# Molecular genetic and epigenetic analysis of osteoarthritis

**Emma Louise Rogers**


A thesis submitted for the degree of

*Doctor of Philosophy*

Institute of Cellular Medicine

Newcastle University

October 2016

## Abstract

Osteoarthritis (OA) is a common degenerative disease of synovial joints that principally affects older individuals. The genetic architecture of OA is highly complex, with heterogeneous pathological pathways converging on a common end phenotype characterised by the painful loss of articular cartilage. In this study, I investigated both genetic and epigenetic aspects of the OA disease process. The genetic investigation focused on the functional analysis of a female hip OA association signal that was discovered in the arcOGEN study, marked by the single nucleotide polymorphism (SNP) rs4836732. The signal surpasses the genome-wide significance threshold, with $p = 6.11x10^{-10}$. The signal encompasses only two other SNPs that have a high degree of linkage disequilibrium (LD) with rs4836732. Functional studies using luciferase reporter assays, electrophoretic mobility shift assays (EMSAs) and a range of transformed cell lines did not identify any differential allelic activity between SNP alleles, but did identify differential EMSA banding patterns for the C allele of rs4836732 when compared to the T allele, suggestive of differential protein complex binding dependent on the allele present at rs4836732. Subsequent investigations focused on three genes in the immediate vicinity of the association signal: *ASTN2*, *PAPPA* and *TRIM32*. All three genes were expressed in cartilage, synovium and fat pad from OA patients, though none displayed differential allelic expression correlating with rs4836732 genotype in these tissues. This lack of positive evidence in end-stage disease tissue may be indicative of the association signal mediating risk during joint development or growth. Of the three genes, *PAPPA* was deemed the most likely candidate to have a functional affect in joint development. Knock-down of the gene in mesenchymal stem cells revealed that *PAPPA* is required for osteogenic and chondrogenic differentiation, providing support

for further investigation of this gene. My second line of investigation focused on OA epigenetics, specifically on subgroups of hip and knee OA patients who displayed altered cartilage methylation of inflammatory genes when analysed by high-density CpG methylation array. As part of this study, I developed pyrosequencing assays to determine the inflammatory subgroup of new patient cartilage samples, negating the need for a high-density array. I also further interrogated the inflammatory status of the patients, analysing serum markers of systemic inflammation and gathering biometric data to ascertain the primary cause of inflammation, with a focus on the metabolic status of patients in relation to body fat. Interestingly, no systemic differences in inflammation, circulating leptin or BMI were observed. This suggests the elevated inflammatory status may be local to the synovial joint, as opposed to being secondary to increased systemic inflammation. Overall, the two investigation routes that I have pursued further emphasise the heterogeneous nature of OA. They also highlight the importance of patient stratification for OA research and provide platforms to improve our comprehension of this debilitating disease.

*To my family*

*and to my friends.*

*It's been an adventure!*

*Thank you*

## Acknowledgements

I would like to thank my supervisors Prof. John Loughlin, Dr Louise Reynard and Dr Kenny Rankin for their support and advice throughout my PhD.

I greatly appreciate the orthopaedic surgeons and research nurses for collecting samples for this work.

Thank you to Dr Emma Raine for the cartilage gene expression data, Dr Kalliope Panoutsopoulou for the SNP data imputation and to Dr Madhushika Ratnayake and Maria Tselepi for the *in vitro* chondrogenesis data.

Thank you to Dr Michael Rushton for providing the cartilage methylation array data and Andrew Skelton for analysis of this data.

My thanks to Dr Amanda Villavilla for provision of nucleic acid from patient osteoblasts and to Dr Rachel Harry for her assistance with the electrochemiluminescence studies.

I am extremely grateful to Dr Katherine Johnson for her experimental advice and friendship throughout, to Dr David Swan for his technical assistance with the immunoblotting and to Dr Dennis Lendrem for his statistical assistance.

Thank you to Lucy Gentles for the collection and processing of patient samples. I would like to extend my thanks to all of the Osteoarthritis Genetics Group and the Musculoskeletal Research Group.

This research would not have been possible without funding from The William Harker Foundation and the generosity of patients at the Freeman Hospital and Royal Victoria Infirmary.

## Contents

**Chapter 2: Materials and Methods**

## Figures and Tables

### Figures

**Tables**

## Abbreviations

| | |
|---|---|
| **3'UTR** | Three prime untranslated region |
| **5'UTR** | Five prime untranslated region |
| **ACL** | Anterior cruciate ligament |
| **AEI** | Allelic expression imbalance |
| **APS** | Ammonium persulphate |
| **arcOGEN** | Arthritis Research Campaign Osteoarthritis Genetics |
| **ASD** | Autism spectrum disorder |
| **BMI** | Body mass index |
| **BMP** | Bone morphogenetic protein |
| **BSA** | Bovine serum albumin |
| **cDNA** | Complementary DNA |
| **ChIP** | Chromatin immunoprecipitation |
| **CI** | Cluster index |
| **COX** | Cyclooxygenase |
| **CRP** | C-reactive protein |
| **CS** | Chondroitin sulphate |
| **DEPC** | Diethylpyrocarbonate |
| **DMEM** | Dulbecco’s modified eagles medium |
| **DMM** | Destabilisation of the medial meniscus |
| **DNA** | Deoxyribonucleic acid |
| **DNMT** | DNA methyltransferase |
| **DTT** | Dithiothreitol |
| **ECM** | Extracellular matrix |
| **EDTA** | Ethylenediaminetetraacetic acid |

| | |
|---|---|
| **EMSA** | Electrophoretic mobility shift assay |
| **eQTL** | Expression quantitative trait locus |
| **FBS** | Fetal bovine serum |
| **FGF** | Fibroblast growth factor |
| **GAG** | Glycosaminoglycan |
| **GAPDH** | Glyceraldehyde 3-phosphate dehydrogenase |
| **gDNA** | Genomic DNA |
| **GH** | Glucosamine hydrochloride |
| **GWAS** | Genome wide association scan |
| **HA** | Hyaluronic acid |
| **HAC** | Human articular chondrocyte |
| **HD** | Hip dysplasia |
| **HDACi** | Histone deacetylase inhibitor |
| **IGF** | Insulin-like growth factor |
| **IHH** | Indian hedgehog |
| **IL** | Interleukin |
| **iPSC** | Induced pluripotent stem cell |
| **IQ** | Interquartile range |
| **ITS+L** | Insulin, transferrin, selenium, linoleic acid premix |
| **KL** | Kellgren-Lawrence |
| **LB** | Lysogeny broth |
| **LD** | Linkage disequilibrium |
| **LGMD2H** | Limb-girdle muscular dystrophy type 2H |
| **LLOD** | Lower limit of detection |
| **meQTL** | Methylation quantitative trait loci |

| | |
|---|---|
| **MMP** | Matrix Metalloproteinase |
| **MRI** | Magnetic resonance imaging |
| **mRNA** | Messenger RNA |
| **MSC** | Mesenchymal stem cell |
| **NDD** | Neurodevelopmental disorder |
| **NICE** | National Institute for Health and Care Excellence |
| **NOF** | Neck of femur fracture |
| **NSAID** | Non-steroidal Anti Inflammatory Drugs |
| **NT** | Non-targeting |
| **OA** | Osteoarthritis |
| **PAGE** | Polyacrylamide gel electrophoresis |
| **PBS** | Phosphate buffer saline |
| **PCR** | Polymerase chain reaction |
| **PFA** | Paraformaldehyde |
| **PTH** | Parathyroid hormone |
| **PTHrP** | Parathyroid hormone-related protein |
| **qPCR** | Quantitative real-time-PCR |
| **RFLP** | Restriction fragment length polymorphism |
| **RISC** | RNA-induced silencing complex |
| **RNA** | Ribose nucleic acid |
| **RNAi** | RNA interference |
| **ROS** | Reactive oxygen species |
| **rpm** | Revolutions per minute |
| **RT-PCR** | Reverse transcription PCR |
| **SEM** | Standard error of the mean |

| | |
|---|---|
| **siRNA** | Small interfering RNA |
| **SNP** | Single Nucleotide Polymorphism |
| **SSM** | Statistical shape model |
| **TBE** | Tris/borate/EDTA |
| **TBS** | Tris-buffered saline |
| **TF** | Transcription factor |
| **TGF** | Transforming growth factor |
| **TIMP** | Tissue inhibitor of metalloproteinases |
| **TLR** | Toll-like receptor |
| **TNAP** | Tissue non-specific alkaline phosphatase |
| **TNF** | Tissue necrosis factor |
| **UV** | Ultraviolet |
| **v/v** | Volume/volume |
| **WOMAC** | Western Ontario and McMaster Universities |
| **w/v** | Weight/volume |

# Chapter 1 - General Introduction

## 1.1 Synovial joint biology

The term joint refers to the location at which two bones meet, and there are three main classes of joint in the human body: fibrous, cartilaginous and synovial. Of these, synovial joints are the most common. Synovial joints are mobile joints, which function to enable articulation between the opposing skeletal surfaces. They facilitate a range of movements, determined by joint location and architecture, and the arrangement of  the tissues in a healthy synovial joint permits friction-free and painless motion. However, there are over 200 defined types of musculoskeletal disorders which hinder normal joint articulation and result in restricted motion and pain [1].

### 1.1.1 Joint structure and function

Synovial joints are highly sophisticated structures, each of which possesses a unique shape according to its biomechanical requirements. Synovial joints can be distinguished from other articulations in the body as they are surrounded by a thick and flexible joint capsule. The fibrous connective tissue of the joint capsule inserts into the two epiphyseal bones to provide both passive and active stability in order to support joint mobility. Synovial joints are further stabilised by surrounding ligaments, comprised primarily of type I collagen, which intersect the joint capsule [2]. Deep to the joint capsule lies the synovium, a soft tissue comprised of two predominant cell types. Type A synoviocytes act as resident macrophages, phagocytosing any detected waste and debris. Type B synoviocytes are fibroblast-like cells, which produce synovial fluid to lubricate the joint and minimise friction [3]. Articular cartilage lines the opposing bone surfaces and acts both a shock absorber during

movement, and provides a smooth articular surface with low friction coefficient [4]. Lubricin (also known as proteoglycan 4) and hyaluronic acid (HA), secreted by the synovium, nourish the articular cartilage by assisting in the exchange of carbon dioxide and metabolic waste for oxygen and nutrients [5].

Specific synovial joints contain additional structures such as adipose tissue pockets, known as articular fat pads. Extrasynovial infrapatellar fat pad (IFP) is a tissue found only in the knee joint and is proposed to protect exposed articular surfaces and to assist in distribution of synovial fluid through this joint. The fat pad is also hypothesised to assist in the repair of the nearby avascular structures, namely the anterior cruciate ligament, given its extensive vascularisation. Though the definite function remains unproven, studies indicate that the IFP is preserved during times of starvation even when subcutaneous fat is depleted, suggesting it is essential for knee function [6].

Chondrocytes are the sole cell type present in articular cartilage and their morphology and spatial arrangement varies according to their location (Figure 1.1). In adult articular cartilage, arrested chondrocytes form distinct zones. These zones are comprised of non-calcified cartilage, which is separated from the deeper zone containing calcified cartilage by a basophilic tidemark [7]. Chondrocytes are surrounded by an extracellular matrix (ECM), comprised of proteoglycans, of which aggrecan is the most abundant, collagen and water. Type II collagen is the predominant collagen in articular cartilage, and acts as a structural scaffolding protein, as well as providing tensile strength. The arrangement of collagen fibres

differs according to location within the articular cartilage. The most superficial fibres lie parallel to the joint surface, while the deepest fibres are radially orientated [8]. As the other abundant protein in the ECM, aggrecan is comprised of a protein core and three globular domains: G1, G2 and G3. Hyaluronic acid forms non covalent bonds with G1, while glycosaminoglycan (GAG) chains, which include chondroitin sulphate, are found between G2 and G3 [9].

Chondrocytes in the superficial zone (zone 1), residing adjacent to the synovial space, appear flattened in morphology. These chondrocytes, like the synovium, produce lubricin. Unlike the densely packed fibres present in the superficial zone, chondrocytes in the middle zone (zone 2) are surrounded by a randomly organised collagen fibril domain. In the final non-calcified zone, the deep zone (zone 3), radially orientated collagen fibrils extend into the zone of calcified cartilage (zone 4) to anchor the layers together. In the deep zone, cells are aligned perpendicular to the subchondral bone.  Zone 4 contains hypertrophic chondrocytes, expressing markers of hypertrophy such as collagen X, which reside within a calcified matrix [10].

The epiphyseal bone lying deep to the articular cartilage, known as the subchondral bone, consists of two zones. Separating calcified cartilage from the bone marrow is the subchondral bone plate. This region is densely calcified, owing to the formation of hydroxyapatite crystals, to support the cartilage and facilitate distribution of mechanical load. Beneath is the lesser calcified trabecular bone with a predominantly metabolic function [11]. Subchondral bone matrix is primarily comprised of type I collagen, in addition to proteins such as osteocalcin. The

calcified matrix contains osteoblasts and osteoclasts, which synthesise and degrade bone matrix components, facilitating bone growth and repair [12].

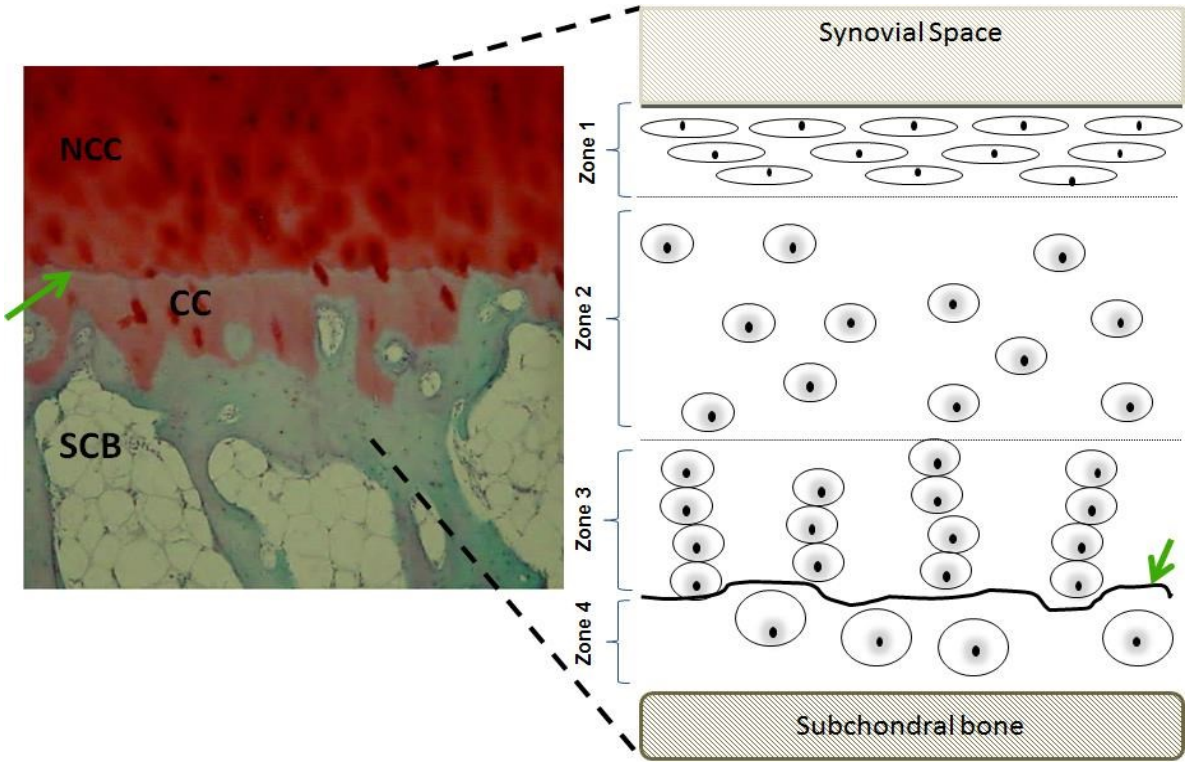



**Figure 1.1**
Histology and diagrammatic representation of adult articular cartilage structure. NCC – non-calcified cartilage; CC – calcified cartilage, SCB – subchondral bone. The photomicrograph depicts the osteochondral junction in an individual without OA, with red Safranin O dye staining of the proteoglycan rich cartilage. Zone 1, or the superficial zone, comprises 10-20% of the cartilage and contains densely packed collagen fibrils. 40-60% of articular cartilage is accounted for by zone 2, the middle zone, which is a region of randomly organised collagen fibres. Zone 3 is the deep zone, around 30% of articular cartilage, from which collagen fibres extend into the calcified zone (zone 4) to anchor the layers together. The tidemark (green arrow) separates the non-calcified and calcified cartilage. Photomicrograph taken from [7].

### 1.1.2 Joint development

Long bone formation during embryogenesis is initiated when mesenchymal cells condense to form distinctive limb buds, subsequently forming a hyaline cartilage core known as the anlage. Formation of synovial joints becomes morphologically evident when the uninterrupted cartilage anlage becomes separated by a region known as the joint interzone. Formation of the interzone is critical for joint formation and microsurgical removal of this region results in failure of joint morphogenesis [13]. Studies in mice have revealed that expression of *Gdf5*, encoding the growth/differentiation factor 5 protein, in the joint interzone induces cavitation of the joint to separate the opposing cartilaginous elements [14]. This is then succeeded by joint capsule formation. As the joint cavity enlarges, the interzone disperses and gives rise to articular cartilage and the synovium [15]. Murine fate mapping studies have identified that articular cartilage, and potentially all joint tissues, form from a specific group of *Gdf5*-expressing progenitor cells, which are distinct from those cells fated to become growth plate cartilage [16].

Chondrogenic mesenchymal stem cells (MSCs) follow one of two pathways during joint development. Resting chondrocytes form articular cartilage on the opposing bones, with expression of cartilage-specific collagens II, IX and XI denoting the transition of MSCs to chondroprogenitor cells. Remaining chondrocytes undergo proliferation, controlled by the IHH/PTHrP axis, followed by terminal differentiation and hypertrophy, to facilitate endochondral ossification [8].

Development of the long bones results in the formation of three distinct zones. The end of the long bones, the epiphysis, is separated from the middle region (the diaphysis) by the transitional zone between the two (the metaphysis). Between the epiphysis and metaphysis is a layer of cartilage known as the growth plate. During development, the long bones are created by endochondral ossification. This process occurs in two steps, firstly with differentiation of mesenchymal stem cells into chondrocytes to form a cartilage model. A centre of ossification forms within the diaphysis, within which chondrocytes are replaced by osteoblasts and undergo mineralisation. Later, a secondary centre of ossification appears in the epiphysis which allows continued growth of the bone as the cartilage between the ossification sites continues to expand until growth stops and the ossification sites fuse. In later life, fracture healing utilises a different process which does not require the formation of a cartilage model, known as intramembranous ossification. In this instance, mesenchymal stem cells differentiate directly into osteoblasts and secrete a mesh of trabecular bone matrix, which is later replaced with mature lamellar bone [17].

### 1.1.3 Joint homeostasis

Tissues of the synovial joint are composed of an extracellular matrix (ECM) containing collagens, glycoproteins and proteoglycans. The components of the ECM are continually synthesised and degraded according to the balance of growth factors, enzymes and cytokines, as well as in response to mechanical factors such as joint loading.

The ECM in articular cartilage supports a sole cell type, the chondrocyte. Synthesis of ECM components by chondrocytes occurs in response to the metabolic state of the chondrocytes, influenced by surrounding chemical and biomechanical factors, which is maintained in a steady state in healthy adult cartilage. However, when this balance is disrupted, the integrity of the articular cartilage is disturbed. Joint development and subsequent growth are normal physiological processes which require matrix synthesis to supersede degradation [18]. Many growth factors, such as transforming growth factor β (TGFβ), bone morphogenetic proteins (BMPs), fibroblast growth factors (FGFs) and insulin-like growth factors (IGFs) are present in the joint during development and have roles in cartilage development and repair. TGF-β1, BMP-2, BMP-7 and IGF-I all stimulate cartilage ECM synthesis, which FGF-18 has been shown to not only stimulate ECM production, but to induce chondrocyte proliferation [19,20].

Cartilage degradation is largely carried out by a two families of metalloproteinases: matrix metalloproteinases (MMPs) and the ADAMTS (a disintegrin and metalloprotease with thrombospondin motifs) family of metalloproteases. These proteinases primarily degrade collagen and aggrecan respectively. MMPs are sub grouped according to their target cleavage sites and their cellular location, and are synthesised in a pro-enzyme form to aid their regulation. Cleavage of an N-terminal fragment is necessary for MMP activation, a process performed by factors such as activated MMPs and reactive oxygen species (ROS). Connective tissue cells also synthesise tissue inhibitors of metalloproteinases (TIMPs), under the control of local growth factors. TIMPs provide additional regulation of MMPs by inhibiting pro-MMP cleavage and by preventing target substrate cleavage [21]. ADAMTS4 and

ADAMTS5 are potent enzymes which cleave the core protein of aggrecan, resulting in aggrecan protein truncation and reduced protein abundance [22].

Subchondral bone undergoes constant remodeling, reliant on the three main cell types present in bone; osteoblasts, osteoclasts and osteocytes. Bone formation is greatest during childhood but with increasing age, homeostasis is disrupted in favour of increased bone resorption. The dynamic formation and resorption of bone is essential to repair damage and maintain metabolic homeostasis of calcium and phosphorus, and relies on the stringent coordination between osteoblasts and osteoclasts. Osteoblasts and osteoclasts operate under an opposite feedback loop, whereby active osteoblasts secrete a protein, RANKL, which binds the RANK receptor on osteoclast precursors, stimulating the formation of mature osteoclasts [23]. Osteoclasts resorb bone to expose the underlying mineral matrix, which releases osteotrophic factors such as IGF1. This stimulates the differentiation of MSCs into osteoblasts to enable new bone formation [24]. Osteoblasts which become isolated in the bone matrix become osteocytes. Osteocytes possess long cellular processes and facilitate communication between cells in response to stimuli such as altered mechanical forces [25].

### 1.2 Osteoarthritis

Osteoarthritis (OA) is defined by NICE guidelines as ‘a clinical syndrome of joint pain accompanied by varying degrees of functional limitation’ [26]. It is the most prevalent chronic joint disease worldwide and is estimated to affect a third of the UK population over the age of 45 [27].

#### 1.2.1 Incidence and distribution of OA

OA can present in any synovial joint, but is most common in the hip, knee and hand. Age and sex adjusted data from the Fallon Community Health Plan in Massachusetts (USA) concluded OA to be most prevalent in the knee, with intermediate rates for hand OA and lowest observed rates for hip OA [28]. A later study in the Dutch population corroborated with these conclusions, finding knee OA to comprise the highest number of OA cases as seen by general practitioners in 2000 [29]. The study reported knee OA figures of 1.18 and 2.8 per 1000 for men and women respectively, while the figures were 0.9 and 1.6 per 1000 for hip OA. More recently, estimates from the 2010 Global Burden of Disease Study found radiographically-defined symptomatic knee OA to be more prevalent worldwide than hip OA, indicating a prevalence of 3.80% and 0.85% respectively [30].

OA prevalence positively correlates with increasing age, with 49% of women and 42% of men over the age of 75 years having sought treatment for OA in the UK [27]. The number of people suffering from OA is therefore anticipated to rise in coming years, reflective of our ageing population. A report commissioned by Arthritis Care estimates 17 million people will be diagnosed with OA in the UK by 2030 [31].

OA epidemiology also varies according to geographical location. Studies by the Community Oriented Program for Control of Rheumatic Disorders (COPCORD) identified knee OA incidence in Asia to vary from as much as 1.4% in urban Filipinos to 19.3% in rural Iranian communities [32]. The COPCORD studies went on to identify an increase in clinically diagnosed knee OA, after adjusting for age and sex, in urban compared to rural Indian communities. This difference may be attributed to differences in the socioeconomic and physical environments.

### 1.2.2 Pathology of synovial joint tissues

OA is often misconceived to be a disease affecting solely the articular cartilage. However, pathological changes have been described in all tissues found within the synovial joint, emphasising the notion that OA is an active response to joint injury as opposed to merely a passive degenerative course. While articular cartilage degradation is a vital feature of OA, many other joint abnormalities are also apparent, such as subchondral bone thickening, synovial inflammation and ligament degeneration [4] (Figure 1.2).

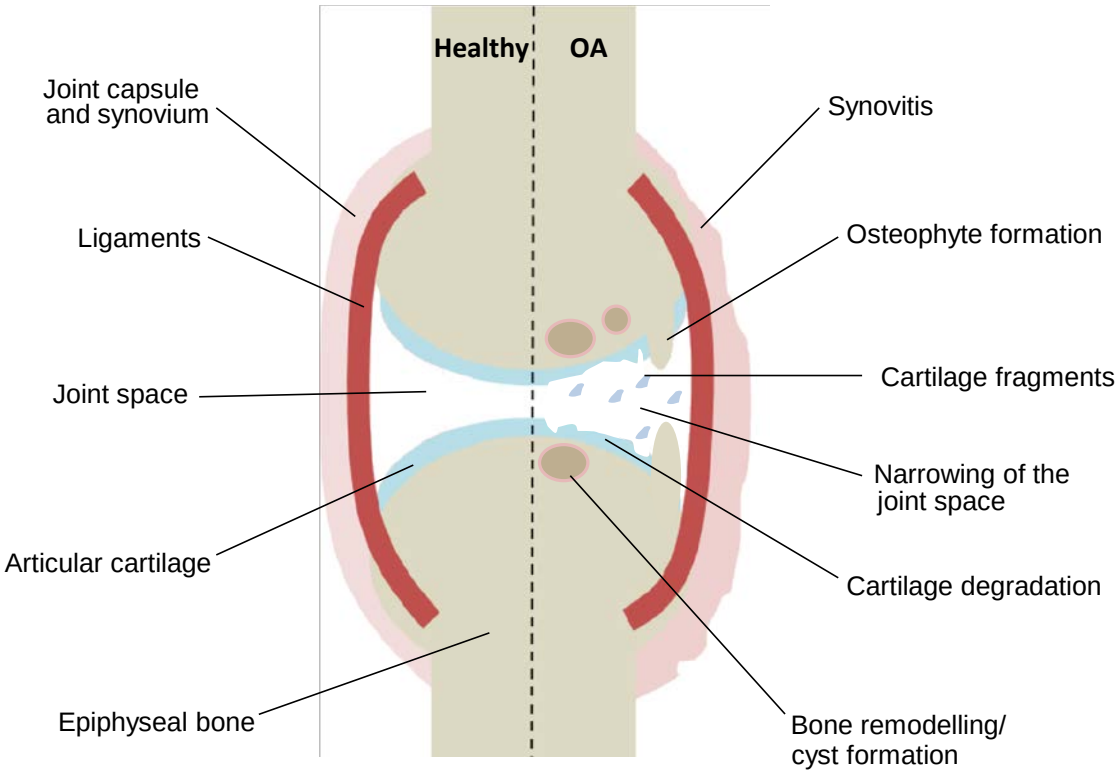



**Figure 1.2**

A comparative diagram of the pathological changes in OA joint tissues. In OA, the articular cartilage is eroded and abnormalities are noted in other structures, such as the formation of subchondral bone cysts and synovitis. In later stages of the disease, there may be narrowing of the joint space and formation of osteophytes, which are postulated to reflect a reparatory response to joint damage, though this remains debated [33,34].

The disrupted homeostasis of OA cartilage results in the abnormal activation of chondrocytes and leads to recapitulation of developmental events, causing chondrocytes to become hypertrophic and calcify [4]. Further to this, OA chondrocytes express MMPs, ADAMTSs and cytokine receptors which upregulate the inflammatory and catabolic response. Degradation of the matrix results in a positive feedback loop, as matrix protein fragments such as collagen and fibronectin stimulate production of MMPs and proinflammatory cytokines which cause further matrix degradation [35].

One of the earliest changes noted in the OA joint is bone remodeling, though it remains disputed as to whether this is a disease trigger factor, or a secondary consequence of cartilage degradation [36]. Changes in subchondral trabecular bone density have been identified in numerous studies and a systematic review of 139 studies concluded tibial bone area and bone marrow lesions to be independently associated with the structural progression of knee OA [37]. The process of skeletal adaptation also gives rise to boney projections, called osteophytes, along the joint margins. Though their function remains unproven, they are suggested to develop in an attempt to stabilise the joint, thus reducing further damage [38].

Synovitis is another common feature of OA and has been shown to correlate with radiographic disease severity [39]. Synovial infiltration of inflammatory mediators, such as macrophages, have been detected in over 50% of knee OA patients, and synovial hyperplasia and fibrosis can also occur [40]. Activated synoviocytes promote a catabolic response in the joint owing to secretion of MMPs and cytokines,

such as IL1β, IL6 and TNFα. Studies have shown this not only results in synovial joint inflammation, but alters the metabolic state of chondrocytes to increase their production of MMPs and aggrecanases, thus promoting cartilage destruction [41].

Our comprehension of the pathological changes associated with OA has been greatly improved by genetic, transcriptomic, epigenetic and proteomic analyses. It is now apparent that numerous pathological pathways converge on the same endpoint of symptomatic OA and as such, there is a large degree of pathologic heterogeneity both between patients and within the affected joint. Inflammation, however, is emerging as a vital feature of OA pathology. Though genetic studies do not support the notion that variation in inflammatory genes increases disease risk, inflammation appears to be highly relevant to disease pathology [42]. Of particular interest, research has suggested that OA patients may separate into those in whom inflammation is a key factor in disease pathology, and those in whom it is not [43]. This has clear implications both in terms of improving our understanding of the pathological changes occurring in synovial tissues in OA and may be vital in guiding future patient stratification in clinical trials.

### 1.2.3 OA diagnosis and treatment

Symptomatically, OA presents as a combination of pain and functional impairment and 80% of patients suffer from decreased joint mobility [33]. Clinical examination of the joint to detect signs such as stiffness and pain, which usually occur in conjunction with significant functional impairment, is required for OA diagnosis. Pain is normally associated with activity, for example walking or rising out of a chair in the case of knee OA [44]. In the more advanced stages of disease progression, OA can also be characterised radiographically, owing to the presence of subchondral bone cysts, subchondral sclerosis, narrowing of the joint space and abnormal bone contour [45]. OA can be clinically subtyped into one of two categories; primary OA has no discernible cause, while secondary OA is diagnosed in the presence of a causative episode, for example injury, infection or avascular necrosis [46].

Various scales have been developed in attempts to quantify OA severity. One of the most commonly used systems is the Kellgren and Lawrence (KL) scale, which grades severity on a scale of 0-4 on plain radiographs. Here, 0 represents no disease-associated joint abnormality and scores greater than 2 are indicative of radiographic OA [47]. The Western Ontario and McMaster Universities Arthritis Index (WOMAC) is another common system, comprising a set of standardised questionnaires which measure three subscales, developed to measure patient-reported outcomes. The WOMAC assesses five items for pain (score range 0–20), two for stiffness (score range 0–8), and 17 for functional limitation (score range 0–68) [48].

To date, no disease-modifying drugs have been approved for the treatment of OA and therapeutic intervention is limited to pain relief, and eventual joint replacement [49]. Consequently, the onus lies with effective disease management and current universal guidelines recommend a combination of pharmacological and non-pharmacological measures [50]. NICE recommends exercise as a core treatment, with the intention to strengthen local muscles and reduce body weight, therefore reducing joint stress and minimising abnormal joint biomechanics [26].

Pain relief is provided using paracetamol, non-selective NSAIDs with selective COX-2 inhibitors and opioids. NSAIDs predominantly target COX-1 and COX-2, the two isoforms of the enzyme cyclooxygenase, which generates prostanoids. These signalling molecules are a biological mediator of pain and inflammation. However, NSAIDs have been identified to cause adverse gastrointestinal (GI) effects, mediated by the widely expressed COX-1 isoform. This prompted the development of COX-2 specific inhibitors, such as celecoxib and rofecoxib [51]. While NSAIDs and COX-2 inhibitors have been found to be marginally more effective than paracetamol at reducing pain, paracetamol is recommended as the first line of pharmacological intervention owing to its lower toxicity. More recent research though indicates 3000mg/day paracetamol may cause similar gastrointestinal blood loss to 1200mg/day ibuprofen, with combination therapy being additive [52]. Further to these safety concerns, an ongoing evidence review by the MHRA (Medicines and Healthcare Products Regulatory Agency) has revealed reduced effectiveness of paracetamol in OA management compared to what was previously believed [26]. As such, NICE intend to carry out a full review of evidence regarding pharmacological recommendations in the near future. It is also to be noted that published evidence on

OA treatment is limited by the fact that most studies are of short duration using a single therapy, primarily in knee OA. Consequently, the genuine effectiveness of interventions is likely to differ given that patients tend to have pain in more than one joint and are on combined therapies [26].

Though pain relief is essential to improve quality of life, the key focus for pharmacological intervention is the generation of disease modifying drugs. Many clinical trials have been conducted in attempts to limit disease progression using various approaches, yet none have proved successful. A recent multicentre, randomised, double-blind trial comparing chondroitin sulphate plus glucosamine hydrochloride (CS+GH) versus celecoxib in 606 knee OA patients concluded the compounds to have comparable effect to celecoxib, but with no improvement in loss of cartilage loss [53]. Another method to reduce cartilage degradation is to inhibit the mediators of breakdown, such as MMPs. Systemic delivery of agents such as tacrolimus, an inhibitor of MMP-13, NF-κB, and bone matrix remodeling, have been shown to reduce pain-associated behaviours in the monoiodoacetic acid (MIA) and medial meniscal tear/medial collateral ligament tear rat OA models [54]. However, the benefits of such drugs are hampered by adverse reactions due to the inevitable broad protease inhibition [55].

For patients with ongoing functional impairment and pain, total joint arthroplasty is indicated and total knee replacement is the most commonly performed arthroplasty procedure in the UK [56]. However, joint replacements are finite and so, are only considered for those individuals who experience prolonged functional limitation and

pain refractory to the non-surgical core treatments. The selection of patients for arthroplasty is highly subjective and remains a key hurdle for clinicians. This is especially true when patients present at younger ages, with elevated BMI or additional associated co-morbidities present [56]. In order to help streamline this process, new screening techniques have been developed, particularly in the assessment of pain and stiffness, in attempts to predict post-operative patient dissatisfaction [57].

Attempts to minimise the current treatment gap between non-surgical interventions and arthroplasty have yielded procedures such as arthroscopic lavage and debridement. However, a randomised controlled trial conducted by Mosely and colleagues concluded there is no difference between these procedures and sham surgery [58]. An alternative surgical option for some younger patients with OA that is confined to either the lateral or medial compartment of the knee, is femoral or tibial osteotomy. This involves realignment of the femur or tibia, in order to reduce the load on the affected compartment [59]. This will often alleviate symptoms for a few years before knee replacement is considered.

Though risk factors for OA have been identified, we are yet to establish any robust prevention strategies. One of the most extensively studied areas for prophylactic intervention against OA is in OA occurring secondary to sports injury. Anterior cruciate ligament (ACL) damage is of particular interest, as an estimated 25% of the knee OA population have sustained previous ACL rupture. Consequently, ACL reconstruction has been investigated as a candidate prophylactic approach. Though

reported to improve short- and mid-term outcomes, longitudinal studies have found 50% of patients with ACL injury develop OA within 20 years of injury [60]. A subsequent literature review also found little evidence to support ACL reconstruction in the prevention of OA [61]. Given that most ACL injury occurs in the young population, evidence implies injury prevention in this subgroup would greatly reduce OA incidence. As such, implementation of evidence-based neuromuscular and biomechanical interventions are an area of great promise [62].

When reviewed as a collective, the absence of disease modifying drugs and the current passive, symptomatic nature of treatment emphasises the requirement for improved knowledge of OA pathogenesis with the aim to divert interventions to more active, prophylactic treatment.

### 1.3 OA risk factors

The complex aetiology of OA is poorly understood, though multiple risk factors have been identified, both modifiable and non-modifiable (Figure 1.3). It is becoming increasingly apparent, through molecular and clinical studies, that numerous risk factors and different pathophysiological pathways contribute to disease susceptibility and progression.

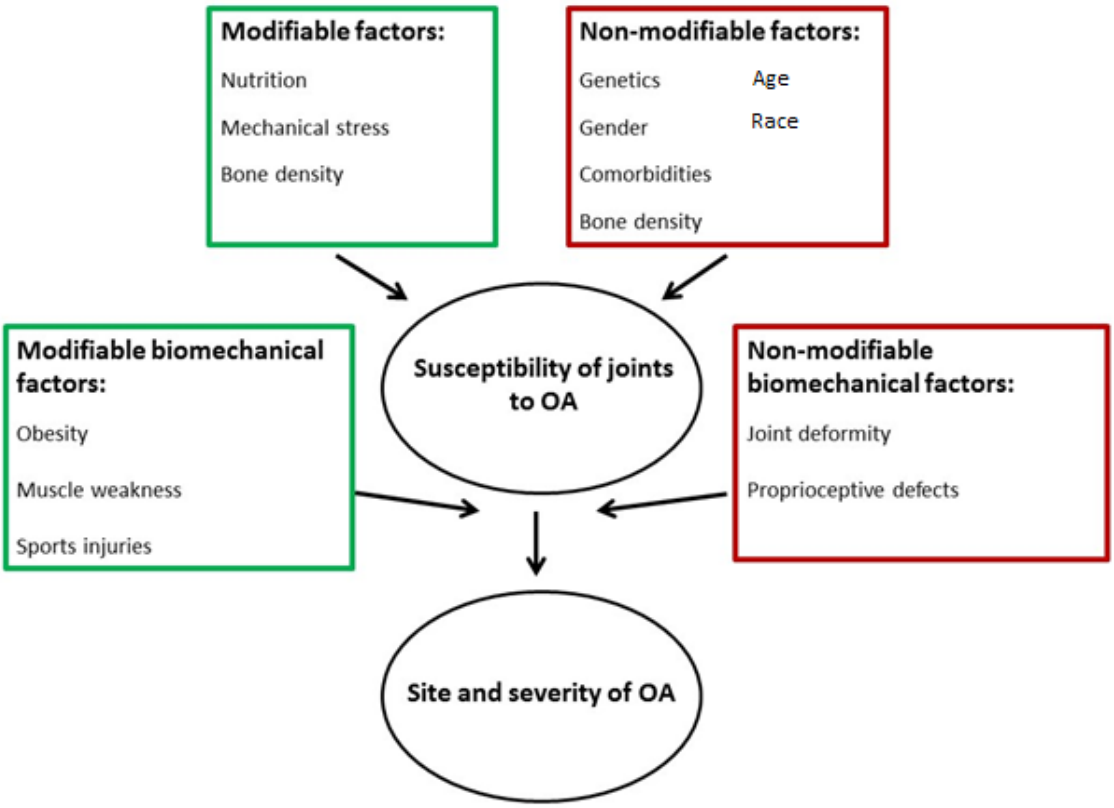



**Figure 1.3**

Factors contributing to increased OA risk. Various modifiable and non-modifiable systemic risk factors increase overall OA susceptibility. Local factors then act to determine site and severity of disease in the form of modifiable risks, such as muscle weakness, and non-modifiable risks, including joint deformities.

Age is the greatest risk factor for all joints, with female sex being a further predominant non-modifiable risk factor [63]. Prevalence of OA increases in women after the menopause, resulting in the suggestion that decreasing oestrogen levels may be a risk factor. However, clinical trials assessing the effect of oestrogen replacement therapy on OA development do not support this hypothesis [64].

Data from the Johnston County Osteoarthritis Project indicates obesity to be one of the main modifiable risks, finding individuals who are obese to be at twice the risk of

developing knee OA compared to those with normal or low BMI [65]. Furthermore, disease risk positively correlates with increasing body mass index (BMI). Though obesity results in abnormal loading through joints, there is also increased incidence of hand OA in obese individuals. As such, the effect of obesity cannot be wholly attributed to an increase in joint loading. It is hypothesised that increased risk is therefore likely to be mediated by the deregulation of metabolic and inflammatory factors [66].

OA risk is also influenced by race and ethnicity, with an increase in hip OA incidence having been observed in the Johnston County cross-sectional study in comparison to the Caucasian cohort [65]. Other ethnic difference have also been identified, such as an increase in knee OA prevalence in Chinese compared to Caucasian females [67].

The importance of understanding the impact of co-morbidity on health outcomes in OA patients has been highlighted in the recent population cohort study by Hawker *et al.* [68]. OA patients with pre-existing cardiovascular disease and increased baseline pain were found to have significantly decreased survival rates. A significant association was also found between OA-related disability and increased all-cause mortality, as well as serious cardiovascular events. This consequently indicates OA associated functional impairment as a potentially modifiable factor for serious cardiovascular events and death.

### 1.4 OA genetics

The genetic component of OA was first highlighted by family aggregation studies in the 1940s, which identified the disproportionately increased occurrence of Heberden's nodes in first degree relatives of patients with this indicator of hand OA [69]. Subsequent twin concordance studies, which allow for the separation of environmental and genetic factors, have identified that between 40% - 70% of OA risk can be accounted for by genetics [70]. What these studies have also identified is that OA is not a monogenic disease, but is caused by inheriting multiple polymorphisms in various genes, each of which contribute a very small amount to the overall risk.

#### 1.4.1 Candidate gene studies

With the confirmation that genetic factors contribute to OA susceptibility, further investigation has sought to identify the molecular basis for these hereditary factors. As such, when prior research indicates a gene has a role in joint biology or in a process potentially important in OA pathology, hypothesis-led candidate gene studies may be used. Perhaps the most compelling candidate gene studies performed have been on *GDF5*, a member of the TGF-β superfamily known to be necessary for skeletogenesis and joint formation [71]. A study by Miyamoto and colleagues identified a significant association between SNP rs143383 in the 5' UTR of *GDF5* and OA within the Asian population, with functional work indicating rs143383 to be a *cis*-acting expression quantitative trait locus (*cis*-eQTL) [72]. Subsequent research identified the T allele to be associated with OA in a Spanish and UK cohort, with studies showing *GDF5* expression in cartilage to be reduced in the presence of the T allele at rs143383 [73]. However, candidate-led studies have

resulted in many false positive results, with a meta-analysis study of 199 genes associated with OA concluding only two, *COL11A1* and *VEGF*, to be significantly associated with OA [74].

The major caveat with the candidate gene approach is that it is reliant on prior comprehension of the disease process. Given our incomplete knowledge of OA pathogenesis, this makes it highly likely that important genes may not be investigated were this to be the sole approach.

### 1.4.2 Genome wide association studies

Candidate genes studies have been largely replaced by genome wide association studies (GWAS). GWAS are superior in that they require no prior gene knowledge and are extremely powerful owing to large cohort sizes and extensive genome coverage. They work by comparing single nucleotide polymorphisms (SNPs) in a cohort of patients with a disease compared to a cohort without, or to population controls. This approach was made possible upon completion of the human genome sequencing project and the International Human HapMap project [75,76]. The phenomenon of linkage disequilibrium (LD), the non-random association of alleles at different loci, can be exploited so fewer SNPs need to be genotyped than the total number in the genome.

To date, several high-powered OA GWAS, genotyping ≥ 500,000 SNPs, have been performed and, in combination with candidate studies, this has facilitated the identification of 14 replicated OA susceptibility loci which reached genome-wide significance (defined as $p \leq 5.0 \times 10^{-8}$) [77]. In addition to this, a further 11 loci have

been identified which are suggestive of OA association ($p \leq 5.0 \times 10^{-5}$). A summary of these findings is shown in Table 1.1, except the association signals identified by the arcOGEN consortium, which are discussed in Chapter 1.5. The identification of these loci has reinforced the genetic heterogeneity of OA, with many showing joint-specific or sex-specific effects (Table 1.1).

| SNP | *p* value | Nearest gene | Stratum | Study |
|---|---|---|---|---|
| rs3204689 | $1.10 \times 10^{-11}$ | *ALDH1A2* | Hand | [78] |
| rs143383 | $6.20 \times 10^{-11}$ | *GDF5* | Knee | [79] |
| rs6094710 | $7.90 \times 10^{-09}$ | *NCOA3* | Hip | [80] |
| rs4730250 | $9.17 \times 10^{-09}$ | *COG5* | Knee | [81] |
| rs11842874 | $2.10 \times 10^{-08}$ | *MCF2L* | Hip & knee | [82] |
| rs143383* | $1.80 \times 10^{-13}$ | *GDF5* | Hip | [72] |
| rs7639618 | $7.30 \times 10^{-11}$ | *DVWA* | Knee | [83] |
| rs7775228 | $2.43 \times 10^{-08}$ | *HLA-DQB1* | Knee | [84] |
| rs10947262 | $6.73 \times 10^{-08}$ | *BTNL2* | Knee | [84] |

| SNP | *p* value | Nearest gene | Stratum | Study |
|---|---|---|---|---|
| rs12982744 | $7.80 \times 10^{-08}$ | *DOT1L* | Hip | [85] |
| rs3815148 | $8.00 \times 10^{-08}$ | *HBP1* | Knee & hand | [86] |
| rs12901499 | $7.50 \times 10^{-06}$ | *SMAD3* | Hip & knee | [87] |
| rs4907986 | $1.29 \times 10^{-05}$ | *COL11A1* | Hip | [74] |
| rs833058 | $1.35 \times 10^{-05}$ | *VEGF* | Male hip | [74] |
| rs225014 | $2.02 \times 10^{-05}$ | *DIO2* | Female hip | [88] |
| D-repeat D13/D14 | $1.30 \times 10^{-06}$ | *ASPN* | Knee | [89] |
| rs10980705 | $2.60 \times 10^{-05}$ | *LPAR1* | Knee | [90] |

**Table 1.1**

Replicated OA susceptibility loci identified by candidate gene studies and GWAS. The top panel shows those which reached genome-wide significance ($p \leq 5.0 \times 10^{-8}$), and the bottom panel are those suggestive of OA association ($p \leq 5.0 \times 10^{-5}$). *p* values shown are the combined discovery and replication values. *Indicates where the same association polymorphism was identified in different population cohorts.

### 1.5 Overview of the arcOGEN study

Funded by Arthritis Research UK, the arcOGEN Consortium was a UK-wide collaboration, which aimed to perform well-powered GWAS studies to identify genetic factors implicated in OA. The initial stage 1 GWAS comprised 3,177 OA cases and 4,894 controls from the UK, but failed to identify any association signals reaching genome-wide significance [91]. In stage 2, a GWAS comparing 7,410 unrelated OA patients and 11,009 unrelated controls from UK cohorts, 129 potential signals were identified [92]. Subsequent *in silico* replication using 5,064 OA cases and 40,619 controls, followed by *de novo* replication of the 26 of the most significant SNPs in 2,409 cases and 2,319 controls from European ancestry identified five loci as having genome-wide significance ($p = 5.0 \times 10^{-8}$) for association with OA (Table 1.2). An additional three loci were just below the significance threshold for OA-association.

| SNP | Chromosomal location | Nearest gene(s) | *p*-value | Stratum |
|---|---|---|---|---|
| rs11177<br>rs6976 | 3p21.1<br>3p21.1 | *GNL3*<br>*GLT8D1* | $7.24 \times 10^{-10}$<br>$1.25 \times 10^{-10}$ | TJR |
| rs4836732 | 9q33.1 | *ASTN2* | $6.11 \times 10^{-10}$ | THR-female |
| rs9350591 | 6q14.1 | *FILIP1*<br>*SENP6* | $2.42 \times 10^{-09}$ | Hip |
| rs10492367 | 12p11.22 | *PTHLH*<br>*KLHDC5* | $1.48 \times 10^{-08}$ | Hip |
| rs835487 | 12q23.3 | *CHST11* | $1.64 \times 10^{-08}$ | THR |

**Table 1.2**

Table of the 5 most significant signals from the arcOGEN study. Five loci display genome wide levels of significance for OA association, $p = 5.0 \times 10^{-8}$. rs11177 and rs6976 are in perfect linkage disequilibrium (LD) ($r^2=1$) and so, represent one signal [92]. TJR=total joint replacement. THR=total hip replacement. Hip = radiographically confirmed OA, prior to joint replacement surgery. *p* values shown are the combined discovery and replication values.

### 1.5.1 Expression quantitative trait loci (eQTL)

SNPs identified to be associated with a disease via GWAS studies mark a susceptibility locus in which other polymorphisms may reside. As such, it may be that a different SNP in high LD with the association SNP is in fact the causal variant. It is therefore necessary to first narrow down the region and identify the causal SNP, prior to then investigating the mechanism driving the disease association. In the event that the causal SNP is non-synonymous, an amino acid change leading to altered protein structure or function may cause the disease association, while synonymous mutations could result in aberrant gene transcription or splicing. The causal SNP may alternatively reside in an intron or be intergenic and thus, be non-

coding but regulate disease risk by regulating gene expression, for example by disrupting transcription factor binding and altering the enhancer or promoter activity of the region [83]. In this scenario, when regulation of gene expression is mediated by the genomic locus, the SNP is known as an expression quantitative trait locus (eQTL). When the genomic locus in which the causal SNP resides is in close proximity to the regulated gene, it is known as a *cis*-eQTL.

### 1.5.2 rs4836732 discovery and stratum

Table 1.3 shows the population data available for rs4836732. The T allele is the most common in the European and Asian populations, while TT homozygotes are uncommon in the West African population.

| | Genotype | | | Alleles | |
|---|---|---|---|---|---|
| **Population group** | **CC** | **CT** | **TT** | **C** | **T** |
| **European** | 0.23 | 0.51 | 0.26 | 0.49 | 0.51 |
| **Han Chinese** | 0.19 | 0.47 | 0.35 | 0.42 | 0.58 |
| **Japanese** | 0.19 | 0.47 | 0.35 | 0.42 | 0.58 |
| **West African** | 0.56 | 0.37 | 0.07 | 0.74 | 0.26 |

**Table 1.3**
Genotype frequencies of rs4836732. The C and T allelic frequencies and homozygote and heterozygote genotype frequencies are displayed according to population group. Population frequency data obtained from dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/).

rs4836732 is a C/T polymorphism residing on chromosome 9q33.1 within intron 18 of the *ASTN2* gene. The arcOGEN study identified the C allele to associate with female hip OA ($p=6.11 \times 10^{-10}$) (Table 1.2) [92]. It does not correlate with any other variants which code for non-synonymous polymorphisms. Therefore, the association signal marked by rs4836732 is anticipated to influence OA-susceptibility by a *cis*-acting regulatory effect on gene expression, as discussed previously. Three genes reside in the LD region demarcated by rs4836732 (Figure 1.4) and as such, are candidates for regulation by the SNP.

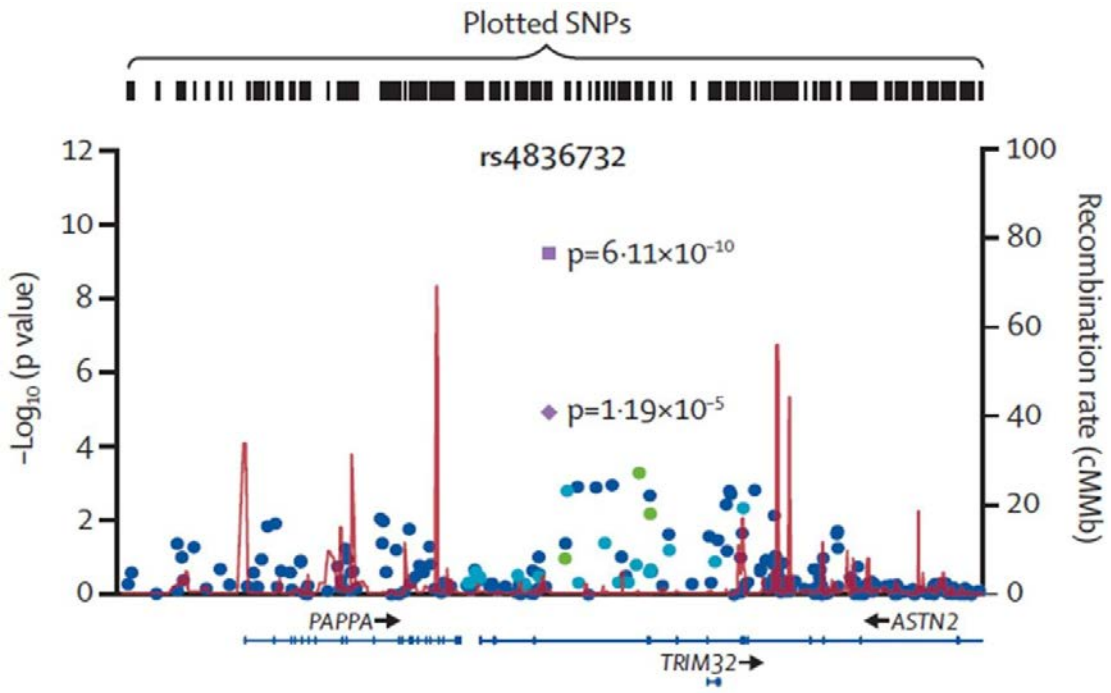



**Figure 1.4**

Plot of the rs4836732 association region and genes within the locus. Case-control association results of genotyped SNPs in the region (shown as circles), for the female hip stratum, are plotted against genomic position. rs4836732 is denoted in the discovery set by the purple diamond and in the final meta-analysis by a purple square. The circle colour is reflective of the $r^2$ of the genotyped SNP and red lines show estimated recombination rates in cM/Mb. rs4836732 is in intron 18 of *ASTN2*, with *PAPPA* residing upstream and *TRIM32* downstream, both of which are transcribed from the opposite strand. Arrows indicate the direction of gene transcription. Figure taken from [92].

### 1.5.3 ASTN2

*ASTN2* (astrotactin 2) is a poorly characterised gene, about which little scientific literature is available. The gene was first described by Wilson and colleagues in 2010 after being identified as a homologue to *ASTN1*, a gene encoding the ASTN1 protein which is required for glial-guided migration of cerebral cortex output neurons [93]. Murine models indicated the gene to be abundantly expressed in cerebral granule neurons during glial-guided neural migration and identified that ASTN2, like ASTN1, is a membrane-bound protein. Further to this, ASTN2 was shown to complex with ASTN1 and consequently facilitate ASTN1 polarisation during migration, thus modulating surface ASTN1 [93].

Bis *et al*. documented an association between nine SNPs within *ASTN2* and hippocampal volume in a GWAS of a young, heterogeneous sample of 7,794 individuals [94]. In the same study, a largely independent sample (n=1563) found the *ASTN2* SNP rs7852872 to associate with accelerated global cognitive decline and accelerated memory loss. This is hypothesised to relate to the function of glial-guided neuronal migration, especially in the dentate gyrus which functions in formation of new episodic memories and is essential for cognition and spatial memory.

A meta-analysis of GWASs interrogating genetic variants influencing schizophrenia and bipolar in European-American samples (653 bipolar cases and 1034 controls, 1172 schizophrenia cases and 1379 controls) noted *ASTN2* variants to be associated with attention deficit/hyperactivity disorder and schizophrenia [95].

Together, research to date suggests *ASTN2* is required for neuronal development and function, with no evidence as of yet for a role in joint development or homeostasis.

### 1.5.4 TRIM32

Tripartite motif containing 32, or *TRIM32*, is a small gene which resides within an intron of *ASTN2,* and is transcribed from the opposite strand. The TRIM32 protein is a member of the ring finger family of zinc finger nucleases and is known to interact with the HIV-1 Tat protein, a regulator of transcriptional activation [96]. However, cotransfection of *TRIM32* with a Tat expression vector did not enhance the function of Tat, meaning no conclusions could be made in relation to a role in transcription [96].

A case-control study of limb-girdle muscular dystrophy type 2H (LGMD2H) and 100 controls identified a mutation at position 487 (D487N) of *TRIM32* to be causative of LGMD2H [97]. This mutation was subsequently identified to be identical to the causative mutation in sacrotubular myopathy, a Hutterite population specific disease [98]. Further analysis revealed TRIM32 is a E3 ubiquitin ligase, acting alongside ubiquitin-conjugating enzymes UbcH5a, UbcH5c, and UbcH6 [99]. The same group identified differential expression of the TRIM32 protein confined mainly to skeletal muscle, with induction in the event of muscle remodelling in response to increased loading. This implies that TRIM32 levels change in response to physiological conditions. TRIM32 specifically interacts with myosin to cause ubiquitination of actin, indicating a likely participation in myofibrillar protein turnover.  A more recent study

found *TRIM32* is expressed in myogenic progenitors in mouse, with global knockout resulting in disrupted myogenic differentiation [100]. A LGMD2H-like phenotype was also observed, concluded to be due to impaired muscle regeneration owing to aberrant regulation of the transcription factor c-myc, a protein which blocks skeletal muscle stem cell differentiation.

TRIM32 is also required during neurogenesis. Four mutations in the gene have been identified as the cause of LGMD2H, while a fifth was identified to result in Bardet-Biedl syndrome type 11 [99]. With higher expression of *Trim32* in the brain compared to skeletal muscle, a mouse knock-out revealed decreased concentration of neurofilaments and a reduction in myelinated motor axon diameters compared to wild type, for the first time defining a role for TRIM32 in neurogenic function.

Schwamborn and colleagues showed polarisation of TRIM32 to daughter cells during mitosis of neural progenitor cells in Drosophila, suggesting that it plays a role in the balance between proliferation and differentiation [101]. Overexpression was shown to induce neural differentiation, while inhibition maintains progenitor status in daughter cells. This balance is suggested to be mediated via control of c-myc degradation, which is required for MSC self-renewal, and TRIM32 ubiquitination by the N terminus targets c-myc for degradation [101]. Studies in mice have also shown mutual transcriptional regulation of TRIM32 and the transcription factor p73, with dysregulation affecting both neural development and potentially oncogenesis [102]. Furthermore, a screen of clinical microarray DNA datasets from 64114 patients with neurodevelopmental disorders (NDDs) and 44 085 population-based controls identified an enrichment of 3′ *ASTN2* deletions (a group of which included *TRIM32*) in the NDD subjects ($p = 0.002$) [103]. NDD individuals with these deletions were

observed to have frequent phenotypes including speech delay, autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD). It is therefore not unreasonable to hypothesise that both ASTN2 and TRIM32 have a key role in human neurodevelopment and psychopathologies.

### 1.5.5 PAPPA

*PAPPA*, or pregnancy-associated plasma protein A, is the most extensively studied of the genes within the rs4836732 susceptibility locus. The first pathologic association with this gene was published in 1985 in relation to stage I breast carcinoma recurrence [104]. The main body of subsequent work focused on circulating PAPPA and it was identified as a maternal serum marker of foetal chromosomal disease. PAPPA screening is now performed routinely in the first trimester of pregnancy in combination with beta-HCG and nuchal translucency measurements to detect Down's syndrome foetuses, together having a reported successful detection rate of 90% [105].

The first publication of direct importance to OA research was provided by Qin and colleagues [106]. They noted that the putative Zn-motif of PAPPA is identical to the active site Zn-motif of matrix metalloproteinases. It was subsequently identified that PAPPA is the IGF-II-dependent IGF-binding protein (IGFBP)-4 protease, with IGFBP2, 4 and 5 being the PAPPA-cleavable IGFBPs. Given that IGFs are one of the key factors which stimulate bone growth, and also modulate cartilage homeostasis by regulation of proteoglycan synthesis and degradation, this finding thus suggests PAPPA may have the capacity to regulate bone and cartilage development and maintenance. As such, mouse models suggest PAPPA cleavable

IGFBs are inhibitory to IGF function in bone, and all three IGFBPs were confirmed to be produced by osteoblasts *in vitro*. Qin *et al.* demonstrated that osteoblast-specific PAPPA expression in mouse increased bone size, indicative of PAPPA-cleavable IGFBPs exerting a regulatory effect over IGF bioavailability in skeletal tissues. This suggests PAPPA to have an anabolic role in bone formation and metabolism.

In a global mouse *PAPPA* knock-out, a significant decrease in trabecular volume was recorded, progressing with age, alongside an age-associated decrease in fracture resistance [107]. Measurement of dynamic serum parameters associated with bone formation indicated reduced bone formation in young knockout mice (4 and 6 months). By 6 months, bone resorption was also impaired compared to controls, although this was followed by an increase in resorption markers at 12 months. Osteoblast numbers were reduced by 45-60% and osteoclasts by 50-60% compared to wild type littermates at 2 and 6 months, and 6 months respectively. As such, *PAPPA* appears to be necessary for both embryonic and postnatal bone development.

Global gene expression profiling of the collagen-induced arthritis mouse, a model for rheumatoid arthritis which display aberrant bone formation at disease onset, identified a fold change increase of 2.3 of *Pappa* in days 0-3 of disease onset compared to during weeks 1-2 [108]. This therefore provides further evidence for an anabolic role in bone formation *in vivo*.

## 1.6 OA epigenetics

As well as heritable genetic variation, epigenetic alterations have also been associated with OA. Coined by developmental biologist Conrad Waddington, the term epigenetics describes mitotically heritable changes in gene expression which occur without changes to the nucleic acid sequence of the DNA [109].

### 1.6.1 Mechanisms of epigenetic modifications

Various types of epigenetic modification have been identified and these can regulate gene expression either at the level of gene transcription, or by exerting post-transcriptional influences. The key mechanisms by which this occurs are histone modifications, non-coding RNAs (ncRNAs) and DNA methylation. Owing to its importance to this thesis, DNA methylation is discussed in more detail in Section 1.6.2, while histone modifications and ncRNAs are summarised below.

Histones are positively charged proteins which associate with DNA to condense it into chromatin. DNA is wrapped around a histone octamer (two copies of histone proteins H2A, H2B, H3, and H4) to create a nucleosome, which then is wrapped into a 30nm solenoid spiral with surrounding nucleosomes. H1 histone proteins, so-called ‘linker’ proteins, maintain the chromosome structure within the condensed chromatin [110]. Over 150 histone modifications, including acetylation, phosphorylation, ubiquitination and methylation of histone residues, have been identified and these can have variable effects on gene transcription [77]. Dysregulation of histone modifications has been described in numerous diseases and global alterations in

histone modifications have been well characterised in diseases such as cancer [111,112].

ncRNAs can interfere with gene expression by affecting transcription, splicing or translation and include microRNAs (miRNAs) and long non-coding RNAs [109]. It is reported that over a third of mammalian mRNAs are regulated by miRNAs, highlighting the vital role of this epigenetic mechanism in gene regulation [113]. miRNAs are short 20- to 23- nucleotide single stranded non-coding RNAs transcribed from intergenic and intronic regions of the genome which regulate transcript stability by promoting degradation of mRNA or repressing subsequent translation [114]. The creation of mouse knock-outs of specific miRNAs has revealed an essential regulatory role for miRNAs in normal development. Relevant to OA, severely impaired skeletogenesis has been described in mice null for the miRNA processing enzyme, Dicer [115].

### 1.6.2 DNA methylation

Occurring primarily at cytosine-guanine (CpG) dinucleotide pairs, DNA methylation of mammalian cells involves addition of a methyl group to a cytosine base by DNA methyltransferases (DNMT) [116]. Different DNMT enzymes are implicated in the establishment of methylation compared to the maintenance of methylation signatures. The mitotic inheritance of DNA methylation in primarily regulated by the maintenance enzyme DNMT1, which is essential to transfer patterns of methylation following DNA replication [114]. *De novo* DNA methylation however, is largely

regulated by DNMT3a and DNMT3B, which are responsible for establishing DNA methylation patterns, primarily during early development [110].

Regions of DNA with a high frequency of CpG sequences are termed ‘CpG islands’, and around 30% of genes have CpG islands near their promoters, the majority of which are unmethylated [114]. There are two main models as to how DNA methylation affects gene transcription. Firstly, they can have a direct effect on gene expression by interfering with transcription factor binding [117]. Further to this, methylated cytosines can act as docking sites for DNA binding proteins, such as MeCP2. This can result in indirect repression of gene expression as the binding proteins recruit co-repressors, which alter chromatin conformation.

Though DNA methylation is an essential mechanism required for normal regulation of cell-specific gene expression, aberrant methylation has been increasingly associated with OA pathophysiology [118]. The majority of DNA methylation studies in OA have focused on articular cartilage. These have highlighted a distinct cartilage methylome not only when comparing non-OA and OA cartilage, but also according to the OA affected joint site [118, 119]. Furthermore, our group published research which identified separation of knee and hip OA patients into distinct clusters according to cartilage CpG site methylation [43]. The distinct methylation landscapes identified both between and within hip and knee cartilage is suggestive of differential gene regulation and cellular characteristics at different joint sites, thus further reinforcing the heterogeneity of OA pathophysiology.

The precise role of DNA methylation in OA is only more recently beginning to be elucidated. DNA methylation changes in both isolated chondrocytes and OA cartilage have been associated with altered gene expression of genes, including *GDF5* [120]. Targeted analyses have demonstrated that methylation of certain CpG sites in cartilage DNA can interfere with transcription factor binding, suggested to then affect gene transcription [109]. Genome-wide DNA methylation studies have revealed a potentially exciting association between CpG site methylation in OA cartilage and the cartilage expression of inflammatory genes [121]. As such, gene regulation in OA, especially in relation to inflammatory genes, appears to be influenced by an interplay between genetic and epigenetic mechanisms.

### 1.6.3 Epigenetics and inflammatory gene expression

One of the most compelling associations between epigenetics and gene expression in OA has come from two microarray studies, performed by our group and Fernandez-Tajes *et al*., which identified separation of OA patients into two distinct clusters, according to inflammatory gene CpG methylation in cartilage DNA [43, 122]. The study by our group, however, observed substantially more differentially methylated CpGs compared to the study by Fernandez-Tajes *et al.*, likely to be reflective of the use of both a larger sample size including both hip and knee cartilage, as well as a more high-powered CpG methylation array.

The study by Rushton *et al*. used the Illumina Infinium HumanMethylation450 BeadChip array, which allows the annotation of ~480,000 CpG sites, to compare methylation in chondrocyte DNA extracted from 23 hip OA patients, 73 knee OA

patients, and 21 non-OA control patients with neck of femur fractures [43]. Comparison of hip OA patients and non-OA controls identified 5,322 differentially methylated loci (DMLs). Further to this, a differential methylation profile was found when comparing hip and knee OA samples, revealing 5,547 DMLs between these groups.

Perhaps most interestingly, the hip OA samples were found to cluster into two distinct groups due to differential methylation at 15,239 loci, with an enrichment of genes involved in inflammation (Figure 1.5). Knee OA samples also formed two clusters, though less well defined compared to the hip samples.

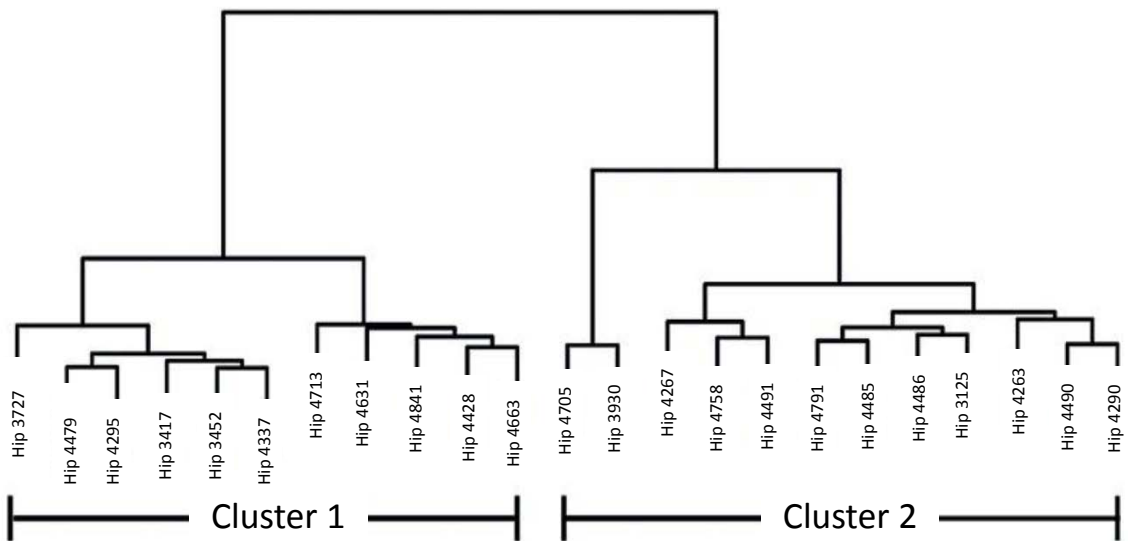



**Figure 1.5**
Unsupervised hierarchical clustering of the global CpG β methylation values in hip OA cartilage samples. Each end branch represents an individual patient. Figure taken from [43].

Further investigation using RNA from the hip cartilage of samples analysed on the array concluded promoter hypomethylation of inflammatory genes in one of the identified groups was accompanied by increased gene expression, as measured by real-time quantitative PCR (qPCR) [121]. Analysis of 6 patients with hypomethylation at the promoters of *TNF*, *IL6*, *IL1A*, *CXCR2*, *CCL5* and *CCL2* found significantly increased expression of these genes in the cartilage compared to both 5 non-OA controls and 5 OA hip cartilage samples from patients in the non-inflammatory group. Together, these two studies therefore identify a distinct subgroup of OA hip patients with epigenetically and transcriptionally unique articular cartilage profiles. This finding may prove important for future clinical studies and for stratification of patients in OA research. It also poses the question as to whether aberrant inflammatory gene expression is also present in the cartilage of the knee subgroup, and vitally, what it is that is mediating this increased inflammation.

### 1.7 Summary

Our ageing population is creating undeniable socioeconomic strain and increasing the global healthcare burden. Osteoarthritis (OA) is a highly prevalent chronic disease, which is particularly associated with increasing age. While characterised by loss of articular cartilage, OA causes pathologic changes to all synovial joint tissues and has been identified to have a distinct inflammatory component. At present, no disease modifying drugs have been established and so, therapeutic options remain inadequate.

OA is a multifactorial disease, influenced by genetic and non-genetic factors. Genetic, epigenetic and transcriptomic studies have demonstrated the heterogeneity in OA pathophysiology, highlighting that multiple different pathways contribute to result in the same disease endpoint. As such, investigations into both genetic and epigenetic aspects of the OA disease process are essential in order to further our understanding of OA and to help guide future clinical trials.

From a genetic stance, OA is highly polygenic, with multiple susceptibility alleles contributing to disease pathogenesis. Through the arcOGEN GWAS, several regions of the genome have been associated with OA including chromosome 9q33.1, marked by the rs4836732 single nucleotide polymorphism (SNP). Demonstrating significant association with female hip osteoarthritis, this SNP resides within an intron of the *ASTN2* gene and in the immediate vicinity of *PAPPA* and *TRIM32*. As such, functional analysis of this locus to establish the mechanism by which it influences OA risk would improve our understanding of the disease process and may potentially yield novel therapeutic possibilities.

Epigenetic research has identified various aberrances in epigenetic regulation in OA patients, both compared to non-OA controls and within the disease group. CpG microarray analysis in OA cartilage DNA has identified a subgroup of OA patients who display divergent gene methylation, with a particular enrichment of inflammatory genes, which is reflected by increased expression of inflammatory genes in OA hip cartilage. Further investigation into the mechanism driving the elevated inflammatory gene expression in these patients, and also to establish whether these patients

display elevated systemic inflammation, would greatly assist in our knowledge of disease mechanisms and would help to direct future clinical trials.

### 1.8 Overall aims

The broad aim of this research is to investigate both genetic and epigenetic pathways which contribute to OA pathophysiology. This will be achieved by investigating:

- Allelic functionality
- Gene expression
- Transcription factor binding
- DNA methylation
- Systemic inflammation

# Chapter 2: Materials and methods

## 2.1 Database searches to characterise the rs4836732 association signal

The UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway) was initially used to characterise the association signal, through which various data including the NCBI RefSeq collection of annotated gene transcripts and data generated by the ENCODE Consortium, was accessed. The dbSNP online database (http://www.ncbi.nlm.nih.gov/SNP/) was used to obtain genotype frequencies for the SNPs, while the Broad Institute online software (http://www.broadinstitute.org/mpg/snap/) was used to identify SNPs in LD $r^2$>0.7 with the association SNP.

## 2.2 Tissue sample collection

Informed written consent for the use of joint tissues was obtained from patients who had undergone knee or hip arthroplasty 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). Following surgery, samples were stored at 4$^{o}$C in Hank's Balanced Salt Solution supplemented with nystatin, penicillin, and streptomycin. Articular cartilage, infrapatellar fat pad and synovium were isolated from the samples and snap frozen at -80$^{o}$C on the day of surgery.

### 2.3 Nucleic acid extraction from human tissue

Having been snap frozen at -80$^{o}$C on the day of surgery, human tissue was subsequently ground under liquid nitrogen with a Retsch Mixer Mill MM 200 (Retsch, Leeds, UK).

#### 2.3.1 Nucleic acid extraction from fat pad and synovium

Extraction of genomic DNA (gDNA) and RNA from fat pad and synovium was performed from 250 mg of ground tissue using the Omega E.Z.N.A. DNA/RNA Isolation Kit (Omega Bio-Tek, Georgia, USA), according to manufacturer protocol. Nucleic acids were quantified using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, USA). Samples were diluted in diethylpyrocarbonate (DEPC)-treated $H_2O$ (Invitrogen, Life Technologies, UK). DNA was stored at -20$^{o}$C at 50 ng/µl and RNA at -80$^{o}$C at a concentration of 250 ng/µl.

#### 2.3.2 Nucleic acid extraction from cartilage

gDNA was extracted from 250 mg ground cartilage tissue using the Omega E.Z.N.A. DNA/RNA Isolation Kit (Omega Bio-Tek, Georgia, USA), according to manufacturer protocol. Briefly, ground tissue was vortexed with 700 µl GTC Lysis Buffer, then centrifuged for 5 minutes (min) at 11,700 revolutions per minute (rpm). The supernatant was transferred to a HiBind® column and centrifuged at 10,300 rpm for 1 min. The centrifugation was repeated after the addition of HB Buffer, and then following addition of DNA Wash Buffer. The column membrane was then dried by centrifugation at 13,000 rpm for 2 min. The DNA was then eluted in 100 µl Elution Buffer.

To extract RNA, 1 ml of TRIzol Reagent (Ambion, Life Technologies, UK) was added to 250 mg of ground cartilage tissue and subsequently homogenised by vortexing thoroughly. Samples were then incubated at room temperature for 15 min and subsequently centrifuged for 3 min at 13,000 rpm at 4$^{o}$C. The supernatant was transferred to a fresh Eppendorf tube and 200 µl of chloroform was added. The tube was shaken vigorously for 15 seconds (sec), then incubated at room temperature for 3 min. The gDNA, total RNA and protein phases were then separated by centrifugation for 15 min at 13,000 rpm and 4$^{o}$C. RNA, contained in the top aqueous phase, was transferred to a fresh tube and total RNA was then extracted using an RNeasy kit (QIAGEN, Crawley, UK) according to the manufacturer's instructions. The remaining phases were discarded. The eluted RNA was quantified using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, USA) and stored at -80$^{o}$C, at a concentration of 250 ng/µl.

### 2.4 Polymerase chain reaction optimisation

For all primer pairs, the polymerase chain reaction (PCR) annealing conditions were optimised prior to use (Appendix A: Table A.1, Table A.2 and Table A.5). AmpliTaq Gold® *Taq* Polymerase (Applied Biosystems, Life Technologies, USA) was used for the basic PCR reactions for genotyping, pyrosequencing and cloning, while Titanium® *Taq* DNA Polymerase (Clontech Laboratories, Inc., France) was used for PCR using bisulphite converted DNA.

### 2.4.1 Optimisation of PCR conditions using AmpliTaq Gold® Taq DNA Polymerase

Annealing temperatures from 55ºC to 70ºC were tested along with $MgCl_2$ concentrations of 1 mM, 1.50 mM and 2 mM to determine the optimum conditions. PCR reactions were carried out in a total volume of 15 µl, containing: 1 x PCR Buffer II (Applied Biosystems, Life Technologies, USA), 0.50 µM of both the forward and reverse primers (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 cycling conditions were as follows: initialisation of reaction for 14 min at 94ºC, denaturation for 30 sec at 94ºC, annealing for 30 sec and extension for 1 min per 1 kb of template, at 72ºC. This was performed for 35 cycles, followed by an elongation step for 5 min at 72ºC. To confirm a product had been amplified, 4 µl of the PCR product was loaded onto a 2% weight/volume (w/v) agarose Tris/borate/EDTA (TBE) gel containing 20 µg ethidium bromide per 100 ml total volume. The gel was run at 100 V for approximately 30 min, and the products were then visualised using a G:BOX gel visualisation system (Syngene, UK) under ultraviolet (UV) light.

### 2.4.2 Optimisation of PCR conditions using Titanium® Taq DNA Polymerase

Annealing temperatures from 55ºC to 70ºC were tested to identify the optimum condition. Each reaction comprised a total volume of 20 µl, containing: 1 x Titanium® *Taq* PCR Buffer (Clontech Laboratories, Inc., France), 0.20 µM of both the forward and reverse primers (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 bisulphite converted DNA (Chapter 2.14). PCR cycling conditions were as follows: initialisation for 5 min at 95ºC, denaturation for 30 sec at 95ºC, annealing for 30 sec and extension for 1 min at 68ºC. This was repeated for 35 cycles, followed by an elongation step for 5 min at 68ºC. To confirm a product had been amplified, 4 µl of the PCR product was loaded onto a 2% weight/volume (w/v) agarose Tris/borate/EDTA (TBE) gel containing 20 µg ethidium bromide per 100 ml total volume. The gel was run at 100 V for approximately 30 min, and the products were then visualised using a G:BOX gel visualisation system (Syngene, UK) under ultraviolet (UV) light.

## 2.5 Genotyping at the sentinel SNP by restriction fragment length polymorphism (RFLP)

### 2.5.1 Restriction fragment length polymorphism (RFLP) primer design to rs4836732

In order to establish the genotype at rs4836732, the DNA sequence flanking ~200 bp either side of the SNP was obtained using the UCSC Genome Browser. Forward and reverse primers were then designed to the SNP using the Primer3 Input online software (http://primer3.wi.mit.edu/; Appendix A: Table A.1) [123]. A BLAST-like alignment tool (BLAT) search of the primers was performed, using the UCSC Genome Browser BLAT software [124], to ensure the primers were specific to the SNP. The NEBcutter online software (http://tools.neb.com/NEBcutter2/) was used to identify that the restriction enzyme *RsaI* cut at the SNP site [125].

### 2.5.2 Restriction fragment length polymorphism (RFLP) assay

DNA amplification was performed using AmpliTaq Gold® Taq DNA Polymerase, as described in Chapter 2.4.1, using 2 mM $MgCl_2$ and an annealing temperature of 55$^{o}$C. 7.50 µl of amplified DNA was then digested in a final volume of 15 µl containing: 5 U *RsaI* restriction enzyme (New England BioLabs, UK) and 1 x reaction buffer (New England BioLabs, UK) as recommended by the manufacturer. The reaction was incubated at 37$^{o}$C overnight and the resulting DNA fragments were the 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). In the presence of a T allele, a 202 bp band would be present, while 36 bp and 166 bp bands would be produced in the presence of the C allele.

## 2.6 SNP genotyping by pyrosequencing

Pyrosequencing primers (Sigma-Aldrich, UK) were designed using PyroMark Assay Design Software 2.0 (QIAGEN, Crawley, UK) and PCR cycling conditions were optimised as described in Chapter 2.4.1 (Appendix A: Table A.1). gDNA was then PCR amplified and 10 µl of the amplified product was added to the wells of a 0.20 ml 24-well PCR plate (STARLAB, Milton Keynes, UK). The product was agitated for 10 min 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. Using a PyroMark Q24 Vacuum Workstation (QIAGEN, Crawley, UK), the sepharose beads were captured on 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. The plate was heated for 2 min at 80$^{o}$C to denature the gDNA, followed by primer annealing for 5 min at room temperature. PyroMark Gold Q24 Reagents (QIAGEN, Crawley, UK) were loaded into a PyroMark Cassette (QIAGEN, Crawley, UK) and pyrosequencing was performed in a PyroMark Q24 Pyrosequencing machine (QIAGEN, Crawley, UK). Prior to genotype analysis of patient samples, the ability of the assays to differentiate between allelic quantities was validated in duplicate by comparing known synthetic allelic ratios with the ratios detected by the Pyrosequencing machine (Chapter 3: Figure 3.15).

## 2.7 Luciferase reporter assays

### 2.7.1 Cloning of DNA into pGL3-Promoter Luciferase Reporter Vectors

Both the minor and major alleles of rs4836732, rs4837613 and rs13283416 were PCR amplified in two 20 µl reactions, using primers introducing *Xho*I and *Mlu*I restriction enzyme sites (for primer sequences and conditions see Appendix A: Table A.2). Each reaction contained 1x PCR Buffer II (Applied Biosystems, Life Technologies, USA), 0.5 µM forward primer (Sigma-Aldrich, UK), 0.5 µM reverse primer (Sigma-Aldrich, UK), 200µM dNTPs (Bioline, UK), 2 mM $MgCl_2$, 0.4 U AmpliTaq Gold® *Taq* Polymerase (Applied Biosystems, Life Technologies, USA) and 50ng DNA. The products were purified using the QIAGEN Gel Extraction Kit, according to manufacturer protocol (QIAGEN, Crawley, UK). A double digest was performed with 1µg of the purified PCR

product and pGL3-Promoter Vector using 10U of *Mlu*1 and *Xho*I restriction enzymes (New England BioLabs, UK), 1 x NEB Buffer 3 (50 mM Tris-HCl, 10 mM $MgCl_2$, 100 mM NaCl and 1 mM DTT, pH 7.9) in a 30µl reaction. The digest was performed overnight at 37°C. Following digestion, the PCR product and pGL3-Promoter Vector were electrophoresed through a 1% TAE EtBr gel and extracted under UV light. Subsequent purification was performed using the QIAquick Gel Extraction Kit (QIAGEN, Crawley, UK) following manufacturer instructions. The products were eluted in 50 µl of DEPC water.

The products were ligated into the digested pGL3-Promoter Vector using a 5:1 fragment to plasmid ratio, along with 400 U T4 ligase (New England BioLabs, UK) and 10x T4 buffer (New England BioLabs, UK) and incubated overnight at 16ºC. Plasmid DNA was then transformed into One Shot® Mach1™ T1 Phage-Resistant Chemically Competent *E. coli* cells (Invitrogen, Life Technologies, UK), according to manufacturer guidelines. Transformed cells were plated onto LB agar plates containing 100 µg/ml ampicillin (Sigma-Aldrich, UK) and incubated overnight at 37°C. Between 8 and 10 colonies were selected and individually transferred into 1ml LB media and 1/1000 ampicillin. Successful insertion of the PCR fragment into the plasmid was confirmed by colony PCR using previously designed primers. Products were run on a 1% agarose TBE EtBr gel and 3 µl of each positive clone was then placed in 3 ml LB broth (Sigma-Aldrich, UK) and 100 µg/ml ampicillin (Sigma-Aldrich, UK), then grown overnight with shaking at 37ºC and 180 rpm.

Glycerol stocks were prepared by adding a 1:1 ratio of 200 µl of bacterial culture to 200µl of glycerol (Sigma-Aldrich, UK) and stored at -80ºC. The identity of the DNA insert was confirmed using sequencing performed by Source BioScience, UK to confirm that the sequences were correct, and only differed at the SNP of interest.

### 2.7.2 Purification of plasmid DNA by miniprep

For transfection, plasmid cultures were purified using the Promega PureYield™ Plasmid Miniprep System (Promega, UK). 5µl of the glycerol bacteria stock was added to 3 ml of LB broth containing ampicillin (100 µg/ml) and incubated at 37°C with shaking for 2-3 hours. 1.5 ml of the overnight culture was centrifuged in an Eppendorf tube at 13,000 rpm for 30 sec and the supernatant discarded. This was repeated with the remaining 1.5 ml of culture. The cell pellet was resuspended in 600 µl DEPC water and 100 µl Cell Lysis Buffer was added, then mixed by inverting the tube 6 times for a maximum of 2 min. Following addition of 350 µl of cold Neutralisation Solution, the tube was inverted again to mix thoroughly, with subsequent centrifugation for 3 min at 13,000 rpm. The supernatant was transferred to a PureYield™ Minicolumn and centrifuged at 13,000 rpm for 15 sec. The flow through was then discarded and 200 µl Endotoxin Removal Wash added. The mix was centrifuged at 13,000 rpm for 15 sec and the flow through discarded. 400 µl Column Wash Solution was added before centrifuging at 13,000 rpm for 30 sec. The Minicolumn was then transferred to a new Eppendorf tube and 30 µl of DEPC water added. Following incubation at room temperature for 1 min, plasmid DNA was eluted by centrifuging at 13,000 rpm for 15 sec. DNA was then diluted to achieve a

concentration of 50 ng/µl, using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, USA), and stored at -20$^{o}$C until required.

### 2.7.3. Cell line culture

Three cell lines were used for luciferase reporter analyses, SW872 (human liposarcoma line from a 36 year old Caucasian male), SW1353 (chondrosarcoma line from a 72 year old Caucasian female) and U2OS (osteosarcoma line from a 15 year old Caucasian female). The cells were cultured in the appropriate culture medium, as follows, in Corning vented T75cm$^2$ flasks (Sigma-Aldrich, UK) at 37°C in a 5% (v/v) $CO_2$ atmosphere to 80% confluency.

SW872 medium- Dulbecco's modified eagle's medium (DMEM): Hams F12 Glutamax medium in a 3:1 ratio (GIBCO, Life Technologies, UK) containing 5% (v/v) fetal bovine serum (FBS), 100 U/ml penicillin and 100 µg/ml streptomycin (Sigma-Aldrich, UK).

SW1353 medium- DMEM: Hams F12 medium in a 1:1 ratio (GIBCO, Life Technologies, UK) containing 10% (v/v) FBS, 100 U/ml penicillin, 100 µg/ml streptomycin and 2 mM of L-glutamine (Sigma-Aldrich, UK).

U2OS medium- DMEM containing 10% v/v FBS, 2 mM L-glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin (Sigma-Aldrich, UK).

### 2.7.4 Transfection of cell lines

SW1353 and U2OS cells were seeded at a density of 6000 cells per well, while SW872 cells were seeded at 10,000 cells per well, in a Corning Costar 96-well cell culture plate (Sigma-Aldrich, UK) and cultured for 24 hours prior to transfection at 37ºC with 5% $CO_2$. Cells in each well were transfected with a 100 µl master mix containing 50 ng luciferase vector DNA, 1.5 ng *Renilla* Luciferase Reporter Vector (Promega, UK) and 0.3 µl per well of FuGene HD Transfection Reagent (Promega, UK) in the appropriate medium. Eight technical and six biological repeats were performed per SNP allele. Cells were then incubated in a 37ºC 5% $CO_2$ incubator overnight.

### 2.7.5 Luciferase assay readings

After 24 hours, transfected 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. 20 µl of the lysate was used to quantify the luciferase and *Renilla* activity using the Dual Luciferase Assay system (Promega, UK) with the MicroLumat Plus LB96V Luminometer (Berthold Technologies UK, Harpenden, UK). Luciferase values were normalised to the appropriate *Renilla* luciferase values and subsequently normalised to the mean of the corresponding empty pGL3 vector control. The Mann-Whitney *U* test was used to assess statistical significance.

## 2.8 Electrophoretic mobility shift assay (EMSA)

### 2.8.1 Prediction of protein binding sites

A publicly available online search tool was used to predict protein binding to rs4836732: PROMO version 3.0 (http://alggen.lsi.upc.es) [126, 127]. The UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway) was also used as a predictor of protein binding, based on experimental cell line chromatin immunoprecipitation (ChIP) assays with sequencing (ChIP-Seq) data from ENCODE [124, 128].

### 2.8.2 Nuclear protein extraction method

Two buffers were used for the preparation of nuclear extract for EMSA experiments:

Hypotonic Buffer -10 mM HEPES, pH 7.6, 1.5 mM $MgCl_2$, 10 mM KCl, 1 mM DTT, 10 mM NaF, 1 mM Na3VO4, 0.1% Tergitol (v/v), 1x complete protease inhibitor cocktail tablet per 50 ml solution (Roche, UK).

High Salt Buffer -20 mM HEPES, pH 7.9, 420 mM NaCl, 20% glycerol (v/v), 1 mM DTT, 10 mM NaF, 1 mM Na3VO4, 1x complete protease inhibitor cocktail tablet per 50 ml of buffer (Roche, UK).

Cells were seeded onto Corning 500 $cm^2$ cell culture dishes (Sigma-Aldrich, UK) until 70-80% confluent. Culture medium was then removed and cells were washed in ice cold PBS. Cells were then removed from the surface into 5 ml of fresh PBS and centrifuged for 30 sec at 10,000g at 4°C. The cell pellet was resuspended in 1 ml of hypotonic buffer and incubated on ice for 15 min. Cells were centrifuged as before

and the pellet was resuspended in ice cold hypotonic buffer, supplemented with 0.25 M sucrose to fractionate the nuclei. Cells were centrifuged and the pellet was resuspended in 1 ml high salt buffer. This was incubated for 30 min on ice, then cells were centrifuged at 10,000g for 2 min at 4°C. The supernatant containing nuclear protein was immediately stored at -80°C.

### 2.8.3 Preparation of probes

Fluorescently labelled (5′DY682) C and T oligonucleotides (Eurofins MWG Operon, Ebersberg, Germany) spanning 15 bp upstream and 15 bp downstream of each allele were resuspended in DEPC $H_20$ (Invitrogen, Life Technologies, UK) at a concentration of 100 pmol/µl (For sequences see Appendix A: Table A.3). The complementary forward and reverse oligonucleotides were then heated for 5 min at 95$^o$C with annealing buffer (100 mM Tris-HCl, 500 mM NaCl, 10 mM EDTA), before cooling to room temperature and storage at -80$^o$C. The annealed probes were subsequently diluted in DEPC to achieve a working stock of 100 fmol, which was stored at -20$^o$C.

### 2.8.4 EMSA reagents

- Nuclear proteins

- Fluorescently labelled oligonucleotides (Eurofins MWG Operon, Ebersberg, Germany)

- EMSA Annealing Buffer: 100 mM Tris HCl, 500 mM NaCl, 10 mM EDTA.

- 5x TBE buffer: 445 mM Tris, 445 mM Boric Acid, 10 mM EDTA, pH8.

- 5% (weight/volume (w/v)) acrylamide gel (prepared with TEMED and ammonium persulfate (APS)).

- Buffers and loading dye: Odyssey Infrared EMSA kit (LiCor Biosciences, Cambridge, UK).

- rs4836732 Optimal Binding Reaction: 1x Binding Buffer, 25 mM DTT, 2.5% Tween-20, 1 µg/µl Poly (dI:dC), 100 mM $MgCl_2$, 200 fmol annealed oligonucleotide and 5 µg nuclear extract.

- 5x (1 pmol), 10x (2 pmol), 25x (5 pmol), 50x (10 pmol) unlabelled competitor oligonucleotide (For competitor sequences see Appendix A: Table A.3).

- 2 µg candidate antibody (Appendix C: Table C.1).

### 2.8.5 EMSA gel preparation and electrophoresis

The native 5% (v/v) acrylamide gel (1 x TBE, 1:1,000 TMED and 0.07% [v/v] ammonium persulphate) was prepared and set at 4°C overnight. The gel was then run for 30 min in 0.5x TBE buffer at 100 V 4ºC prior to loading of samples in order to remove traces of APS and equilibrate ions in the running buffer. The 20 µl binding reactions were prepared using the above Odyssey Infrared EMSA kit components (LiCor Biosciences, Cambridge, UK) and incubated in the dark at room temperature for 20 min. Each 20 µl reaction contained 200 fmol of labelled probe. For competition EMSAs, unlabelled competitors were added at 5 x (1 pmol), 10 x (2 pmol), 25 x (5 pmol), and 50 x (10 pmol) that of the labelled probe concentration. For supershift EMSAs, 3 µg or 6 µg of antibody was added to the reaction (Appendix C: Table C.1).

Appropriate control reactions were used containing either no competitor or a species-matched IgG antibody for competition EMSAs and supershift EMSAs respectively. Orange G loading dye was then added (1x final concentration) and the samples were loaded onto the gel. Electrophoresis was performed at 100V at 4°C in the dark for approximately 4 hours/ until the dye front reached the end of the gel. The gel was visualised using the LiCor Odyssey Infrared Imager (LiCor Biosciences, Cambridge, UK).

## 2.9 Complementary DNA (cDNA) synthesis

### 2.9.1 Reverse transcriptase polymerase chain reaction (RT-PCR) method

In order to remove contaminating DNA, 1 µg of total RNA was incubated with 1 U TURBOTM DNase (Invitrogen, Life Technologies, UK) and 1 x TURBOTM DNase Buffer (Invitrogen, Life Technologies, USA) at 37$^{o}$C for 30 min. The DNase was then inactivated by addition of 100 mM EDTA and incubation at 75$^{o}$C for 10 min. The total RNA was then incubated at 65$^{o}$C for 5 min in a volume of 8 µl containing: 1 µg random primers (Invitrogen, Life Technologies, USA), 10 µM dNTP mix and the required volume of diethylpyrocarbonate (DEPC)-treated water (Invitrogen, Life Technologies, USA). The reaction was then incubated for 1 min at 25$^{o}$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 RNaseOUTTM (Invitrogen, Life Technologies, USA). Reverse transcription of the total RNA into complementary DNA (cDNA) was then performed following the addition of 200 U SuperScriptTM II Reverse Transcriptase (Invitrogen, Life Technologies, USA). The reaction was incubated for 10 min at 25$^{o}$C, 50 min at

42°C and 10 min at 70°C. Any remaining complementary RNA was then degraded by incubation at 37°C for 20 min, following the addition of 2 U *E. coli* RNase H (Invitrogen, Life Technologies, USA). cDNA was stored at -20°C.

### 2.9.2 Polymerase chain reaction (PCR) to assess complementary DNA (cDNA) integrity

After synthesis by RT-PCR, cDNA integrity was assessed by PCR amplification with *HBP1* primers, which span an intron:exon boundary to enable cDNA and residual gDNA contamination to be distinguished (for primer sequences see Appendix A: Table A.4). In a total volume of 15 µl, 0.50 µl of cDNA added to a 14.50 µl master mix containing 2 mM $MgCl_2$ and PCR amplified using an annealing temperature of 60°C as described in Chapter 2.4.1. Only cDNA containing no contaminating gDNA was carried forward for downstream applications.

## 2.10 Allelic expression imbalance analysis by pyrosequencing

### 2.10.1 Online database search for transcript single nucleotide polymorphisms (SNPs)

Transcript SNPs in each gene of interest were identified using the online software from the UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway) [124]. The Broad Institute (http://www.broadinstitute.org/mpg/snap/) online software was used to conduct pairwise searches to assess the LD between the transcript SNPs and rs4836732 [129].

### 2.10.2 Pyrosequencing method

PyroMark Assay Design Software 2.0 (QIAGEN, Crawley, UK) was used to design pyrosequencing primers for transcript SNPs within the genes of interest (for primer sequences see Appendix A: Table A.1). PCR conditions for the primers (Sigma-Aldrich, UK) were optimised as described in Chapter 2.4.1. The ability of the 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 (Chapter 3: Figure 3.15). gDNA and cDNA from cartilage, fat pad and synovium samples were PCR amplified and the products were then taken forward for pyrosequencing, as described in Chapter 2.6. Each reaction was performed in triplicate for DNA and cDNA and significance was assessed using a Mann-Whitney *U* test.

## 2.11 Cell culture for *PAPPA* siRNA knock down analyses

### 2.11.1 SW1353 cell culture

SW1353 (chondrosarcoma line from a 72 year old Caucasian female) cells were cultured in DMEM: Hams F12 medium in a 1:1 ratio (GIBCO, Life Technologies, UK) containing 10% (v/v) FBS, 100 U/ml penicillin, 100 µg/ml streptomycin and 2 mM of L-glutamine (Sigma-Aldrich, UK). Cells were cultured in Corning vented T75cm$^2$ flasks (Sigma-Aldrich, UK) at 37°C in a 5% (v/v) $CO_2$ atmosphere to 80% confluency.

### 2.11.2 Human mesenchymal stem cell (hMSC) culture

hMSCs from OA patients were isolated from trabecular bone under sterile conditions in a laminar flow hood. Extracted bone was passed through a 100µm Cell Strainer (BD Biosciences, Oxford, UK) with PBS containing 100U/ml penicillin and 100µg/ml streptomycin (Sigma-Aldrich, UK). hMSCs were isolated via Ficoll density centrifugation (GE Healthcare, UK) and seeded with hMSC medium containing: 1g/L glucose DMEM (GIBCO, Life Technologies, UK), supplemented with 10% (v/v) FBS, 100 U/ml penicillin, 100 µg/ml streptomycin and 2 mM of L-glutamine (Sigma-Aldrich, UK). Cells were cultured in Corning vented T75cm$^2$ flasks (Sigma-Aldrich, UK) at 37°C in a 5% (v/v) $CO_2$ atmosphere to 80% confluency. Cell purity was assessed by flow cytometric analysis (FACs) to confirm a single phenotypic population positive for the defined hMSC cell surface markers CD29, CD44, CD71 and CD106.

hMSCs from individuals without OA (Lonza, UK) were seeded at a density of 17,500 cells/cm$^2$ and cultured at 37$^o$C 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 then replaced with 1 g/L glucose Dulbecco's modified Eagle's medium (DMEM) (GIBCO, Life Technologies, UK), supplemented with 20% (volume/volume (v/v)) fetal bovine serum (FBS), 8 ng/µl bovine- fibroblast growth factor (FGF) (Millipore, UK), 2 mM of L-glutamine, 100 U/ml penicillin and 100 µg/ml streptomycin (Sigma-Aldrich, UK). Cells were incubated in standard cell culture conditions: a sterile, 37°C, humidified, 5% $CO_2$/95% air environment, with cell medium refreshed every 72 hours.

### 2.11.3 RNA mediated interference (RNAi) of *PAPPA* in human mesenchymal stem cell (hMSC) and SW1353 cells

On-TARGETplus SMART pool human *PAPPA* siRNA (Dharmacon, GE Healthcare, UK, cat no. L-005130-00-0005) was resuspended in 250 µl of 1x siRNA Buffer (Dharmacon, GE Healthcare, UK) to make 20 µM siRNA. On-TARGETplus Non-targeting (NT) pool siRNA (Dharmacon, GE Healthcare, UK, cat no. D-001810-10-20) at 20µM was used as a control (for sequences see Appendix D: Table D.1).

hMSCs and SW1353 were seeded at a density of 1.2 million cells into three 10 cm tissue culture dishes per cell line (for untransfected, NT siRNA and *PAPPA* siRNA). The cells were cultured at 37 °C for 24 hours in DMEM media to reach 80% confluency. 35 µl siRNA and 665 µl serum free DMEM media (SFM) were combined and incubated for 5 min at room temperature as were 14 µl Dharmafect transfection reagent 1 (Dharmacon, GE Healthcare, UK) and 686 µl SFM. The two reaction mixes were then combined (total volume 1.4 ml) and incubated for a further 20 min at room temperature. Cells were transfected by removing the appropriate amount of media before adding the transfection mix, followed by incubation at 37 °C for 48 hours. Subsequently, cells were lifted off the surface using trypsin and counted. hMSCs were seeded into 12 well plates for osteogenesis (100,000 cells/well in 1 ml osteogenic medium) or into 96 well v-bottom plates for chondrogenesis (30,000 cells/well in 200 µl chondrogenic medium) for differentiation according to the previously described protocols. Untransfected cells and cells transfected with NT siRNA were used as controls and three repeats were performed per condition in each cell line, with an additional well seeded for all three treatment groups on day three for immunoblot analysis (Chapter 2.12).

### 2.11.4 Human mesenchymal stem cell (hMSC) differentiation into osteoblasts

hMSCs were seeded in a 12 well culture plate at a density of 1.3 x $10^4$/$cm^2$ in standard hMSC medium and incubated for three hours to adhere, before adding 1 ml fresh osteogenic media (1 g/L DMEM, 2 mM of L-glutamine, 100 U/ml penicillin, 100 µg/ml streptomycin, 10% FBS, 0.1 µM dexamethasone, 10 mM β-glycerophosphate, 50 µM ascorbic acid). Media was refreshed every 72 hours, with ascorbic acid added freshly to the medium each time. Cells were harvested on days 0, 3 and 10 of differentiation. On day 3, differentiated cells were stained for alkaline phosphatase to confirm successful osteogenic differentiation. Wells were rinsed in ice-cold PBS followed by 70% ethanol before fixing in 4% PFA for 10 min. Alkaline phosphatase activity was detected following incubation in a 0.1% Fast blue RR salt/ napthol phosphate solution (Sigma-Aldrich, UK) at room temperature for 45 min. RNA was extracted from cells and used to synthesise cDNA in confirm differentiation had occurred by real-time gene expression analysis of *ALPL*, *RUNX2* and *COL1A1* (Chapter 2.13).

### 2.11.5 Human mesenchymal stem cell (hMSC) differentiation into chondrocytes

Following monolayer culture to 80% confluency, trypsinised hMSCs were seeded at a density of 0.3x$10^5$/well into a UV-irradiated V bottomed 96 well plate (Greiner Bio-One Ltd, Stonehouse, UK) with 150 µl of chondrogenic differentiation media (1g/L DMEM, 2 mM L-glutamine, 100 U/ml penicillin, 100 µg/ml streptomycin, 1% ITSL premix, 40 µg/mL L-proline and 100 nM dexamethasone). The plate was then centrifuged at 500g for 5 min and incubated at 37°C in a humidified, 5% $CO_2$/ 95% air environment. Media was refreshed every 3 days and cells were harvested for

analysis on days 0, 3 and 10 of chondrogenesis. Successful differentiation was confirmed by real-time gene expression analysis of *ACAN* and *SOX9* (Chapter 2.13).

## 2.12 Immunoblotting

### 2.12.1 Protein extraction and quantification

Total protein was extracted from cells on day 3 of differentiation for all conditions using a spin column extraction kit according to the manufacturer's instructions (Nucleospin RNA/protein, Macherey-Nagel, supplied by Thermo Fisher Scientific, UK). The extracted total protein was snap-frozen on dry ice, before storage at -80$^{o}$C. Protein was quantified in a microplate using the MACHEREY-NAGEL Protein Quantification Assay, according to the manufacturer's microplate assay procedure protocol. Absorbances were read at 595 nm using a Tecan Sunrise Microplate Absorbance Reader (Tecan, Reading, UK).

### 2.12.2 Immunoblotting method

To confirm successful PAPPA depletion, equal amounts of extracted protein were resolved on a NUPAGE 4-12% Bis-Tris GEL (Thermo Fisher Scientific, UK) before being transferred to a Immobilon-P polyvinylidene fluoride (PVDF) membrane (Merck Millipore, UK) by electroblotting in a Scie-Plas V20-SDB 20x20 semi-dry blotter (Scie-Plas, Cambridge, UK). The membrane was then blocked for non-specific protein binding by incubation at room temperature for 30 min in 1x PBS containing 5% (v/v) Marvel milk and 0.02% (v/v) Tween-20. The membrane was washed for 5 min in 1 x PBS containing 0.02% (v/v) Tween-20, then incubated at 4°C overnight with the PAPPA (B-7) mouse monoclonal primary antibody (Santa Cruz

Biotechnology, USA; Catalogue number sc-365226), raised against amino acids 666-840, diluted 1:2,000 in 1 x PBS containing 5% (w/v) Marvel milk and 0.02% (v/v) Tween-20. The membrane was washed as previously described, followed by incubation with the polyclonal horse anti-mouse horseradish peroxidise (HRP) conjugated secondary antibody (Cell Signalling Technologies inc., MA, USA; Catalogue number 7076S), diluted 1:2,000, for 1 hour at room temperature. A rabbit monoclonal GAPDH primary antibody (Cell Signalling Technologies inc.; Catologue number 5174S) was used as a loading control (1:10,000) with subsequent incubation using goat anti-rabbit horseradish peroxidise (HRP) conjugated secondary antibody (Cell Signalling Technologies inc., MA, USA; Catalogue number 7074S), diluted 1:2,000. Detection was carried out using X-ray film (Kodak Film, Sigma-Aldrich, UK) with Immobilon Western Chemiluminescent HRP Substrate (Millipore, UK) in a G:BOX gel visualisation system (Syngene, UK).

### 2.13 Quantitative real-time polymerase chain reaction (qPCR) following *PAPPA* siRNA knock down

#### 2.13.1 cDNA synthesis

For the siRNA experiments, cells were lysed in 30 µl Cells-to-cDNA II cell lysis buffer (Ambion, Life Technologies, UK) and transferred to a 96 well plate before being heated to 75$^{o}$C for 15 min. 10 ng random primers (Invitrogen, Life Technologies, UK) and 10 mM dNTP mix (Invitrogen, Life Technologies, UK) were added to 8 µl of the cell lysate and incubated at 70$^{o}$C for 5 min. To the reaction, 1x reaction buffer (50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM $MgCl_2$; Invitrogen), 100 U MMLV reverse transcriptase (Invitrogen, Life Technologies, UK) and 10 mM DTT (Invitrogen, Life

Technologies, UK) were added, along with DEPC water to create a final reaction volume of 20 µl. The samples were incubated at 37 °C for 50 min and 75 °C for 15 min. Then 30 µl of water was added and samples were stored at -20ºC.

### 2.13.2 Quantitative Real-Time Polymerase Chain Reaction (qPCR) using PrimeTime Quantitative Real-Time-PCR (qPCR) Assays

cDNA was diluted 1:20 and 2.5µl was then used in 10 µl total volume reactions with TaqMan® Fast Universal PCR Master Mix (Applied Biosystems, Life Technologies, USA) and PrimeTime qPCR Assays (IDT, Iowa, USA; For primer and probe sequences, see Appendix A: Table A.4). The qPCR reaction was carried out using the ABI PRISM 7900HT Sequence Detection System (Applied Biosystems, Life Technologies, USA) with fast cycling conditions (95°C for 20 sec followed by 40 cycles of 95°C for 1 sec and 60°C for 20 sec) in MicroAmp® Fast Optical 96-well Reaction Plates (Applied Biosystems, Life Technologies, USA). Gene expression was normalised to the housekeeping genes *18S*, *HPRT1* and *GAPDH* using the delta delta ct method (2-(ct test gene)-(ct mean of housekeepers)). Each sample was analysed in triplicate.

## 2.14 CpG methylation analysis

For methylation analysis, 1 µg E.Z.N.A – extracted cartilage genomic DNA (Omega Bio-Tek, Georgia, USA) was bisulphite converted using the EZ DNA Methylation™ Kit (Zymo Research, California, USA) according to manufacturer's instructions, and eluted in a volume of 20 µL. Bisulphite DNA was stored at -20ºC.

### 2.14.1 Polymerase chain reaction (PCR) of bisulphite converted DNA

PyroMark Assay Design Software 2.0 (QIAGEN, Crawley, UK) was used to design pyrosequencing primers to the selected CpG sites (for primer sequences see Appendix A: Table A.5). PCR conditions for the primers (Sigma-Aldrich, UK) were optimised as described in Chapter 2.4.2. PCR was performed on 50 ng bisulphite converted DNA using 1 x Titanium® *Taq* DNA Polymerase (Chapter 2.4.2).

### 2.14.2 CpG methylation analysis by pyrosequencing method

DNA methylation analysis was performed on a PyroMark Q24 workstation (QIAGEN, Crawley, UK) and analysed with the PyroMark CpG software (QIAGEN, Crawley, UK), as previously described (Chapter 2.6). Prior to methylation analysis of patient samples, the ability of the assays to detect methylation levels was validated by comparing expected methylation values, as determined by the high-density CpG methylation array, with the methylation detected by the Pyrosequencing machine. Each sample was analysed in duplicate and the mean methylation value calculated.

### 2.14.3 Batch correction

The methylation values derived from the β values provided by the high-density CpG methylation array at CpG sites cg12245706 and cg21944234 were batch corrected, using a regression of the difference in observed and expected values as a function of beta values. The correction factor (as determined below) was then applied to the methylation value derived from the original β value:

cg12245706 correction factor = (-29.83 x β value) – 1.77

cg21944234 correction factor = (66.83 x β value) – 19.20

### 2.14.4 Statistics

Statistical analysis of CpG methylation were performed with the aid of JMP Pro Version 11.2 Statistical Data Visualization Software (JMP, Marlow, Buckinghamshire).

## 2.15 Isolation of patient serum

Prior to joint arthroplasty surgery, approximately 3 ml of peripheral blood was collected in a BD Gold-top serum separating Vacutainer ®, containing clotting activator (Becton, Dickinson and Company, Franklin Lakes, New Jersey) from consenting patients. This was spun at room temperature at 1800g for 12 min, and the clear serum was aliquoted into 1.2 ml Eppendorf tubes. Serum was frozen at -80°C until subsequent use.

## 2.16 Electrochemiluminescent Meso Scale Discovery (MSD) assays

### 2.16.1 Method

For the electrochemiluminescence studies of human serum, the V-PLEX Human leptin kit, V-PLEX custom Human Cytokine, V-PLEX Human Proinflammatory Panel 1 and V-PLEX Human CRP kit (catalogue numbers K151BYC-1, K151A0H-1, K15049D-1 and K151STD-1 respectively) were used according to the manufacturer's instructions (Mesco Scale Discovery (MSD), Gaithersburg, USA).

MSD plates were analysed on the MSD Analysis-SECTOR imager (MSD, Gaithersburg, USA). Standards were measured in duplicate and patient samples in singlicate.

### 2.16.2 Statistics

Statistical analyses were performed using Graph-Pad Prism 5.0 (GraphPad Software, La Jolla, USA). Analysis of two groups of data was performed using the Wilcoxon–Mann–Whitney test. *p<0.05 **p<0.01 ***p<0.001. For comparison of multiple groups, Kruskal-Wallis one-way ANOVA was used to calculate statistical significance. *p<0.05 **p<0.01 ***p<0.001. Where appropriate, Bonferroni multiple comparison correction was applied.

## 2.17 Quantitative real-time polymerase chain reaction (qPCR) of inflammatory genes in OA cartilage

### 2.17.1 Quantitative Real-Time Polymerase Chain Reaction (qPCR) method using SYBR® Green technology

Cartilage RNA was extracted as previously described (Chapter 2.3.2) and reverse transcriptase polymerase chain reaction (RT-PCR) performed to syntheses cDNA (Chapter 2.9). Subsequently, inflammatory gene expression was quantified in a 10µl total volume containing: KiCqStart® SYBR® Green qPCR ReadyMix™ with ROX (Sigma-Aldrich, UK), 0.5 µM of both the forward and reverse primers (Sigma-Aldrich, UK) and 2.5µl of 1:20 diluted cDNA. For primer sequences see Appendix A: Table A.6. Each reaction was performed in triplicate using the ABI PRISM 7900HT Sequence Detection System (Applied Biosystems, Life Technologies, USA). The

thermal cycling conditions were: 95$^{o}$C for 10 min, 40 cycles of 95$^{o}$C for 3 sec and 60$^{o}$C for 35 seconds, followed by a dissociation stage of 95$^{o}$C for 15 sec, 60$^{o}$C for 15 sec and 95$^{o}$C for 15 sec. Expression of target genes was measured relative to the housekeeping gene *GAPDH*, and relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method, as described previously (Chapter 2.13.2).

### 2.17.2 Statistics

Statistical analyses of inflammatory gene expression were performed using Graph-Pad Prism 5.0 (GraphPad Software, La Jolla, USA) using the Wilcoxon–Mann–Whitney test or the Student's *t* test. *p<0.05 **p<0.01 ***p<0.001.

# Chapter 3: Functional analysis of the rs4836732 signal

## 3.1 Introduction

One of the five genome-wide significant loci associated with OA in the arcOGEN GWAS was marked by the C allele of the rs4836732 SNP ($p = 6.11 \times 10^{-10}$), on chromosome 9q22.32 [92]. The SNP was associated with female hip OA in a European patient cohort, with an odds ratio (OR) of 1.20.

In order to assess this signal and identify the SNP most likely to cause the OA risk association, it was necessary to identify SNPs in high LD (defined as $r^2 \geq 0.7$) with rs4836732. This chapter then sought to functionally analyse candidate SNPs by investigating their capacity to modulate enhancer or repressor activity in relevant cell lines and by interrogating differential *trans*-acting factor binding to the SNP alleles.

There are no non-synonymous polymorphisms which associate with rs4836732 and therefore, I hypothesise that OA risk is likely to be modulated by regulation of gene transcription. Identification of candidate genes in the association region that may be regulated by the potential *cis*-acting eQTL was consequently required to functionally dissect the association signal. Gene expression of candidate genes was previously quantified by Dr Emma Raine (Prof. Loughlin's group, Institute of Cellular Medicine, Newcastle University), which confirmed their expression in OA cartilage. This chapter will characterise the rs4836732 OA association signal by investigating the expression of candidate genes according to rs4836732 genotype.

### 3.2 Aims

The overall aim of this chapter was to characterise the female hip OA association signal marked by rs4836732 and its functionality. In order to achieve this, I sought to:

- Perform database searches to characterise the rs4836732 association region.
- Clone DNA fragments of the two rs4836732 alleles and of other polymorphisms in high LD with this SNP into pGL3-promoter vectors.
- Identify any allele-specific differences in SNP activity in chondrosarcoma, liposarcoma and osteosarcoma cell lines, assessed by luciferase activity.
- Use online databases to identify proteins predicted to differentially bind to the two rs4836732 alleles.
- Assess differential nuclear protein binding to the rs4836732 C/T alleles using EMSA assays in chondrosarcoma, liposarcoma and osteosarcoma cell lines.
- Identify candidate genes in the LD region which may be regulated by the potential *cis*-eQTL marked by rs4836732.
- Characterise the expression profiles of *ASTN2*, *PAPPA* and *TRIM32* in OA hip and knee cartilage, using data previously generated by Dr Emma Raine.
- Investigate whether rs4836732 marks a *cis*-eQTL by analysing allelic expression imbalance of *ASTN2*, *PAPPA* and *TRIM32* transcription in cartilage, fat pad and synovium samples.

## 3.3 Results

### 3.3.1 Identification and initial database searches of candidate SNPs

To identify the SNP within the rs4836732 locus causing the susceptibility signal, the SNAP online search tool was used to identify SNPs in linkage disequilibrium (LD; $r^2 \geq 0.7$) with rs4836732 [129]. The search only identified two other SNPs in high LD with rs4836732, both of which were in intron 17 of *ASTN2* and so, may influence OA risk by acting as eQTLs (Table 3.1; Figure 3.1).

| SNP | Effect allele | Genomic location | $r^2$ relative to sentinel SNP | D' relative to sentinel SNP |
|---|---|---|---|---|
| rs4836732 | C | *ASTN2* intron 18 | Sentinel | Sentinel |
| rs4837613 | C | *ASTN2* intron 17 | 0.807 | 0.929 |
| rs13283416 | G | *ASTN2* intron 17 | 0.712 | 1.0 |

**Table 3.1**
All SNPs in LD $r^2 \geq 0.7$ with rs4836732. The D′ considers the recombination rate only, with complete LD (D′ = 1.00) denoting no recombination between the two loci. $r^2$ however, also considers the allelic frequencies. Perfect LD (D′ = 1.00, $r^2 = 1.00$) therefore denotes two loci with identical allelic frequencies, between which there is no recombination. Both D′ and $r^2$ are measured on a scale of 0.00 to 1.00.

Online database searches, as described in Chapter 2.1, were used to perform primary analyses of the candidate SNPs prior to functional analyses (Figure 3.2). None of the SNPs reside within a DNase I hypersensitivity region, though rs4837613 does lie less than 50bp upstream to a region of hypersensitivity (Figure 3.2 B) [130]. Sensitivity to DNase I digestion is indicative of a region which is in an open

conformation and so, is more likely to be exposed to transcription factor binding that could regulate nearby gene expression. Indeed, several transcription factors are predicted to bind to the rs4837613 region, including USF1 (upstream transcription factor 1) and TCF12 (transcription factor 12) [131]. No transcription factors are known to bind to either rs4836732 or rs13283416 (Figures 3.2 A and 3.2 C). Chromatin state modelling of the regions shows clear differences in functionality dependant on the cell line under investigation, with the green boxes of the track representing an area of active transcription, orange boxes denoting a region with strong enhancer activity, yellow weak enhancer activity and grey boxes indicating regions of heterochromatin [132]. Both rs4836732 and rs4837613 reside in regions of active transcription and in predicted enhancer regions, while rs13283416 is in a region of heterochromatin. Consequently, the database searches indicate rs4836732 and rs4837613 are the most promising candidates, as rs13283416 appears to be in an inactive region.

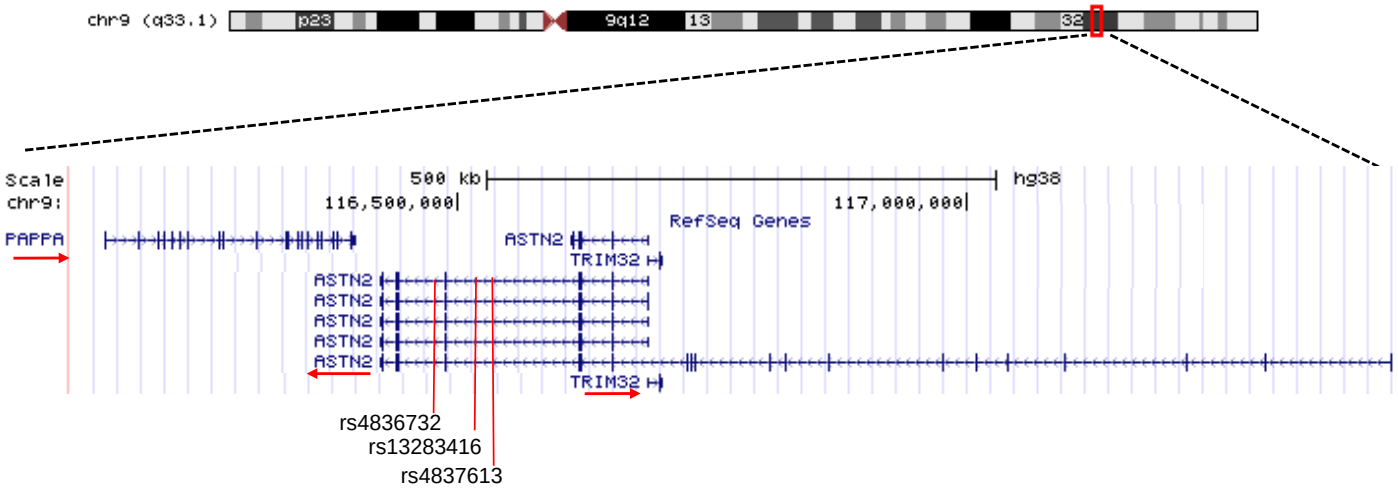



**Figure 3.1**

UCSC Genome Browser screenshot of the 9q33.1 locus and representation of SNP location. The position of each candidate SNP is denoted by a red line. *ASTN2*, *PAPPA* and *TRIM32* all reside in the rs4836732 association region. The top panel represents the p and q arms of chromosome 9 respectively. The first track displays RefSeq genes, with the scale line depicting 200 kb in distance. The multiple transcripts indicate known isoforms of the genes, while vertical blue lines indicate exons and horizontal blue lines indicate the gene introns. The red arrows indicate the direction of gene transcription, with *PAPPA* and *TRIM32* transcribed from the opposite strand to *ASTN2*. Adapted from UCSC genome browser using the hg19 reference genome (http://genome.ucsc.edu/cgi-bin/hgGateway).


**Figure 3.2**

UCSC Genome Browser screenshot of the candidate SNPs rs4836732 (A), rs4837613 (B) and rs13283416 (C). The tracks for DNase I-hypersensitivity (1), where increasing colour gradients of boxes from grey to black indicate increasing hypersensitivity, Transcription Factor ChIP-Seq (2), indicating transcription factors identified to bind to the region, and Chromatin State Segmentation (3), where green boxes indicate an area of active transcription, grey boxes regions of heterochromatin and yellow/orange boxes denoting a region with medium/strong enhancer activity respectively, are shown for each SNP. The images were obtained using the hg19 reference genome.

### 3.3.2 Imputation of novel candidate SNP association with OA

Although rs13283416 and rs4837613 were not analysed in the arcOGEN GWAS [92], imputed data was obtained for the association of their effect allele with OA (data generated by Dr Kalliope Panoutsopoulou, Wellcome Trust Sanger Institute). It was only possible to impute *p* values for the discovery analysis of the arcOGEN study. Based on the *p* values, this data indicates rs4836732 is the most likely causal SNP, though it does not comprehensively eliminate the other two SNPs (Table 3.2).

| **SNP** | **Location** | **Effect allele** | ***p* value (discovery)** | ***p* value (replication)** |
|---|---|---|---|---|
| rs4836732 (sentinel) | *ASTN2* intron 18 | C | $1{\cdot}19\times10^{-05}$ | $6.11\times10^{-10}$ |
| rs4837613 | *ASTN2* intron 17 | C | $3.59x^{-05}$ | |
| rs13283416 | *ASTN2* intron 17 | G | $3.11x10^{-4}$ | |

**Table 3.2**
Imputed discovery *p* values for association of rs4837613 and rs13283416 with female hip OA in the discovery analysis component of the arcOGEN study. The data for rs4836732 is that from the direct genotyping of this SNP and includes the replication analysis *p* value.

### 3.3.3 The investigation of SNP allele functionality using luciferase reporter assays

To investigate the functionality of the two alleles of each of the three candidate SNPs in relation to their regulatory capacity, pGL3 luciferase promoter vector constructs were created. The pGL3 promoter vector contains a minimal SV40 promoter and a multiple cloning site upstream of the luciferase gene, but lacks enhancers and repressors. This enables analysis of the functional effects of a target region on gene expression.

Regions surrounding each of the SNPs were cloned into a pGL3 luciferase promoter vector (Chapter 2.7.1) by ligation into the *MluI* and *Xho*I restriction enzyme sites, using primers listed in Appendix A: Table A.2. Sequencing then confirmed the ligation of the fragments into the plasmid. The plasmids were then transfected, along with *Renilla* as an internal control, into the SW1353 chondrosarcoma cell line, the U2OS osteosarcoma cell line and the SW872 liposarcoma cell line. Luciferase and *Renilla* reporter gene expressions were then quantified after 24 hours, with the empty pGL3 promoter vector used as a control. The luciferase value was divided by the *Renilla* value of the control plasmid to give a value of 1 for normalisation of the data (Figure 3.3).

In general, it appears that all three regions analysed display functional activity in the cell lines analysed. In the SW872 liposarcoma cell line, all three SNPs exhibited enhancer functionality when compared to the empty pGL3 promoter vector, indicated by normalised *y*-axis values >1 (Figure 3.3 A). An allele-specific difference was seen

for rs4837613 in this cell line, with the C allele displaying greater enhancer activity compared to the G allele ($p$ = 0.0023). However, this differential allelic activity was not present in SW1353 (Figure 3.3 B) or U2OS (Figure 3.3 C) cell lines. In these two cell lines, only rs4837613 displayed enhancer activity, while both rs13283416 and rs4836732 acted as repressors, with the exception of the rs13283416 T allele in U2OS cells.

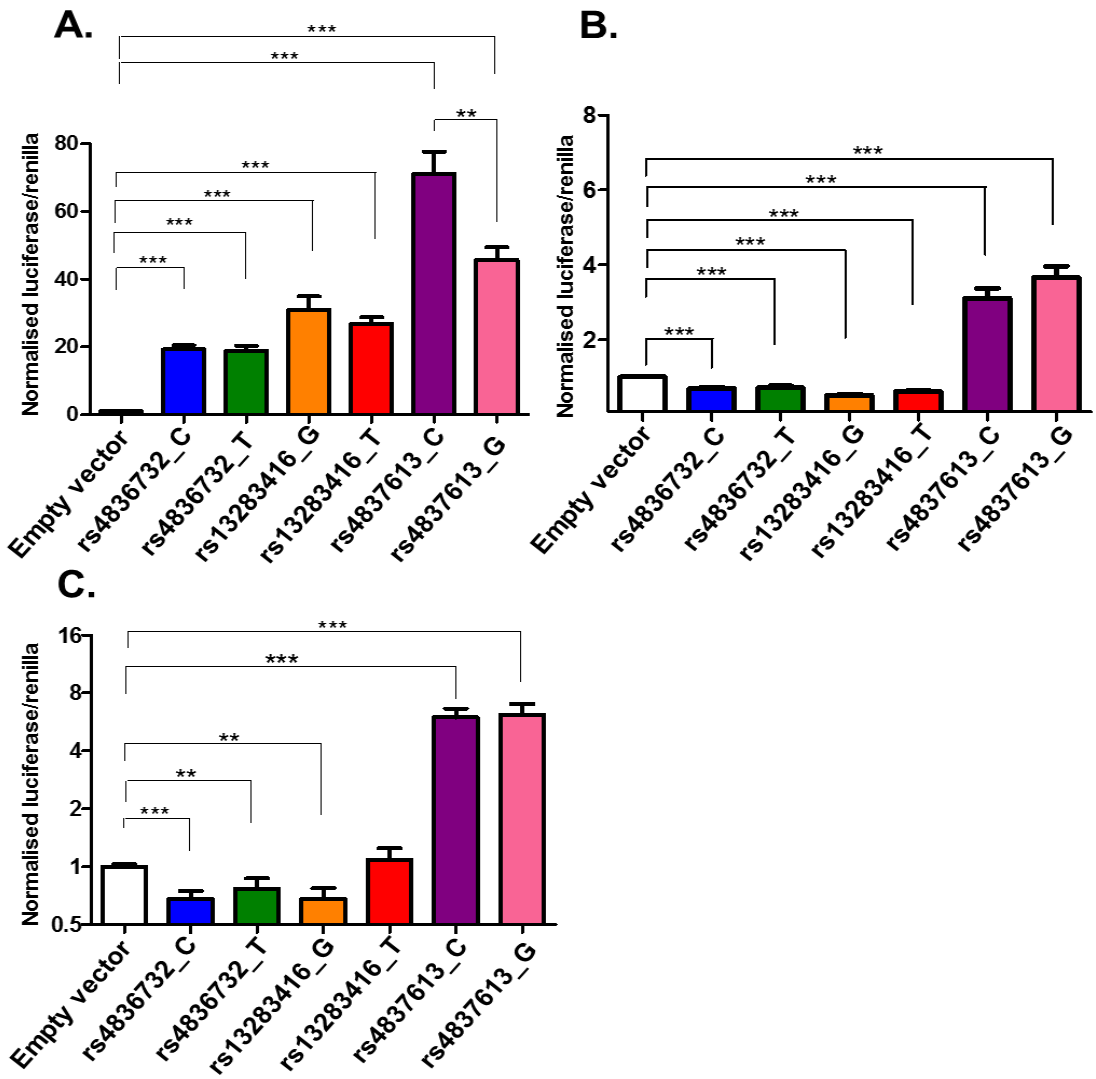



**Figure 3.3**

Luciferase reporter gene expression following transfection with pGL3 promoter vector containing each allele of rs4836732, rs4837613 or rs13283416. The experiment was performed in SW872 cells (A), SW1353 cells (B) and U2OS cells (C). Luciferase activity was quantified 24 hours post transfection and normalised to the *Renilla* internal control. The data was then normalised to the empty pGL3 promoter vector. Eight technical replicates and six biological repeats were performed for each SNP allele in each cell line. Statistical significance was assessed using the Mann-Whitney *U* test. Error bars = SEM. ** $p < 0.01$, *** $p < 0.001$.

### 3.3.4 *In Silico* characterisation of the binding region

Having identified enhancer and repressor functionality of polymorphisms within the association interval, I next aimed to investigate the transcription factors that could be responsible for the functional effects. Though no consistent differential allelic activity was evidenced in the luciferase reporter assays, it is possible that differential transcription factor binding exerts functional effects at specific times in synovial joint development or during the disease process. As such, the next stage of the investigation sought to further characterise the association region by assessing if there are any differences in protein binding to the C and T alleles of rs4836732. Only rs4836732 was taken forward for analysis owing to the fact that this signal had been replicated successfully in the arcOGEN GWAS [92]. Online databases were used to identify proteins predicted to bind to the C and T allele of rs4836732 according to sequence complementarity, as described in Chapter 2.8.1. Table 3.3 shows the factors predicted to bind, their consensus sequences and the database used.

| Transcription Factor | Consensus sequence(s) | Corresponding rs4836732 SNP sequence (C allele) | Corresponding rs4836732 SNP sequence (T allele) | Database |
|---|---|---|---|---|
| GRα | CCTAC/TGAGG | GAGAGACAGCACCTA**C**TTTCTGAGGTCTAAG | GAGAGACAGCACCTA**T**TTTCTGAGGTCTAAG | PROMO 3.0 |
| GRβ | CTATT | | GAGAGACAGCACCTA**T**TTTCTGAGGTCTAAG | PROMO 3.0 |

**Table 3.3**

Consensus sequences of proteins predicted to bind to rs4836732. Shown are the *trans*-acting factors predicted to bind to the rs4836732 sequence, their consensus sequence and the database used for identification. The sequence flanking rs4836732 is displayed in black and the two SNP alleles are underlined.

### 3.3.5 Optimisation of EMSA binding conditions

Having identified predicted transcription factor binding to the rs4836732 locus, I next used EMSAs, a technique which analyses protein binding to specific DNA sequences, to further investigate this in cell lines. The EMSA reaction was performed as described in Chapter 2.8 and the labelled EMSA probes and competitor sequences can be viewed in Appendix A: Table A.3.

Optimisation of EMSA binding conditions was performed using 200 fmol of either the C or T allele labelled probe and 10µg SW872 nuclear protein. This was incubated with binding buffer, DTT/2.5% Tween 20 and Poly (dI:dC) (Figure 3.4). Also added to individual lanes was 50% glycerol, 1% NP40, 1M KCl, 100mM $MgCl_2$ or 200mM EDTA. NP40, $MgCl_2$ and EDTA resulted in increased abundance of protein binding and therefore, further optimisations were performed using combinations of these conditions (Figure 3.5). Of the different combinations of conditions tested, protein complex binding in the presence of NP40 and $MgCl_2$ combined appeared to be optimal. Consequently subsequent EMSAs were performed in the presence of NP40 and $MgCl_2$.

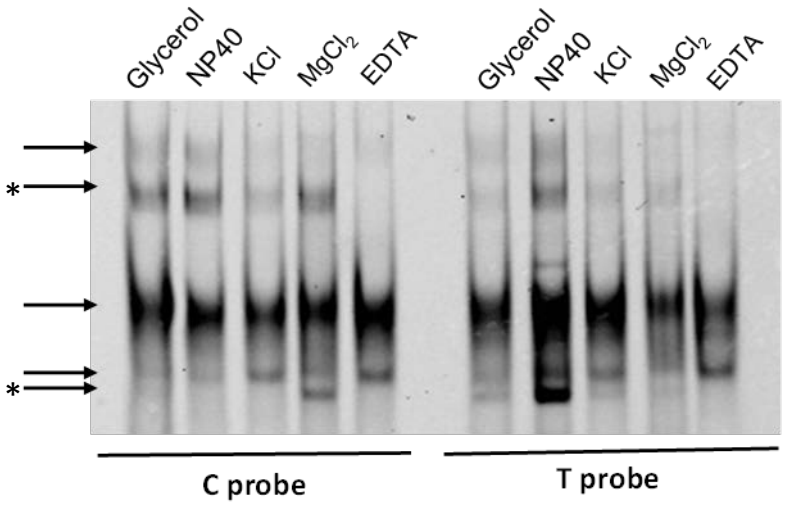



**Figure 3.4**

Optimisation of EMSA binding conditions with SW872 nuclear protein in the presence of rs4836732 C or T labelled probes. Present in each lane is labelled probe, SW872, binding buffer, DTT/2.5% Tween 20 and Poly (dI:dC). The additional additive is indicated at the top of each lane. Arrows indicate SW872 protein complex binding to the probes. * Indicates a complex present in some, but not all conditions.

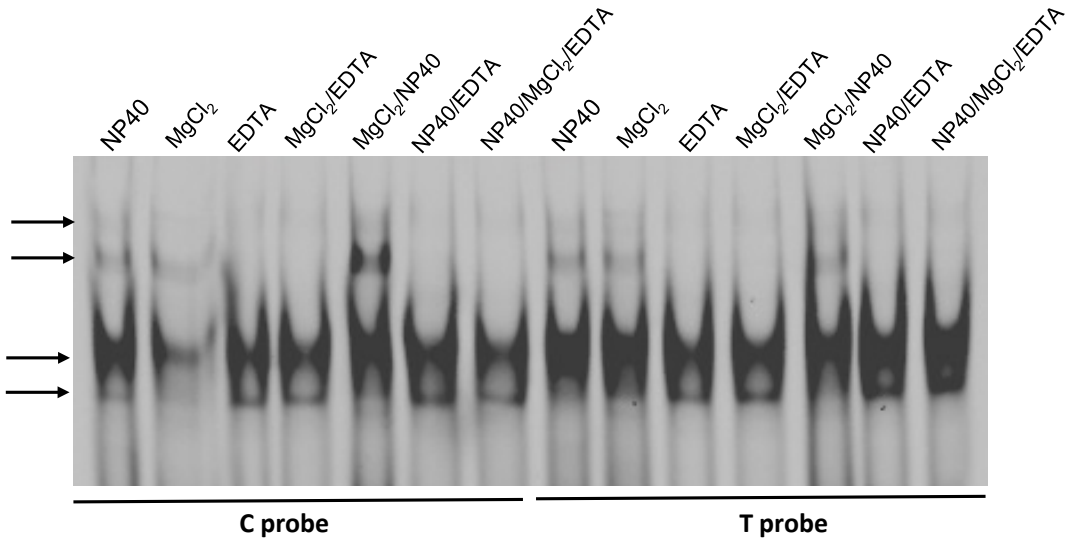



**Figure 3.5**

Optimisation of binding conditions of proteins to the C and T allele EMSA probes in the presence of NP40, $MgCl_2$ and EDTA. Lanes 1-7 contain the C allele labelled probe and lanes 8-14, the T allele probe. The arrows indicate SW872 nuclear protein complexes binding to the probes.

### 3.3.6 Investigation of protein complex binding to rs4836732 using chondrosarcoma, liposarcoma and osteosarcoma cell line nuclear protein

Having optimised the EMSA binding conditions and noted protein complex binding to the C and T probes of rs4836732, the specificity of these complexes was then investigated. This was achieved by introducing increasing concentrations (5-, 10-, 25- or 50- fold excess relative to the labelled probe) of unlabelled competitors with identical sequences to the labelled probes (Figure 3.6).

Having noted specific binding of protein complexes, the region of the labelled probe involved in binding the protein complex was next investigated. Competitor sequences consisting of one region identical to the labelled probe and the remaining sequence entirely random were created to achieve this (Figure 3.7). Complex binding to the C allele probe was outcompeted by competitor 1 in all three cell lines. This therefore suggests it is the 3' region of the C allele probe sequence that is required for protein binding. This was also the case with the T allele probe with U2OS nuclear protein binding. SW872 and SW1353 protein complex binding to the T allele probe was also outcompeted by competitor 1, but not as strongly as compared to U2OS nuclear protein.

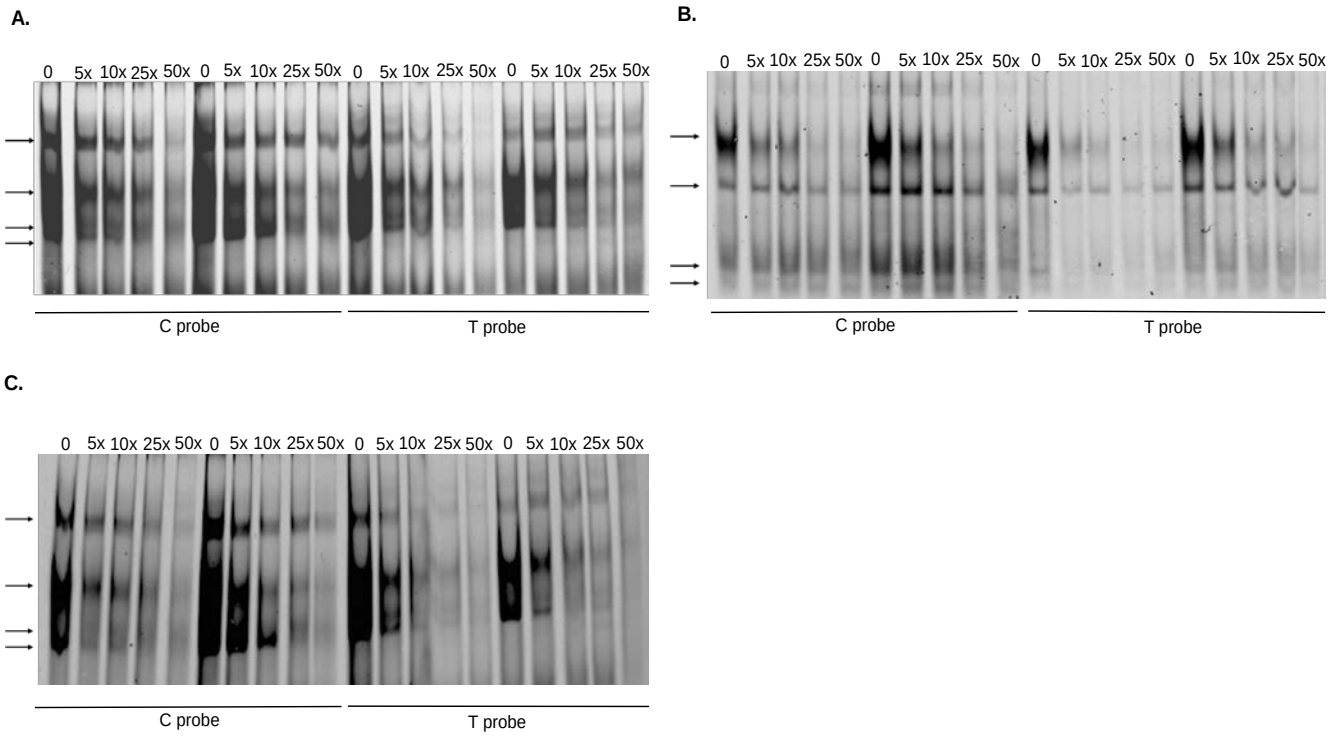



**Figure 3.6**

Competition EMSAs to investigate allele-specific binding of SW872, SW1353 and U2OS nuclear extract to the C and T alleles of rs4836732. Unlabelled competitors were added at 0-, 5-, 10-, 25- and 50x concentrations, relative to the labelled probes. The experiment was performed with SW872 (A), SW1353 (B) and U2OS (C) nuclear protein. Protein: probe complexes which were outcompeted are indicated by the arrows, suggesting specific binding to the probes.

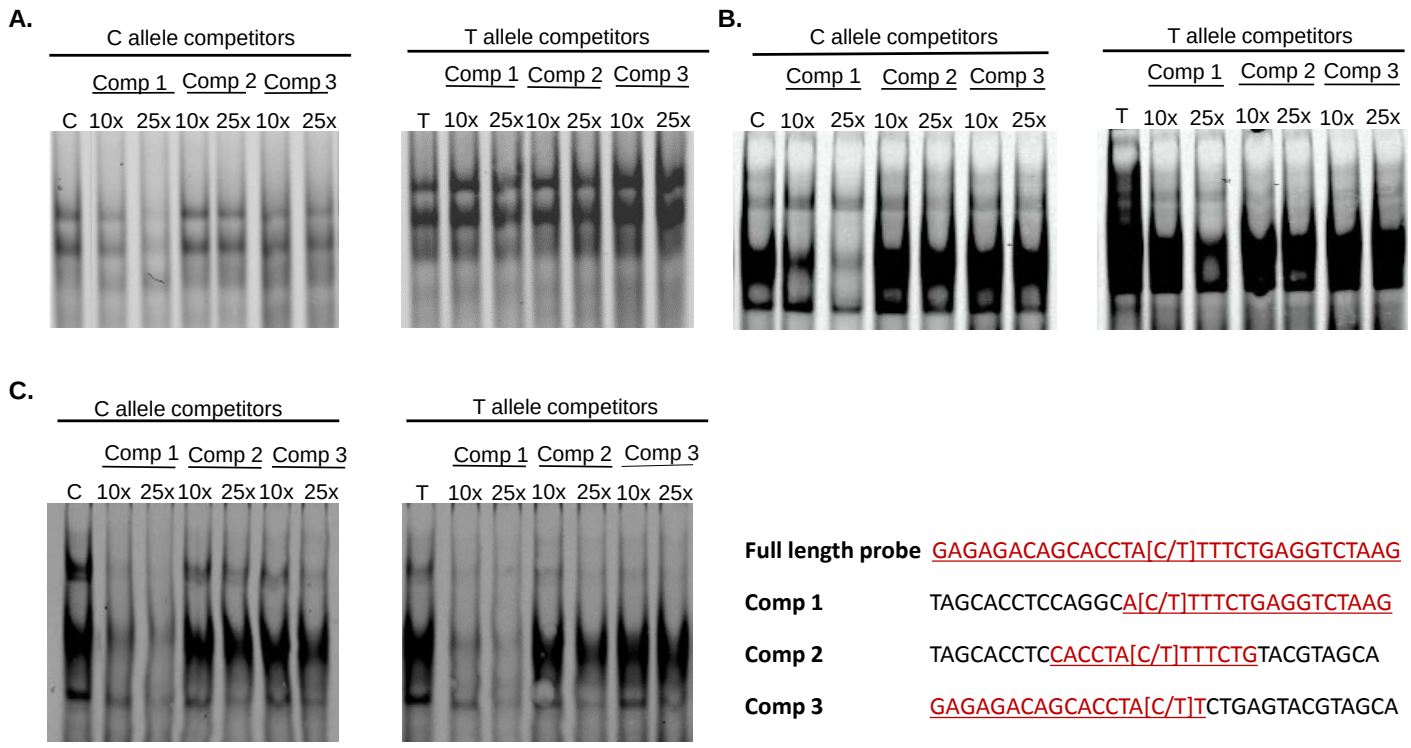



**Figure 3.7**

Competition EMSAs to investigate the regions of the C and T allele probes of rs4836732 binding protein complexes in SW872 (A), SW1353 (B) and U2OS (C) nuclear extract. The protein and labelled probe mixes were incubated with unlabelled competitors containing sections of random sequences (black text) replacing some of the original competitor sequence (underlined in red). Competitor concentrations = 10 or 25 x probe concentration. Each primer was annealed to its reverse complement to create double stranded DNA (dsDNA), prior to use in EMSAs. Full length probe = 31 bp; Comp 1 = 5′ region of the competitor replaced by a 14 bp random sequence; Comp 2 = 9 bp random sequence flanking 13 bp competitor; Comp 3 = 3′ region of the competitor replaced by a 14 bp random sequence.

### 3.3.7 Predicted differentially binding transcription factor competition EMSAs

Having used databases to identify transcription factors predicted to bind to the rs4836732 SNP region (Section 3.3.4), it was necessary to investigate whether these predictions were correct and whether binding is disrupted in the presence of either SNP allele.

Increasing concentrations of consensus sequence competitors for GRα and GRβ were added to the optimised EMSA reaction in the presence of SW872, SW1353 or U2OS nuclear protein (Table 3.4, Figure 3.8). This did not generate evidence for the differential binding of the GRβ consensus sequence to the T allele, as predicted using the PROMO database. No competition is seen in the presence of the GRβ consensus sequence in any of the cell lines with either the C or T allele probe. GRα appears to bind to both the C and T allele probes in the presence of SW872 nuclear protein, with increasing concentrations of the competitor resulting in both probes being out competed (Figure 3.8 A). However, this was not seen in the presence of SW1353 or U2OS nuclear protein.

| Transcription Factor | Competitor sequence (5'-3') |
|---|---|
| GRα | TAGCACCTCCAGGCTGTCT**TGAGG**CGTAGCA |
| GRβ | TAGCACCTCCAG**CTATTT**TGAGTACGTAGCA |

**Table 3.4**

Transcription factors selected for competition EMSAs to investigate binding to rs4836732. After identifying transcription factors predicted to bind the SNP, competitor sequences were designed. These include the protein consensus sequence at the site predicted to bind (predicted by the PROMO database), flanked by a random sequence. Transcription factor consensus sequences are underlined in **bold** and the random sequences are in black text.

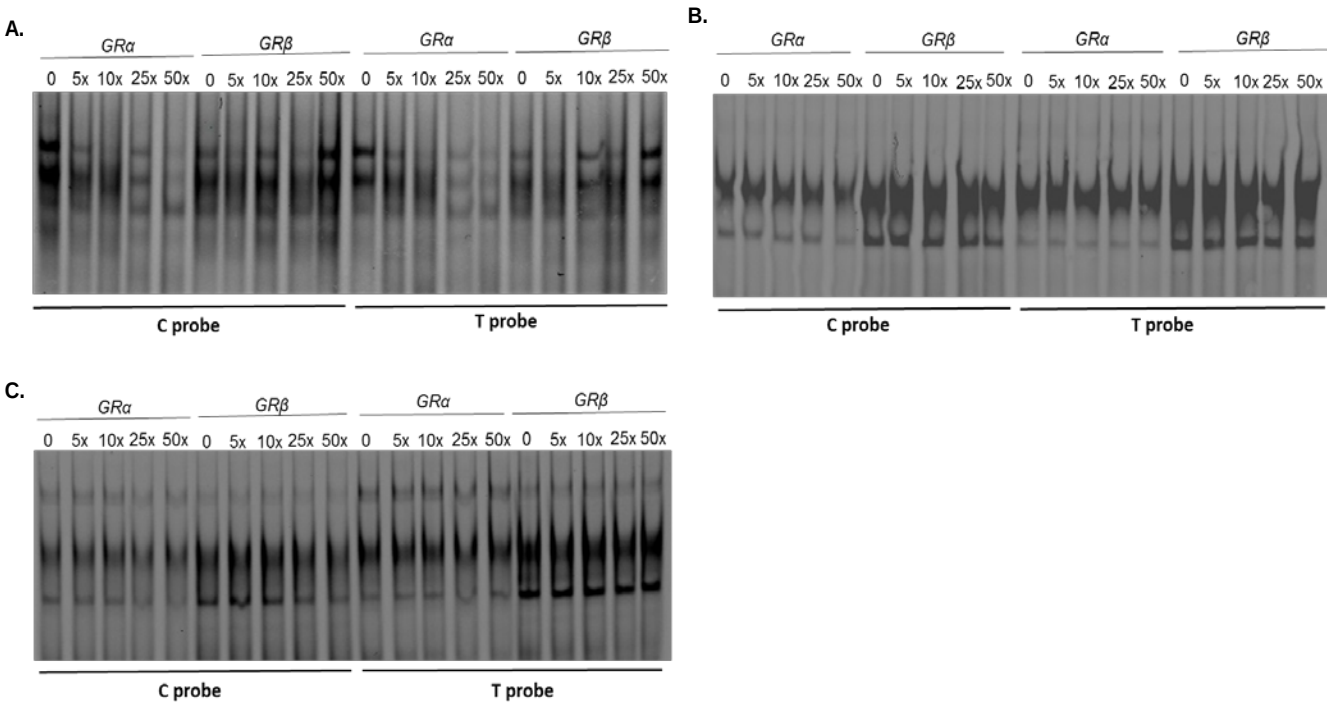



**Figure 3.8**

EMSA competition analysis of GRα and GRβ competition binding to the rs4836732 C and T probes in the presence of SW872 (A), SW1353 (B) and U2OS (C) nuclear protein. The protein:probe mixes were combined with increasing concentrations (0, 5-, 10-, 25- and 50- fold compared to the probe concentration) of competitors containing either the GRα or GRβ consensus sequence. Increasing concentrations of GRα competitor outcompeted C and T probe binding in the presence of SW872 nuclear protein (A). No changes in banding patterns was seen upon addition of the GRβ competitor sequence in any cell line with either the C or T labelled probe.

### 3.3.8 Supershift analysis

Having noted protein complex binding to the rs4836732 probes, this was further investigated by using supershift EMSAs in an attempt to identify the proteins. Upon addition of antibodies raised against specified transcription factors, if an antibody were specific for a transcription factor present, the antibody would bind the *trans*-acting factor:DNA complex and result in the complex migrating more slowly through the gel which would be observed as a 'supershift'. As such, the binding of the two predicted transcription factors, GRα and GRβ, as well as a panel of *trans*-acting factors known to be expressed in cartilage was investigated. For a list of antibodies used, see Appendix C: Table C.1.

No supershifts were observed in any cell line for either of the rs4836732 probes (Figures 3.9, 3.10 and 3.11). This therefore means we are not able to identify the protein complexes binding to the probes. However, the pattern of protein complex binding does differ between the C and T probe in all three cell lines, suggesting that the probes do have differential binding capacity.

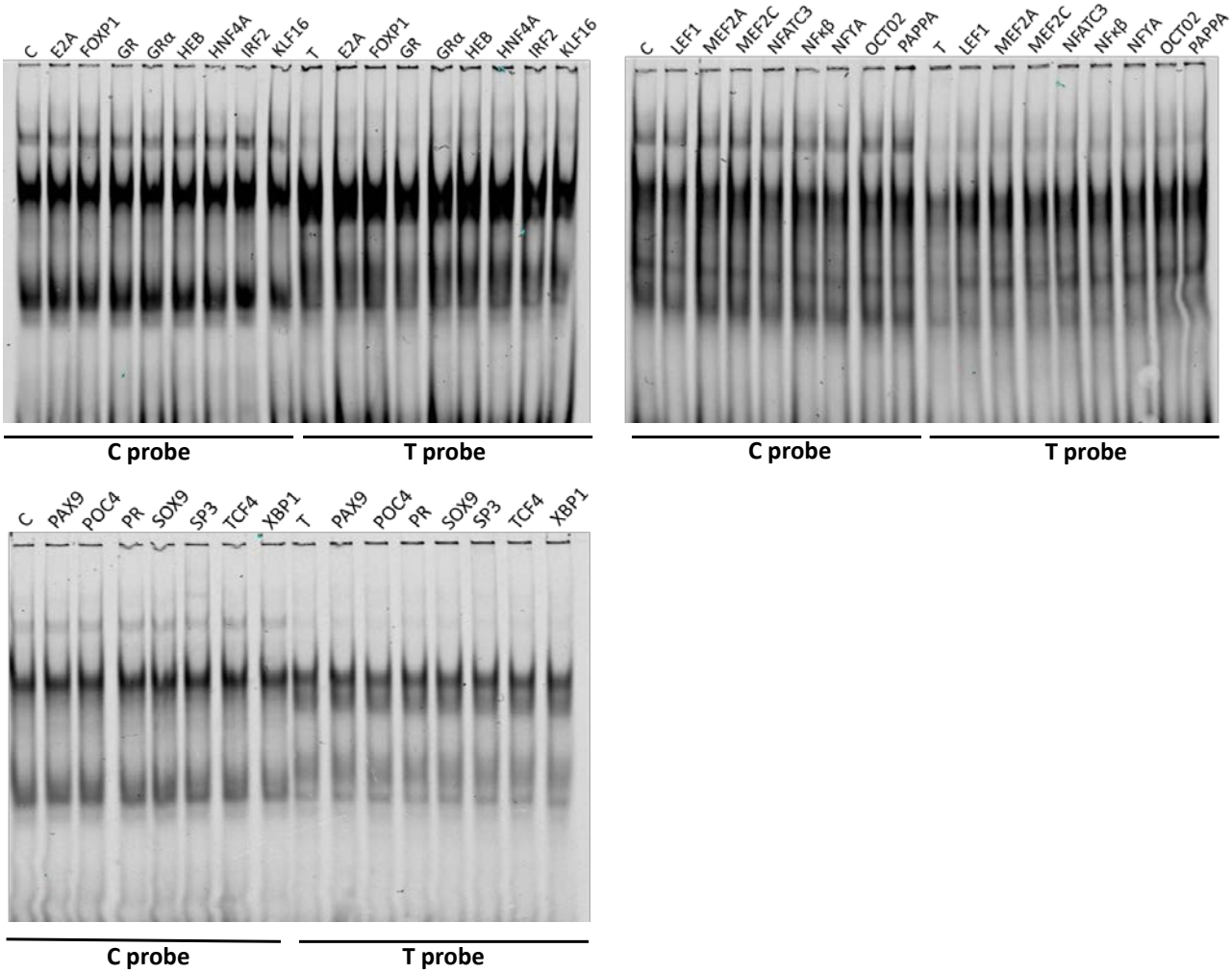



**Figure 3.9**

Supershift EMSAs to investigate the transcription factors of SW872 nuclear extract binding to the C and T alleles of rs4836732. The protein:probe mixes were incubated with 2 µg of antibody, indicated at the top of each well. No changes in the banding patterns were observed after incubation with any of the antibodies. C/T indicates control wells containing just the labelled probe with no antibody.

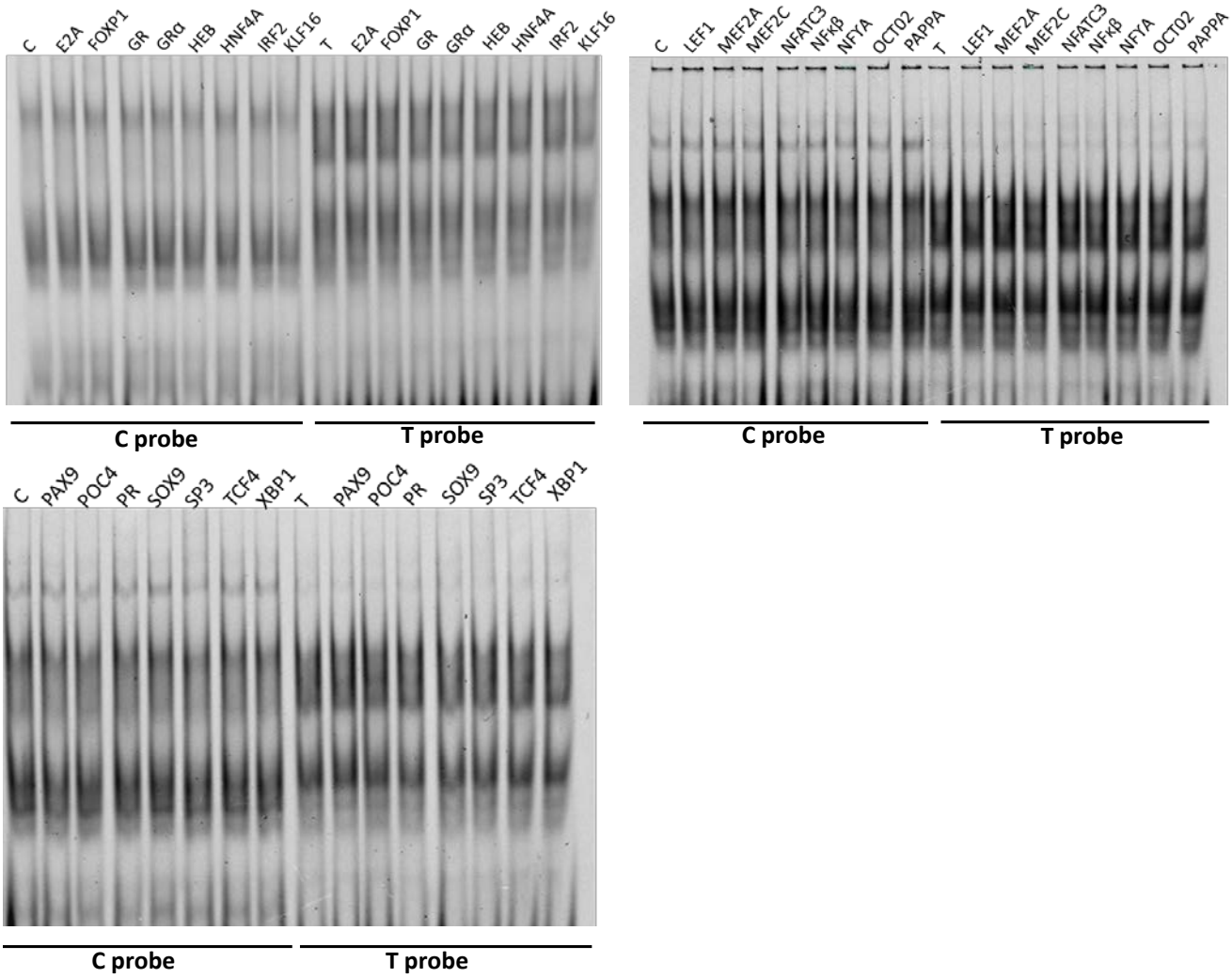



**Figure 3.10**

Supershift EMSAs to investigate the transcription factors of SW1353 nuclear extract binding to the C and T alleles of rs4836732. The protein:probe mixes were incubated with 2 µg of antibody, indicated at the top of each well. No changes in the banding patterns were observed after incubation with any of the antibodies. C/T indicates control wells containing just the labelled probe with no antibody.

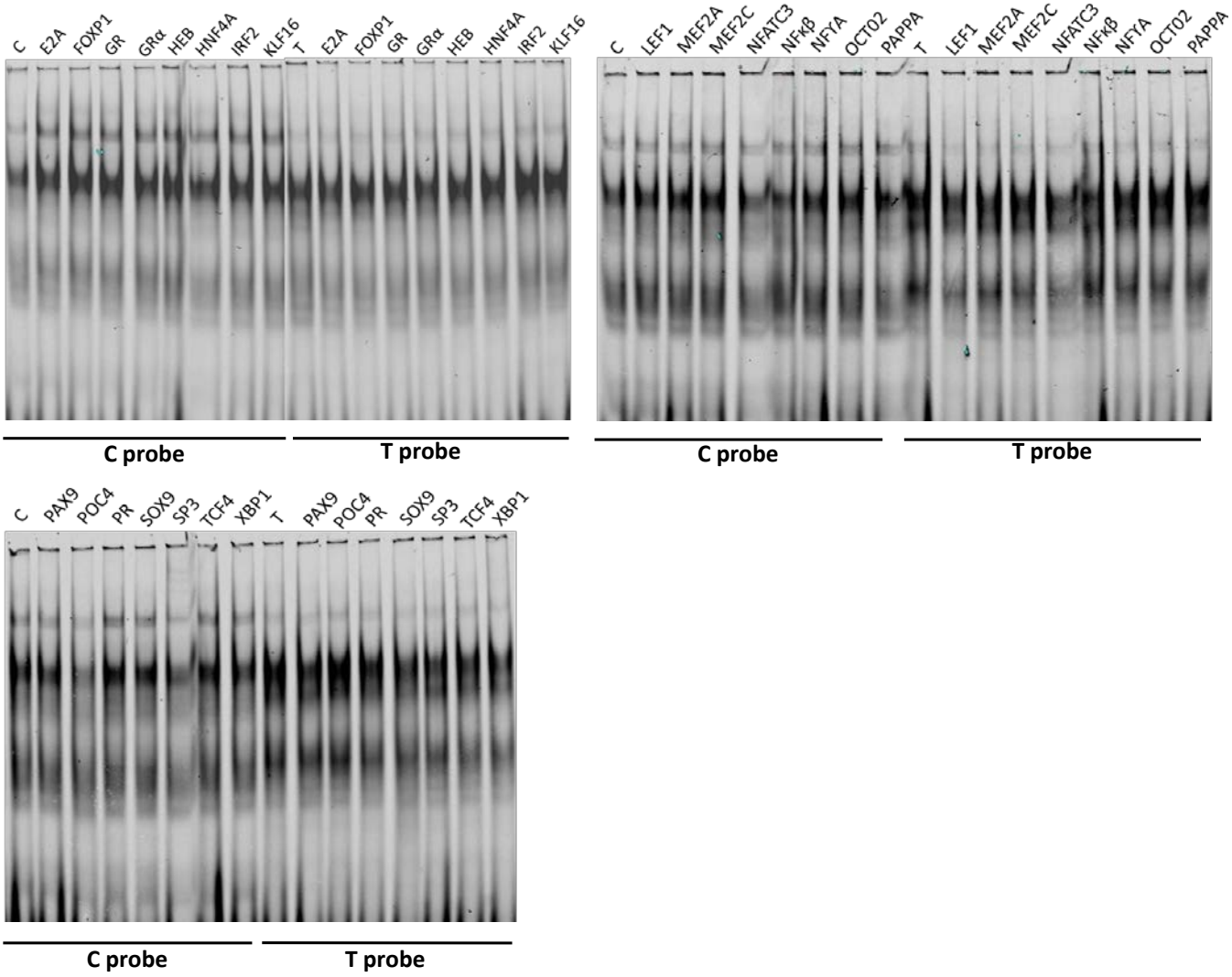



**Figure 3.11**

Supershift EMSAs to investigate the transcription factors of U2OS nuclear extract binding to the C and T alleles of rs4836732. The protein:probe mixes were incubated with 2 µg of antibody, indicated at the top of each well. No changes in the banding patterns were observed after incubation with any of the antibodies. C/T indicates control wells containing just the labelled probe with no antibody.

### 3.3.9 Identification of candidate genes within the LD block

The candidate SNPs of interest in the 9q33.1 susceptibility locus all reside within introns of *ASTN2* (Table 3.1). It is therefore anticipated that the OA susceptibility will be mediated through a *cis*-acting expression quantitative trait locus (*cis*-eQTL), as opposed to the effect of a non-synonymous polymorphism. The UCSC Genome Browser, a database providing information regarding the reference sequence of the human genome, was therefore utilised to confirm the genes which reside closest to the SNPs and so, which may be subject to regulation by the association signal. Figure 3.1 is a screenshot from the website covering the 9q33.1 locus. The first track displays RefSeq genes and confirms rs4836732, rs4837613 and rs13283416 are intronic SNPs, residing within *ASTN2*. *TRIM32* is situated downstream within intron 12 of *ASTN2*, while *PAPPA* is upstream, and both are orientated in the opposite direction to *ASTN2*.

### 3.3.10 Stratification of overall cartilage gene expression by rs4836732 genotype using qPCR

Prior to the start of my PhD, expression data in OA hip and OA knee cartilage was generated for *ASTN2*, *PAPPA* and *TRIM32* by Dr Emma Raine (Institute of Cellular Medicine, Newcastle University) and normalised to the housekeeping genes *18S*, *GAPDH* and *HPRT1* (Chapter 2.13.2). The assay positions are detailed in Figure 3.12 (housekeeping genes) and Figure 3.13 (target genes). All probes and primers were designed to the cDNA sequences, excluding intronic regions.

For each donor, I compiled the genotype at rs4836732, age (in years) at the time of arthroplasty surgery, joint replaced and sex (Appendix B: Table B1). I then analysed the expression profiles of *ASTN2*, *PAPPA* and *TRIM32* in cartilage samples from patients according to genotype at rs4836732 (Table 3.5 – Table 3.7; Figure 3.14). The rs4836732 association signal was discovered when the arcOGEN data was stratified by joint and by sex and was highly significant only in the female hip-specific stratum. This does not though imply that the functional effect of the association signal will only be active in female-hip tissues; the functional effect could be much broader but it is only a risk factor for OA development in the hips of females. An excellent example of this phenomenon is the association of SNPs at the *ALDH1A2* gene with OA [78]. These SNPs are associated with hand OA but not with hip or knee OA and yet they correlate with reduced expression of *ALDH1A2* in cartilage from hip and knee OA patients. Based on such observations, I therefore included analyses of *ASTN2*, *PAPPA* and *TRIM32* expression in knee OA samples, and of neck of femur (NOF) fracture donors, though only *TRIM32* expression data was available from these samples. Analysis of the expression data according to genotype

was assessed using Mann-Whitney U tests and Kruskal Wallis one way ANOVA tests.

There was no statistically significant difference in *ASTN2, PAPPA*, or *TRIM32* gene expression in OA cartilage when stratified by rs4836732 genotype, though the data did confirm that all genes are expressed in cartilage (Tables 3.5, 3.6 and 3.7; Figure 3.14). Comparison of *TRIM32* expression in OA compared to NOF cartilage also did not yield any significant differences. Analysis according to the OA risk stratum (female hip) in comparison to NOF females also did not indicate any differential *TRIM32* expression.

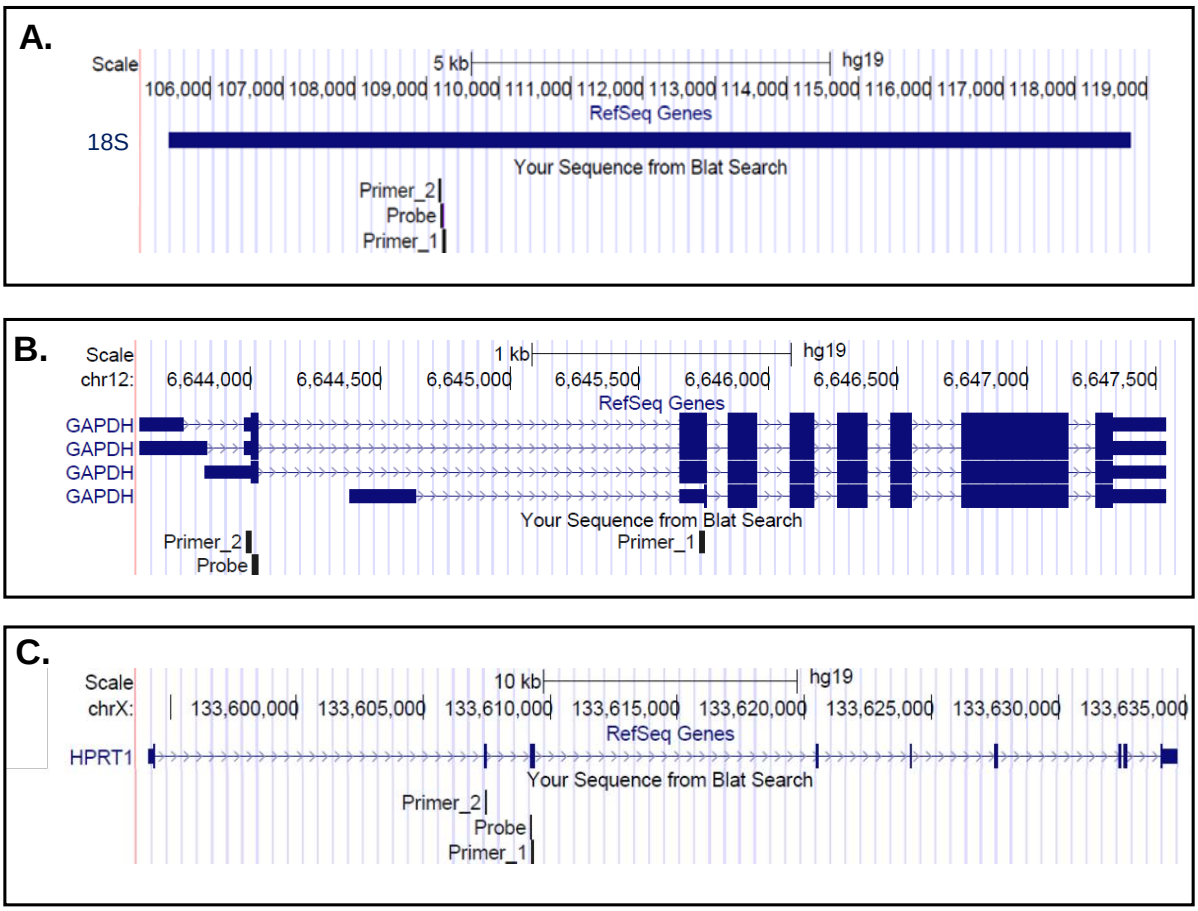



**Figure 3.12**

Location of qPCR housekeeping primers and probes used for quantitative real-time PCR. Assays were designed to the exons of A) *18S*, B) *GAPDH* and C) *HPRT1.* Indicated on each image is the genomic location and scale bar. The images were produced using the hg19 reference genome in the UCSC Genome Browser.

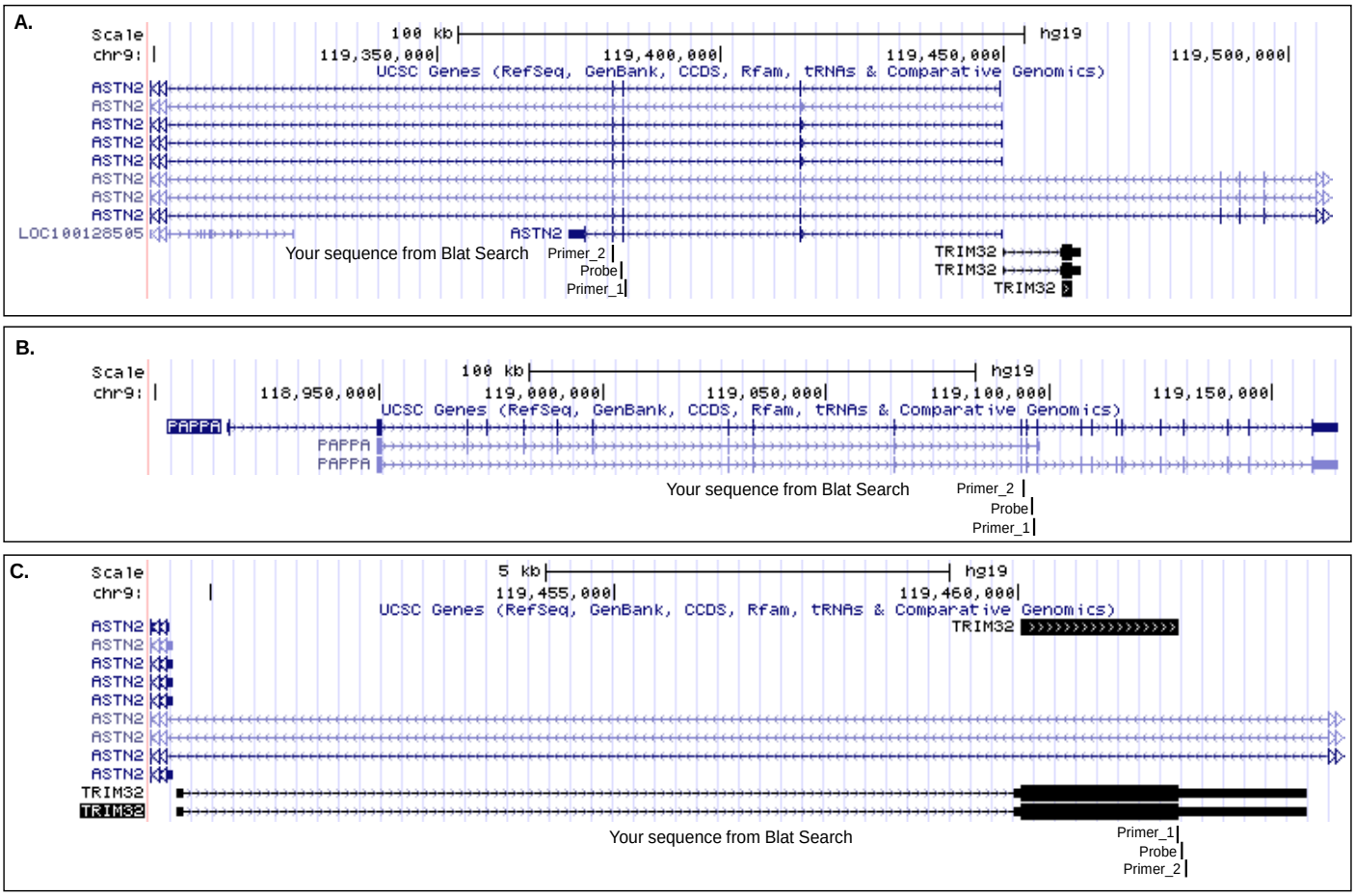



**Figure 3.13**

Location of qPCR primers and probes used for quantitative gene expression analysis of candidate genes. PrimeTime Quantitative Real-Time-PCR (qPCR) assays were purchased, designed for the exons of *ASTN2* (A), *PAPPA* (B) and *TRIM32* (C) (IDT, Iowa, USA). Assays cover each transcript variant of all three genes. Images were generated using the UCSC Genome Browser and hg19 reference genome.

| **Stratification for *ASTN2* qPCR data** | ***p*-value** |
|---|---|
| OA CT (*n* = 22) vs. OA CC (*n* = 5) vs. OA TT (*n* = 14) | 0.063 |
| OA C carriers (*n* = 14) vs. OA TT (*n* = 27) | 0.862 |
| OA hip (*n* = 13) vs. OA knee (*n* = 28) | 0.529 |
| OA females (*n* = 23) vs. OA males (*n* = 18) | 0.572 |
| OA female hip (*n* = 11) vs. OA female knee (*n* = 12) | 0.601 |

**Table 3.5**

Analysis of *ASTN2* expression in OA hip and OA knee cartilage. Data was stratified by rs4836732 genotype or by affected joint. There were no significant differences in gene expression for any of the stratifications, though individuals with the C risk allele at rs4836732 did show a trend toward increased *ASTN2* expression. Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons and a Kruskal Wallis one-way analysis of variance for a comparison of more than two groups. *n* = number of patients in the comparison.

| **Stratification for *PAPPA* qPCR data** | ***p*-value** |
|---|---|
| OA CT (*n* = 17) vs. OA CC (*n* = 3) vs. OA TT (*n* = 12) | 0.969 |
| OA C carriers (*n* = 20) vs. OA TT (*n* = 12) | 0.546 |
| OA hip (*n* = 10) vs. OA knee (*n* = 22) | 0.823 |
| OA females (*n* = 17) vs. OA males (*n* = 15) | 0.104 |
| OA female hip (*n* = 8) vs. OA female knee (*n* = 9) | 0.321 |

**Table 3.6**

Analysis of *PAPPA* expression in OA hip and OA knee cartilage. Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons or a Kruskal Wallis one-way analysis of variance for a comparison of more than two groups. *n* = the number of individuals in the comparison group.

| **Stratification for *TRIM32* qPCR data** | **p-value** |
|---|---|
| OA CT (*n* = 43) vs. OA CC (*n* = 20) vs. OA TT (*n* = 20) | 0.636 |
| OA C carriers (*n* = 63) vs. OA TT (*n* = 20 ) | 0.954 |
| All OA (*n* = 83) vs. NOF (*n* = 15) | 0.375 |
| OA hip (*n* = 23) vs. NOF (*n* = 15) | 0.300 |
| OA female hip (*n* = 17) vs. NOF female (*n* = 15) | 0.082 |
| OA hip (*n* = 23) vs. OA knee (*n* = 60) | 0.407 |
| OA females (*n* = 46) vs. OA males (*n* = 37) | 0.555 |
| OA female hip (*n* = 17) vs. OA female knee (*n* = 29) | 0.228 |
| OA male hip (*n* = 6) vs. OA male knee (*n* = 31) | 0.789 |

**Table 3.7**

Analysis of *TRIM32* expression in OA hip, OA knee and NOF cartilage. There were no significant differences in gene expression when stratified according to affected joint, disease status or rs4836732 genotype. Statistical significance was assessed using a Mann-Whitney *U* test for two-way comparisons or a Kruskal Wallis one-way analysis of variance for a comparison of more than two groups. *n* = the number of individuals in the comparison group.

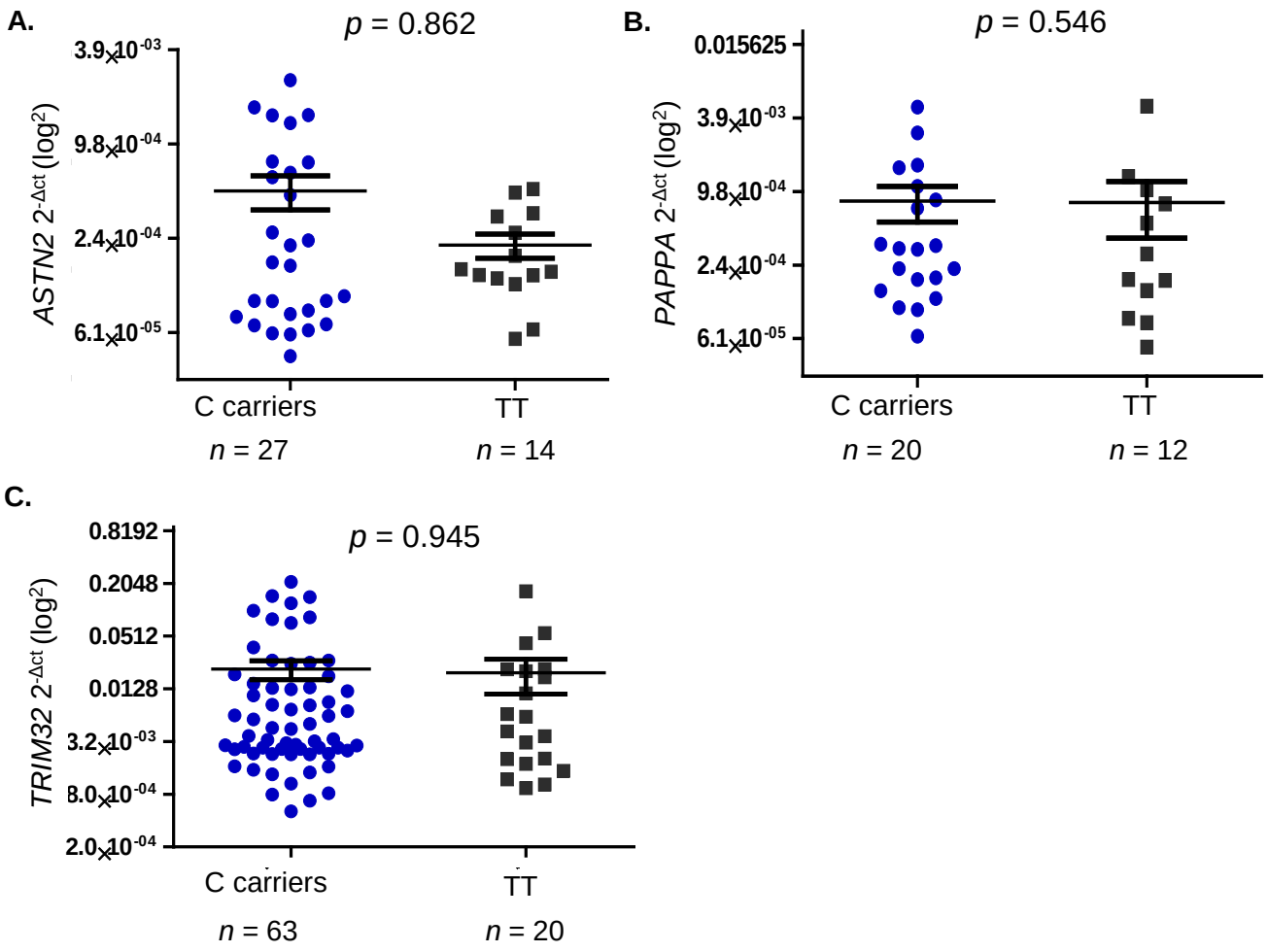



**Figure 3.14**

Analysis of candidate gene expression in OA hip and OA knee cartilage. There were no significant differences in gene expression of *ASTN2* (A), *PAPPA* (B), or *TRIM32* (C) between carriers of the rs4836732 C risk allele and TT homozygotes. 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.

### 3.3.11 Identification of gene transcript SNPs for AEI

In the previous section of this chapter, I did not observe any evidence for rs4836732 marking a *cis*-eQTL in this locus within OA cartilage. However, it is possible that this may be the result of inter-individual variability concealing the effect of the SNP. As such, I therefore deemed it prudent to utilise a more specific line of investigation and analysed gene expression at the allelic level. To achieve this, I used pyrosequencing to quantify the allelic output of the gene transcripts. Given that rs4836732 is intronic, transcript SNPs present in the European population within the candidate genes were identified using the UCSC Genome Browser to act as discriminators of messenger RNA (mRNA) output at the allelic level [124]. A SNAP Pairwise LD search was performed to assess the correlation between transcript SNPs and rs4836732 [129], which failed to identify any transcript SNP to be in high LD with the sentinel SNP (*data not shown*).

As AEI analysis requires transcript SNP heterozygotes to investigate differential allelic outputs, heterozygote frequencies for transcript SNPs were obtained from the dbSNP online database [133]. Consequently, transcript SNPs were selected based on the heterozygote frequency in order to maximise the AEI sample size. Another key consideration was to select transcript SNPs which were present in as many transcript isoforms of a gene as possible. In light of this, after transcript SNPs had been identified, some were excluded as the heterozygote frequency was low (< 20%) and others were rejected owing to the fact they were not present in all transcripts. The selected transcript SNPs are all

located in the 3'UTR of their respective genes (Table 3.8). For *ASTN2* rs7518 (A/G) was selected, for *PAPPA* rs2565 (G/T) and for *TRIM32* rs2281627 (C/T).

| Gene | Transcript SNP | $r^2$ | D' | Heterozygote frequency (%) |
|---|---|---|---|---|
| *ASTN2* | rs7518 | 0.31 | 0.917 | 35 |
| *PAPPA* | rs2565 | 0.17 | 1 | 40 |
| *TRIM32* | rs2281627 | 0.395 | 0.884 | 43 |

**Table 3.8**
Transcript SNPs within the candidate genes which were selected for use in AEI analysis. The $r^2$ and D' were calculated relative to rs4836732.

### 3.3.12 Design and validation of AEI assays

AEI was performed using pyrosequencing, a real-time sequencing-by-synthesis technique (Chapter 2.10.2). This method uses an additional sequencing primer, which extends over the SNP in the direction of the biotinylated primer. The first step for AEI was therefore to create and then validate an assay for each transcript SNP, to ensure the sequencing primers are able to provide an accurate measurement of both alleles relative to each other. Validation was achieved by mixing DNA homozygous at the SNP to create specified ratios of the major and minor allele of each SNP. Subsequent PCR and pyrosequencing was performed and the experimental values were compared to the expected values. The three assays were successfully validated, providing evidence that the two alleles of each SNP could be reliably distinguished (Figure 3.15). Primer sequences can be found in Appendix A: Table A.1.

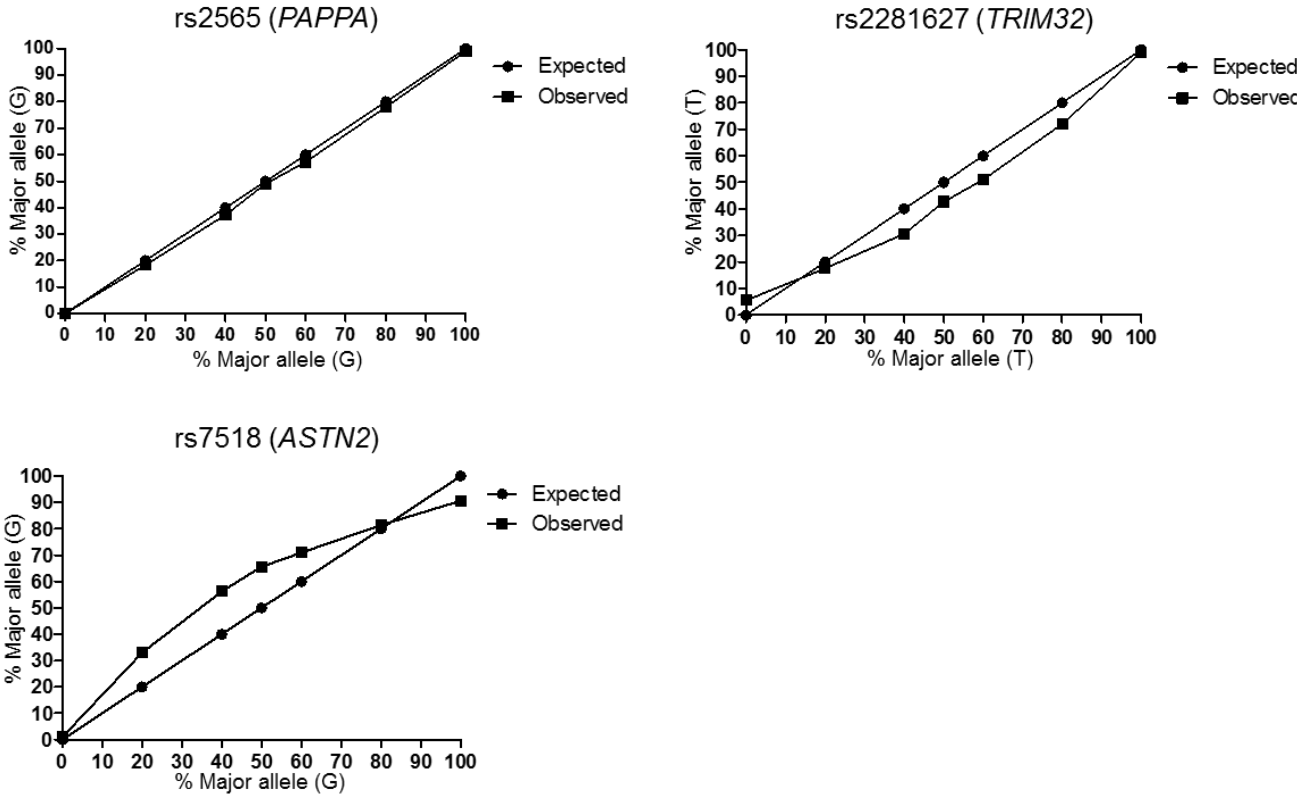



**Figure 3.15**

Validation of transcript SNP pyrosequencing assays in the three candidate genes. In each reaction 50ng of patient DNA was used which was homozygous at the specified transcript SNP. Different ratios of major allele and minor allele homozygous DNA were created and measured by pyrosequencing.

### 3.3.13 Statistical analysis of AEI in synovium, cartilage and fat pad

Genomic DNA and RNA were extracted from the fat pad, cartilage and synovium of OA patients. All donors were genotyped at the sentinel SNP, rs4836732, using an RFLP assay (Chapter 2.5.2) and at the three transcript SNPs using pyrosequencing assays (Chapter 2.6). For patients who were heterozygous at the transcript SNP, RNA was reverse transcribed to cDNA. Allelic expression imbalance (AEI) analysis was performed by pyrosequencing and the allelic ratio of cDNA was normalised to genomic DNA.

Stratification by rs4836732 genotype enabled the investigation of any correlation between AEI and the sentinel SNP. In the event of AEI being present, clustering of compound heterozygotes would occur. This would create two groups, one with an allelic ratio of greater than 1:1 and the other less than 1:1. The transcript SNPs are not in high LD with the sentinel SNP, so recombination events mean it would not be possible to determine the causal allele if AEI were detected.

No evidence was identified which indicates rs4836732 correlates with a *cis*-eQTL acting on *ASTN2*, *PAPPA* or *TRIM32* in the three tissue types analysed. In all tissues, rs4836732 heterozygotes display clustering around the 1:1 expression ratio for the analysed genes (Figure 3.16).

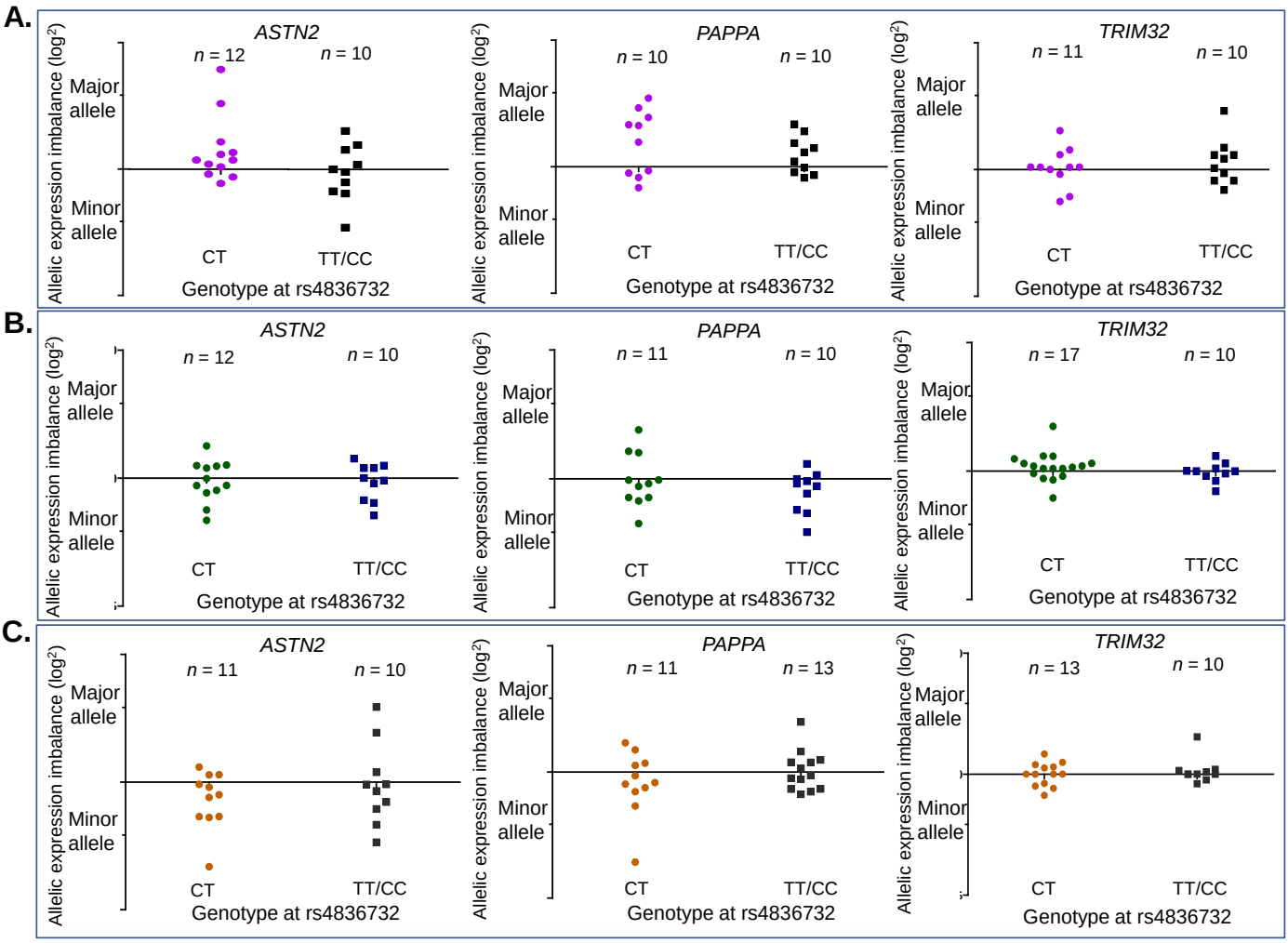



**Figure 3.16**

Allelic expression imbalance of *ASTN2*, *PAPPA* and *TRIM32* transcript SNPs stratified by rs4836732 genotype in cartilage (A), fat pad (B) and synovium (C) from OA patients. No AEI was observed to correlate with rs4836732 genotype. Statistical significance was assessed using a Mann-Whitney *U* test, though none of the comparisons proved statistically significant. *n* indicates the number of patients in each comparison group.

### 3.4 Discussion

In Chapter 3, I have attempted to characterise the 9q33.1 OA susceptibility region marked by rs4836732. GWAS association signals often mark disease risk influenced by another polymorphism, which acts to modulate gene transcription. Consequently, I hypothesised that rs4836732 marks a functional SNP which regulates expression of a gene residing in the association locus.

I began by performing database searches to identify SNPs in high LD with the association signal, which identified two further candidate SNPs. To dissect any functional allelic differences between SNP alleles, I created luciferase reporter constructs [134]. This confirmed all three SNPs to have enhancer or repressor functions in the cell lines analysed, and identified differential SNP activity for rs4837613 in SW872 cells, but did not identify consistent differences in allelic activity for any of the three SNPs. Imputed *p* values for association with the rs4836732 association signal were subsequently generated by Dr Kalliope Panoutsopoulou (Wellcome Trust Sanger Institute) for rs13283416 and rs4837613, as neither were present on the arcOGEN array [92]. Given that both of these were less significant than for rs4836732, and the fact that the association for rs4836732 had already been robustly replicated, this SNP appeared to be the most compelling candidate. Furthermore, a true causal polymorphism in high enough LD with rs4836732 to detect the effect on a GWAS should assert a strong enough affect to be detected via rs4836732.

Though differential allelic activity was not seen using the luciferase reporter assays, it is potential that allelic differences may be regulated by differential transcription factor binding, but that the effects of this are only functional at specific times in joint development or repair. EMSAs are a commonly used method to characterise protein:DNA binding and so, were performed using rs4836732 C and T allele probes with nuclear protein from chondrosarcoma, liposarcoma and osteosarcoma cell lines. In particular, binding of two transcription factors, GRα and GRβ, which had been predicted to bind the SNP by the online PROMO database was analysed, with GRβ predicted to bind the T allele only and GRα to both alleles. These EMSAs suggested GRβ did not bind either probe allele, while GRα did appear to outcompete probe:protein binding with increasing concentration in the SW872 cell line in the presence of both the C and T probes. Protein complexes binding the probes were analysed by introduction of antibodies to try and observe supershifts. However, no supershifts were observed with the antibodies investigated. Different banding patterns were observed in all cell lines in the presence of the C allele probe compared to the T allele. This suggests differential protein complex binding dependent on the allele present at rs4836732. However, it has not been possible to identify these proteins. Furthermore, it is important to highlight that despite numerous replications, the EMSAs produced variable results in terms of banding patterns. This is potentially a result of the artificial nature of the technique, in that the probe DNA is linear. As such, it is not an accurate reflection of the natural chromatin conformation and this is likely to affect transcription factor binding. Further experiments using techniques such as ChIP, which offers a native representation of the cellular interactions therefore allowing investigation of transcription factor binding to the endogenous loci rather than the artificial conditions utilised in EMSAs, would

therefore be advisable. Additionally, it would be of interest to investigate nuclear protein from difference sources, such as HACs, as transcription factors may be present in different abundances.

Given its intronic position, if functional, rs4836732 is likely to mediate OA risk by acting as an eQTL. I therefore investigated genes residing within the association region to try and identify whether rs4836732 was acting as a *cis*-eQTL. Overall gene expression was quantified in OA cartilage for *ASTN2*, *PAPPA* and *TRIM32,* but this did not reveal any difference in gene expression according to rs4836732 genotype. Gene expression was subsequently analysed at the allelic level using AEI analysis, a method which allows the quantification of mRNA outputs in relation to the different alleles of an individual who is heterozygous at the SNP of interest. AEI was performed using nucleic acids from synovium, fat pad and cartilage from OA patients. However, the results did not yield any significant differences in gene expression in relation to rs4836732 genotype. This may be indicative of the SNP acting at a different time in joint development or repair which does not occur in end-stage disease tissues. Alternatively, it may be that rs4836732 is regulating the expression of a different gene given that *cis*-eQTLs can act up to megabases in distance [135]. It is also possible that no effect has been seen as this is not the functional polymorphism.

To summarise, in this chapter, I have demonstrated that the three candidate SNPs displayed enhancer or repressor effects in the cell lines analysed, and that the main functional candidate, rs4836732, displayed allelic differences in EMSA analyses.

However, the data indicate that the OA association signal marked by rs4836732 does not assert its effects in end-stage OA tissues. It is possible that disease risk is modulated during joint development and as such, this chapter supports further investigation into the candidate gene which is most likely to have a role in joint formation.

# Chapter 4: Investigation of *PAPPA* as a key candidate gene

## 4.1 Introduction

Although there was no correlation observed between rs4836732 genotype and expression of candidate genes within the association region, this does not mean that the SNP is not functional in OA susceptibility. It is possible that the association signal may exert its effects in other tissue types, during joint development or earlier in the disease process. It is also possible that the rs4836732 signal modulates gene expression outside of the LD region.

A large body of evidence supports abnormal synovial joint development as a key predisposing factor for OA risk. Indeed, a cross sectional study identified hip dysplasia (HD) as the only individual risk factor to associate with hip OA in males upon analysis of the influence of hip dysplasia, body mass index, smoking and occupational exposure to repeated daily lifting [136]. Analysis into the association between OA susceptibility loci and variation in proximal femur shape, represented by statistical shape model (SSM) modes, in patients with unilateral hip OA identified association between rs4836732 and SSM mode 5 ($p = 0.0016$) [137]. Mode 5, a female SSM, captures hip shape variation of the superior aspect of the femoral head. As such, this suggests rs4836732 may contribute to OA susceptibility by altering the shape of the femoral head, and so promotes further investigation into the association signal in relation to joint development.

Of the three candidate genes investigated in Chapter 3, *PAPPA* is arguably the most compelling candidate in relation to synovial joint development. PAPPA is the IGF-II-dependent IGF-binding protein (IGFBP)-4 protease, with IGFBP2, 4 and 5 being the PAPPA-cleavable IGFBPs. Murine studies have demonstrated that osteoblast-specific *PAPPA* expression causes increased bone size, indicative of PAPPA-cleavable IGFBPs exerting a regulatory effect over IGF bioavailability in skeletal tissues [106].

Furthermore, in a global mouse PAPPA knock-out, a significant decrease in trabecular volume was recorded, progressing with age, alongside an age-associated decrease in fracture resistance [107]. Measurement of dynamic serum parameters associated with bone formation indicated reduced bone formation in young knockout mice (4 and 6 months) accompanied by a 45-60% reduction in osteoblasts and 50-60% reduction in osteoclasts. As such, *PAPPA* appears to be necessary for both embryonic and postnatal bone development and is a promising candidate for regulation by the rs4836732 signal during joint development.

A genome wide expression study comparing the gene expression profile of paired lesioned and preserved OA cartilage identified increased PAPPA expression in lesioned tissue, particularly in hip OA which demonstrated a 2.6-fold increase in the lesioned cartilage in the 33 patients analysed [138]. As such, this suggests there is aberrant *PAPPA* expression in OA-affected cartilage during end-stage disease and may be indicative of a role in joint repair.

Based on the above evidence, I hypothesise that the rs4836732 signal influences risk of hip OA in females by regulating *PAPPA* expression during joint development and repair. I consequently sought to characterise *PAPPA* expression in hMSC differentiation into cells found in the synovial joint and to further investigate the association between rs4836732 genotype and *PAPPA* expression.

## 4.2 Aims

The aim of this chapter was to explore the potential that rs4836732 confers OA risk through a role in synovial joint development owing to aberrant *PAPPA* regulation. This was achieved by:

- Analysis of *PAPPA* expression during hMSC differentiation into chondrocytes and osteoblasts.
- Performing *in vitro PAPPA* siRNA knockdowns in hMSCs, followed by chondrogenic and osteogenic differentiation.
- Characterisation of hMSC differentiation capacity following *PAPPA* knock-down by qPCR and histological staining analysis.
- Quantification of bone-specific *PAPPA* expression in OA patient samples and stratification by rs4836732 genotype to assess whether differential expression correlates with the presence of the risk allele.

## 4.3 Results

### 4.3.1 Examination of the expression profile of *PAPPA* during hMSC differentiation

In order to examine the developmental role of *PAPPA*, I first sought to investigate expression throughout the differentiation of hMSCs into osteoblasts and chondrocytes. Expression during differentiation was essential to support the hypothesis that the functional effects of rs4836732 could be exerted at early stages of joint development, thus resulting in altered joint structure and increased OA risk.

I tracked overall gene expression throughout chondrogenesis and osteogenesis, irrespective of rs4836732 genotype, in both OA and non-OA patients (Table 4.1). For details of hMSC isolation and expansion see Chapter 2.11.2. The purpose of this was to clarify whether *PAPPA* had the potential to contribute to joint development, reflected by expression during hMSC differentiation. Dr Madhushika Ratnayake and Maria Tselepi of Prof. Loughlin's group (Institute of Cellular Medicine, Newcastle University) performed *in vitro* chondrogenesis and I performed *in vitro* osteogenesis (Chapter 2.11.4). I then performed the qPCR and data analysis of *PAPPA* expression (Chapter 2.13.2). qPCR primer and probe sequences can be found in Appendix A: Table A.4, while assay positions are detailed in Figure 3.11 (housekeeping genes) and Figure 3.12 (target genes).

During chondrogenesis, gene expression was quantified on days 3, 7 and 14 as at day 0, MSCs were undifferentiated (Figure 4.1 A). To confirm successful

chondrogenic differentiation, expressions of the chondrogenic markers *ACAN* (aggrecan), *COL2A1* and *SOX9* were confirmed at each time point by Dr Madhushika Ratnayake and Maria Tselepi. *PAPPA* was expressed throughout chondrogenesis, though a distinct pattern of expression associated with specific stages of differentiation was not seen.

During osteogenesis, *PAPPA* expression was quantified on days 0, 3 and 14 of differentiation (Figure 4.1 B). In this instance, cells had been allowed to adhere in osteogenic media for 6 hours prior to harvesting on day 0. Again, *PAPPA* expression was detected at all time points but there was no distinctive pattern of expression.

Together, these data confirm dynamic expression of *PAPPA* in an *in vitro* model of joint development and therefore, support further investigation into the developmental role of this gene and association with rs4836732 genotype.

| Donor | Sex | Joint | Age (years) |
|---|---|---|---|
| 2454E * | Female | Hip | 24 |
| N2980 | Female | Hip | 82 |
| 3503 | Female | Hip | 61 |
| 4147 | Female | Knee | 51 |

**Table 4.1**
Characteristics of donors whose hMSCs were used for chondrogenesis and osteogenesis experiments. The age listed is that at the time of joint replacement, or

when the aspirate was taken. The donor marked with an asterisk (*) was a young, non-OA donor purchased from Lonza, UK.

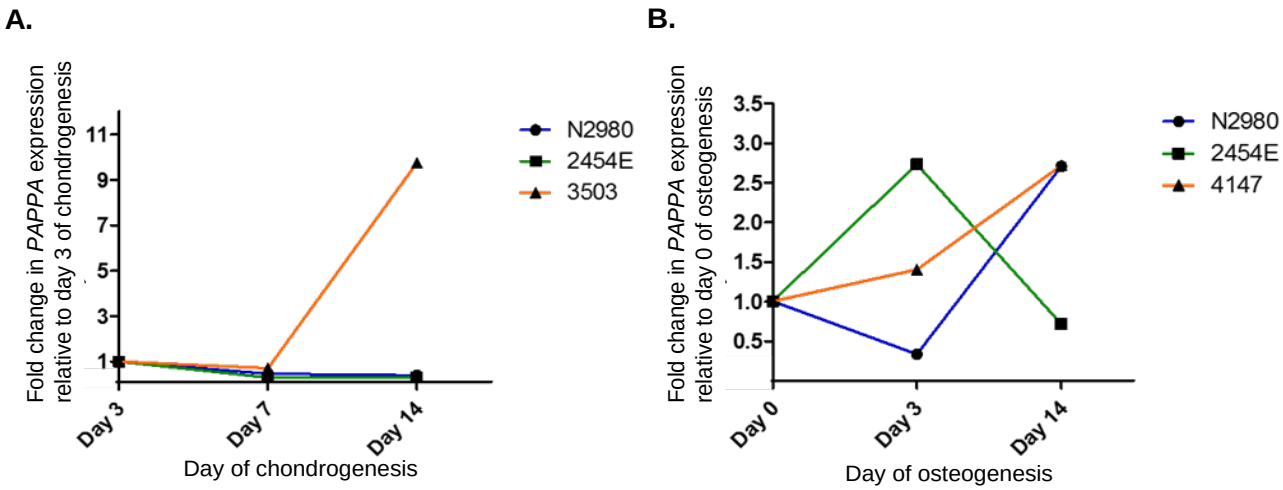



**Figure 4.1**

Expression of *PAPPA* during chondrogenesis and osteogenesis. (A) hMSCs were differentiated into chondrocytes and gene expression was quantified on days 3, 7 and 14. (B) hMSCs were differentiated into osteoblasts and *PAPPA* expression was quantified on days 0, 3 and 14 of differentiation. In differentiation into both lineages, *PAPPA* expression was detected at all time points. The notable inter-individual variation and the dynamic expression patterns support a potential functional role during cartilage development. Blue line = non-OA patient (NOF fracture; patient N2980), green line = healthy control (Lonza, UK; patient 2454E) and orange = two independent OA patients (patients 3503 and 4147).

### 4.3.2 Knockdown of *PAPPA* in human cells

In order to further investigate the role of *PAPPA* in human joint development, I aimed to deplete *PAPPA* in human MSCs prior to differentiation into cells found within the synovial joint. This was achieved using On-TARGETplus SMART pool human *PAPPA* siRNA (Dharmacon, GE Healthcare, UK), targeted to *PAPPA* mRNA, as described in Chapter 2.11.3. Details of the siRNAs used can be found in Appendix D: Table D.1.

The ability of the siRNA to knockdown *PAPPA* was first assessed in the SW1353 cell line. The cells were cultured and then transfected with *PAPPA* siRNA (Dharmacon, GE Healthcare, UK). Following 48 hours, efficiency of the knockdown was assessed at both the gene expression and protein levels. To ensure the transfection reagent was not interfering with PAPPA expression, cells were transfected with non-targeting siRNA (NT siRNA).

*PAPPA* mRNA depletion in SW1353 cells was confirmed by qPCR. A significant decrease in *PAPPA* expression was observed in comparison to both the untransfected cells and the NT siRNA controls ($p < 0.001$) (Figure 4.2 A). No statistically significant difference was seen when comparing the untransfected and NT siRNA treated cells. Depletion of the protein was confirmed by immunoblotting, as described in Chapter 2.12. This confirmed PAPPA protein depletion in *PAPPA* siRNA transfected cells relative to the two control conditions (Figure 4.2 B).

These data confirm PAPPA expression at both the mRNA and protein levels is depleted following transfection with the selected *PAPPA* siRNA. As such, the On-TARGETplus SMART pool human *PAPPA* siRNA (Dharmacon, GE Healthcare, UK) was used for all subsequent *PAPPA* siRNA knockdown experiments.

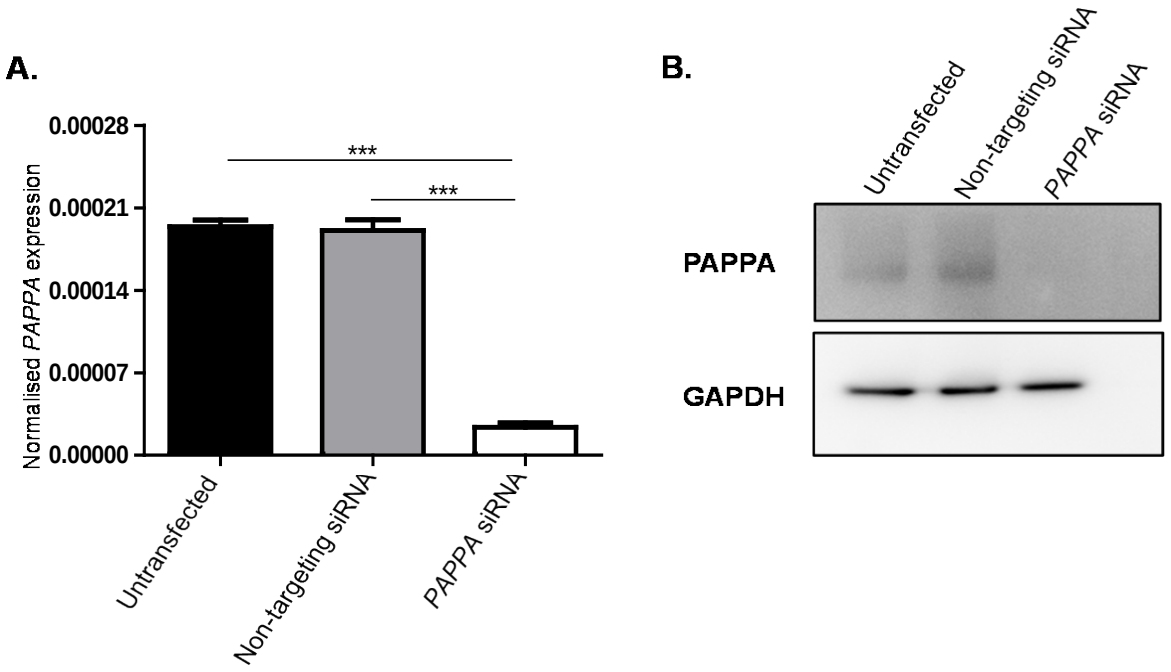



**Figure 4.2**

Knockdown of *PAPPA* in SW1353 cells using On-TARGETplus SMART pool human *PAPPA* siRNA (Dharmacon, GE Healthcare, UK). Cells were expanded to 80% confluency and transfected with 100nM non-targeting siRNA or 100nM siRNA targeting *PAPPA*. For each condition, gene expression was quantified using qPCR and normalised against *18S*, *GAPDH* and *HPRT1* housekeeper expression (A). *PAPPA* expression was significantly downregulated relevant to both untransfected and NT siRNA controls ($p < 0.001$). Immunoblotting was used to confirm protein depletion, with GAPDH used as a loading control (B). This confirmed PAPPA knockdown at the protein level in cells transfected with *PAPPA* siRNA. Three technical repeats were performed for each qPCR reading and two biological repeats were performed, with one representative immunoblot shown. Error bars represent the SEM. Statistical significance was assessed using an unpaired Student's *t* test. *** $p < 0.001$.

### 4.3.3 siRNA Knockdown in MSCs undergoing osteogenic differentiation

Having confirmed the capacity of the *PAPPA* siRNA to reduce *PAPPA* mRNA and protein expression, the next step was to perform knockdowns using hMSCs from healthy donors (Table 4.2). hMSCs from three healthy individuals (purchased from Lonza, UK.) were expanded under standard tissue culture conditions to 80% confluency and then transfected with siRNA targeting *PAPPA* or NT siRNA (Chapter 2.11.3). Untransfected cells were used as a control to confirm successful differentiation.

| Donor | Sex | Age when aspirate was taken (years) |
|---|---|---|
| 2454E | Female | 24 |
| 26601 | Female | 41 |
| 26798 | Male | 23 |

**Table 4.2**
hMSCs used for siRNA knockdown experiments. All were iliac crest bone marrow aspirates from young, healthy, non-OA individuals, purchased from Lonza UK.

*PAPPA* was successfully knocked down in hMSCs relative to expression in the NT siRNA transfection control (Figure 4.3). The knockdown was observed following the 48 hour transfection (day 0 of osteogenic differentiation) and persisted through to day 10 of differentiation. Depletion of PAPPA was observed at the protein level, as assessed by immunoblotting, with quantification relative to the GAPDH loading control confirming that this was not a result of unequal protein loading (Figure 4.4).

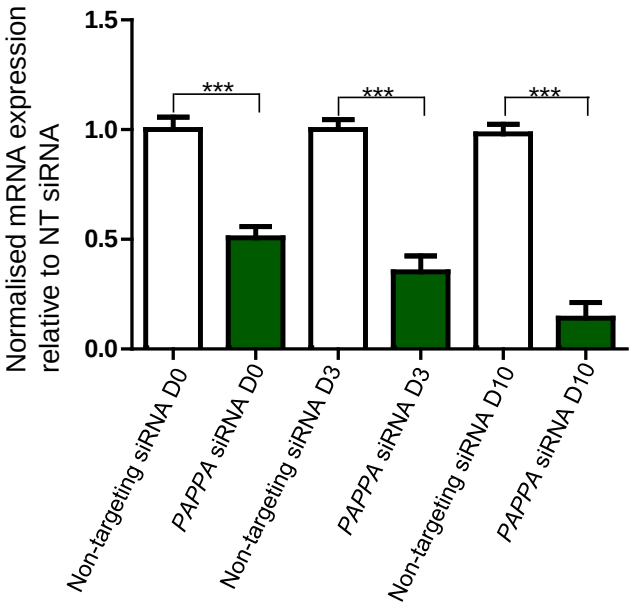



**Figure 4.3**

*PAPPA* expression in hMSCs differentiating into osteoblasts following siRNA knockdown. Gene expression was quantified using qPCR and normalised to the corresponding non-targeting siRNA control for each time point. Gene expression was significantly decreased at all time points relative to the non-targeting control. D0 = day 0, D3 = day 3 and D10 = day 10. *n* = Three independent non-OA donors (Lonza, UK.). Three technical repeats per donor per time point. qPCR was carried out as standard, with three technical replicates. Error bars indicate the SEM and statistical significance was assessed using the Mann-Whitney *U* test. *** $p < 0.001$.

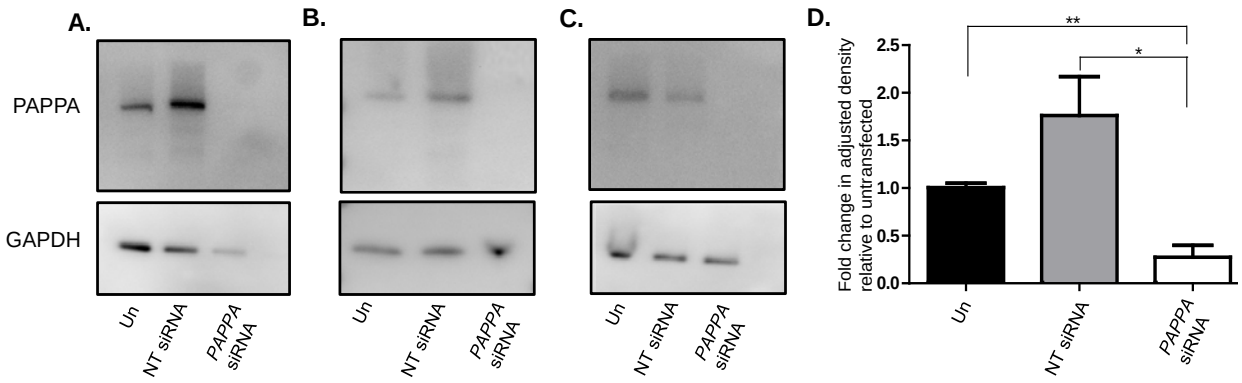



**Figure 4.4**

Knockdown of PAPPA protein in hMSCs on day 3 of osteogenic differentiation following *PAPPA* siRNA knockdown. Immunoblotting confirmed protein depletion in three independent non-OA donors (A = 2454E, B = 26601 & C = 26798). Each of the three knockdowns resulted in a decrease in PAPPA protein levels. The loading control, GAPDH, is shown directly beneath the corresponding knockdowns. The untransfected control is denoted by 'un'. Densitometry to calculate the adjusted density relative to untransfected cells was performed using IMAGEJ software in order to account for unequal protein loading (D). The relative adjusted densities were combined for the three donors and confirms successful protein knockdown in cells transfected with *PAPPA* siRNA. NT (non-targeting siRNA control). For image analysis the freeware ImageJ v1.4 was used (accessed via the NIH website (http://rsb.info.nih.gov/ij). Statistical significance was assessed using an upaired Student's *t* test and error bars indicate the SEM. * $p < 0.05$; ** $p < 0.01$.

Having confirmed the successful knockdown of PAPPA at the mRNA and protein level, the next necessity was to ensure the osteogenic differentiation had been successful. A key feature of osteogenic differentiation of MSCs using the *in vitro* monolayer model is alkaline phosphatase transcription and protein activity, which peaks between days 5 to 14 of differentiation [139]. Expression of the *ALPL* gene, coding for alkaline phosphatase, was measured on days 0, 3 and 10 of osteogenesis by qPCR and this confirmed significantly increased expression in the untransfected and NT controls compared to *PAPPA* siRNA transfected cells (Figure 4.5). Increased translation of the protein was investigated using Fast blue RR salts as an indicator of alkaline phosphatase activity (Figure 4.6). Quantification of staining in the three donors on day 3 of differentiation confirmed the enzyme's activity, indicating osteogenic differentiation was successful.

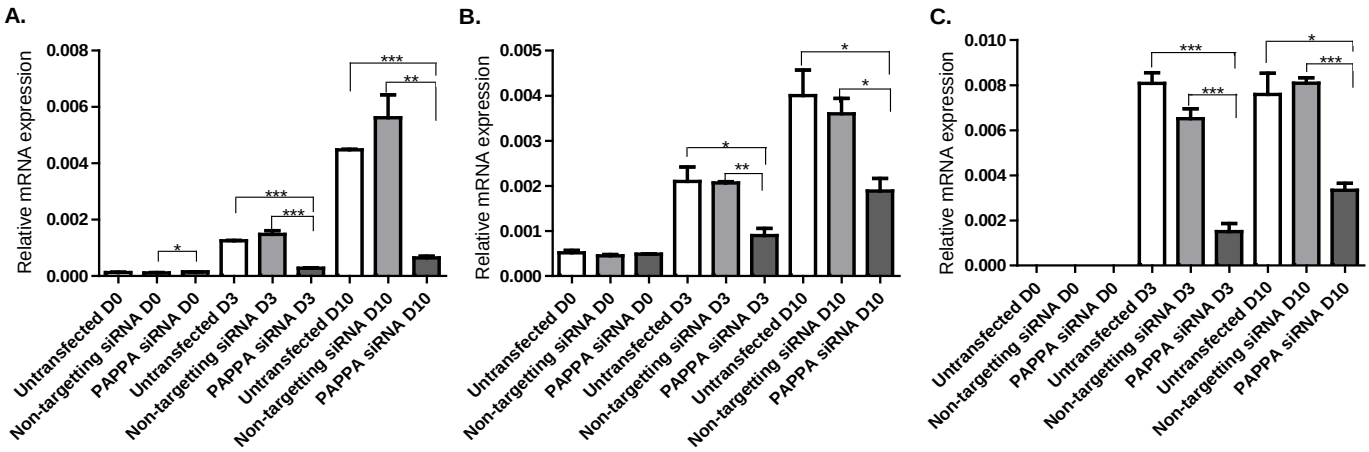



**Figure 4.5**

Alkaline phosphatase (*ALPL*) gene expression in hMSCs undergoing osteogenic differentiation following siRNA transfection. Gene expression was quantified for each condition on days 0 (D0), 3 (D3) & 10 (D10) of differentiation using qPCR and normalised to the *18S*, *GAPDH* and *HPRT1* housekeeping genes in the hMSCs from the three Lonza donors (A = 2454E, B = 26601 & C = 26798). *ALPL* expression increased in all three donors by day 10, indicating successful osteogenic differentiation. Three technical repeats were performed per time point, per condition and qPCR was performed as standard, including three technical replicates. Error bars represent the SEM. Statistical significance was assessed using an unpaired Student's *t* test. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

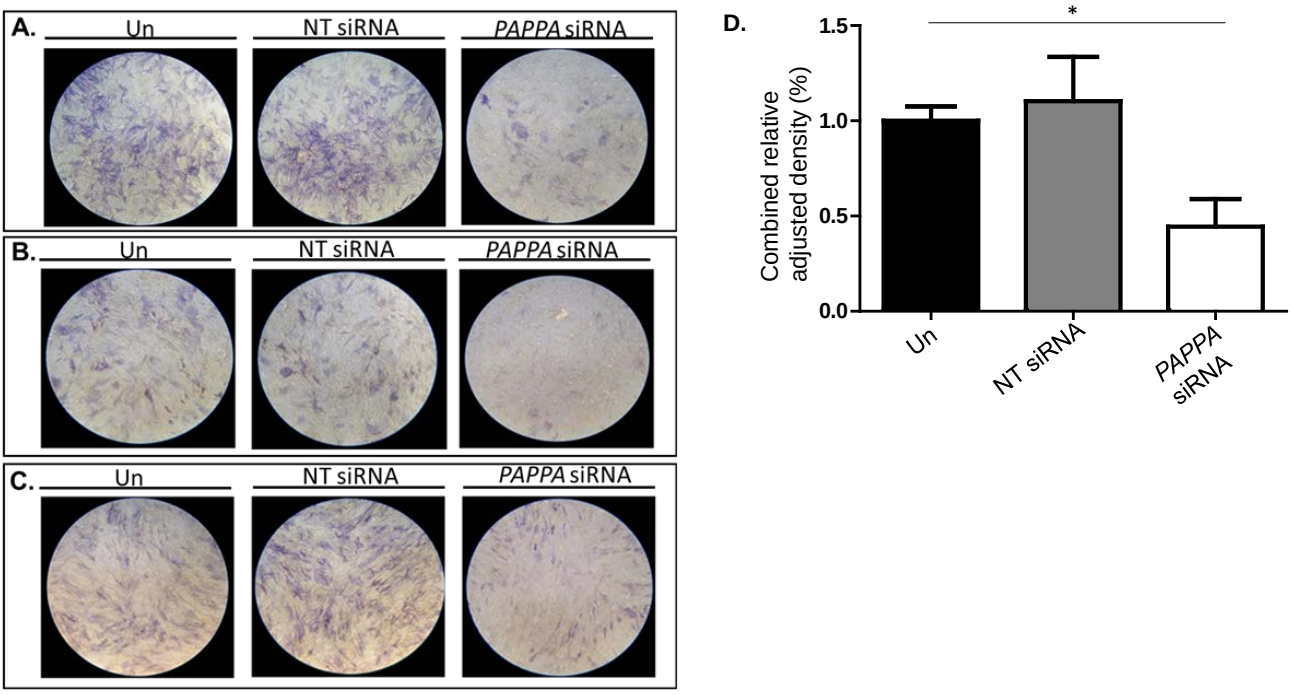



**Figure 4.6**

Tissue nonspecific alkaline phosphatase (TNAP) staining of hMSCs on day 3 of osteogenic differentiation. After being fixed in 4% paraformaldehyde, hMSCs from three donors (A = 2454E, B = 26601 & C = 26798) were stained for alkaline phosphatase enzymatic activity using TNAP staining (required components: napthol AS-MX-phosphate, dimethylformamide and Fast blue RR salt) on day 3 of osteogenesis. This macroscopically indicated successful differentiation, leading to increased alkaline phosphatase translation and therefore activity, in untransfected cells (un) and cells transfected with non-targeting (NT) siRNA. Staining appears far weaker in *PAPPA* siRNA transfected cells. Quantification of the percentage of the area stained, defined as the number of pixels over a user-determined threshold intensity, was measured using ImageJ V1.4 software (accessed via the NIH website (http://rsb.info.nih.gov/ij) and is shown in panel D relative to untransfected cells. Statistical significance was assessed using the Student's *t* test; * $p < 0.05$.

Having confirmed successful osteogenic differentiation, the next step was to assess the osteogenic differentiation capacity of *PAPPA* siRNA transfected cells in comparison to the NT siRNA control cells. To achieve this, expression of *ALPL*, *COL1A1* and *RUNX2* relative to NT siRNA were investigated by qPCR. All three genes have a well-defined role in osteogenesis and are induced following osteogenic differentiation of hMSCs [140]. Expression of *ALPL* was re-analysed as, in this phase of the experiment, we were attempting to investigate the osteogenic capacity of the *PAPPA* depleted cells as opposed to confirming overall successful differentiation in the experiment. mRNA expression of the three osteogenic markers was significantly lower in *PAPPA* siRNA treated cells compared to the NT siRNA control ($p \leq 0.05$) (Figure 4.7). This therefore suggests osteogenic differentiation is impaired in hMSCs following *PAPPA* depletion.

To further investigate the reduced osteogenic capacity of *PAPPA* siRNA transfected cells, mRNA levels of *KLF4* were measured. KLF4 is a neural transcription factor identified by Takahashi and Yamanaka as one of the core factors required for somatic cell reprogramming to induced pluripotent stem (iPS) cells and is therefore a genetic marker of pluripotency [141]. *KLF4* expression was significantly higher in *PAPPA* siRNA transfected cells compared to the NT siRNA control on days 3 and 10 (Figure 4.7). This consequently supports the hypothesis that depletion of *PAPPA* reduces the osteogenic capacity of hMSCs and that these cells, rather than differentiation towards a different mesenchymal lineage, maintain their pluripotent differentiation capacity.

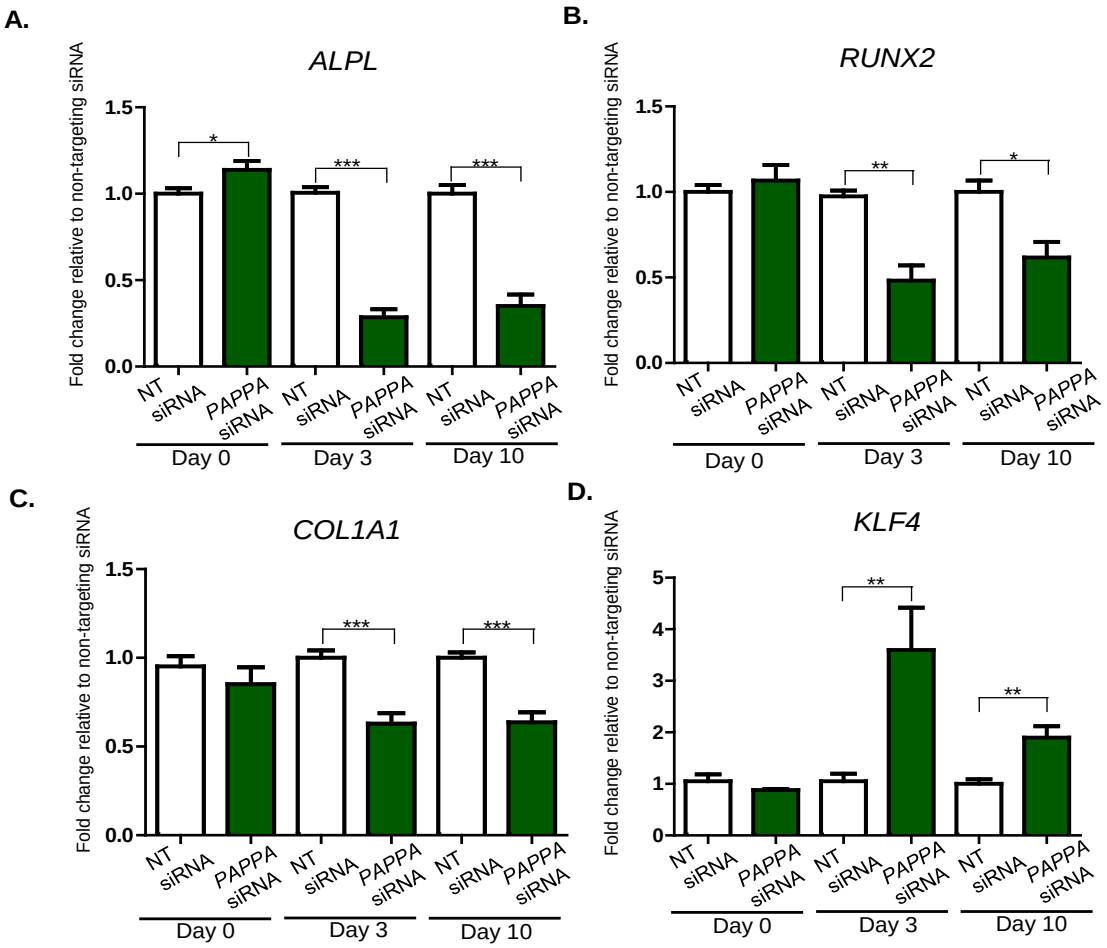



**Figure 4.7**

Expression of genetic markers of hMSC differentiation in non-OA hMSCs following *PAPPA* siRNA knockdown. hMSCs from three separate non-OA donors (Lonza, UK) were transfected with non-targeting siRNA, or siRNA targeting *PAPPA* and cultured in osteogenic differentiation media for 10 days. Gene expression of the osteogenic markers *ALPL* (A), *RUNX2* (B) and *COL1A1* (C), as well as *KLF4* (D), a marker of pluripotency, were quantified using qPCR and normalised to the corresponding expression in the non-targeting siRNA control (*y*-axis = 1). *n* = knockdown of three independent donors (biological replicates) with three technical replicates each. qPCR was performed as standard, including three technical replicates. Error bars represent the SEM. Statistical significance was assessed using the Mann-Whitney *U* test. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

### 4.3.4 siRNA Knockdown in MSCs undergoing chondrogenic differentiation

Having identified that *PAPPA* depletion results in impaired osteogenic differentiation of hMSCs, it was then of interest to investigate whether differentiation into other cells of the synovial joint was impaired. *In vitro* pellet chondrogenesis was performed following siRNA transfection, using hMSCs from the three same donors as in 4.3.2, as described in Chapter 2.11.5.

qPCR confirmed successful knock down of *PAPPA* mRNA in cells following transfection and throughout the 10 day chondrogenic differentiation relative to the corresponding NT siRNA control ($p < 0.001$) (Figure 4.8). Furthermore, the resulting levels of protein after the siRNA knockdowns were assessed by immunoblotting and PAPPA protein depletion was confirmed in all 3 donors in protein extracted from cells on day 3 of chondrogenesis, relative to untransfected and NT siRNA controls (Figure 4.9). Though the GAPDH loading control indicated unequal quantities of protein between conditions, densitometry analysis confirmed the depletion of the PAPPA protein relative to the controls (Figure 4.9 D).

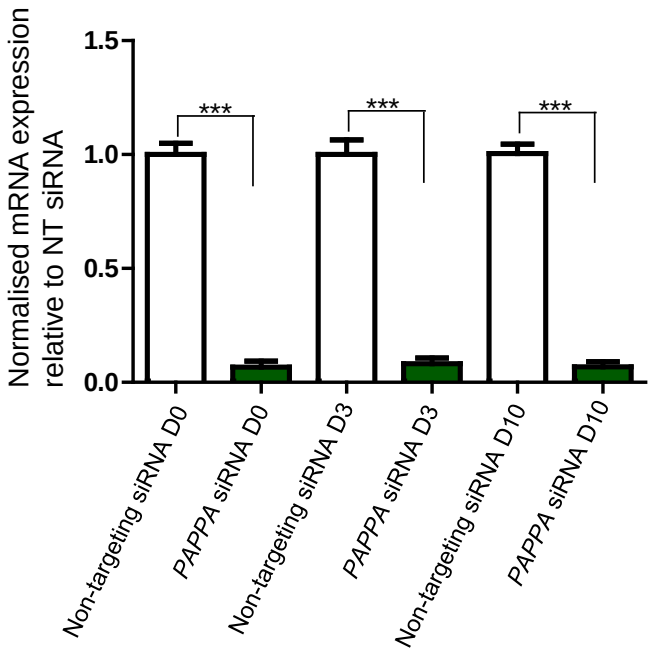



**Figure 4.8**

*PAPPA* expression in hMSCs differentiating into chondrocytes following siRNA knockdown. Gene expression was quantified using qPCR and normalised to the corresponding non-targeting siRNA control for each time point. Gene expression was significantly decreased at all time points relative to the non-targeting control. D0 = day 0, D3 = day 3 and D10 = day 10. *n* = Three independent non-OA donors (Lonza, UK.), three technical repeats per donor per time point. qPCR was carried out as standard, with three technical replicates. Error bars indicate the SEM and statistical significance was assessed using the Mann-Whitney *U* test; *** $p < 0.001$.

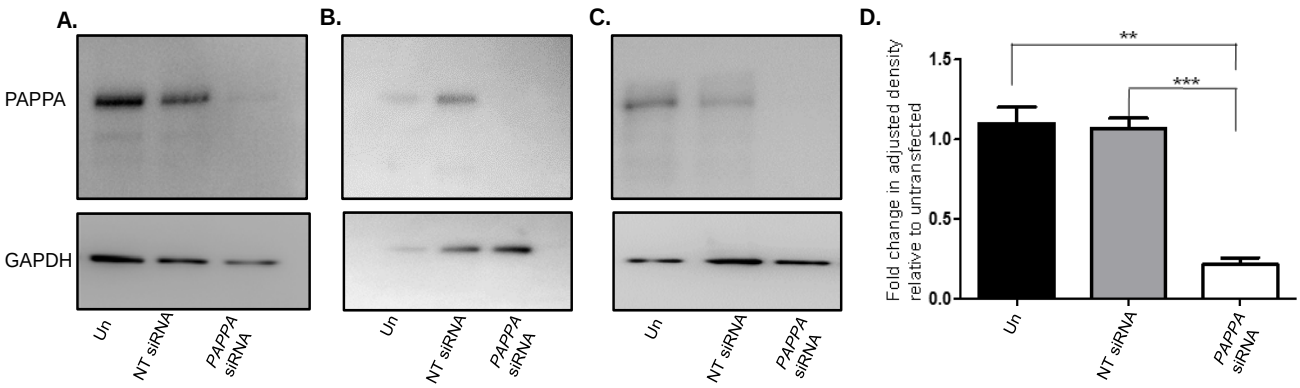



**Figure 4.9**

Knockdown of PAPPA protein in hMSCs on day 3 of chondrogenic differentiation following *PAPPA* siRNA knockdown. Immunoblots were performed to confirm protein depletion in three independent non-OA donors (A = 2454E, B = 26601 & C = 26798). Each of the three knockdowns resulted in a decrease in PAPPA protein levels. The loading control, GAPDH, is shown directly beneath the corresponding knockdowns. The untransfected control is denoted by 'un'. Densitometry to calculate the adjusted density relative to untransfected cells was performed using IMAGEJ software in order to account for unequal protein loading (D). The relative adjusted densities were combined for the three donors and confirms successful protein knockdown in cells transfected with *PAPPA* siRNA. NT (non-targeting siRNA control). For image analysis the freeware ImageJ v1.4 was used (accessed via the NIH website (http://rsb.info.nih.gov/ij). Error bars indicate the SEM and statistical significance was assessed using an unpaired Student's *t* test. ** $p < 0.01$; *** $p < 0.001$.

In order to confirm successful chondrogenic differentiation of the hMSCs, *ACAN* mRNA expression was quantified by qPCR. *ACAN* codes for aggrecan, a well-characterised hyaline matrix molecule, which transiently increases during *in vitro* cartilage development, peaking at differentiation day 14 [142]. *ACAN* expression increased in all three donors under all three conditions, with the highest recorded expression relative to the housekeeping genes recorded on day 10 of differentiation (Figure 4.10).

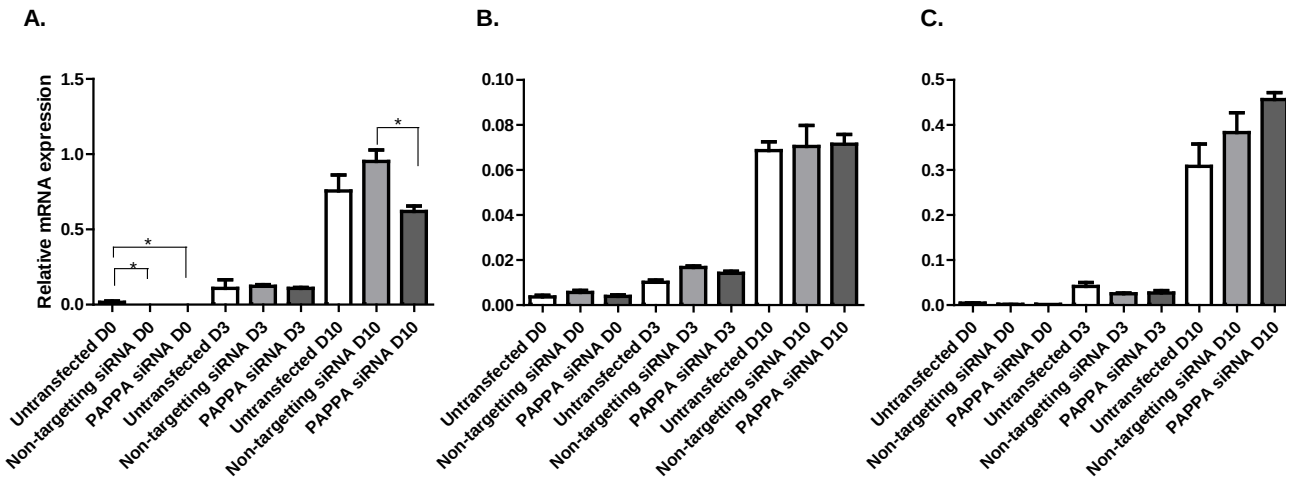



**Figure 4.10**

Aggrecan (*ACAN*) gene expression in hMSCs undergoing chondrogenic differentiation following siRNA transfection. Gene expression was quantified for each condition on days 0 (D0), 3 (D3) & 10 (D10) of differentiation by qPCR and normalised to the *18S*, *GAPDH* and *HPRT1* housekeeping genes in the hMSCs from the three Lonza donors (A = 2454E, B = 26601 & C = 26798). *ACAN* expression increased in all three donors by day 10, indicating successful chondrogenic differentiation. Three technical repeats were performed per time point, per condition and qPCR was performed as standard, including three technical replicates. Error bars represent the SEM. Statistical significance was assessed using an unpaired Student's *t* test. * $p < 0.05$.

Having confirmed successful PAPPA knock-down and chondrogenic differentiation, we next sought to interrogate the chondrogenic differentiation capacity of *PAPPA* siRNA transfected cells in comparison to the NT siRNA control cells. This was achieved by analysis of *ACAN* and *SOX9* by qPCR. Expression of *ACAN* was re-analysed as, in this phase of the experiment, we were attempting to investigate the chondrogenic capacity of the *PAPPA* depleted cells relative to the NT siRNA controls, as opposed to confirming overall successful differentiation in the experiment. *SOX9* was selected along with *ACAN* as it is a chondrogenic lineage marker [143]. No significant difference was found at any time point when analysing *ACAN* expression relative to the NT siRNA control (Figure 4.11 A). A significant increase in *SOX9* on day 10 of differentiation was noted, however ($p$ = 0.007) (Figure 4.11 B). This suggests that, unlike during osteogenesis, *PAPPA* depletion in hMSCs does not prevent chondrogenic differentiation and may actually result in enhanced chondrogenic capacity.

In order to interrogate the chondrogenic capacity of *PAPPA* depleted hMSCs further, gene expression of *COL10A1* was also quantified. *Col10A1* is expressed by hypertrophic chondrocytes, is upregulated in OA cartilage and the encoded protein (type X collagen) is required for the developmental process of endochondral ossification [144]. No significant difference was seen in *COL10A1* expression on days 0 or 10 of differentiation between *PAPPA* siRNA and NT siRNA transfected cells (Figure 4.11 C). Overall, this investigation revealed that, *in vitro*, PAPPA depletion does not inhibit the chondrogenic capacity of hMSCs, and may increase differentiation potential towards normal, non-hypertrophic, chondrocytes.

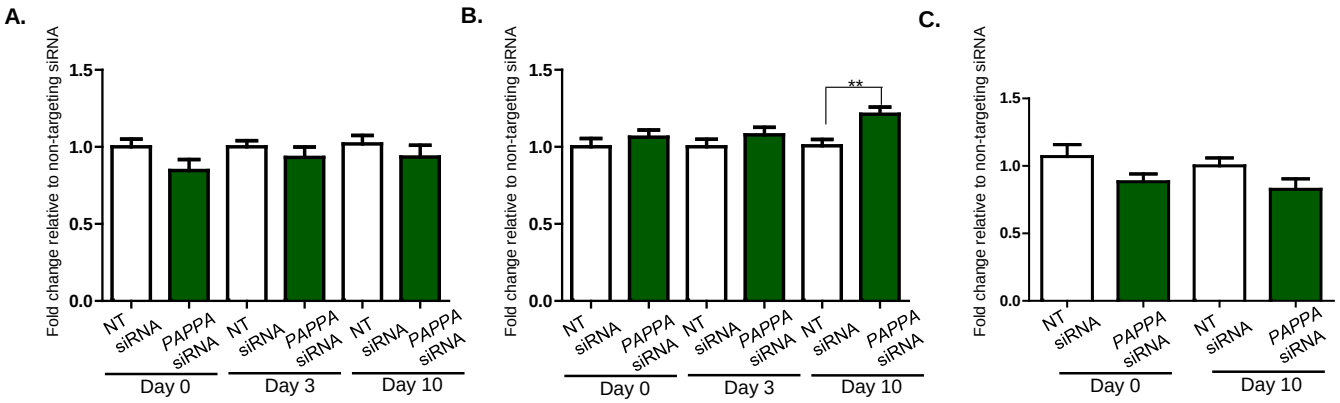



**Figure 4.11**

Expression of genetic markers of chondrogenic differentiation in hMSCs following *PAPPA* siRNA knockdown. hMSCs from three separate non-OA donors (Lonza, UK) were transfected with non-targeting siRNA, or siRNA targeting *PAPPA* and cultured in chondrogenic differentiation media for 10 days (donor information Table 4.1). Gene expression of chondrogenic markers *ACAN* (A), *SOX9* (B) and *COL10A1* (C) were quantified using qPCR and normalised to the corresponding expression in the non-targeting siRNA control (*y*-axis = 1). No difference in expression of *ACAN* and *COL10A1* was detected at any time point, while an increase in *SOX9* expression was seen on day 10 in *PAPPA* siRNA transfected cells. *n* = knockdown of three independent donors (biological replicates) with three technical replicates each. qPCR was performed as standard, including three technical replicates. Error bars represent the SEM. Statistical significance was assessed using the Mann-Whitney *U* test. ** $p < 0.01$.

### 4.3.5 *PAPPA* expression in OA bone stratified by rs4836732 genotype

Having confirmed a potential role for *PAPPA* in synovial joint development in an *in vitro* model of osteogenesis, it was subsequently of interest to investigate *PAPPA* expression in OA bone. Stratification by rs4836732 genotype is necessary to investigate if an eQTL is operating at this locus. Dr Amanda Villavilla of Prof. Loughlin's group (Institute of Cellular Medicine, Newcastle University) extracted and expanded osteoblasts from OA patients who had undergone joint arthroplasty, before then extracting nucleic acids. I then genotyped the samples (Chapter 2.5.2) and performed qPCR (Chapter 2.13.2). Table 4.3 provides a summary of donor information, including genotype at rs4836732. qPCR primer and probe sequences can be found in Appendix A: Table A.4, and the assay positions are detailed in Figure 3.11 (housekeeping genes) and Figure 3.12 (*PAPPA*).

| Sample Number | Sex | Affected joint | rs4836732 genotype |
|---|---|---|---|
| 5918 | Male | Hip | TT |
| 5928 | Male | Knee | CT |
| 5930 | Male | Hip | CT |
| 5931 | Male | Knee | CC |
| 5935 | Male | Knee | TT |
| 5937 | Female | Knee | TT |
| 6002 | Male | Knee | CC |
| 6018 | Male | Knee | CT |
| 6019 | Female | Knee | CC |
| 6062 | Female | Knee | TT |
| 6111 | Male | Knee | CT |
| 6132 | Female | Hip | TT |
| 6168 | Female | Hip | CT |
| 6192 | Female | Knee | CT |
| 6197 | Male | Knee | CC |
| 6201 | Male | Hip | CT |
| 6209 | Male | Knee | CT |
| 6213 | Female | Knee | CT |
| 6222 | Female | Knee | CT |
| 6223 | Male | Knee | CT |
| 6241 | Female | Knee | CT |
| 6247 | Male | Knee | TT |
| 6252 | Female | Hip | CC |

**Table 4.3**
Demographic data and rs4836732 genotype in OA patients used for *PAPPA* expression analysis in bone.

qPCR of *PAPPA* expression in OA bone did not yield any statistically significant differences when stratified by rs4836732 genotype (Figure 4.12). Analysis according to the genotype at the association signal risk strata, female hip, by rs4836732 genotype was not possible owing to low sample numbers (Table 4.3). When stratified by overall expression, irrespective of genotype, there was no statistically significant gene expression in any individual stratum (Figure 4.13).

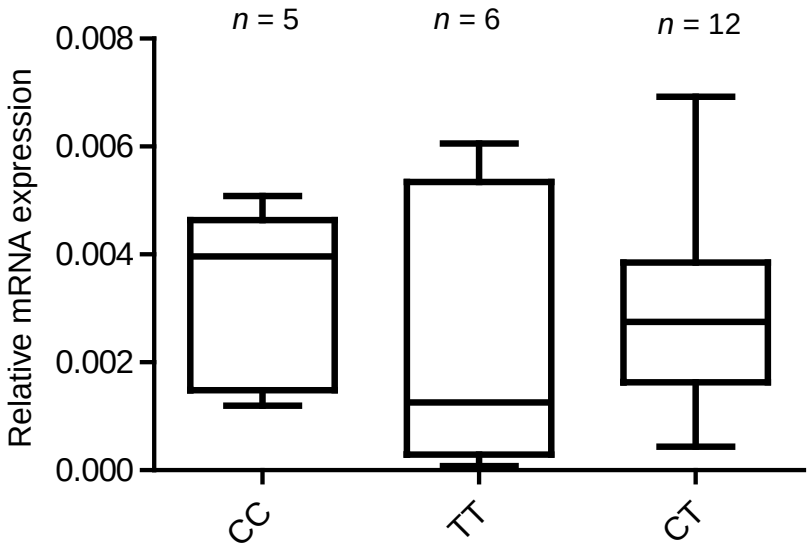



**Figure 4.12**
*PAPPA* expression in OA bone stratified according to rs4836732 genotype. Gene expression is shown relative to *18S*, *GAPDH* and *HPRT1* housekeeper expression. A box and whisker plot is shown representing the interquartile range (IQR) with median, 25th and 75th percentile, minimum and maximum values. No significant difference was seen in *PAPPA* gene expression between the three genotypes. Statistical significance was assessed using the Mann-Whitney *U* test for two-way comparisons and a one-way analysis of variance for a comparison of more than two groups. *n* = the number of OA patients in the comparison group.

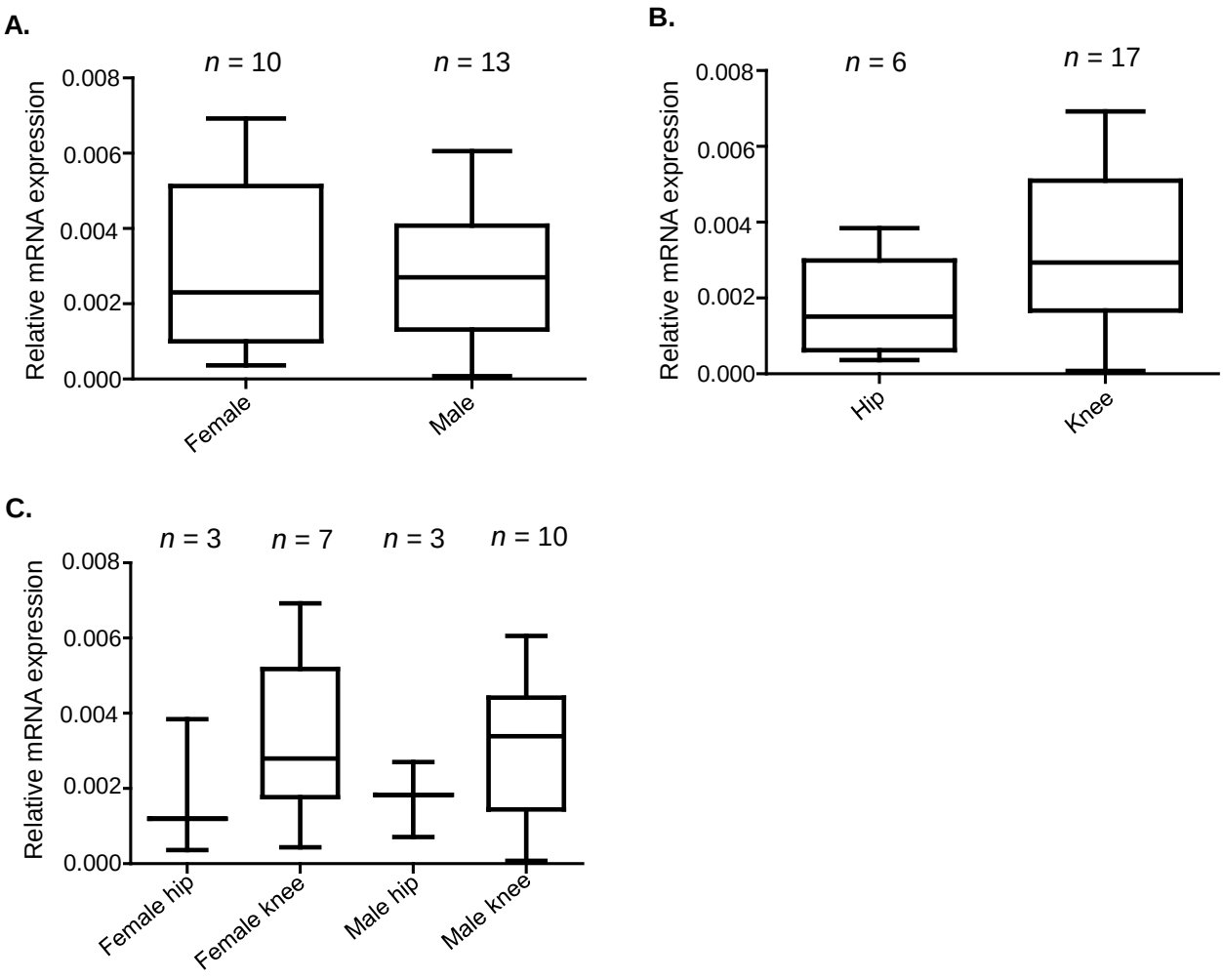



**Figure 4.13**

*PAPPA* expression in OA bone stratified according to sex (A), affected joint (B) and both sex and affected joint (C). Gene expression is shown relative to *18S*, *GAPDH* and *HPRT1* housekeeper expression. A box and whisker plot is shown representing the interquartile range (IQR) with median, 25th and 75th percentile, minimum and maximum values. No significant difference was seen in *PAPPA* gene expression in any of the comparison groups. Statistical significance was assessed using the Mann-Whitney *U* test for two-way comparisons, using all possible comparison combinations, and a one-way analysis of variance for a comparison of more than two groups. *n* = the number of OA patients in the comparison group.

## 4.4 Discussion

The aim of this chapter was to investigate the role of PAPPA in human synovial joint development by using lineage-specific hMSC differentiation as an *in vitro* model of joint embryogenesis. Having not identified any evidence for rs4836732 acting as an eQTL in end-stage OA tissues, I hypothesised that the SNP mediates disease risk by functioning during joint development. Of the genes within the association region, *PAPPA* is arguably the most promising candidate and so, this chapter sought to interrogate the role of the gene in a model of human skeletogenesis.

I utilised lineage-specific hMSC differentiation as an *in vitro* model of synovial joint development in order to explore the functional role of *PAPPA* in human joint cell differentiation. I depleted *PAPPA* gene and protein expression in hMSCs using siRNA before inducing osteogenic or chondrogenic differentiation. This confirmed the studies performed in murine models, finding that *PAPPA* depletion resulted in impaired osteogenesis, likely resultant of aberrant IGF regulation. Chondrogenesis, however, was not inhibited when assessed by qPCR analysis of chondrogenic markers. Conversely, an increase in *SOX9* expression was seen on day 10 of chondrogenesis in *PAPPA* depleted cells. Given that *SOX9* is an upstream regulator of *ACAN*, this suggests *PAPPA* deletion leads to enhanced chondrogenesis [145]. As such, these experiments suggest *PAPPA* is functional during the differentiation of cell types which have essential roles in skeletogenesis. It is well-characterised that aberrant skeletogenesis, for example in the instance of congenital hip dysplasias, results in increased risk of OA [136]. Consequently, based on the above findings, I hypothesised that the rs4836732 signal may increase the risk of female hip OA by acting as a *cis*-eQTL, modulating *PAPPA* expression during joint development.

Although a role for *PAPPA* in an *in vitro* model of skeletogenesis has been identified, further interrogation into the *in vivo* function of *PAPPA* during human embryogenesis is not possible. However, given that global mouse PAPPA knock outs display age-associated trabecular deficiencies, it is possible that aberrant *PAPPA* expression, regulated by the rs4836732 association signal, may also contribute to the abnormal bone remodelling present in end-stage OA [107]. Osteoblasts isolated from the bone of end-stage OA patients undergoing joint arthroplasty were therefore used to analyse overall *PAPPA* expression, and this data was then stratified by rs4836732 genotype. Though no statistically significant difference in expression was identified in individuals homozygous for the C risk allele, the small sample size and limitations caused by inter-individual differences when studying overall gene expression could be making any true correlations. Consequently, it would be prudent for future research to analyse allelic expression imbalance of *PAPPA* in bone according to rs4836732 genotype.

In Chapters three and four, I have functionally analysed the rs4836732 female hip-associated OA signal that emerged from the arcOGEN GWAS [92]. I have concluded, through both data imputation and luciferase reporter assay studies, that the association signal is the most compelling candidate driving the increased OA risk. I have not found any evidence of a correlation between the association SNP genotype and expression of candidate genes within the LD region (*ASTN2*, *PAPPA* and *TRIM32*) in end-stage disease tissue. Based on published studies, I have identified *PAPPA* as the most compelling gene candidate and subsequently, I have identified a functional role for *PAPPA* in hMSC differentiation into cell types present in the synovial joint. Taken together, the data presented in these chapters therefore

support the hypothesis that the association signal may regulate OA risk during joint development and cell differentiation, as opposed to in end-stage disease tissue. Owing to the difficulties presented in studying gene expression during human skeletogenesis, I consequently decided to pursue alternative research targets during the remainder of my PhD.

# Chapter 5: Identification of hip and knee OA patients with increased cartilage inflammation using methylation analysis by pyrosequencing

## 5.1 Introduction

A high-density CpG methylation array, which analysed cartilage DNA from 23 hip OA patients, 73 knee OA patients and 21 hip control patients with a neck of femur fracture (NOF), identified numerous differentially methylated loci (DMLs) between the patient cohorts [43]. Interestingly, hip OA samples formed two distinctive clusters, displaying a total of 15,239 DMLs enriched for genes involved in inflammation and immunity. Knee OA samples were also found to form two distinct clusters, though less defined than in the hip samples. In knee OA, 5,769 DMLs were identified, many of which are within genes coding for inflammatory factors. This indicates inflammation is crucial to OA pathogenesis and potentially presents an exciting therapeutic opportunity for anti-inflammatory treatments in OA. In order to carry forward these intriguing findings, which were generated by our group, it is vital to identify whether the inflammatory signature is present in blood, thus enabling for patient stratification in future clinical trials and patient diagnostics.

In light of this, I sought to create pyrosequencing assays to CpG sites which have the capacity to differentiate OA cartilage samples into the inflammatory or non-inflammatory clusters. The principle being that it would then be possible to negate the requirement for a costly and time-consuming genome-wide methylation array, and new samples could be easily clustered solely using methylation analysis by pyrosequencing. This would therefore provide the capacity to further interrogate the inflammatory status of the cartilage samples, and of matched blood, in an attempt to

better comprehend the mechanism and implications of the increased inflammation. This chapter therefore focuses on the identification of CpG sites which are able to discriminate between the non-inflammatory and inflammatory cluster, creation of a logistic function in order to predict cluster membership using methylation at these sites and the design of pyrosequencing assays against the selected CpGs. This would consequently enable quantification of CpG methylation in new cartilage DNA samples and subsequent prediction of cluster membership.

Further to differential methylation, quantitative real-time PCR (qPCR) of non-inflammatory and inflammatory OA hip samples confirmed these subgroups to not only display epigenetic differences, but to also be transcriptiomically distinct [121]. In the inflammatory group, promoter hypomethylation of inflammatory factors was associated with increased expression of these genes, as was discussed in Chapter 1.6.3. Consequently, this chapter also sought to replicate and confirm these findings in cartilage samples available from the microarray, and to subsequently analyse inflammatory gene expression in new hip cartilage samples to assist in the validation of our new methylation assays and discriminant function. Moreover, this chapter aimed to also analyse inflammatory gene expression in OA knee samples present on the microarray, in order to confirm whether the transcriptome of knee OA inflammatory samples also differs when compared to non-inflammatory knee OA samples.

## 5.2 Aims

This chapter sought to assist in the investigation of an inflammatory subgroup of OA patients who have altered methylation of inflammatory genes, by creating a method to determine cluster membership of new patients. It also intended to further interrogate the inflammatory transcriptome of patients in the divergent methylation clusters. This was achieved by:

- Selection of CpG methylation sites analysed on the high-density CpG methylation array for categorisation of new hip OA cartilage samples into the non-inflammatory or inflammatory subgroup.
- Selection of CpG methylation sites analysed on the high-density CpG methylation array for categorisation of new knee OA cartilage samples into the non-inflammatory or inflammatory subgroup.
- Creation and testing of multiple logistic regression models predicting cluster membership, based on methylation at the selected CpG sites, as determined by CpG methylation array, to determine cluster membership in hip and knee OA patients, performed with JMP Pro Version 11.2 Statistical Data Visualization Software.
- Design of pyrosequencing assays to the selected hip and knee CpGs for analysis of CpG methylation by pyrosequencing.
- Validation of CpG assays using bisulphite converted DNA from cartilage samples analysed on the high-density methylation array.
- CpG site methylation analysis by pyrosequencing using new cohorts of both hip and knee OA patients, with subsequent application of the logistic functions to predict cluster membership.

- Replication of inflammatory gene expression analysis in hip OA cartilage samples analysed on the high-density methylation array.
- Validation of the multiple logistic regression models predicting cluster membership by inflammatory gene expression analysis in new hip OA cartilage.
- Inflammatory gene expression analysis in knee OA cartilage samples analysed on the high-density methylation array and in new knee OA cartilage samples.

## 5.3 Results

### 5.3.1 Selection of CpG sites for cluster differentiation in OA hip cartilage

CpG sites analysed on the high-density methylation array were assessed according to both their *p* values and β values, in order to determine the best candidates to take forward and use to determine cluster membership in OA hip cartilage. Methylation data from the array was provided as β values, ranging from 0 (no methylation) to 1 (100% methylation). Large differences in the average β values between the clusters would increase the likelihood of successfully designing pyrosequencing assays that have the sensitivity to also discriminate between the clusters, whilst significant *p* values would provide reassurance that the methylation differences are genuine and not artefactual. Based on these criteria, cg14818279 and cg03278514 were selected (Table 5.1). It is to be noted that the CpG sites were selected based purely on their *p* values and β values when comparing methylation in the non-inflammatory and inflammatory subgroups in hip OA cartilage, and their genomic location is entirely unrelated to the study.

A subsequent post-publication re-analysis of the CpG methylation array data by our group (performed by Andrew Skelton and Dr Louise Reynard; Institute of Cellular Medicine, Newcastle University), resulted in the removal of six samples from the original 23 hip OA cohort. However, this has not altered the capacity of the selected CpG sites to discriminate between the subgroups; Figure 5.1 presents the data for the discriminatory capacity of the two CpG in the 17 hip OA patients subsequently studied.

| Probe | Chromosome | Adjusted *p* value | Non-inflammatory cluster mean β value | Inflammatory cluster mean β value |
|---|---|---|---|---|
| cg14818279 | 5 | $0.07x10^{-7}$ | 0.32 | 0.61 |
| cg03278514 | 15 | 0.002 | 0.71 | 0.44 |

**Table 5.1**

Hip CpG sites selected from the high-density array to take forward for logistic analysis. The data shown was generated by Dr Michael Rushton using the Illumina Infinium HumanMethylation450 BeadChip array and is derived from all 23 hip OA patient samples analysed on the array [43]. The *p* values shown are those following Benjamini-Hochberg correction for multiple testing when comparing methylation in the non-inflammatory and inflammatory subgroups.

**A.**

| Probe | Chromosome | *p* value | Non-inflammatory cluster mean β value | Inflammatory cluster mean β value |
|---|---|---|---|---|
| cg14818279 | 5 | $0.08 \times 10^{-2}$ | 0.43 | 0.58 |
| cg03278514 | 15 | $0.01 \times 10^{-2}$ | 0.71 | 0.44 |

**B.**

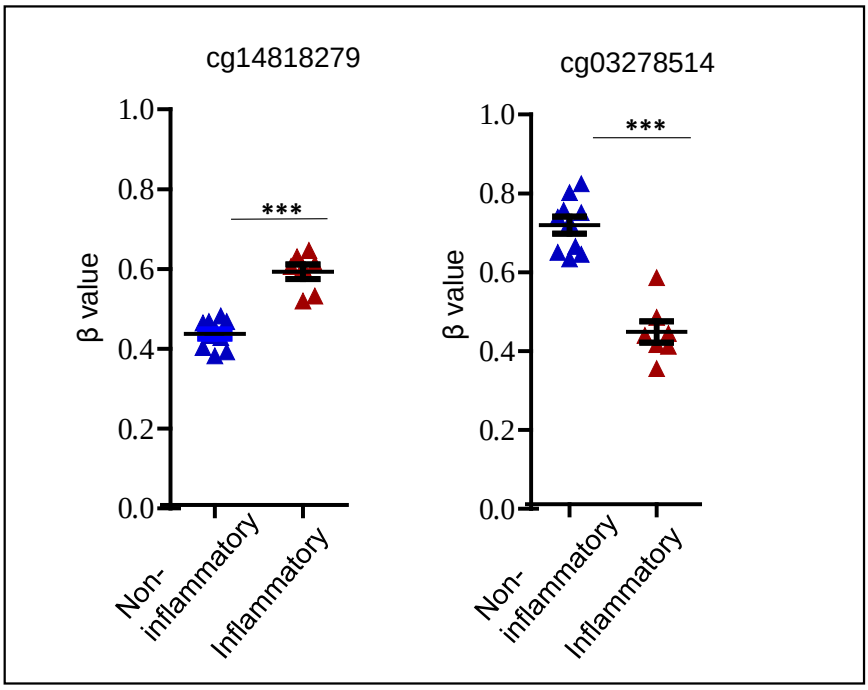



**Figure 5.1**

Analysis of methylation at the selected CpG sites in the revised cohort of 17 hip OA patient samples. The data shown was generated by Dr Michael Rushton using the Illumina Infinium HumanMethylation450 BeadChip array [43]. The mean cluster β values and *p* values are shown in panel (A), following removal of six of the original cohort of 23 patients. β values for the 17 patients are plotted in panel (B) according to the inflammatory subgroup for both CpG sites. Each data point represents a single patient sample; horizontal lines and error bars show the mean ± SEM. Statistical significance was assessed using a Mann-Whitney *U* test. *** $p \leq 0.001$. Non-inflammatory $n = 10$, inflammatory $n = 7$ (where $n$ = number of patient samples in each comparison).

### 5.3.2 Logistic regression modelling to predict hip cartilage inflammatory cluster membership

Having identified CpG sites which appear to discriminate between the inflammatory subgroups in hip OA cartilage samples, based on the β values from the high-density methylation array, I then sought to create a logistic function utilising these CpG sites in order to predict cluster membership for future samples. Statistical analyses were performed with JMP Pro Version 11.2 Statistical Data Visualization Software with the assistance of Dr Dennis Lendrem (Institute of Cellular Medicine).

A multiple logistic regression model predicting cluster membership was built using β values at both CpG sites from the cohort of 10 non-inflammatory and 7 inflammatory cluster patients analysed on the high-density methylation array. This model is a multivariable function which predicts classification, in this case the non-inflammatory or inflammatory hip OA groups, based on the observed values of continuous variables, in this case methylation at the selected CpG sites. A model using cg14818279 and cg03278514 methylation permitted successful prediction of cluster membership using these two sites ($p \leq 0.001$) (Figure 5.2). Consequently, a full logistic regression model using a linear combination of both cg14818279 and cg03278514 was created and demonstrates separation of clusters according to methylation at these sites (Figure 5.3).

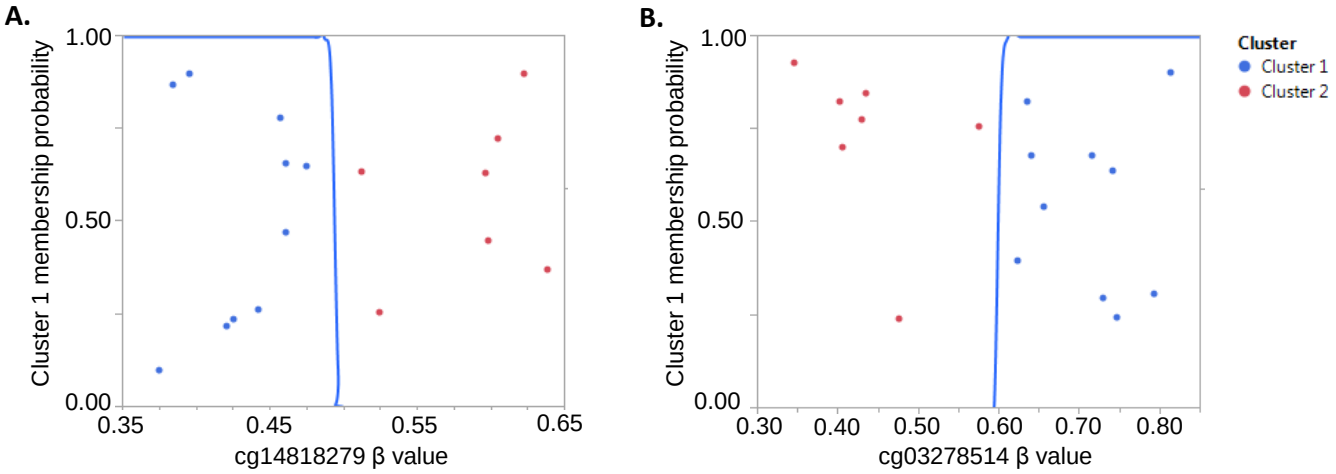



**Figure 5.2**

Logistic regression plots for the selected CpG sites: cg14818279 (A) and cg03278514 (B). The y axis shows the probability of belonging to cluster 1 (non-inflammatory) according to methylation at cg14818279 (A) or cg03278514 (B), and refers only to the blue regression line, not to the individual data points. Samples are plotted according to the β value at cg14818279 (A) or cg03278514 (B), demarcated on the x axis, as determined by the high-density CpG methylation array [43]. Each data point represents a single sample, where blue donates non-inflammatory cluster membership (cluster 1; $n$ = 10) and red the inflammatory cluster samples (cluster 2; $n$ = 7). The logistic regression analyses show that methylation at cg14818279 and cg03278514 both provide perfect separation of inflammatory and non-inflammatory samples ($p \leq 0.001$).

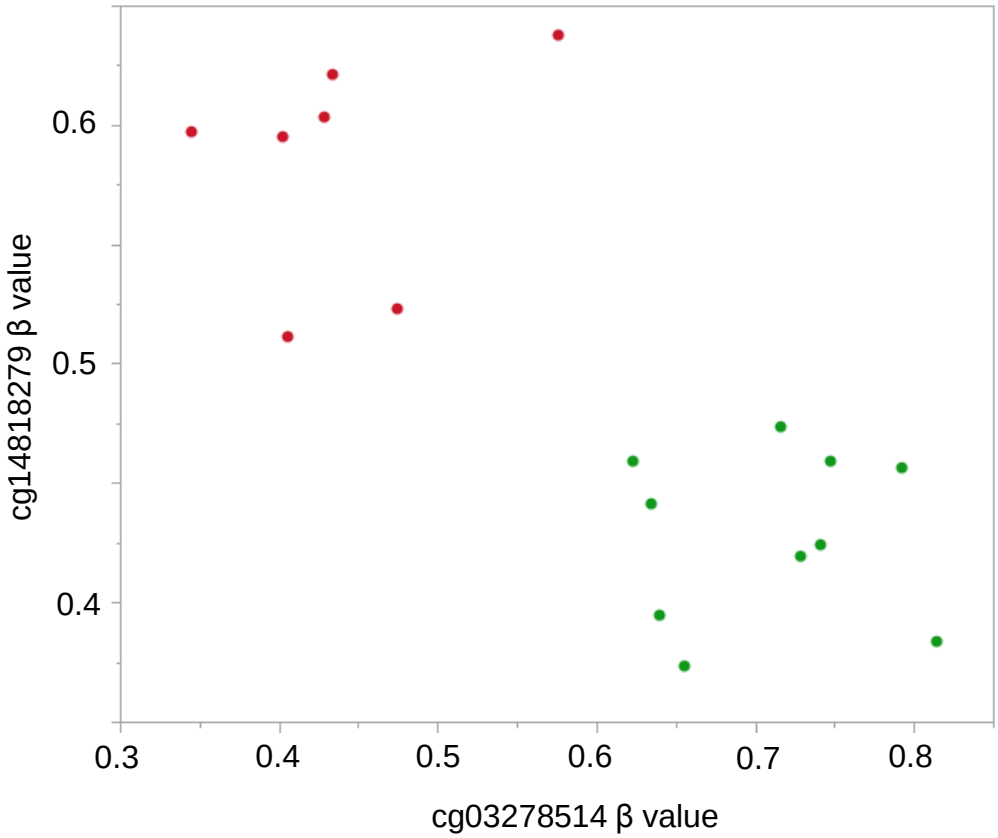



**Figure 5.3**

Hip OA cartilage cluster separation achieved using a linear combination of methylation at cg03278514 and cg14818279. Hip sample β values from the high-density CpG methylation array are plotted for cg03278514 (x axis) and cg14818279 (y axis). Non-inflammatory samples are denoted in green (cluster 1; *n* = 10) and inflammatory samples in red (cluster 2; *n* = 7). Methylation at these two sites provides clear separation of samples into the inflammatory and non-inflammatory clusters.

While the logistic regression model permits prediction of cluster membership, inspection of the plotted data permitted the creation of a simple heuristic decision tool. This would enable the cluster prediction model to be applied more easily by other researchers. Consequently, a heuristic cluster index (CI), was formulated using hip OA cartilage β values from the high-density CpG methylation array (Figure 5.4). Methylation β values were converted to percentage methylation values (0 = 0% methylation and 1 = 100% methylation) in order to create the CI. This allows for the CI to be applied to future samples, which will be analysed by pyrosequencing, a method that provides CpG methylation as a percentage value. Subsequent application of the CI successfully predicted cluster membership of all 17 hip OA samples, with zero misclassification (Figure 5.5).

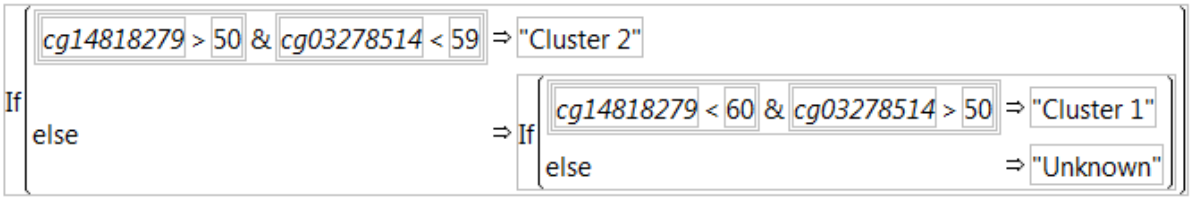



**Figure 5.4**

The cluster index (CI) created to predict hip cluster membership. The CI is based on methylation at cg14818279 and cg03278514 in hip OA cartilage, as determined based on the β values previously generated using the high-density CpG methylation array. The CI dictated that when methylation at cg14818279 is greater than 50% **and** at cg03278514 is less than 59%, samples are assigned as inflammatory (cluster 2). If methylation at cg14818279 is less than 60% **and** is greater than 50% at cg03278514, the sample is assigned as non-inflammatory (cluster 1). If methylation values fit neither category, the sample is not classified as belonging to either subgroup (unknown).

**Contingency table**

Most likely cluster

| CI | Count | Cluster 1 | Cluster 2 | Total |
|---|---|---|---|---|
| | Cluster 1 | 10 | 0 | 10 |
| | Cluster 2 | 0 | 7 | 7 |
| | Total | 10 | 7 | 17 |

**Figure 5.5**

JMP screenshot of cluster index (CI) cluster membership predictions for patients analysed on the high-density CpG methylation array. The CI correctly identified 10/10 samples as non-inflammatory (cluster 1) and 7/7 as inflammatory (cluster 2).

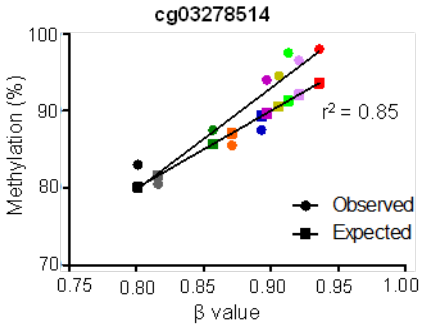
### 5.3.3 Design of pyrosequencing assays to determine methylation at cg14818279 and cg03278514

Having identified two CpGs that can successfully determine the inflammatory cluster of hip OA cartilage, the next step was to design pyrosequencing assays to cg14818279 and cg03278514 (Chapter 2.14.1). Primer sequences for the assays can be found in Appendix A: Table A.5.

After designing the primer sequences, validation of the assays was performed. This was necessary as pyrosequencing assays use a third primer, known as the sequencing primer, along with the forward and reverse primers of a standard PCR. The sequencing primer binds adjacent to the CpG and then extends over the site in the direction of the captured biotinylated primer. It is consequently necessary to validate the ability of the sequencing primer to quantify methylation at the CpG site. Validation was performed, in duplicate, using genomic DNA (gDNA) extracted from the cartilage of neck of femur fracture (NOF) patients analysed on the original genome-wide methylation array (patients used are listed in the tables in Figure 5.6). The gDNA was bisulphite treated, to convert unmethylated cytosine bases to thymine bases, allowing the reaction to quantify methylation at the sites. Comparison of the detected proportion of protected cytosine bases to the expected percentage, as determined by the high-density array, therefore allows for assay validation (Chapter 2.14.2; Figure 5.6). A positive correlation was seen for the observed methylation values, each with a goodness of fit $r^2 \geq 0.85$, and so the assays were deemed suitable for further use.

A.

| Patient | β value | Expected methylation (%) | Average observed methylation (%) |
|---|---|---|---|
| T172 | 0.801 | 80.1 | 83.0 |
| T168 | 0.816 | 81.6 | 80.5 |
| T152 | 0.857 | 85.7 | 87.5 |
| T199 | 0.871 | 87.1 | 85.5 |
| T191 | 0.893 | 89.3 | 87.5 |
| T150 | 0.897 | 89.7 | 94.0 |
| T177 | 0.906 | 90.6 | 94.5 |
| T117 | 0.913 | 91.3 | 97.5 |
| T193 | 0.921 | 92.1 | 96.5 |
| T151 | 0.936 | 93.6 | 98.0 |




B.

| Patient | β value | Expected methylation (%) | Average observed methylation (%) |
|---|---|---|---|
| T150 | 0.278 | 27.8 | 20.5 |
| T117 | 0.335 | 33.5 | 35.5 |
| T184 | 0.347 | 34.7 | 29.0 |
| T172 | 0.415 | 41.5 | 33.5 |
| T178 | 0.448 | 44.8 | 45.5 |
| T193 | 0.459 | 45.9 | 48.5 |
| T199 | 0.470 | 47.0 | 50.5 |
| T152 | 0.529 | 52.9 | 56.0 |
| T177 | 0.531 | 53.1 | 51.5 |
| T197 | 0.544 | 54.4 | 53.5 |

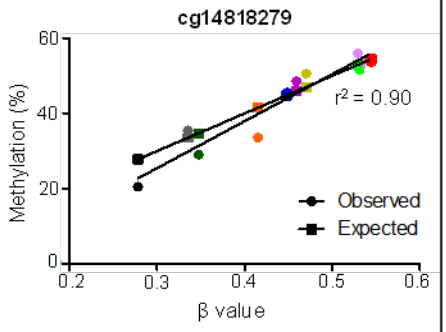



**Figure 5.6**

Validation of pyrosequencing assays used for methylation quantification in hip cartilage at CpG sites cg03278514 (A) and cg14818279 (B). gDNA from NOF patient cartilage samples analysed on the high-density methylation array were bisulphite converted and the region surrounding the selected CpG site amplified by PCR. The average observed methylation levels as determined by the pyrosequencer, were then compared to the expected methylation β values from the high-density methylation array. The table lists the patient sample, β value from the array, expected methylation value (derived from the β value) and average observed methylation value, as determined by pyrosequencing. The data for both CpG assays is also presented as a graph, with the corresponding expected methylation value and average observed methylation value for each patient represented in the same colour. Statistical analyses confirmed both validations had a positive correlation, with a goodness of fit $r^2 \geq 0.85$.

### 5.3.4 Cluster membership analysis of new hip OA cartilage samples

The pyrosequencing assays validated in Chapter 5.3.3 were subsequently used to determine methylation at cg14818279 and cg03278514 in a cohort of 25 new hip OA cartilage DNA samples, with subsequent application of the CI (Chapter 5.3.2) to determine inflammatory cluster membership (Table 5.2). This identified 15 samples as non-inflammatory (cluster 1), 6 as inflammatory (cluster 2), but was unable to classify 4 of the samples. A scatter plot analysis of the average methylation values at cg14818279 and cg03278514 according to the predicted cluster membership is shown in Figure 5.7.

| Patient | cg14818279 average methylation (%) | cg03278514 average methylation (%) | Predicted cluster according to CI |
|---|---|---|---|
| 4963 | 43 | 50.5 | Cluster 1 |
| 5072 | 29 | 91.5 | Cluster 1 |
| 5152 | 42.5 | 69 | Cluster 1 |
| 5216 | 29 | 83 | Cluster 1 |
| 5268 | 45 | 82 | Cluster 1 |
| 5343 | 61.5 | 37.5 | Cluster 2 |
| 5405 | 33 | 62.5 | Cluster 1 |
| 5430 | 43 | 58.5 | Cluster 1 |
| 5431 | 49 | 58 | Cluster 1 |
| 5483 | 63.5 | 51.5 | Cluster 2 |
| 5667 | 37 | 94 | Cluster 1 |
| 5707 | 49 | 70.5 | Cluster 1 |
| 5749 | 45.5 | 48 | Unknown |
| 5756 | 46 | 48.5 | Unknown |
| 5865 | 47 | 60 | Cluster 1 |
| 5866 | 67.5 | 53.5 | Cluster 2 |
| 5870 | 40 | 92.5 | Cluster 1 |
| 5924 | 68 | 59 | Unknown |
| 5977 | 70 | 40.5 | Cluster 2 |
| 6024 | 64.5 | 51.5 | Cluster 2 |
| 6101 | 24.5 | 49 | Unknown |
| 6109 | 30.5 | 64 | Cluster 1 |
| 6110 | 27.5 | 89 | Cluster 1 |
| 6129 | 40 | 53 | Cluster 1 |
| 6130 | 52 | 50 | Cluster 2 |

**Table 5.2**

Table showing methylation values at cg14818279 and cg03278514 in bisulphite converted OA hip cartilage DNA from 25 new patients, along with cluster membership as predicted using the CI. Cluster 1 denotes the non-inflammatory subgroup and cluster 2 the inflammatory. ‘Unknown’ indicates patients that the CI was unable to classify.

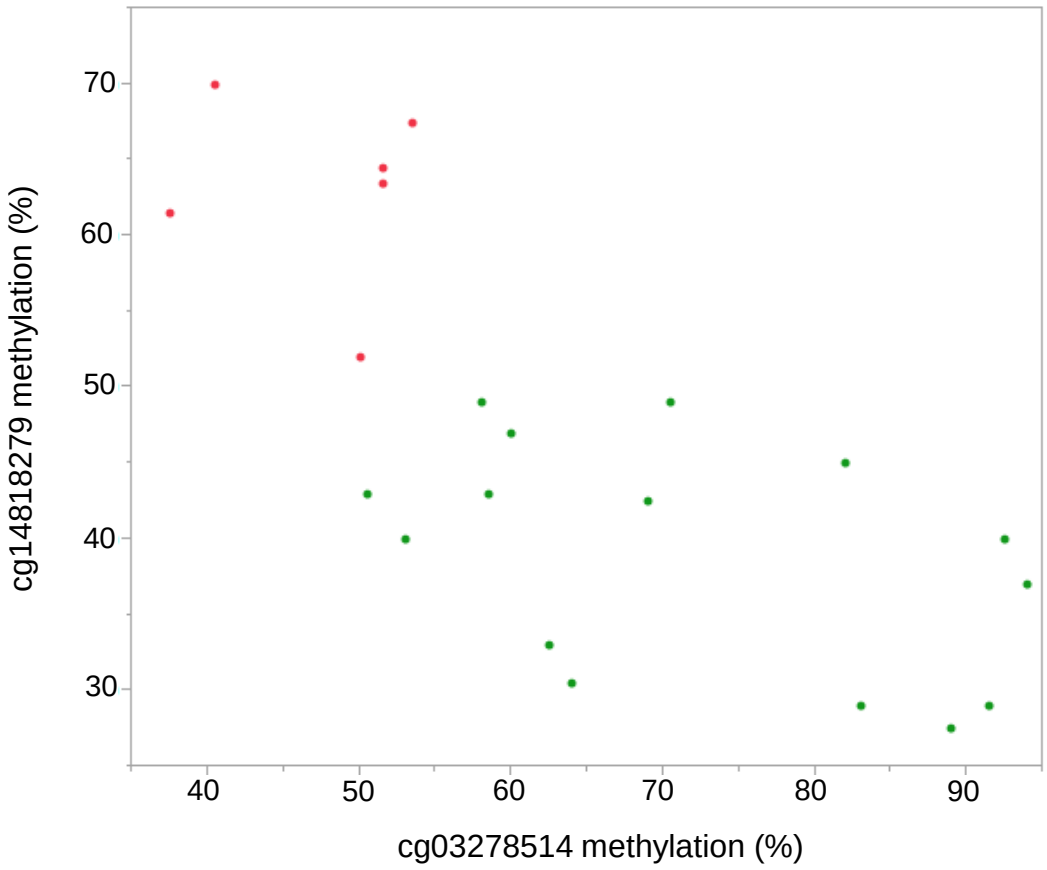



**Figure 5.7**

Average methylation, as determined by pyrosequencing, at cg03278514 and cg14818279 for the new OA hip cartilage patient samples. The 15 samples predicted by the CI to belong to the non-inflammatory cluster (cluster 1) are denoted in green and the six inflammatory samples in red (cluster 2). The y axis indicates average methylation at cg14818279 and the x axis, cg03278514.

### 5.3.5 Differences in inflammatory gene expression between the predicted hip clusters

Having created a discriminant function based on methylation analysis by pyrosequencing to predict cluster membership of new hip cartilage samples, I next sought to test for differences in inflammatory gene expression. I first replicated the inflammatory gene expression qPCR performed by Rushton *et al* (2015) on hip OA patient samples analysed on the array (Chapter 2.17) [121]. This was performed using RNA from the cartilage of 10 of the revised cohort of 17 hip OA patients analysed on the array, as RNA was only available from these individuals. (Figure 5.8). Primer sequences for the assays are listed in Appendix A: Table A.6.

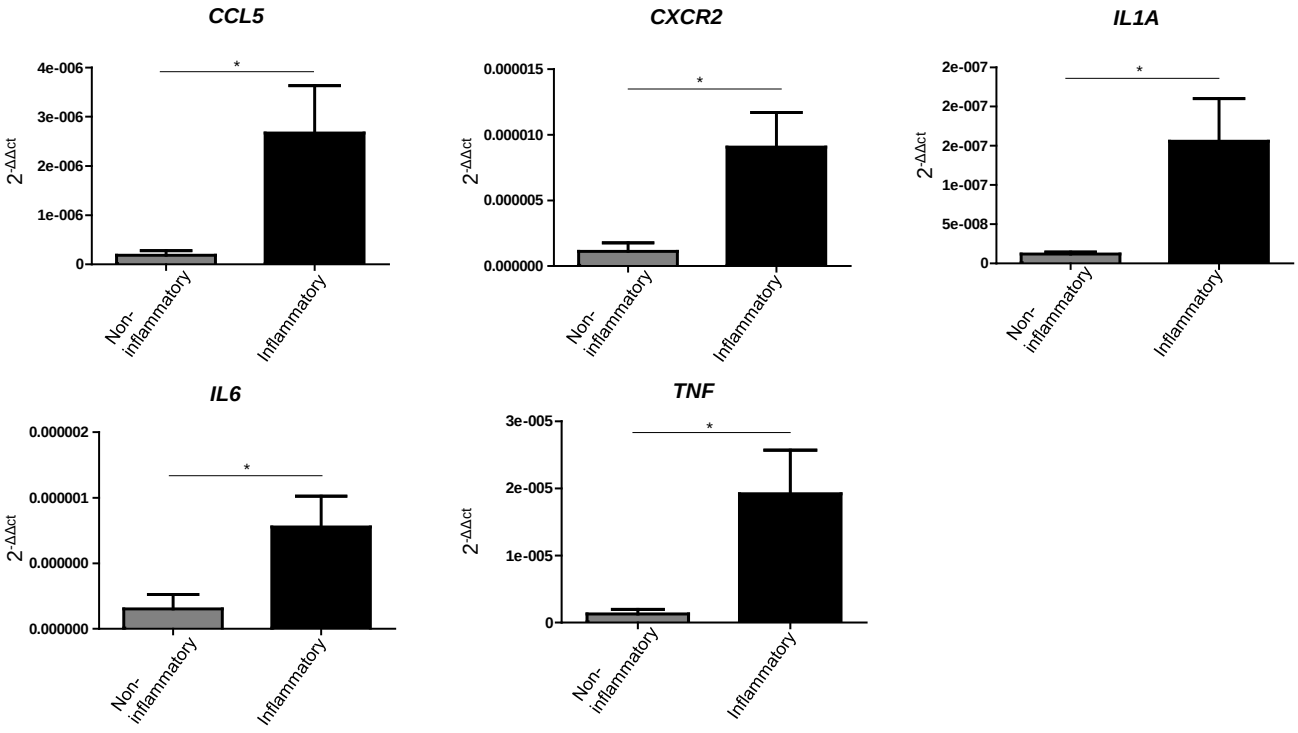



**Figure 5.8**

Expression of inflammation-associated genes in hip OA cartilage in five non-inflammatory cluster patients and five inflammatory cluster patients who were analysed on the high-density methylation array. Gene expression was measured by qPCR in five genes previously shown to display differential expression between the clusters [119]. Data are presented as $2^{-\Delta\Delta Ct}$, having been normalised to the *GAPDH* housekeeper gene. Error bars represent the SEM. Statistical significance was assessed using the Student's *t* test. *p<0.05.

Subsequently, expression of these genes was analysed by qPCR using RNA extracted from the cartilage of the 21 new hip OA patients, who had been subgrouped using the CI (Figure 5.9). All five genes demonstrated greater expression in the samples that had been subgrouped into the inflammatory cluster and for *IL1A*, *IL6* and *TNF* the increases were significant. This data therefore supports the capacity of the CI to place hip OA cartilage samples into the non-inflammatory or inflammatory cluster.

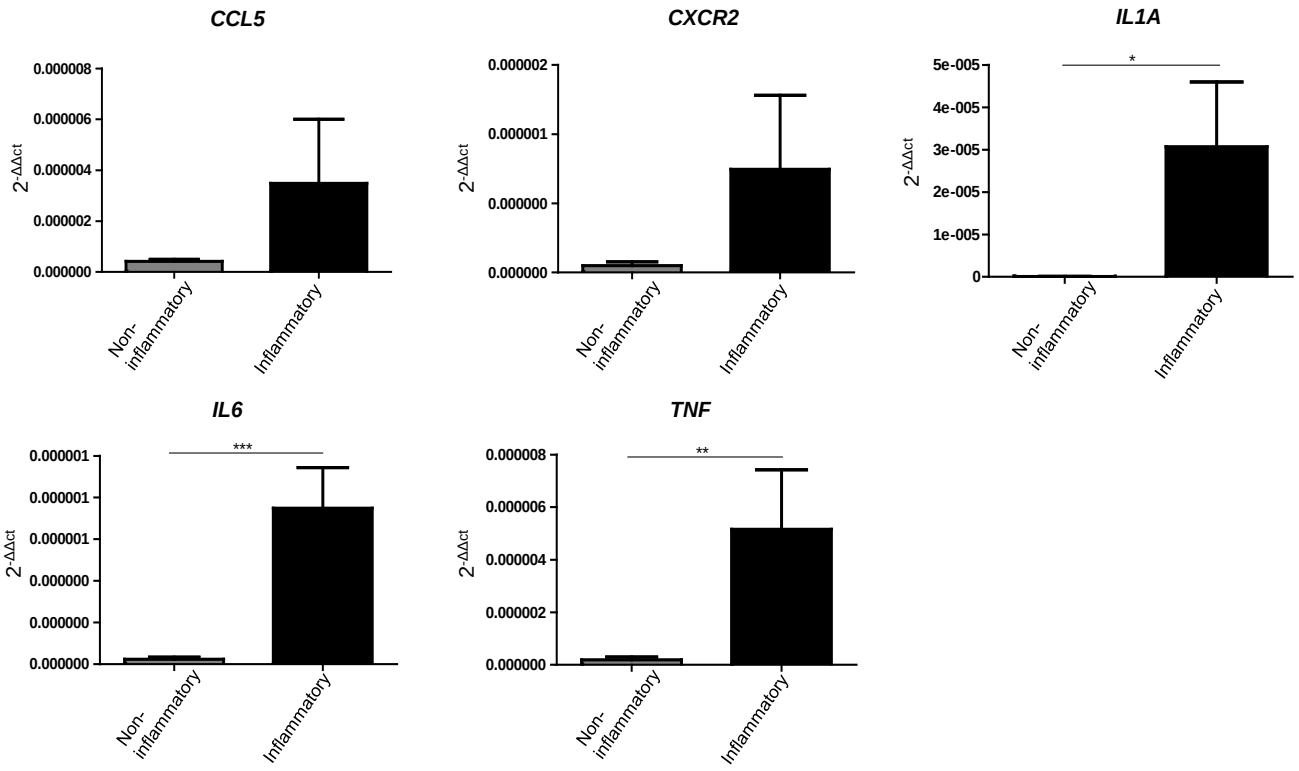



**Figure 5.9**

Expression of inflammation-associated genes in the 21 new hip OA patients who had been categorised into the non-inflammatory or inflammatory cluster using the CI, based on methylation at cg14818279 and cg03278514. Gene expression was measured by qPCR and is presented as $2^{-\Delta\Delta Ct}$, having been normalised to the *GAPDH* housekeeper gene. Error bars represent the SEM. Statistical significance was assessed using the Student's *t* test. *$p$<0.05 ** $p$ < 0.01; *** $p$ < 0.001. Non-inflammatory $n$ = 15, inflammatory $n$ = 6 (where $n$ = number of samples in each comparison).

### 5.3.6 Selection of CpG sites for cluster differentiation in OA knee cartilage

Despite both forming distinct non-inflammatory and inflammatory clusters, the knee and hip cartilage methylomes displayed clear heterogeneity when analysed by high-density CpG methylation array [43]. As such, it was not possible to use the same cluster-differentiating CpG sites in hip and knee cartilage and separate sites had to be identified for subgrouping the knee OA patients. Consequently, CpG sites analysed on the array in 73 knee OA cartilage samples were assessed according to both their *p* values and β values, in order to determine the best candidates to design pyrosequencing assays to. Based on these criteria, cg12245706 and cg21944234 were selected (Table 5.3).

As for our hip samples, subsequent analysis of the array data by our group resulted in 14 of the 73 knee OA samples being removed from the data set. Re-analysis of the data for the 59 remaining samples confirmed the capacity of the selected CpG sites to still discriminate between the inflammatory subgroups (Figure 5.10).

| Probe | Chromosome | Adjusted *p* value | Non-inflammatory cluster mean β value | Inflammatory cluster mean β value |
|---|---|---|---|---|
| cg12245706 | 2 | 4.08E-11 | 0.34 | 0.53 |
| cg21944234 | 3 | 7.01E-11 | 0.29 | 0.48 |

**Table 5.3**

Knee CpG sites selected from the high-density array to take forward for logistic analysis. The data shown was generated by Dr Michael Rushton using the Illumina Infinium HumanMethylation450 BeadChip array and is derived from all 73 of the knee OA patient samples [43]. The *p* values shown are those following Benjamini-Hochberg correction for multiple testing when comparing methylation in the non-inflammatory and inflammatory subgroups.

**A.**

| Probe | Chromosome | *p* value | Non-inflammatory cluster mean β value | Inflammatory cluster mean β value |
|---|---|---|---|---|
| cg12245706 | 2 | < 0.0001 | 0.37 | 0.57 |
| cg21944234 | 3 | < 0.0001 | 0.43 | 0.53 |

**B.**

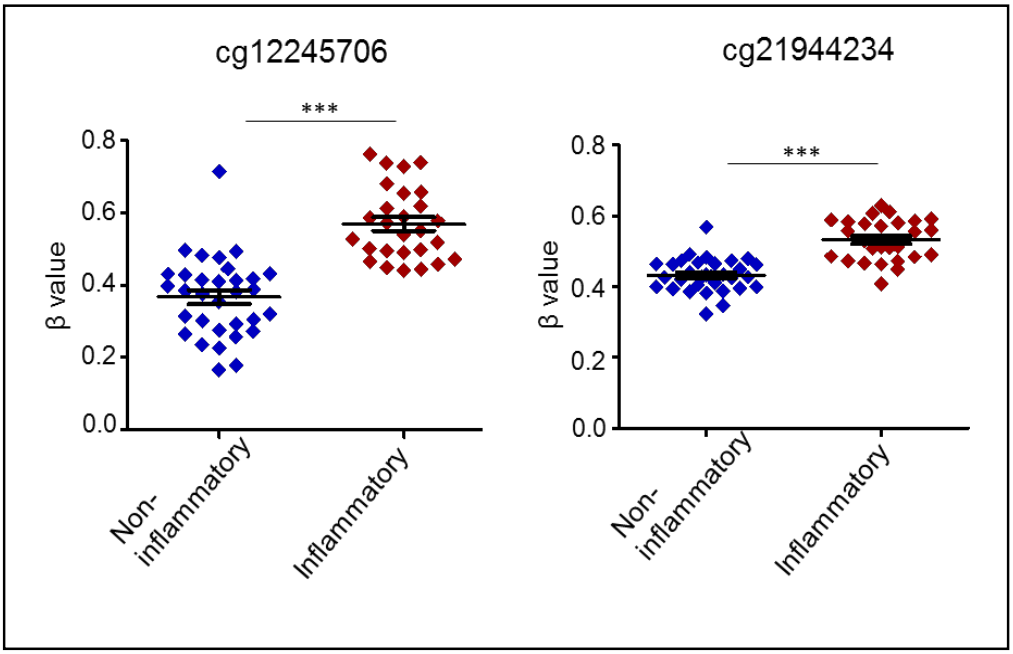



**Figure 5.10**

Analysis of methylation at the selected CpG sites in the revised cohort of 59 knee OA patient samples. The data shown was generated by Dr Michael Rushton using the Illumina Infinium HumanMethylation450 BeadChip array [43]. The mean cluster β values and *p* values are shown in panel (A), following the removal of 14 of the original 73 patients. The β values for the 59 patients are plotted in panel (B), according to the inflammatory subgroup, for both CpG sites. Each data point represents a single patient sample; horizontal lines and error bars show the mean ± SEM. Statistical significance was assessed using a Mann-Whitney *U* test. *** $p \leq 0.001$. Non-inflammatory $n = 32$, inflammatory $n = 27$ (where $n$ = number of patient samples in each comparison).

### 5.3.7 Logistic regression modelling to predict knee cartilage inflammatory cluster membership

Logistic regression analyses were performed with the two selected CpG sites, based on the β values for the 59 samples analysed on the high-density methylation array, in order to assess their capacity to predict inflammatory cluster membership in knee OA cartilage. Statistical analyses were performed with JMP Pro Version 11.2 Statistical Data Visualization Software with the assistance of Dr Dennis Lendrem (Institute of Cellular Medicine).

Stepwise discriminant analysis identified both cg12245706 and cg21944234 to successfully predict cluster membership to a statistically significant degree ($p < 0.05$) (Figure 5.11). A linear combination of both cg12245706 and cg21944234 demonstrates separation of the inflammatory clusters according to methylation at these sites, though the separation is not total and a degree of overlap is apparent (Figure 5.12).

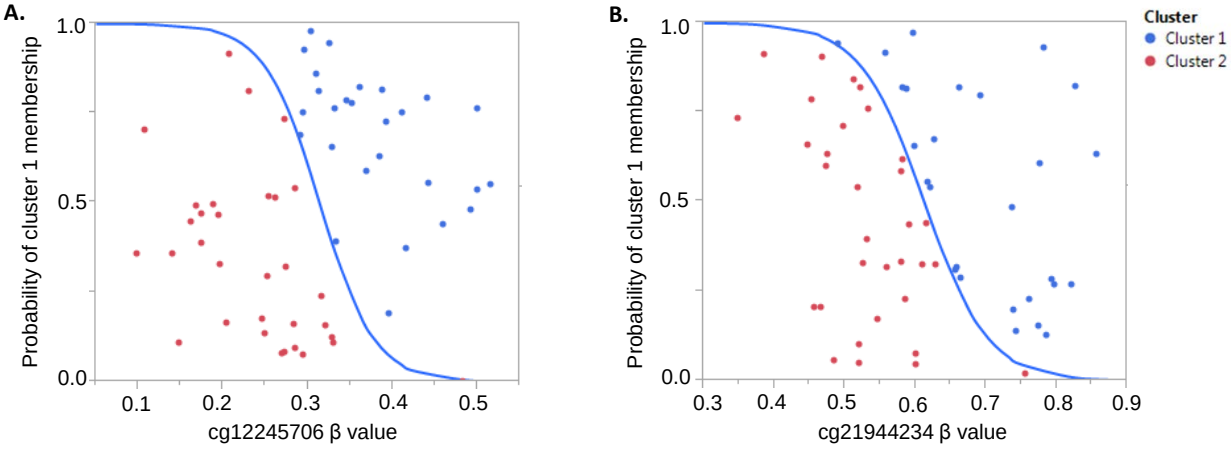



**Figure 5.11**

Logistic regression plots for the selected CpG sites: cg12245706 (A) and cg21944234 (B). The y axis shows the probability of belonging to cluster 1 (non-inflammatory) according to methylation at cg12245706 (A) or cg21944234 (B), and refers only to the blue regression line, not to the individual data points. Samples are plotted according to the β value at cg12245706 (A) or cg21944234 (B), demarcated on the x axis, as determined by high-density CpG methylation array. Each data point represents a single sample, where red donates non-inflammatory cluster membership (cluster 1; $n$ = 32) and blue the inflammatory cluster samples (cluster 2; $n$ = 27). The logistic regression analyses show that methylation at cg12245706 and cg21944234 both provide good separation of inflammatory and non-inflammatory samples ($p < 0.05$).

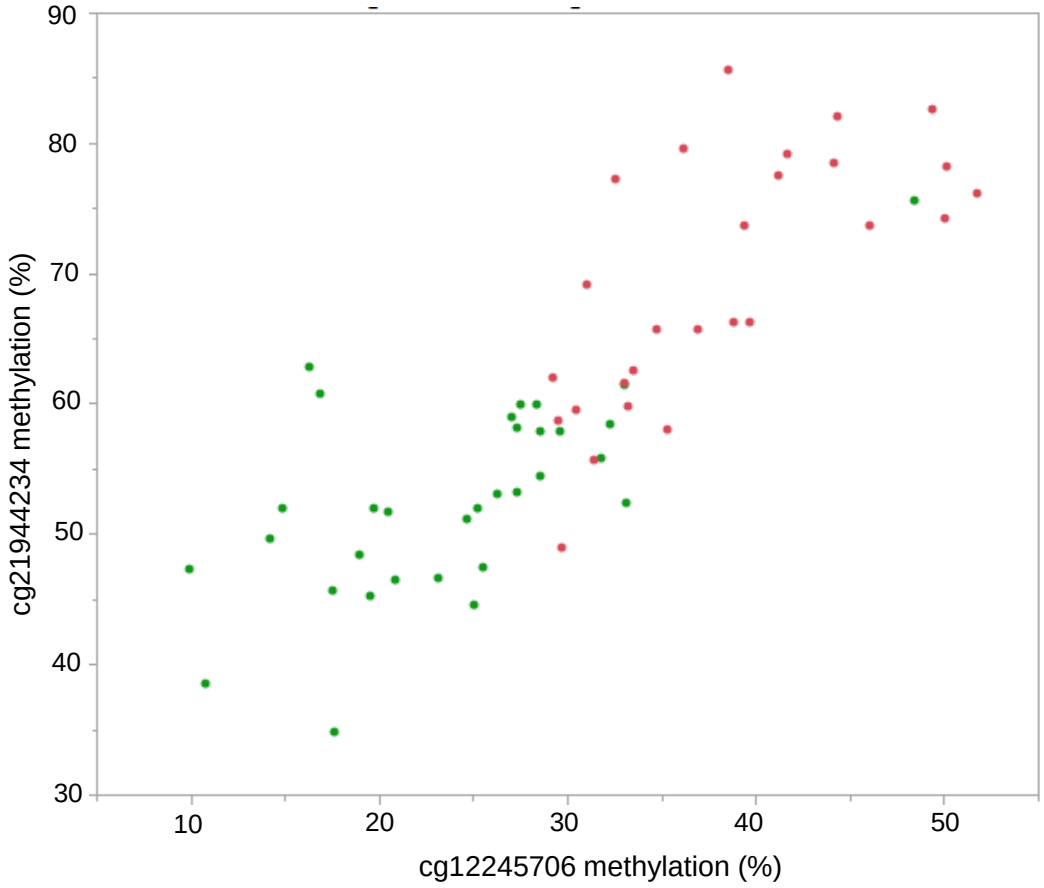



**Figure 5.12**

Knee OA cartilage cluster separation achieved using a linear combination of methylation at cg12245706 and cg21944234. The 59 patients analysed on the high-density CpG methylation array are plotted according to their β values at cg12245706 (x axis) and cg21944234 (y axis). Non-inflammatory samples are denoted in green (cluster 1; *n* = 32) and inflammatory samples in red (cluster 2; *n* = 27).

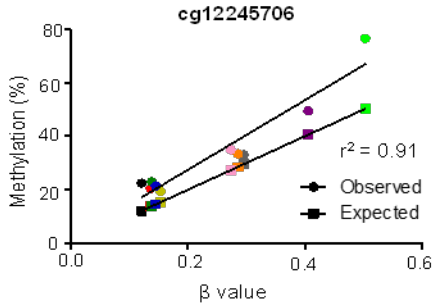
### 5.3.8 Design of pyrosequencing assays to determine methylation at cg12245706 and cg21944234

Having identified two CpGs that can discriminate between the knee OA cartilage inflammatory clusters, the next step was to design pyrosequencing assays to cg12245706 and cg21944234 (Chapter 2.14.1). Primer sequences for the assays can be found in Appendix A: Table A.5.

As discussed when previously designing pyrosequencing assays to the selected hip CpG sites (Chapter 5.3.3), it was necessary to validate the assays to ensure the sequencing primer was able to correctly quantify methylation at the CpG site. Validation was performed, in duplicate, using gDNA extracted from the cartilage of neck of NOF patients analysed on the genome-wide methylation array (Chapter 2.14.2; patients are listed in the tables in Figure 5.13). Both validations had a positive correlation, with $r^2$ values of ≥ 0.91, and were therefore deemed suitable for use.

A.

| Patient | β value | Expected methylation (%) | Average observed methylation (%) |
|---|---|---|---|
| T151 | 0.122 | 12.2 | 22.5 |
| T178 | 0.137 | 13.7 | 20.5 |
| T193 | 0.139 | 13.9 | 23 |
| T184 | 0.146 | 14.6 | 21.5 |
| T150 | 0.154 | 15.4 | 19.5 |
| T152 | 0.274 | 27.4 | 33 |
| T197 | 0.286 | 28.6 | 33.5 |
| T191 | 0.296 | 29.6 | 35 |
| T168 | 0.406 | 40.6 | 49.5 |
| T100 | 0.503 | 50.3 | 76.5 |




B.

| Patient | β value | Expected methylation (%) | Average observed methylation (%) |
|---|---|---|---|
| T151 | 0.335 | 33.5 | 29.5 |
| T193 | 0.343 | 34.3 | 31.5 |
| T150 | 0.381 | 38.1 | 33.5 |
| T117 | 0.409 | 40.9 | 34 |
| T178 | 0.415 | 41.5 | 33.5 |
| T184 | 0.417 | 41.7 | 33.5 |
| T192 | 0.442 | 44.2 | 34 |
| T177 | 0.451 | 45.1 | 34.5 |
| T168 | 0.508 | 50.8 | 36.5 |
| T100 | 0.549 | 54.9 | 39.5 |

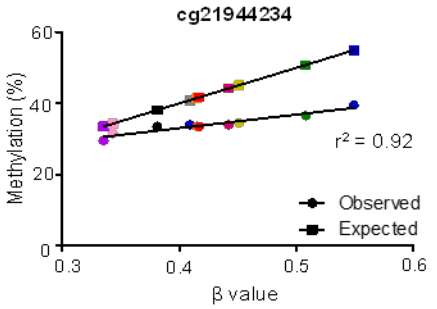



**Figure 5.13**

Validation of pyrosequencing assays used for methylation quantification in knee cartilage at CpG sites cg12245706 (A) and cg21944234 (B). gDNA from NOF patient cartilage samples analysed on the high-density methylation array were bisulphite converted and the region surrounding the selected CpG site amplified by PCR. The average observed methylation levels as determined by the pyrosequencer (observed) were then compared to the expected methylation β values from the high-density methylation array (expected). The table lists the patient sample, β value from the array, expected methylation value (derived from the β value) and average observed methylation value, as determined by pyrosequencing, for cg12245706 (A) and cg21944234 (B). The data for both CpG assays is also presented as a graph, with the corresponding expected methylation value and average observed methylation value for each patient represented in the same colour. Statistical analyses confirmed both validations had a positive correlation, with a goodness of fit $r^2 \geq 0.91$.

### 5.3.9 Batch correction of knee cartilage β values and logistic regression modelling to predict cluster membership

The pyrosequencing assays designed to determine methylation at cg12245706 and cg21944234 both demonstrated positive correlations between the array methylation β value and observed methylation. However, the observed average methylation values, as determined by the pyrosequencer, differed to the expected values from the array, and these differences were greater at higher β values (Figure 5.13). Accordingly, a batch correction was performed from a regression of the difference in observed and expected values as a function of the β values, for both cg12245706 and cg21944234 in the 59 knee samples (Chapter 2.14.3). This permits batch correction of methylation values as determined by the array prior to using the prediction model to assign cluster membership for new samples analysed by pyrosequencing.

After batch correction, a cluster prediction formula was generated from the methylation at cg12245706 and cg21944234 in the 59 knee OA samples analysed on the array using the JMP software (JMP, Version 11, SAS Institute Inc., Cary, NC). Briefly, the software creates a weighted function minimizing the unexplained variability assuming a logistic relationship between the probability of cluster membership based on methylation at cg12245706 and cg21944234. This value is subsequently used to generate a probability for membership to cluster 1 and for membership to cluster 2. When applied to the 59 batch corrected knee samples analysed on the array, the prediction formula correctly predicted cluster membership in 49 of the 59 knee samples on the array (Figure 5.14).

**Contingency table**

| | Most likely cluster | | |
|---|---|---|---|
| Count | Cluster 1 | Cluster 2 | |
| Cluster 1 | 31 | 1 | 32 |
| Cluster 2 | 9 | 18 | 27 |

**Figure 5.14**

JMP screenshot of cluster membership predictions for knee OA patients analysed on the high-density CpG methylation array, according to the prediction formula created using JMP (JMP, Version 11, SAS Institute Inc., Cary, NC). This correctly identified the cluster of 31/32 non-inflammatory patients (cluster 1), and 18/27 inflammatory patients (cluster 2).

### 5.3.10 Cluster membership analysis of new knee OA cartilage samples

The pyrosequencing assays validated in Chapter 5.3.8 were subsequently used to determine methylation at cg12245706 and cg21944234 in a cohort of 24 new knee cartilage DNA samples, with subsequent cluster prediction analysis using the formula created in Chapter 5.3.9. This identified 14 samples as non-inflammatory (cluster 1) and 10 as inflammatory (cluster 2), as shown in Table 5.4. A scatter plot analysis of the average methylation values at cg12245706 and cg21944234 according to the predicted cluster membership is shown in Figure 5.15.

| Sample | cg12245706 average methylation (%) | cg21944234 average methylation (%) | Predicted cluster |
|---|---|---|---|
| 4851 | 14 | 16.5 | Cluster 1 |
| 4943 | 41.5 | 24.5 | Cluster 1 |
| 4969 | 67 | 52 | Cluster 2 |
| 5003 | 50.5 | 36 | Cluster 2 |
| 5007 | 48.5 | 40.5 | Cluster 2 |
| 5012 | 47 | 37 | Cluster 2 |
| 5057 | 59 | 38.5 | Cluster 2 |
| 5066 | 34 | 33.5 | Cluster 1 |
| 5101 | 59 | 42.5 | Cluster 2 |
| 5139 | 25 | 21 | Cluster 1 |
| 5140 | 46 | 34.5 | Cluster 2 |
| 5146 | 36 | 27 | Cluster 1 |
| 5166 | 50.5 | 35.5 | Cluster 2 |
| 5272 | 46 | 37 | Cluster 2 |
| 5654 | 26 | 21.5 | Cluster 1 |
| 5846 | 40 | 11.5 | Cluster 1 |
| 5854 | 42.5 | 17 | Cluster 1 |
| 6014 | 43.5 | 18 | Cluster 1 |
| 6018 | 44.5 | 15.5 | Cluster 1 |
| 6052 | 41.5 | 16.5 | Cluster 1 |
| 6062 | 20 | 17 | Cluster 1 |
| 6111 | 41.5 | 15.5 | Cluster 1 |
| 6125 | 55 | 42 | Cluster 2 |
| 6135 | 38 | 32.5 | Cluster 1 |

**Table 5.4**

Table showing methylation values at cg12245706 and cg21944234 in bisulphite converted OA knee cartilage DNA from 24 new patients, along with cluster membership as predicted using the prediction model created using JMP (JMP, Version 11, SAS Institute Inc., Cary, NC). Cluster 1 denotes the non-inflammatory subgroup and cluster 2 the inflammatory.

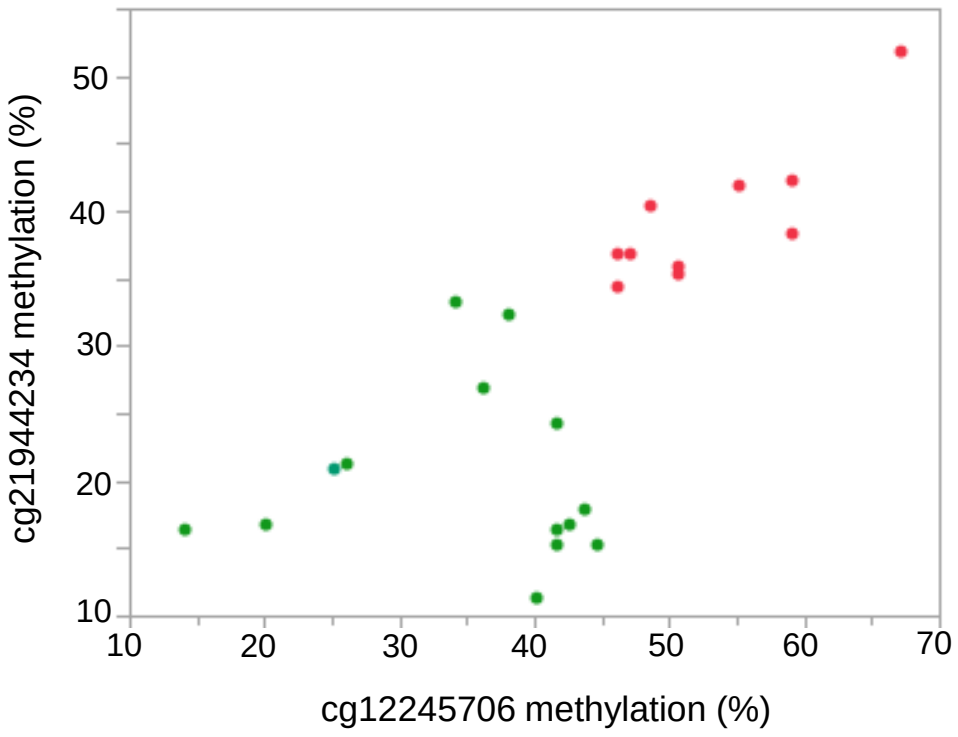



**Figure 5.15**

Average methylation, as determined by pyrosequencing, at cg12245706 and cg21944234 for the 24 new OA knee cartilage patient samples. The 14 samples predicted by the JMP prediction model to belong to the non-inflammatory cluster (cluster 1) are denoted in green and the 10 inflammatory samples in red (cluster 2). The y axis indicates average methylation at cg21944234 and the x axis, cg12245706.

### 5.3.11 Differences in inflammatory gene expression between the predicted knee clusters

As noted earlier, Rushton *et al*. concluded that the hip cartilage methylome identifies an inflammatory subgroup of individuals with hip OA, who also display concurrent differential cartilage expression of inflammatory genes [121]. It was therefore of interest to investigate whether differential inflammatory gene expression was also present in OA knee, as this assessment had not been performed by Rushton *et al*. The expression of the inflammatory genes analysed in hip OA cartilage (Chapter 5.3.5) were measured by qPCR in available cDNA from 16 non-inflammatory and 11 inflammatory patients from the revised cohort of 59 who had been analysed on the high-density array (Chapter 2.17). For primer sequences see Appendix A: Table A.6. *CXCR2* and *IL1A* displayed significantly higher expression in the inflammatory cluster patients, while expression of *CCL5*, *IL1B*, *IL6* and *TNF* was not significantly different between the two clusters (Figure 5.16).

Analysis of the new knee OA cartilage samples listed in Table 5.4 did not show any significant differences between the subgroups in expression of any of the genes analysed, with a large degree of inter-individual variability in the inflammatory group (Figure 5.17). However, a trend was seen toward higher expression of all genes analysed in the predicted inflammatory samples when compared to the predicted non-inflammatory samples. This indicates the inflammatory transcriptome is not as consistently altered in OA knee patients as in OA hip patients, thus re-emphasising the disparity between knee and hip OA cartilage pathology. The considerable inter-individual variability within the knee cartilage may also reflect the heterogeneous nature of the pathophysiological changes occurring.

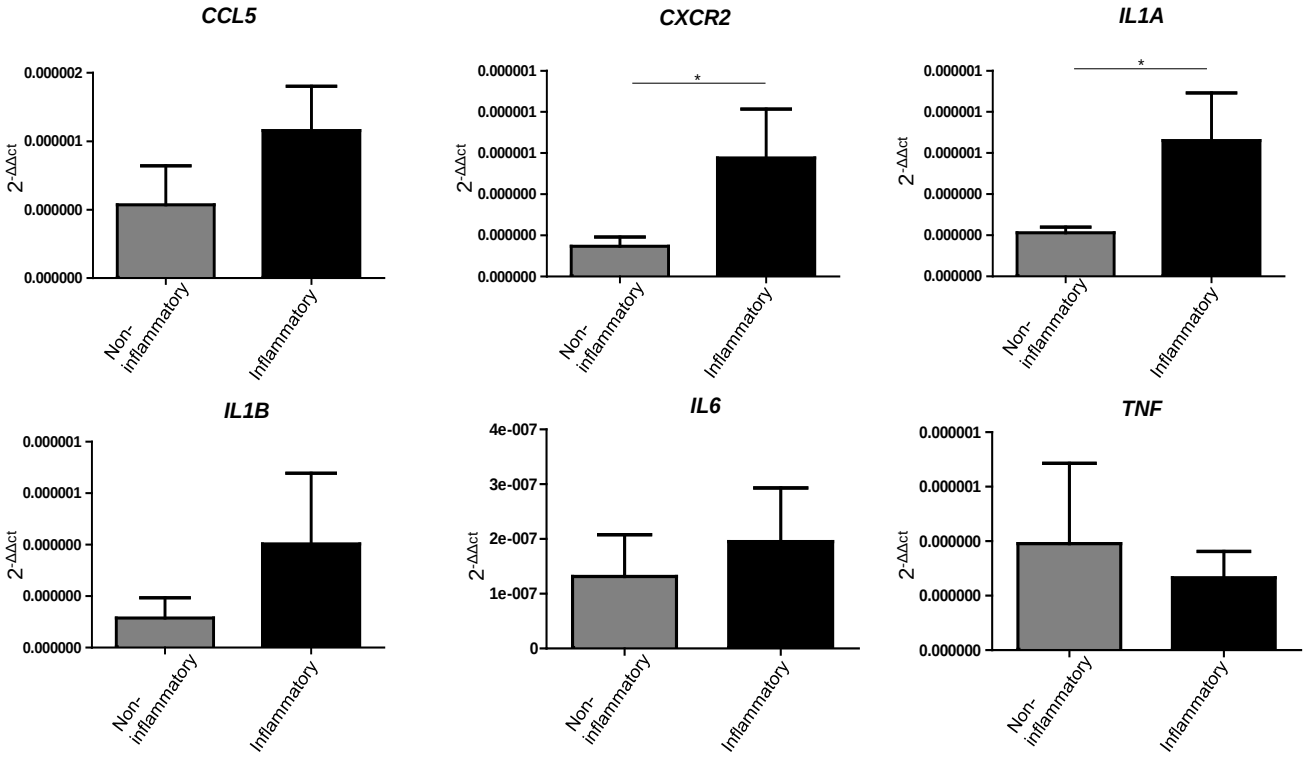



**Figure 5.16**

Expression of inflammation-associated genes in knee OA cartilage in 16 non-inflammatory cluster patients and 11 inflammatory cluster patients who were analysed on the high-density methylation array. Data are presented as $2^{-\Delta\Delta Ct}$, having been normalised to the *GAPDH* housekeeper gene. Error bars represent the SEM. Statistical significance was assessed using the Student's *t* test. *p<0.05.

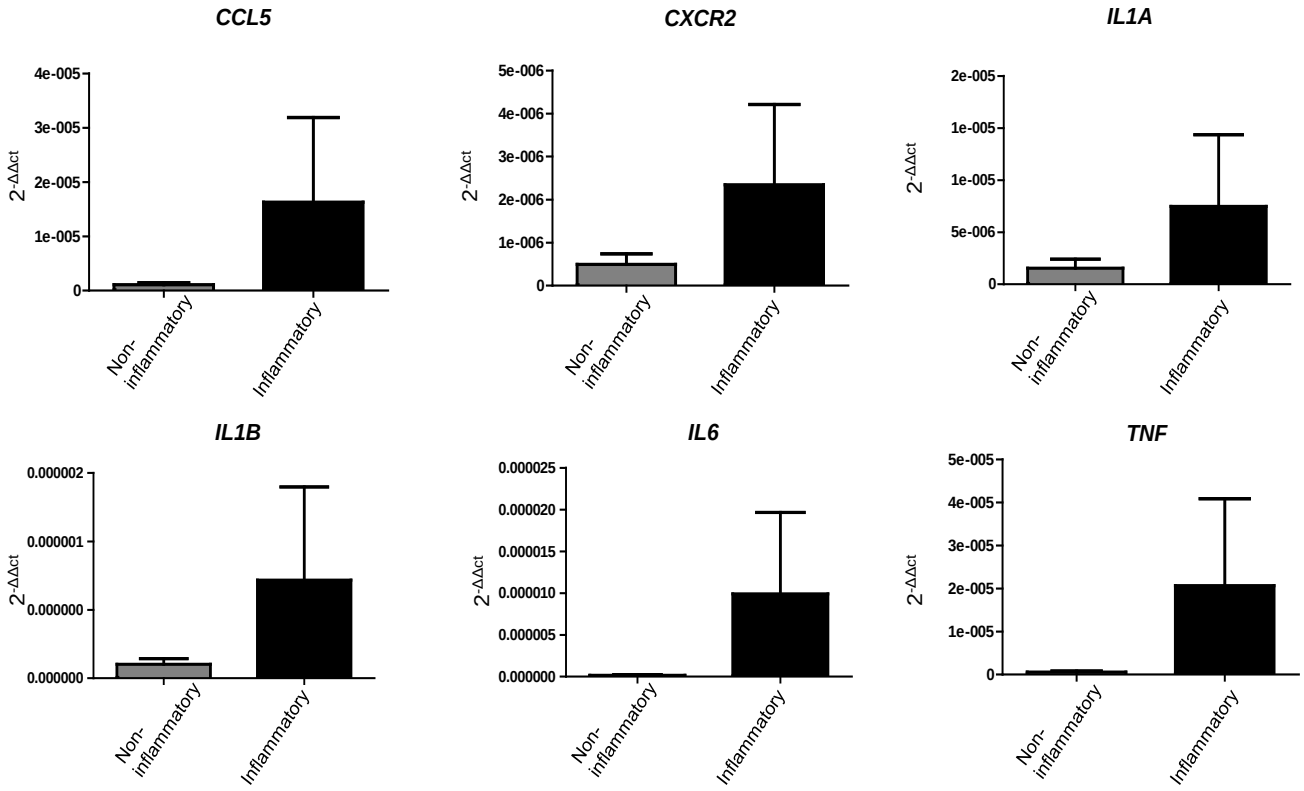



**Figure 5.17**

Expression of inflammation-associated genes in the 24 new knee OA patients who had been categorised into the non-inflammatory or inflammatory cluster, based on methylation at cg12245706 and cg21944234. Gene expression was measured by qPCR and is presented as $2^{-\Delta\Delta Ct}$, having been normalised to the *GAPDH* housekeeper gene. Error bars represent the SEM. Statistical significance was assessed using the Student's *t* test. Non-inflammatory *n* = 14, inflammatory *n* = 10 (where *n* = number of samples in each comparison).

## 5.4 Discussion

The overarching aim of this chapter was to generate a method to enable the classification of new hip and knee OA patient samples into the non-inflammatory or inflammatory subgroups previously identified by our group, without the need for a genome-wide methylation array. This was to be achieved by identifying CpG sites that in the original array analysis, showed a significant difference in methylation between the two hip subgroups and the two knee subgroups, followed by logistic analyses to create models predicting cluster membership based on methylation at these sites. Subsequent design, and validation, of pyrosequencing assays to the selected CpG sites would then enable for methylation analysis and clustering of new cartilage samples.

I first selected CpG sites which displayed differential methylation between the non-inflammatory and inflammatory groups in the high-density array and performed logistic analyses in order to assess the capacity of methylation at the sites to predict cluster membership. This was performed separately for the hip and knee OA patients. A heuristic cluster index (CI) using methylation at cg14818279 and cg03278514 was formulated to identify cluster membership in hip OA patients. Owing to the knee subgroups separating less well according to methylation, a prediction formula generated using statistical software was used to assign samples to a cluster according to methylation at cg12245706 and cg21944234. Pyrosequencing assays were designed to the selected CpG sites and then validated using bisulphite converted DNA from samples present on the array. Following validation, the assays were used on a new cohorts of hip and knee OA patient

cartilage DNA and the samples were then assigned to the non-inflammatory or inflammatory subcluster by the CI or prediction formula respectively.

I then went on to analyse overall inflammatory gene expression in OA hip cartilage samples. Firstly, I confirmed that there was significantly higher expression of the inflammatory genes *CCL5*, *CXCR2*, *IL1A*, *IL6* and *TNF* in five inflammatory OA cartilage samples which had been analysed on the array, when compared to five non-inflammatory samples, as previously reported by Rushton *et al*. [121]. Inflammatory gene expression analysis was then performed in 21 new hip OA patient cartilage samples, subgrouped using the CI, in order to help determine whether the discriminatory function had correctly identified the inflammatory samples. This confirmed significantly increased inflammatory gene expression of *IL1A*, *IL6* and *TNF* in the six samples categorised as inflammatory when compared to the 15 samples categorised as non-inflammatory. Though not statistically significant, this was accompanied by elevated expression of *CCL5* and *CXCR2* in the inflammatory samples. There was a degree of inter-individual expression variability in the inflammatory cohort, which may explain why statistical significance was not achieved for all five genes analysed. However, the increase in inflammatory gene expression seen in the new samples acts to further validate the cluster index and its capacity to accurately determine cluster membership.

Subsequently, I performed the same inflammatory gene expression analysis in knee OA samples. Analysis of the knee cartilage transcriptome according to inflammatory subgroup had not previously been performed. I initially ascertained whether the

divergent knee cartilage methylome was translated into an altered inflammatory transcriptome, as had been demonstrated in hip cartilage. Both *CXCR2* and *IL1A* displayed significantly increased expression in the inflammatory patient group when analysed in 16 non-inflammatory and 11 inflammatory patients who were studied on the high-density methylation array. However, a statistically significant difference in expression was not seen for *CCL5*, *IL1B*, *IL6* or *TNF*. Analysis of the new patient samples did not identify any significant differences in gene expression, when analysed in 14 samples categorised as non-inflammatory and the 10 categorised as inflammatory. However, in all cases the predicted inflammatory samples demonstrated a trend towards higher inflammatory gene expression when compared to the predicted non-inflammatory samples. Taken together, these data indicate that, although they possess a distinct cartilage inflammatory methylome, knee OA cartilage is not as transcriptionally distinct when comparing the non-inflammatory and inflammatory subgroups, especially in comparison to hip OA.

Overall, this chapter has successfully seen the creation of a methodology to assign new hip or knee OA patients into the non-inflammatory or inflammatory subgroup based on cartilage methylation. It has also emphasised the fact that hip and knee OA cartilage display significant differences in reference to gene methylation and expression, reflecting the heterogeneity of disease pathology occurring within the joints. The capacity to distinguish cluster membership of new patient samples will enable the further interrogation of the subgroups and the upstream factor, or factors, driving the divergent methylomes.

# Chapter 6: Investigation into the systemic inflammation status and metabolic status in individuals displaying increased cartilage inflammation

## 6.1 Introduction

Having developed methylation assays and a discriminant function to categorise new hip and knee OA patients into the inflammatory or non-inflammatory subgroups, I subsequently sought to analyse a new cohort of patients. Cartilage and blood serum was obtained from 19 individuals undergoing hip arthroplasty and 23 knee arthroplasty patients. The patients were designated as either inflammatory or non-inflammatory using the methodology developed in Chapter 5 in order to further analyse these individuals.

The main intention of this chapter was to investigate the systemic inflammatory status of OA patients according to their cartilage inflammatory group. The purpose of this was to determine whether those patients displaying aberrant methylation of inflammatory genes also had higher systemic levels of inflammation. As such, this would assist in our comprehension of the inflammatory group of patients and would suggest local cartilage inflammation may be occurring secondary to elevated systemic inflammation.

We also attempted to investigate the upstream regulator of the disparate cartilage methylomes and inflammatory gene expression profiles. Studies have shown

cultured OA chondrocytes express higher levels of leptin compared to non-OA cartilage and that leptin can stimulate chondrocytes to express increased levels of inflammatory genes such as *TNF*, *IL1B* and *IL6* [146]. Interestingly, increased leptin concentrations in mice are associated with DNA hypomethylation of inflammation-associated genes [147]. Further to this, in a meta-analysis of 13 studies, patients with rheumatoid arthritis, a systemic inflammatory arthritis, were found to have significantly higher circulating leptin expression when compared to healthy controls [148]. Consequently, I intended to analyse the levels of circulating leptin in our inflammatory and non-inflammatory OA patients in order to investigate whether elevated circulating leptin may be influencing the different inflammatory status in the two groups. In addition, patient demographics were collected in order to identify whether there was any association between elevated serum leptin and sex, or BMI. This therefore sought to explore the hypothesis that increased circulating leptin, resultant of increased age, female sex or high BMI, may result in increased systemic inflammation and aberrant inflammatory gene expression in OA cartilage in the inflammatory subgroup of patients.

Further to improving our comprehension of the inflammatory subgroup, the indication that inflammation plays a critical role in OA pathogenesis has important therapeutic implications. Specifically, it is possible that the inflammatory subgroup of patients could benefit from anti-inflammatory drug treatment. It is therefore essential to identify an accessible biomarker, for example in blood, which distinguishes between the subgroups in order to facilitate patient stratification for future clinical trials and diagnostic purposes.

## 6.2 Aims

The aim of this chapter was to explore the inflammatory sub-group of both hip and knee OA patients and investigate systemic inflammation, as well as potential regulators of the distinct inflammatory methylomes. Further to this, this chapter sought to identify circulating biomarkers which could distinguish between the inflammatory subgroups in order to facilitate patient stratification in future clinical trials. These aims were achieved by:

- Categorising new OA hip and knee patients into the inflammatory or non-inflammatory subgroups using the methods developed in Chapter 5.
- Analysis of markers of systemic inflammation, such as CRP, in patient serum and stratification according to inflammatory subgroup.
- Analysis of patient demographic data according to inflammatory subgroup.
- Quantification of circulating serum leptin and stratification by inflammatory subgroup, sex and BMI.

## 6.3 Results

### 6.3.1 Assigning new hip OA patients into the inflammatory or non-inflammatory subgroup according to cartilage methylation

Matched cartilage and blood serum were obtained for 19 patients who had undergone hip arthroplasty (patient details are listed in Appendix B: Table B.3). Cartilage methylation at cg14818279 and cg03278514 were quantified by pyrosequencing, and patients were subsequently assigned to the inflammatory or non-inflammatory cluster using the discriminant function created in Chapter 5 (Chapter 5.3.2). Nine patients were identified as belonging to the non-inflammatory group (cluster 1) and eight to the inflammatory group (cluster 2), while two patients could not be categorised (Table 6.1).

| Patient | cg14818279 average methylation (%) | cg03278514 average methylation (%) | Cluster |
|---|---|---|---|
| 5989 | 60 | 50 | Cluster 2 |
| 5996 | 45 | 60 | Cluster 1 |
| 6011 | 67 | 34 | Cluster 2 |
| 6013 | 41.5 | 62 | Cluster 1 |
| 6027 | 59.5 | 78.5 | Cluster 1 |
| 6031 | 54.5 | 72 | Cluster 1 |
| 6033 | 45.5 | 51 | Cluster 1 |
| 6089 | 59 | 41 | Cluster 2 |
| 6105 | 55 | 38.5 | Cluster 2 |
| 6127 | 59 | 38.5 | Cluster 2 |
| 6132 | 30.5 | 61.5 | Cluster 1 |
| 6136 | 30 | 52.5 | Cluster 1 |
| 6148 | 28 | 56.5 | Cluster 1 |
| 6163 | 45.5 | 49 | Unknown |
| 6166 | 44 | 35 | Unknown |
| 6191 | 54 | 45 | Cluster 2 |
| 6207 | 52 | 51 | Cluster 2 |
| 6226 | 51 | 57.5 | Cluster 2 |
| 6229 | 55 | 62 | Cluster 1 |

**Table 6.1**
Average cartilage methylation at cg14818279 and cg03278514 in 19 OA hip patients and designated inflammatory subgroup, where cluster 1 is the non-inflammatory and cluster 2 the inflammatory group. CpG methylation was analysed by pyrosequencing and statistical analyses were performed with JMP Pro Version 11.2 Statistical Data Visualization Software (SAS Institute Inc., Cary, NC) to determine the subcluster by using the discriminant function described in Chapter 5.3.2.

### 6.3.2 Investigation of systemic markers of inflammation in hip OA patients

Systemic inflammation was assessed by analysis of serum CRP, which is a well-established and clinically used marker of systemic inflammation [149]. This was achieved using the MSD V-PLEX Human CRP Kit, as described in Chapter 2.16. A calibration curve was created using dilutions generated using the Vascular Injury Panel 2 (human) calibrator blend (Mesco Scale Discovery (MSD), Gaithersburg, USA) (Figure 6.1). Having created a calibration curve in order to allow for quantification of unknown samples, patient samples were analysed to ensure the patient samples generated a signal within the limits detected using the calibrator blend, from which the concentration is derived. Patient samples were diluted 1000 fold as recommended by MSD in order to ensure all signals generated were within the detectable limits. The majority of values were recorded above the lower limit of detection (LLOD), which is the lowest concentration of a substance (in this instance CRP) that can be distinguished from the absence of the substance (the blank value) and is derived using the signal strength. It was deemed acceptable to take forward all data, including those falling above and below the limits of detection, given that the data is not being used in a clinical diagnostic context.

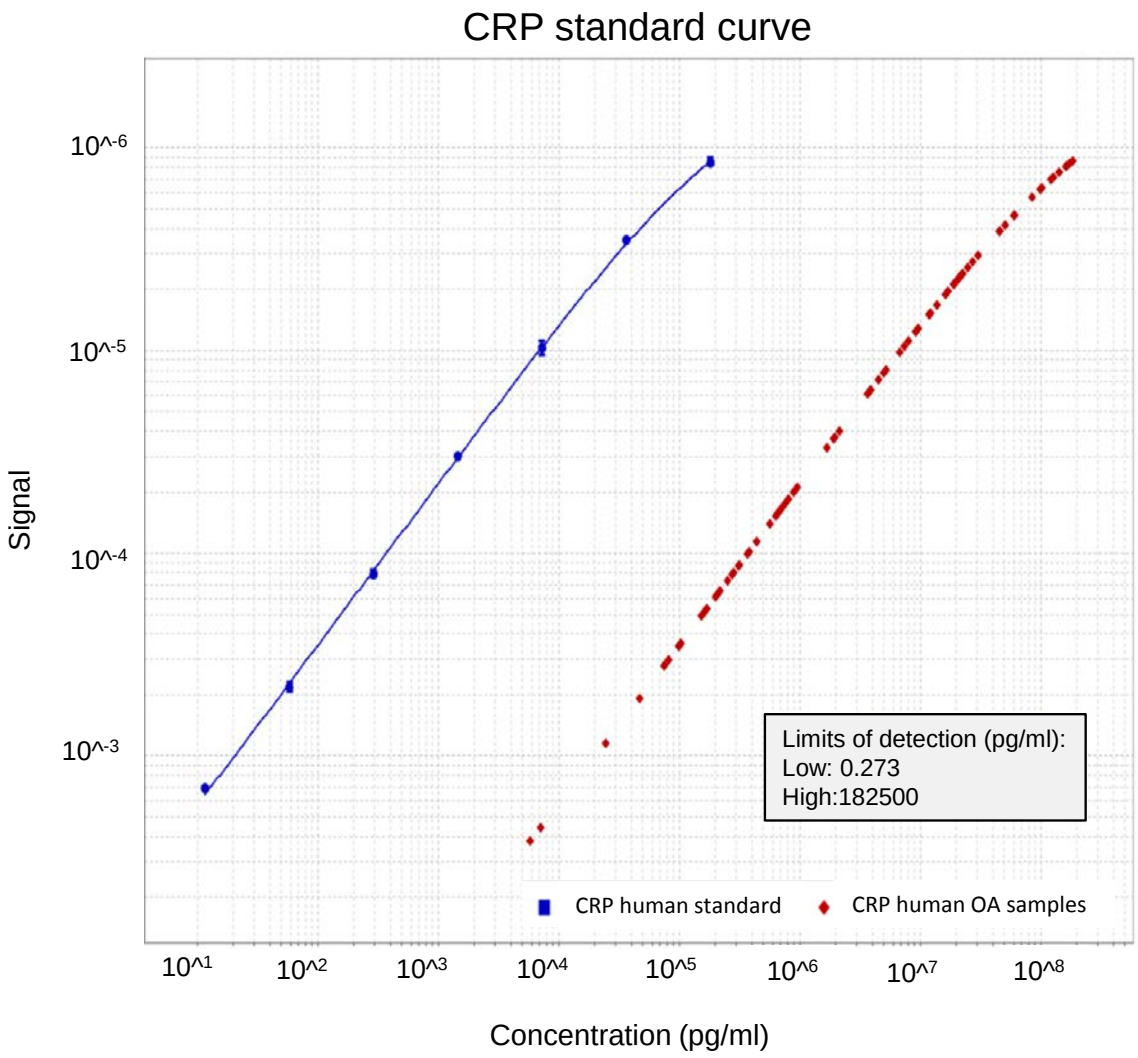



**Figure 6.1**

CRP standard curve generated using the Vascular Injury Panel 2 (human) calibrator blend (MSD, Gaithersburg, USA). Blue squares indicate CRP calibration standards of known concentration, with each standard run in duplicate. OA patient samples, are depicted by the red diamonds and had been diluted 1000 fold, as recommended by MSD to ensure the signal is within the detectable limits. Both knee and hip arthroplasty patients shown on the graph. The concentration (X axis) is derived using the signal strength (y axis).

Analysis of circulating serum CRP levels did not reveal any difference between the non-inflammatory and inflammatory OA patients (Figure 6.2). The data did indicate a number of patients as having a CRP outside of the healthy range (0-10mg/dL), however these patients were present in both in the inflammatory and non-inflammatory subcluster. Given that CRP is a non-specific marker of systemic inflammation, it is not possible to speculate the reason for the elevated readings in these individuals.

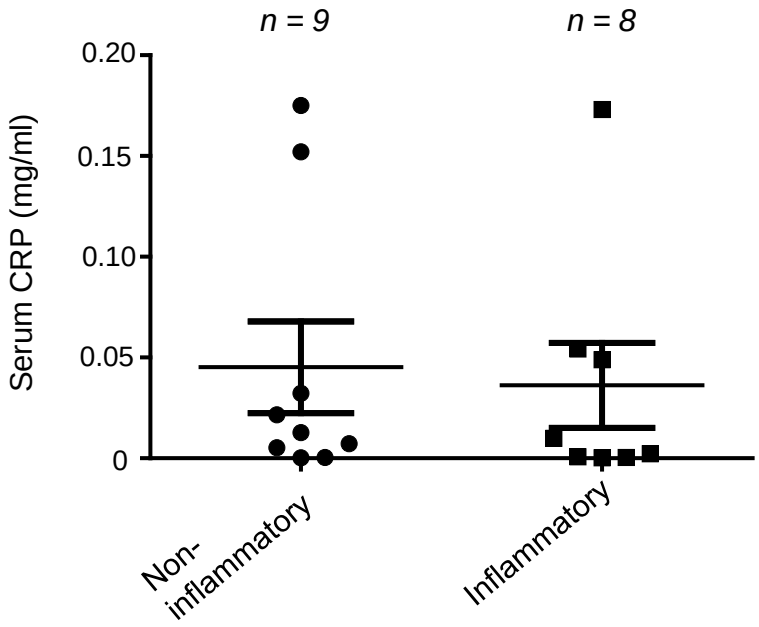



**Figure 6.2**
Analysis of serum CRP levels in hip OA patients, stratified according to inflammatory subcluster. There was no significant difference in serum CRP between the two groups, as assessed using a Mann-Whitney *U* test. *n* represents the number of individuals in the comparison group; error bars represent the mean ± SEM.

Though no difference in serum CRP was seen in hip OA when comparing the patient subgroups, I next investigated circulating levels of inflammatory proteins which had displayed significant differences in gene expression between the groups when analysed in cartilage.

Given that *IL6* displayed the most significant difference in expression in hip cartilage between the non-inflammatory and inflammatory groups (Chapter 5.3.3), circulating IL6 was selected for analysis using the V-PLEX Human Proinflammatory Panel 1 Kit (MSD, Gaithersburg, USA). A calibration curve was generated and all hip OA patient serum samples contained a detectable IL6 concentration, which was recorded within the limits of detection (Figure 6.3).

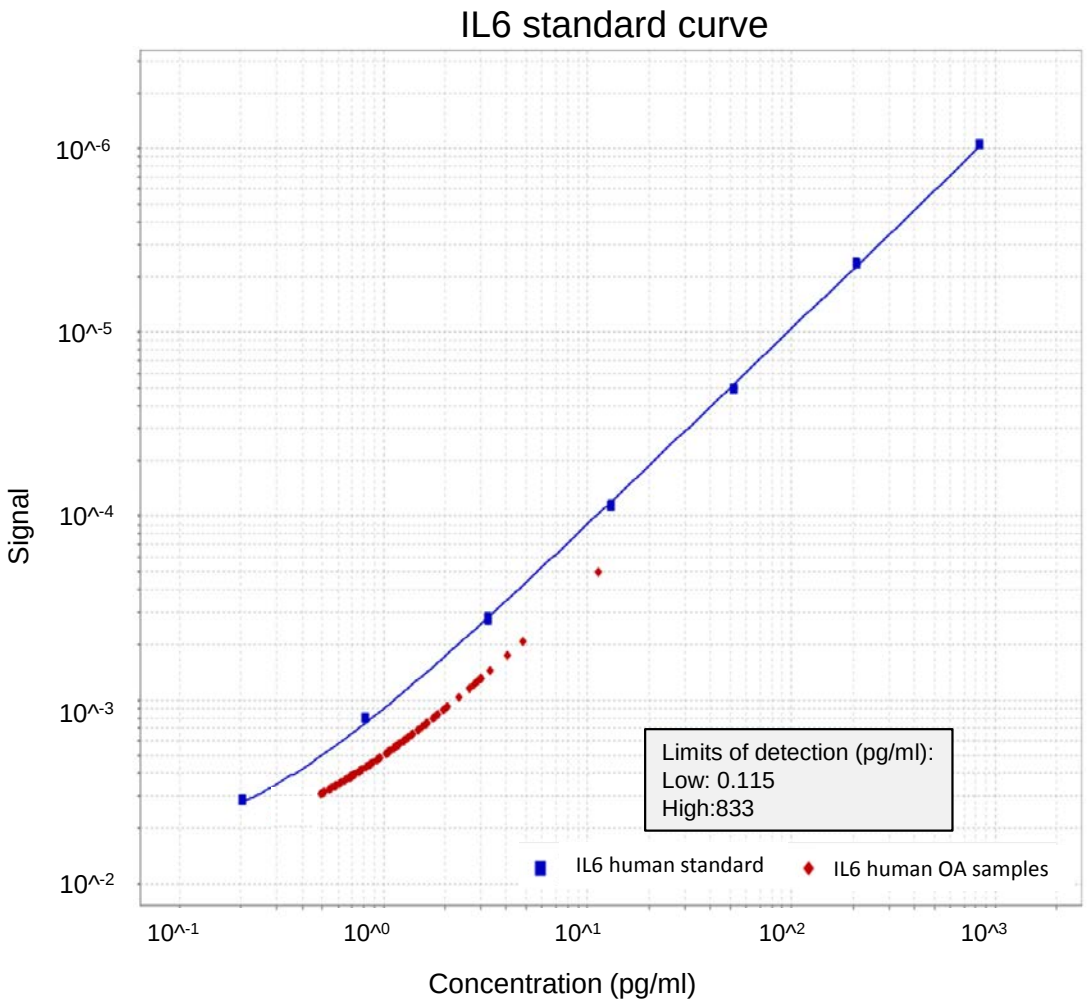



**Figure 6.3**

IL6 standard curve generated using the Proinflammatory Panel 1 (human) calibrator blend (MSD, Gaithersburg, USA). Blue squares indicate IL6 calibration standards of known concentration, with each standard run in duplicate. OA patient samples are depicted by the red diamonds, with both knee and hip arthroplasty patient serum samples shown on the graph. All patient sample values fall within the reference range.

Serum IL6 concentrations displayed no difference between the non-inflammatory and inflammatory groups (Figure 6.4 A). Quantification of *IL6* gene expression in hip cartilage from matched patients, where available, by qPCR (Chapter 2.17). For primer sequences see Appendix A: Table A.6. The gene displayed significantly higher expression in patients in the inflammatory group (Figure 6.4 B).

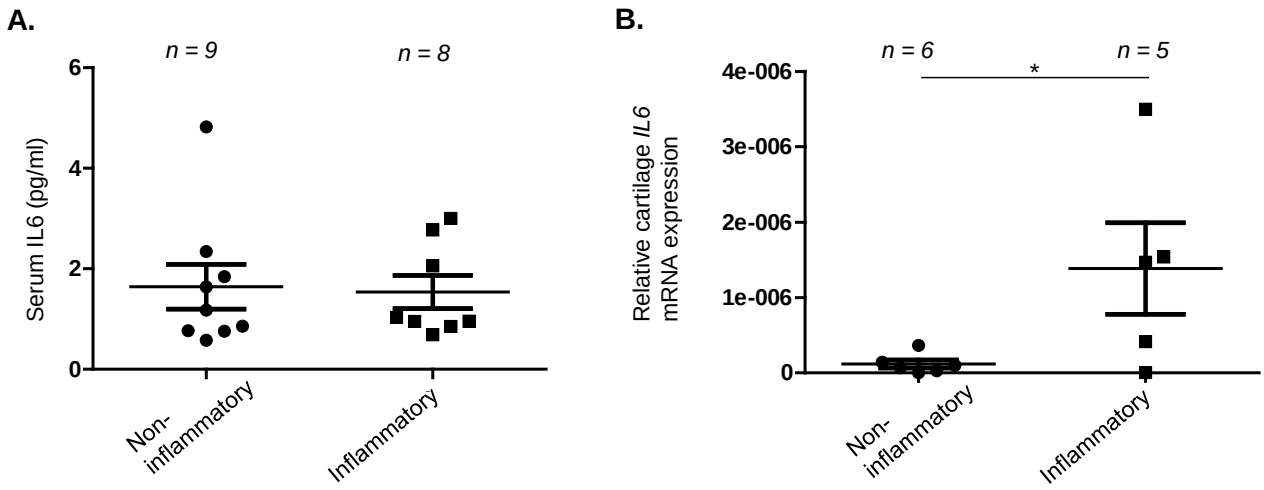



**Figure 6.4**

Analysis of serum IL6 concentration (A) and cartilage *IL6* gene expression in hip OA patients (B). There was no significant difference in circulating IL6 between the non-inflammatory and inflammatory groups (A). *IL6* gene expression was measured by qPCR in hip cartilage samples from matched patients, where nucleic acid was available, and displayed significantly higher expression in patients in the inflammatory group (B). Statistical significance was assessed using a Mann-Whitney *U* test; * $p < 0.05$. *n* represents the number of individuals in the comparison group; error bars represent the mean ± SEM.

TNF-α was selected for analysis as *TNF* had shown a reproducible significant difference in gene expression between the inflammatory and non-inflammatory hip OA patient groups (Chapter 5.3.3). A calibration curve was generated using dilutions of the Proinflammatory Panel 1 (human) calibrator blend (MSD, Gaithersburg, USA) and all analysed patient samples provided values within the calculated limits of detection (Figure 6.5).

Analysis of circulating TNF-α did not reveal any significant difference in concentration when comparing the inflammatory and non-inflammatory hip OA patients (Figure 6.6 A). *TNF* expression was also measured in matched hip cartilage, where nucleic acid was available, by qPCR (Chapter 2.17; for primer sequences see Appendix A: Table A.6). This confirmed, as previously demonstrated, that there was a significant difference in gene expression between the two patient subgroups (Figure 6.6 B).

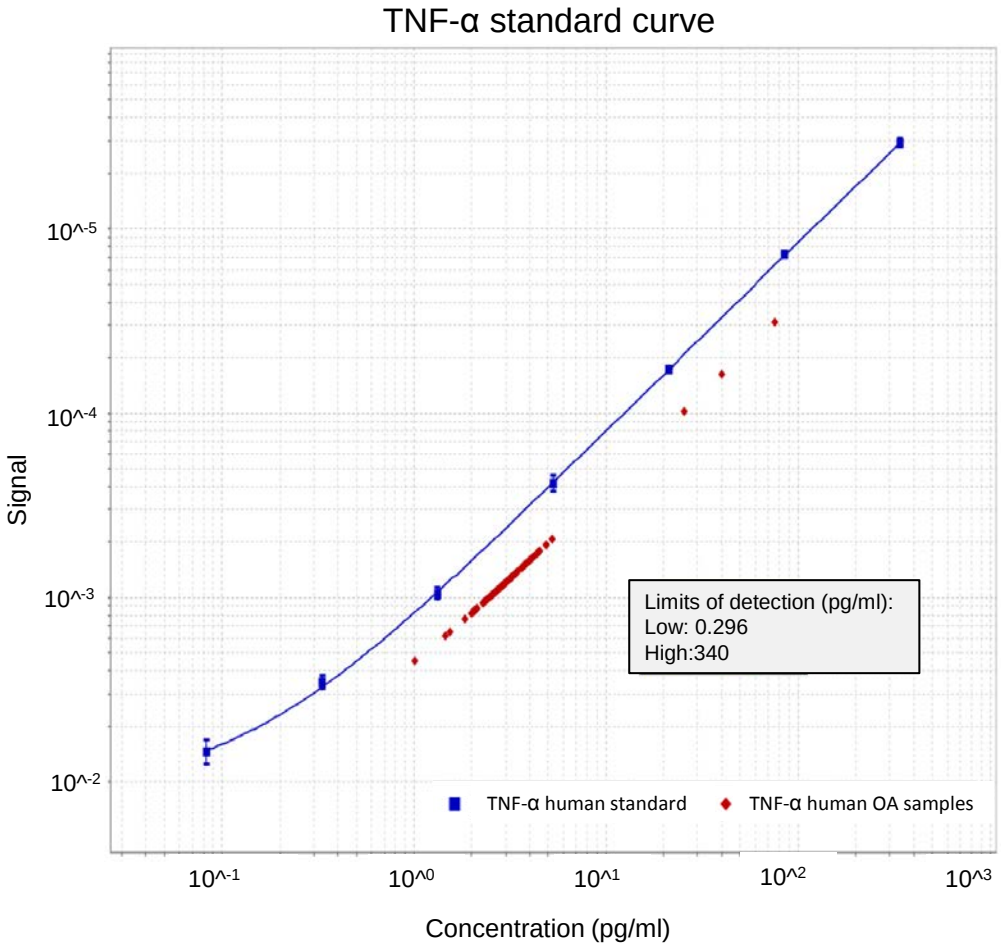



**Figure 6.5**

TNF-α standard curve generated using the Proinflammatory Panel 1 (human) calibrator blend (MSD, Gaithersburg, USA). Blue squares indicate TNF-α calibration standards of known concentration, with each standard run in duplicate. OA patient samples are depicted by the red diamonds, with both knee and hip arthroplasty patients shown on the graph. All patient sample values fall within the reference range.

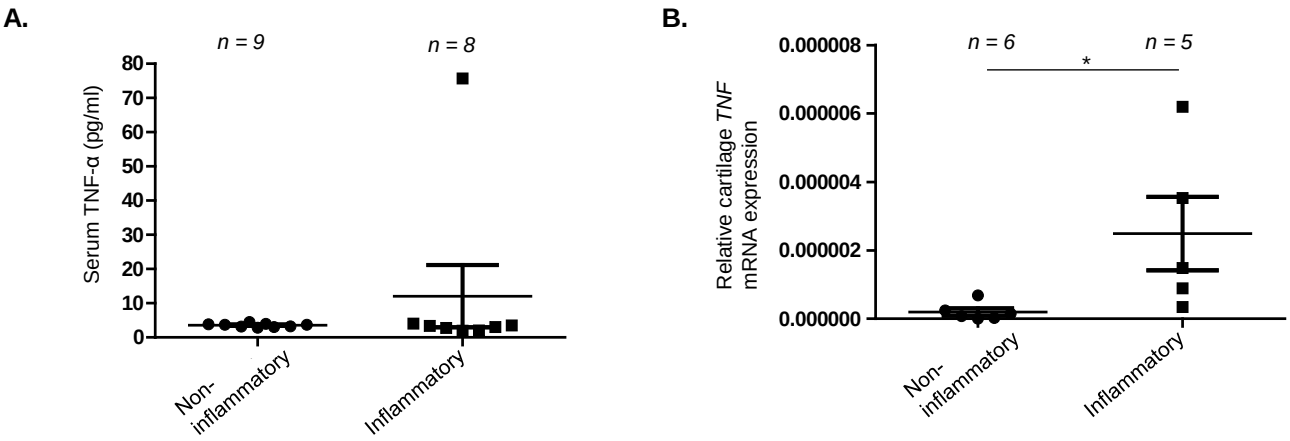



**Figure 6.6**

Analysis of serum TNF-α concentration (A) and cartilage *TNF* gene expression in hip OA patients (B). There was no significant difference in circulating TNF-α between the non-inflammatory and inflammatory groups (A). *TNF* gene expression was measured by qPCR in hip cartilage samples from matched patients, where nucleic acid was available, and displayed a significant difference in expression between the two groups. Statistical significance was assessed using a Mann-Whitney *U* test; * $p < 0.05$. *n* represents the number of individuals in the comparison group; error bars represent the mean ± SEM.

### 6.3.3 Analysis of circulating leptin and BMI in hip OA patients

Having confirmed that changes in the cartilage methylome translate into altered cartilage transcription of inflammatory genes in a subgroup of hip OA patients, we sought to probe the upstream influences driving this altered inflammatory status. The focus of this investigation was BMI and serum leptin. BMI, as a surrogate marker of body fat percentage and so, increased serum leptin, has been shown to have a regulatory role in the inflammatory state of certain patient groups [150].

Serum leptin was measured using the MSD V-PLEX Human Leptin Kit (MSD, Gaithersburg, USA), as detailed in Chapter 2.16 (Figure 6.7). Patient BMI data was recorded prior to arthroplasty surgery and was provided by Mr Kenneth Rankin (Honorary Consultant, Northern Institute for Cancer Research). Neither circulating leptin nor BMI showed a significant difference between the two patient subgroups (Figure 6.8). Given that BMI varies between males and females, it was not appropriate to further stratify data by sex owing to low patient numbers ($n$ = 2 for non-inflammatory males, $n$ = 1 for inflammatory females).

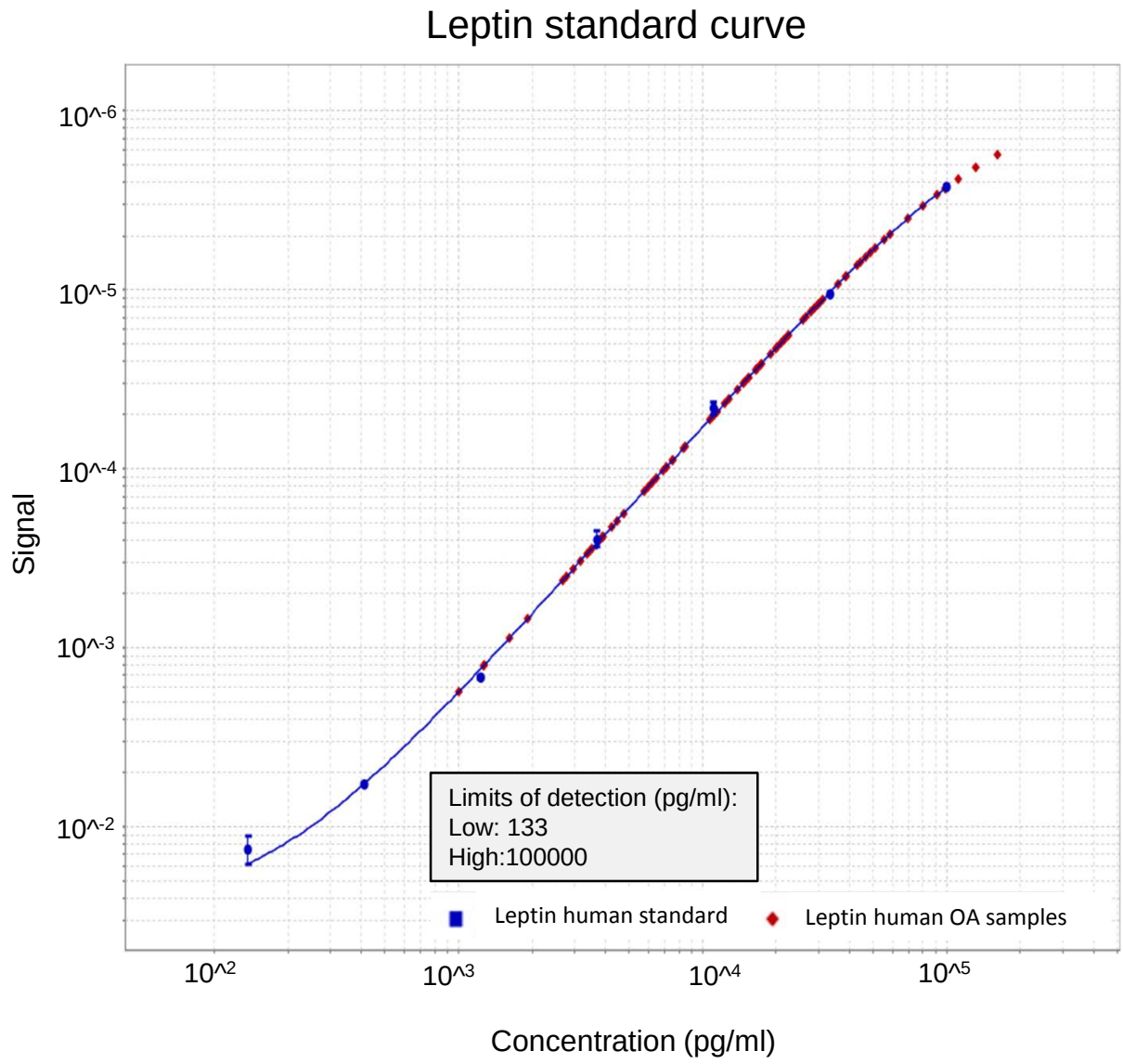



**Figure 6.7**

Leptin standard curve generated using the Leptin (human) Calibrator blend (MSD, Gaithersburg, USA). Blue squares indicate Leptin calibration standards of known concentration, with each standard run in duplicate. OA patient samples are depicted by the red diamonds, with both knee and hip arthroplasty patients shown on the graph. One hip OA patient recorded a serum leptin concentration above the limit of detection, but given that the data is not being used for clinical diagnostic purposes, was included in the analysis.

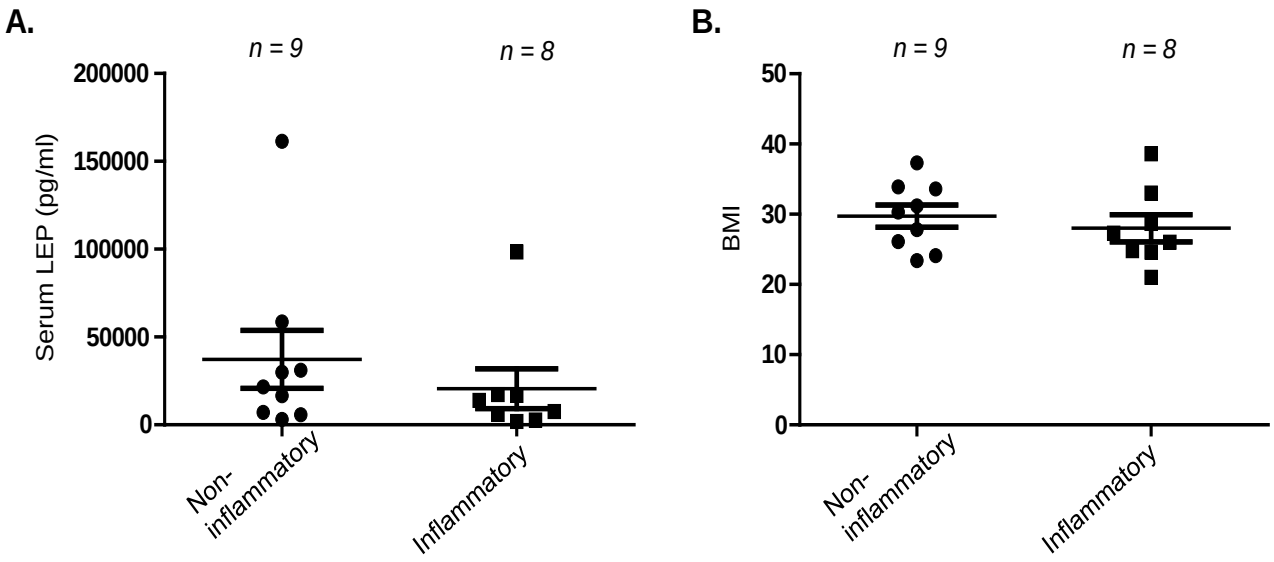



**Figure 6.8**

Analysis of serum leptin levels (A) and BMI (B) in hip OA patients, stratified according to inflammatory subcluster. There were no significant differences in either serum leptin or BMI between the two groups, as assessed using a Mann-Whitney *U* test. *n* represents the number of individuals in the comparison group; error bars represent the mean ± SEM.

### 6.3.4 Assigning new knee OA patients into the inflammatory or non-inflammatory subgroup according to cartilage methylation

Matched cartilage and blood serum were obtained for 29 patients who had undergone total knee arthroplasty. Patient details are listed in Appendix B: Table B.3. Cartilage methylation at cg12245706 and cg21944234 was quantified by pyrosequencing (Chapter 2.14.2), and patients were subsequently assigned to the inflammatory or non-inflammatory cluster using the discriminant function created in Chapter 5 (Chapter 5.3.5). This identified 18 patients as belonging to the non-inflammatory group (cluster 1) and 11 to the inflammatory group (cluster 2), with none of the samples failing to be categorised (Table 6.2).

| Sample | cg12245706 average methylation (%) | cg21944234 average methylation (%) | Cluster |
|---|---|---|---|
| 5994 | 58 | 37 | Cluster 2 |
| 6002 | 36.5 | 25.5 | Cluster 1 |
| 6003 | 65 | 29 | Cluster 2 |
| 6006 | 18.5 | 22.5 | Cluster 1 |
| 6007 | 59.5 | 33.5 | Cluster 2 |
| 6008 | 43 | 34.5 | Cluster 1 |
| 6012 | 75.5 | 40 | Cluster 2 |
| 6015 | 51.5 | 27.5 | Cluster 1 |
| 6054 | 59 | 36.5 | Cluster 2 |
| 6057 | 35.5 | 33.5 | Cluster 1 |
| 6090 | 31.5 | 27 | Cluster 1 |
| 6095 | 45.5 | 21.5 | Cluster 1 |
| 6104 | 22 | 23.5 | Cluster 1 |
| 6114 | 27 | 32.5 | Cluster 1 |
| 6146 | 62 | 27 | Cluster 2 |
| 6150 | 18.5 | 25 | Cluster 1 |
| 6153 | 66 | 28 | Cluster 2 |
| 6164 | 60.5 | 23 | Cluster 2 |
| 6165 | 41.5 | 26.5 | Cluster 1 |
| 6175 | 49 | 24 | Cluster 1 |
| 6190 | 20.5 | 22 | Cluster 1 |
| 6203 | 40 | 25 | Cluster 1 |
| 6204 | 63.5 | 30.5 | Cluster 2 |
| 6208 | 54.5 | 30 | Cluster 2 |
| 6209 | 43 | 21.5 | Cluster 1 |
| 6223 | 38 | 22 | Cluster 1 |
| 6225 | 67.5 | 34 | Cluster 2 |
| 6227 | 45 | 20 | Cluster 1 |
| 6228 | 36 | 23.5 | Cluster 1 |

**Table 6.2**

Average cartilage methylation at cg12245706 and cg21944234 in 29 knee OA patients and designated inflammatory subgroup, where cluster 1 is the non-inflammatory and cluster 2 the inflammatory group. CpG methylation was analysed by pyrosequencing and statistical analyses were performed with JMP Pro Version 11.2 Statistical Data Visualization Software (SAS Institute Inc., Cary, NC) to determine the subcluster by using the discriminant function described in Chapter 5.3.5.

### 6.3.5 Investigation of systemic markers of inflammation in knee OA patients

Although knee OA patients did not display significant differences in cartilage inflammatory gene expression when stratified according to the non-inflammatory and inflammatory clusters (Chapter 5.3.6), the two groups are clearly distinguished according to their cartilage methylome [43]. This may have important implications for patient subgrouping in future clinical trials, in respect to having a potential impact on response to treatment. Analysis of the cartilage methylome however, requires a cartilage sample and would limit subgrouping of patients to those who had already undergone arthroplasty. As such, identification of a circulating factor that can discriminate between the two clusters could prove highly beneficial for patient subgrouping for future research.

Given that methylation differences were enriched in inflammatory genes, the focus for a circulating biomarker was centred on circulating inflammatory factors. Serum CRP, IL6 and TNF-α were analysed using the MSD MULTI-SPOT V-PLEX Assay System (MSD, Gaithersburg, USA), as detailed in Chapter 2.16. Samples were analysed along with the previously discussed hip OA samples (Figures 6.1, 6.3 and 6.5 respectively). Comparison of the 18 non-inflammatory and 11 inflammatory knee OA patients did not reveal any difference in circulating levels of CRP, nor of circulating levels of the selected cytokines (Figure 6.9).

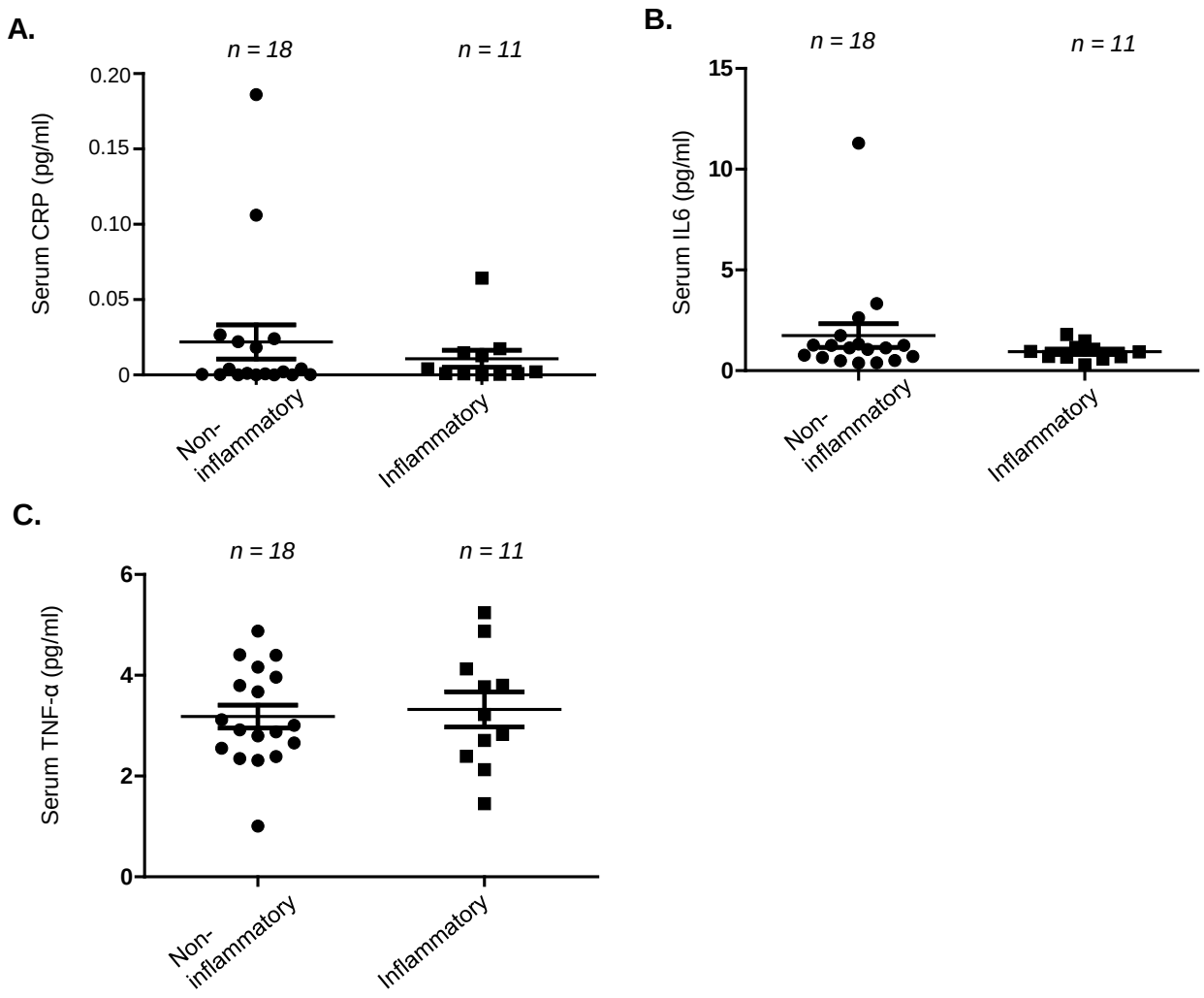



**Figure 6.9**

Analysis of serum CRP (A), IL6 (B) and TNF-α (C) levels in knee OA patients, stratified according to inflammatory subcluster. There were no significant differences in circulating CRP, IL6 or TNF-α between the two groups, as assessed using a Mann-Whitney *U* test. *n* represents the number of individuals in the comparison group; error bars represent the mean ± SEM.

### 6.3.6 Analysis of circulating leptin and BMI in knee OA patients

Having not identified a circulating inflammatory factor which distinguishes the two knee OA inflammatory groups, circulating leptin levels were next interrogated. As noted earlier, circulating leptin may be driving the differential gene methylation in the two knee OA groups and would prove an easily quantifiable marker for patient subgrouping. Further to this, leptin is known to correlate with BMI [151]. Owing to the fact that it is an easy and non-invasive measurement, BMI would be a preferential measurement to use for future patient stratification.

Serum leptin was measured using the MSD V-PLEX Human Leptin Kit (MSD, Gaithersburg, USA), as detailed in Chapter 2.16 (Figure 6.7). Patient BMI data was recorded prior to arthroplasty surgery and was provided by Mr Kenneth Rankin. As with circulating inflammatory markers, neither serum leptin nor BMI showed a significant difference between the patient subgroups (Figure 6.10). Further stratification by sex, to account for the natural variation in BMI between males and females, also did not identify any significant differences in either parameter (Figure 6.11).

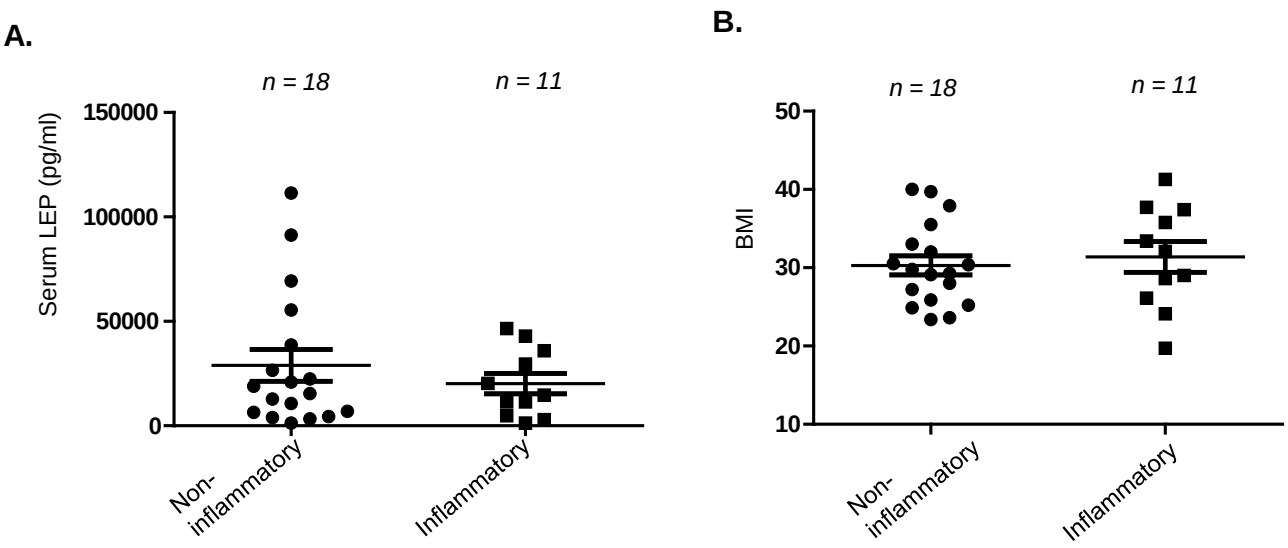



**Figure 6.10**

Analysis of serum leptin levels (A) and BMI (B) in knee OA patients, stratified according to inflammatory subcluster. There were no significant differences in either serum leptin or BMI between the two groups, as assessed using a Mann-Whitney *U* test. *n* represents the number of individuals in the comparison group; error bars represent the mean ± SEM.

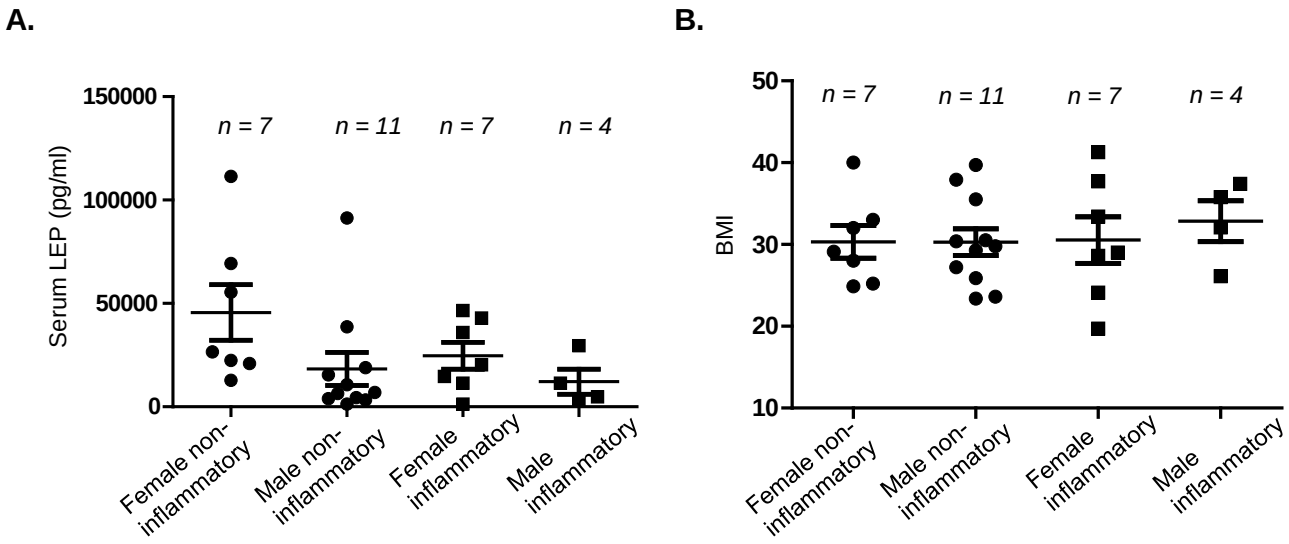



**Figure 6.11**

Analysis of serum leptin levels (A) and BMI (B) in knee OA patients, stratified according to both inflammatory subcluster and sex. There were no significant differences in neither serum leptin nor BMI, having performed two-way comparisons for all possible group combinations, as assessed using a Mann-Whitney *U* test. *n* represents the number of individuals in the comparison group; error bars represent the mean ± SEM.

## 6.4 Discussion

The aim of the research in this chapter was to investigate circulating inflammatory and metabolic factors in hip and knee OA patients, in response to the identification that a subgroup of patients display altered cartilage methylation of inflammatory genes [43]. Given that increased leptin has been shown to correlate with a heightened inflammatory state [150], I hypothesised that, in hip patients, increased cartilage expression of inflammatory genes may be driven by increased systemic inflammation, potentially caused by high BMI and so, high circulating leptin. I also hypothesised that a circulating inflammatory or metabolic marker may be able to differentiate between our two patient subgroups in both hip and knee OA patients, thus allowing patient stratification for future clinical trials and diagnostic purposes. This may also have implications for patient treatment, for example the inflammatory subgroup may benefit from repurposing of existing anti-inflammatory drugs. Both lines of investigation were achieved by methylation analysis of new patient samples to create a new patient hip and knee cohort, followed by analysis of selected factors in blood serum using Electrochemiluminescent Meso Scale Discovery (MSD) assays (MSD, Gaithersburg, USA).

I first theorised that increased cartilage inflammatory gene expression in hip OA patients belonging to the inflammatory subgroup may be secondary to increased systemic inflammation. I analysed serum CRP, a general marker of inflammation, as well as IL6 and TNF-α in nine non-inflammatory and eight inflammatory hip OA patients. This did not identify any difference in circulating inflammatory factors, though qPCR analysis of matched cartilage did show a significant difference in inflammatory gene expression in the inflammatory subgroup, as shown previously.

Consequently, this implies increased inflammation in the inflammatory hip subgroup is a primary event in the cartilage as opposed to a result of elevated systemic inflammation. It also indicates that the selected markers do not distinguish the patient subgroups in view of subgrouping future patients by use of blood sample analysis.

I next analysed serum leptin and BMI in the non-inflammatory and inflammatory hip patients to investigate if elevated BMI or circulating leptin may be driving aberrant methylation and cartilage gene expression in the inflammatory group, or may be a method to determine subcluster membership of future patients. Circulating leptin nor BMI were able to distinguish our hip OA subgroups, therefore suggesting that altered metabolic status is not what is driving the divergent methylome and cartilage gene expression in our inflammatory subgroup.

Subsequently, I analysed serum inflammatory markers in knee OA patients. Methylation analysis by pyrosequencing and subsequent discriminant analysis using JMP Pro Version 11.2 Statistical Data Visualization Software (SAS Institute Inc., Cary, NC) was performed. This identified a new cohort of 18 non-inflammatory and 11 inflammatory knee OA patients. I then explored whether the inflammatory subgroup could be identified based on blood inflammatory and metabolic markers. Serum CRP, IL6, TNF-α and leptin were unable to distinguish the inflammatory from the non-inflammatory groups. Likewise, BMI was not significantly different between the subgroups, even following further stratification by sex to account for the fact that females have naturally higher BMI.

Overall, this chapter indicates that abnormal inflammatory gene regulation is a local, as opposed to a systemic, event in both knee and hip OA inflammatory cluster patients. We did not find any evidence of elevated systemic inflammation in the blood serum of the inflammatory subgroup cohort. This research also shows that circulating CRP, IL6, TNF-α and leptin are not suitable biomarkers to enable for stratification of the inflammatory subgroup in hip or knee OA patients. However, given that peripheral blood DNA methylation levels have been shown to correlate to that of internal tissues including skeletal muscle and brain, further investigations into the blood methylome are warranted, particularly in reference to hip OA where we have demonstrated increased cartilage inflammatory gene expression [152,153]. A DNA methylation signature in peripheral blood would provide an easy, accessible method to distinguish inflammatory subgroup patients and therefore, facilitate easy patient stratification for OA clinical trials.

# Chapter 7. General discussion

## 7.1 Perspective

Osteoarthritis is a progressive, painful disease which is increasing in prevalence as the global expected life expectancy increases [154]. Both pharmacologic and non-pharmacologic therapeutic options, however, remain suboptimal [155]. To improve the likelihood of successful clinical trials being designed, a more comprehensive understanding of OA pathophysiology is vital. Given the highly heterogeneous nature of the disease, molecular research must therefore retain a broad focus, encompassing genetics, epigenetics and transcriptomics, in order to provide the most comprehensive disease understanding possible.

Genetic studies have entailed from candidate gene studies, linkage studies and GWAS, which have indicated several regions of the genome are harbouring OA susceptibility. However, association alone does not benefit our comprehension of the disease and it is therefore vital to functionally analyse each signal in an attempt to dissect the mechanisms driving OA association. As the most powerful OA GWAS performed to date, the arcOGEN consortium study identified five genomic regions to be associated with OA [92]. Subsequently, our group has focused on the functional investigation of these loci. One of the association SNPs has been identified to act as a *cis*-eQTL, with the genotype associating with *GNL3* and *SPCS1* gene expression [156]. Combined molecular and bioinformatics studies into the association signal marked by rs835487, residing in an enhancer of *CHST11,* have identified the polymorphism to mediate OA risk owing to differential protein binding to its alleles, which may therefore affect expression of *CHST11* or an alternative gene [157].

Analysis of two other signals identified in the arcOGEN study did not find any evidence for functionality in end-stage disease tissue, suggesting the OA associations are modulated in joint development or early disease [158, 159].

Residing in an intron of the *ASTN2* gene, the final signal discovered by the arcOGEN study is rs4836732. The region demarcated by this polymorphism has been identified to be associated with height in individuals of European descent [160]. Furthermore, genome-wide analysis for copy number variants (CNVs) using single-nucleotide polymorphism arrays in patients with short stature identified CNVs in the rs4836732 association region which segregated with short stature [161]. These findings provide evidence that the locus is important in skeletal development and, with the results presented in this thesis, may indicate the OA association signal regulates OA susceptibility during joint development.

Epigenetic research in OA has gained momentum over the past 10 years and there is a growing body of work focusing on genome-wide methylation studies, particularly in cartilage [77]. These have identified that not only is the OA cartilage methylome distinct compared to non-OA cartilage, but also that patients with the same affected joint can display distinct differences in their cartilage methylome [43, 118, 122]. Having identified a subgroup of OA patients who have altered methylation and expression of inflammatory genes in cartilage, the study by Rushton *et al*., supported by the research generated for this thesis, provides an exciting direction for future clinical trials, with specific patients potentially benefiting from re-purposing of current anti-inflammatory treatments [43, 121].

## 7.2 Key results

My PhD had two overall aims: to functionally dissect the OA associated locus marked by rs4836732 (aim one) and to characterise a subgroup of OA patients displaying differential methylation of inflammatory genes in cartilage DNA (aim two). The key findings of this thesis are as follows:

Aim one

- Two SNPs are in LD at $r^2 > 0.7$ with rs4836732, but functional studies using luciferase reporter assays, electrophoretic mobility shift assays (EMSAs) and a range of transformed cell lines did not identify any differential allelic activity between any of the SNP alleles
- Differential EMSA banding patterns for the C allele of rs4836732 were observed, when compared to the T allele
- All three candidate genes, *ASTN2*, *PAPPA* and *TRIM32,* are expressed in synovial joint tissues
- No correlation between the association SNP and expression of candidate genes was observed in end-stage disease tissue
- *PAPPA* expression is required for osteogenic and chondrogenic differentiation of human mesenchymal stem cells

Aim two

- CpG methylation analysis by pyrosequencing using cartilage DNA can be used to predict inflammatory cluster membership of OA patients
- There is no difference in circulating leptin or BMI when comparing patients according to their inflammatory subgroup

- Inflammatory subgroup patients with knee OA display increased cartilage expression of some inflammatory genes, though this is not as pronounced as for the hip OA patients
- Increased local expression of inflammatory genes is not accompanied by increased systemic inflammation

Together, the data in this thesis emphasise the heterogeneity of OA pathophysiology and support the notion that various different pathological pathways are converging onto the same end disease phenotype. Specifically, this research will allow for further analysis of the rs4836732 association signal in relation to joint development, with a particular focus on *PAPPA*. It also supports continued work to fully characterise the inflammatory phenotype of a subgroup of OA patients, having found that elevated expression of inflammatory genes appears to be localised to the joint, rather than being a systemic event.

### 7.3 Investigating the functionality of the rs4836732 OA susceptibility locus

In Chapter 3, I hypothesised that rs4836732 marks the action of a *cis*-eQTL which increases OA susceptibility by mediating the expression of a nearby gene. It is plausible for SNPs to regulate genes megabases in distance away, but many functional studies of signals have found them to act as eQTLs on nearby genes. An excellent example of this is the regulation of *RUNX3* expression by rs4648889 in association with ankylosing spondylitis, whereby presence of the risk allele resulted in impaired transcription factor binding, thus reducing *RUNX3* expression [162]. In relation to OA, this has been demonstrated with *GNL3* and *SPCS1*, as well as *GDF5*, *ALDH1A2* and *NCOA3* [78, 120, 153, 163].

Consequently, investigations were focused on *ASTN2*, *PAPPA* and *TRIM32*, which reside nearby to rs4836732. However, quantification of overall gene expression and allelic expression in OA cartilage did not identify any such correlations, as has been the case with other similar studies [159, 164]. Analysis by Raine *et al*., of the 7q22 OA association signal identified tissue-specific eQTL effects [165]. As such, I also performed allelic expression analysis of gene expression in synovium and fat pad samples, though this also did not yield any positive correlations. A key limitation of analysing overall gene expression is that inter-individual variability may conceal any true correlations, which could only be overcome by using a very large sample size. Both the overall gene expression analysis and allelic analysis may also have been limited by the fact that, owing to availability, a variety of OA patient samples were used, rather than only hip OA samples. Taken together, it cannot be dismissed that an eQTL acting in end stage disease tissues may have gone undetected. A further important limitation is that the research was conducted using end-stage disease tissues. As such, the possibility remains that susceptibility may be acting during joint development, or disease initiation/ progression.

Differential protein binding has been identified to mediate the rs835487 OA susceptibility signal within intron 2 of *CHST11* and differential transcription factor binding has been identified in the context of *GDF5* regulation by rs143383 [157, 166]. As such, I hypothesised that the rs4836732 susceptibility signal may function in a similar manner. While EMSA analysis indicated differential protein binding to the SNP when comparing binding to either the C or T rs4836732 allele, supershifts did not manage to identify the proteins. Consequently, this may suggest that the region is functional, but verifying the identity of the protein complexes differentially binding

the alleles has not been possible during the course of my PhD. However, this could be achieved in the future using an oligonucleotide pull down assay, followed by mass spectrometry [166].

### 7.4 The role of PAPPA in joint development

A possible explanation as to why no eQTL was detected in this study is that the effects are being exerted during joint development, as opposed to in end stage disease tissue. A study analysing radiographs and genotype data from a subset of 929 of the arcOGEN patients with unilateral hip OA generated statistical shape modes to capture the shape variation of the OA-unaffected proximal femur [137]. A significant association was found between rs4836732 and one of the female modes, suggesting the SNP may have an effect on the superior orientation and size of the femoral head. Of the candidate genes, *PAPPA* was deemed the most compelling gene for a role in hip joint development, given that it has been described to have an anabolic role in bone formation both *in vitro* and *in vivo* [106, 107].

Knock-down of *PAPPA* in hMSCs resulted in impaired osteogenic differentiation, as assessed by expression analysis of genes upregulated during osteogenesis and cellular staining to assess alkaline phosphatase activity [139]. Assessment of chondrogenic differentiation also indicated *PAPPA* knock-down resulted in alterations to the chondrogenic differentiation capacity of hMSCs, reflected in increased *SOX9* expression during late stages of pellet chondrogenesis. *Pappa* expression has been identified to be higher during very early-stage bone development compared to later stages of joint development, when analysed in

collagen-induced arthritis mouse models [108]. Combined with the data presented in this thesis, this therefore invites the hypothesis that dysregulation of *PAPPA* during early joint development may affect normal joint shape formation. It may therefore be the case that, although the OA association marked by rs4836732 does not contribute to disease susceptibility by modulating gene expression in end-stage diseased synovial joint tissue, it may act during joint development and as such, *PAPPA* is the most promising gene candidate analysed to date for this locus.

### 7.5 Characterising an inflammatory subgroup of OA patients

Heterogeneity within OA patients extends beyond pure genetic variation, and subgroups of OA patients display altered genome-wide methylation of inflammatory genes in cartilage [43]. In Chapter 5, I proposed that these patients could be identified by analysis of individual differentially methylated CpG sites and accordingly selected appropriate CpGs for both hip and knee cartilage analysis. The subsequent design of both assays and statistical methodology to determine cluster membership has enabled me to further characterise the inflammatory subgroup in new samples which were not analysed on the high-density Illumina array. This was not only beneficial in acting to increase my sample size, but it will enable future research to continue without requiring ongoing high-density CpG methylation arrays to determine the inflammatory group of new patients.

Altered expression of inflammatory genes in the subgroup of OA hip patients in the absence of increased systemic inflammation presents the potential that the abnormal inflammatory response is localised to the affected joint. Indeed, a recent novel study

employing fluorescence optical imaging (FOI) of wrist and finger joints in 23 OA patients detected active local inflammation [167]. This poses the question as to whether patients in the inflammatory subgroup perhaps display either more severe, or a different pattern, of detectable inflammation compared to the non-inflammatory subgroup. Were it possible to characterise patients belonging to the inflammatory cluster using a non-invasive method such as FOI, this may present an exciting clinical opportunity to identify such patients and to help assist in determining the course of therapeutic intervention. These same differences in cartilage expression of inflammatory genes were not seen to as great an extent in knee cartilage, indicating that investigation of a different panel of inflammatory genes may be indicated in inflammatory cluster knee OA.

Further to identifying inflammatory patients in order to inform therapeutic options, it is also vital to understand the cause of increased localised inflammation in these patients. Along with altered inflammatory gene expression, Rushton *et al*. also identified differential methylation and expression of zinc transporter genes, such as *ZIP8* [121]. Not only do matrix-degrading enzymes such as MMPs and ADAMTSs require $Zn^{2+}$ for their catalytic activity, but research in mice has indicated the zinc transporter ZIP8 plays a vital role in upregulating matrix-degrading enzymes, MMP3, MMP13, and ADAMTS5, thus inducing cartilage damage [167]. During the normal wound-healing response, damage to articular cartilage is succeeded by a so called ‘auto-debridement response’. Research in chondrocytes has identified that mediators such as IL1 can be induced by cartilage damage and result in increased abundance of ZIP8, ultimately enhancing MMP activity to induce cartilage break-down and assist in the removal of damaged cartilage [168]. Having not found any evidence for

increased systemic inflammation, it may therefore be the case that the inflammatory subgroup of patients I studied are demonstrating an impaired wound-healing response, mediated by abnormal zinc transporter expression [169].

Looking to the future, a key aim is to transfer our expanding knowledge of inflammation in OA to help develop effective therapeutic treatment. While previous clinical trials of anti-inflammatory therapeutics in OA have proved unfruitful, a recent review identified that trials thus far have failed to identify or stratify patients accordingly [170]. Failing to select appropriate patient populations in clinical trials can clearly have a negative impact on the trial outcome, highlighted by studies into epidermal growth factor receptor (EGFR)-targeted therapy for non–small cell lung cancer patients [171]. Here, the drug Iressa (AstraZeneca), initially failed to progress beyond Phase II trials, yet was later approved when subsequent analysis showed those patients with an EGFR mutation had improved survival. Given that a vast bank of evidence already exists as to the heterogeneity of OA, there is clear need for patient stratification in clinical trials if they are to prove successful. As such, despite no evidence of altered systemic inflammatory markers, continued investigations using peripheral blood of inflammatory subgroup patients is essential in attempts to identify patient-selection biomarkers.

### 7.6 Future work

In order to identify the SNP marked by the rs4836732 association signal which is mediating increased disease risk, I would extend the analyses performed in chapter three. Owing to the timeframe of my PhD, only rs4836732 was taken forward to investigate protein binding using EMSA analysis. While this was an informed

decision, based on lack of evidence favouring either of the other identified SNPs over the discovery SNP, it would be prudent to perform these experiments on rs4837613 and rs13283416 as well. Not only this, but in all cases the EMSA experiments could be furthered by using nuclear protein extracted from hip cartilage, in light of the fact that it is a female hip association signal. An additional improvement to the EMSA experiments could be to increase the length of the probes in an attempt to ensure proteins which bind over larger distances are also identified. However, the benefits of this approach may be offset by increased non-specific binding. As mentioned previously, a hypothesis-free approach to identify proteins differentially binding the EMSA probes would be to perform an oligonucleotide pull down assay followed by quantitative mass spectrometry [166]. This would therefore enable both the identification and quantification of any proteins binding the different SNP alleles.

Recently, DNA and cDNA from OA patient subchondral bone has been made available within our group. As such, it would be advisable to extend the allelic expression imbalance analysis to include this tissue type. Targeted deletions of the candidate SNPs using the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system are already being performed within our group in relevant cell lines [172]. This approach will allow for the comparison of candidate gene expression in the cell lines on the same genetic background, except one will have the deletion and the other will not. As such, it may provide new evidence as to both the functional SNP, and the gene regulated by the potential *cis*-eQTL.

A fundamental limitation of the functional analysis of the association signal is that the hypothesis employed is that rs4836732 marks a *cis*-eQTL. Consequently, the investigation focused solely on the genes surrounding the signal. While a sensible line of research, the lack of positive data correlating genotype at the signal with gene expression may support the notion of also looking outside of the region. The *in vivo* chromosome conformation is essential to transcriptional regulation [173]. As such, the gene regulated by the association signal may indeed be located outside of the association locus, even up to megabases away [174]. Employing Chromosome Conformation Capture (3C), Circularized Chromosome Conformation Capture (4C), or Carbon-Copy Chromosome Conformation Capture (5C) would facilitate the characterisation of chromosome folding and identify other genomic loci that the association region may interact with [175]. Were interactions identified, this would greatly influence future research depending on the genes present in these regions. Subsequent experiments could then include knock-down of the identified proteins in order to assess the effect on the expression of candidate genes.

Having identified that knock-down of *PAPPA* in mesenchymal stem cells impairs osteogenic and chondrogenic differentiation, the next stage of this line of investigation would be to attempt to identify whether the susceptibility region is perhaps functional during joint development. One approach would therefore be to quantify expression of the candidate gene(s) during mesenchymal stem cell differentiation and also, to perform AEI during differentiation. Furthermore, highly efficient *Pappa* gene deletion has been achieved in bone in a conditional knock-out mouse model, though studies with this model to date have been limited to analyses

of vascular injury [176, 177]. As such, this model presents another mechanism to assess *PAPPA* functionality in the skeleton at different developmental stages.

In relation to Chapters 5 and 6, it is essential to definitively confirm the ability of the developed methodology to determine inflammatory cluster membership of new patient samples. To achieve this, new cartilage DNA samples assigned to the inflammatory or non-inflammatory cluster using methylation data generated by pyrosequencing should also be analysed by high-density CpG methylation array. Further to this, a basic improvement to the systemic inflammation study would be to increase the sample size.

Though higher expression of some of the analysed inflammatory genes was seen in knee OA cartilage when comparing the two subgroups, the differential expression was not as striking as in hip OA. This may reflect a higher degree of heterogeneity within the knee inflammatory group, which would be supported by the fact that fewer differentially methylated loci were identified between the clusters in comparison to the hip samples [43]. However, it would be best to tailor the inflammatory gene expression analysis to the knee methylation data, as the inflammatory genes displaying the greatest methylation difference in hip samples are not necessarily the same as in the knee samples.

Having not identified any association between metabolic status, in terms of body fat and systemic inflammation, and the inflammatory OA subgroup, future work should focus on characterising the upstream regulator of the localised inflammation. As

discussed in Chapter 6.5, inflammatory cluster patients also demonstrated altered methylation and expression of zinc transporter genes [121]. Consequently, it is important to determine whether zinc transporter gene expression is increased in OA knee cartilage, as has been identified in hip cartilage. I have shown in my research that *IL1A* is differentially expressed in inflammatory cluster patients. Given that IL-1α increases intracellular IL-1β in chondrocytes *in vitro* and is proposed to be an essential mediator of early OA, it would be interesting to determine whether IL-1α can also mediate the *in vivo* ZIP8 inflammatory response in mouse models, as has been shown for IL-1β [169].

Research is currently limited to end-stage OA cartilage, obtained post operatively. Evidently, to truly exploit the findings to benefit patients, a biomarker to identify inflammatory subgroup patients is essential. As such, future research must continue to include analysis of patient blood to try and characterise a circulating factor, or methylation signature, which allows for early non-invasive subgrouping of OA patients and to potentially enable for differential clinical treatment of individuals. As such, this would allow for the true translation of this research from bench to bedside.

### 7.7 Concluding summary

In overall conclusion, I have achieved the aims of my PhD: to functionally analyse the OA association signal marked by the rs4836732 polymorphism and to investigate a subgroup of OA patients who display altered methylation of inflammatory genes in cartilage DNA. The functionality of the OA association region could not be correlated with the genotype at the rs4836732, and no *cis*-eQTL was identified in end-stage OA

tissue. As such, it is more likely that the OA risk is mediated during joint development, or early-stage disease, and as such, I have provided evidence to support further research focusing on the *PAPPA* gene. In reference to my second aim, I have created methodology to determine the inflammatory subgroup of new patient cartilage samples and provided evidence that elevated inflammation appears to be local to the synovial joint, as opposed to a secondary event caused by increased systemic inflammation.

## Appendix A. Primer sequences

| SNP | Forward primer (5′ - 3′) | Reverse primer (5′ - 3′) | Sequencing primer (5′ - 3′) | $MgCl_2$ (mM) | Anneal (°C) |
|---|---|---|---|---|---|
| rs2565 | [Btn]TGACTGGGCAGTGCTTTCT | AATAAAGAGATCCCCAAAGTGGT | CCCAAAGTGGTGGG | 2.0 | 57.5 |
| rs7518 | [Btn]GCTTGGTCTCCACCTGAA | TTTCTTATGGATGGCTGTCAATC | ATGGCTGTCAATCCC | 2.0 | 57.5 |
| rs2281627 | [Btn]GCTATGCCATTTTGCTTCCTTAT | GGGGGATGATTCTGTACACATG | GGGAGAAAGATTTGGATTA | 2.0 | 57.5 |

| SNP | Forward primer (5′ - 3′) | Reverse primer (5′ - 3′) | Restriction enzyme | $MgCl_2$ (mM) | Anneal (°C) |
|---|---|---|---|---|---|
| rs4836732 | GAAAGAAGGGAAGGTAAGATGGGAGAGACAGCACCT | CTATCTTACACTCTCAAG | *RsaI* | 2.0 | 55.0 |

**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. Bottom panel: primer sequences and restriction enzymes used for rs4836732 genotyping by restriction fragment length polymorphism analysis.

| SNP | Forward primer (5′ - 3′) | Reverse primer (5′ - 3′) | $MgCl_2$ (mM) | Anneal (°C) |
|---|---|---|---|---|
| rs13283416 | GGGGACGCGTTAGGTTAGTGGTTGCATAGGGTT | GGGGCTCGAGCTGTGCATCGTTCTTGTGCTATA | 2.0 | 63.9 |
| rs4836732 | GGGGACGCGTTGTCTACAATTGGAATGGTGAGC | GGGGCTCGAGCTACAGATGCACGATACAACACC | 2.0 | 69.7 |
| rs4837613 | GGGGACGCGT CCTCTTTGTACTTTACCCAGCTG | GGGGCTCGAGATTCCAAAGTTCAGCCTCTTAGC | 2.0 | 63.9 |

**Table A.2 Primer sequences and conditions used for the cloning of fragments into pGL3 promoter vectors.** PCR primers and conditions used for cloning into pGL3 vectors between the restriction sites *MluI* and *XhoI*. Restriction sites are in red text, preceded by a 5′ GGGG.

| Sequence name | Forward primer (5′ - 3′) | Reverse primer (5′ - 3′) |
|---|---|---|
| Random primer full length | TAGCACCTCCAGGCTGTCTGAGTACGTAGCA | TGCTACGTACTCAGACAGCCTGGAGGTGCTA |
| C allele probe and competitor | GAGAGACAGCACCTA**C**TTTCTGAGGTCTAAG | CTTAGACCTCAGAAA**G**TAGGTGCTGTCTCTC |
| T allele probe and competitor | GAGAGACAGCACCTA**T**TTTCTGAGGTCTAAG | CTTAGACCTCAGAAA**A**TAGGTGCTGTCTCTC |
| Competitor 1: C allele | TAGCACCTCCAGGCA**C**TTTCTGAGGTCTAAG | CTTAGACCTCAGAAA**G**TGCCTGGAGGTGCTA |
| Competitor 1: T allele | TAGCACCTCCAGGCA**T**TTTCTGAGGTCTAAG | CTTAGACCTCAGAAA**A**TGCCTGGAGGTGCTA |
| Competitor 2: C allele | TAGCACCTCCACCTA**C**TTTCTGTACGTAGCA | TGCTACGTACAGAAA**G**TAGGTGGAGGTGCTA |
| Competitor 2: T allele | TAGCACCTCCACCTA**T**TTTCTGTACGTAGCA | TGCTACGTACAGAAA**A**TAGGTGGAGGTGCTA |
| Competitor 3: C allele | GAGAGACAGCACCTA**C**TCTGAGTACGTAGCA | TGCTACGTACTCAGA**G**TAGGTGCTGTCTCTC |
| Competitor 3: T allele | GAGAGACAGCACCTA**T**TCTGAGTACGTAGCA | TGCTACGTACTCAGA**A**TAGGTGCTGTCTCTC |

**Table A.3 Primer sequences used for EMSAs to investigate SW872, SW1353 and U2OS nuclear protein binding to the rs4836732 SNP.** The original competitor sequences are underlined in red and the random sequences are in black text.

| Gene | Forward primer (5′ - 3′) | Reverse primer (5′ - 3′) | cDNA/gDNA fragment size (bp) | $MgCl_2$ (mM) | Anneal (°C) |
|---|---|---|---|---|---|
| *HBP1* | TCGAAGAGTGAACCAGCCTT | GAAGGCCAGGAATTGCACCATCC | 152/570 | 2.0 | 60 |

| Gene | Probe sequence (5′-3′) | Primer 1 (5′-3′) | Primer 2 (5′-3′) |
|---|---|---|---|
| *18S* | 56-FAM/TCCTTTGGTCGCTCGCTCCTCTCCC/TAMRA | TATTAGCTCTAGAATTACCACAGTTA | CGAATGGCTCATTAAATCAGTTATG |
| *ACAN* | 56-FAM/CTGGGTTTT/ZEN/CGTGACTCTGAGGGT/3IABkFQ | TGTGGGACTGAAGTTCTTGG | AGCGAGTTGTCATGGTCTG |
| *ALPL* | 56-FAM/AGTGGGAGTGCTTGTATCTCGGTTTG/3IABkFQ | GATGTGGAGTATGAGAGTGACG | GGTCAAGGGTCAGGAGTTC |
| *ASTN2* | 56-FAM/CCTCAGCGC/ZEN/CTTTGATGCCAATG/IABkFQ | TGATCGATGACTGGTGCAG | TCCAAGGAGACCACAGTACT |
| *COL1A1* | 56-FAM/TTCCGGGCA/ZEN/ATCCTCGAGCA/3IABkFQ | CCCCTGGAAAGAATGGAGATG | TCCAAACCACTGAAACCTCTG |
| *COL10A1* | 56-FAM/CCAAGACACACAGTTCTTCATTCCCTACACC/36-TAMSp | GTACCTTGCTCTCCTCTTACTG | CATAAAAGGCCCACTACCCA |
| *GAPDH* | 56-FAM/AAGGTCGGAGTCAACGGATTTGGTC/IABkFQ/36-TAMSp | TGTAGTTGAGGTCAATGAAGGG | ACATCGCTCAGACACCATG |
| *HPRT1* | 56-FAM/AGGACTGAACGTCTTGCTCGAGATG/36-TAMSp | ACAGAGGGCTACAATGTGATG | TGCTGAGGATTTGGAAAGGG |
| *KLF4* | 56-FAM/TGGAAATTCGCCCGCTCAGATGA/3IABkFQ | GTTTACGGTAGTGCCTGGTC | AAGAGTTCCCATCTCAAGGC |
| *PAPPA* | 56-FAM/ACTTTTTGG/ZEN/CTCCGGGCGTATTTTTC/3IABkFQ | TGATGGTCTCCTGCTTTTGG | CATCAGCTACCCATATTCCCAG |
| *RUNX2* | 56-FAM/CTGTTGGTC/ZEN/TCGGTGGCTGGTAG/3IABkFQ | TGTTTGATGCCATAGTCCCTC | AATGGTTAATCTCCGCAGGTC |
| *SOX9* | 56-FAM/TCTGGAGAC/ZEN/TTCTGAACGAGAGCGA/3IABkFQ | ACTTGCACAACGCCGAG | CTGGTACTTGTAATCCGGGTG |
| *TRIM32* | 56-FAM/TGGCTTGGG/ZEN/AGCAGAAGAAACTGAT/IABkFQ | TGGACTGTTGGATCATTGC | ACAACCAAGAACTAAGGGAAGG |

**Table A.4 Primer and probe sequences used for gene expression analysis by real-time PCR using PrimeTime® predesigned assays.** Top panel: primer sequences used for validation of cDNA synthesis. Bottom panel: primer and probe sequences used for quantification of gene expression

| CpG | Forward primer (5′ - 3′) | Reverse primer (5′ - 3′) | Sequencing primer (5′ - 3′) | Anneal (ºC) |
|---|---|---|---|---|
| cg03278514 | TGAGGAGTTATTTTTATTGATGGGAAATT | [Btn]ACCTCTACTCTCTCTAACCTCAATCATAATAT | ATGGGGATTTATTTTGAGATA | 60 |
| cg14818279 | GGGAGAAAAGTTTTTTTAATTTAGTAGGT | [Btn]AAAACTATCCCTTACAACACTCT | TTGGGTTTAGGGTTAGAT | 60 |
| cg12245706 | AGAGAATTTTTTTAGGTTTTAGTAAGG | [Btn]AAAATACTACCAACACCAATACAAC | ATAAGTATGAGGGTTTTAGAT | 60 |
| cg21944234 | AGTAAGGTTGATTAATAGAAGAGTTAGAT | [Btn]ACCAAAAATTCTTTAACTAAACTAACACT | TTTTGTTTTGGTATTAGTTGT | 60 |

**Table A.5 Primer sequences and conditions used for PCR and methylation quantification.** PCR primer sequences and conditions used for methylation quantification by pyrosequencing; primers target bisulphite converted DNA.

| Gene | Forward primer (5′ - 3′) | Reverse primer (5′ - 3′) |
|---|---|---|
| *CCL5* | CCCTCGCTGTCATCCTCAT | AGTGGGCGGGCAATGTA |
| *CXCR2* | TCTGGATGCCACCGAGATTCT | AGTCCATGGCGAAACTTC |
| *IL1A* | AACCAGTGCTGCTGAAGGA | TTCTTAGTGCCGTGAGTTTCC |
| *IL6* | GGTACATCCTCGACGGCATCT | GTGCCTCTTTGCTGCTTTCAC |
| *TNF* | GGAGAAGGGTGACCGACTCA | CTGCCCAGACTCGGCAA |

**Table A.6 Primer sequences used for measuring inflammatory gene expression by real-time PCR using SYBR® green chemistry.**

## Appendix B. Patient details and genotypes

| Patient | Sex | Age (years) | Joint | rs4836732 genotype |
|---|---|---|---|---|
| 669 | F | 75 | H | TT |
| 988 | M | 69 | K | CT |
| 1059 | F | 77 | H | CC |
| 1349 | F | 73 | K | CC |
| 1350 | M | 71 | K | CC |
| 1520 | F | 73 | K | CT |
| 1595 | F | 51 | H | CT |
| 1604 | M | 71 | H | TT |
| 1651 | F | 60 | K | CT |
| 1658 | F | 67 | K | CT |
| 1724 | M | 74 | K | TT |
| 1744 | M | 50 | K | CC |
| 1820 | F | 76 | H | TT |
| 1874 | F | 71 | K | CT |
| 1922 | F | 67 | H | TT |
| 1939 | F | 70 | H | CT |
| 1951 | F | 70 | H | CT |
| 1955 | F | 60 | H | CC |
| 3028 | F | 78 | H | TT |
| 3033 | F | 81 | K | TT |
| 3060 | M | 76 | K | TT |
| 3085 | F | 55 | K | CT |
| 3244 | M | 57 | K | CC |
| 3288 | M | 57 | H | CT |
| 3330 | M | 69 | K | CT |
| 3335 | M | 74 | K | CT |
| 3340 | F | 67 | K | CT |
| 3343 | F | 69 | K | CT |
| 3355 | M | 57 | K | CT |
| 3365 | F | 61 | H | CT |
| 3377 | F | 66 | K | CT |
| 3379 | M | 63 | K | CT |
| 3388 | M | 82 | K | CC |
| 3389 | u | u | K | CT |
| 3390 | F | 78 | K | CT |
| 3391 | M | 82 | K | TT |
| 3396 | M | 56 | K | TT |
| 3399 | F | 54 | K | CT |
| 3411 | M | 71 | K | TT |

**Table B.1 Characteristics and genotypes of patients whose cartilage was used for gene expression analysis.** Donors denoted by a ‘T’ in the patient number are NOF patients, all other donors are OA patients. Age (age at joint replacement); Joint (joint replaced). (u) indicates information unavailable. Continued overleaf.

| Patient | Sex | Age (years) | Joint | rs4836732 genotype |
|---|---|---|---|---|
| 3417 | F | 58 | H | CT |
| 3452 | F | 69 | H | CT |
| 3472 | M | 63 | K | CC |
| 3475 | F | 71 | H | CT |
| 3478 | M | 70 | K | TT |
| 3502 | M | 67 | K | TT |
| 3510 | F | 80 | K | TT |
| 3555 | M | 71 | K | CC |
| 3558 | F | 46 | K | CC |
| 3661 | F | 62 | K | CT |
| 3664 | F | 58 | K | CT |
| 3666 | M | u | K | CT |
| 3667 | M | 59 | K | TT |
| 3672 | M | 64 | K | CT |
| 3682 | F | 81 | K | CT |
| 3686 | F | 80 | K | CC |
| 3697 | F | 64 | K | CC |
| 3700 | F | 78 | K | CT |
| 3701 | F | 61 | K | CC |
| 3703 | F | 80 | K | CC |
| 3705 | F | 80 | K | CC |
| 3714 | F | 59 | K | CT |
| 3727 | F | 71 | H | CT |
| 3729 | M | 74 | K | TT |
| 3771 | M | 72 | K | CT |
| 3772 | M | 72 | K | CC |
| 3773 | M | 68 | H | CC |
| 3777 | F | 72 | H | CC |
| 3802 | F | 63 | H | CT |
| 3834 | M | 77 | K | CT |
| 3838 | M | 71 | H | CC |
| 3862 | M | 71 | H | CC |
| 3912 | F | 73 | K | TT |
| 3916 | F | 82 | K | CT |
| 3920 | M | 71 | K | CT |
| 3921 | M | 80 | K | TT |
| 3930 | M | 66 | H | CT |
| 3931 | M | 69 | K | TT |
| 3933 | F | 58 | K | CT |

**Table B.1 Characteristics and genotypes of patients whose cartilage was used for gene expression analysis.** Donors denoted by a ‘T’ in the patient number are NOF patients, all other donors are OA patients. Age (age at joint replacement); Joint (joint replaced). (u) indicates information unavailable. Continued overleaf.

| Patient | Sex | Age (years) | Joint | rs4836732 genotype |
|---|---|---|---|---|
| 3975 | F | 60 | K | TT |
| 4012 | F | 62 | K | CT |
| 4064 | F | 67 | H | CT |
| 207218 | M | 68 | K | CT |
| 333031 | M | 75 | K | CT |
| T007 | F | 71 | H | CT |
| T012 | F | 81 | H | TT |
| T013 | F | 72 | H | CT |
| T014 | F | 84 | H | TT |
| T018 | F | 94 | H | CT |
| T020 | F | 84 | H | CC |
| T021 | F | 84 | H | CT |
| T023 | F | 52 | H | CC |
| T024 | F | 80 | H | CT |
| T028 | F | u | H | CC |
| T031 | F | 89 | H | TT |
| T034 | F | u | H | CT |
| T035 | F | 82 | H | CC |
| T037 | F | u | H | CT |
| T059 | F | 82 | H | TT |

**Table B.1 Characteristics and genotypes of patients whose cartilage was used for gene expression analysis.** Donors denoted by a ‘T’ in the patient number are NOF patients, all other donors are OA patients. Age (age at joint replacement); Joint (joint replaced). (u) indicates information unavailable.

| Patient | Sex | Age (years) | Joint | rs4836732 genotype | rs2565 genotype | rs7518 genotype | rs2281627 genotype |
|---|---|---|---|---|---|---|---|
| 1828 | F | 72 | K | CT | | | CT |
| 1874 | F | 71 | K | CT | | AG | CT |
| 1888 | F | 42 | K | CT | GT | | |
| 1902 | M | 61 | K | TT | | | CT |
| 3055 | M | 69 | K | CT | GT | | CT |
| 3082 | M | 82 | K | CC | | | CT |
| 3085 | F | 55 | K | CT | | AG | CT |
| 3136 | F | 81 | K | CT | GT | AG | CT |
| 3210 | M | 75 | K | TT | GT | | |
| 3281 | F | 64 | K | CC | GT | | |
| 3297 | F | 78 | K | CC | GT | | |
| 3305 | F | 60 | K | CT | GT | | |
| 3312 | M | 63 | K | CC | GT | | |
| 3330 | M | 69 | K | CT | GT | | |
| 3335 | M | 75 | K | CT | GT | | CT |
| 3348 | M | 56 | K | CT | GT | | CT |
| 3383 | M | 77 | K | CT | GT | | |
| 3384 | F | 71 | K | CT | GT | | CT |
| 3410 | F | 61 | K | CT | GT | AG | |
| 3411 | M | 71 | K | TT | | | CT |
| 3458 | F | 71 | K | CC | GT | | |
| 3662 | M | 67 | K | CT | GT | AG | CT |
| 4244 | M | 78 | K | CT | GT | AG | CT |
| 4277 | F | 79 | K | CT | GT | | |
| 4278 | M | 67 | K | TT | GT | | CT |
| 4285 | F | 64 | K | CC | | AG | CT |
| 4286 | M | 65 | K | CT | | | CT |
| 4288 | M | 85 | K | CT | GT | | |

**Table B.2 Characteristics and genotypes of patients whose tissues were used for allelic expression imbalance analysis.**

Age (age at joint replacement); Joint (joint replaced). Individuals heterozygous for the *PAPPA* (rs2565), *ASTN2* (rs7518) or *TRIM32* (rs2281627) were analysed. Continued overleaf.

| Patient | Sex | Age (years) | Joint | rs4836732 genotype | rs2565 genotype | rs7518 genotype | rs2281627 genotype |
|---|---|---|---|---|---|---|---|
| 4290 | M | 71 | H | TT | | | CT |
| 4295 | F | 68 | H | TT | GT | | CT |
| 4304 | M | 70 | K | CT | | AG | CT |
| 4308 | F | 55 | K | CT | | AG | CT |
| 4356 | F | 48 | K | CC | | AG | |
| 4383 | F | 64 | K | CC | GT | | |
| 4388 | M | 67 | K | TT | | | CT |
| 4432 | F | 72 | K | CT | GT | AG | |
| 4448 | F | 61 | K | CT | | | CT |
| 4452 | M | 60 | K | CT | | AG | |
| 4457 | F | 54 | K | TT | GT | | |
| 4459 | F | 76 | K | TT | GT | | CT |
| 4459 | F | 76 | K | TT | GT | | |
| 4475 | F | 70 | K | CT | GT | AG | CT |
| 4480 | F | 71 | K | CT | | | CT |
| 4483 | F | 72 | K | CT | GT | | |
| 4495 | F | 78 | K | CT | | AG | CT |
| 4507 | F | 73 | K | TT | | AG | |
| 4510 | F | 68 | K | CC | | AG | |
| 4545 | M | 81 | K | CT | GT | AG | CT |
| 4552 | F | 82 | K | CT | GT | AG | CT |
| 4569 | F | 68 | K | TT | GT | | CT |
| 4576 | F | 81 | K | CT | | | CT |
| 4580 | F | 70 | K | TT | | | CT |
| 4588 | F | 34 | K | TT | | AG | |
| 4608 | M | 57 | K | CT | | | CT |
| 4631 | F | 73 | H | CT | GT | AG | CT |
| 4637 | F | 79 | K | CC | | AG | |

**Table B.2 Characteristics and genotypes of patients whose tissues were used for allelic expression imbalance analysis.**

Age (age at joint replacement); Joint (joint replaced). Individuals heterozygous for the *PAPPA* (rs2565), *ASTN2* (rs7518) or *TRIM32* (rs2281627) were analysed. Continued overleaf.

| Patient | Sex | Age (years) | Joint | rs4836732 genotype | rs2565 genotype | rs7518 genotype | rs2281627 genotype |
|---|---|---|---|---|---|---|---|
| 4639 | F | 78 | K | TT | GT | AG | |
| 4645 | F | 77 | K | CC | GT | | |
| 4657 | M | 69 | K | CC | | AG | |
| 4659 | F | 62 | K | CT | | | CT |
| 4660 | F | 69 | K | CC | GT | | |
| 4661 | F | 79 | K | CC | | AG | |
| 4664 | M | 69 | K | CT | GT | AG | CT |
| 4676 | M | 67 | K | TT | GT | | |
| 4706 | M | 65 | K | CT | | AG | |
| 4708 | M | 77 | K | TT | GT | AG | |
| 4761 | F | 82 | K | TT | GT | AG | |
| 4773 | F | 76 | K | TT | GT | | CT |
| 4803 | M | 54 | K | TT | | AG | CT |
| 4804 | M | 54 | K | TT | | AG | CT |
| 4805 | F | 65 | K | TT | | | CT |
| 4836 | M | 73 | K | TT | | | CT |
| 4849 | F | 67 | H | CT | GT | AG | |
| 4851 | M | 58 | K | CC | GT | AG | |
| 4864 | F | 67 | H | TT | | | CT |
| 4901 | F | 68 | K | TT | GT | | |
| 4923 | M | 82 | K | CC | | AG | |
| 4962 | M | 67 | K | CT | GT | | CT |
| 4963 | F | 68 | H | CT | GT | | |
| 4966 | F | 87 | K | CT | | AG | |
| 4984 | M | 62 | K | CT | GT | AG | |
| 5000 | M | 54 | K | CT | | | CT |
| 5009 | F | 54 | K | CT | | | CT |
| 5019 | M | 50 | K | CT | | | CT |

**Table B.2 Characteristics and genotypes of patients whose tissues were used for allelic expression imbalance analysis.**

Age (age at joint replacement); Joint (joint replaced). Individuals heterozygous for the *PAPPA* (rs2565), *ASTN2* (rs7518) or *TRIM32* (rs2281627) were analysed. Continued overleaf.

| Patient | Sex | Age (years) | Joint | rs4836732 genotype | rs2565 genotype | rs7518 genotype | rs2281627 genotype |
|---|---|---|---|---|---|---|---|
| 5023 | M | 75 | K | CC | GT | | |
| 5050 | F | 61 | K | CC | GT | AG | |
| 5097 | M | 66 | K | CT | | AG | |
| 5139 | F | 64 | K | CT | | AG | CT |
| 5160 | M | 74 | H | CT | | AG | CT |
| 5171 | F | 68 | K | TT | GT | AG | |
| 5240 | F | 61 | K | CT | | | CT |
| 5254 | F | 62 | K | CT | | AG | CT |
| 5271 | F | 68 | K | CC | GT | AG | CT |
| 5281 | F | 59 | K | CT | | AG | CT |
| 5291 | M | 77 | K | CT | GT | AG | CT |
| 5299 | F | 75 | K | CT | GT | | |
| 5300 | M | 87 | K | CC | | AG | |
| 5380 | M | 72 | K | CC | | AG | |
| 5482 | M | 91 | K | CC | | AG | CT |
| 5504 | M | 68 | K | CT | | | CT |
| 5509 | M | 62 | K | CT | | AG | CT |
| 5541 | M | 53 | K | CT | GT | | |
| 5550 | M | 57 | K | TT | | | CT |
| 5551 | M | 49 | K | CT | GT | | |
| 5569 | F | 63 | K | TT | GT | | CT |
| 5613 | M | 75 | K | CC | GT | AG | |
| 5643 | M | 76 | K | TT | GT | | |
| 5645 | M | 76 | H | CT | | AG | |
| 5648 | F | 66 | K | TT | GT | | |
| 5652 | F | 82 | K | CC | | AG | |
| 5655 | F | 49 | H | CT | | AG | |
| 5667 | F | 68 | H | CT | | AG | |

**Table B.2 Characteristics and genotypes of patients whose tissues were used for allelic expression imbalance analysis.**

Age (age at joint replacement); Joint (joint replaced). Individuals heterozygous for the *PAPPA* (rs2565), *ASTN2* (rs7518) or *TRIM32* (rs2281627) were analysed. Continued overleaf.

| Patient | Sex | Age (years) | Joint | rs4836732 genotype | rs2565 genotype | rs7518 genotype | rs2281627 genotype |
|---|---|---|---|---|---|---|---|
| 5677 | M | 54 | K | TT | GT | | |
| 5679 | F | 65 | H | CT | | AG | |
| 5680 | F | 70 | K | TT | | | CT |
| 5681 | F | 51 | K | CC | | | CT |
| 5694 | F | 81 | K | CC | GT | | |
| 5710 | F | 70 | K | CC | | AG | |
| 5725 | M | 81 | K | TT | | | CT |
| 5729 | F | 48 | K | CC | | AG | CT |
| 5747 | F | 87 | K | TT | | | CT |
| 5748 | M | 67 | K | CC | | AG | |
| 5809 | M | 79 | K | CC | | AG | |
| 5834 | M | 71 | K | TT | | | CT |
| 333031 | M | 75 | K | CT | | AG | CT |
| 406459 | | | K | CT | GT | | CT |
| 8157410 | F | 82 | K | CT | | AG | |

**Table B.2 Characteristics and genotypes of patients whose tissues were used for allelic expression imbalance analysis.**

Age (age at joint replacement); Joint (joint replaced). Individuals heterozygous for the *PAPPA* (rs2565), *ASTN2* (rs7518) or *TRIM32* (rs2281627) were analysed.

| Patient | Sex | Age (years) | Joint | BMI |
|---|---|---|---|---|
| 5989 | M | 68 | H | 24.6 |
| 5996 | F | 82 | H | 24.1 |
| 6011 | M | 66 | H | 26.0 |
| 6013 | F | 70 | H | 30.3 |
| 6027 | F | 46 | H | 26.1 |
| 6031 | M | 66 | H | 23.4 |
| 6033 | F | 61 | H | 33.6 |
| 6089 | M | 77 | H | 27.3 |
| 6105 | M | 81 | H | 21.0 |
| 6127 | F | 87 | H | 28.7 |
| 6132 | F | 61 | H | 37.3 |
| 6136 | F | 57 | H | 33.9 |
| 6148 | M | 50 | H | 27.8 |
| 6163 | F | 75 | H | 27.0 |
| 6166 | M | 65 | H | 27.4 |
| 6191 | M | 72 | H | 24.8 |
| 6207 | M | 67 | H | 33.0 |
| 6226 | M | 58 | H | 38.6 |
| 6229 | F | 63 | H | 31.2 |

**Table B.3 Characteristics of patients for whom matched cartilage and blood serum were analysed.** Age (age at joint replacement); Joint (joint replaced); BMI (body mass index at time of pre-operative assessment). Continued overleaf.

| Patient | Sex | Age (years) | Joint | BMI |
|---|---|---|---|---|
| 5994 | M | 58 | K | 37.4 |
| 6002 | M | 58 | K | 37.9 |
| 6003 | M | 65 | K | 32.1 |
| 6006 | F | 75 | K | 28.0 |
| 6007 | F | 64 | K | 37.7 |
| 6008 | M | 68 | K | 27.2 |
| 6012 | F | 79 | K | 41.3 |
| 6015 | M | 74 | K | 23.6 |
| 6054 | F | 60 | K | 33.4 |
| 6057 | F | 76 | K | 24.9 |
| 6090 | F | 72 | K | 40.0 |
| 6095 | M | 84 | K | 25.9 |
| 6104 | F | 53 | K | 25.2 |
| 6114 | F | 56 | K | 33.0 |
| 6146 | F | 72 | K | 28.6 |
| 6150 | F | 78 | K | 29.1 |
| 6153 | F | 72 | K | 19.7 |
| 6164 | M | 73 | K | 26.1 |
| 6165 | M | 64 | K | 29.3 |
| 6175 | M | 80 | K | 30.4 |
| 6190 | F | 70 | K | 32.0 |
| 6203 | M | 47 | K | 35.5 |
| 6204 | F | 71 | K | 29.0 |
| 6208 | F | 73 | K | 24.1 |
| 6209 | M | 73 | K | 39.7 |
| 6223 | M | 78 | K | 29.8 |
| 6225 | M | 60 | K | 35.8 |
| 6227 | M | 56 | K | 30.5 |
| 6228 | M | 78 | K | 23.4 |

**Table B.3 Characteristics of patients for whom matched cartilage and blood serum were analysed.** Age (age at joint replacement); Joint (joint replaced); BMI (body mass index at time of pre-operative assessment).

## Appendix C. Antibodies

| Antibody | Cat. number |
|---|---|
| E2A | sc-349 |
| FOXP1 | sc-66896 |
| GR | sc-1003 |
| GRα | Sc-1002 |
| HEB | sc-357 |
| HNF4α | sc-8987 |
| IRF2 | sc-498 |
| KLF16 | sc-131168 |
| LEF1 | sc-28687 |
| MEF2A | sc-10794 |
| MEF2C | sc-13268 |
| NFATC3 | sc-1152 |
| NFkB | sc-372 |
| NFYA | sc-10779 |
| Oct2 | sc-25400 |
| PAPPA | sc-365226 |
| Pax9 | sc-7746 |
| Poc4 | sc-48778 |
| PR | sc-538 |
| Sox9 | sc-20095 |
| Sp3 | sc-644 |
| TCF4 | sc-13027 |
| XBP1 | sc-7160 |

**Table C.1 Antibodies used for supershift EMSAs to investigate protein binding to rs4836732.** All antibodies were purchased from Santa Cruz Biotechnology (USA).

## Appendix D. siRNAs

| siRNA | Cat. number | siRNA target sequences |
|---|---|---|
| SMARTpool: ON-TARGETplus *PAPPA* siRNA | L-005130-00-0005 | GAACCAAGGUGAUAGAUCU |
| | | AGACAYGGAUCUAAAUCUU |
| | | GUACAGAGGGCAAGUGGAA |
| | | GCAAAGUGCUCAUGUUAGG |
| ON-TARGETplus Non-targeting Pool | D-001810-10-20 | UGGUUUACAUGUCGACUAA |
| | | UGGUUUACAUGUUGUGUGA |
| | | UGGUUUACAUGUUUUCUGA |
| | | UGGUUUACAUGUUUUCCUA |

**Table D.1 siRNA (Dharmacon, GE Healthcare, UK) used for the knockdown of *PAPPA*.**

## Presentations and publications

Presentations

- April 2014: World Congress on Osteoarthritis, Paris, France (*poster presentation*).
- September 2014: UK-German Connective Tissue Meeting, Allendale, UK (*oral presentation*).
- September 2015: 21st Northern and Yorkshire Rehumatology Meeting, York, UK (*oral presentation*).
- October 2015: North East Postgraduate Conference, Newcastle Upon Tyne, UK (*oral presentation*).
- February 2016: MRC Dermatology/Rheumatology Proximity to Discovery (PtD) week, Newcastle Upon Tyne, UK (*poster presentation*).
- April 2016: World Congress on Osteoarthritis, Amsterdam, Netherlands (*poster presentation*).
- April 2017: World Congress on Osteoarthritis, Las Vegas, USA (*oral presentation*).

Publications

- Rogers, E.L, Reynard, L.N., Loughlin, J. (2015) The role of inflammation-related genes in osteoarthritis. Osteoarthritis Cartilage 23(11):1933-8.
- Rogers, E.L., Reynard, L.N., Rankin, K., Loughlin, J. (2016) Functional Investigation of an OA susceptibility locus at chromosome 9q33.1 associated with female hip osteoarthritis. Presented at the 2016 World Congress on Osteoarthritis. Osteoarthritis and Cartilage 24:S232.
- Rogers, E.L, Lendrem, D. Using methylation data for the predictive modelling of increased cartilage inflammation in hip osteoarthritis (*manuscript in preparation*).
- Rogers E.L., Reynard, L., Loughlin, J. Functional analysis of the osteoarthritis susceptibility locus marked by the polymorphism rs4836732 (*manuscript in preparation*).

- Rogers, E.L., Reynard, L., Loughlin, J. Investigation into the metabolic status and systemic inflammation in a subgroup of osteoarthritis patients displaying elevated cartilage inflammation (*manuscript in preparation*).

## References


1. Arthritis Research UK. What is Arthritis? [Online] Available at: http://www.arthritisresearchuk.org/arthritis-information/conditions/arthritis/what-is-arthritis.aspx. [Accessed 23rd June 2016].
2. Woo, S.Y., Fenwick, J.A., Kanamori, A., Gil, J.E., Chan Saw, S.S., Vogrin, T.M. (2001) Biomechanical considerations of joint function. In: Rubash, H.E. Osteoarthritis: Diagnosis and Medical/Surgical Management. 3rd Ed, Philadelphia: WB Saunders.
3. Iwanaga, T., Shikichi, M., Kitamura, H., Yanase, H., Nozawa-Inoue, K. (2000) Morphology and functional roles of synoviocytes in the joint. Arch Histol Cytol 63(1):17-31.
4. Loeser, R., Goldring, S., Scanzello, C., Goldring, M. (2012) Osteoarthritis: A Disease of the Joint as an Organ. Arthritis and Rheumatism 64(6): 1697-1707.
5. Scanzello, C.R., Goldring, S.R. (2012) The role of synovitis in osteoarthritis pathogenesis. Bone 51(2):249-57.
6. Ioan-Facsinay, A., Kloppenburg M. (2013) An emerging player in knee osteoarthritis: the infrapatellar fat pad. Arthritis Res Ther 15(6):225.
7. Suri S. and Walsh D.A. (2012) Osteochondral alterations in osteoarthritis. Bone 51: 204-211.
8. Goldring, M. (2012) Chondrogenesis, chondrocyte differentiation, and articular cartilage metabolism in health and osteoarthritis. Therapeutic Advances in Musculoskeletal Disease 4(4): 269–285.
9. Knudson, C.B., Knudson, W. (2001) Cartilage proteoglycans. Semin Cell Dev Biol 12: 69-78.
10. Grogan, S.P., Duffy, S.F., Pauli, C., Koziol, J.A., Su, A.I., D'Lima, D.D.,Lotz, M.K. (2013) Zone specific gene expression patterns in articular cartilage. Arthritis and Rheumatism 65(2):418-28.
11. Li, G., Yin, J., Gao, J., Cheng, T.S., Pavlos, N. J., Zhang, C., Zheng, M. H. (2013) Subchondral bone in osteoarthritis: insight into risk factors and microstructural changes. Arthritis Res Ther 15(6): 223.
12. Buckwalter, J.A., Cooper, R.R. (1987) Bone structure and function. Instr Course Lect 36: 27-48.

13. Holder, N. (1977) An experimental investigation into the early development of the chick elbow joint. J Embryol Exp Morphol 39:115-27.
14. Guo, X., Day, T.F., Jiang, X., Garrett-Beal, L., Topol, L., Yang, Y. (2004) Wnt/beta-catenin signaling is sufficient and necessary for synovial joint formation. Genes and Development 18(19):2404-17.
15. Archer, C.W., Dowthwaite, G.P., Francis-West, P. (2003) Development of Synovial Joints. Birth Defects Research 69:144-155.
16. Koyama, E., Shibukawa, Y., Nagayama, M., Sugito, H., Young, B., Yuasa, T., *et al*. (2008) A distinct cohort of progenitor cells participates in synovial joint and articular cartilage formation during mouse limb skeletogenesis. Birth Defects Research (Part C) 69:144–15.
17. Rivas, R., Shapiro, F. (2002) Structural stages in the development of the long bones and epiphyses: a study in the New Zealand white rabbit. J Bone Joint Surg Am 84-A(1):85-100.
18. Mort, J.S., Billington, C.J. (2001) Articular cartilage and changes in arthritis: matrix degradation. Arthritis Res 3(6):337-41.
19. Ellman, M.B., An, H.S., Muddasani, P., Im, H., (2008) Biological impact of the fibroblast growth factor family on articular cartilage and intervertebral disc homeostasis. Gene 420(1): 82–89.
20. Fortier, L.A., Barker, J.U., Strauss, E.J., McCarrel, T.M., Cole, B.J. (2011) The role of growth factors in cartilage repair. Clin Orthop Relat Res 469(10):2706-15.
21. Fox, A. J. S., Bedi, A., Rodeo, S. A. (2009) The Basic Science of Articular Cartilage. Sports Health 1(6): 461–468.
22. 22. Mort, J.S., Geng, Y., Fisher, W.D., Roughley, P.J. (2016) Aggrecan heterogeneity in articular cartilage from patients with osteoarthritis. BMC Musculoskelet Disord. 17:89.
23. Liu, W., Zhang, X. (2015) Receptor activator of nuclear factor-κB ligand (RANKL)/RANK/osteoprotegerin system in bone and other tissues (review). Mol Med Rep. 11(5):3212-8.
24. Zhen, G, Cao, Z. (2014) Targeting TGFβ Signaling in Subchondral Bone and Articular Cartilage Homeostasis. Trends Pharmacol Sci 35(5): 227–236.
25. Cawston, T.E., Young, D.A. (2010) Proteinases involved in matrix turnover during cartilage and bone breakdown. Cell Tissue Res 339(1):221-35.

26. NICE. NICE clinical guidelines: CG177 Osteoarthritis: care and management (2014) [Online] Available at https://www.nice.org.uk/guidance/cg177 [Accessed 20th June 2016].
27. Arthritis Research UK, 2013. Osteoarthritis in general practise. Chesterfield: Arthritis Research UK.
28. Oliveria, S.A., Felson, D.T., Reed, J.I., Cirillo, P.A., Walker, A.M. (1995) Incidence of symptomatic hand, hip, and knee osteoarthritis among patients in a health maintenance organization. Arthritis Rheum 38(8):1134-41.
29. Bijlsma, J.W., Knahr, K. (2007) Strategies for the prevention and management of osteoarthritis of the hip and knee. Best Pract Res Clin Rheumatol 21(1):59-76.
30. 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.
31. Arthritis Care, 2012. Annual Report & Accounts 2012. London: Arthritis Care.
32. Haq, S.A., Davatchi, F. (2011) Osteoarthritis of the knees in the COPCORD world. Int J Rheum Dis 14(2):122-9.
33. Wieland, H.A., Michaelis, M., Kirschbaum, B.J., Rudolphi, K.A. (2005) Osteoarthritis: an untreatable disease? Nat Rev Drug Discov 4: 331-344.
34. Junker, S., Krumbholz, G., Frommer, K.W., Rehart, S., Steinmeyer, J., Rickert, M., *et al*. (2016) Differentiation of osteophyte types in osteoarthritis - proposal of a histological classification. Joint Bone Spine 83(1):63-7.
35. Loeser, R.F. (2014) Integrins and chondrocyte-matrix interactions in articular cartilage. Matrix Biol. 39:11-6.
36. Li, G., Yin, J., Gao, J., Cheng, T.S., Pavlosm N.J., Zhang, C., Zheng, M.H. (2013) Subchondral bone in osteoarthritis: insight into risk factors and microstructural changes. Arthritis Res Ther. 15(6):223.
37. Barr, A.J., Campbell, T.M., Hopkinson, D., Kingsbury, S.R., Bowes, M.A., Conaghan, P.G. (2015) A systematic review of the relationship between subchondral bone features, pain and structural pathology in peripheral joint osteoarthritis. Arthritis Res Ther.17:228.
38. van der Kraan, P.M., van den Berg, W.B. (2007) Osteophytes: relevance and biology. Osteoarthritis Cartilage 15(3):237-44.

39. Krasnokutsky, S., Belitskaya-Lévy, I., Bencardino, J., Samuels, J., Attur, M., Regatte, R., Rosenthal, P., Greenberg, J., Schweitzer, M., Abramson, S.B., Rybak, L. (2011) Quantitative magnetic resonance imaging evidence of synovial proliferation is associated with radiographic severity of knee osteoarthritis. Arthritis Rheum 63(10):2983-91.
40. Pearle, A.D., Scanzello, C.R., George, S., Mandl, L.A., DiCarlo, E.F., Peterson, M., Sculco, T.P., Crow, M.K. (2007) Elevated high-sensitivity C-reactive protein levels are associated with local inflammatory findings in patients with osteoarthritis. Osteoarthritis Cartilage 15(5):516-23.
41. Fernandes, J.C., Martel-Pelletier, J., Pelletier, J.P. (2002) The role of cytokines in osteoarthritis pathophysiology. Biorheology 39: 237-246.
42. Rogers, E.L., Reynard, L.N., Loughlin, J. (2015) The role of inflammation-related genes in osteoarthritis. Osteoarthritis Cartilage 23(11):1933-8.
43. Rushton, M.D., Reynard, L.N., Barter, M.J., Refaie, R., Rankin, K.S., Young, D.A., Loughlin, J. (2014) Characterization of the cartilage DNA methylome in knee and hip osteoarthritis. Arthritis Rheumatol 66(9):2450-60.
44. Felson, D.T. (2006) Clinical practice. Osteoarthritis of the knee. New England Journal of Medicine 54:841-848.
45. Arden, N. and Nevitt, M.C. (2006) Osteoarthritis: epidemiology. Best Pract Res Clin Rheumatol 20: 3-25.
46. Altman, R., Asch, E., Bloch, D., Bole, G., Borenstein, D., Brandt, K., Christy, W., Cooke, T. D., Greenwald, R., Hochberg, M., Howell, D., Kaplan, D., Koopman, W., Longley, S., Mankin, H., McShane, D. J., Medsger, T., Meenan, R., Mikkelsen, W., Moskowitz, R., Murphy, W., Rothschild, B., Segal, M., Sokoloff, L., Wolfe, F. (1986) Development of criteria for the classification and reporting of osteoarthritis: Classification of osteoarthritis of the knee. Arthritis Rheum 29: 1039-1049.
47. Kellgren, J.H., Lawrence J.S. (1957) Radiographic assessment of osteoarthritis. Ann Rheum Dis. 16: 494-502.
48. Bellamy, N., Buchanan, W.W., Goldsmith, C.H., Campbell, J., Stitt, L.W. (1988) Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 15(12):1833-40.

49. Pulsatelli, L., Addimanda, O., Brusi, V., Pavloska, B. and Meliconi, R. (2013) New findings in osteoarthritis pathogenesis: therapeutic implications. Therapeutic Advances in Chronic Disease 4(1):23-43.
50. Zhang, W., Moskowitz, R.W., Nuki, G., Abramson, S., Altman R.D., Arden, N., *et al*. (2007) OARSI recommendations for the management of hip and knee osteoarthritis, Part I: Critical appraisal of existing treatment guidelines and systematic review of current research evidence. Osteoarthritis and Cartilage 15(9):981-1000.
51. Lane, N.E. (1997) Pain management in osteoarthritis: the role of COX-2 inhibitors. J Rheumatol Suppl 49:20-4.
52. Doherty, M., Hawkey, C., Goulder, M., Gibb, I., Hill, N., Aspley, S. and Reader, S. (2011) A randomised controlled trial of ibuprofen, paracetamol or a combination tablet of ibuprofen/paracetamol in community-derived people with knee pain. Arthritis & Rheumatology 70(9):1534-41.
53. Hochberg, M.C., Martel-Pelletier, J., Monfort, J., Möller, I., Castillo, J.R., Arden, N., Berenbaum, F., Blanco, F.J., Conaghan, P.G., Doménech, G., Henrotin, Y., Pap, T., Richette, P., Sawitzke, A., du Souich, P., Pelletier, J.P., MOVES Investigation Group. (2016) Combined chondroitin sulfate and glucosamine for painful knee osteoarthritis: a multicentre, randomised, double-blind, non-inferiority trial versus celecoxib. Ann Rheum Dis 75(1):37-44.
54. TenBroek, E.M., Yunker, L., Nies, M.F., Bendele, A.M. (2016) Randomized controlled studies on the efficacy of antiarthritic agents in inhibiting cartilage degeneration and pain associated with progression of osteoarthritis in the rat. Arthritis Res Ther 21;18:24.
55. Baragi, V.M., Baragi, V.M., Becher, G., Bendele, A.M., Biesinger, R., Bluhm, H., Boer, J., Deng, H., Dodd, R., Essers, M., Feuerstein, T., Gallagher, B.M Jr., Gege. C., Hochgürtel, M., Hofmann, M., Jaworski, A., Jin, L., Kiely, A., Korniski, B., Kroth, H., Nix, D., Nolte, B., Piecha, D., Powers, T.S., Richter, F., Schneider, M., Steeneck, C., Sucholeiki, I., Taveras, A., Timmermann, A., Van Veldhuizen, J., Weik, J., Wu, X., Xia, B. (2009) A new class of potent matrix metalloproteinase 13 inhibitors for potential treatment of osteoarthritis: Evidence of histologic and clinical efficacy without musculoskeletal toxicity in rat models. Arthritis Rheum 60: 2008-2018.

56. Birrell, F., Howells, N. and Porcheret, M. (2011) Osteoarthritis: pathogenesis and prospects for treatment. Arthritis Research UK Topical Reviews, Reports on the Rheumatic Diseases 6(10).
57. Lundblad, H., Kreichbergs, A. and Jansson, K.A. (2008) Prediction of persistent pain after total knee replacement for osteoarthritis. Journal of Bone and Joint Surgery 90(2):166-71.
58. 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: 81-88.
59. Sabzevari, S., Ebrahimpour, A., Roudi, M.K., Kachooei, A.R. (2016) High Tibial Osteotomy: A Systematic Review and Current Concept. Arch Bone Jt Surg. 4(3):204-12.
60. Lohmander, L.S., Englund, P.M., Dahl, L.L., Roos, E.M. (2007) The long-term consequence of anterior cruciate ligament and meniscus injuries: osteoarthritis. Am J Sports Med 35(10):1756-69.
61. Wong, J.M., Jayadey, C.S., Khan, W., Johnstone, D. (2012) Anterior cruciate ligament rupture and osteoarthritis progression. Open Orthopaedic Journal 6: 295-300.
62. Ratzlaff, C.R., Liang, M.H. (2010) New developments in osteoarthritis. Prevention of injury-related knee osteoarthritis: opportunities for the primary and secondary prevention of knee osteoarthritis. Arthritis Res Ther 12(4):215.
63. Zhang, Y. and Jordan, J.M. (2010) Epidemiology of osteoarthritis. Clinical Geriatric Medicine 26(3): 355-69.
64. Reginster, J.Y., Kvasz, A., Bruyere, O., Henrotin, Y. (2003) Is there any rationale for prescribing hormone replacement therapy (HRT) to prevent or to treat osteoarthritis? Osteoarthritis Cartilage 11(2):87-91.
65. Murphy, L., Schwartz, T.A., Helmick, C.G., Renner, J.B., Tudor, G., Koch, G. *et al*. (2008) Lifetime risk of symptomatic knee osteoarthritis. Arthritis & Rheumatology 59:1207e13.
66. Yusuf, E. (2012) Metabolic factors in osteoarthritis: obese people do not walk on their hands. Arthritis Res Ther 4(4):123.
67. Felson, D.T., Nevitt, M.C., Zhang, Y., Aliabadi, P., Baumer, B., Gale, D., Li, W., Yu, W., Xu, L. (2002) High prevalence of lateral knee osteoarthritis in

Beijing Chinese compared with Framingham Caucasian subjects. Arthritis Rheum 46(5):1217-22.

68. Hawker, G.A., Croxford, R., Bierman, A.S., Harvey, P.J., Ravi, B., Stanaitis, I., Lipscombe, L.L. (2014) All-cause mortality and serious cardiovascular events in people with hip and knee osteoarthritis: a population based cohort study. PLoS One 9(3):e91286.
69. Stecher, R.M. (1941) HEBERDEN'S NODES: Heredity in Hypertrophic Arthritis of the Finger Joints. Am J Med Sci 201: 801-809.
70. Goldring, M.B. and Goldring, S. R. (2007) Osteoarthritis. Journal of Cellular Physiology 213(3):62634.
71. Francis-West, P.H., Abdelfattah, A., Chen, P., Allen, C., Parish, J., Ladher, R., Allen, S., MacPherson, S., Luyten, F.P., Archer, C.W. (1999) Mechanisms of GDF-5 action during skeletal development. Development 126: 1305-1315.
72. Miyamoto, Y., Mabuchi, A., Shi, D., Kubo, T., Takatori, Y., Saito, S., Fujioka, M., Sudo, A., Uchida, A., Yamamoto, S., Ozaki, K., Takigawa, M., Tanaka, T., Nakamura, Y., Jiang, Q., Ikegawa, S. (2007) A functional polymorphism in the 5´UTR of GDF5 is associated with susceptibility to osteoarthritis. Nat Genet 39: 529-533.
73. 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: 2226-2232.
74. Rodriguez-Fontenla, C., Calaza, M., Evanelou, E., *et al*. (2014) Assessment of Osteoarthritis Candidate Genes in a Meta-Analysis of Nine Genome-Wide Association Studies. Arthritis & Rheumatology 66(4):940-9.
75. International HapMap Consortium. (2003) The International HapMap Project. Nature 426: 789-796.
76. Human Genome Sequencing. (2004) Finishing the euchromatic sequence of the human genome. Nature 431: 931-945.
77. Reynard, L.N. (2016) Analysis of genetics and DNA methylation in osteoarthritis: What have we learnt about the disease? Semin Cell Dev Biol. pii: S1084-9521(16)30121-5.

78. Styrkarsdottir, U., Thorleifsson, G., Helgadottir, H.T., Bomer, N., Metrusty, S., Bierma-Zeinstra, S. *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(5):498-502.
79. Valdes, A.M., Evangelou, E., Kerkof, H.J, Tamm, A., Doherty, S.A., Kisand, K. *et al*. (2011). The GDF5 rs143383 polymorphism is associated with osteoarthritis of the knee with genome-wide statistical significance. Annals of Rheumatic Diseases 70(5):873-5.
80. Evangelou, E., Kerkhof, H.J., Styrkarsdottir, U., Ntzani, E.E., Bos, S.D., Esko, T., *et al*. (2014). A meta-analysis of genome-wide association studies identifies novel variants associated with osteoarthritis of the hip. Annals of Rheumatic Diseases 73(12):2130-6.
81. Evangelou, E., Valdes, A.M., Kerkhof, H.J., Styrkarsdottir, U., Zhu, Y., Meulenbelt, I., *et al*. (2011) Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. Annals of Rheumatic Diseases 70(2):349-55.
82. Day-Williams, A.G., Southam, L., Panoutsopoulou, K., Rayner, N.W., Esko, T., Estrada, K., *et al*. (2011) A variant in MCF2L is associated with osteoarthritis. Am J Hum Genet. 2011 Sep 9;89(3):446-50.
83. Miyamoto, Y., Shi, D., Nakajima, M., Ozaki, K., Sudo, A., Kotani, A., *et al*. (2008) Common variants in DVWA on chromosome 3p24.3 are associated with susceptibility to knee osteoarthritis. Nat Genet. 40(8):994-8.
84. Nakajima, M., Takahashi, A., Kou, I., Rodriguez-Fontenla, C., Gomez-Reino, J.J., Furuichi, T., Dai, J., Sudo, A., Uchida, A., Fukui, N., Kubo, M., Kamatani, N., Tsunoda, T., Malizos, K.N., Tsezou, A., Gonzalez, A., Nakamura, Y., Ikegawa, S. (2010) New sequence variants in hla class ii/iii region associated with susceptibility to knee osteoarthritis identified by genome-wide association study. PLoS ONE 18;5(3):e9723.
85. Evangelou, E., Valdes, A.M., Castano-Betancourt, M.C., Doherty, M., Doherty, S., Esko, T., *et al*. (2013) The DOT1L rs12982744 polymorphism is associated with osteoarthritis of the hip with genome-wide statistical significance in males. Ann Rheum Dis. 72(7):1264-5.
86. Kerkhof, H.J., Lories, R.J., Meulenbelt, I., Jonsdottir, I., Valdes, A.M., Arp, P., *et al*. (2010) A genome-wide association study identifies an osteoarthritis

susceptibility locus on chromosome 7q22. Arthritis and Rheumatism 62(2):499-510.

87. Valdes, A.M., Spector, T.D., Tamm, A., Kisand, K., Doherty, S.A., Dennison, E.M., *et al*. (2010) Genetic variation in the SMAD3 gene is associated with hip and knee osteoarthritis. Arthritis Rheum. 62(8):2347-52.
88. Meulenbelt, I., Min, J.L., Bos, S., Riyazi, N., Houwing-Duistermaat, J.J., van der Wijk, H.J., *et al*. (2008) Identification of DIO2 as a new susceptibility locus for symptomatic osteoarthritis. Hum Mol Genet. 17(12):1867-75.
89. Nakamura, T., Shi, D., Tzetis, M., Rodriguez-Lopez, J., Miyamoto, Y., Tsezou, A., *et al*. (2007) Meta-analysis of association between the ASPN D-repeat and osteoarthritis. Hum Mol Genet. 6(14):1676-81.
90. Mototani, H., Iida, A., Nakajima, M., Furuichi, T., Miyamoto, Y., Tsunoda, T., *et al*. (2008) A functional SNP in EDG2 increases susceptibility to knee osteoarthritis in Japanese. Hum Mol Genet. 17(12):1790-7.
91. 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: 864-867.
92. arcOGEN Consortium: arcOGEN collaborators (2012) Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. Lancet 380: 815-823.
93. Wilson, P., Fryer, R., Fang, Y., Hatten, M. (2010) Astn2, A Novel Member of the Astrotactin Gene Family, Regulates the Trafficking of ASTN1 During Glial-Guided Neuronal Migration. Journal of Neuroscience 30(25):8529-40.
94. Bis, C., DeCarli, C., Smith, A.V., van der Lijn, F., Crivello, F., Fornage, M., Debette, S. *et al*. (2011) Common variants at 12q14 and 12q24 are associated with hippocampal volume. Nature Genetics 44(5): 545–551.
95. Wang, K., Liu, X., Aragam, N. (2010) A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. Schizophrenia Research 124(1-3):192-9.
96. Fridell, R., Harding, L., Bogerd, H., Cullen, B. (1995) Identification of a novel human zinc finger protein that specifically interacts with the activation domain of lentiviral Tat proteins. Virology 209(2):347-57.

97. Frosk. P, Weiler. T, Nylen. E, Sudha. T, Greenberg. C.R., Morgan. K., Fujiwara. T.M., Wrogemann. K. (2002) Limb-girdle muscular dystrophy type 2H associated with mutation in TRIM32, a putative E3-ubiquitin-ligase gene. American Journal of Human Genetics 70(3):663-7.
98. Schoser, B.G., Frosk, P., Engel, A.G., Klutzny, U., Lochmüller, H., Wrogemann, K. (2005) Commonality of TRIM32 mutation in causing sarcotubular myopathy and LGMD2H. Annals of Neurology 57(4):591-5.
99. Nicklas, S., Otto, A., Wu, X., Miller, P., Stelzer, S., Wen, Y., Kuang, S., Wrogemann, K., Patel, K., Ding, H., Schwamborn, J.C. (2012). TRIM32 regulates skeletal muscle stem cell differentiation and is necessary for normal adult muscle regeneration. PLoS One 7(1):e30445.
100. Kudryashova, E., Wu, J., Havton, L.A., Spencer, M.J. (2009) Deficiency of the E3 ubiquitin ligase TRIM32 in mice leads to a myopathy with a neurogenic component. Human Molecular Genetics 18(7):1353-67.
101. Schwamborn, J., Berezikov, E., Knoblich, J. (2009) The TRIM-NHL Protein TRIM32 Activates MicroRNAs and Prevents Self-Renewal in Mouse Neural Progenitors. Cell 136(5): 913–925.
102. Gonzalez-Cano, L., Hillje, L., Fuertes-Alvarez, L., Marques, M., Blanch, A., Ian, R., Irwin, M. S., Schwamborn, J., Marín, M. (2013) Regulatory feedback loop between TP73 and TRIM32. Cell Death and Disease 4(7): e704.
103. Lionel, A.C., Tammimies, K., Vaags, A.K., Rosenfeld, J.A., Ahn, J.W., Merico, D., *et al*. (2014) Disruption of the ASTN2/TRIM32 locus at 9q33.1 is a risk factor in males for autism spectrum disorders, ADHD and other neurodevelopmental phenotypes. Hum Mol Genet 23(10):2752-68.
104. Kuhajda, F., Eggleston, J. (1985) Pregnancy-associated plasma protein A. A clinically significant predictor of early recurrence in stage I breast carcinoma is independent of estrogen receptor status. American Journal of Pathology 121(2): 342–348.
105. Christiansen, M., Jaliashvili, I. (2003) Total pregnancy-associated plasma protein A-a first trimester maternal serum marker for Down's syndrome: clinical and technical assessment of a poly-monoclonal enzyme immunoassay. Scand J Clin Lab Invest 63(6):407-15.
106. Qin, X., Wergedal, J.E., Rehage, M., Tran, K., Newton, J., Lam, P., Baylink, D.J., Mohan, S. (2006) Pregnancy-associated plasma protein-A increases

osteoblast proliferation *in vitro* and bone formation *in vivo*. Endocrinology 147(12):5653-61.

107. Tanner, S.J., Hefferan, T.E., Rosen, C.J., Conover, C.A. (2008) Impact of pregnancy-associated plasma protein-a deletion on the adult murine skeleton. Journal of Bone and Mineral Research 23(5):655-62.
108. Denninger, K.C., Litman, T., Marstrand, T., Moller, K., Svensson, L., Labuda, T., Andersson, Å. (2015) Kinetics of gene expression and bone remodelling in the clinical phase of collagen-induced arthritis. Arthritis Res Ther 5;17:43.
109. Allis, C.D., Jenuwein, T. (2016) The molecular hallmarks of epigenetic control. Nat Rev Genet. 17(8):487-500.
110. Dupont, C., Armant, D.R., Brenner, C.A. (2009) Epigenetics: definition, mechanisms and clinical perspective. Semin Reprod Med 27(5):351-7.
111. Shao, Y., Chernaya, V., Johnson, C., Yang, W.Y., Cueto, R., Sha, X., Zhang, Y., Qin, X., Sun, J., Choi, E.T., Wang, H., Yang, X.F. (2016) Metabolic Diseases Downregulate the Majority of Histone Modification Enzymes, Making a Few Upregulated Enzymes Novel Therapeutic Targets-"Sand Out and Gold Stays". J Cardiovasc Transl Res 9(1):49-66.
112. Peña-Altamira, L.E., Polazzi, E., Monti, B. (2013) Histone post-translational modifications in Huntington's and Parkinson's diseases. Curr Pharm Des19(28):5085-92.
113. Lewis, B.P., Burge, C.B., Bartel, D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120(1):15-20.
114. Barter, M.J., Bui, C., Young, D.A. (2011) Epigenetic mechanisms in cartilage and osteoarthritis: DNA methylation, histone modifications and microRNAs. Osteoarthritis Cartilage 20(5):339-49.
115. Yu, X.M., Meng, H.Y., Yuan, X.L., Wang, Y., Guo, Q.Y., Peng, J., *et al*. (2015) MicroRNAs' Involvement in Osteoarthritis and the Prospects for Treatments. Evid Based Complement Alternat Med. 236179.
116. Zhang, M., Wang, J. (2015) Epigenetics and Osteoarthritis. Genes Dis. 2(1):69-75.
117. den Hollander, W., Meulenbelt, M. (2015) DNA Methylation in Osteoarthritis. Curr Genomics. 16(6): 419–426.

118. Alvarez-Garcia, O., Fisch, K.M., Wineinger, N.E., Akagi, R., Saito, M., Sasho, T., *et al*. (2016) Increased DNA Methylation and Reduced Expression of Transcription Factors in Human Osteoarthritis Cartilage. Arthritis Rheumatol. 68(8):1876-86.
119. den Hollander, W., Ramos, Y.F., Bos, S.D., Bomer, N., van der Breggen, R., Lakenberg, N., *et al*. (2014) Knee and hip articular cartilage have distinct epigenomic landscapes: implications for future cartilage regeneration approaches. Ann Rheum Dis. 73(12):2208-12.
120. Reynard, L.N., Bui, C., Syddall, C.M., Loughlin, J. (2014) CpG methylation regulates allelic expression of *GDF5* by modulating binding of SP1 and SP3 repressor proteins to the osteoarthritis susceptibility SNP rs143383. Hum Genet. 133(8):1059-73.
121. Rushton, M.D., Young, D.A., Loughlin, J., Reynard, L.N. (2015) Differential DNA methylation and expression of inflammatory and zinc transporter genes defines subgroups of osteoarthritic hip patients. Ann Rheum Dis 74(9):1778-82.
122. Fernandez-Tajes, J., Soto-Hermida, A., Vazquez-Mosquera, M.E., Cortes-Pereira, E., Mosquera, A., Fernandez-Moreno, M., *et al*. (2014) Genome-wide DNA methylation analysis of articular chondrocytes reveals a cluster of osteoarthritic patients. Ann Rheum Dis 73:668e77.
123. Rozen, S., Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 132: 365-86.
124. Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., Haussler, D. (2002) The human genome browser at UCSC. Genome Res. 12: 996-1006.
125. Vincze, T., Posfai, J., Roberts, R. J. (2003) NEBcutter: A program to cleave DNA with restriction enzymes. Nucleic Acids Res. 31: 3688-91.
126. Messeguer, X., Escudero, R., Farre, D., Nunez, O., Martinez, J., Alba, M. M. (2002) PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics. 18: 333-4.
127. Farre, D., Roset, R., Huerta, M., Adsuara, J. E., Rosello, L., Alba, M. M., Messeguer, X. (2003) Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. Nucleic Acids Res. 31: 3651-3.

128. Encode Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature. 489: 57-74.
129. Johnson A. D., Handsaker R. E., Pulit S., Nizzari M. M., O'Donnell C. J., de Bakker P.I.W. (2008) SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics, 24: 2938-2939.
130. Song, L., 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.
131. Wang, J., Zhuang, J., Iyer, S., Lin, X. Y., Greven, M. C., Kim, B. H., *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.
132. Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., *et al*. (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 473: 43-9.
133. Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29: 308-11.
134. 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.
135. Nica, A. C., Dermitzakis, E. T. (2013) Expression quantitative trait loci: present and future. Philos Trans R Soc Lond B Biol Sci. 368: 20120362.
136. Jacobsen, S., Sonne-Holm, S. (2005) Hip dysplasia: a significant risk factor for the development of hip osteoarthritis. A cross-sectional survey. Rheumatology (Oxford) 44(2):211-8.
137. Lindner, C., Thiagarajah, S., Wilkinson, J.M., Panoutsopoulou, K., Day-Williams, A.G.; arcOGEN Consortium, Cootes, T.F., Wallis, G.A. (2015) Investigation of association between hip osteoarthritis susceptibility loci and radiographic proximal femur shape. Arthritis Rheumatol. 67(8):2076-84.
138. Ramos, Y., den Hollander, W., Bovee, J., Bomer, N., van der Breggen, R., Lakenberg, N., *et al*. (2014) Genes involved in the osteoarthritis process identified through genome wide expression analysis in articular cartilage; the RAAK study. Arthritis Rheumatol 63:391e400.

139. Birmingham, E., Niebur, G.L., McHugh, P.E., Shaw, G., Barry, F.P., McNamara, L.M. (2012) Osteogenic differentiation of mesenchymal stem cells is regulated by osteocyte and osteoblast cells in a simplified bone niche. Eur Cell Mater. 23:13-27.
140. Westhrin, M., Xie, M., Olderøy, M.Ø., Sikorski, P., Strand, B.L., Standal, T. (2015) Osteogenic differentiation of human mesenchymal stem cells in mineralized alginate matrices. PLoS One 10(3):e0120374.
141. Takahashi K, Yamanaka S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126:663–676.
142. James, C.G., Appleton, C.T., Ulici, V., Underhill, T.M., Beier, F. (2005) Microarray analyses of gene expression during chondrocyte differentiation identifies novel regulators of hypertrophy. Mol Biol Cell. 16(11):5316-33.
143. Papadimitriou, N., Thorfve, A., Brantsing, C., Junevik, K., Baranto, A., Barreto Henriksson, H. (2014) Cell viability and chondrogenic differentiation capability of human mesenchymal stem cells after iron labeling with iron sucrose. Stem Cells Dev. 23(21):2568-80.
144. Dreier, R.(2010) Hypertrophic differentiation of chondrocytes in osteoarthritis: the developmental aspect of degenerative joint disorders. Arthritis Res Ther. 12(5):216.
145. Hino, K., Saito, A., Kido, M., Kanemoto, S., Asada, R., Takai, T., *et al*. (2014) Master regulator for chondrogenesis, Sox9, regulates transcriptional activation of the endoplasmic reticulum stress transducer BBF2H7/CREB3L2 in chondrocytes. J Biol Chem. 289(20):13810-20.
146. Poonpet, T., Honsawek, S. (2014) Adipokines: Biomarkers for osteoarthritis? World J Orthop. 5(3): 319–327.
147. Ding, Y., Li, J., Liu, S., Zhang, L., Xiao, H., Li, J., *et al*. (2014) DNA hypomethylation of inflammation-associated genes in adipose tissue of female mice after multigenerational high fat diet feeding. Int J Obes (Lond). (2):198-204.
148. Lee YH, Bae SC. (2016) Circulating leptin level in rheumatoid arthritis and its correlation with disease activity: a meta-analysis. Z Rheumatol. [Epub ahead of print].

149. Nielung, N., Christensen, R., Danneskiold-Samsøe, B., Bliddal, H., Cato Holm, C., Ellegaard, K., *et al*. (2015) Validity and Agreement between the 28-Joint Disease Activity Score Based on C-Reactive Protein and Erythrocyte Sedimentation Rate in Patients with Rheumatoid Arthritis. Arthritis. 2015: 401690.
150. van Dielen, F.M., van't Veer, C., Schols, A.M., Soeters, P.B., Buurman, W.A., Greve, J.W. (2001) Increased leptin concentrations correlate with increased concentrations of inflammatory markers in morbidly obese individuals. Int J Obes Relat Metab Disord. 25(12):1759-66.
151. Lönnqvist, F., Wennlund, A., Arner, P. (1997) Relationship between circulating leptin and peripheral fat distribution in obese subjects. Int J Obes Relat Metab Disord. 21(4):255-60.
152. Slieker, R.C., Bos, S.D., Goeman, J.J., Bovée, J.V.M.G., Talens, R.P., van der Breggen, R. *et al*. (2013) Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. Epigegenetics Chromatin 6(1):26.
153. Walton, E., Hass, J., Liu, J., Roffman, J.L., Bernardoni, F., Roessner, V. *et al*. (2016) Correspondence of DNA methylation between blood and brain tissue and its application to Schizophrenia research. Schizophr Bull. 42(2):406-14.
154. Freitag,J., Bates, D., Boyd, R., Shah, K., Barnard, A., Huguenin, L., Tenen, A. (2016) Mesenchymal stem cell therapy in the treatment of osteoarthritis: reparative pathways, safety and efficacy – a review. BMC Musculoskelet Disord. 17: 230.
155. Sharma, L. (2016) Osteoarthritis year in review 2015: clinical. Osteoarthritis Cartilage. 24(1):36-48.
156. Gee, F., Clubbs, C. F., Raine, E. V., Reynard, L. N., 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.
157. Reynard, L.N., Ratnayake, M., Santibanez-Koref, M., Loughlin, J. (2016) Functional Characterization of the Osteoarthritis Susceptibility Mapping to CHST11-A Bioinformatics and Molecular Study. PLoS One 11(7):e0159024.

158. Johnson, K., Reynard, L.N., Loughlin, J. (2014) Functional analysis of the osteoarthritis susceptibility locus marked by the polymorphism rs10492367. Osteoarthritis and Cartilage 22: S236 - S237.
159. Johnson, K., Reynard, L.N., Loughlin, J. (2015) Functional characterisation of the osteoarthritis susceptibility locus at chromosome 6q14.1 marked by the polymorphism rs9350591. BMC Med Genet. 16:81.
160. Lettre, G., Jackson, A.U., Gieger, C., Schumacher, F.R., Berndt, S.I., Sanna, S., *et al*. (2008) Identification of ten loci associated with height highlights new biological pathways in human growth. Nat Genet. 40(5):584-91.
161. van Duyvenvoorde, H.A., Lui, J.C., Kant, S.G., Oostdijk, W., Gijsbers, A.C., Hoffer, M.J. *et al*. (2014) Copy number variants in patients with short stature. Eur J Hum Genet. 22(5):602-9.
162. Vecellio, M., Roberts, A.R., Cohen, C.J., Cortes, A., Knight, J.C., Bowness, P., Wordsworth, B.P. (2016) The genetic association of RUNX3 with ankylosing spondylitis can be explained by allele-specific effects on IRF4 recruitment that alter gene expression. Ann Rheum Dis. 75(8):1534-40.
163. Lawrenson, K., Li, Q., Kar, S., Seo, J., Tyrer, J., Spindler, T.J. *et al*. (2015) Cis-eQTL analysis and functional validation of candidate susceptibility genes for high-grade serous ovarian cancer. Nat Commun. 6: 8234.
164. Ratnayake, M., Reynard, L. N., Raine, E. V., Santibanez-Koref, M., Loughlin, J. (2012) Allelic expression analysis of the osteoarthritis susceptibility locus that maps to MICAL3. BMC Med Genet. 13: 12.
165. Raine, E. V., Wreglesworth, N., Dodd, A. W., Reynard, L. N., 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.
166. Syddall, C. M., Reynard, L. N., Young, D. A., 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.
167. Glimm, A.M., Werner, S.G., Burmester, G.R., Backhaus, M., Ohrndorf, S. (2015) Analysis of distribution and severity of inflammation in patients with osteoarthitis compared to rheumatoid arthritis by ICG-enhanced fluorescence optical imaging and musculoskeletal ultrasound: a pilot study. Ann Rheum Dis doi:10.1136/annrheumdis-2015-207345.

168. Kim, J.H., Jeon, J., Shin, M., Won, Y., Lee, M., Kwak, J.S. *et al.* (2014) Regulation of the catabolic cascade in osteoarthritis by the zinc-ZIP8-MTF1 axis. Cell 156(4):730-43.
169. Kraus, V.B. (2014) Osteoarthritis: The zinc link. Nature 507(7493):441-2.
170. Philp, A.M., Davis, E.T., Jones, S.W. (2016) Developing anti-inflammatory therapeutics for patients with osteoarthritis. Rheumatology (Oxford) pii: kew278.
171. Armour, A.A., Watkins, C.L. (2010) The challenge of targeting EGFR: experience with gefitinib in nonsmall cell lung cancer. Eur Respir Rev. 19(117):186-96.
172. Sander, J. D., Joung, J. K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol. 32: 347-55.
173. Plank, J.L., Dean, A. (2014) Enhancer function: mechanistic and genome-wide insights come together. Mol Cell. 55(1):5-14.
174. Smemo, S., Tena, J.J., Kim, K.H., Gamazon, E.R., Sakabe, N.J., Gómez-Marín, C. *et al.* (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. Nature 507(7492):371-5.
175. Ramos, Y.F., Meulenbelt, I. (2015) Implementation of Functional Genomics for Bench-to-Bedside Transition in Osteoarthritis. Curr Rheumatol Rep.17(8):53.
176. Conover, C.A., Bale, L.K., Powell, D.R. (2013) Inducible Knock Out of Pregnancy-Associated Plasma Protein-A Gene Expression in the Adult Mouse: Effect on Vascular Injury Response. Endocrinology 154(8):2734-8.
177. Bale, L.K., Chakraborty, S., Conover, C.A. (2014) Inducible reduction in pregnancy-associated plasma protein-A gene expression inhibits established atherosclerotic plaque progression in mice. Endocrinology 155(4):1184-7.