 1299-304

Corr, M. (2008) Wnt-beta-catenin signaling in the pathogenesis of osteoarthritis. *Nat Clin Pract Rheumatol*. **4**: 550-6

Croft, P., Cooper, C., Wickham, C. and Coggon, D. (1992) Osteoarthritis of the hip and occupational activity. *Scand J Work Environ Health*. **18**: 59-63

Cross, M., Smith, E., Hoy, D., Nolte, S., Ackerman, I., Fransen, M., Bridgett, L., *et al.* (2014) The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. *Ann Rheum Dis*. **73**: 1323-30

Cubukcu, D., Sarsan, A. and Alkan, H. (2012) Relationships between Pain, Function and Radiographic Findings in Osteoarthritis of the Knee: A Cross-Sectional Study. *Arthritis*. **2012**: 984060

Cummings, C. M., Bentley, C. A., Perdue, S. A., Baas, P. W. and Singer, J. D. (2009) The Cul3/Klhdc5 E3 ligase regulates p60/katanin and is required for normal mitosis in mammalian cells. *J Biol Chem*. **284**: 11663-75

Dai, H., Zhao, Y., Qian, C., Cai, M., Zhang, R., Chu, M., Dai, J., *et al.* (2013) Weighted SNP set analysis in genome-wide association study. *PLoS One*. **8**: e75897

Das, C., Hizume, K., Batta, K., Kumar, B. R., Gadad, S. S., Ganguly, S., Lorain, S., *et al.* (2006) Transcriptional coactivator PC4, a chromatin-associated protein, induces chromatin condensation. *Mol Cell Biol*. **26**: 8303-15

Davies, M. N., Volta, M., Pidsley, R., Lunnon, K., Dixit, A., Lovestone, S., Coarfa, C., *et al.* (2012) Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biol*. **13**: R43

De Bari, C., Dell'Accio, F., Tylzanowski, P. and Luyten, F. P. (2001) Multipotent mesenchymal stem cells from adult human synovial membrane. *Arthritis Rheum*. **44**: 1928-42

De Jager, P. L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L. C., Yu, L., Eaton, M. L., *et al.* (2014) Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat Neurosci*. **17**: 1156-63

de Sousa, E. B., Casado, P. L., Moura Neto, V., Duarte, M. E. and Aguiar, D. P. (2014) Synovial fluid and synovial membrane mesenchymal stem cells: latest discoveries and therapeutic perspectives. *Stem Cell Res Ther*. **5**: 112

Decker, R. S., Koyama, E. and Pacifici, M. (2014) Genesis and morphogenesis of limb synovial joints and articular cartilage. *Matrix Biol*. **39**: 5-10

den Hollander, W., Bos, S. D., Ramos, Y. F., Lakenberg, N., van der Breggen, R., Duijnisveld, B. J., Nelissen, R. G., *et al.* (2012) CpG sites of osteoarthritis susceptibility gene DIO2 are differentially methylated in arthritic compared to preserved cartilage. *Osteoarthritis Cartilage*. **20 Suppl**: S196

Deng, Z. L., Sharff, K. A., Tang, N., Song, W. X., Luo, J., Luo, X., Chen, J., *et al.* (2008) Regulation of osteogenic differentiation during skeletal development. *Front Biosci*. **13**: 2001-21

Department for Work and Pensions (2015) Tabulation Tool Disability Living Allowance - cases in payment [Web Page]. Available at: http://tabulation-tool.dwp.gov.uk/100pc/dla/nonscrpt/tabtool.html (Accessed: 25th June 2015).

Dhanoa, B. S., Cogliati, T., Satish, A. G., Bruford, E. A. and Friedman, J. S. (2013) Update on the Kelch-like (KLHL) gene family. *Hum Genomics*. **7**: 13

Dodd, A. W., Syddall, C. M. and Loughlin, J. (2013) A rare variant in the osteoarthritis-associated locus GDF5 is functional and reveals a site that can be manipulated to modulate GDF5 expression. *Eur J Hum Genet*. **21**: 517-21

Donahue, L. R., Chang, B., Mohan, S., Miyakoshi, N., Wergedal, J. E., Baylink, D. J., Hawes, N. L., *et al.* (2003) A missense mutation in the mouse Col2a1 gene causes spondyloepiphyseal dysplasia congenita, hearing loss, and retinoschisis. *J Bone Miner Res*. **18**: 1612-21

Dublet, B., Oh, S., Sugrue, S. P., Gordon, M. K., Gerecke, D. R., Olsen, B. R. and van der Rest, M. (1989) The structure of avian type XII collagen. Alpha 1 (XII) chains contain 190-kDa non-triple helical amino-terminal domains and form homotrimeric molecules. *J Biol Chem*. **264**: 13150-6

Dumond, H., Presle, N., Terlain, B., Mainard, D., Loeuille, D., Netter, P. and Pottie, P. (2003) Evidence for a key role of leptin in osteoarthritis. *Arthritis Rheum*. **48**: 3118-29

Egli, R. J., Southam, L., Wilkins, J. M., Lorenzen, I., Pombo-Suarez, M., Gonzalez, A., Carr, A., *et al.* (2009) Functional analysis of the osteoarthritis susceptibility-associated GDF5 regulatory polymorphism. *Arthritis Rheum*. **60**: 2055-64

Elsner, M., Rauser, S., Maier, S., Schone, C., Balluff, B., Meding, S., Jung, G., *et al.* (2012) MALDI imaging mass spectrometry reveals COX7A2, TAGLN2 and S100-A10 as novel prognostic markers in Barrett's adenocarcinoma. *J Proteomics*. **75**: 4693-704

Encode Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*. **489**: 57-74

Englund, M., Guermazi, A. and Lohmander, L. S. (2009) The meniscus in knee osteoarthritis. *Rheum Dis Clin North Am*. **35**: 579-90

Ensslen-Craig, S. E. and Brady-Kalnay, S. M. (2004) Receptor protein tyrosine phosphatases regulate neural development and axon guidance. *Dev Biol*. **275**: 12-22

Ernst, J. and Kellis, M. (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol*. **28**: 817-25

Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., Zhang, X., *et al.* (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. **473**: 43-9

Esbrit, P. and Alcaraz, M. J. (2013) Current perspectives on parathyroid hormone (PTH) and PTH-related protein (PTHrP) as bone anabolic therapies. *Biochem Pharmacol*. **85**: 1417-23

Esko, J. D., Kimata, K. and Lindahl, U. (2009) 'Proteoglycans and Sulfated Glycosaminoglycans', in A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart and M. E. Etzler (eds.) Essentials of Glycobiology. 2nd edn. Cold Spring Harbor (NY).

Esteller, M., Garcia-Foncillas, J., Andion, E., Goodman, S. N., Hidalgo, O. F., Vanaclocha, V., Baylin, S. B., *et al.* (2000) Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents. *N Engl J Med*. **343**: 1350-4

Estrada, K., Styrkarsdottir, U., Evangelou, E., Hsu, Y. H., Duncan, E. L., Ntzani, E. E., Oei, L., *et al.* (2012) Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat Genet*. **44**: 491-501

Evangelou, E., Valdes, A. M., Kerkhof, H. J., Styrkarsdottir, U., Zhu, Y., Meulenbelt, I., Lories, R. J., *et al.* (2011) Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. *Ann Rheum Dis*. **70**: 349-55

Farooq, M., Nakai, H., Fujimoto, A., Fujikawa, H., Kjaer, K. W., Baig, S. M. and Shimomura, Y. (2013) Characterization of a novel missense mutation in the prodomain of GDF5,

which underlies brachydactyly type C and mild Grebe type chondrodysplasia in a large Pakistani family. *Hum Genet.* **132**: 1253-64

Farre, D., Roset, R., Huerta, M., Adsuara, J. E., Rosello, L., Alba, M. M. and Messeguer, X. (2003) Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Res.* **31**: 3651-3

Fehrmann, R. S., Jansen, R. C., Veldink, J. H., Westra, H. J., Arends, D., Bonder, M. J., Fu, J., *et al.* (2011) Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet.* **7**: e1002197

Fei, W., Shui, G., Zhang, Y., Krahmer, N., Ferguson, C., Kapterian, T. S., Lin, R. C., *et al.* (2011) A role for phosphatidic acid in the formation of "supersized" lipid droplets. *PLoS Genet.* **7**: e1002201

Felbor, U., Gehrig, A., Sauer, C. G., Marquardt, A., Kohler, M., Schmid, M. and Weber, B. H. (1998) Genomic organization and chromosomal localization of the interphotoreceptor matrix proteoglycan-1 (IMPG1) gene: a candidate for 6q-linked retinopathies. *Cytogenet Cell Genet.* **81**: 12-7

Felson, D. T. (2006) Clinical practice. Osteoarthritis of the knee. *N Engl J Med.* **354**: 841-8

Felson, D. T., Gale, D. R., Elon Gale, M., Niu, J., Hunter, D. J., Goggins, J. and Lavalley, M. P. (2005) Osteophytes and progression of knee osteoarthritis. *Rheumatology (Oxford).* **44**: 100-4

Felson, D. T., Lawrence, R. C., Dieppe, P. A., Hirsch, R., Helmick, C. G., Jordan, J. M., Kington, R. S., *et al.* (2000) Osteoarthritis: new insights. Part 1: the disease and its risk factors. *Ann Intern Med.* **133**: 635-46

Feng, Y. and Walsh, C. A. (2004) The many faces of filamin: a versatile molecular scaffold for cell motility and signalling. *Nat Cell Biol.* **6**: 1034-8

Fernandez-Moreno, M., Rego, I., Carreira-Garcia, V. and Blanco, F. J. (2008) Genetics in osteoarthritis. *Curr Genomics.* **9**: 542-7

Fernandez-Tajes, J., Soto-Hermida, A., Vazquez-Mosquera, M. E., Cortes-Pereira, E., Mosquera, A., Fernandez-Moreno, M., Oreiro, N., *et al.* (2014) Genome-wide DNA methylation analysis of articular chondrocytes reveals a cluster of osteoarthritic patients. *Ann Rheum Dis.* **73**: 668-77

Fiaschi-Taesch, N. M. and Stewart, A. F. (2003) Minireview: parathyroid hormone-related protein as an intracrine factor--trafficking mechanisms and functional consequences. *Endocrinology.* **144**: 407-11

Finnson, K. W., Chi, Y., Bou-Gharios, G., Leask, A. and Philip, A. (2012) TGF-b signaling in cartilage homeostasis and osteoarthritis. *Front Biosci (Schol Ed).* **4**: 251-68

Folmer, D. E., Mok, K. S., de Wee, S. W., Duijst, S., Hiralall, J. K., Seppen, J., Oude Elferink, R. P., *et al.* (2012) Cellular localization and biochemical analysis of

mammalian CDC50A, a glycosylated beta-subunit for P4 ATPases. *J Histochem Cytochem*. **60**: 205-18

Fox, J. W., Lamperti, E. D., Eksioglu, Y. Z., Hong, S. E., Feng, Y., Graham, D. A., Scheffer, I. E., *et al.* (1998) Mutations in filamin 1 prevent migration of cerebral cortical neurons in human periventricular heterotopia. *Neuron*. **21**: 1315-25

Frank, D. J., Noguchi, T. and Miller, K. G. (2004) Myosin VI: a structural role in actin organization important for protein and organelle localization and trafficking. *Curr Opin Cell Biol*. **16**: 189-94

Franke, L. and Jansen, R. C. (2009) eQTL analysis in humans. *Methods Mol Biol*. **573**: 311-28

Fullwood, M. J., Han, Y., Wei, C. L., Ruan, X. and Ruan, Y. (2010) Chromatin interaction analysis using paired-end tag sequencing. *Curr Protoc Mol Biol*. **Chapter 21**: Unit 21 15 1-25

Furukawa, M., He, Y. J., Borchers, C. and Xiong, Y. (2003) Targeting of protein ubiquitination by BTB-Cullin 3-Roc1 ubiquitin ligases. *Nat Cell Biol*. **5**: 1001-7

Gaffney, D. J., Veyrieras, J. B., Degner, J. F., Pique-Regi, R., Pai, A. A., Crawford, G. E., Stephens, M., *et al.* (2012) Dissecting the regulatory architecture of gene expression QTLs. *Genome Biol*. **13**: R7

Gambari, R. (2011) Recent patents on therapeutic applications of the transcription factor decoy approach. *Expert Opin Ther Pat*. **21**: 1755-71

Gavrilov, A., Eivazova, E., Priozhkova, I., Lipinski, M., Razin, S. and Vassetzky, Y. (2009) Chromosome conformation capture (from 3C to 5C) and its ChIP-based modification. *Methods Mol Biol*. **567**: 171-88

Gee, F., Clubbs, C. F., Raine, E. V., Reynard, L. N. and Loughlin, J. (2014) Allelic expression analysis of the osteoarthritis susceptibility locus that maps to chromosome 3p21 reveals cis-acting eQTLs at GNL3 and SPCS1. *BMC Med Genet*. **15**: 53

Gee, F., Rushton, M. D., Loughlin, J. and Reynard, L. N. (2015) Correlation of the Osteoarthritis Susceptibility Variants That Map to Chromosome 20q13 With an Expression Quantitative Trait Locus Operating on NCOA3 and With Functional Variation at the Polymorphism rs116855380. *Arthritis Rheumatol*. **67**: 2923-32

Gehrig, A., Felbor, U., Kelsell, R. E., Hunt, D. M., Maumenee, I. H. and Weber, B. H. (1998) Assessment of the interphotoreceptor matrix proteoglycan-1 (IMPG1) gene localised to 6q13-q15 in autosomal dominant Stargardt-like disease (ADSTGD), progressive bifocal chorioretinal atrophy (PBCRA), and North Carolina macular dystrophy (MCDR1). *J Med Genet*. **35**: 641-5

Geisbrecht, E. R. and Montell, D. J. (2002) Myosin VI is required for E-cadherin-mediated border cell migration. *Nat Cell Biol*. **4**: 616-20

Geiss-Friedlander, R. and Melchior, F. (2007) Concepts in sumoylation: a decade on. *Nat Rev Mol Cell Biol*. **8**: 947-56

Gelse, K., Ekici, A. B., Cipa, F., Swoboda, B., Carl, H. D., Olk, A., Hennig, F. F., *et al.* (2012) Molecular differentiation between osteophytic and articular cartilage--clues for a transient and permanent chondrocyte phenotype. *Osteoarthritis Cartilage*. **20**: 162-71

Gerecke, D. R., Olson, P. F., Koch, M., Knoll, J. H., Taylor, R., Hudson, D. L., Champliaud, M. F., *et al.* (1997) Complete primary structure of two splice variants of collagen XII, and assignment of alpha 1(XII) collagen (COL12A1), alpha 1(IX) collagen (COL9A1), and alpha 1(XIX) collagen (COL19A1) to human chromosome 6q12-q13. *Genomics*. **41**: 236-42

Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K. K., Cheng, C., Mu, X. J., *et al.* (2012) Architecture of the human regulatory network derived from ENCODE data. *Nature*. **489**: 91-100

Gil-Bernabe, P., D'Alessandro-Gabazza, C. N., Toda, M., Boveda Ruiz, D., Miyake, Y., Suzuki, T., Onishi, Y., *et al.* (2012) Exogenous activated protein C inhibits the progression of diabetic nephropathy. *J Thromb Haemost*. **10**: 337-46

Glish, G. L. and Vachet, R. W. (2003) The basics of mass spectrometry in the twenty-first century. *Nat Rev Drug Discov*. **2**: 140-50

Goldring, M. B. and Goldring, S. R. (2010) Articular cartilage and subchondral bone in the pathogenesis of osteoarthritis. *Ann N Y Acad Sci*. **1192**: 230-7

Grand, E. K., Grand, F. H., Chase, A. J., Ross, F. M., Corcoran, M. M., Oscier, D. G. and Cross, N. C. (2004) Identification of a novel gene, FGFR1OP2, fused to FGFR1 in 8p11 myeloproliferative syndrome. *Genes Chromosomes Cancer*. **40**: 78-83

Graneli, C., Thorfve, A., Ruetschi, U., Brisby, H., Thomsen, P., Lindahl, A. and Karlsson, C. (2014) Novel markers of osteogenic and adipogenic differentiation of human bone marrow stromal cells identified using a quantitative proteomics approach. *Stem Cell Res*. **12**: 153-65

Grogan, S. P., Duffy, S. F., Pauli, C., Koziol, J. A., Su, A. I., D'Lima, D. D. and Lotz, M. K. (2013) Zone-specific gene expression patterns in articular cartilage. *Arthritis Rheum*. **65**: 418-28

Guex, N. and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. **18**: 2714-23

Hall, E., Volkov, P., Dayeh, T., Esguerra, J. L., Salo, S., Eliasson, L., Ronn, T., *et al.* (2014) Sex differences in the genome-wide DNA methylation pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets. *Genome Biol*. **15**: 522

Hannan, M. T., Felson, D. T. and Pincus, T. (2000) Analysis of the discordance between radiographic changes and knee pain in osteoarthritis of the knee. *J Rheumatol*. **27**: 1513-7

Hatzis, P., van der Flier, L. G., van Driel, M. A., Guryev, V., Nielsen, F., Denissov, S., Nijman, I. J., *et al.* (2008) Genome-wide pattern of TCF7L2/TCF4 chromatin occupancy in colorectal cancer cells. *Mol Cell Biol*. **28**: 2732-44

Hegan, P. S., Giral, H., Levi, M. and Mooseker, M. S. (2012) Myosin VI is required for maintenance of brush border structure, composition, and membrane trafficking functions in the intestinal epithelial cell. *Cytoskeleton (Hoboken)*. **69**: 235-51

Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A. E., Kel, O. V., Ignatieva, E. V., *et al.* (1998) Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. *Nucleic Acids Res*. **26**: 362-7

Hirai, T., Chagin, A. S., Kobayashi, T., Mackem, S. and Kronenberg, H. M. (2011) Parathyroid hormone/parathyroid hormone-related protein receptor signaling is required for maintenance of the growth plate in postnatal life. *Proc Natl Acad Sci U S A*. **108**: 191-6

Hojo, H., Ohba, S., Yano, F. and Chung, U. I. (2010) Coordination of chondrogenesis and osteogenesis by hypertrophic chondrocytes in endochondral bone development. *J Bone Miner Metab*. **28**: 489-502

Hui, W., Rowan, A. D. and Cawston, T. (2001) Modulation of the expression of matrix metalloproteinase and tissue inhibitors of metalloproteinases by TGF-beta1 and IGF-1 in primary human articular and bovine nasal chondrocytes stimulated with TNF-alpha. *Cytokine*. **16**: 31-5

International HapMap Consortium (2003) The International HapMap Project. *Nature*. **426**: 789-96

International HapMap Consortium (2005) A haplotype map of the human genome. *Nature*. **437**: 1299-320

Ishikawa, M., Sawada, Y. and Yoshitomi, T. (2015) Structure and function of the interphotoreceptor matrix surrounding retinal photoreceptor cells. *Exp Eye Res*. **133**: 3-18

Izu, Y., Sun, M., Zwolanek, D., Veit, G., Williams, V., Cha, B., Jepsen, K. J., *et al.* (2011) Type XII collagen regulates osteoblast polarity and communication during bone formation. *J Cell Biol*. **193**: 1115-30

Jerosch, J. (2011) Effects of Glucosamine and Chondroitin Sulfate on Cartilage Metabolism in OA: Outlook on Other Nutrient Partners Especially Omega-3 Fatty Acids. *Int J Rheumatol*. **2011**: 969012

Jiang, J., Leong, N. L., Mung, J. C., Hidaka, C. and Lu, H. H. (2008) Interaction between zonal populations of articular chondrocytes suppresses chondrocyte mineralization and this process is mediated by PTHrP. *Osteoarthritis Cartilage*. **16**: 70-82

Jobert, A. S., Zhang, P., Couvineau, A., Bonaventure, J., Roume, J., Le Merrer, M. and Silve, C. (1998) Absence of functional receptors for parathyroid hormone and parathyroid hormone-related peptide in Blomstrand chondrodysplasia. *J Clin Invest*. **102**: 34-40

Jodoin, J. N., Shboul, M., Sitaram, P., Zein-Sabatto, H., Reversade, B., Lee, E. and Lee, L. A. (2012) Human Asunder promotes dynein recruitment and centrosomal tethering to the nucleus at mitotic entry. *Mol Biol Cell*. **23**: 4713-24

Johnson, A. D., Handsaker, R. E., Pulit, S. L., Nizzari, M. M., O'Donnell, C. J. and de Bakker, P. I. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*. **24**: 2938-9

Jones, P. L., Veenstra, G. J., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., *et al.* (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet*. **19**: 187-91

Junker, S., Krumbholz, G., Frommer, K. W., Rehart, S., Steinmeyer, J., Rickert, M., Schett, G., *et al.* (2015) Differentiation of osteophyte types in osteoarthritis - proposal of a histological classification. *Joint Bone Spine*.

Kalhammer, G. and Bahler, M. (2000) Unconventional myosins. *Essays Biochem*. **35**: 33-42

Karagiannis, G. S., Petraki, C., Prassas, I., Saraon, P., Musrap, N., Dimitromanolakis, A. and Diamandis, E. P. (2012) Proteomic signatures of the desmoplastic invasion front reveal collagen type XII as a marker of myofibroblastic differentiation during colorectal cancer metastasis. *Oncotarget*. **3**: 267-85

Karlsson, C., Dehne, T., Lindahl, A., Brittberg, M., Pruss, A., Sittinger, M. and Ringe, J. (2010) Genome-wide expression profiling reveals new candidate genes associated with osteoarthritis. *Osteoarthritis Cartilage*. **18**: 581-92

Kato, U., Inadome, H., Yamamoto, M., Emoto, K., Kobayashi, T. and Umeda, M. (2013) Role for phospholipid flippase complex of ATP8A1 and CDC50A proteins in cell migration. *J Biol Chem*. **288**: 4922-34

Katoh, Y. and Katoh, M. (2004) Identification and characterization of CDC50A, CDC50B and CDC50C genes in silico. *Oncol Rep*. **12**: 939-43

Kellgren, J. H. and Lawrence, J. S. (1957) Radiological assessment of osteo-arthrosis. *Ann Rheum Dis*. **16**: 494-502

Kellgren, J. H., Lawrence, J. S. and Bier, F. (1963) Genetic Factors in Generalized Osteo-Arthrosis. *Ann Rheum Dis*. **22**: 237-55

Kemaladewi, D. U., Pasteuning, S., van der Meulen, J. W., van Heiningen, S. H., van Ommen, G. J., Ten Dijke, P., Aartsma-Rus, A., *et al.* (2014) Targeting TGF-beta Signaling by Antisense Oligonucleotide-mediated Knockdown of TGF-beta Type I Receptor. *Mol Ther Nucleic Acids*. **3**: e156

Kemp, J. P., Medina-Gomez, C., Estrada, K., St Pourcain, B., Heppe, D. H., Warrington, N. M., Oei, L., *et al.* (2014) Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. *PLoS Genet*. **10**: e1004423

Kent, W. J. (2002) BLAT--the BLAST-like alignment tool. *Genome Res*. **12**: 656-64

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res*. **12**: 996-1006

Kerkhof, H. J., Lories, R. J., Meulenbelt, I., Jonsdottir, I., Valdes, A. M., Arp, P., Ingvarsson, T., *et al.* (2010) A genome-wide association study identifies an osteoarthritis susceptibility locus on chromosome 7q22. *Arthritis Rheum*. **62**: 499-510

Kessler, J. D., Kahle, K. T., Sun, T., Meerbrey, K. L., Schlabach, M. R., Schmitt, E. M., Skinner, S. O., *et al.* (2012) A SUMOylation-dependent transcriptional subprogram is required for Myc-driven tumorigenesis. *Science*. **335**: 348-53

Kho, C., Lee, A., Jeong, D., Oh, J. G., Chaanine, A. H., Kizana, E., Park, W. J., *et al.* (2011) SUMO1-dependent modulation of SERCA2a in heart failure. *Nature*. **477**: 601-5

Kim, I. F., Mohammadi, E. and Huang, R. C. (1999) Isolation and characterization of IPP, a novel human gene encoding an actin-binding, kelch-like protein. *Gene*. **228**: 73-83

Klomp, L. W., Vargas, J. C., van Mil, S. W., Pawlikowska, L., Strautnieks, S. S., van Eijk, M. J., Juijn, J. A., *et al.* (2004) Characterization of mutations in ATP8B1 associated with hereditary cholestasis. *Hepatology*. **40**: 27-38

Klug, M. and Rehli, M. (2006) Functional analysis of promoter CpG methylation using a CpG-free luciferase reporter vector. *Epigenetics*. **1**: 127-30

Knight, J. C. (2005) Regulatory polymorphisms underlying complex disease traits. *J Mol Med (Berl)*. **83**: 97-109

Knudson, C. B. and Knudson, W. (2001) Cartilage proteoglycans. *Semin Cell Dev Biol*. **12**: 69-78

Kobayashi, H., Hirata, M., Saito, T., Itoh, S., Chung, U. I. and Kawaguchi, H. (2013) Transcriptional induction of ADAMTS5 protein by nuclear factor-kappaB (NF-kappaB) family member RelA/p65 in chondrocytes during osteoarthritis development. *J Biol Chem*. **288**: 28620-9

Kovacs, C. S., Lanske, B., Hunzelman, J. L., Guo, J., Karaplis, A. C. and Kronenberg, H. M. (1996) Parathyroid hormone-related peptide (PTHrP) regulates fetal-placental calcium transport through a receptor distinct from the PTH/PTHrP receptor. *Proc Natl Acad Sci U S A*. **93**: 15233-8

Kriajevska, M., Fischer-Larsen, M., Moertz, E., Vorm, O., Tulchinsky, E., Grigorian, M., Ambartsumian, N., *et al.* (2002) Liprin beta 1, a member of the family of LAR transmembrane tyrosine phosphatase-interacting proteins, is a new target for the metastasis-associated protein S100A4 (Mts1). *J Biol Chem*. **277**: 5229-35

Kuehn, M. H. and Hageman, G. S. (1999) Expression and characterization of the IPM 150 gene (IMPG1) product, a novel human photoreceptor cell-associated chondroitin-sulfate proteoglycan. *Matrix Biol*. **18**: 509-18

Kurmasheva, R. T., Peterson, C. A., Parham, D. M., Chen, B., McDonald, R. E. and Cooney, C. A. (2005) Upstream CpG island methylation of the PAX3 gene in human rhabdomyosarcomas. *Pediatr Blood Cancer*. **44**: 328-37

Kuroda, T., Matsumoto, T., Mifune, Y., Fukui, T., Kubo, S., Matsushita, T., Asahara, T., *et al.* (2011) Therapeutic strategy of third-generation autologous chondrocyte implantation for osteoarthritis. *Ups J Med Sci*. **116**: 107-14

Lambrinoudaki, I., Kaparos, G., Armeni, E., Alexandrou, A., Damaskos, C., Logothetis, E., Creatsa, M., *et al.* (2011) BsmI vitamin D receptor's polymorphism and bone mineral density in men and premenopausal women on long-term antiepileptic therapy. *Eur J Neurol*. **18**: 93-8

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., Willer, C. J., *et al.* (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*. **467**: 832-8

Lawrence, R. C., Felson, D. T., Helmick, C. G., Arnold, L. M., Choi, H., Deyo, R. A., Gabriel, S., *et al.* (2008) Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. *Arthritis Rheum*. **58**: 26-35

Leder, B. Z., O'Dea, L. S., Zanchetta, J. R., Kumar, P., Banks, K., McKay, K., Lyttle, C. R., *et al.* (2015) Effects of abaloparatide, a human parathyroid hormone-related peptide analog, on bone mineral density in postmenopausal women with osteoporosis. *J Clin Endocrinol Metab*. **100**: 697-706

Lee, J. W., Chen, Q., Rayborn, M. E., Shadrach, K. G., Crabb, J. W., Rodriguez, I. R. and Hollyfield, J. G. (2000) SPACR in the interphotoreceptor matrix of the mouse retina: molecular, biochemical and immunohistochemical characterization. *Exp Eye Res*. **71**: 341-52

Levano, K., Punia, V., Raghunath, M., Debata, P. R., Curcio, G. M., Mogha, A., Purkayastha, S., *et al.* (2012) Atp8a1 deficiency is associated with phosphatidylserine externalization in hippocampus and delayed hippocampus-dependent learning. *J Neurochem*. **120**: 302-13

Li, G., Yin, J., Gao, J., Cheng, T. S., Pavlos, N. J., Zhang, C. and Zheng, M. H. (2013) Subchondral bone in osteoarthritis: insight into risk factors and microstructural changes. *Arthritis Res Ther*. **15**: 223

Li, N. (2010) Critical Role of Transcription Cofactor PC4 in Mammals. Ph.D thesis. Ludwig-Maximilians-Universität München: Fakultät für Chemie und Pharmazie.

Libioulle, C., Louis, E., Hansoul, S., Sandor, C., Farnir, F., Franchimont, D., Vermeire, S., *et al.* (2007) Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet*. **3**: e58

Lievense, A. M., Bierma-Zeinstra, S. M., Verhagen, A. P., van Baar, M. E., Verhaar, J. A. and Koes, B. W. (2002) Influence of obesity on the development of osteoarthritis of the hip: a systematic review. *Rheumatology (Oxford)*. **41**: 1155-62

Lima, C. D. and Reverter, D. (2008) Structure of the human SENP7 catalytic domain and poly-SUMO deconjugation activities for SENP6 and SENP7. *J Biol Chem*. **283**: 32045-55

Lin, A., Hokugo, A., Choi, J. and Nishimura, I. (2010) Small cytoskeleton-associated molecule, fibroblast growth factor receptor 1 oncogene partner 2/wound inducible transcript-3.0 (FGFR1OP2/wit3.0), facilitates fibroblast-driven wound closure. *Am J Pathol*. **176**: 108-21

Lin, E. H. (2008) Depression and osteoarthritis. *Am J Med*. **121**: S16-9

Lippuner, K. (2012) The future of osteoporosis treatment - a research update. *Swiss Med Wkly*. **142**: w13624

Little, A. G., Kocha, K. M., Lougheed, S. C. and Moyes, C. D. (2010) Evolution of the nuclear-encoded cytochrome oxidase subunits in vertebrates. *Physiol Genomics*. **42**: 76-84

Little, G. H., Noushmehr, H., Baniwal, S. K., Berman, B. P., Coetzee, G. A. and Frenkel, B. (2012) Genome-wide Runx2 occupancy in prostate cancer cells suggests a role in regulating secretion. *Nucleic Acids Res*. **40**: 3538-47

Liu, C., Chen, J., Guo, Y., Yang, L., Zhao, C. and Bai, L. (2011) The expression of PTHLH in human gastric mucosa enterochromaffin-like cells. *Dig Dis Sci*. **56**: 993-8

Lloyd-Roberts, G. C. (1953) The role of capsular changes in osteoarthritis of the hip joint. *J Bone Joint Surg Br*. **35-B**: 627-42

Loeser, R. F. (2009) Aging and osteoarthritis: the role of chondrocyte senescence and aging changes in the cartilage matrix. *Osteoarthritis Cartilage*. **17**: 971-9

Lohmander, L. S., Englund, P. M., Dahl, L. L. and Roos, E. M. (2007) The long-term consequence of anterior cruciate ligament and meniscus injuries: osteoarthritis. *Am J Sports Med*. **35**: 1756-69

Lokk, K., Modhukur, V., Rajashekar, B., Martens, K., Magi, R., Kolde, R., Koltsina, M., *et al.* (2014) DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome Biol*. **15**: r54

Loughlin, J., Dowling, B., Mustafa, Z. and Chapman, K. (2002a) Association of the interleukin-1 gene cluster on chromosome 2q13 with knee osteoarthritis. *Arthritis Rheum*. **46**: 1519-27

Loughlin, J., Irven, C., Fergusson, C. and Sykes, B. (1994) Sibling pair analysis shows no linkage of generalized osteoarthritis to the loci encoding type II collagen, cartilage link protein or cartilage matrix protein. *Br J Rheumatol*. **33**: 1103-6

Loughlin, J., Mustafa, Z., Dowling, B., Southam, L., Marcelline, L., Raina, S. S., Ala-Kokko, L., *et al.* (2002b) Finer linkage mapping of a primary hip osteoarthritis susceptibility locus on chromosome 6. *Eur J Hum Genet*. **10**: 562-8

Loughlin, J., Mustafa, Z., Irven, C., Smith, A., Carr, A. J., Sykes, B. and Chapman, K. (1999) Stratification analysis of an osteoarthritis genome screen-suggestive linkage to chromosomes 4, 6, and 16. *Am J Hum Genet*. **65**: 1795-8

Ma, B., Zhong, L., van Blitterswijk, C. A., Post, J. N. and Karperien, M. (2013) T cell factor 4 is a pro-catabolic and apoptotic factor in human articular chondrocytes by potentiating nuclear factor kappaB signaling. *J Biol Chem*. **288**: 17552-8

Maass, P. G., Wirth, J., Aydin, A., Rump, A., Stricker, S., Tinschert, S., Otero, M., *et al.* (2010) A cis-regulatory site downregulates PTHLH in translocation t(8;12)(q13;p11.2) and leads to Brachydactyly Type E. *Hum Mol Genet*. **19**: 848-60

Macgregor, P. F. and Squire, J. A. (2002) Application of microarrays to the analysis of gene expression in cancer. *Clin Chem*. **48**: 1170-7

Mackie, E. J., Ahmed, Y. A., Tatarczuch, L., Chen, K. S. and Mirams, M. (2008) Endochondral ossification: how cartilage is converted into bone in the developing skeleton. *Int J Biochem Cell Biol*. **40**: 46-62

Madry, H., van Dijk, C. N. and Mueller-Gerbl, M. (2010) The basic science of the subchondral bone. *Knee Surg Sports Traumatol Arthrosc*. **18**: 419-33

Maleki-Fischbach, M. and Jordan, J. M. (2010) New developments in osteoarthritis. Sex differences in magnetic resonance imaging-based biomarkers and in those of joint metabolism. *Arthritis Res Ther*. **12**: 212

Manaster, B. J. (2000) From the RSNA Refresher Courses. Radiological Society of North America. Adult chronic hip pain: radiographic evaluation. *Radiographics*. **20 Spec No**: S3-S25

Manes, G., Meunier, I., Avila-Fernandez, A., Banfi, S., Le Meur, G., Zanlonghi, X., Corton, M., *et al.* (2013) Mutations in IMPG1 cause vitelliform macular dystrophies. *Am J Hum Genet*. **93**: 571-8

Mann, M. J. (2005) Transcription factor decoys: a new model for disease intervention. *Ann N Y Acad Sci*. **1058**: 128-39

Mannstadt, M., Juppner, H. and Gardella, T. J. (1999) Receptors for PTH and PTHrP: their biological importance and functional properties. *Am J Physiol*. **277**: F665-75

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**: 351-84

Martin, J. A., Martini, A., Molinari, A., Morgan, W., Ramalingam, W., Buckwalter, J. A. and McKinley, T. O. (2012) Mitochondrial electron transport and glycolysis are coupled in articular cartilage. *Osteoarthritis Cartilage*. **20**: 323-9

Matthews, B. G., Naot, D., Callon, K. E., Musson, D. S., Locklin, R., Hulley, P. A., Grey, A., *et al.* (2014) Enhanced osteoblastogenesis in three-dimensional collagen gels. *Bonekey Rep*. **3**: 560

Messeguer, X., Escudero, R., Farre, D., Nunez, O., Martinez, J. and Alba, M. M. (2002) PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics*. **18**: 333-4

Messner, K. and Gao, J. (1998) The menisci of the knee joint. Anatomical and functional characteristics, and a rationale for clinical treatment. *J Anat*. **193 (Pt 2)**: 161-78

Miao, D., He, B., Karaplis, A. C. and Goltzman, D. (2002) Parathyroid hormone is essential for normal fetal bone formation. *J Clin Invest*. **109**: 1173-82

Milligan, J. N. and Jolly, E. R. (2012) Identification and characterization of a Mef2 transcriptional activator in schistosome parasites. *PLoS Neglected Tropical Diseases*. **6**: e1443

Miyamoto, Y., Mabuchi, A., Shi, D., Kubo, T., Takatori, Y., Saito, S., Fujioka, M., *et al.* (2007) A functional polymorphism in the 5' UTR of GDF5 is associated with susceptibility to osteoarthritis. *Nat Genet*. **39**: 529-33

Moffatt, M. F., Kabesch, M., Liang, L., Dixon, A. L., Strachan, D., Heath, S., Depner, M., *et al.* (2007) Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature*. **448**: 470-3

Mondal, A. K., Sharma, N. K., Elbein, S. C. and Das, S. K. (2013) Allelic expression imbalance screening of genes in chromosome 1q21-24 region to identify functional variants for Type 2 diabetes susceptibility. *Physiol Genomics*. **45**: 509-20

Montgomery, S. B. and Dermitzakis, E. T. (2011) From expression QTLs to personalized transcriptomics. *Nat Rev Genet*. **12**: 277-82

Morris, A. P., Voight, B. F., Teslovich, T. M., Ferreira, T., Segre, A. V., Steinthorsdottir, V., Strawbridge, R. J., *et al.* (2012) Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet*. **44**: 981-90

Moseley, J. B., O'Malley, K., Petersen, N. J., Menke, T. J., Brody, B. A., Kuykendall, D. H., Hollingsworth, J. C., *et al.* (2002) A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med*. **347**: 81-8

Mukhopadhyay, D., Arnaoutov, A. and Dasso, M. (2010) The SUMO protease SENP6 is essential for inner kinetochore assembly. *J Cell Biol*. **188**: 681-92

Mundy, G. R. and Edwards, J. R. (2008) PTH-related peptide (PTHrP) in hypercalcemia. *J Am Soc Nephrol*. **19**: 672-5

Munoz-Martinez, F., Torres, C., Castanys, S. and Gamarro, F. (2010) CDC50A plays a key role in the uptake of the anticancer drug perifosine in human carcinoma cells. *Biochem Pharmacol*. **80**: 793-800

Murdoch, A. D., Grady, L. M., Ablett, M. P., Katopodi, T., Meadows, R. S. and Hardingham, T. E. (2007) Chondrogenic differentiation of human bone marrow stem cells in transwell cultures: generation of scaffold-free cartilage. *Stem Cells*. **25**: 2786-96

Murphy, G. and Nagase, H. (2008) Reappraising metalloproteinases in rheumatoid arthritis and osteoarthritis: destruction or repair? *Nat Clin Pract Rheumatol*. **4**: 128-35

Nagano, T., Morikubo, S. and Sato, M. (2004) Filamin A and FILIP (Filamin A-Interacting Protein) regulate cell polarity and motility in neocortical subventricular and intermediate zones during radial migration. *J Neurosci*. **24**: 9648-57

Nagano, T., Yoneda, T., Hatanaka, Y., Kubota, C., Murakami, F. and Sato, M. (2002) Filamin A-interacting protein (FILIP) regulates cortical cell migration out of the ventricular zone. *Nat Cell Biol*. **4**: 495-501

Nakajima, K., Tamai, M., Okaniwa, S., Nakamura, Y., Kobayashi, M., Niwa, T., Horigome, N., *et al.* (2013) Humoral hypercalcemia associated with gastric carcinoma secreting parathyroid hormone: a case report and review of the literature. *Endocr J*. **60**: 557-62

Nakajima, M., Takahashi, A., Kou, I., Rodriguez-Fontenla, C., Gomez-Reino, J. J., Furuichi, T., Dai, J., *et al.* (2010) New sequence variants in HLA class II/III region associated with susceptibility to knee osteoarthritis identified by genome-wide association study. *PLoS One*. **5**: e9723

National Institute for Health and Care Excellence (2014) Osteoarthritis: Care and management in adults [Web Page]. Available at: https://www.nice.org.uk/guidance/cg177 (Accessed: 24th June 2015).

Nguyen, M., He, B. and Karaplis, A. (2001) Nuclear forms of parathyroid hormone-related peptide are translated from non-AUG start sites downstream from the initiator methionine. *Endocrinology*. **142**: 694-703

NHS England (2015) Programme Budgeting [Web Page]. Available at: https://www.networks.nhs.uk/nhs-networks/health-investment-network/news/2012-13-programme-budgeting-data-is-now-available (Accessed: 24th June 2015).

Nica, A. C. and Dermitzakis, E. T. (2013) Expression quantitative trait loci: present and future. *Philos Trans R Soc Lond B Biol Sci*. **368**: 20120362

Nilsson, E., Jansson, P. A., Perfilyev, A., Volkov, P., Pedersen, M., Svensson, M. K., Poulsen, P., *et al.* (2014) Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. *Diabetes*. **63**: 2962-76

Norrmen, C., Vandevelde, W., Ny, A., Saharinen, P., Gentile, M., Haraldsen, G., Puolakkainen, P., *et al.* (2010) Liprin (beta)1 is highly expressed in lymphatic vasculature and is important for lymphatic vessel integrity. *Blood*. **115**: 906-9

Nussinov, R. (2013) The spatial structure of cell signaling systems. *Phys Biol*. **10**: 045004

O'Connell, K., Knight, H., Ficek, K., Leonska-Duniec, A., Maciejewska-Karlowska, A., Sawczuk, M., Stepien-Slodkowska, M., *et al.* (2015) Interactions between collagen gene variants and risk of anterior cruciate ligament rupture. *Eur J Sport Sci*. **15**: 341-50

Oda, Y., Chalkley, R. J., Burlingame, A. L. and Bikle, D. D. (2010) The transcriptional coactivator DRIP/mediator complex is involved in vitamin D receptor function and regulates keratinocyte proliferation and differentiation. *J Invest Dermatol*. **130**: 2377-88

Office for National Statistics (2014) Full Report: Sickness Absence in the Labour Market, February 2014 [Web Page]. Available at: http://www.ons.gov.uk/ons/rel/lmac/sickness-absence-in-the-labour-market/2014/rpt---sickness-absence-in-the-labour-market.html (Accessed: 24th June 2015).

Office for National Statistics (2015) Ageing of the UK population [Web Page]. Available at: http://www.ons.gov.uk/ons/rel/pop-estimate/population-estimates-for-uk--england-and-wales--scotland-and-northern-ireland/mid-2014/sty-ageing-of-the-uk-population.html (Accessed: 7th July 2015).

Oh, S. P., Griffith, C. M., Hay, E. D. and Olsen, B. R. (1993) Tissue-specific expression of type XII collagen during mouse embryonic development. *Dev Dyn*. **196**: 37-46

Okada, Y., Awaya, G., Ikeda, T., Tada, H., Kamisato, S. and Futami, T. (1989) Arthroscopic surgery for synovial chondromatosis of the hip. *J Bone Joint Surg Br*. **71**: 198-9

Oldershaw, R. A. (2012) Cell sources for the regeneration of articular cartilage: the past, the horizon and the future. *Int J Exp Pathol*. **93**: 389-400

Olsson, A. H., Ronn, T., Ladenvall, C., Parikh, H., Isomaa, B., Groop, L. and Ling, C. (2011) Two common genetic variants near nuclear-encoded OXPHOS genes are associated with insulin secretion in vivo. *Eur J Endocrinol*. **164**: 765-71

Ondrouch, A. S. (1963) Cyst Formation in Osteoarthritis. *J Bone Joint Surg Br*. **45**: 755-60

Ouma, W. Z., Mejia-Guerra, M. K., Yilmaz, A., Pareja-Tobes, P., Li, W., Doseff, A. I. and Grotewold, E. (2015) Important biological information uncovered in previously unaligned reads from chromatin immunoprecipitation experiments (ChIP-Seq). *Sci Rep*. **5**: 8635

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**: 237-48

Pacifici, M., Koyama, E., Shibukawa, Y., Wu, C., Tamamura, Y., Enomoto-Iwamoto, M. and Iwamoto, M. (2006) Cellular and molecular mechanisms of synovial joint and articular cartilage formation. *Ann N Y Acad Sci*. **1068**: 74-86

Panagiotou, O. A., Ioannidis, J. P. and Genome-Wide Significance, P. (2012) What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. *Int J Epidemiol*. **41**: 273-86

Panoutsopoulou, K., Southam, L., Elliott, K. S., Wrayner, N., Zhai, G., Beazley, C., Thorleifsson, G., *et al.* (2011) Insights into the genetic architecture of osteoarthritis from stage 1 of the arcOGEN study. *Ann Rheum Dis*. **70**: 864-7

Paulusma, C. C. and Oude Elferink, R. P. (2005) The type 4 subfamily of P-type ATPases, putative aminophospholipid translocases with a role in human disease. *Biochim Biophys Acta*. **1741**: 11-24

Peng, H. and Huard, J. (2004) Muscle-derived stem cells for musculoskeletal tissue regeneration and repair. *Transpl Immunol*. **12**: 311-9

Pierce, J. W., Lenardo, M. and Baltimore, D. (1988) Oligonucleotide that binds nuclear factor NF-kappa B acts as a lymphoid-specific and inducible enhancer element. *Proceedings of the National Academy of Sciences of the United States of America*. **85**: 1482-6

Pintard, L., Willems, A. and Peter, M. (2004) Cullin-based ubiquitin ligases: Cul3-BTB complexes join the family. *EMBO J*. **23**: 1681-7

Poole, A. R. (2012) Osteoarthritis as a whole joint disease. *HSS J*. **8**: 4-6

Poole, C. A. (1997) Articular cartilage chondrons: form, function and failure. *J Anat*. **191 (Pt 1)**: 1-13

Posthumus, M., September, A. V., O'Cuinneagain, D., van der Merwe, W., Schwellnus, M. P. and Collins, M. (2010) The association between the COL12A1 gene and anterior cruciate ligament ruptures. *Br J Sports Med*. **44**: 1160-5

Provencal, N., Suderman, M. J., Guillemin, C., Massart, R., Ruggiero, A., Wang, D., Bennett, A. J., *et al.* (2012) The signature of maternal rearing in the methylome in rhesus macaque prefrontal cortex and T cells. *J Neurosci*. **32**: 15626-42

Pruitt, K. D., Tatusova, T. and Maglott, D. R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. **33**: D501-4

Qi, Y., Feng, G. and Yan, W. (2012) Mesenchymal stem cell-based treatment for cartilage defects in osteoarthritis. *Mol Biol Rep*. **39**: 5683-9

Qian, D., Zhang, B., Zeng, X. L., Le Blanc, J. M., Guo, Y. H., Xue, C., Jiang, C., *et al.* (2014) Inhibition of human positive cofactor 4 radiosensitizes human esophageal squmaous cell carcinoma cells by suppressing XLF-mediated nonhomologous end joining. *Cell Death Dis*. **5**: e1461

Raine, E. V., Dodd, A. W., Reynard, L. N. and Loughlin, J. (2013) Allelic expression analysis of the osteoarthritis susceptibility gene COL11A1 in human joint tissues. *BMC Musculoskelet Disord*. **14**: 85

Raine, E. V., Wreglesworth, N., Dodd, A. W., Reynard, L. N. and Loughlin, J. (2012) Gene expression analysis reveals HBP1 as a key target for the osteoarthritis susceptibility locus that maps to chromosome 7q22. *Ann Rheum Dis*. **71**: 2020-7

Ratnayake, M., Reynard, L. N., Raine, E. V., Santibanez-Koref, M. and Loughlin, J. (2012) Allelic expression analysis of the osteoarthritis susceptibility locus that maps to MICAL3. *BMC Med Genet*. **13**: 12

Responte, D. J., Natoli, R. M. and Athanasiou, K. A. (2007) Collagens of articular cartilage: structure, function, and importance in tissue engineering. *Crit Rev Biomed Eng*. **35**: 363-411

Reynard, L. N. and Loughlin, J. (2013) The genetics and functional analysis of primary osteoarthritis susceptibility. *Expert Rev Mol Med*. **15**: e2

Reynard, L. N., Ratnayake, M., Santibanez-Koref, M. and Loughlin, J. (2014) Functional analysis of the osteoarthritis susceptibility locus residing at the carbohydrate sulfotransferase 11 gene *CHST11*. *Osteoarthritis Cartilage*. **22 Suppl**: S228

Ricard-Blum, S. and Ruggiero, F. (2005) The collagen superfamily: from the extracellular matrix to the cell membrane. *Pathol Biol (Paris)*. **53**: 430-42

Robertson, A. K., Geiman, T. M., Sankpal, U. T., Hager, G. L. and Robertson, K. D. (2004) Effects of chromatin structure on the enzymatic and DNA binding functions of DNA methyltransferases DNMT1 and Dnmt3a in vitro. *Biochem Biophys Res Commun*. **322**: 110-8

Rodriguez-Fontenla, C., Calaza, M., Evangelou, E., Valdes, A. M., Arden, N., Blanco, F. J., Carr, A., *et al.* (2014) Assessment of osteoarthritis candidate genes in a meta-analysis of nine genome-wide association studies. *Arthritis Rheumatol*. **66**: 940-9

Rodriguez, S., Jafer, O., Goker, H., Summersgill, B. M., Zafarana, G., Gillis, A. J., van Gurp, R. J., *et al.* (2003) Expression profile of genes from 12p in testicular germ cell tumors of adolescents and adults associated with i(12p) and amplification at 12p11.2-p12.1. *Oncogene*. **22**: 1880-91

Romanuik, T. L., Ueda, T., Le, N., Haile, S., Yong, T. M., Thomson, T., Vessella, R. L., *et al.* (2009) Novel biomarkers for prostate cancer including noncoding transcripts. *Am J Pathol*. **175**: 2264-76

Ronn, K., Reischl, N., Gautier, E. and Jacobi, M. (2011) Current surgical treatment of knee osteoarthritis. *Arthritis*. **2011**: 454873

Rosenberg, K., Olsson, H., Morgelin, M. and Heinegard, D. (1998) Cartilage oligomeric matrix protein shows high affinity zinc-dependent interaction with triple helical collagen. *J Biol Chem*. **273**: 20397-403

Ross, C. T., Roodgar, M. and Smith, D. G. (2015) Evolutionary distance of amino acid sequence orthologs across macaque subspecies: identifying candidate genes for SIV resistance in Chinese rhesus macaques. *PLoS One*. **10**: e0123624

Roughley, P. J. and Lee, E. R. (1994) Cartilage proteoglycans: structure and potential functions. *Microsc Res Tech*. **28**: 385-97

Rozen, S. and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*. **132**: 365-86

Ruano-Ravina, A. and Jato Diaz, M. (2006) Autologous chondrocyte implantation: a systematic review. *Osteoarthritis Cartilage*. **14**: 47-51

Rushton, M. D., Reynard, L. N., Barter, M. J., Refaie, R., Rankin, K. S., Young, D. A. and Loughlin, J. (2014) Characterization of the cartilage DNA methylome in knee and hip osteoarthritis. *Arthritis Rheumatol*. **66**: 2450-60

Sacks, J. J., Helmick, C. G., Luo, Y. H., Ilowite, N. T. and Bowyer, S. (2007) Prevalence of and annual ambulatory health care visits for pediatric arthritis and other rheumatologic conditions in the United States in 2001-2004. *Arthritis Rheum*. **57**: 1439-45

Safran, M., Dalah, I., Alexander, J., Rosen, N., Iny Stein, T., Shmoish, M., Nativ, N., *et al.* (2010) GeneCards Version 3: the human gene integrator. *Database: The Journal of Biological Databases and Curation*. **2010**: baq020

Saito, T., Fukai, A., Mabuchi, A., Ikeda, T., Yano, F., Ohba, S., Nishida, N., *et al.* (2010) Transcriptional regulation of endochondral ossification by HIF-2alpha during skeletal growth and osteoarthritis development. *Nat Med*. **16**: 678-86

Sambrook, P. N., MacGregor, A. J. and Spector, T. D. (1999) Genetic influences on cervical and lumbar disc degeneration: a magnetic resonance imaging study in twins. *Arthritis Rheum*. **42**: 366-72

Samuels, J., Krasnokutsky, S. and Abramson, S. B. (2008) Osteoarthritis: a tale of three tissues. *Bull NYU Hosp Jt Dis*. **66**: 244-50

Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W. W. and Lenhard, B. (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res*. **32**: D91-4

Sandell, L. J. and Aigner, T. (2001) Articular cartilage and changes in arthritis. An introduction: cell biology of osteoarthritis. *Arthritis Res*. **3**: 107-13

Sander, J. D. and Joung, J. K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*. **32**: 347-55

Sanggaard, K. M., Kjaer, K. W., Eiberg, H., Nurnberg, G., Nurnberg, P., Hoffman, K., Jensen, H., *et al.* (2008) A novel nonsense mutation in MYO6 is associated with progressive nonsyndromic hearing loss in a Danish DFNA22 family. *Am J Med Genet A*. **146A**: 1017-25

Sato, M. and Nagano, T. (2005) Involvement of filamin A and filamin A-interacting protein (FILIP) in controlling the start and cell shape of radially migrating cortical neurons. *Anat Sci Int*. **80**: 19-29

Sawitzke, A. D., Shi, H., Finco, M. F., Dunlop, D. D., Harris, C. L., Singer, N. G., Bradley, J. D., *et al.* (2010) Clinical efficacy and safety of glucosamine, chondroitin sulphate, their combination, celecoxib or placebo taken to treat osteoarthritis of the knee: 2-year results from GAIT. *Ann Rheum Dis*. **69**: 1459-64

Scanzello, C. R. and Goldring, S. R. (2012) The role of synovitis in osteoarthritis pathogenesis. *Bone*. **51**: 249-57

Sebastian, T. T., Baldridge, R. D., Xu, P. and Graham, T. R. (2012) Phospholipid flippases: building asymmetric membranes and transport vesicles. *Biochim Biophys Acta*. **1821**: 1068-77

Sellers, R. S., Luchin, A. I., Richard, V., Brena, R. M., Lima, D. and Rosol, T. J. (2004) Alternative splicing of parathyroid hormone-related protein mRNA: expression and stability. *J Mol Endocrinol*. **33**: 227-41

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M. and Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. **29**: 308-11

Silveira, S. M., da Cunha, I. W., Marchi, F. A., Busso, A. F., Lopes, A. and Rogatto, S. R. (2014) Genomic screening of testicular germ cell tumors from monozygotic twins. *Orphanet J Rare Dis*. **9**: 181

Simmons, D. L., Botting, R. M. and Hla, T. (2004) Cyclooxygenase isozymes: the biology of prostaglandin synthesis and inhibition. *Pharmacol Rev*. **56**: 387-437

Small, K. W., Weber, J. L., Roses, A., Lennon, F., Vance, J. M. and Pericak-Vance, M. A. (1992) North Carolina macular dystrophy is assigned to chromosome 6. *Genomics*. **13**: 681-5

Smith, A. K., Kilaru, V., Kocak, M., Almli, L. M., Mercer, K. B., Ressler, K. J., Tylavsky, F. A., *et al.* (2014) Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. *BMC Genomics*. **15**: 145

Song, L. and Crawford, G. E. (2010) DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc*. **2010**: pdb prot5384

Souslova, V., Townsend, P. A., Mann, J., van der Loos, C. M., Motterle, A., D'Acquisto, F., Mann, D. A., *et al.* (2010) Allele-specific regulation of matrix metalloproteinase-3 gene by transcription factor NFkappaB. *PLoS One*. **5**: e9902

Southam, L., Dowling, B., Ferreira, A., Marcelline, L., Mustafa, Z., Chapman, K., Bentham, G., *et al.* (2004) Microsatellite association mapping of a primary osteoarthritis susceptibility locus on chromosome 6p12.3-q13. *Arthritis Rheum*. **50**: 3910-4

Southam, L., Rodriguez-Lopez, J., Wilkins, J. M., Pombo-Suarez, M., Snelling, S., Gomez-Reino, J. J., Chapman, K., *et al.* (2007) An SNP in the 5'-UTR of GDF5 is associated with osteoarthritis susceptibility in Europeans and with in vivo differences in allelic expression in articular cartilage. *Hum Mol Genet*. **16**: 2226-32

Sowers, M. R. and Karvonen-Gutierrez, C. A. (2010) The evolving role of obesity in knee osteoarthritis. *Curr Opin Rheumatol*. **22**: 533-7

Spector, T. D., Cicuttini, F., Baker, J., Loughlin, J. and Hart, D. (1996) Genetic influences on osteoarthritis in women: a twin study. *BMJ*. **312**: 940-3

Spector, T. D. and MacGregor, A. J. (2004) Risk factors for osteoarthritis: genetics. *Osteoarthritis Cartilage*. **12 Suppl A**: S39-44

Sprooten, E., Knowles, E. E., McKay, D. R., Goring, H. H., Curran, J. E., Kent, J. W., Jr., Carless, M. A., *et al.* (2014) Common genetic variants and gene expression associated with white matter microstructure in the human brain. *Neuroimage*. **97**: 252-61

Srikanth, V. K., Fryer, J. L., Zhai, G., Winzenberg, T. M., Hosmer, D. and Jones, G. (2005) A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. *Osteoarthritis Cartilage*. **13**: 769-81

St-Jacques, B., Hammerschmidt, M. and McMahon, A. P. (1999) Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes Dev*. **13**: 2072-86

Stecher, R. M. (1941) Heberden's nodes. Heredity in hypertrophic arthritis of the finger joints. *Am J Med Sci*. **201**: 801-9

Stogios, P. J. and Prive, G. G. (2004) The BACK domain in BTB-kelch proteins. *Trends Biochem Sci*. **29**: 634-7

Strick, D. J. and Elferink, L. A. (2005) Rab15 effector protein: a novel protein for receptor recycling from the endocytic recycling compartment. *Mol Biol Cell*. **16**: 5699-709

Styrkarsdottir, U., Thorleifsson, G., Helgadottir, H. T., Bomer, N., Metrustry, S., Bierma-Zeinstra, S., Strijbosch, A. M., *et al.* (2014) Severe osteoarthritis of the hand associates with common variants within the ALDH1A2 gene and with rare variants at 1p31. *Nat Genet*. **46**: 498-502

Suri, S. and Walsh, D. A. (2012) Osteochondral alterations in osteoarthritis. *Bone*. **51**: 204-11

Suzuki, A., Ogura, T. and Esumi, H. (2006) NDR2 acts as the upstream kinase of ARK5 during insulin-like growth factor-1 signaling. *J Biol Chem*. **281**: 13915-21

Syddall, C. M. (2013) Functional analysis of the OA associated variant rs143383. Ph.D thesis. Newcastle University.

Syddall, C. M., Reynard, L. N., Young, D. A. and Loughlin, J. (2013) The identification of trans-acting factors that regulate the expression of GDF5 via the osteoarthritis susceptibility SNP rs143383. *PLoS Genet*. **9**: e1003557

Symmons, D., Turner, G., Webb, R., Asten, P., Barrett, E., Lunt, M., Scott, D., *et al.* (2002) The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. *Rheumatology (Oxford)*. **41**: 793-800

Tan, A. L., Toumi, H., Benjamin, M., Grainger, A. J., Tanner, S. F., Emery, P. and McGonagle, D. (2006) Combined high-resolution magnetic resonance imaging and histological examination to explore the role of ligaments and tendons in the phenotypic expression of early hand osteoarthritis. *Ann Rheum Dis*. **65**: 1267-72

Tate, P. H. and Bird, A. P. (1993) Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr Opin Genet Dev*. **3**: 226-31

Terkeltaub, R., Lotz, M., Johnson, K., Deng, D., Hashimoto, S., Goldring, M. B., Burton, D., *et al.* (1998) Parathyroid hormone-related proteins is abundant in osteoarthritic cartilage, and the parathyroid hormone-related protein 1-173 isoform is selectively induced by transforming growth factor beta in articular chondrocytes and suppresses generation of extracellular inorganic pyrophosphate. *Arthritis Rheum*. **41**: 2152-64

Tew, S. R., Murdoch, A. D., Rauchenberg, R. P. and Hardingham, T. E. (2008) Cellular methods in cartilage research: primary human chondrocytes in culture and chondrogenesis in human bone marrow stem cells. *Methods*. **45**: 2-9

Ushita, M., Saito, T., Ikeda, T., Yano, F., Higashikawa, A., Ogata, N., Chung, U., *et al.* (2009) Transcriptional induction of SOX9 by NF-kappaB family member RelA in chondrogenic cells. *Osteoarthritis Cartilage*. **17**: 1065-75

Vahlensieck, M., Linneborn, G., Schild, H. and Schmidt, H. M. (2002) Hoffa's recess: incidence, morphology and differential diagnosis of the globular-shaped cleft in the infrapatellar fat pad of the knee on MRI and cadaver dissections. *Eur Radiol*. **12**: 90-3

Valdes, A. M., Styrkarsdottir, U., Doherty, M., Morris, D. L., Mangino, M., Tamm, A., Doherty, S. A., *et al.* (2011) Large scale replication study of the association between HLA class II/BTNL2 variants and osteoarthritis of the knee in European-descent populations. *PLoS One*. **6**: e23371

Van Beeumen, J. J., Van Kuilenburg, A. B., Van Bun, S., Van den Bogert, C., Tager, J. M. and Muijsers, A. O. (1990) Demonstration of two isoforms of subunit VIIa of cytochrome c oxidase from human skeletal muscle. Implications for mitochondrial myopathies. *FEBS Lett*. **263**: 213-6

van Blitterswijk, W. J. and Verheij, M. (2008) Anticancer alkylphospholipids: mechanisms of action, cellular sensitivity and resistance, and clinical prospects. *Curr Pharm Des*. **14**: 2061-74

van der Kraan, P. M. and van den Berg, W. B. (2012) Chondrocyte hypertrophy and osteoarthritis: role in initiation and progression of cartilage degeneration? *Osteoarthritis Cartilage*. **20**: 223-32

van der Mark, V. A., Elferink, R. P. and Paulusma, C. C. (2013) P4 ATPases: Flippases in Health and Disease. *Int J Mol Sci*. **14**: 7897-922

van der Velden, L. M., Wichers, C. G., van Breevoort, A. E., Coleman, J. A., Molday, R. S., Berger, R., Klomp, L. W., *et al.* (2010) Heteromeric interactions required for abundance and subcellular localization of human CDC50 proteins and class 1 P4-ATPases. *J Biol Chem*. **285**: 40088-96

van Dongen, J., Slagboom, P. E., Draisma, H. H., Martin, N. G. and Boomsma, D. I. (2012) The continuing value of twin studies in the omics era. *Nat Rev Genet*. **13**: 640-53

Verzi, M. P., Hatzis, P., Sulahian, R., Philips, J., Schuijers, J., Shin, H., Freed, E., *et al.* (2010) TCF4 and CDX2, major transcription factors for intestinal function, converge on the same cis-regulatory regions. *Proc Natl Acad Sci U S A*. **107**: 15157-62

Vincze, T., Posfai, J. and Roberts, R. J. (2003) NEBcutter: A program to cleave DNA with restriction enzymes. *Nucleic Acids Res*. **31**: 3688-91

von Bulow, M., Heid, H., Hess, H. and Franke, W. W. (1995) Molecular nature of calicin, a major basic protein of the mammalian sperm head cytoskeleton. *Exp Cell Res*. **219**: 407-13

Vortkamp, A., Lee, K., Lanske, B., Segre, G. V., Kronenberg, H. M. and Tabin, C. J. (1996) Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein. *Science*. **273**: 613-22

Vreugde, S., Ferrai, C., Miluzio, A., Hauben, E., Marchisio, P. C., Crippa, M. P., Bussi, M., *et al.* (2006) Nuclear myosin VI enhances RNA polymerase II-dependent transcription. *Mol Cell*. **23**: 749-55

Waddington, R. J., Roberts, H. C., Sugars, R. V. and Schonherr, E. (2003) Differential roles for small leucine-rich proteoglycans in bone formation. *Eur Cell Mater*. **6**: 12-21; discussion 21

Walchli, C., Koch, M., Chiquet, M., Odermatt, B. F. and Trueb, B. (1994) Tissue-specific expression of the fibril-associated collagens XII and XIV. *J Cell Sci*. **107 (Pt 2)**: 669-81

Wang, B., Lin, D., Li, C. and Tucker, P. (2003) Multiple domains define the expression and regulatory properties of Foxp1 forkhead transcriptional repressors. *J Biol Chem*. **278**: 24259-68

Wang, H. and Elbein, S. C. (2007) Detection of allelic imbalance in gene expression using pyrosequencing. *Methods Mol Biol*. **373**: 157-76

Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T. W., Greven, M. C., Pierce, B. G., *et al.* (2012) Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res*. **22**: 1798-812

Wang, J., Zhuang, J., Iyer, S., Lin, X. Y., Greven, M. C., Kim, B. H., Moore, J., *et al.* (2013) Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. *Nucleic Acids Res*. **41**: D171-6

Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. **10**: 57-63

Watt, S. L., Lunstrum, G. P., McDonough, A. M., Keene, D. R., Burgeson, R. E. and Morris, N. P. (1992) Characterization of collagen types XII and XIV from fetal bovine cartilage. *J Biol Chem*. **267**: 20093-9

Wehrli, F. W. (2007) Structural and functional assessment of trabecular and cortical bone by micro magnetic resonance imaging. *J Magn Reson Imaging*. **25**: 390-409

Wei, S., Dunn, T. A., Isaacs, W. B., De Marzo, A. M. and Luo, J. (2008) GOLPH2 and MYO6: putative prostate cancer markers localized to the Golgi apparatus. *Prostate*. **68**: 1387-95

Wells, A. L., Lin, A. W., Chen, L. Q., Safer, D., Cain, S. M., Hasson, T., Carragher, B. O., *et al.* (1999) Myosin VI is an actin-based motor that moves backwards. *Nature*. **401**: 505-8

Wenham, C. Y. and Conaghan, P. G. (2010) The role of synovitis in osteoarthritis. *Ther Adv Musculoskelet Dis*. **2**: 349-59

West, J. A., Cook, A., Alver, B. H., Stadtfeld, M., Deaton, A. M., Hochedlinger, K., Park, P. J., *et al.* (2014) Nucleosomal occupancy changes locally over key regulatory regions during cell differentiation and reprogramming. *Nat Commun*. **5**: 4719

Wieland, H. A., Michaelis, M., Kirschbaum, B. J. and Rudolphi, K. A. (2005) Osteoarthritis - an untreatable disease? *Nat Rev Drug Discov*. **4**: 331-44

Willoughby, D. A., Moore, A. R. and Colville-Nash, P. R. (2000) COX-1, COX-2, and COX-3 and the future treatment of chronic inflammatory disease. *Lancet*. **355**: 646-8

Wluka, A. E., Cicuttini, F. M. and Spector, T. D. (2000) Menopause, oestrogens and arthritis. *Maturitas*. **35**: 183-99

Wolff, K. J., Ramakrishnan, P. S., Brouillette, M. J., Journot, B. J., McKinley, T. O., Buckwalter, J. A. and Martin, J. A. (2013) Mechanical stress and ATP synthesis are coupled by mitochondrial oxidants in articular cartilage. *J Orthop Res*. **31**: 191-6

Woolf, A. D. and Pfleger, B. (2003) Burden of major musculoskeletal conditions. *Bull World Health Organ*. **81**: 646-56

Wysolmerski, J. J. (2012) Parathyroid hormone-related protein: an update. *Journal of Clinical Endocrinology and Metabolism*. **97**: 2947-56

Xin, Z., Liu, W., Tian, L., Yuan, Y., Xin, H., Fu, J., Lin, G., *et al.* (2003) [Studies on the testis regression related gene profile in aged male]. *Beijing Da Xue Xue Bao*. **35**: 364-8

Xu, Q., Yang, G. Y., Liu, N., Xu, P., Chen, Y. L., Zhou, Z., Luo, Z. G., *et al.* (2012a) P4-ATPase ATP8A2 acts in synergy with CDC50A to enhance neurite outgrowth. *FEBS Lett*. **586**: 1803-12

Xu, Y., Barter, M. J., Swan, D. C., Rankin, K. S., Rowan, A. D., Santibanez-Koref, M., Loughlin, J., *et al.* (2012b) Identification of the pathogenic pathways in osteoarthritic hip cartilage: commonality and discord between hip and knee OA. *Osteoarthritis Cartilage*. **20**: 1029-38

Yan, D., Davis, F. J., Sharrocks, A. D. and Im, H. J. (2010) Emerging roles of SUMO modification in arthritis. *Gene*. **466**: 1-15

Yang, R., Kong, E., Jin, J., Hergovich, A. and Puschel, A. W. (2014) Rassf5 and Ndr kinases regulate neuronal polarity through Par3 phosphorylation in a novel pathway. *J Cell Sci*. **127**: 3463-76

Yasuda, H. and de Crombrugghe, B. (2009) Joint formation requires muscle formation and contraction. *Dev Cell*. **16**: 625-6

Yoshida, H., Cheng, W., Hung, J., Montell, D., Geisbrecht, E., Rosen, D., Liu, J., *et al.* (2004) Lessons from border cell migration in the Drosophila ovary: A role for myosin VI in dissemination of human ovarian cancer. *Proc Natl Acad Sci U S A*. **101**: 8144-9

Zhang, Y. and Jordan, J. M. (2010) Epidemiology of osteoarthritis. *Clin Geriatr Med*. **26**: 355-69

Zou, Y., Zwolanek, D., Izu, Y., Gandhy, S., Schreiber, G., Brockmann, K., Devoto, M., *et al.* (2014) Recessive and dominant mutations in COL12A1 cause a novel EDS/myopathy overlap syndrome in humans and mice. *Hum Mol Genet.* **23**: 2339-52