Expression analysis of osteoarthritis susceptibility loci

Emma V. A. Raine

Institute of Cellular Medicine

Thesis submitted for the degree of Doctor of Philosophy

June 2014

Abstract

Osteoarthritis (OA) is a common disease characterised by the progressive loss of the articular cartilage of synovial joints. It is multifactorial and polygenic. Candidate gene-based association studies and genome-wide association scans (GWAS) have been used to identify OA risk alleles with over 10 regions of the human genome so far reported as harbouring OA susceptibility.

Expression quantitative trait loci (eQTLs) demonstrate a correlation between gene expression levels and DNA polymorphism, the assumption being that the polymorphism resides within a regulatory element of the gene. The functional polymorphism can be *cis*-acting (within or close to the gene) or *trans*-acting (located a distance from the gene).

It has become apparent for many common diseases that the majority of risk alleles are eQTLs rather than polymorphisms that alter amino acid sequences.

The aim of this thesis was to investigate whether genes proximal to or harbouring OA-associated polymorphisms identified by GWAS are subject to *cis*-regulation, which could therefore be contributing to the association signal.

The expression levels of the genes were analysed in joint tissues from OA patients who had undergone joint replacement. Overall gene expression was measured by quantitative PCR whilst allelic expression was assessed using assays that can distinguish mRNA output from each allele of a transcript SNP. The functional effect of SNPs was tested *in vitro* using luciferase reporter assays.

Expression analysis of genes at the OA-associated chromosome 7q22 locus identified *HBP1* as the likely target of the OA susceptibility mapping to this region. Expression analysis of *SMAD3* and *GNL3* identified eQTLs active in OA joint tissues for both genes, whilst luciferase reporter assays identified two 3´UTR SNPs as potential mediators of the *SMAD3* eQTL.

Overall, this thesis demonstrates that OA tissues are subject to a range of regulatory elements that can influence disease during development or ageing.

Dedication

I dedicate this thesis to my parents and my husband for all the love and support they have given me throughout.

Acknowledgments

I would like to thank my supervisor Professor John Loughlin for all his support and encouragement throughout my 6 years in his group and especially for all his help and advice in writing this thesis. I would also like to thank my second supervisor Dr Louise Reynard for all her suggestions and her invaluable experimental help and advice.

I would like to thank all members of the Osteoarthritis Genetics Group and the Musculoskeletal Research Group past and present.

Thank you to Drs David Young and Matthew Barter for kind donation of stem cells, Drs Aida Bertoli-Avella and Ingrid van de Laar for kind donation of vascular smooth muscle cells, and to Professor Jeremy Lakey for modelling the GNL3 protein.

A huge thank you to Andrew Waugh for his wonderful illustrations.

Finally, I'd like to thank Newcastle University and NIHR Newcastle Biomedical Research Centre for funding this research.

Table of Contents

List of Tables and Figures

Tables

Figures

Abbreviations

ADAMTS	A Disintegrin and Metalloproteinase with Thrombospondin Motifs
AEI	Allelic Expression Imbalance
AOS	Aneurysms and Osteoarthritis Syndrome
APS	Ammonium Persulfate
ATP	Adenosine Triphosphate
BMD	Bone Mineral Density
BME	β-Mercaptoethanol
BMP	Bone Morphogenic Protein
Bn	Cancellous Bone
BSA	Bovine Serum Albumin
cDNA	Complementary DNA
Cn	Osteoarthritic Cartilage
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Ct	Cycle Threshold
COX	Cyclooxygenase
ddNTP	Dideoxynucleotide Triphosphate
DEPC	Diethylpyrocarbonate
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl Sulfoxide
dNTP	Deoxynucleotide Triphosphate
DSB	Double Strand Breaks
DTT	Dithiothreitol
ECM	Extra Cellular Matrix
EDTA	Ethylenediaminetetraacetic Acid
EGTA	Ethylene Glycol Tetraacetic Acid
EMSA	Electrophoretic Mobility Shift Assay
ENCODE	Encyclopaedia of DNA Elements
eQTL	Expression Quantitative Trait Locus
FBS	Fetal Bovine Serum
Fp	Fat Pad
FTAAD	Familial Thoracic Aortic Aneurysm and Aortic Dissections
GDF	Growth Differentiation Factor

gDNA	Genomic DNA
GWAS	Genome Wide Association Scan
GWS	Genome Wide Significant
HAC	Human Articular Chondrocytes
HDR	Homology-Directed Repair
HMG	High Mobility Box Group
IGF	Insulin-Like Growth Factor
IL	Interleukin
IPFP	Infrapatellar Fat Pad
JSW	Joint Space Width
LB	Lysogeny Broth
LD	Linkage Disequilibrium
LEF	Lymphoid Enhancer-Binding Factor
Li	Ligament
MAPK	Mitogen-Activated Protein Kinase
Me	Meniscus
miRNA	MicroRNA
MMP	Matrix Metalloproteinase
MRI	Magnetic Resonance Imaging
mRNA	Messenger RNA
MSC	Mesenchymal Stem Cells
NCAM	Neural Cell Adhesion Molecule
NGF	Nerve Growth Factor
NHEJ	Non-Homologous End Joining
NICE	National Institute for Health and Care Excellence
NOF	Neck-of-Femur
NOS	Nitric Oxide Synthase
NSAID	Non Steroidal Anti Inflammatory
OA	Osteoarthritis
OARSI	Osteoarthritis Research Society International
OR	Odds Ratio
OSM	Oncostatin M
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction

PGE	Prostaglandin E
PKA	Protein Kinase A
PLB	Passive Lysis Buffer
PTH	Parathyroid Hormone
PVDF	Polyvinylidene Fluoride
qPCR	Quantitative PCR
RFLP	Restriction Fragment Length Polymorphism
rRNA	Ribosomal RNA
RT-PCR	Reverse Transcriptase PCR
SAP	Shrimp Alkaline Phosphatase
SBE	Single Base Extension
SDS	Sodium Dodecyl Sulfate
siRNA	Small interfering RNA
SNP	Single Nucleotide Polymorphism
SOC	Super Optimal Broth with Catabolite Repression
Sy	Synovium
TALENS	Transcription Activator-Like Effector Nuclease
TBS-T	Tris-Buffered Saline-Tween
TCF	T-Cell Factor
TBE	Tris Borate EDTA
Te	Tendon
TEMED	Tetramethylethylenediamine
TGF-β	Transforming Growth Factor- β
THR	Total Hip Replacement
TJR	Total Joint Replacement
TKR	Total Knee Replacement
TNF	Tumour Necrosis Factor
U	Units
UTR	Untranslated Region
VEGF	Vascular Endothelial Growth Factor
VSMC	Vascular Smooth Muscle Cell
WT	Wild Type
ZFN	Zinc-Finger Nucleases

Chapter 1: General Introduction

1.1 Introduction to Osteoarthritis

Osteoarthritis (OA) affects nearly 9 million people in the UK, over half of which are affected by OA of the knee (1). In 2012 an estimated 2.3 million working days were lost to musculoskeletal disorders of the lower limbs, the most common of which is OA (2). In the same year, over 200,000 people claimed incapacity or severe disablement benefit for musculoskeletal disorders (3) and over 160,000 joint replacement surgeries took place (4). As a result of these factors, OA is of considerable economic burden to the UK.

OA is characterised by progressive destruction of articular cartilage, but it affects the entire joint, including the synovial membrane, joint capsule, ligaments, muscles, tendons and subchondral bone, and can occur at any site but is most common in the hand, knee, hip and foot (5). In 2013 it was estimated that 4.71 million people in the UK had knee OA, 2.12 million had hip OA, 1.77 million had foot and ankle OA, and 1.56 had hand and wrist OA (6). The eventual destruction of cartilage results from a failure of chondrocytes to maintain a homeostatic balance between matrix synthesis and degradation (7). Clinical manifestations of OA include pain, stiffness and impairment of joint motion, which lead to disability and finally, the need for joint replacement (5).

1.2 Disease Definition and Diagnosis

OA is not a single disease entity but comprises a heterogeneous group of syndromes with differing underlying pathophysiological mechanisms (8). All the tissues of the joint are involved in OA (9), although the articular cartilage and the adjacent bone often show the most prominent changes (10). OA can be divided into primary and secondary forms; primary or idiopathic OA is defined as a condition developing in previously undamaged joints in the absence of an apparent triggering event, and secondary OA is caused by well-recognised predisposing conditions (10). These conditions include: anatomic abnormalities, such as congenital dislocation of the hip; trauma, such as joint

fracture; inflammatory disorders, such as septic arthritis; and metabolic disorders, such as haemochromatosis (11).

The epidemiological criteria for OA are based upon radiographically observed changes originally described by Kellgren and Lawrence in 1957 (12). The criteria are appearance of osteophytes, narrowing of the space between the joint resulting from the loss of articular cartilage, subchondral bone sclerosis, and cysts (11, 13-15). Studies have shown that there are many asymptomatic people with advanced radiographic OA and many others with severe symptoms but minimal pathology, so as the severity of OA pathology has little to do with the severity of the symptoms there are no strict diagnostic criteria for OA (16). The American College of Rheumatology has produced classification criteria based on differentiating OA from other rheumatic diseases, stating that 'anyone over the age of 50 with regional joint pain, short-lasting 'gelling' of joints and crepitus should be considered to have OA' (17). The initial clinical diagnosis of OA is usually made by general practitioners, predominantly on symptoms of pain, stiffness after inactivity, swelling and associated functional impairment (14, 15). Radiography is used to define OA for epidemiological studies and is the usual method of diagnosis, however inconsistencies in the descriptions of radiographic features of OA have led to studies being performed using criteria that are conflicting (11). Magnetic resonance imaging (MRI) has the advantage over radiography of being able to image all the structures within the OA joint, thereby improving our understanding of the association between structural changes and clinical symptoms, but MRI remains largely for use in research studies rather than clinical practice (18). Ultrasound has also provided greater insight into OA pathology than radiographic methods, although it does not assess cartilage surfaces as efficiently as MRI and cannot assess the subchondral bone (19).

1.3 Management and Treatment of OA

Since the relationship between the severity of radiographic changes seen in OA patients and the severity of their symptoms is weak, decisions about management of disease are usually not made on the severity of the pathology but on the degree of pain, disability and distress felt by the individual (16). Current treatments for OA focus on pain relief (symptom modifying drugs), rather than treating the cause of the disease (20, 21), since as yet there are no licenced treatments that can stop the progression of structural

changes within the joint (structure modifying drugs; 18). Current guidelines, set down by Osteoarthritis Research Society International (OARSI) and National Institute for Health and Care Excellence (NICE) for the treatment of OA, recommend a combination of both pharmacological and non-pharmacological treatments (22, 23). Pharmacological treatments include non-steroidal anti-inflammatory drugs (NSAIDs), opioids, corticosteroid injections and paracetamol. Non-pharmacological treatments include physiotherapy, splints, acupuncture and the use of a walking aid or brace (18). Current established treatments such as opioids, NSAIDs such as Ibuprofen, and Cyclooxygenase-2 (COX-2) inhibitors, are limited by side effects (22, 24), although COX-2 inhibitors have fewer gastrointestinal side effects in comparison to Ibuprofen (5, 16). Due to the adverse side effects of most pain-relieving drugs, there is a distinct lack of safe treatments for those with OA pain (18). Another issue facing many symptom-modifying drugs is their lack of efficacy, as demonstrated by intra-articular corticosteroid injections, the effectiveness of which is short-lived (25). Anti-synovial analgesic agents and antibodies against tumour necrosis factor (TNF)-α have however shown good efficacy for some OA patients (26). In addition to this, Duloxetine, a serotonin-noradrenaline reuptake inhibitor also shows a significant improvement in pain and function for people with painful knee OA who are already taking NSAIDs (27). While the presence of cartilage damage from the earliest stages of OA is undisputed, it seems unusual that this should result in pain, as healthy cartilage is aneural (5, 18). A number of studies have demonstrated that nerve and vascular in-growth occurs in damaged areas of OA cartilage at the osteochondral junction (28), which may explain the pain in OA. Increased nerve growth factor (NGF) expression has been observed in these vascular channels, and may offer a target for treatment by a novel class of drugs, the anti-NGFs, such as Tanezumb, which has demonstrated good analgesic efficacy and improvement in function in a study of people with knee OA (29).

Glucosamine and chondroitin are commonly used dietary supplements that are claimed to reduce the symptoms and retard the progression of OA (30). Another naturally occurring agent is hyaluronic acid, a glycosaminoglycan that makes up the viscous component of synovial fluid and acts as a lubricant and shock absorber (16). Injections of hyaluronic acid have been shown to improve the biomechanical function in OA (31).

Although there are currently no drugs available with proven structure-modifying efficacy in OA (18, 21, 32) there has been some research using Strontium ranelate, an

osteoporosis treatment, which can reduce bone resorption as well as increasing bone formation, and has been shown to cause a significant reduction in both joint space narrowing and pain (33, 34). Another study looking at Calcitonin, a thyroid hormone, in knee OA, showed improved knee function and increased cartilage volume compared to a placebo, although side effects were common (35).

Of the non-pharmacological treatments for OA, exercise is one of the most frequently recommended (16). Muscle strengthening and aerobic exercises have both been shown to lead to a moderate decrease in pain (36, 37). Other non-pharmacological approaches for treating cartilage damage, such as micro-fractures, abrasions or chondrocyte implantation require a regenerative response from the patient, and healing is often incomplete leaving cartilage susceptible to the degradation processes that lead to OA (38). One surgical treatment, joint distraction, has been effectively applied to treat OA. Joint distraction is based upon the theory that OA cartilage has some reparative activity when the damaged cartilage is no longer subject to mechanical loading (39). The joint is unloaded using an external frame which partially transfers the load through the frame rather than the bone, eliminating mechanical contact between the degenerating cartilage surfaces, preventing further wear and tear, whilst the pressure changes in the synovial fluid essential for the nutrition of cartilage are maintained (39, 40). Joint distraction has been used successfully in ankle OA, demonstrating decreased subchondral bone density that persisted for at least 2 years (41), and has been shown to reverse tissue structure damage in patients with knee OA who demonstrated decreased exposed bone area and joint space width (JSW) due to an increase in cartilage thickness (42). The side effects of joint distraction are stiffness of the joint, and the potential for pain or infection at the points at which the frame is affixed (43).

At the end stage of OA replacement of joints by endoprotheses is the last possible curative approach (20). However, there are many problems with total joint replacements including morbidity, and the need for revisions, resulting in an estimated 15% of patients being unsatisfied following joint replacement (44). Despite all the available therapies many people treated for OA still live with significant symptoms (5, 18).

1.4 Normal Structure and Function of the Synovial Joint

The function of the synovial joint is to facilitate articulation where two skeletal elements meet (45). Synovial joints are made up of articular cartilage, subchondral bone, ligament, tendon, synovial fluid, and fat pad and meniscus in the knee, enclosed in a sleeve of dense fibrous connective tissue known as the joint capsule (46). The ends of the opposing skeletal elements at the synovial joint are covered with a layer of articular cartilage that provides a low-friction, wear-resistant surface for load-bearing and smooth movement of the joints (9, 47).

1.4.1 Articular Cartilage

The articular cartilage is composed of a sparse population of cells named chondrocytes (7). Chondrocytes make up only 5% of the tissue and are relatively metabolically inactive due to the aneural and avascular nature of cartilage (5, 45) Chondrocytes are responsible for the synthesis and maintenance of the extra cellular matrix (ECM; 13). The ECM is composed of a high concentration of highly sulphated proteoglycans, mainly the large chondroitin-sulfate proteoglycan, aggrecan, which attaches to hyaluronic acid via link protein to form large aggregates (8, 20; Figure 1.1A). These are entangled in a dense network of type II collagen fibrils together with other collagens, predominantly type IX and type XI, and a large amount of water (8, 7). The composition of ECM gives articular cartilage its ability to function as a cushion, absorbing the impact of excessive loads (13, 47). Under compressive forces, water molecules are forced from the matrix until the repulsive forces between the proteoglycan molecules balance the load (9). On removing the load, the hydrophilic aggregates rehydrate, restoring the elasticity of the articular cartilage (8). This ability to support compressive forces is due to the stratified nature of articular cartilage (48), which is divided into four zones (Figure 1.1B): the superficial zone, the middle zone, the deep zone and the calcified cartilage zone (7, 8). Each zone is characterised by its organisation of the chondrocytes and collagen network, and by the composition and type of the proteoglycans (20). The superficial zone is at the surface of the articular cartilage, in contact with the synovial fluid. Here the chondrocytes are elongated and orientated with the collagen fibres parallel to the surface, the proteoglycan content is low and the water content is high. In the middle zone, the largest of the zones,

comprising 40-60% of the total articular cartilage, the chondrocytes are larger, more rounded in shape and randomly distributed throughout the collagen network. The chondrocytes of the deep zone are similar in size and shape to the middle zone, but are arranged with the collagen fibres in columns perpendicular to the surface. In the deep zone, collagen is at its thickest, the proteoglycan content is high and the chondrocytes are very sparse and hypertrophic. The deep zone is separated from the layer of calcified cartilage that sits above the subchondral bone by a 'tidemark' (7, 8, 14).

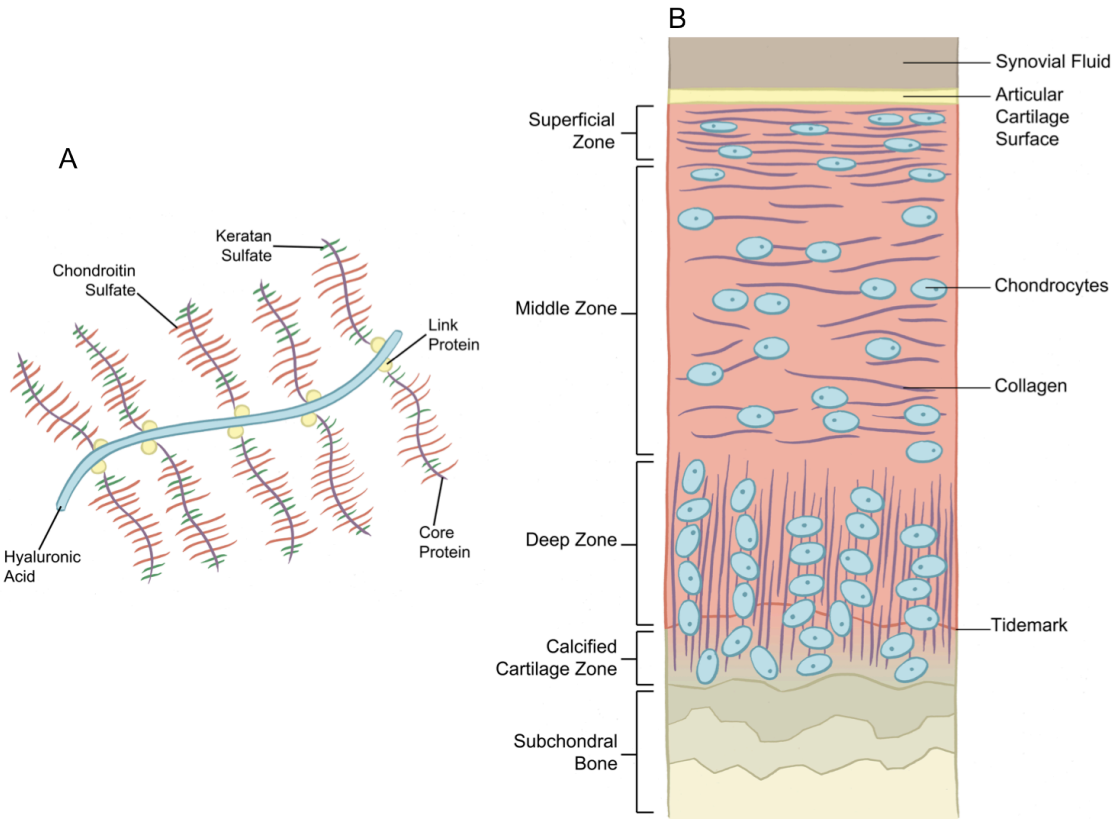

Figure 1.1. The structure of aggrecan and the cartilage extracellular matrix. A) The structure of aggrecan. Aggrecan is comprised of monomers non-covalently bound to a hyaluronic acid backbone. The small glycoprotein link protein helps to stabilize aggregate formation. The side chains are mostly chondroitin sulfate, but also include keratan sulfate. B) The chondrocyte and collagen composition of the four zones of articular cartilage: the superficial zone, the middle zone, the deep zone and the calcified cartilage zone. Figure reproduced with kind permission of Andrew Waugh.

1.4.2 Subchondral Bone

Subchondral bone is the term used for the region encompassing the subchondral bone plate, consisting of the relatively non-porous and poorly vascularised cortical bone, and the richly vascularised subchondral trabecular bone beneath it (49-51). As articular cartilage has no intrinsic vasculature or lymphatic system, it therefore relies on the

richly vascularised adjacent tissues, the subchondral trabecular bone and synovial membrane to provide nutrients that are essential for maintaining the health of the chondrocyte and articular cartilage (52, 53).

Bone is a specialised tissue made up of several components including: osteoblasts, which are responsible for bone formation; osteoclasts, which are responsible for degradation of bone matrix; non-collagenous substances such as proteoglycans; and a collagenous component of which type I collagen is the major constituent (51). In the adult skeleton, bone is in a state of dynamic equilibrium between bone formation and resorption, which is perfectly balanced i.e. bone resorption by osteoclasts, is exactly compensated by an equal amount of new bone deposition by osteoblasts (54). Following remodelling is the process of mineralisation (49). The mechanical properties of bone are dependent on the mineral content, and the degree of mineralisation is highly dependent on the rate of bone formation and remodelling (50).

1.4.3 Synovium

The synovial joint capsule consists of dense fibrous connective tissue that forms a sleeve around the articulating bones providing a seal that keeps synovial fluid in position around the cartilage (46). The synovial membrane lines the joint capsule and consists of two layers, a lining layer directly next to the joint space, named the intima, and an outer layer that merges with the fibrous capsule, named the subsynovial or sub-intima (51). The cells of the synovial membrane secrete the synovial fluid components, which contribute to the unique functional properties of articular surfaces and modulate chondrocyte activity (53). The synovial cells consist mainly of two main types: type A, macrophage-like synoviocytes and type B, the synovial fibroblasts (51). These cells produce two important synovial fluid components, proteoglycan 4, also known as lubricin, and hyaluronic acid (55). Together these two molecules reduce friction by providing boundary lubrication at the articular surface (53). The synovial membrane also functions as a semi-permeable membrane allowing the removal of the products of chondrocyte metabolism and matrix turnover (55).

1.4.4 Infrapatellar Fat Pad

The main role of the infrapatellar fat pad (IPFP) is to facilitate the distribution of synovial fluid and distribute mechanical forces through the knee (56). The IPFP is composed of a fibrous scaffold upon which adipose tissue is embedded (57). It is located in the space between the patellar tendon, the femoral condyles and the tibial plateau and its surface is covered by a layer of synovial membrane (58). Adipose tissue can function as an endocrine organ actively secreting cytokines such as TNFα and interleukin 6 (IL-6) in humans (58).

1.4.5 Meniscus

The menisci are fibrocartilaginous wedge shaped discs that are positioned on the tibial plateau and act as functional extensions of the tibia (59, 60). The meniscus is composed of a solid phase of collagen and ECM and a fluid phase consisting of water and electrolytes (59). Compressive forces facilitate the flow of fluid in and out of the meniscus lubricating the joint, providing nutrition and dissipating the compressive energy, making the meniscus is an integral part of the biomechanical system of the knee (51, 59).

1.4.6 Ligaments and Tendons

Ligament and tendons are viscoelastic structures with unique mechanical properties. Tendons serve primarily to transfer the pull of muscles to bone while ligaments connect one bone to another (61). Both ligaments and tendons are pliant and flexible, allowing natural movements of bones but are strong enough to sustain high tensile forces and offer resistance (62). Ligaments and tendons are composed of tightly packed collagen fibre bundles, mainly collagen I, and proteoglycan threads interspersed with fibroblasts and orientated in a parallel fashion to provide stability to joints (63, 64). There are four major ligaments in the knee: the medial and lateral collateral ligaments and the anterior and posterior cruciate ligaments (63). There are also ligaments providing peripheral attachments between the tibial plateau and both menisci preventing the menisci from being extruded from the joint during compression (60).

1.5 Development of the Synovial Joint

Long bones of the limbs are formed through a process named endochondral ossification, which consists of four stages (65; Figure 1.2). Initially, epithelium-mesenchymal interactions induce the existing mesenchymal cells to aggregate and form a densely packed cell condensation at the sites where bones will form (66, 38). The condensed mesenchymal cells then proliferate, followed by differentiation into hypertrophic chondrocytes, forming the cartilaginous anlagen of the future bones, which is then vascularised by the invasion of blood vessels from the perichondrium (66). Finally osteoblasts are transported into the cartilage by the blood vessels and replace the cartilage with mineralised trabecular bone (20, 65-69). This centre of mineralised bone splits into two opposite growth plates, where the maturation of cartilage continues and chondrocyte proliferation causes bone growth, separating the growth plates by an increasing amount of space that eventually becomes filled by bone marrow (68).

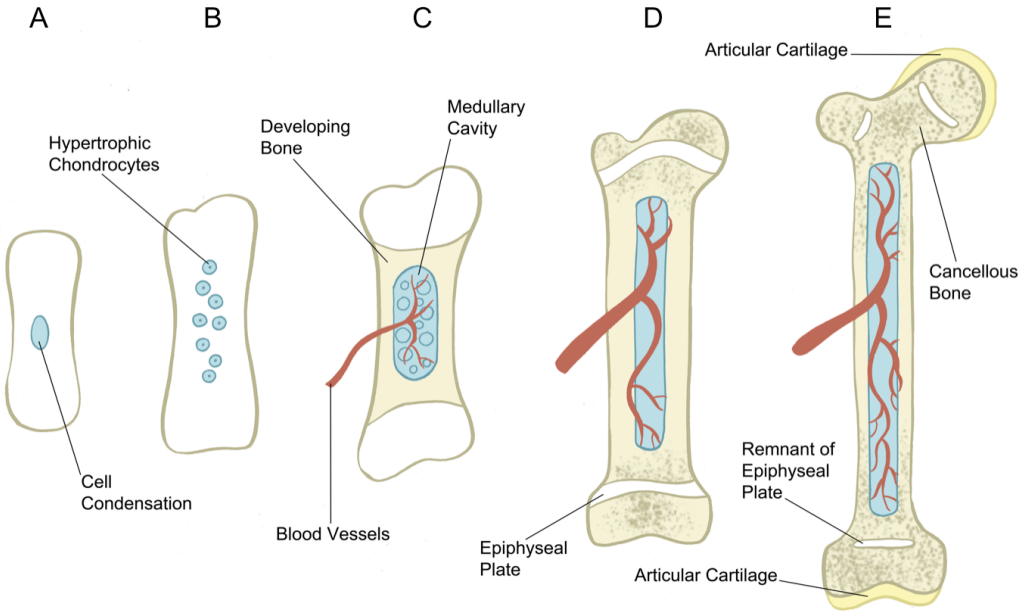

Figure 1.2 The stages of endochondral ossification. A) In the cartilage model the mesenchymal cells form a condensation. B) Cells proliferate and differentiate into hypertrophic chondrocytes and die as matrix calcifies. C) Newly formed osteoblasts cover the shaft in a layer of bone and the medullary cavity is invaded by blood vessels. D) The bone of the shaft thickens as osteoblasts replace cartilage with mineralised bone. E) The mature bone, with the remnants of the epiphyses. Figure reproduced with kind permission of Andrew Waugh.

The development of the joint is dependent on signalling from: members of the Wnt signalling pathway such as *WNT4*, which regulates growth and differentiation of the chondrocytes (70); members of the hedgehog signalling pathway such as Indian hedgehog (*IHH*), which controls chondrocyte proliferation and maturation (71); Notch signalling, which is essential for chondrocyte proliferation and for the normal progression of hypertrophic chondrocyte differentiation into bone in the developing skeleton (72); and transforming growth factor β (TGF-β) signalling.

1.6 TGF-β Pathway

The TGF-β superfamily of secreted growth and differentiation factors consists of over 30 ligands that have regulatory roles both during embryogenesis and in adult homeostasis (73). The superfamily comprises amongst others, TGF-βs 1, 2 and 3, and members of the bone morphogenic protein (BMP) and growth differentiation factor (GDF) subfamilies (74). TGF-β signals are conveyed through type I and type II receptors that are transmembrane serine/threonine kinases (74). The type I receptor family comprises seven activin-like kinases (ALK1-7), whilst the type II receptor family has four members (BMPRII, ActRIIA, ActRIIB and TβRII; 73,74).

During the first stage in skeletal development, chondroprogenitor cells begin to upregulate expression of TGF-β and fibronectin, which in turn upregulates expression of the cell-cell adhesion molecules N-cadherin and neural cell adhesion molecule (N-CAM; 75), initiating cell condensation and cartilage differentiation (57, 76). BMPs are also required for the initiation of chondrocyte differentiation by inducing *SOX9* expression (69). Progressing from condensation to differentiation of chondroblasts or osteoblasts requires down-regulation of N-CAM expression, resulting in cells being released from strong interactions with each other (75). The subsequent transition from proliferating chondroblasts to chondrocytes and hypertrophic chondrocytes is regulated via BMPs interacting with IHH, which induces Parathyroid hormone (PTH) and PTH-related protein (PTHrP), which in turn act via negative feedback to regulate IHH, so controlling the size of the condensation (67, 69). Undifferentiated mesenchymal stem cells (MSCs) produce an ECM rich in collagen type I, hyaluronan, tenascin and fibronectin (66). In the third stage of the differentiation of MSCs into chondrocytes,

TGF-β promotes chondrocytes to change this ECM composition and they begin producing articular cartilage specific collagen II and aggrecan (66). During endochondral bone formation, BMP signalling is required for the commitment of MSCs toward the osteoblast lineage (69), specifically BMP2, which stimulates the osteogenic differentiation of MSCs by inhibiting their differentiation into the myoblastic lineage (73). In the final stage of skeletal development, BMP2 induces expression of key osteogenic transcription factors in osteoblast precursor cells, like *RUNX2*, and osteoblast markers such as alkaline phosphatase, promoting the mineralisation of the bone matrix (73).

1.7 Pathobiology of OA

OA is pathologically characterised by damage to the articular cartilage, changes in subchondral bone, synovitis and capsular thickening (15; Figure 1.3). In primary OA there is an imbalance between the biological properties resisting cartilage wear and mechanical stress, which results in the breakdown of the ECM (14). The articular cartilage becomes fibrillated, in that cracks develop along the surface, and at later stages of the disease fibrocartilage forms, possibly as a consequence of unsuccessful attempts by chondrocytes to repair the damaged tissue (77). Finally osteophytes, which are irregular bony outgrowths capped with cartilage, form at the periphery of the joint surface (14). The similarities in the pathologic progression of OA between patients, even when the initiating events are different, indicates that there may be a common molecular sequence of events underlying OA progression (77).

1.7.1 OA Cartilage

The earliest indication of pathological change is at the articular cartilage surface, which becomes rough and yellowish in colour and fibrillations become visible where chondrocytes cluster together as a result of increased cell proliferation (5, 8, 77). Cartilage is progressively lost until in the later stages the eburnated subchondral bone plate is visible (8; Figure 1.3). Gradual loss of proteoglycans also occurs at the surface region of articular cartilage and this is followed by type II collagen degradation, resulting in disruption of the integrity of the collagen network, allowing hyperhydration

of articular cartilage and loss of tensile strength in the cartilage matrix as the lesion progresses (5, 9, 77). The loosening of the collagen network results in further loss of proteoglycans, leading to mechanical overload, which in turn further damages and loosens the collagen network (14). It is this loosening of the collagen network that causes the hyperhydration of the cartilage giving the soft swollen macroscopic appearance that is visible early in OA (14). Chondrocytes appear unable to compensate fully for this proteoglycan loss even in the presence of increased synthesis, resulting in net loss of matrix (7). With the loss of the matrix components the cartilage is no longer able to function properly (13).

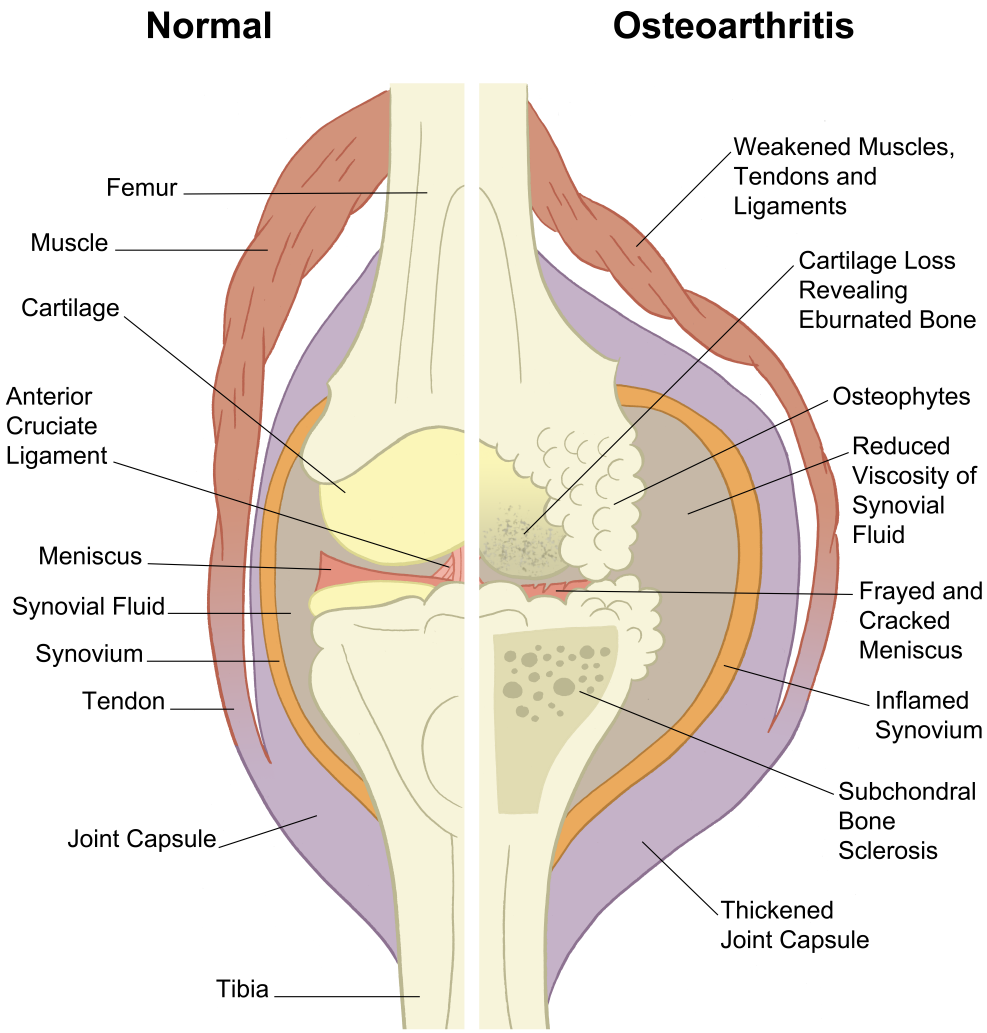

Figure 1.3 Structure of the synovial joint under normal and osteoarthritic conditions. Figure reproduced with kind permission of Andrew Waugh.

The chondrocyte has clinical importance in the context of the pathogenesis of OA as cartilage matrix turnover is mediated by a multitude of factors that act on the chondrocytes and which could lead to repair, remodelling or catabolic processes (5). At a certain point these processes are altered, resulting in matrix changes including abnormal biochemical pathways that contribute to cartilage degradation (7). The gradual self-destruction of the matrix is induced as the chondrocytes respond to catabolic mediators such as IL-1 or to fragments of matrix components including fibronectin, collagen and hyaluronan (13). Synthesis of cartilage components is dependent on a number of growth factors including insulin-like growth factor 1 (IGF-1) and TGF-β (32). These factors have been seen to induce changes in chondrocyte metabolism or induce mechanisms leading to matrix degradation (20). In an attempt by the chondrocytes to initiate repair, new aggrecan, which has a composition more similar to juvenile cartilage than adult, with decreased keratan-sulfate content and increased chondroitin-4-sulfate relative to chondroitin-6-sulfate, is synthesised (78). This new aggrecan forms aggregates that are retained in cartilage during the early phases of OA, but in later stages they diffuse into the synovial fluid (79). Degradative proteinases secreted by the chondrocytes such as the matrix metalloproteinases (MMPs), enzymes that catalyse both collagen and proteoglycan degradation, and which are responsible for matrix turnover and have an increased concentration in OA cartilage (80). Collagenases (MMP1), gelatinases (MMP2) and stromelysin (MMP3) are synthesised by chondrocytes and their activity can be upregulated by certain stimuli, for example cytokines such as IL-1, hormones, and mechanical or growth factors (81). Multiple members of the MMP family acting in unison are responsible for the degradation of the collagen fibrils, which result in the irreversible fibrillation of the articular cartilage in the OA joint (7). In the early phases of OA, MMP13 is predominantly expressed in the lower-middle and deep layers of cartilage, whereas aggrecanases are expressed in the superficial zone (7). An increase in expression of cartilage degrading proteinases and matrix proteins such as these, suggests that chondrocytes attempt to repair the damaged cartilage (77). Another protease that attacks the proteoglycan monomer is Oncostatin M (OSM), a member of the IL-6 family, which plays a role in the articular tissue (82). It is the only member of this cytokine family to cause proteolytic release of proteoglycans and collagen from human articular cartilage (7).

1.7.2 OA Bone

Changes in subchondral bone are closely involved in disease progression. Subchondral bone changes are often seen on radiographs in patients with established OA. Increasingly data suggests that subchondral bone lesions may precede cartilage alterations and that it is the site causing the most significant pathophysiological events occurring in cartilage, rather than as a consequence of cartilage damage (32). Cartilage thins with advancing age as a result of the slow but progressive advance of the ossification front (48). As the cartilage breaks down, changes occur in the underlying bone, which thickens with the formation of bony outgrowths from the calcified cartilage layer and the bone surface (21; Figure 1.3). Increased bone mineral density (BMD), is associated with increased risk of incident OA (83) but some longitudinal studies have shown that decreased BMD is associated with progression of OA (84), implying that increased BMD, although a risk factor for OA, may protect against disease progression once it is established (85).

1.7.3 OA Synovium

The synovium becomes inflamed as a result of excessive catabolism of articular cartilage, causing the release of matrix breakdown products including chondroitin and keratin sulfate peptides, proteoglycan fragments, type II collagen peptides and chondrocyte membranes into the synovial fluid, all of which are antigenic and can elicit an inflammatory response in the synovial membrane (9). As soon as degradation occurs, the inflammatory mediators are either rapidly diffused into the synovial fluid or they act on chondrocytes to cause further catabolism by creating a positive amplification loop leading to more protease production, as well as cytokines and other inflammatory factors (7). Among the mediators responsible for the progression of the disease, evidence points to the proinflammatory cytokine IL-1β as being the most important factor responsible for the catabolic process in OA, as it reduces expression of type II collagen (86) and increases production of MMPs (87, 88). Other proinflammatory cytokines include TNF-α, which is present in OA but only at severe end stage disease, and Nitric oxide (NO) generated by the inducible NO synthase (iNOS), which has been shown to be involved in the promotion of cartilage catabolism in OA through a number of mechanisms (7). For example, it can inhibit the major anabolic processes such as the synthesis of cartilage matrix macromolecules and enhance catabolic processes such as

MMP activity (7). Other major inflammatory factors involved in OA pathophysiology are the prostaglandins, synthesised from arachidonic acid via the actions of the COX enzymes (89). The most abundant of these is prostaglandin E2 (PGE2), which, in addition to exacerbating joint inflammation, can potentiate the effects of other mediators of inflammation, affect cartilage remodelling directly, and contribute to cartilage damage by promoting MMP production (7).

1.7.4 OA Infrapatellar Fat Pad

Inflammation plays an important role in the aetiopathogenesis of OA, and the IPFP secretes adipokines, cytokines, and growth factors that are capable of altering cartilage and synovium (57). The adipokine IL-6 is increased in the IPFP of OA patients compared to their subcutaneous adipose tissue (90), and it has been shown to induce the release of MMPs in vitro (91). In addition to this human OA IPFP contains significant protein levels of vascular endothelial growth factor (VEGF) and the cytokine TNF-α which both contribute to OA (58).

1.7.5 OA Meniscus

OA menisci are often dark yellow/light brown in colour and have calcium deposited on roughened, cracked and frayed surfaces (92; Figure 1.3). Microscopically, OA meniscus ECM shows fine fibrillations, a loss of structure due to swelling of the tissue, large variation in proteoglycan content with some areas high in proteoglycan and others devoid of it, and abnormal cell clustering close to meniscal tears and frays (92). Meniscal extrusion can also occur in the OA joint, where the meniscus is pushed out and extends beyond the tibial plateau, and is caused by articular changes such as joint space narrowing (93).

1.7.6 OA Ligaments and Tendons

Joint space narrowing caused by loss of cartilage causes laxity of the ligaments and tendons reducing their joint restraining abilities and contributing to the progression of OA (94).

1.8 Risk Factors

OA is a complex disease that is likely the result of a combination of risk factors. These risk factors can be divided into two categories: systemic risk factors such as age and gender, or local risk factors such as injury or occupation (15, 95).

1.9 Systemic Risk Factors

1.9.1 Age

Age is the strongest risk factor for both the prevalence and incidence of OA in all joints (32, 96, 97) and it is estimated that radiological evidence of OA is found in 70% of those over 65 years of age (97). The increase in prevalence of OA with age can be attributed to the many age-related changes that occur in the joint including: mechanical stress on cartilage that may arise from local factors such as altered gait; increased joint laxity caused by muscle weakness (98); and changes in proprioception, which is the perception of position and movement of a joint (99). Chondrocytes also undergo age-related changes including decreased synthesis, decreased responsiveness to anabolic growth factors, and synthesis of smaller and less uniform proteoglycans (98) all contributing to the chondrocytes inability to protect itself from biomechanical stress (95).

1.9.2 Sex

Gender has been associated with OA in a joint-specific manner (96). Knee OA is more common in females than in males, and although hip OA is slightly more common in men there is an increased prevalence of OA in all joints in post-menopausal women (13, 32, 98, 100). This high incidence of OA in women who have undergone menopause has been attributed to oestrogen deficiency (85, 98), as adult articular chondrocytes have oestrogen receptors, and it has been demonstrated that oestrogen can have an effect on proteoglycan synthesis (101). In addition to this, the Framingham study reported that post-menopausal women taking oestrogen have a decreased prevalence and incidence of radiographic OA (102). However many studies have failed to detect functional

oestrogen receptors in chondrocytes (101) and the findings of case-control studies of oestrogen use in women with OA have been inconsistent (85).

1.9.3 Ethnicity

Different ethnic populations have been shown to have differences in prevalence of OA; in the Johnston County study it has been demonstrated that African-Americans have a higher prevalence of knee OA than Caucasians but a lower prevalence of hand OA (103). The Beijing osteoarthritis study showed that Chinese men and Caucasian men have a similar prevalence of knee OA whereas Chinese women have a significantly higher prevalence of both radiographic and symptomatic knee OA than Caucasian women from the Framingham study (104). However, in contrast the same study looking at severe radiographic OA and a study of radiographic hip OA both demonstrated lower prevalence in Chinese populations compared to Caucasian populations (104, 105).

1.9.4 Joint Deformity and Injury

Developmental defects that affect the composition of the joint or the alignment angle of the hip, knee and ankle, can affect the weight distribution of the joint and may make an individual more likely to develop OA than people with structurally normal joints (5, 21, 95, 106). Any defect that affects joint stability, such as muscle weakness, or articular cartilage surface fractures, can also lead to development of OA (85, 95). Anterior cruciate ligament injury also leads to joint laxity, which leaves the knee more susceptible to meniscal tears, and surgery to treat meniscal tears has been shown to significantly increase the long-term risk of developing OA (95, 107).

1.10 Local Risk Factors

1.10.1 Obesity

Obesity is strongly linked to both the development and progression of OA, particularly in the knee but also in the hip and the hand (32, 95, 108). The increase in mechanical stress across load-bearing joints in obese individuals has been thought to account for

this risk (98). Overloading of the joints could lead to failure of the ligaments and structural support leading to joint misalignment and knee deformities that accelerate cartilage degeneration and sclerosis of the subchondral bone (32, 85, 98, 109). Although this is probably the primary factor accounting for progression of OA in obese individuals (98), it does not account for the difference in risk for the non-weight-bearing joints such as the hands, wrist and shoulder (108, 109), or for the fact that women are more susceptible to the risks of OA associated with obesity than men (32). It is because of these incongruities that metabolic pathways are also thought to be involved in the risk from obesity (98). It is hypothesised that excess fat may have a metabolic effect on cartilage (109), however some studies have not supported this and controlling for hypertension, diabetes, high serum cholesterol and high uric acid levels did not reduce the risk of OA (110). It has also been suggested that the association between obesity and OA may be partially mediated by the inflammatory effects of adipokines such as leptin, adiponectin and resistin, produced by the excess fat (111).

1.10.2 Sports and Exercise

Epidemiological studies have shown that participation in high-level intense sport is associated with increased risk of lower limb OA (32, 95). However studies have shown that individuals who participate in lower-impact recreational activity, such as jogging do not have an increased risk of OA provided the joint is biomechanically normal (32, 112). This probably reflects the fact that normal moderate movement and loading of the joint is required to maintain healthy cartilage (48, 112).

1.10.3 Occupation

Occupations involving repetitive tasks such as kneeling, bending, squatting and climbing stairs, overworking the joints and fatiguing the muscles, have been shown to increase the incidence of knee OA (32, 85, 95). Whilst manual labour jobs that require heavy lifting such as farming have been associated with increased incidence of hip OA (32).

1.11 Genetic Contribution to OA

OA is a multifactorial disease with a clear genetic component (113) that was first reported as far back as the 1940s (114). Following this several twin studies were carried out in which monozygotic twins demonstrated a consistently higher concordance for OA than dizygotic twins (115), giving the first estimate of the heritability of OA; that is the extent to which genetic variation determines variation of a disease or trait (116). These studies translated into heritability estimates of 39% for OA of the knee and 60% for OA of the hip (96).

1.11.1 Linkage Analysis

Genetic linkage is the co-segregation of characteristics, such as a phenotype or allele in a family, because their determinants are not separated by recombination (116, 117). Families that have at least one sibling pair affected by primary OA are analysed to see if these alleles are segregating with incidence of OA (118, 119). Linkage studies often suffer from several shortcomings, including: analysis of relatively small sized cohorts; genotyping only a proportion of the variants within a gene; and late onset of disease meaning that often no parents are available for extended family studies (14). The most significant result from linkage studies has been the identification of the region in chromosome 14q harbouring deiodinase, iodothyronine, type II (*DIO2*; 120), which mediates thyroid hormone signalling, important in skeletogenesis (119). The common *DIO2* haplotype containing the minor allele of rs225014 and the major allele of rs12885300 was shown to be associated with female hip OA (OR 1.79, $P < 2.02 \times 10^{-5}$; 120).

1.11.2 Candidate Gene Studies

Candidate genes are chosen based on prior knowledge of OA, and since our knowledge of the aetiology of OA is poor, and the power of small cohorts is generally low, these studies have yielded inconsistent results (14, 118). Initial candidate genes focused on several cartilage ECM structural genes i.e. the type II, IV, V, VI and XI collagen genes on the basis that they code for regulators of joint formation maintenance (119), and

some studies focussed on genes that are important in bone homeostasis such as *VDR,* encoding the vitamin D receptor gene (121) and *ESR1,* which codes for the oestrogen receptor α gene (122), as subchondral sclerosis and increased bone density appears early in OA. Genes encoding inflammatory cytokines have also been widely investigated, based on the inflammation of the synovium in OA and the possible autocrine and paracrine role of cytokines synthesised by chondrocytes; in particular the *IL1* cluster which contains members of the *IL1* family including *IL1A*, *IL1B*, *IL1R1, IL1R2 IL1RN* (encoding IL-1α, IL-1β, IL-1 receptors type 1 and 2, and IL-1 receptor antagonist, respectively; 14, 123, 124, 125), Some of the reasons for a lack of consistency in replication of candidate gene studies are: difference in case ascertainment and phenotype criteria; difference in ethnicity; the sites of OA studied; the sex of the patients; and the occurrence of false positives in the initial studies or false negatives in the replication study (126). Insufficient power due to sample size, which can cause over-estimation of genetic effects, is one likely source of false positives in initial studies (127). Heterogeneity in control ascertainment may also explain some inconsistent results across study populations, as what constitutes an appropriate control population for OA genetics studies remains unclear; for example, genes important in OA pain may be detected using controls with asymptomatic OA but genes important in joint structural changes may only be detected when the controls have no OA. Additionally, because of the late onset and high prevalence of OA, individuals who are considered to be controls at one time-point may subsequently develop OA and become cases in later studies (127).

There is only one candidate gene to have reached genome-wide significance (GWS; $P < 5x10^{-8}$); *GDF5* coding for growth differentiation factor 5, part of the TGF-β superfamily (128). *GDF5* mutations cause a variety of rare inherited skeletal dysplasias such as chondrodysplasia grebe type and brachydactyly type C (118) and common functional variation, affecting transcription of this gene is significantly associated with OA in both Europeans and Asians ($P < 5x10^{-8}$; 119).

1.11.3 Genome Wide Association Studies

Genome wide association studies (GWAS) involve a hypothesis-free, systematic screen of large case-control cohorts with dense maps of millions of polymorphic markers (119, 129). GWAS has been made possible by advances in high-throughput SNP genotyping

methods, the completion of the Human Genome Project and databases such as HapMap with extensive information about the linkage disequilibrium (LD) structure of chromosomes (116), as the genotypes of the SNPs missing from an array can by imputed, or predicted, from the LD relationship with SNPs on the array (130). However, GWAS are also subject to drawbacks such as lack of defined patient stratification criteria, as OA associations are often to a particular joint rather than to several different joints (119), the need for a large number of subjects and the need for replication of positive associations (116). In recent years a few such studies have overcome these drawbacks to successfully identify OA-associated loci with GWS ($P < 5x10^{-8}$):

7q22

A region on chromosome 7q22 encompassing six different genes, *PRKAR2B*, *COG5, GPR22, HBP1, DUS4L* and *BCAP29* was found to be associated with knee and/or hand OA in a Dutch Caucasian population with an odds ratio (OR) of 1.14 ($P = 8x10^{-8}$; 131).

NCOA3

A meta-analysis comprising the largest study of hip OA to date identified a locus near nuclear receptor co-activator 3 (*NCOA3)* that is associated with OA (OR = 1.8, $P = 7.9x10^{-9}$;). *NCOA3* recruits histone acetyltransferases and methyltransferases for chromatin remodeling, and is involved in the co-activation of nuclear receptors for many hormones that have been implicated in skeletal metabolism such as thyroid hormone and vitamin D3 (113).

DOT1L

A GWAS analysing cartilage thickness in hip OA identified a SNP in disruptor of telomeric silencing 1-like (*DOT1L*) histone H3 methyltransferase associated with larger JSW (OR = 1.10, $P = 4.8x10^{-10}$) and with a decreased risk of hip OA but not at a GWS level (OR = 1.14, $P = 1.5x10^{-4}$; 132). A meta-analysis then demonstrated that the SNP was associated with increased risk of hip OA (OR = 1.10, $P = 8.1x10^{-8}$; 133)

MCF2L

In a European cohort a SNP in MCF.2 cell-line-derived transforming sequence-like (*MCF2L*) was found to be associated with OA (OR 1.17, $P = 2.1x10^{-8}$). *MCF2L*

regulates nutrophin-3, a member of the nerve growth factor family, which are known to be effective in the treatment of pain in knee OA (134).

arcOGEN

In the largest GWAS to date, eight novel signals, five of which were GWS and three that were approaching it were identified (135). The most significant signal, after replication was on chromosome 3p marked by two SNPs in perfect LD with each other in glycosyltransferase 8 Domain Containing 1 (*GLT8D1*; OR = 1.09, P = 7.24×10^{-11}) and Guanine Nucleotide Binding Protein-Like 3 (*GNL3*; OR = 1.09, P = 5.13×10^{-9}). The two SNPs were significant in all joints, whereas the remaining four GWS signals (Table 1.1) were significant only in hip OA.

Table 1.1 The five genome-wide significant signals from the arcOGEN GWAS.

SNP	Chromosome	Nearest Gene	Stratum	Odds ratio	P-value
rs6976/ rs11177	3	GLT8D1/ GNL3	All joint	1.12 (1.08-1.16)	7.24×10^{-11}
rs4836732	9	ASTN2	Female hips	1.20 (1.13-1.27)	6.11×10^{-10}
rs9350591	6	FILIP1/ SENP6	All hips	1.18 (1.12-1.25)	2.42×10^{-9}
rs835487	12	CHST11	All hips	1.13 (1.09-1.18)	1.64×10^{-8}
rs10492367	12	KLHDC5/ PTHLH	All hips	1.14 (1.09-1.20)	1.48×10^{-8}

Interestingly there is no overlap between genes and regions identified by linkage scans and those found by GWAS (116). For example, *GDF5* which is the only candidate gene found to be robustly associated with OA, has not been identified in any of the GWAS so far (131).

1.12 Expression Quantitative Trait Loci

Ideally the loci identified by GWAS would implicate functional variants and disease pathways that could be experimentally studied in follow-up functional studies, such as non-synonymous SNPs that change the amino acid sequence or truncate the protein, or insertions/deletions that can interrupt splicing sites, changing the mRNA isoform (136). However the OA-associated SNPs identified by GWAS will not usually be the causal variants. Instead the surrounding genomic regions of high LD with the associated SNPs will most likely contain the causal variants with biological function (137). Most of these SNPs fall into non-coding regions of the genome according to the GWAS catalogue (138). Projects such as the Encyclopedia of DNA Elements (ENCODE) have collected and annotated genomic features, which are valuable for the prioritisation of SNPs (137). Among the genomic features catalogued in this database are, DNase I hypersensitivity sites, which are markers of accessible chromatin indicating regulatory roles in the transcription process (136), transcription start sites, enhancers, repressors and promoter regions (139). In addition to these regions it has been demonstrated that trait-associated SNPs will often alter gene expression rather than alter the protein (140). By treating gene expression as a quantitative trait it is possible to correlate genotypes of disease-associated SNPs with transcriptional levels of the genes (141). Regions that demonstrate a correlation between expression levels and genotype at trait-associated SNPs are known as expression quantitative trait loci (eQTLs). Polymorphisms that regulate eQTLs can be *cis*-acting SNPs (within or close to the gene) or *trans*-acting (located a distance from the gene or on a different chromosome; 142). Variants can regulate expression post-transcriptionally in 3′UTR regions by altering microRNA binding sites (143), by altering polyadenylation sites resulting in transcripts with different 3′UTR lengths (143, 144) or by altering splice sites causing the splicosome to either employ the next available legitimate splice site, known as exon skipping, or select the next best, pseudo splice site in the surrounding sequence, known as cryptic splice-site utilisation (145). This alternative splicing can affect gene expression by removing or inserting regulatory elements controlling translational efficiency, mRNA stability or localisation (143, 146). Once an eQTL is identified there has been a good level of success pinpointing the nearest regulatory element (147), but in the majority of cases, however, it is unclear whether such variants are of direct functional significance, as opposed to simply being in LD with another as yet unidentified SNP in the vicinity (148).

1.13 Aims of this Study

The overall aim of this study is to analyse the expression of genes from a number of OA susceptibility loci to identify eQTLs and regulatory variants. Each of the chapters in this study further explores a locus identified as being associated with OA. Initially, for each locus, the allelic output of a gene was measured in synovial joint tissues, working under the hypothesis that if one allele produces significantly more mRNA than the other allele, then a *cis*-regulatory polymorphism is present. Where possible the results were further analysed in order to identify the precise regulatory variants affecting the expression of genes at the locus.

Chapter 2: Materials and Methods

2.1 Tissue Samples

Cartilage (Cn) was obtained from patients undergoing either total hip replacement (THR) surgery, total knee replacement (TKR) surgery or from trauma patients presenting with a neck-of-femur fracture (NOF). Additionally a number of synovial joint tissues were obtained; infrapatellar fat pad (Fp), meniscus (Me), patellar tendon (Te), anterior cruciate ligament (Li), synovium (Sy), and cancellous bone (Bn). Ethical approval for the collection of these tissues was granted by local ethics committees from both Oxford and Newcastle upon Tyne and informed consent was obtained from each donor. Patient details can be found in the appendix (Table A1).

2.2 Tissue Grinding

On the day of surgery tissue samples were removed from the joint, washed in phosphate-buffered saline (PBS; Lonza, Slough, UK) and snap frozen at -80°C. The frozen tissue was placed into a grinding vial, pre-cooled in liquid nitrogen, with a grinding ball and ground to a powder using a Retsch mixermill 200 (Retsch Ltd, Castleford, UK). Samples we subjected to 90 seconds of grinding at a frequency of 20 Hz, cooling the vials in liquid nitrogen between each cycle, until completely powderised. The ground tissue was then stored at -80°C.

2.3 Nucleic Acid Extraction and Quantification

Genomic DNA (gDNA) and RNA were extracted simultaneously from the ground tissue using a protocol adapted from the Omega EZNA DNA/RNA isolation Kit (Omega Bio-Tek, Georgia, USA). Between 0.25-0.30g of ground tissue was combined with 700 µl of GTC lysis buffer (Omega) mixed with 14 µl of β-mercaptoethanol (BME; Sigma-

Aldrich, Gillingham, UK). The sample was then thoroughly homogenised and centrifuged at 11700 rpm for 5 minutes. To extract DNA from the tissue the resulting supernatant was passed through a Hi-bind DNA column (Omega), and to extract RNA from the tissue the supernatant was mixed with 1 volume of ice-cold 70% ethanol made up with diethylpyrocarbonate (DEPC)-treated H_2O (Invitrogen, Life Technologies, Paisley, UK), and passed through a Hi-bind RNA column (Omega). The DNA and RNA columns were washed with DNA wash buffer or RNA wash buffer I (Omega) respectively, centrifuged at 10300rpm for 1 minute and the filtrate discarded. The DNA and RNA columns were further washed with buffer HB or RNA wash buffer II (Omega) respectively, centrifuged at 10300 rpm, and the filtrate discarded. The column-bound DNA was eluted with elution buffer (Omega) and the column-bound RNA was eluted with DEPC-treated H_2O (Invitrogen). The concentration of the nucleic acids was determined using the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, USA). To the RNA, 3 µl of RNaseOUT recombinant ribonuclease inhibitor (40 U/µl; Invitrogen) was added to prevent degradation, and along with the DNA was centrifuged using the DNA 120 SpeedVac system (Thermo Scientific, Epsom, UK). The pellets were resuspended in DEPC-treated H_2O (Invitrogen) to a final concentration of 50 ng/µl for DNA and 250 ng/µl for RNA. DNA was stored at -20°C and RNA at -80°C until required.

2.4 Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR)

RNA extracted from patient tissue was reverse transcribed to synthesise complementary DNA (cDNA). Prior to RT-PCR a DNase step was carried out to remove any remaining gDNA from the sample. Four microlitres of RNA suspended in DEPC-treated H_2O (Invitrogen) were mixed with 1 µl of 10X TURBO DNase buffer (Ambion, Life Technologies) and 1 µl of TURBO DNase enzyme (2 U/µl; Ambion) and incubated at 37°C for 30 minutes. Following this, 1.8 µl of ethylenediaminetetraacetic acid (EDTA; 100mM; Sigma) was added and incubated at 75°C for 10 minutes.

RT-PCR was carried out with the SuperScript First-Strand Synthesis System (Invitrogen). The DNase-treated RNA was mixed with 1 µl of random primers (50 ng/µl; Invitrogen), 1µl of dNTPs (10 mM each dATP, dCTP, dGTP, dTTP; Invitrogen) and 1 µl of DEPC-treated H_2O (Invitrogen) and incubated at 65°C for 5 minutes

followed by incubation on ice for 1 minute. Four microlitres of 5X first-strand buffer (250 mM Tris-HCl pH 8.3, 375 mM KCl, 15 mM $MgCl_2$; Invitrogen), 4 µl of $MgCl_2$ (25 mM; Life Technologies), 2 µl of dithiothreitol (DTT; 0.1 M; Invitrogen), and 1 µl of RNaseOUT recombinant ribonuclease inhibitor (40 U/µl; Invitrogen) were then added and incubated at 25°C for 1 minute. Following this step, 1 µl of SuperScript II reverse transcriptase enzyme (200 U/µl; Invitrogen) diluted 1:3 with DEPC-treated H_2O (Invitrogen) was added. This mix was then incubated at 25°C for 10 minutes, 42°C for 50 minutes and 70°C for 15 minutes. Finally 1 µl of RNase H (5 U/µl; New England Biolabs; NEB, Hitchin, UK) diluted 2:3 with H_2O (Sigma) was added and the product was incubated at 37°C for 20 minutes.

Quality and integrity of the cDNA was assessed using primers specific for the gene *HBP1* that concurrently amplify both gDNA and cDNA producing different fragment sizes. Primer sequences can be found in the appendix (Table A2). PCR products were electrophoresed through a 2% weight/volume agarose gel containing ethidium bromide (10 mg/ml; Sigma) at a concentration of 1 µl/100 ml agarose and UV visualised using a G:BOX gel doc system (Syngene, Cambridge, UK). Samples that produced the correct sized band for the cDNA fragment and no gDNA fragment were taken forward for analysis.

2.5 Polymerase Chain Reaction (PCR)

PCR was carried out in 15 µl reactions containing 0.5 µl of cDNA or 50 ng of DNA, 0.075 µl of each amplimer (100 µM; Sigma), 1.5 µl of 10X PCR buffer (500 mM KCl, 100 mM Tris-HCl pH 8.3; Life Technologies), 0.8 µl, 1.2 µl or 1.8 µl of 25mM $MgCl_2$ (Life Technologies) to a concentration of 1 mM, 2 mM or 3 mM respectively, 0.375 µl of dNTPs (10 mM each dATP, dCTP, dGTP, dTTP; Bioline, London, UK), 0.08 µl of AmpliTaq Gold DNA Polymerase (5 U/µl; Life Technologies), and made up to 15 µl with H_2O (Sigma). Thermocycling conditions were an initial denaturation at 96°C for 14 minutes, followed by 35 cycles of 96°C for 30 seconds, annealing for 30 seconds at temperatures between 55°C and 65°C depending on the specific conditions required for the amplimers, extension at 72°C for 1 minute/kb then a final extension at 72°C for 5 minutes. Primers and reaction conditions are listed in the appendix (Table A2). PCR

products were electrophoresed through a 2% agarose gel stained with ethidium bromide (10 mg/ml; Sigma).

2.6 Gene Expression Analysis using Real Time Quantitative PCR (qPCR)

Gene expression analysis was carried out using ready-made assays consisting of specific primer-probe mixes (Integrated DNA Technologies: IDT, Leuven, Belgium) or a combination of primers (Sigma) and probes custom designed using the real time PCR Assay Design Tool (IDT). Primer and probe sequences are listed in the appendix (Table A3). The pre-designed primer-probe mixes were dissolved in IDTE buffer (10mM Tris, 0.1mM EDTA pH 8.0) and amplifications were performed in triplicate in 10 µl reactions containing 5 µl of TaqMan Gene Expression Master Mix (Life Technologies), 1 µl of 10X primer-probe mix (IDT) specific to each gene, 1.5 µl of H_2O (Sigma) and 2.5 µl of cDNA at a dilution of 1:20 with DEPC-treated H_2O. Reactions for custom-designed primers and probes were also carried out in a final volume of 10 µl containing 0.2 µl of both forward and reverse primer (10 mM; Sigma), 0.1 µl of custom-designed probe (IDT), 5 µl of TaqMan Gene Expression Master Mix (Life Technologies), 2 µl of H_2O (Sigma), and 2.5 µl of cDNA in a 1:20 dilution. Cycling conditions were 50°C for 2 minutes, 95°C for 10 minutes, 95°C for 15 seconds repeated for 40 cycles and 60°C for 1 minute. The cycles were performed on a 7900HT Fast Real-Time PCR system (Life Technologies). Gene expression was measured relative to the housekeeping genes *HPRT1, GAPDH* and Human *18S* ribosomal RNA (rRNA).

2.7 Statistical Analysis of Real-Time qPCR

Analysis of these data was carried out using Sequence Detection Systems Version 2.4 (Life Technologies). Differences in the threshold cycle (Ct) value were calculated using the equation $2^{-\Delta Ct}$ where ΔCt is the difference between the mean Ct of the three housekeeping genes *HPRT1, GAPDH* and *18S,* and the Ct of the gene of interest. GraphPad Prism Software Version 5.0 (GraphPad) was used to generate graphs and to perform statistical analysis. The $2^{-\Delta Ct}$ values were then grouped together according to the genotype at the associated SNP in question and plotted on a graph. Mann-Whitney

U tests were used to determine a statistical difference between two genotypic groups. Kruskal-Wallis tests were used to determine a statistical difference between three genotypic groups. A $P \leq 0.05$ was considered significant.

2.8 Restriction Fragment Length Polymorphism (RFLP) Analysis

PCR was performed in a total reaction volume of 15 µl as previously described (section 2.5). Primers and reaction conditions are listed in the appendix (Table A4). For the restriction digests 10 µl of PCR product were digested for 3 hours at the appropriate temperature with 5 µl of digestion mixture containing 5 units of the specific restriction enzyme, 1.5 µl of enzyme buffer (10X; NEB), 0.15 µl of bovine serum albumin (BSA; 100X; NEB) if required, and H_2O (Sigma). Enzymes, buffers and reaction conditions are listed in the appendix (Table A5). Digested products were electrophoresed through 3% agarose gels stained with ethidium bromide (10 mg/ml; Sigma) and scored following UV visualisation.

2.9 Database Search for Polymorphisms

BROAD Institute, single nucleotide polymorphism (SNP) annotation and proxy search version 2.2 (SNAP; http://www.broadinstitute.org/mpg/snap/) was used to query the pairwise LD between two SNPs, and to identify transcript SNPs in high LD with associated SNPs that could be used as proxies. In addition to this, HapMap genome browser (http://hapmap.ncbi.nlm.nih.gov), Ensembl (http://www.ensembl.org/index.html), the University of California, Santa Cruz (UCSC) genome browser (http://genome.ucsc.edu), RegulomeDB (http://www.regulomedb.org) and Mutation Taster (http://www.mutationtaster.org) were used to select SNPs for analysis. All SNPs investigated in this study are listed in the appendix (Table A6).

2.10 Allelic Expression Imbalance (AEI) Analysis using a SNaPshot Multiplex Kit

To analyse the differences in mRNA output from the two copies of a gene, allelic expression imbalance (AEI) analysis was carried out on patients who were heterozygous for transcript SNPs used as markers for the associated SNPs. RT-PCR was performed as previously described (section 2.4) to synthesise cDNA from the extracted RNA. This cDNA was then PCR amplified using primers specific for the region encompassing the SNP. Primer sequences and reaction conditions are listed in the appendix (Table A7). For each cDNA and gDNA sample amplifications were carried out in 15 µl reactions as previously described (section 2.5) with 10 technical replicates. Five microlitres of these PCR products were incubated with 0.15 µl Exonuclease I (*ExoI*; 20,000 U/ml; NEB), 0.65 µl of Exonuclease I Reaction Buffer (1X; 67 mM Glycine-KOH, 6.7 mM $MgCl_2$, 10 mM β-ME, pH 9.5; NEB) and 0.7 µl of H_2O (Sigma) at 37°C for 1 hour then inactivated at 80°C for 20 minutes. Following this the samples were further incubated with 1 µl of shrimp alkaline phosphatase (SAP; 1 U/µl; GE Healthcare, Little Chalfont, UK) at 37°C for 1 hour then inactivated at 75°C for 15 minutes. These products were then used as a template in the single base extension (SBE) reaction using a SNaPshot Multiplex Kit (Life Technologies). SBE utilises a primer that anneals immediately adjacent to the SNP of interest. SBE primer sequences are listed in the appendix (Table A8). The primer is extended into the polymorphic site using fluorescently labelled dideoxynucleotides (ddNTPs). Each ddNTP has a different fluorescent tag allowing the 3130*xl* Genetic Analyzer (Life Technologies) to discriminate between the two alleles. SBE reactions were carried out in a total reaction volume of 3 µl containing 1.2 µl of *ExoI*/SAP-treated PCR product, 1.5 µl of SNaPshot Multiplex Kit reaction mix (Life Technologies), 0.1 µl of SBE (100 µM; Sigma), and 0.2 µl of H_2O (Sigma). Positive and negative SNaPshot Multiplex Kit controls were carried out in a total reaction volume of 10 µl containing 5 µl of SNaPshot reaction mix, 1 µl of control primer, 2 µl of control DNA was added to the positive control only and the volumes made up to 10 µl with H_2O (Sigma). Thermocycling conditions were 25 cycles at 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 30 seconds. Finally, following this reaction, extension products were incubated with 1 µl of SAP (1 U/µl; GE Healthcare) at 37°C for 1 hour then inactivated at 80°C for 20 minutes. One microlitre of this final product was combined with 9 µl of GS120 LIZ size standard (Life Technologies) at a 1:100 dilution with HiDi Formamide (Life Technologies). These products were separated by capillary

electrophoresis through a 36 cm capillary array with POP-7 polymer (Life Technologies) at 60°C using a 3130*xl* Genetic Analyzer (Life Technologies). These data were analysed using GeneMapper 4.0 software (Life Technologies), which produces electropherograms of the peak heights for the two alleles.

2.11 AEI Analysis using Taqman Real-Time qPCR Genotyping Assays

Quantitative real time PCR genotyping assays are standard real time assays containing probes specific for each of the alleles, labelled with either FAM or VIC dye. Real time PCR was carried out in a 10 μl reaction containing 5 μl of TaqMan Universal Master Mix II no UNG (Life technologies), 0.25 μl of TaqMan assay (40X; Life technologies), 0.75 μl of H_2O (Sigma) and either 20 ng of gDNA or 4 μl of cDNA at 1:20 dilution with DEPC-treated H_2O (Invitrogen). Samples were amplified using an ABI PRISM 7900HT Sequence Detection System (Life Technologies) under the thermocycling conditions 50°C for 2 minutes, 95°C for 10 minutes followed by 40 cycles of 92°C for 15 seconds and 60°C for 1 minute. Reactions were carried out with five technical replicates for both gDNA and cDNA.

2.12 Statistical Analysis of AEI Data

For SNaPshot, the peak heights produced by the GeneMapper 4.0 software (Life Technologies) are proportional to the amount of each allele present and allelic ratios were calculated with the formula (peak height allele 1)/(peak height allele 2). For TaqMan, allelic ratios were calculated using the formula $(2^{-FAM\ Ct})/(2^{-VIC\ Ct})$. For all methods the average peak height ratio for gDNA (gDNA), which represents a 1:1 ratio was used to normalise the cDNA peak height ratios using the formula; normalised allelic ratio = cDNA average allelic ratio/ gDNA average allelic ratio. A particular patient sample was considered to demonstrate AEI if following a two-tailed Mann-Whitney U Test, the difference in the normalised allelic ratio versus the gDNA allelic ratio generated a $P \leq 0.05$. Statistical analysis was carried out using GraphPad Prism 5.0 software (GraphPad Software Inc. La Jolla, USA). For each assay a standard curve

experiment using mixtures of known amounts of homozygous gDNA was carried out to verify that differences in allelic output could be accurately measured.

2.13 Functional Analysis of the *SMAD3* 3′UTR

2.13.1 Construction of SMAD3 3′UTR Luciferase Reporter Plasmids

To generate the constructs for use in the luciferase reporter gene assays, gDNA was PCR amplified using gene-specific primers containing restriction sites at the 5′ ends of both primers. Primer sequences and restriction enzymes are listed in the appendix (Tables A9 and A10). PCR was carried out in 50 μl reactions containing 1 μl of 50X Titanium *Taq* DNA polymerase (Clontech, Oxford, UK), 5 μl of 10X Titanium *Taq* PCR buffer (400 mM Tricine-KOH pH 8.0, 160 mM KCl, 35 mM $MgCl_2$, 37.5 μg/ml BSA; Clontech), 1 μl of dNTPs (10 mM each dATP, dCTP, dGTP, dTTP; Bioline), 1 μl of each primer (100 μM; Sigma), 12.5 μl of Betaine solution (5M; Sigma), 26.5 μl of H_2O (Sigma) and 100 ng of DNA. Thermocycling conditions were an initial denaturing step at 95°C for 1 minute, followed by 35 cycles of 95°C for 30 seconds and annealing at 68°C for 1 minute, then a final extension at 68°C for 3 minutes. After PCR amplification the PCR product was purified using the QIAquick PCR purification kit (Qiagen) following the manufacturer's protocol.

The purified PCR product was then doubled digested in a 50 μl reaction with 25 units of *SpeI* (10,000 U/ml; NEB) and 25 units of *HindIII* (20,000 U/ml; NEB) for amplicons 1 and 3 or 25 units of *MluI* (10,000 U/ml; NEB) and 25 units of *SacI* (20,000 U/ml) for amplicon 2, and 5 μl of restriction buffer (10X; NEB), made up to 50 μl with H_2O (Sigma). The reaction mix was incubated overnight at 37°C. The digested products were then further purified using the QIAquick PCR purification kit (Qiagen) according to the manufacturer's protocol.

In preparation of cloning 5 μg of p-MIR-REPORT miRNA expression reporter vector (Promega, Southampton, UK) was double-digested as previously described, with *SpeI* (NEB) and *HindIII* (NEB) for amplicon 1 and 3, and *MluI* (NEB) and *SacI* (NEB) for amplicon 2. The digested products were electrophoresed through a 0.8% agarose gel stained with ethidium bromide (10 mg/ml; Sigma). The digested vectors were excised

from the gel and purified using the QIAquick gel extraction kit (Qiagen) according to the manufacturer's instructions.

The purified PCR fragments were then ligated into the purified pMIR-REPORT vector by incubating 17 μl of DNA in a ratio of 4:1 (PCR fragments: plasmid), 400 units of T4 DNA ligase (400,000 U/ml; NEB) and 2 μl of 10X T4 DNA ligase reaction buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT pH 7.5; NEB), at 16°C overnight.

2.13.2 Bacterial Cell Transformation

Two microlitres of the ligated construct were transformed into one 50 μl vial of One Shot Mach1 T1 phage-resistant chemically competent *E.coli* cells (Invitrogen) by incubating the mixture on ice for 20 minutes and heat shocking the cells at 42°C for 30 seconds followed by incubation on ice for a further 2 minutes. The transformed cells were placed at 37°C in a shaking incubator at 200 rpm for 1 hour with 200 μl of super optimal broth with catabolite repression (SOC) medium (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose; Invitrogen). After incubation this mixture was spread onto two Lysogeny broth (LB; Sigma) Bacto™ agar (BD Biosciences, Oxford, UK) plates containing ampicillin (100 mg/ml; Sigma), 50 μl on one plate and 100 μl on the other. Plates were incubated overnight at 37°C.

Ten colonies were picked from each plate and each one added to a 1.5 ml centrifuge tube containing 15 μl of LB (Sigma) and ampicillin (100 mg/ml; Sigma) to a final concentration of 100 μg/ml. This mixture was then amplified using the same PCR primers and conditions used to amplify the original inserts. Clones displaying the correct size PCR product were then grown overnight by adding 10 μl of the colony mixture to a 15 ml centrifuge tube containing 3.5 ml LB (Sigma) and 3.5 μl of ampicillin (100 mg/ml; Sigma), and incubating at 37°C in a shaking incubator at 200 rpm. Glycerol stocks were also made from the positive colonies by adding 200 μl of the colony mixture to 200 μl of glycerol (Sigma), with storage at -80°C.

Plasmid DNA was purified from the overnight bacterial cultures using a Qiagen Plasmid Mini Kit (Qiagen) according to the manufacturer's instructions. Ten microlitres

of the purified plasmid DNA was then sent to Source Bioscience (Nottingham, UK) to be Sanger sequenced. Sequencing primers are listed in the appendix (Table A11).

One plasmid with the correct DNA sequence was grown overnight. Briefly, 10 μl of the bacterial culture glycerol stock was added to 5 ml of LB containing 5 μl of ampicillin (100 mg/ml; Sigma) and incubated at 37°C in a shaking incubator at 200 rpm for 4 hours. This was then added to 250 ml of LB (Sigma) and grown overnight at 37°C in a shaking incubator at 200 rpm. The DNA was then purified using a Qiagen plasmid maxi kit (Qiagen) according to the manufacturer's instructions.

2.13.3 SW1353 Cell Culture

SW1353 human chondrosarcoma cells were cultured in complete growth medium made up of Dulbecco's modified Eagle medium (DMEM)/F-12 (1:1; Life Technologies), supplemented with 10% fetal bovine serum (FBS; Life Technologies), 2mM L-glutamine (Sigma), and 100 U/ml penicillin, 100 μg/ml streptomycin (Sigma). SW1353 cells were maintained in 10 ml of this complete growth medium in 75 cm^2 cell culture flasks (Corning, USA), at 37°C in a humidified 5% CO_2 incubator. Cells were passaged at 80% confluence using standard tissue culture techniques.

2.13.4 MG63 Cell Culture

MG63 human osteosarcoma cells were cultured in DMEM (Life Technologies) supplemented with 10% FBS (Life Technologies), 100 U/ml penicillin, 100 μg/ml streptomycin (Sigma). The cells were maintained in 10 ml of this complete growth medium in 75 cm^2 cell culture flasks (Corning), at 37°C in a humidified 5% CO_2 incubator. Cells were passaged at 80% confluence using standard tissue culture techniques.

2.13.5 Transfection of Cells with SMAD3 3′UTR pMIR-REPORT miRNA Expression Vector Construct

Both SW1353 and MG63 cells were seeded at a density of 17,500 cells/well in a 48-well cell culture plate (Corning) with 300 μl of complete growth medium. Once cells

had reached 80% confluence they were transfected, with a master mix containing 1.65 µl of ExGen 500 *in vitro* transfection reagent (Fermentas, Thermo Scientific), 500 ng of the *SMAD3* 3′UTR luciferase reporter construct and 15 ng of pRL *Renilla* luciferase control reporter vector (Promega) made up to 30 µl with 0.9M NaCl. This was incubated at room temperature for 10 minutes, the media was removed from the wells, and 30 µl of the master mix was added to each well followed by 270 µl of medium. The plate was incubated at 37°C in a humidified 5% CO_2 incubator. For each construct, transfections were performed using three different passages each with 6 biological replicates.

2.13.6 Luciferase Reporter Assay

Medium was removed from the cells and the wells were washed twice with 500 µl of non-sterile PBS (Lonza). Sixty-five microlitres of 1X passive lysis buffer (PLB; Promega) was added to each well. The plates were incubated on a rocker at room temperature for 20 minutes and stored at -20°C for at least 4 hours until the PLB froze.

Once the PLB was frozen the plate was removed and defrosted on a rocker at room temperature for 10-15 minutes. The cell lysate was assayed for firefly luciferase activity by transferring 20 µl of the cell lysate to a luminometer plate, adding 50 µl of Luciferase Assay Reagent II (Promega) to each well and reading the luciferase activity using a MicroLumatPlus Luminometer (Berthold Technologies, Hertfordshire, UK). Following this step 50 µl of Stop and Glo Reagent (Promega) was added to each well and the *Renilla* luciferase activity was measured as a control for transfection efficiency. The experiment was performed using three biological replicates with six technical replicates for each.

2.13.7 Statistical Analysis of Luciferase Activity

Relative luciferase activity was measured by calculating the ratio of the firefly luciferase readings to the *Renilla* luciferase readings to account for differences in transfection efficiency. Luciferase activity of the constructs containing the polymorphisms was compared to the luciferase activity of constructs containing the

wild-type (WT) alleles. Statistical comparison was performed on the average of the six technical and three biological repeats using the two-tailed Student's *t*-test.

2.13.8 Site-Directed Mutagenesis

Site-directed mutagenesis of the *SMAD3* 3′UTR amplicon 3-pMIR-REPORT miRNA expression vector construct was performed using the QuikChange II Site-Directed Mutagenesis Kit (Agilent, Cheshire, UK). Primers sequences are listed in the appendix (Table A12). Mutagenesis was carried out in 50 µl reactions containing 5 µl of 10X reaction buffer (Agilent), 50 ng of DNA, 1.25 µl of both primers (100 µM; Sigma), 1 µl of dNTP mix (Agilent), and 1 µl of *PfuUltra* HF DNA polymerase (2.5 U/µl; Agilent) made up to 50 µl with H_2O (Sigma). Thermal cycling conditions were 95°C for 30 seconds, followed by 18 cycles of 95°C for 30 seconds, 55°C for 1 minute and 68°C for 7 minutes, then cooling at 37°C for 2 minutes. To digest the non-mutated parental product, 1 µl of the *DpnI* restriction enzyme (10 U/µl; Agilent) was added and incubated at 37°C for 1 hour. One microlitre of the digested product was then transformed into 50 µl of XL1-Blue Supercompetent cells (Agilent) by placing on ice for 30 minutes, heat shocking for 45 seconds at 42°C and incubating on ice for a further 2 minutes. The cells were then added to 200 µl of SOC medium (Invitrogen), transferred to a 15 ml centrifuge tube and incubated in a shaking incubator at 37°C for 1 hour at 250 rpm. Following incubation, 250 µl of the cells were plated onto an ampicillin plate treated with 40 µl of 5-bromo-4-chloro-indolyl-β-D-galactopyranoside (X-gal; 40 mg/ml; Sigma) in dimethyl sulfoxide (DMSO; Sigma) and incubated overnight at 37°C. Ten white colonies were selected for PCR amplification as previously described. Three positive colonies were sent away to Source Bioscience for Sanger sequencing. Sequencing primers are listed in the appendix (Table A11).

2.14 Database Search of miRNA Binding Sites

PolymiRTS (http://compbio.uthsc.edu/miRSNP), TargetScan (http://www.targetscan.org/), MiRBase (http://www.mirbase.org/), PicTAR (http://pictar.mdc-berlin.de/) and miRWalk (http://www.umm.uni-

heidelberg.de/apps/zmf/mirwalk/) were used to identify sequence variations in putative miRNA binding sites and to predict binding sites for miRNAs encompassing rs12595334 and rs3743342.

2.15 Cell Culture of Vascular Smooth Muscle Cells (VSMCs)

VSMC (a kind gift from Drs Aida Bertoli-Avella and Ingrid van de Laar, Rotterdam) from aortic tissue were cultured in complete smooth muscle cell basal medium (Lonza) supplemented with 5 % FBS (Lonza), 0.1% human recombinant epidermal growth factor (rhEGF; Lonza), 0.1% human recombinant insulin, 0.2% human recombinant fibroblast growth factor B (hrFGF-B; Lonza), and 500 µl of GA-100 (gentamicin sulphate/amphotericin-B; Lonza). Cells were maintained in 10 cm tissue culture dishes coated in 0.1% gelatin, in a humidified 5% CO_2 incubator. Cells were passaged at 80% confluence.

2.16 Haplotype Analysis by Cloning

2.16.1 PCR of VSMC DNA

DNA and RNA were extracted from vascular smooth muscle cells and cDNA was synthesised as previously described in sections 2.3 and 2.4. Primers were designed to amplify a cDNA fragment containing the SNPs to be haplotyped. Primer sequences are listed in the appendix (Table A13). PCR was carried out in 50 µl reactions containing 2 µl of cDNA, 0.25 µl of each primer (100 µM; Sigma), 5 µl of 10X PCR buffer II (Life Technologies), 4 µl of $MgCl_2$ (25mM; Life Technologies), 1.25 µl of dNTPs (10 mM each dATP, dCTP, dGTP, dTTP; Bioline), and 0.27 µl of AmpliTaq Gold DNA Polymerase (5 U/µl; Life Technologies), made up to 50 µl with H_2O (Sigma). Thermocycling conditions were an initial denaturation at 96°C for 14 minutes, followed by 35 cycles of 96°C for 30 seconds, annealing at 70°C for 30 seconds, and extension at 72°C for 1 minute then a final extension at 72°C for 7 minutes.

2.16.2 Cloning

Cloning was carried out using the TOPO TA Cloning Kit (Invitrogen). Briefly, 1 μl of PCR product was incubated at room temperature for 30 minutes with 1 μl of salt solution (1.2 M NaCl, 0.06 M $MgCl_2$; Invitrogen), 3 μl of H_2O (Sigma), and 1 μl of pCR 4-TOPO vector (Invitrogen).

2.16.3 Transforming Competent Cells

Two microlitres of the TOPO Cloning reaction were added to one 50 μl vial of One Shot Mach1 cells (Invitrogen) and incubated on ice for 20 minutes. The cells were then heat shocked at 42°C for 30 seconds followed by incubation on ice for a further 2 minutes. The transformed cells were incubated in a shaking incubator at 37°C for 1 hour at 200 rpm with 200 μl of SOC medium (Invitrogen). This broth was then spread onto two room temperature LB (Sigma) agar plates containing ampicillin (100 mg/ml; Sigma), 50 μl on one plate and 100 μl on the other. Plates were incubated overnight at 37°C.

2.16.4 Analysing Transformants

Three colonies were picked for each patient and PCR amplified as previously described. Clones displaying the correct size PCR product were then grown overnight and sent to Source Bioscience to be DNA sequenced. Primer sequences are listed in the appendix (Table A11). The WT and mutated construct sequences were compared using the Clustal Omega alignment tool (http://www.ebi.ac.uk/Tools/msa/clustalo/).

2.17 Bioinformatic Analysis of Protein Prediction Databases

PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/), SNPs3D (http://www.snps3d.org) and the Craig J. Venter Institute's SIFT database (http://sift.jcvi.org), were used to predict the structural and functional effects of the non-synonymous rs11177 SNP, on human protein.

2.18 siRNA Knockdown

2.18.1 Cartilage Digestion and HAC Culture

Articular cartilage was removed from the joint using a scalpel blade either on the day of, or up to 24 hours after total joint replacement surgery. The removed cartilage was washed thoroughly in PBS, cut very finely with a scalpel, transferred to a 50 ml centrifuge tube and weighed. Cartilage was digested by incubating with 4 ml/g of hyaluronidase (type I-S from bovine testes; Sigma) in PBS (Lonza) to a concentration of 1 mg/ml. The enzyme was filter-sterilised on to the cartilage and placed at 37°C for 30 minutes. The cartilage was washed twice with PBS (Lonza) and incubated with 4 ml/g of filter-sterilised trypsin (type III from porcine pancreas; Sigma), in PBS (Lonza) to a concentration of 2.5 mg/ml, at 37°C for 30 minutes. The cartilage was then washed with DMEM (Life Technologies) supplemented with 10% FBS (Life Technologies), 2 mM L-glutamine (Sigma), 100 U/ml penicillin, 100 μg/ml streptomycin (Sigma) and 50 U/ml nyastatin (Sigma) to inactivate the trypsin. The cartilage was then incubated in a shaking incubator at 35°C overnight with 3 ml/g of bacterial collagenase (type I from *Clostridium histolyticum*; Sigma) in DMEM (Life Technologies) to a concentration of 2.5 mg/ml. The supernatant was removed and centrifuged at 1100 rpm for 10 minutes to collect the cells. The cell pellet was then washed with PBS (Lonza) and resuspended in serum containing DMEM (Life Technologies). The cells were counted in a 1:10 dilution with PBS (Lonza) using a haemocytometer. Cells were seeded at a density of 3 million cells/ 75 cm^2 tissue culture flask (Corning) in 10 ml DMEM (Life Technologies) supplemented with 10% FBS (Life Technologies), 2 mM L-glutamine (Sigma), 100 U/ml penicillin and 100 μg/ml streptomycin (Sigma), and 50 U/ml nyastatin (Sigma). Cells were passaged after 10-12 days.

2.18.2 Human MSC Culture

Human bone marrow stem cells from human bone marrow mononuclear cells taken from the iliac crest of young Caucasian donors between the ages of 19-24 (Lonza) were isolated by adherence to 75cm^2 tissue culture flasks (Corning) over 24 hours. Patient details are listed in the appendix (Table A1). Remaining cells were then expanded in monolayer culture in MSC basal medium (Lonza) supplemented with 50 ml MSC

growth supplement (Lonza), 10 ml L-glutamine (Lonza), 500 µl GA-100 (gentamicin sulphate amphotericin-B; Lonza), and 5 ng/ml fibroblast growth factor-2 (FGF2; R&D Systems Europe Ltd, Abingdon, UK). Cultures were maintained in a humidified 5% CO_2 incubator at 37°C. Cells were split 1:6 once they had reached 80% confluence.

2.18.3 siRNA Transfection for Gene Expression Analysis

Once the HACs and MSCs (Lonza) had become 80% confluent they were passaged using 0.25% 1X trypsin-EDTA (Life Technologies). The cells were seeded at a density of 10,000 cells/well in 96-well culture plates (Corning), with 200 µl of serum-positive DMEM (Life Technologies) for HACs, and serum-positive MSC growth medium for MSCs. Empty wells surrounding the cells were filled with 200 µl PBS to minimise evaporation. After 24 hours HACs and MSCs were transfected with 0.5 µl of *GNL3* siRNA (20 µM; Sigma) and 0.26 µl of DharmaFECT transfection reagent 1 (Thermo Scientific) made up to 30 µl with serum-free DMEM (Life Technologies). Cells transfected with non-targeting siRNA (20 µM; Sigma), untransfected cells treated with serum-free medium and cells treated with transfection reagent only were used as controls. siRNA sequences are listed in the appendix (Table A14). Firstly, the required amount of siRNA (*GNL3* or non-targeting), and the required amount of transfection reagent were each mixed with serum-free medium and incubated at room temperature for 5 minutes before combining the siRNA with the transfection reagent. The reaction mixture containing siRNA, transfection reagent and serum-free medium was then incubated at room temperature for 20 minutes. Each transfection condition was carried out in 6 biological replicates. Plates were incubated in a humidified 5% CO_2 incubator for 48 hours then lysed and reverse transcribed into cDNA for gene expression analysis.

2.18.4 Cell Lysis and Reverse Transcription

Medium was removed from the wells and the cells were washed with PBS (Lonza). The cells were placed on ice and 30 µl of Cells-to-cDNA II Cell Lysis Buffer (Ambion) was added to each well. The cells were mixed and centrifuged and the suspension was transferred to a 96-well PCR plate. The cell suspension was heated to 75°C for 15 minutes following which 8 µl was transferred to a fresh 96-well PCR plate. The mix

was then incubated with 1 μl of random primers (0.2 μg/μl; Invitrogen) and 3 μl of dNTP mix (10 mM each of dATP, dCTP, dGTP and dTTP; Bioline) at 70°C for 5 minutes. The RNA was then reverse transcribed by adding 4 μl of First-strand Buffer (5X; Invitrogen), 0.5 μl of SuperScript II RT enzyme (200 U/μl; Invitrogen), 2 μl of DTT (0.1 M; Invitrogen) and 1.5 μl of DEPC-treated H_2O (Invitrogen). The thermocycling conditions were 42°C for 50 minutes then 75°C for 15 minutes. When the reverse transcription was complete the cDNA product was diluted in 50 μl of H_2O (Sigma) and stored at -20°C until required.

2.18.5 siRNA Transfection for Western Blot Analysis

HACs and MSCs (Lonza) were seeded at a density of 300,000 cells/well in 6-well culture plates (Corning), with 2 ml of serum positive DMEM (Life Technologies) for HACs, and serum-positive MSC growth medium for MSCs. After 24 hours HACs and MSCs were transfected as previously described with 5 μl of *GNL3* siRNA (20 μM; Sigma), 6.5 μl of DharmaFECT transfection reagent 1 (Thermo Scientific) made up to 500 μl with serum-free medium. Cells transfected with non-targeting siRNA (20 μM; Sigma), untransfected cells treated with serum-free medium and cells treated with transfection reagent only were used as controls. siRNA sequences are listed in the appendix (Table A14). Cells were incubated in a humidified 5% CO_2 incubator for 48 hours then lysed. Each condition was carried out on three biological replicates each with six technical replicates.

2.18.6 Protein Extraction

Medium was aspirated from the wells and the cells were washed with PBS (Lonza). The plate was placed on ice and 120 μl of protein lysis buffer was added, consisting of 50 mM Tris-HCl (Sigma), 50 mM sodium fluoride (NaF; Sigma), 10 mM sodium orthovanadate (Na_3VO_4; Sigma), 0.5 mM sodium pyrophosphate ($Na_4P_2O_7$; Sigma), 1 mM ethyleneglycoltetraacetic acid (EGTA; Sigma), 0.0146g EDTA (Sigma), 10 mM glycerol phosphate (Sigma), 10 % glycerol (Sigma), 10 ml Triton X-100 (Sigma), 1X complete inhibitor cocktail (Roche, Indiana, USA), 1 μM microcystin-LR (Sigma), and made up to 50 ml with deionised H_2O. The cells were scraped in the protein lysis buffer

using a cell scraper (Corning) and the lysate was transferred to a 1.5 ml centrifuge tube. The cell lysate was incubated on ice for 20 minutes and mixed by flicking after 10 minutes. Following this the tubes were centrifuged at 13,000 rpm at 4°C for 10 minutes and the supernatant containing the protein was transferred to a fresh 1.5 ml centrifuge tube and stored at -80°C until required.

2.18.7 Protein Quantification using Bradford Ultra

To ensure equal protein loading in Western blot analysis, the amount of protein was quantified using a standard curve of BSA (2 mg/ml; Thermo Scientific) from 0 to 1.5 mg/ml in protein lysis buffer. Each of the proteins to be measured were diluted 1:10 with protein lysis buffer and 20 µl added to two wells of a microtitre plate (Corning). To the protein, 300 µl of Bradford Ultra (Expedeon, Harsdon, UK) was added and the absorbance at 595 nm was measured using a Tecan microplate reader (Tecan, Reading, UK). The absorbance values were used to plot a BSA standard curve and the equation y = mx + b from the graph was used to calculate the volume of protein where y is the absorbance and x is the protein concentration.

2.19 Western Blotting

2.19.1 Gel Preparation

A 12% separating gel was made, consisting of 3.6 ml of acrylamide bis-acrylamide (40%; NBS Biologicals, Huntingdon, UK), 3 ml of lower buffer (pH 8.8; 1.5 M Tris, 0.4% w/v sodium dodecyl sulphate; SDS; Sigma), 5.4 ml of deionised H_2O, 30 µl of ammonium persulphate (APS; 20%; Sigma) and 20 µl of N, N, N′, N′-tetramethylethylenediamine (TEMED; Sigma). The space between the glass plates was filled with 6 ml of the separating gel and 1 ml of propan-2-ol was added to the top to prevent evaporation of the gel. After 30 minutes a 4.5% stacking gel was made consisting of 1.35 ml of acrylamide bis-acrylamide (40%; Sigma), 3 ml of upper buffer (pH 6.8; 0.5M Tris, 0.4% w/v SDS; Sigma), 7.65 ml of deionised H_2O, 30 µl of APS (20%) and 20 µl of TEMED (Sigma). The propan-2-ol was removed, the top of the gel was washed with deionised H_2O, 3 ml of stacking gel was added to the top of the gel

and the comb was inserted. After 15 minutes the comb was removed from the set stacking gel and the glass plates were placed into a gel tank (Bio-Rad, Hertfordshire, UK) with running buffer (1X; pH 8) made up of 5.86 g glycine (Sigma), 20 g SDS (Sigma) and 60.6 g Tris base (pH 7-9; Sigma) in 2 L deionised H_2O. For loading, 10 μg of the sample was mixed with 4 μl of Laemmli buffer (5X; 0.1M Tris-HCl; Sigma; 0.35 M SDS; Sigma; 20% (v/v) glycerol; Sigma; 0.01% (v/v) bromophenol blue; Sigma; 5% (v/v) BME; Sigma) and made up to 20 μl with H_2O (Sigma). Samples were boiled at 95°C for 5 minutes to denature the proteins. The protein was loaded and electrophoresed at 100 V for 15 minutes at room temperature increasing to 150 V for 1 hour.

2.19.2 Electrophoretic Transfer of the Proteins to PVDF Membrane

A piece of polyvinylidene fluoride (PVDF; GE Healthcare) and four pieces of blotting paper (GE Healthcare) were cut to the same size as the gel. The blotting paper was soaked for 5 minutes in transfer buffer consisting of 5.86 g glycine (Sigma) 0.65 g SDS (Sigma) 11.62 g Tris base (pH 7-9; Sigma) 400 ml methanol (100%) made up to 2 L with deionised H_2O. The PVDF membrane was soaked for 3 minutes in methanol (100%). The gel had finished plates were carefully removed and the stacking gel discarded. The separating gel was then soaked in transfer buffer for 3 minutes. The blotting paper, membrane and gel were assembled on a semi-dry transfer machine (Bio-Rad); two pieces of blotting paper then the membrane followed by the gel and the remaining two pieces of blotting paper. Air bubbles were removed with a roller, and the proteins were transferred to the membrane at 80 mA per gel for 1 hour 15 minutes.

2.19.3 Blocking

Once the proteins were completely transferred, the membrane was blocked for 30 minutes in 20 ml Tris-buffered saline-Tween (TBS-T; pH 7.4; 12.1 g Tris; Sigma; 90 g NaCl; Sigma; 10 ml 10X Tween-20; Sigma made up to 1 L with deionised H_2O) with 5% milk powder (Marvel, Hertfordshire, UK).

2.19.4 Incubation with Antibody

The membranes were incubated on a rocker overnight at 4°C with anti-nucleostemin rabbit polyclonal antibody (Abcam, Cambridge, UK; 1:2000) in TBS-T with 5% milk powder (Marvel), or anti-β-actin mouse polyclonal antibody (1:10,000; Sigma) in TBS-T with 5% milk powder (Marvel), as a control. The membranes were then incubated on a rocker at room temperature for 1 hour with polyclonal goat anti-rabbit secondary antibody or polyclonal goat anti-mouse (1:2000; 0.25 g/L; Dako, Ely, UK) in TBS-T with 5% milk. Following this the membranes were washed with TBS-T for 30 minutes, changing the buffer every 5 minutes.

2.19.5 Detection

Membranes incubated with anti-GNL3 were detected using Enhanced Chemiluminescent reagent (ECL) prime (GE Healthcare) and membranes incubated with anti-β-actin were detected using ECL (GE Healthcare). The membranes were incubated at room temperature for 5 minutes with 1 ml of ECL or ECL prime reagents 1 and 2 in a 1:1 ratio. The membrane was exposed 5 times at 20-second intervals using a G:BOX gel doc system (Syngene). If the protein was undetectable after this exposure time then the exposure time was increased to 1-minute or 5-minute intervals.

Chapter 3: Gene Expression Analysis of the Osteoarthritis Susceptibility Locus Mapping to Chromosome 7q22

3.1 Introduction

A GWAS of hip, knee and hand OA was performed on female Dutch Caucasian OA cases and controls from the Rotterdam study (131). The GWAS tested 500,510 SNPs and identified 18 SNPs at 12 loci with associations to OA. The 12 loci were tested in 12 further studies of hip, knee and hand cases as well as controls. Only one locus reached GWS after replication; a locus on 7q22 that was associated with knee and/or hand OA, with an OR of 1.14 and a P-value of 8×10^{-8} for the G allele of SNP rs3815148 within intron 12 of *COG5*. A subsequent study of knee cases and controls from 8 additional European cohorts to those originally investigated in the Dutch report generated more significant evidence for association of OA to the 7q22 locus (149). This meta-analysis identified an additional signal at 7q22 with an OR of 1.17 and a P-value of 9.2×10^{-9} for the G allele of SNP rs4730250 within intron 3 of *DUS4L.*

The 7q22 locus contains six genes in a 500kb region of high LD with pair-wise r^2 and D′ values between rs3815148 and rs4730250, of 0.82 and 1.0, respectively. In physical order the six genes from within the association signal are*: PRKAR2B, HBP1, COG5, GPR22, DUS4L*, and *BCAP29* .

3.1.1 PRKAR2B

The cyclic adenosine monophosphate (cAMP)-dependent protein kinase (PKA) is a serine/threonine kinase consisting of two catalytic subunits and a regulatory subunit dimer made up of RI and RII subunits. Activation of PKA occurs when two cAMP molecules bind the regulatory dimer, releasing the active catalytic subunits and allowing them to phosphorylate specific substrate proteins on serine and threonine residues (150-152). Both the catalytic and the regulatory subunits exist as multiple isoforms: Cα, Cβ, Cγ, RIα, RIβ, RIIα and RIIβ (150). *PRKAR2B* encodes the RIIβ regulatory subunit of PKA. *PRKAR2B* has limited expression in most tissues but is abundantly expressed in the brain and in brown and white adipose tissue (153, 154). Activation of white adipose tissue is decreased in RIIβ

knockout mice whereas activation of brown adipose is increased, leading to an increased metabolic rate, which protects against diet-induced obesity (153-155). Overexpression studies of the RIIβ regulatory subunit in cancer cell lines have shown that increased RIIβ results in growth arrest, differentiation and maturation of the cells (152).

RIIβ has also been implicated in the autoimmune disease systemic lupus erythematosus (SLE). Deficient PKA activity in the T cells is estimated to affect 37% of SLE patients and is associated with a significant increase in RIIβ levels (156). It is possible that RIIβ acts as a negative regulator of T cell activation as it was demonstrated to bind cAMP response element-binding protein (CREB) *in vitro* leading to inhibition of CREB transcription activity and CRE-dependent transcription in T cells (156).

3.1.2 HBP1

HBP1 is a member of the high mobility group (HMG) box transcription factors and is one of few transcriptional repressors within this family (157, 158). Other HMG box transcription factor members include lymphoid enhancer factor (LEF) and T cell factor (TCF), which are transcription factors activating oncogenes such as Cyclin D1 and c-Myc, via the Wnt/β catenin pathway (158). HBP1 has been shown to be a negative regulator of the Wnt/β-catenin pathway (159) through binding TCF4 and preventing subsequent binding of TCF4 to its target genes (160).

It has been shown that HBP1 interacts with proteins of the retinoblastoma (RB) family, which are growth suppressor proteins with a role in cell cycle regulation (161, 162). *HBP1* is the target of RB in the establishment of terminal differentiation and a large increase in *HBP1* mRNA has been seen during terminal differentiation (163) indicating that *HBP1* has a role in the progression of the cells through the cell cycle (158). RB along with another tumour suppressor gene, p53, is an important factor in both replicative and premature senescence. RB binding is essential for HBP1-induced premature senescence triggered by either RAS or p38 mitogen-activated protein kinase (MAPK) signalling (159).

HBP1 has also been implicated in the regulation of superoxide production whereby nicotinamide adenine dinucleotide phosphate (NADPH) oxidase catalyses the reduction of O_2 to the superoxide anion O_2^-. NADPH oxidase comprises cytoplasmic regulatory subunits p47phox, p67phox and p40phox that combine with a membrane complex containing a

catalytic subunit. HBP1 has been demonstrated to bind to and repress the gene for the p47phox subunit reducing cell cycle progression, and superoxide levels (164).

3.1.3 COG5

The conserved oligomeric Golgi (COG) complex is essential for Golgi structure (165). Mammalian COG is a large complex consisting of eight polypeptides (COG1-8; 166). These eight subunits are divided into two lobes COG1- 4 and COG5-8 bridged by a interaction of COG1 and 8; *COG5* encodes the COG5 subunit. Mutations in COG genes can cause congenital disorders of glycosylation (CDG; 165), which can cause milder phenotypes such as psychomotor retardation, language delay and dysmorphism, or more severe phenotypes such as hormonal disorders, coagulopathies, multiple organ failure and skeletal malformations (166, 167). Mutations within COG5 have been shown to cause the milder phenotypes (166). Around 70% of all proteins are glycosylated and these glycoproteins have a key role in metabolism including correct protein folding, cell recognition, and cell adhesion (166). The Golgi apparatus is a hub for membrane trafficking pathways and plays a key role in the secretion of glycoproteins, glycolipids and proteoglycans (168). Vesicular trafficking of proteins and lipids through the Golgi requires that vesicles recognize and fuse to the membrane compartments (169). It is thought that the COG complex acts as a tether for this vesicular transport through the Golgi compartments (167).

3.1.4 GPR22

GPR22 is referred to as an orphan G protein-coupled receptor as the natural ligand remains undiscovered (170). However, *GPR22* mRNA has been shown to be expressed in the cardiac myocytes and coronary arteries and the protein couples to receptor Gi/Go, which suggests a role in cardiac contractile function (170, 171). Increased Gi protein expression during heart failure has been associated with the decreased contractile response to activation of the β-adrenergic pathway, and this, coupled with the fact that *GPR22* knockout mice have increased susceptibility to heart failure, indicates that there may be a role for *GPR22* in the transition process from hypertrophy to heart failure (170).

3.1.5 DUS4L

DUS4L, also known as *PP35,* encodes dihydrouridine synthase 4-like. Dihydrouridine is one of the most common modified nucleosides (172), and is formed by the reduction of the carbon-carbon double bond by dihydrouridine synthases, mostly at the D loop of tRNAs. Functionally they may protect the tRNA against degradation as it has been demonstrated that tRNAs degrade significantly faster when there are no dihydrouridines present (173). *DUS4L/PP35* has also been linked to Fanconi anaemia (FA), an autosomal recessive immunodeficiency, after being shown to interact with the FA complementation group A (FANCA) protein. Amongst the characteristics of FA are skeletal abnormalities and growth retardation (174).

3.1.6 BCAP29

BCAP29, or *BAP29,* codes for a 29kDa B cell receptor-associated protein. BCAP29 forms a complex with the closely related BCAP31 (175, 176). BCAP29 and BCAP31 both possess a dilysine motif (KKXX) at the C terminus, which is required for localisation of proteins to the endoplasmic reticulum (ER; 177-179) so it is thought that the BCAP29/BCAP31 complex is involved in the process of protein exportation from the ER (179). Immature B cells express only IgM antigen receptors and, once they leave the bone marrow they mature and begin to express both IgM and IgD (176). The antigen receptor on mature B cells (BCR) comprises one membrane bound immunoglobulin (mIg) for example IgM or IgD and an Igα/Igβ heterodimer (178). It has been demonstrated in mice that when co-expressed with BCAP29/31, the transmembrane form of the IgD BCR is not exported to the surface in the absence of the Igα/Igβ heterodimer, however when there is no BCAP29/BCAP31 complex the incomplete BCR is exported to the surface (179). It was therefore concluded that BCAP29/BCAP31 complexes are involved in the ER retention of mIg molecules in the absence of Igα/Igβ (178).

The Dutch 7q22 association report investigated the expression of *PRAKR2B, HBP1, COG5* and *DUS4L* in cultured synovial fibroblasts and chondrocytes, where they detected abundant *HBP1* and *COG5* expression but very low expression levels of *PRKAR2B, DUS4L* and *GPR22 (*131). In their discovery of the 7q22 OA susceptibility locus, Kerkhof and colleagues noted that *GPR22* harbored a common SNP that was associated with *GPR22*

expression in transformed lymphoblast cell lines. Whether this is relevant to the pathology of OA or simply a coincidence reflecting the existence of cell-type specific eQTLs was not resolved. Based upon this, they performed immunohistochemistry on joints from normal and osteoarthritic mice and on osteophytes from mice with instability-induced OA. GPR22-positive chondrocytes were found in the osteophytes of the mice with OA but not in normal mice chondrocytes (131). Evangelou *et al*. analysed the expression of all six genes in chondrocyte pellet cultures compared to monolayer cultures and observed that *HBP1, COG5, DUS4L,* and *BCAP29* were all expressed at higher levels in pellet cultures, which replicate the environment seen in articular cartilage, whereas *PRKAR2B* and *GPR22* showed no difference in expression between pellet cultures and monolayer (149). They also analysed expression in zebrafish to explore possible roles of these genes during embryogenesis and detected expression of all six genes. The overall conclusion from these expression studies was that all of the genes showed near universal expression patterns in the tissues examined, except *GPR22*, which was only seen to be expressed in cultured chondrocytes and in cartilage from mice with instability-induced OA (131).

3.2 Aim

The aim of this chapter is to assess whether the six genes are subject to *cis*-acting regulatory polymorphisms that are active in joint tissues, which could contribute to the association signal.

3.3 Results

3.3.1 Expression of the Genes in Human Synovial Joint Tissues. Each gene was analysed for expression in joint tissues by standard qualitative PCR (Chapter 2, section 2.5). Five of the six genes were expressed in all joint tissue cDNAs tested (Figure 3.1), with the exception being *GPR22*. This gene was not expressed in any joint tissue but was expressed in heart, which was used as a positive control for its expression. Subsequently, expression was tested by Real-Time qPCR (Chapter 2, section 2.6-7). Again, all genes demonstrated expression except for *GPR22* (Figure 3.2).

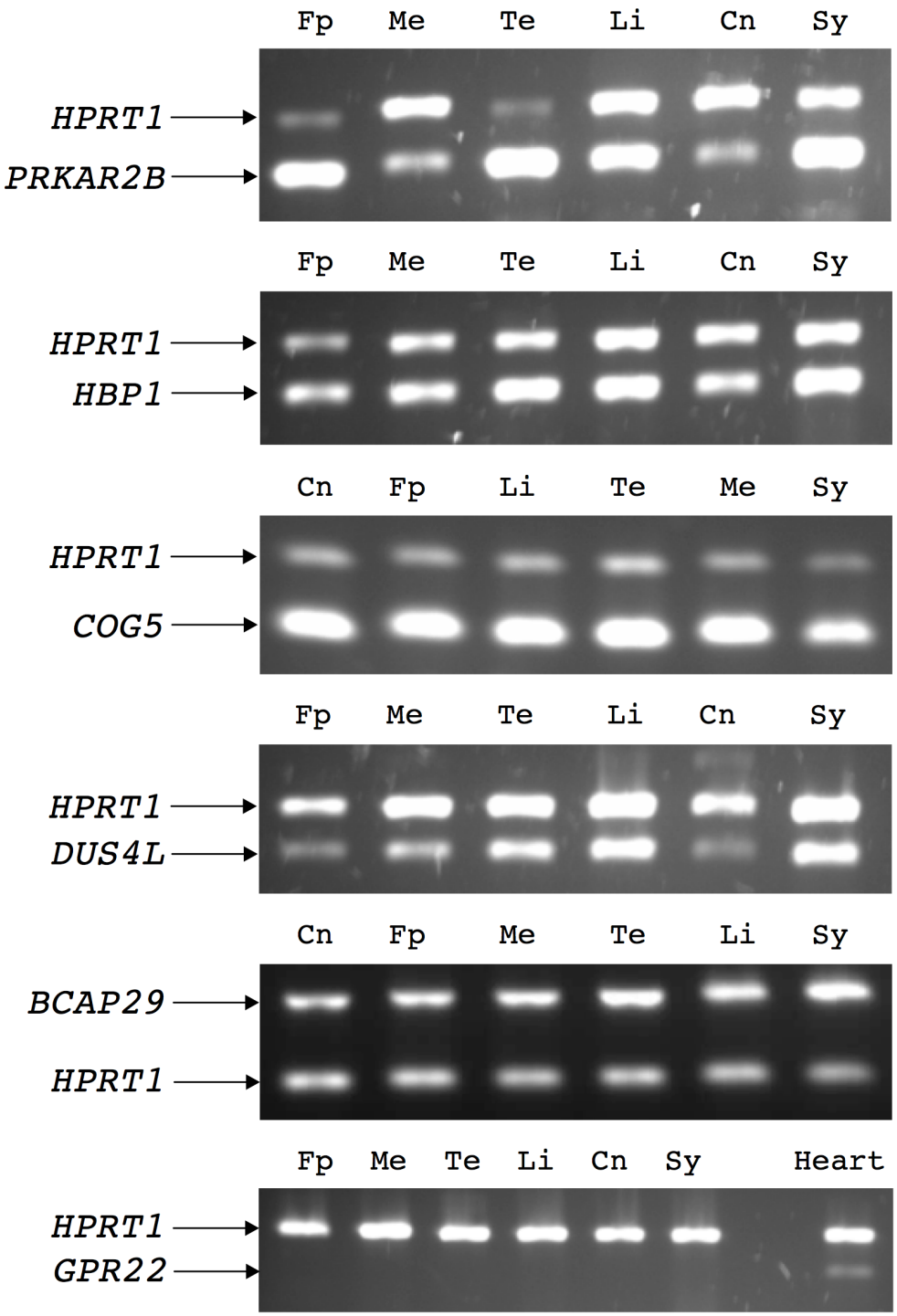

Figure 3.1. Expression analysis of the six chromosome 7q22 genes by standard qualitative PCR. The analysis was performed on pooled cDNA synthesized from RNA extracted from OA knees. The tissues analysed were fat pad (Fp), meniscus (Me), tendon (Te), ligament (Li), cartilage (Cn) and synovium (Sy). The housekeeping gene *HPRT1* was included to confirm the integrity of each cDNA sample. Heart cDNA from a cDNA library was used as a positive control for the *GPR22* primers.

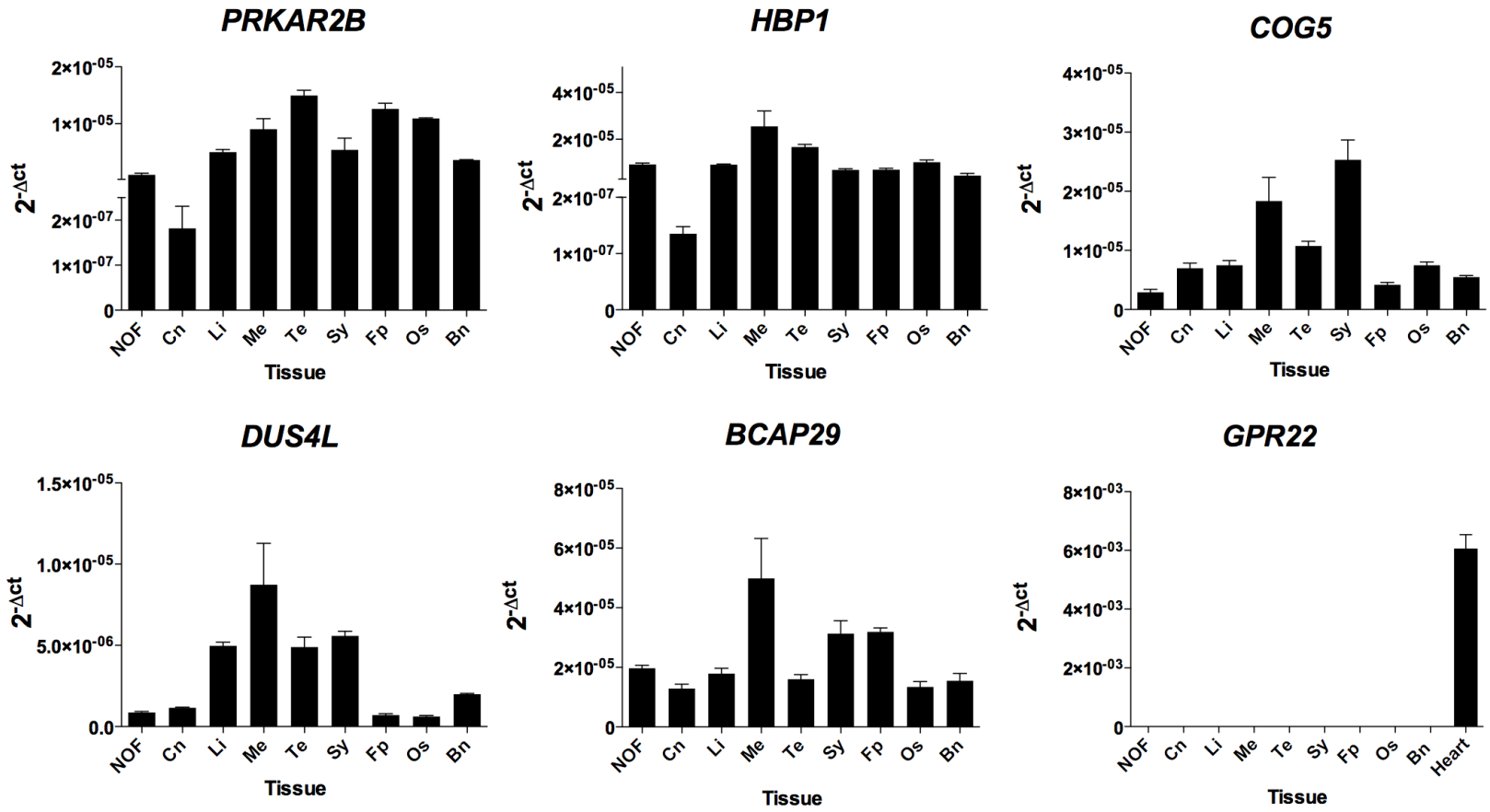

Figure 3.2. Column graphs of expression levels of the six 7q22 genes by Real-Time qPCR in: OA cartilage (Cn), cartilage from a patient who had undergone surgery for a neck of femur fracture (NOF), ligament (Li), meniscus (Me), tendon (Te), synovium (Sy), fat pad (Fp), osteophyte (Os) and cancellous bone (Bn). The reactions were carried out in triplicate using cDNA from three different individuals. The results were normalized against the *18S* housekeeping gene. There is no expression of *GPR22* in synovial joint tissues. Heart cDNA was used as a positive control for the *GPR22* assay. The error bars represent the standard error of the mean.

3.3.2 Quantitative Comparison of the Expression of the Genes between the Cartilage from OA and NOF Patients. Expression of the six genes was compared between cartilage from OA patients who had undergone a THR and control cartilage from patients who had undergone hip surgery due to a NOF fracture (Figure 3.3). *GPR22* was not expressed in any of the over 100 cartilage samples tested but it was expressed in heart. Therefore we can conclude that *GPR22* is not expressed at detectable levels in the patient samples used. As already noted, in the discovery of the 7q22 locus, the investigators were able to detect expression of *GPR22* in RNA isolated from cultured chondrocytes (131). It is known that the culturing of cells can alter the expression pattern of genes (180) so it is therefore possible that this previous report on the expression of *GPR22* reflects an artefact of the use of cultured cells. Based on the cartilage results, and on the lack of expression of *GPR22* in the other joint tissues tested, it was decided to exclude this gene from subsequent experiments. Significant reductions in expression in OA cartilage relative to control cartilage were seen for all five of the remaining genes. Next, expression of the genes in cartilage of patients who had undergone TKR was compared with expression of the genes in cartilage from patients who had undergone THR (Figure 3.4). A significant reduction in expression of *PRKAR2B* ($P = 0.04$), *COG5* ($P = 0.003$) and *BCAP29* ($P = 0.03$) in hip cartilage compared to knee cartilage was observed. Conversely, a significant reduction in expression of *DUS4L* in knee cartilage was observed compared to hip cartilage ($P < 0.0001$). Having observed that *GPR22* was not detectable in our tissue samples, that the five remaining genes showed reduced expression in OA cartilage, and that four of the five (*PRKAR2B*, *COG5*, *BCAP29,* and *DUS4L*) showed significantly different expression levels between OA hip cartilage and OA knee cartilage, the samples were analysed for a correlation between expression levels and genotype at the associated SNP to ascertain whether the 7q22 association signal could be operating by modulating expression of one or more of the five genes in joint tissues from patients undergoing total joint replacement (TJR).

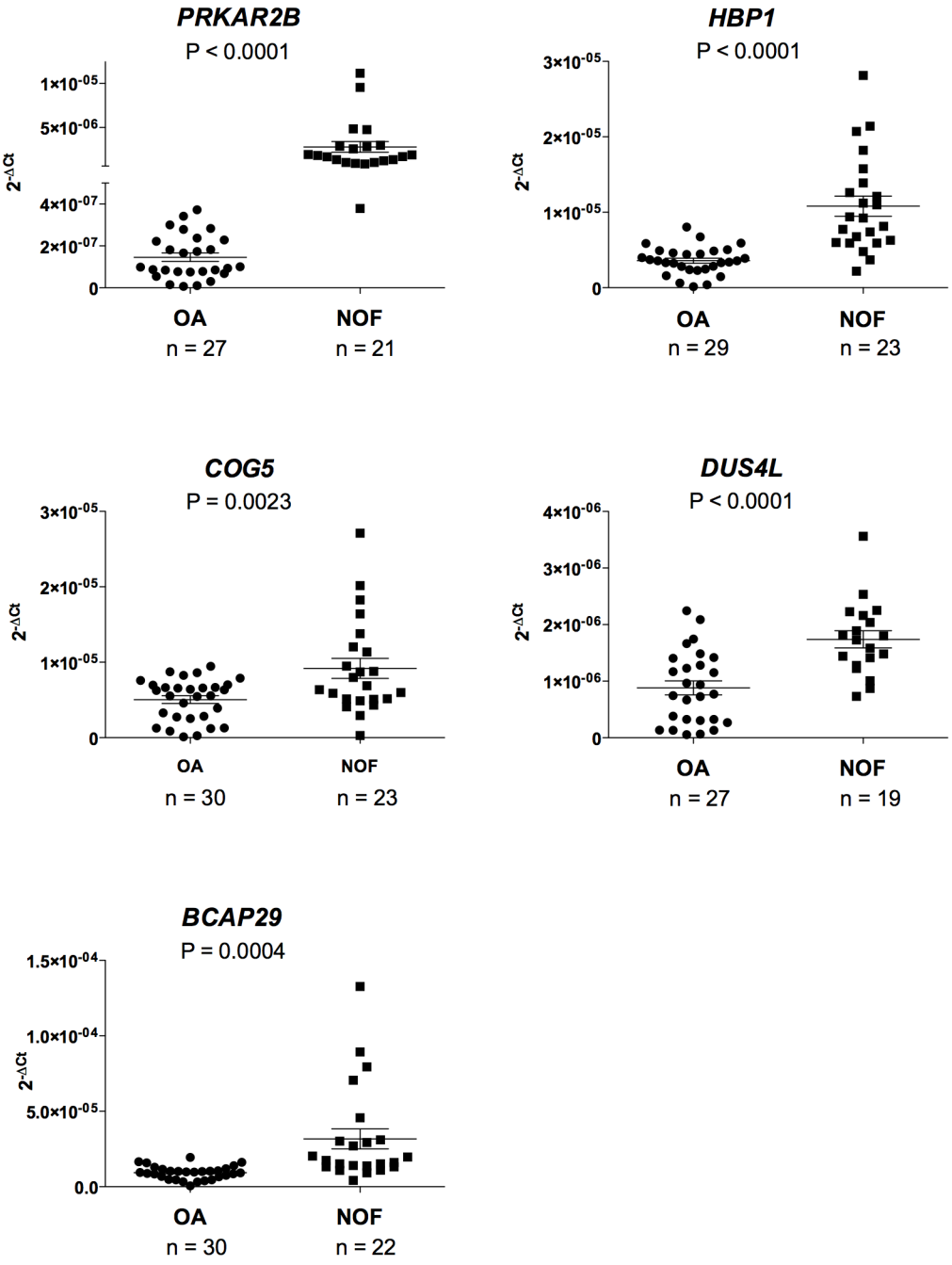

Figure 3.3 Columnar scatter plots of the quantitative expression of the five genes between cDNA from OA hip cartilage and cDNA from control hip cartilage. The control cartilage was collected from patients who had undergone surgery for a NOF fracture. n is the number of patients studied. The horizontal lines in each plot represent the mean and the standard error of the mean. P-values were calculated using a Mann-Whitney U test.

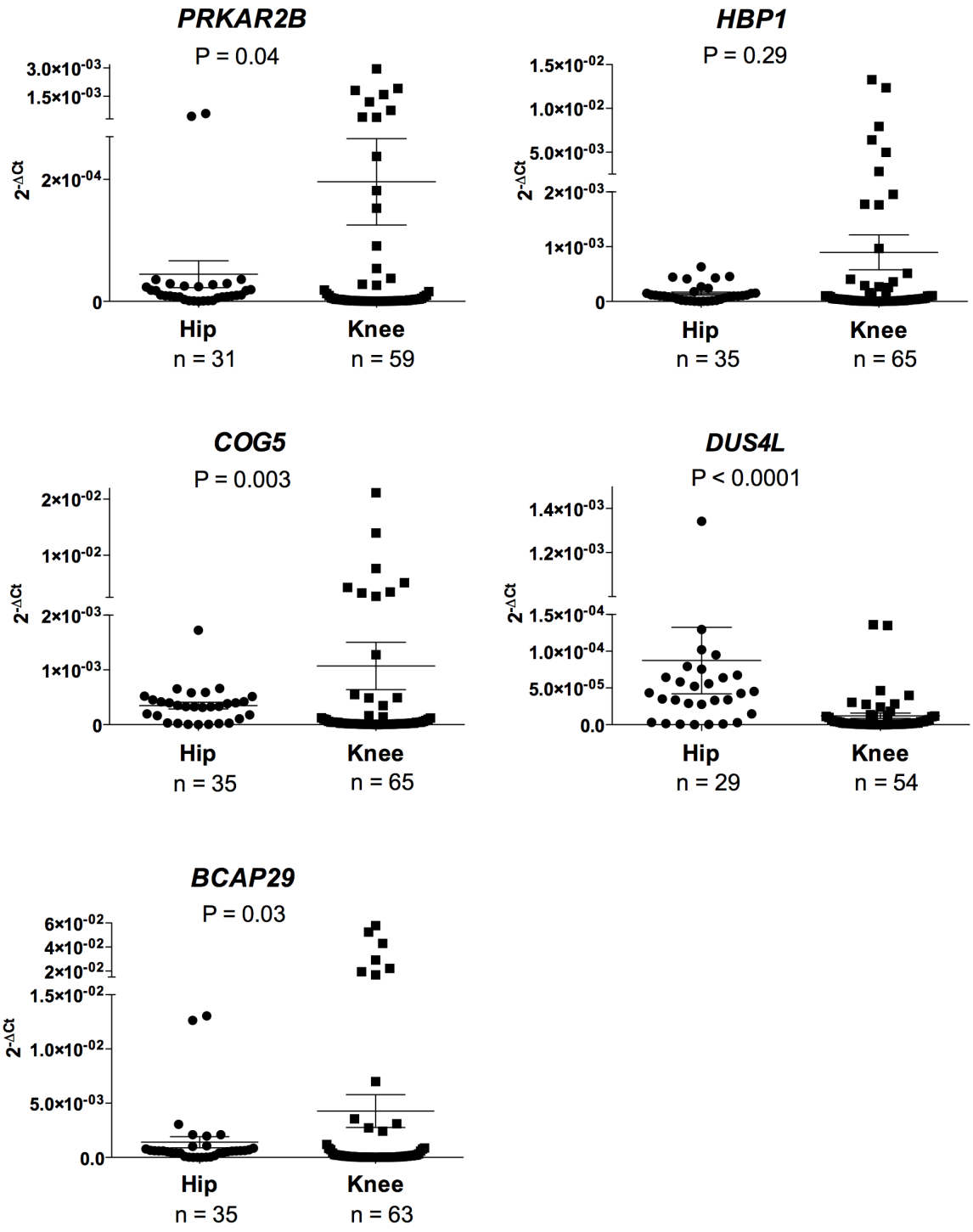

Figure 3.4 Columnar scatter plots of the quantitative expression of the five genes between cDNA from OA hip cartilage and cDNA from OA knee cartilage. n is the number of patients studied. The horizontal lines in each plot represent the mean and the standard error of the mean. P-values were calculated using a Mann-Whitney U test.

3.3.3 Quantitative Expression of PRKAR2B, HBP1, COG5, DUS4L and BCAP29 in Cartilage, Fat Pad, and Synovium Stratified by Genotype at the Associated SNPs. To initially determine whether one or more of the five genes harboured cartilage eQTLs that correlated with the OA association signal, overall levels of expression of each gene were quantified in cartilage samples and the data were stratified by genotype at the two OA associated SNPs, rs3815148 and rs4730250 (Chapter 2, section 2.8). None of the OA patients tested were homozygous for the minor allele of either SNP. This is a reflection of the relatively low minor allele frequencies of the two SNPs; in HapMap CEU rs3815148 has a GG frequency of 4.4% whilst rs4730250 has a GG frequency of 3.5%. The analysis therefore involved a comparison between individuals that were homozygous for the major allele and those that were heterozygous. *HBP1* demonstrated a significant difference in expression between the two genotypic groups for rs3815148 (P = 0.02; Figure 3.5) and rs4730250 (P = 0.003; Figure 3.6). In both instances there was reduced expression of the gene in individuals who carried the OA-associated G alleles at the SNPs. When the analysis was repeated on the TKR cases only, *HBP1* again demonstrated a significant difference in expression in cartilage between the two genotypic groups for rs3815148 (Figure 3.7) and for rs4730250 (Figure 3.8), and once again it was the OA associated G alleles that correlated with reduced expression.

The P-values were however more significant in this TKR analysis compared to the combined analysis, with P-values of 0.004 for rs3815148 and 0.0002 for rs4730250. This clearly aligns with the fact that the 7q22 OA association signal is particularly relevant to knee OA (131, 149). *COG5* also demonstrated a significant difference for rs4730250 in the TKR analysis (Figure 3.8), but the P-value was 100-fold less significant than that seen for *HBP1* (0.02 versus 0.0002), whilst *COG5* was not significant for rs3815148 (P = 0.99; Figure 3.7). All but one of the THR patients were homozygous TT for rs3815148 and homozygous AA for rs4730250, so no genotype comparisons could be made for the THR patients. Fat pad and synovial tissue was also collected from a number of TKR patients, providing the opportunity to assess whether any of the five genes harboured eQTLs relevant to these two joint tissues. As with the cartilage analysis, in fat pad there were no homozygotes for the minor allele of either of the associated SNPs.

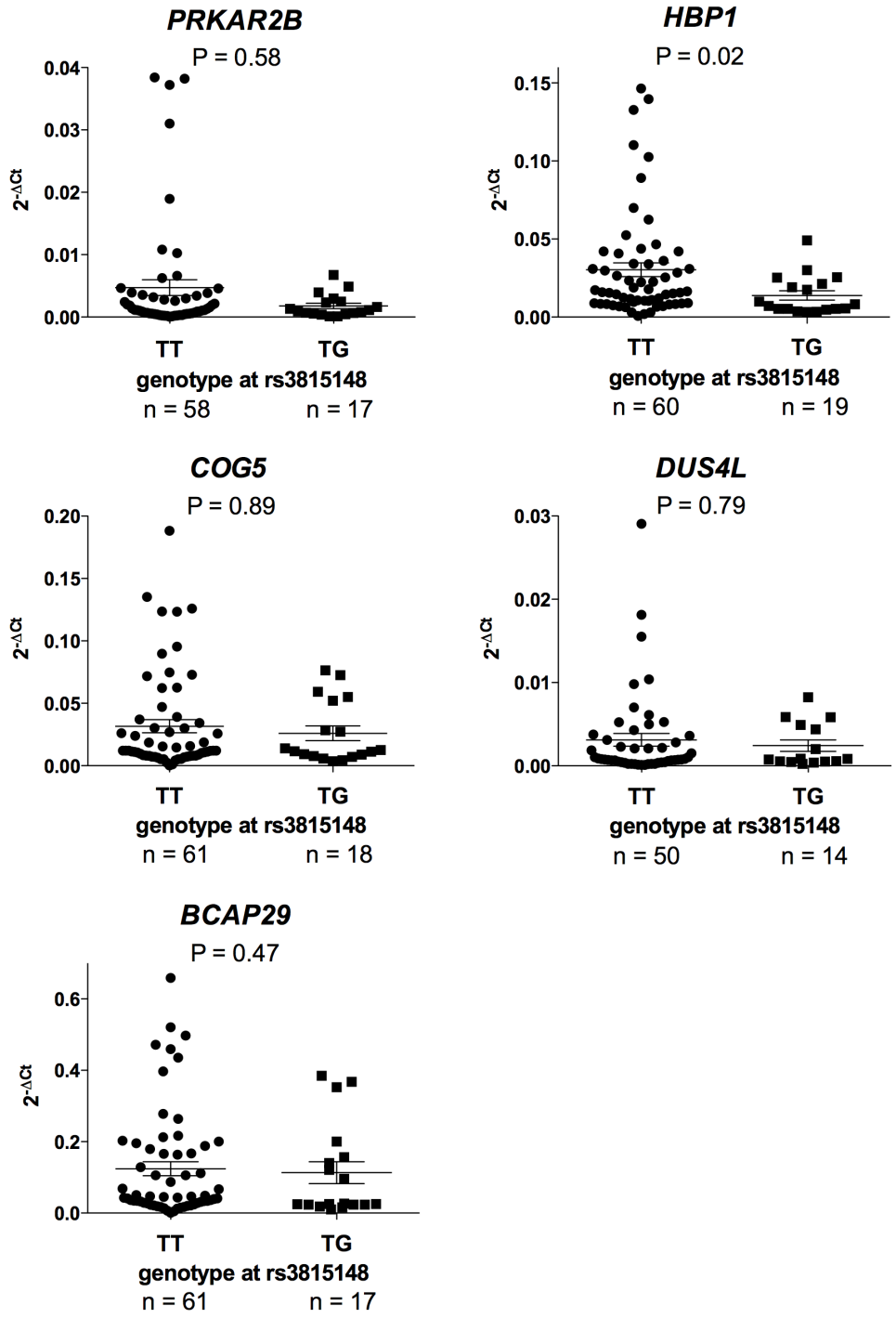

Figure 3.5 Columnar scatter plots of the quantitative gene expression of the 7q22 genes in cDNA from OA hip and OA knee cartilage, with patients stratified by genotype at the associated SNP rs3815148. n is the number of patients studied. Due to the absence of GG homozygotes at rs3815148, the analysis was between TT homozygotes and G-allele carriers (TG). The horizontal lines in each plot represent the mean and the standard error of the mean. P-values were calculated using a Mann-Whitney U test.

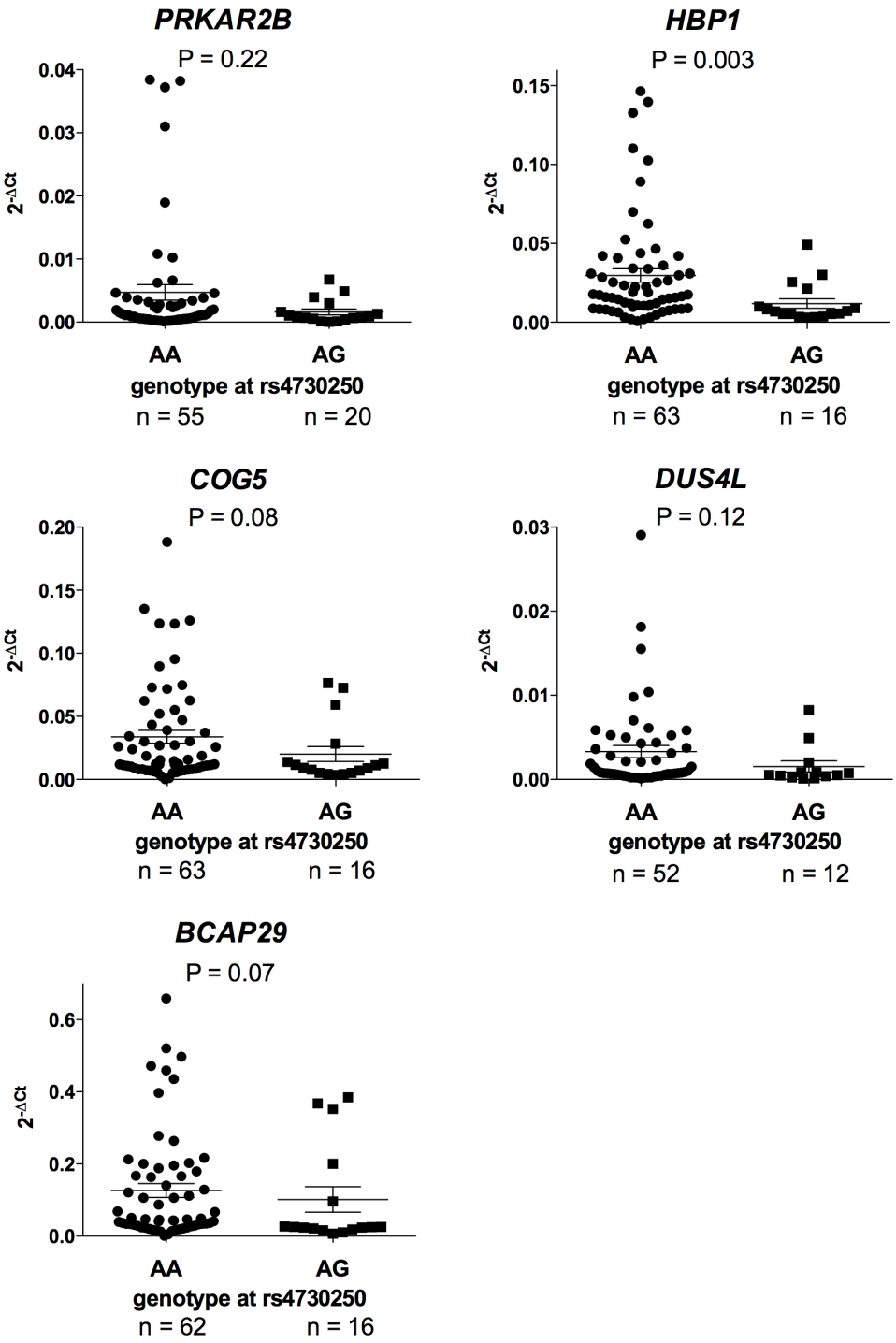

Figure 3.6 Columnar scatter plots of the quantitative gene expression of the 7q22 genes in cDNA from OA hip and OA knee cartilage, with patients stratified by genotype at the associated SNP rs4730250. n is the number of patients studied. Due to the absence of GG homozygotes at rs4730250, the analysis was between AA homozygotes and G-allele carriers (AG). The horizontal lines in each plot represent the mean and the standard error of the mean. P-values were calculated using a Mann-Whitney U test.

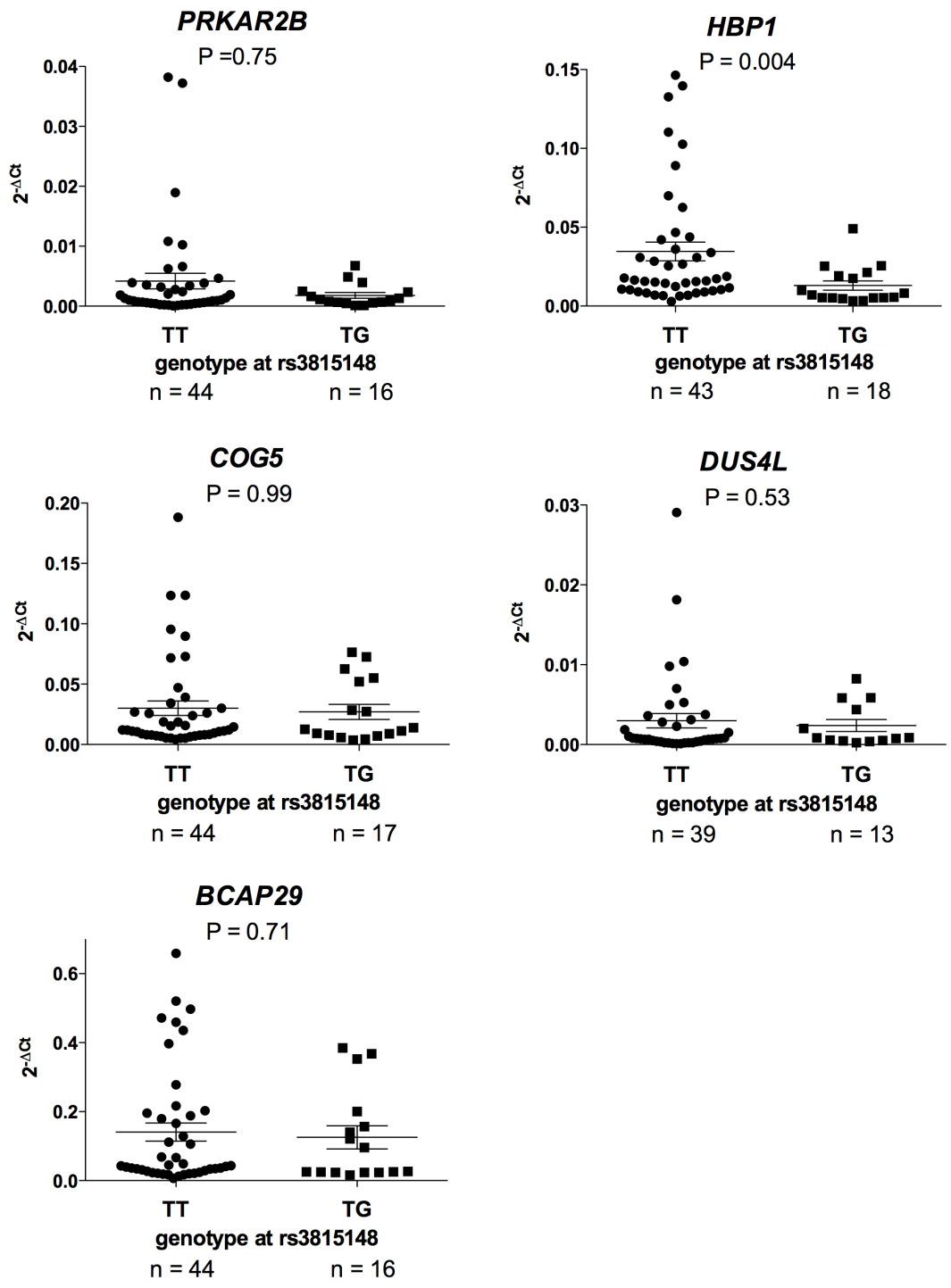

Figure 3.7 Columnar scatter plots of the quantitative gene expression of the 7q22 genes in OA knee cartilage cDNA, with patients stratified by genotype at the associated SNP rs3815148. n is the number of patients studied. Due to the absence of GG homozygotes at rs3815148, the analysis was between TT homozygotes and G-allele carriers (TG). The horizontal lines in each plot represent the mean and the standard error of the mean. P-values were calculated using a Mann-Whitney U test.

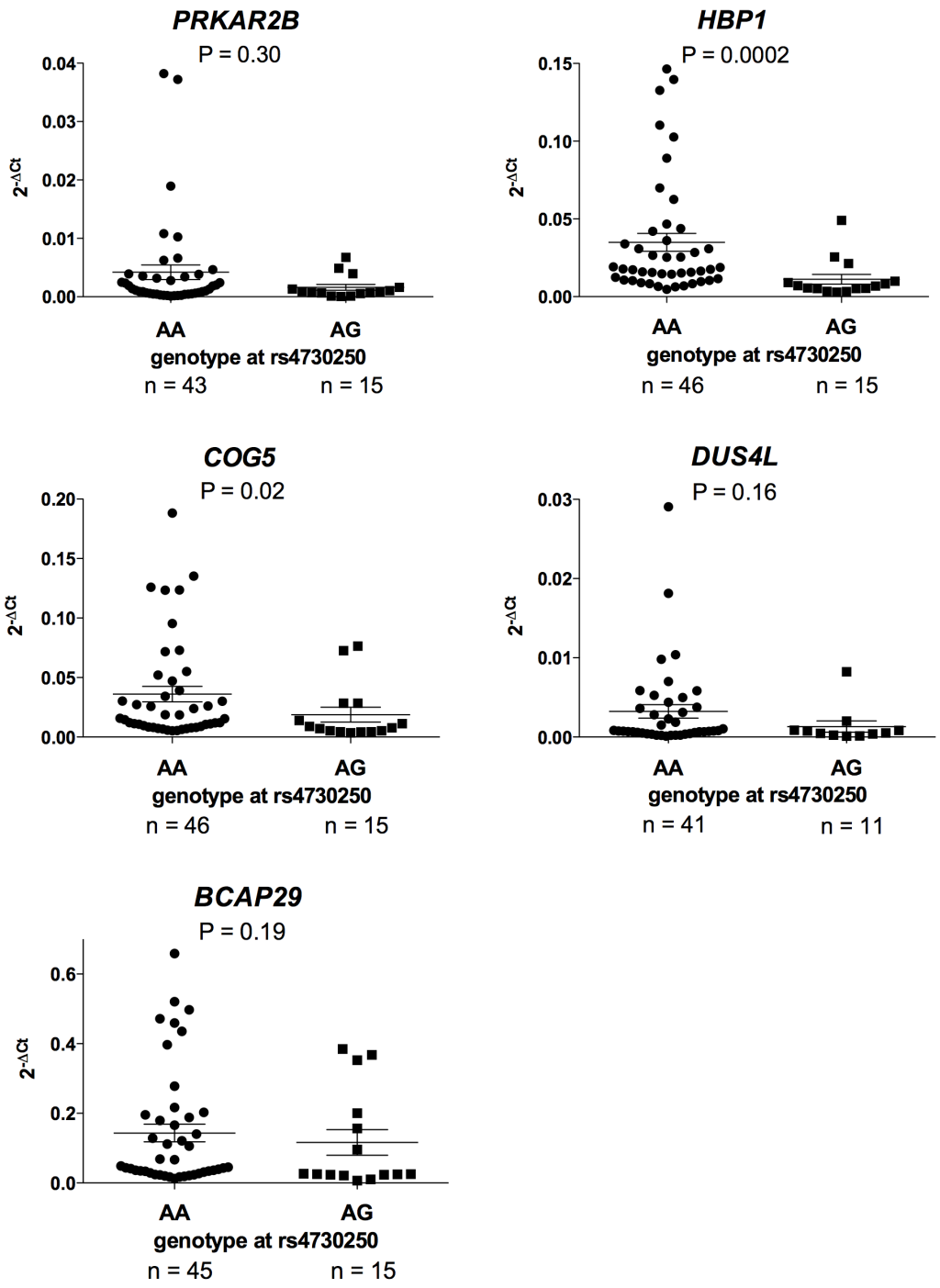

Figure 3.8 Columnar scatter plots of the quantitative gene expression of the 7q22 genes in OA knee cartilage cDNA, with patients stratified by genotype at the associated SNP rs4730250. n is the number of patients studied. Due to the absence of GG homozygotes at rs4730250, the analysis was between AA homozygotes and G-allele carriers (AG). The horizontal lines in each plot represent the mean and the standard error of the mean. P-values were calculated using a Mann-Whitney U test.

The *DUS4L* expression levels in fat pad stratified by rs3815148 approached significance (P = 0.07; Figure 3.9), and a significant difference in expression for *DUS4L* was observed between the two genotypic groups for rs4730250 (P = 0.04, Figure 3.10). In both cases the OA associated G alleles correlated with reduced expression. In the analysis of synovium, only one of the patients was a GG homozygote for rs3815148 and rs4730250. This person was combined with the TG (rs3815148) and AG (rs4730250) heterozygotes and a comparison was made between carriers of the G allele at each SNP and non-carriers. Carriage of the OA associated G allele at rs3815148 correlated with significantly reduced expression of *HBP1* (P = 0.02, Figure 3.11), whilst the data for rs4730250 approached significance for this gene (P = 0.07: Figure 3.12).

This analysis demonstrated that carriers of the OA-associated G allele at both rs3815148 and rs4730250 show reduced *HBP1* expression levels in cartilage, that carriers of the G allele at rs3815148 show reduced *HBP1* expression levels in synovium, and that carriers of the OA-associated G allele at rs4730250 show reduced *DUS4L* expression in fat pad. Both *HBP1* and *DUS4L* are therefore subject to *cis*-regulatory eQTLs that correlate with the OA association signal at 7q22. To further investigate these eQTLs the relative allelic expression levels of *HBP1* and *DUS4L* were then measured in various joint tissues.

3.3.4 AEI Analysis in Joint Tissues. As both rs3815148 and rs4730250 are intronic SNPs, AEI was measured using two transcript SNPs in high LD with the associated SNPs; the *HBP1* SNP rs7794598 and the *DUS4L* SNP rs711442 (Chapter 2, section 2.9; LD values can be found in Table 3.1, and SNP details in the appendix, Table A6). AEI was measured using SNaPshot (Chapter 2, section 2.10) and for each transcript SNP, individuals who were heterozygous for the SNP and heterozygous for one or both of the associated SNPs were studied. This enabled the assessment of whether the OA associated alleles at rs3815148 and rs4730250 correlated with increased or decreased allelic expression of the two genes. If this were the case, AEI would be observed in a compound heterozygote. Individuals who were heterozygous for the transcript SNP but homozygous for one or both of the associated SNPs were also studied. This allowed for the direct assessment of whether *HBP1* or *DUS4L* were subject to a *cis*-eQTL that was acting independently of the association signal.

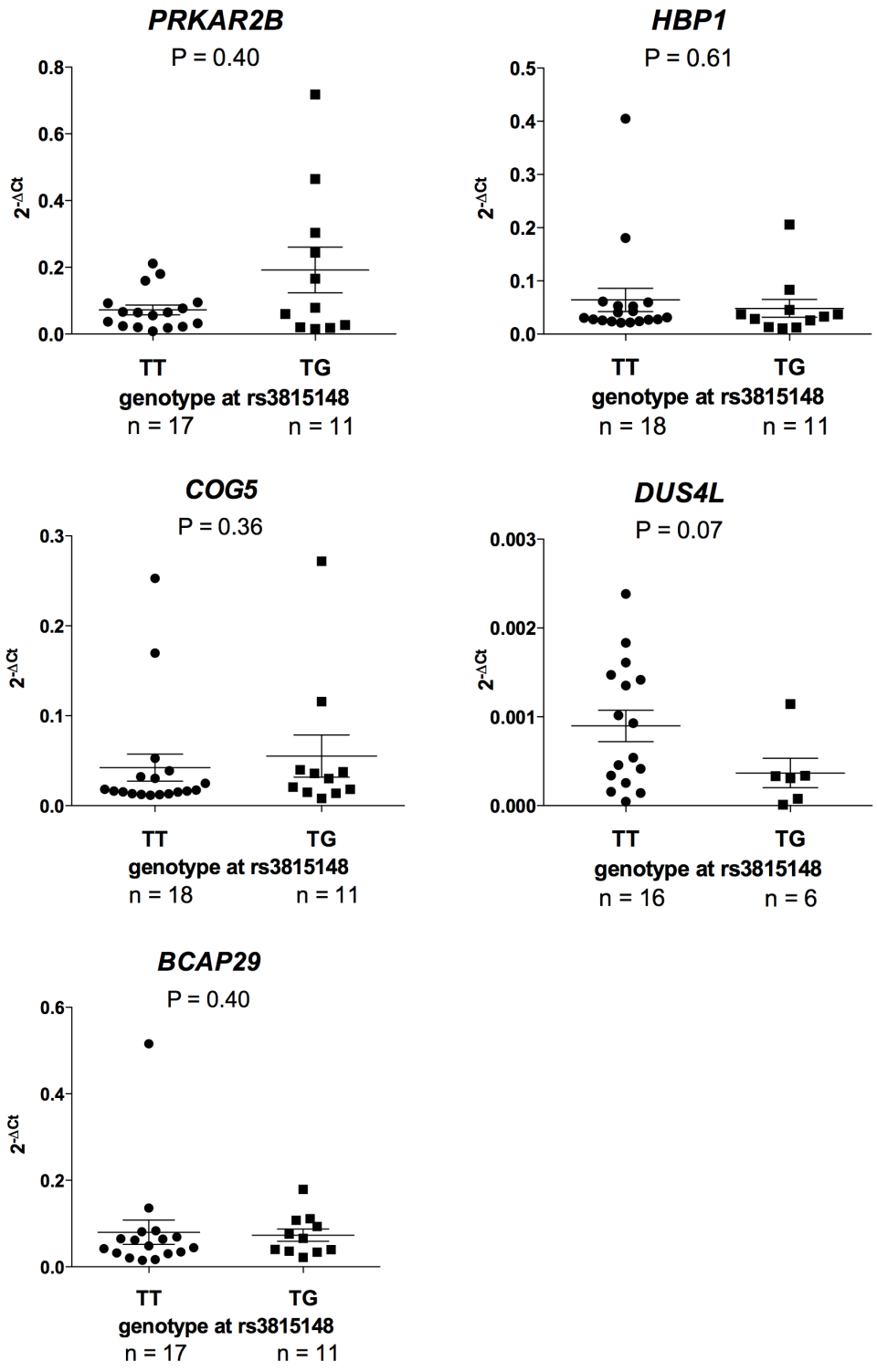

Figure 3.9 Columnar scatter plots of the quantitative gene expression in fat pad cDNA from OA knees, with patients stratified by genotype at the associated SNP rs3815148. n is the number of patients studied. Due to the absence of GG homozygotes at rs3815148, the analysis was between TT homozygotes and G-allele carriers (TG). The horizontal lines in each plot represent the mean and the standard error of the mean. P-values were calculated using a Mann-Whitney U test.

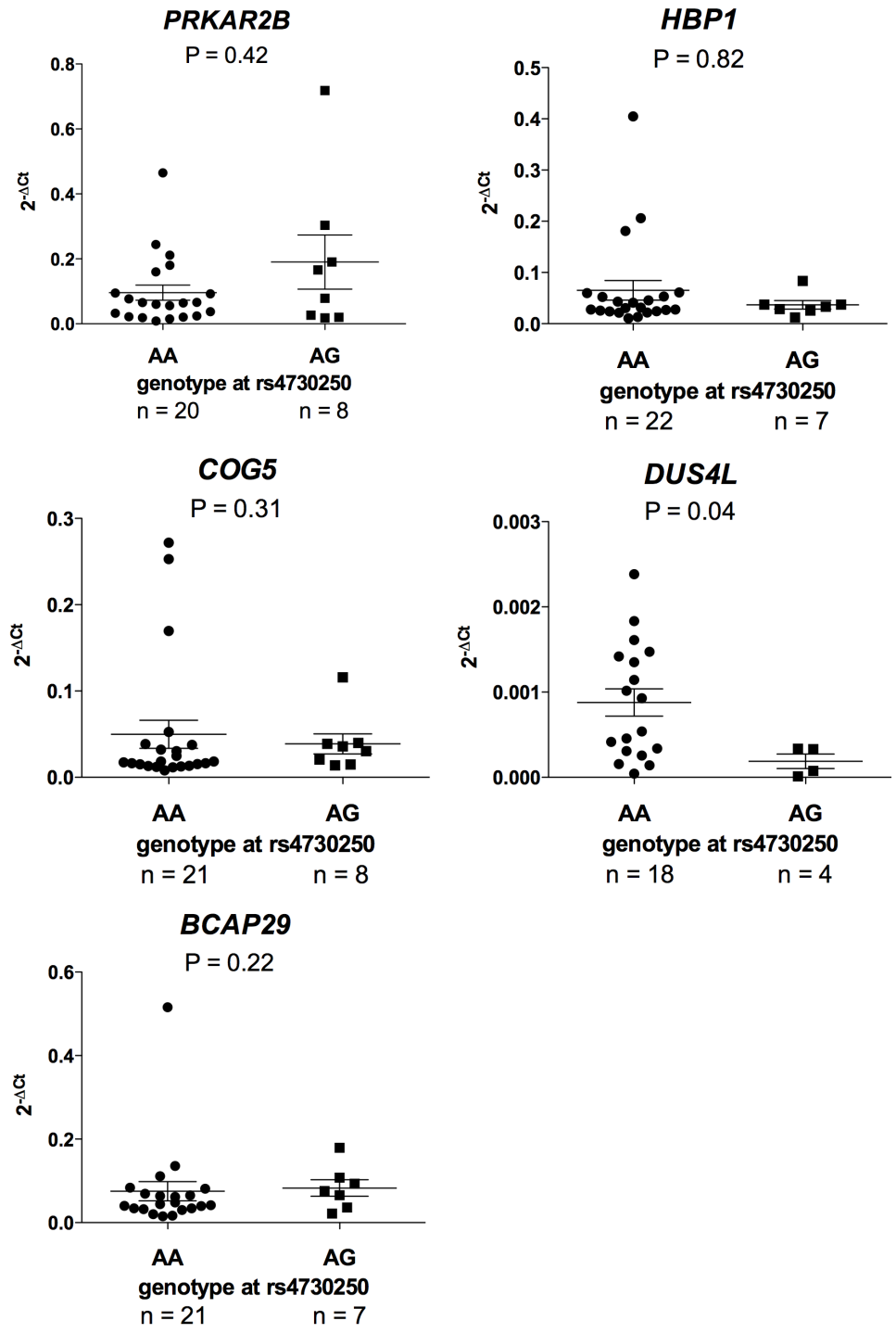

Figure 3.10 Columnar scatter plots of the quantitative gene expression in fat pad cDNA from OA knees, with patients stratified by genotype at the associated SNP rs4730250. n is the number of patients studied. Due to the absence of GG homozygotes at rs4730250, the analysis was between AA homozygotes and G-allele carriers (AG). The horizontal lines in each plot represent the mean and the standard error of the mean. P-values were calculated using a Mann-Whitney U test.

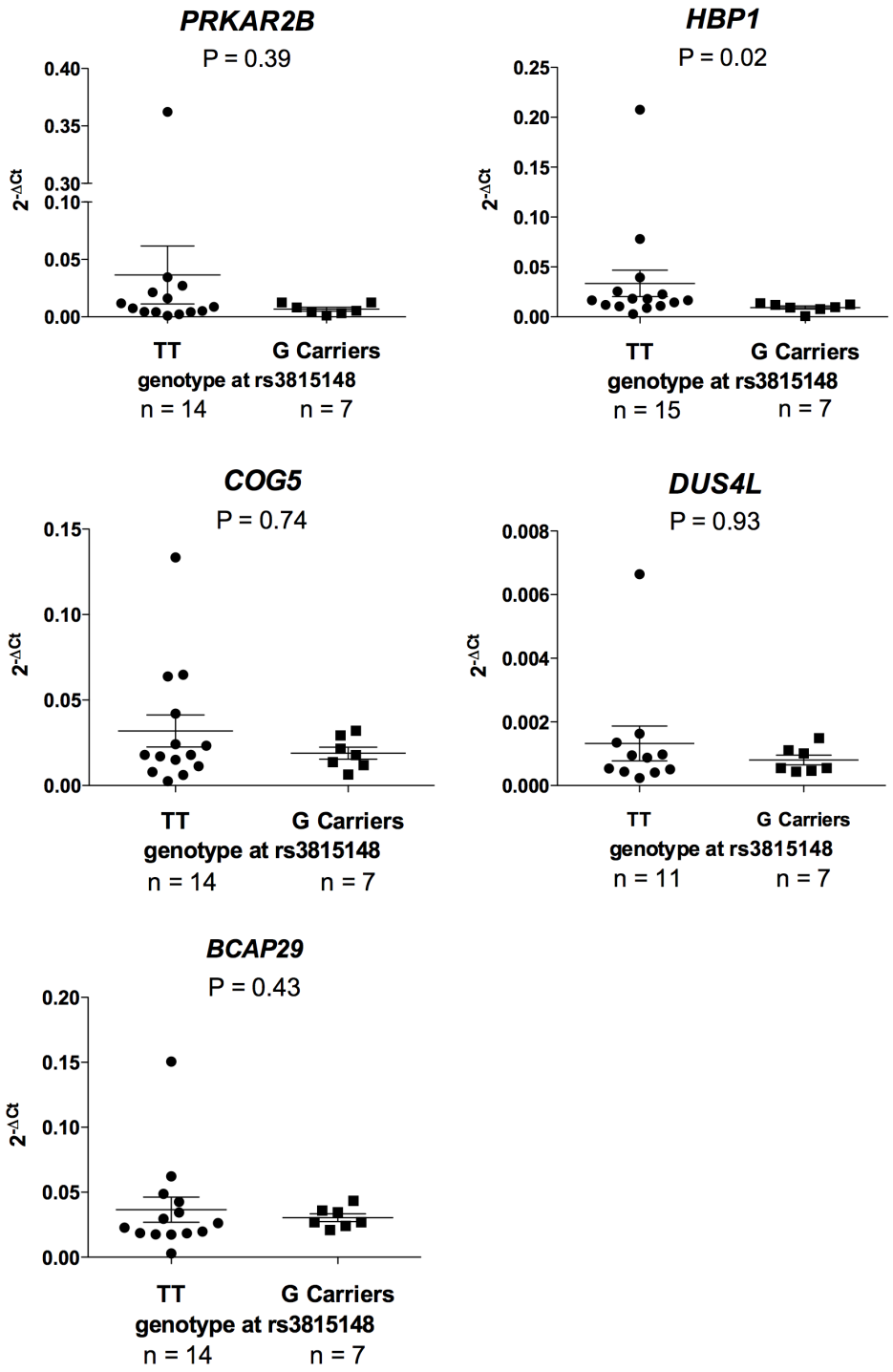

Figure 3.11. Columnar scatter plots of the quantitative gene expression in synovium cDNA from OA knees, with patients stratified by genotype at the associated SNP rs3815148. n is the number of patients studied. Due to the low frequency of GG homozygotes at rs3815148 the analysis was between TT homozygotes and G-allele carriers (TG and GG). The horizontal lines in each plot represent the mean and the standard error of the mean. P-values were calculated using a Mann-Whitney U test.

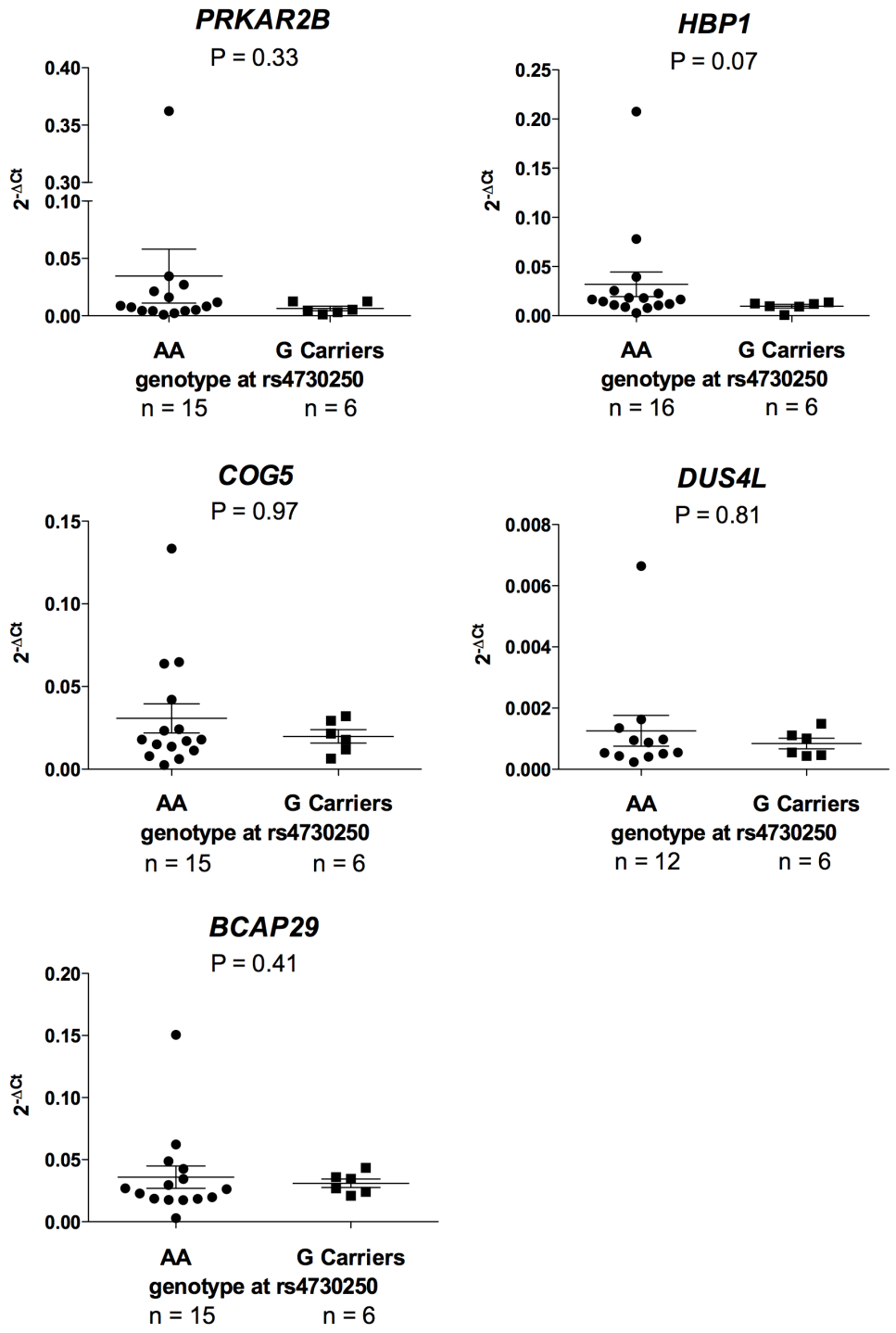

Figure 3.12. Columnar scatter plots of the quantitative gene expression in synovium cDNA from OA knees, with patients stratified by genotype at the associated SNP rs4730250. n is the number of patients studied. Due to the low frequency of GG homozygotes at rs4730250 the analysis was between AA homozygotes and G-allele carriers (AG and GG). The horizontal lines in each plot represent the mean and the standard error of the mean. P-values were calculated using a Mann-Whitney U test.

Table 3.1 The two transcript SNPs used for AEI analysis and their pair-wise D′ and r^2 values relative to the two associated SNPs rs3815148 and rs4730250.

Gene	Transcript SNP	Alleles	MAF[1]	Pairwise linkage disequilibrium relative to rs3815148		relative to rs4730250	
				D′	r^2	D′	r^2
HBP1	rs7794598	C/T	0.219	1	1	1	0.82
DUS4L	rs711442	C/T	0.469	0.59	0.10	1	0.21

[1]Minor allele frequency in Europeans (HapMap CEU)

HBP1

AEI analysis was carried out on a number of different joint tissues taken from twenty patients (Table A1). Twelve of the twenty were heterozygous for both rs7794598 and rs4730250, whilst all twenty were heterozygous for rs7794598 and rs3815148. This is a reflection of the fact that rs7794598 and rs4730250 are in complete linkage disequilibrium (D′ = 1.0, r^2 = 0.82; Table 3.1), whilst rs7794598 and rs3815148 are in perfect linkage disequilibrium (D′ = 1.0, r^2 = 1.0). The G allele of rs4730250 will reside on a haplotype carrying the T allele of rs7794598, as will the G allele of rs3815148. Seventeen of the twenty patients demonstrated a significant AEI ($P < 0.05$; Figure 3.13 and Table 3.2), and these significant imbalances observed operated in both directions, that is, in eight patients (26, 45, 52, 55, 62, 128, 148 and 189) the C-allele at rs7794598 produced more transcript than the T-allele whilst the opposite was the case for nine patients (26, 28, 33, 82, 86, 92, 134, 138 and 142).

The allelic ratios for all 20 patients were plotted on a scatter plot and a Mann-Whitney U test was performed to identify any possible correlation between the associated SNP genotypes and the level or direction of AEI as measured with rs7794598 (Figure 3.14). Since rs7794598 and rs3815148 are in perfect LD all patients were compound heterozygous, thus disallowing the ability to study genotype correlations for rs3815148. rs7794598 and rs4730250 are not in perfect LD and a correlation analysis could therefore be performed for this SNP, revealing a significant difference in the allelic ratios between individuals homozygous AA and heterozygous AG at rs4730250 ($P = 0.008$; Figure 3.14), highlighting that the *HBP1* eQTL correlates with genotype at this associated SNP.

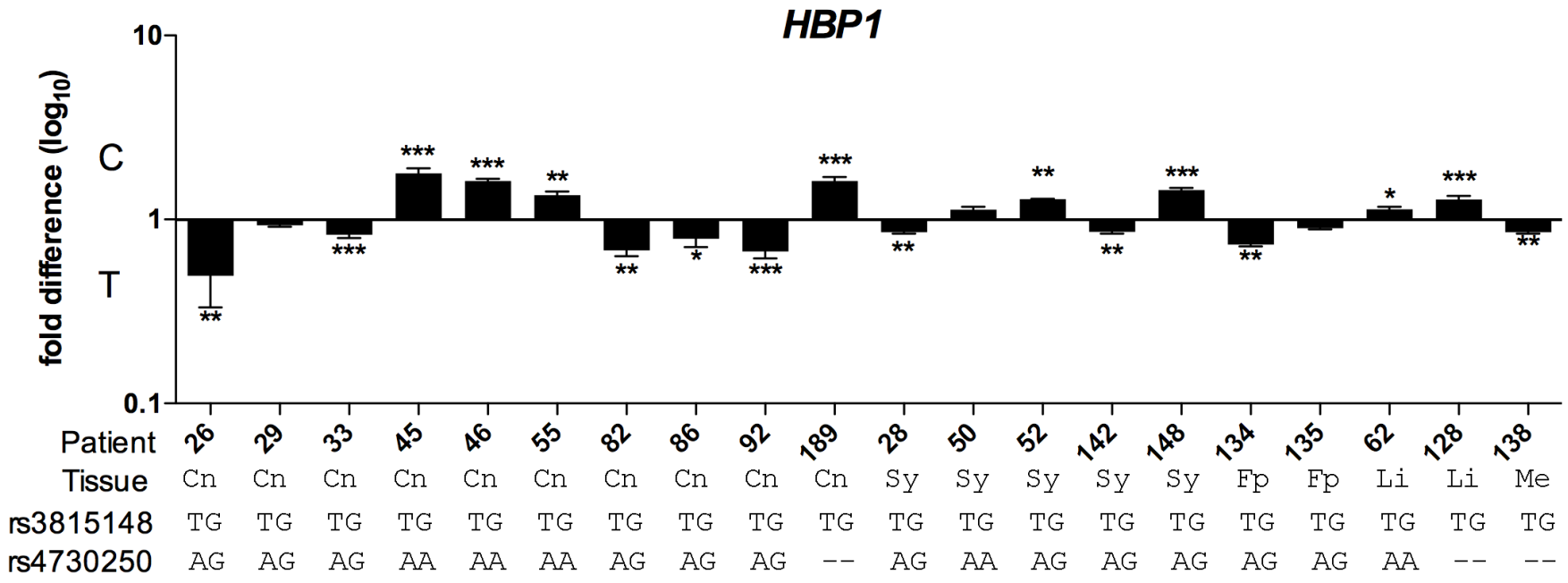

Figure 3.13 Allelic expression analysis of the transcript SNP rs7794598 in *HBP1* using cDNA from joint tissues: Cn, cartilage; Sy, synovium; Fp, fat pad; Li, ligament; Me, meniscus. The genotypes of the patients for the two associated SNPs rs3815148 and rs4730250 are shown (the rs4730250 genotypes were undetermined for patients 128, 138 and 189). The cDNA allelic ratio in each sample was compared to the DNA ratio using a Mann-Whitney U test. *** $P < 0.001$, ** $P < 0.01$ and * $P < 0.05$.

Table 3.2 The allelic ratios for the samples analysed for AEI at rs7794598. Allelic ratios are shown ± the standard error of the mean. P-values were calculated using a Mann-Whitney U test. P-values < 0.05 are highlighted in bold.

HBP1		
rs7794598		
Patient	Allelic ratio (C/T)	*P* value
26	**0.53 ± 0.17**	**<0.01**
29	0.93 ± 0.02	0.39
33	**0.83 ± 0.03**	**<0.01**
45	**1.78 ± 0.11**	**<0.01**
46	**1.62 ± 0.05**	**<0.01**
55	**1.36 ±0.06**	**<0.01**
82	**0.68 ± 0.05**	**<0.01**
86	**0.78 ± 0.08**	**0.01**
92	**0.67 ± 0.05**	**<0.01**
189	**1.62 ± 0.08**	**<0.01**
28	**0.85 ± 0.02**	**<0.01**
50	1.13 ± 0.04	0.11
52	**1.29 ± 0.01**	**0.03**
142	**0.86 ± 0.02**	**<0.01**
148	**1.44 ± 0.04**	**< 0.01**
134	**0.73 ± 0.01**	**<0.01**
135	0.90 ± 0.01	0.33
62	**1.13 ± 0.04**	**0.04**
128	**1.28 ± 0.06**	**<0.01**
138	**0.85 ± 0.01**	**<0.01**

The fact that not all rs4730250 heterozygotes demonstrated a significant allelic imbalance in the same direction (patients 26, 28, 33, 82, 86, 92, 142, and 134 show an imbalance in the direction of the C allele of rs7794598 whereas patients 52 and 148 show an imbalance in the opposite direction), suggests that the correlation is not complete. This is supported by the fact that several rs4730250 AA homozygotes also demonstrate significant AEI, which may also be highlighting the existence of more than one eQTL operating on *HBP1*. One of the cartilage samples studied was derived from a NOF patient (patient 189). This patient also demonstrated significant HBP1 AEI ($P < 0.001$), highlighting that the eQTL is not dependent on OA status.

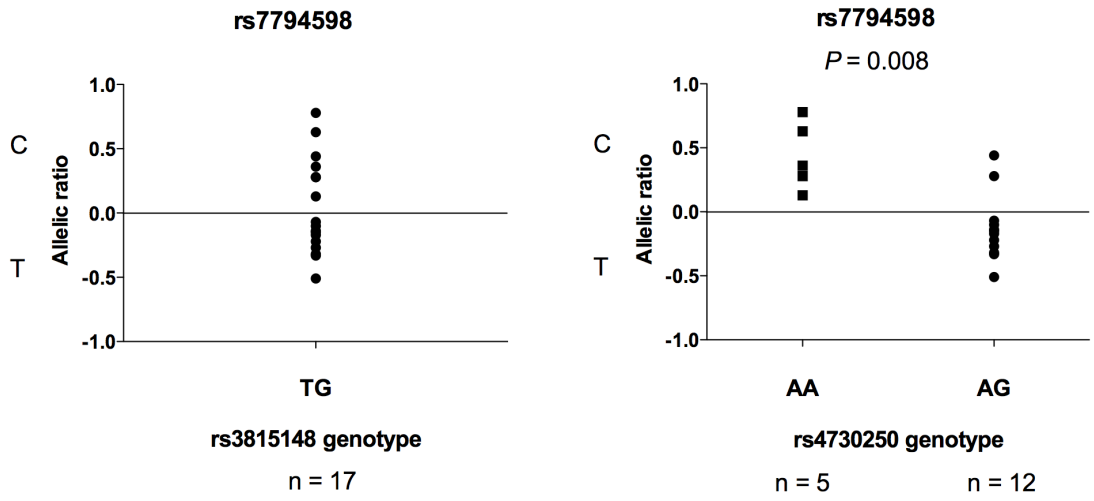

Figure 3.14. Scatter plots of the allelic ratios from AEI analysis of rs7794598 in *HBP1* stratified by genotype at the associated SNPs rs3815148 and rs4730250. TG is the single genotypic group at rs3815148, as the SNPs are in perfect LD, and AA and AG are the genotypic groups at rs4730250. n is the number of patients in each of the genotypic groups. P-values were calculated using a Mann-Whitney U test.

DUS4L

AEI analysis was carried out on a number of different joint tissues taken from eighteen OA patients (Table A1), eight of whom were heterozygous for rs3815148, whilst five were heterozygous for rs4730250. Nine of the eighteen patients demonstrated a significant AEI ($P < 0.05$; Figure 3.15 and Table 3.3). The majority of the imbalances observed operated in the same direction, with the C allele at rs711442 producing more transcript than the T allele.

In the scatter plot analysis there was no significant evidence ($P < 0.05$) for a correlation between the associated SNP genotypes and the level or direction of AEI as measured with rs711442 (Figure 3.16), although rs4730250 approached significance ($P = 0.07$).

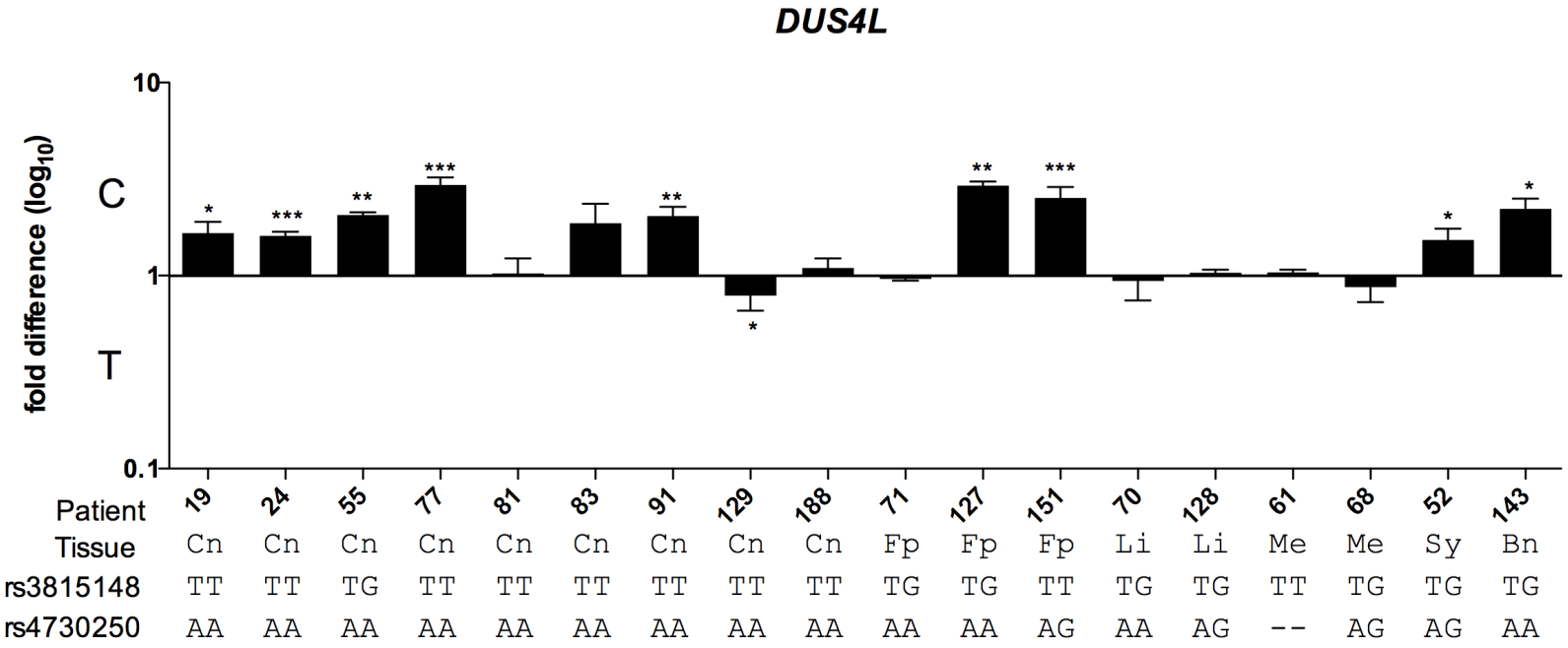

Figure 3.15 Allelic expression analysis of the transcript SNP rs711442 in *DUS4L* using cDNA from joint tissues: Cn, cartilage; Sy, synovium; Fp, fat pad; Li, ligament; Me, meniscus; Bn, cancellous bone. The genotypes of the patients for the two associated SNPs rs3815148 and rs4730250 are shown (the rs4730250 genotype was undetermined for patient 61). The cDNA allelic ratio in each sample was compared to the DNA ratio using a Mann-Whitney U test. *** $P < 0.001$, ** $P < 0.01$ and * $P < 0.05$.

Table 3.3 The allelic ratios for the samples analysed for AEI at rs711442. Allelic ratios are shown ± the standard error of the mean. P-values were calculated using a Mann-Whitney U test. P-values < 0.05 are highlighted in bold.

DUS4L		
rs711442		
Patient	allelic ratio (G/A)	*P* value
19	**1.66 ± 0.23**	**0.01**
24	**1.61 ± 0.08**	**<0.01**
55	**2.02 ± 0.11**	**<0.01**
77	**2.95 ± 0.27**	**<0.01**
81	1.03 ± 0.20	0.89
83	1.87 ± 0.49	0.05
91	**2.03 ± 0.24**	**<0.01**
129	**0.79 ± 0.13**	**0.04**
188	1.09 ± 0.14	0.89
71	0.97 ± 0.03	1.00
127	**2.93 ± 0.14**	**<0.01**
151	**2.53 ± 0.34**	**<0.01**
70	0.94 ± 0.19	0.94
128	1.03 ± 0.09	0.53
61	1.04 ± 0.03	0.25
68	0.87 ± 0.15	0.60
52	**1.53 ± 0.23**	**0.03**
143	**2.22 ± 0.29**	**0.02**

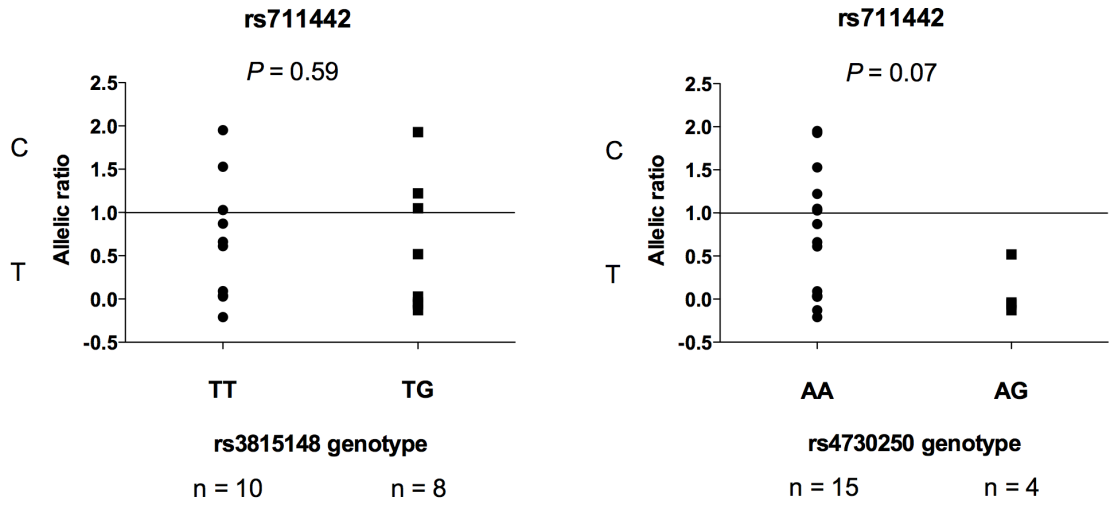

Figure 3.16 Scatter plots of the allelic ratios from AEI analysis of rs711442 in *DUS4L,* stratified by genotype at the associated SNPs rs3815148 and rs4730250. TT and TG are the genotypic groups at rs3815148 and AA and AG are the genotypic groups at rs4730250. n is the number of patients in each of the genotypic groups. P-values were calculated using a Mann-Whitney U test.

3.4 Discussion

This expression analysis of the six genes that map to the chromosome 7q22 OA association signal has revealed that: 1) *PRKAR2B, HBP1, COG5, DUS4L* and *BCAP29* are expressed in joint tissues and that all five are expressed at a significantly lower level in OA cartilage compared to control cartilage, 2) carriage of the OA associated alleles at rs3815148 and rs4730250 correlates significantly with reduced *HBP1* expression in cartilage 3) this correlation is particularly pronounced in knee cartilage, the joint most relevant to the association signal 4) reduced *HBP1* expression in OA synovium correlates with carriage of the OA associated allele at rs3815148 5) reduced *DUS4L* expression in OA fat pad correlates with carriage of the OA associated allele at rs4730250 and 6) all the correlations seen can be directly observed at an allelic level. These results therefore indicate that *HBP1* should be prioritised from amongst the six as being the target of the OA susceptibility that maps to chromosome 7q22.

As previously mentioned, *HBP1* codes for HMG-box transcription factor 1, which has been implicated in Wnt signaling suppression, premature cellular senescence, and superoxide production, but most research to date has focused on the transcriptions factor's activity in a variety of human cancers (181-183). There have been no functional studies yet reported on the role of the transcription factor in arthritis. However the Wnt pathway, cellular senescence and oxidative stress are known to be contributors to OA etiology (184-186).

Wnt pathway activation is initiated by Wnt protein binding to frizzled-receptors and low-density lipoprotein receptor 5/6 co-receptors. In the canonical Wnt pathway this binding inhibits the phosphorylation of β-catenin. This stabilizes the β-catenin and allows it to accumulate and translocate to the nucleus to interact with LEF/TCF transcription factors, which in turn activate Wnt target genes (187). Numerous Wnt factors have been shown to be upregulated in OA: Wnts 2b, 4, 6, 7a, 7b, 9b, 10a, 16 and particularly 3a, which, when used to stimulate chondrocytes, upregulated MMP3 and a disintegrin and metalloproteinase with thrombospondin motifs-4 (ADAMTS4; 188). Furthermore an experiment activating the Wnt/β-catenin pathway in chick embryos triggered expression of MMP13 leading to degradation of the proteoglycan matrix with subsequent loss of tissue structure and function (189) indicating that induction of the pathway reprogrammes chondrocytes towards a catabolic phenotype (184, 187). Therefore a low level of β-catenin is necessary for the

maintenance of the differentiated chondrocyte phenotype (190) preventing this stimulation of chondrocyte catabolic action (187). Since *HBP1* has been shown to be an efficient repressor of Wnt/β-catenin signaling (158), altered *HBP1* expression could therefore have a role in OA aetiology.

Price *et al.* showed that osteoarthritic chondrocytes expressed the marker of senescence, β-galactosidase, whereas non-osteoarthritic chondrocytes from the same donor did not express markers of cellular senescence (191). *HBP1* is essential for initiation of premature cellular senescence as part of both the retinoblastoma and MAPK pathways (159), so again, it can be concluded that *HBP1*-mediated cellular senescence may contribute to OA aetiology.

Since adult articular cartilage is avascular, the mature chondrocytes must survive in a relatively hypoxic environment (192). Chondrocytes express NADPH oxidase, nitrogen oxide synthase (NOS) family members and oxygenases that generate reactive oxygen species (ROS), NO and the superoxide anion. ROS cause cartilage damage by cleaving collagen and aggrecan or by modulating the redox-sensitive pathways that control MMP expression (186). As discussed previously, *HBP1* represses the p47phox subunit of NADPH oxidase (164), which is essential for the regulation of ROS and is therefore another possible mechanism by which altered expression of *HBP1* could contribute to OA.

Since the *HBP1* expression data highlights that reduced expression of the gene correlates with OA susceptibility, then activation of the Wnt pathway, attenuation of senescence and repression of NADPH oxidase could all be risk factors for OA.

The prioritisation of *HBP1* means that this gene, and its encoded protein, merits particularly detailed investigation in the context of OA and joint tissues. The results of the AEI, showing an incomplete correlation between genotype and AEI suggests that there may be multiple *cis*-acting regulatory polymorphisms interacting to modulate the expression of *HBP1*. Other examples of this have been reported. For instance, in their investigation of the gene *PYDN* Babbitt *et al.* (193) demonstrated that the expression of *PYDN* was under the regulation of at least six *cis*-acting variants. These results also emphasise that it is important not to limit a functional investigation of OA genetic susceptibility solely to cartilage; *HBP1* demonstrated significant expression effects in a variety of non-cartilaginous joint tissues,

including the synovium, which plays a key role in the disease (194). Furthermore, this investigation highlighted a second gene, *DUS4L*, as harbouring an eQTL that correlated with the 7q22 association signal at rs4730250 in fat pad, albeit not as compelling as that seen for *HBP1*. This signal would not have been detected had cartilage been the sole tissue investigated. IPFP has been hypothesised to have a role in the development and progression of OA (57) as it can secrete inflammatory mediators, such as TNFα and IL-6, which exacerbate the OA disease process (56).

In addition to this, one of the patients studied was a control sample taken from a patient undergoing surgery for NOF fracture. This patient also demonstrated significant AEI and so the *HBP1* eQTL is therefore acting in the manner of a DNA change that leads to an amino acid substitution; so long as the individual carries the allele, the amino acid will be substituted/the eQTL will mediate allelic imbalance, irrespective of the individuals case or control status. It is the fact that the allele is more prevalent in a case cohort versus a control cohort that matters, namely the standard polygenic model.

Overall, the data reported in this chapter clearly demonstrates the need for a comprehensive molecular analysis of potential *cis*-acting regulatory elements of *HBP1* in order to identify the actual functional variant(s) and to elucidate their haplotypic relationship with rs7794598 and with the OA-associated SNPs rs4730250 and rs3815148.

Chapter 4: Identification and Analysis of a *SMAD3 cis*-acting eQTL Operating in Primary Osteoarthritis

4.1 Introduction

Mothers against decapentaplegic homologue 3 (SMAD3) is member of the TGF-β superfamily (195). The TGF-β signalling pathway plays a critical role in joint development and joint maintenance; TGFβ initiates chondrogenesis through stimulation of mitogen-activated protein (MAP) kinase and wnt pathways (196), the proliferation of the chondroblasts and the subsequent synthesis of the components of the ECM such as aggrecan and type II collagen. During the formation of articular cartilage, the terminal differentiation of the chondrocyte is inhibited by TGF-β and the prehypertrophic phenotype is stabilised by TGF-β, which results in permanent cartilage residing at the end of the long bones. However during OA the chondrocytes undergo phenotypic changes that resemble terminally differentiating chondrocytes (196). SMAD2 and 3 signalling is essential for the inhibitory effect of TGF-β on the terminal differentiation of chondrocytes; without SMAD2/3 signalling chondrocytes leave their quiescent state and undergo premature terminal differentiation (196,197). SMAD2 and 3 are therefore essential for cartilage maintenance (198).

TGF-β is secreted as an inactive complex, which once activated, binds to its cell surface type II receptor (TGFβRII; 197, 199). TGFβRII then activates and recruits the type I receptor (TGFβRI), which phosphorylates the receptor-activated SMADs (R-SMADs) SMAD2 and SMAD3 (199). The phosphorylated SMADs form complexes with the co-mediator SMAD (Co-SMAD) SMAD4 (SMAD2/4 or SMAD3/4), which translocate to the nucleus where they regulate transcription (199, 200).

The SMAD family has eight members; two TGFβ R-SMADs (SMAD2 and SMAD3), three bone morphogenic protein (BMP) R-SMADs (SMAD1, SMAD5 and SMAD8), one Co-SMAD (SMAD4) and two inhibitor-SMADs (I-SMADs; SMAD6 and SMAD7; 201).

There is accumulating evidence from mouse models with disrupted TGF-β signalling that demonstrate joint abnormalities, of the involvement of the TGF-β pathway in OA aetiology. For example, mice overexpressing a dominant negative form

of TGF-β type II receptor in skeletal tissue develop progressive skeletal degeneration that strongly resembles human OA (202), mice lacking latent TGF-β binding protein 3, a structural component of the ECM that regulated TGF-β availability, show altered chondrocyte differentiation and early-onset OA (203), and overexpression of Smurf2 in transgenic mice, an E3 ubiquitin ligase and negative regulator of TGFβ, causes articular cartilage fibrillation, osteophyte formation and increased expression of type X collagen and MMP13, all of which are observed in OA (204). In addition to these models, *Smad3*-knockout mice have been created by targeted disruption of *Smad3* exon 8. These $Smad3^{ex8/ex8}$ mice are characterised by hypertrophic chondrocytes expressing type X collagen, progressive loss of cartilage and formation of large osteophytes, much like human OA (198).

Based on evidence linking the TGF-β pathway to joint abnormalities, a study was performed to investigate the role of common genetic variation in *SMAD3* to the risk of large joint OA in humans (205). The study was performed on 313 TKR and 214 THR cases and 520 controls, then replicated in 2,537 knee cases, 1,074 hip cases and 1,221 controls. All cases and controls were European. It was reported that the SNP rs12901499, which is located in intron 1 of *SMAD3*, was associated with both hip (OR 1.22, $P < 4x10^{-4}$) and knee OA (OR 1.22, $P < 7.5x10^{-6}$;8). Since rs12901499 is not in LD with a non-synonymous polymorphism, this association may be operating on gene expression rather than having an effect on the protein.

4.2 Aim

The aim of this chapter is to test whether this OA association is marking a *cis*-eQTL that operates on *SMAD3*.

4.3 Results

4.3.1 Expression of SMAD3 in Human Synovial Joint Tissues. *SMAD3* was initially assessed for expression by Real-Time qPCR (Chapter 2, section 2.6) in eight joint tissues collected from three OA patients who had undergone joint replacement surgery and in control cartilage from three non-OA patients who had undergone hip joint replacement due to a NOF fracture (Figure 4.1). *SMAD3* demonstrated expression in all joint tissues tested.

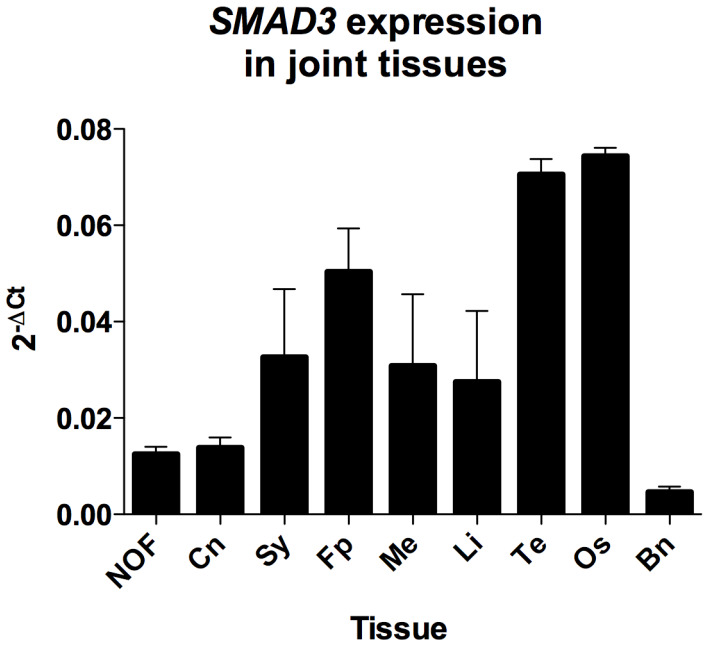

Figure 4.1 Column graph of the levels of expression of *SMAD3* by Real-Time qPCR in tissue from three patients and three controls who had undergone total joint replacement. The tissues used were: neck of femur fracture (NOF), OA cartilage (Cn), synovium (Sy), fat pad (Fp), meniscus (Me), ligament (Li), tendon (Te), osteophyte (Os) and cancellous bone (Bn). The reactions were carried out in triplicate using cDNA from three different individuals. The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The error bars represent the standard error of the mean.

4.3.2 Quantitative Comparison of SMAD3 Expression between Cartilage from OA and NOF Patients. Expression of *SMAD3* was then compared between hip cartilage from 44 patients who had undergone THR for OA and hip cartilage from 16 NOF patients (Figure 4.2), revealing a significantly lower level of *SMAD3* expression in OA hip cartilage (P = 0.005). *SMAD3* expression was then compared between knee cartilage from OA patients who had undergone TKR and hip cartilage from OA patients who had undergone THR (Figure 4.3). This revealed a significantly lower level of *SMAD3* expression in knee cartilage (P = 0.0008). Overall, *SMAD3* showed a lower level of expression in OA versus NOF cartilage, and in knee versus hip cartilage. Having shown that *SMAD3* demonstrates significantly different expression levels between OA and NOF and between OA hip cartilage and OA knee cartilage, the samples were analysed for a correlation between expression levels and genotype at rs12901499.

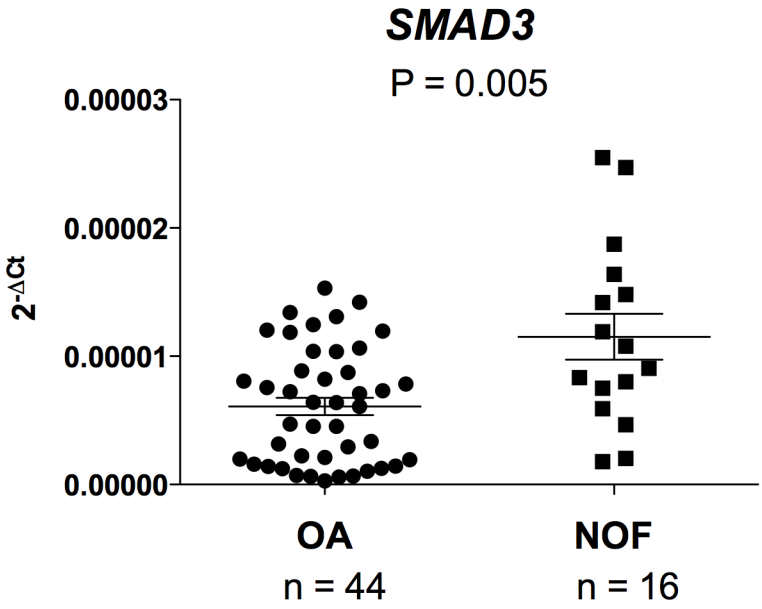

Figure 4.2 Columnar scatter plot of the quantitative expression of *SMAD3* between cDNA from OA hip cartilage and cDNA from control hip cartilage. The control cartilage was collected from patients who had undergone surgery for a neck of femur (NOF) fracture. n is the number of patients studied. The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The horizontal lines in each plot represent the mean and the standard error of the mean. The P-value was calculated using a Mann-Whitney U test.

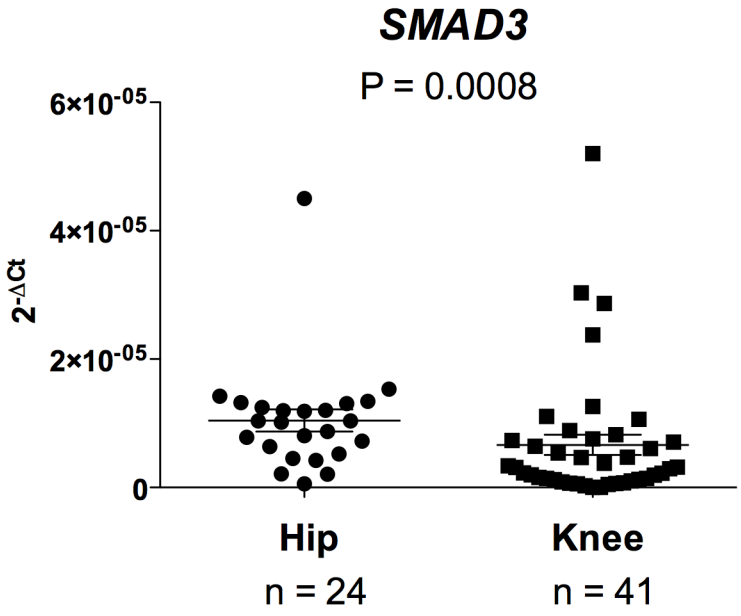

Figure 4.3. Columnar scatter plot of the quantitative expression of *SMAD3* between cDNA from OA hip cartilage and cDNA from OA knee cartilage. n is the number of patients studied. The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The horizontal lines in each plot represent the mean and the standard error of the mean. The P-value was calculated using a Mann-Whitney U test.

4.3.3 Quantitative Expression of SMAD3 in Cartilage, Fat Pad and Synovium Stratified by Genotype at rs12901499. *SMAD3* expression levels were quantified in the OA cartilage samples using Real-Time qPCR (Chapter 2, section 2.6) and then the data were stratified by genotype at rs12902499 (Chapter 2, section 2.8, Figure 4.4).

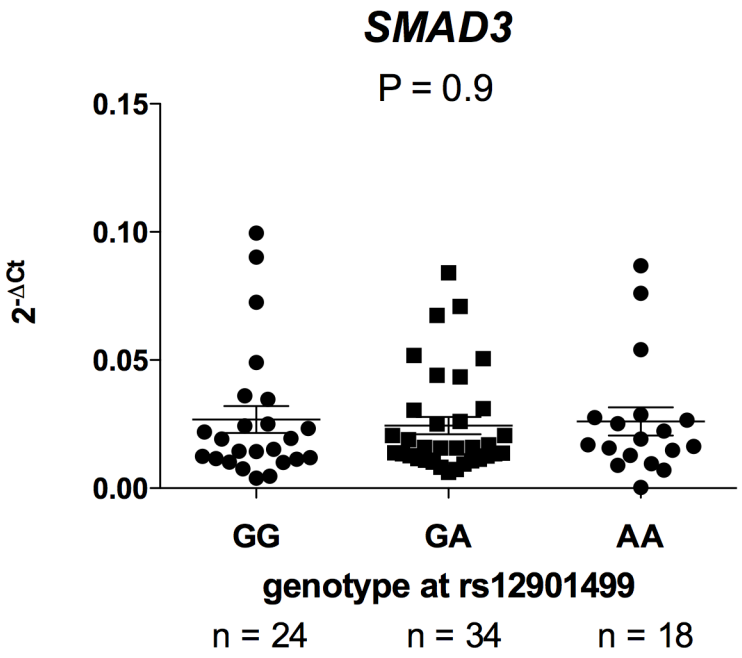

Figure 4.4 Columnar scatter plots of the quantitative gene expression of *SMAD3*, with patients stratified by genotype at the associated SNP rs12901499. n is the number of patients studied. The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The horizontal lines in each plot represent the mean and the standard error of the mean. The P-value was calculated using a Kruskal-Wallis test.

There was no correlation between genotype and expression (*P* = 0.9). The analysis was subsequently repeated after separating the knee and hip cases (Figure 4.5). Again there was no correlation in either the knee cartilage (P = 0.44) or the hip cartilage (P = 0.31). As part of this study, fat pad and synovial tissue were also collected from several of the OA knee patients, providing the opportunity to assess whether *SMAD3* expression in these two tissues correlated with genotype at rs12901499 (Figure 4.6). Once again there was no correlation between genotype and *SMAD3* expression in either fat pad (P = 0.67) or synovium (P = 0.23).

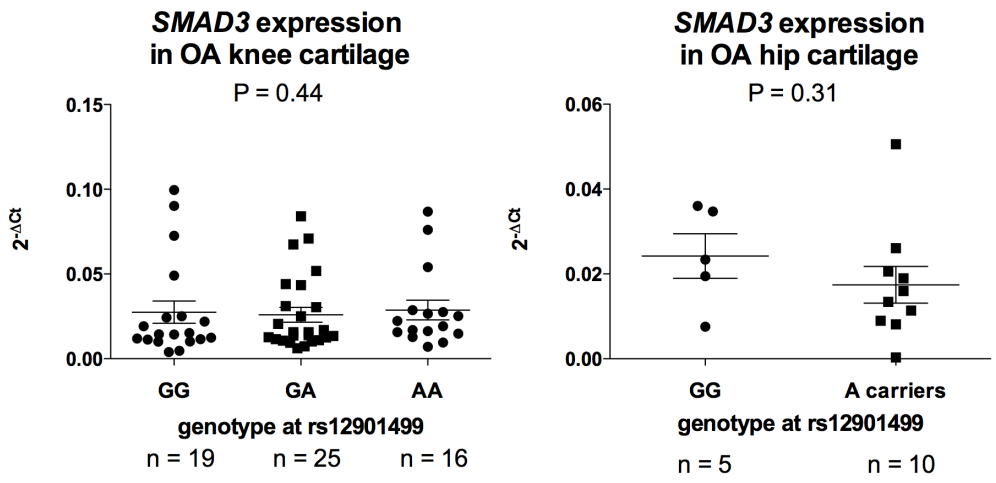

Figure 4.5 Columnar scatter plots of the quantitative gene expression of *SMAD3* in cDNA from OA hip and OA knee cartilage, with patients stratified by genotype at the associated SNP rs12901499. n is the number of patients studied. Due to the low frequency of AA homozygotes at rs12901499 in the hip cohort, the analysis was between GG homozygotes and A-allele carriers (AG and AA). The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The horizontal lines in each plot represent the mean and the standard error of the mean. The P-value was calculated using a Kruskal-Wallis test for the knees and a Mann-Whitney U test for the hips.

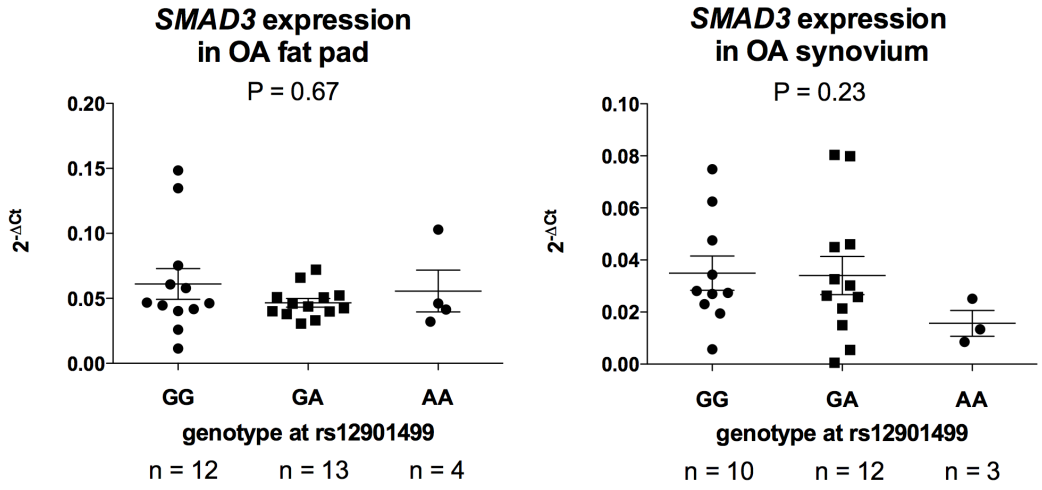

Figure 4.6 Columnar scatter plots of the quantitative expression of *SMAD3* in fat pad and synovium from OA knees. Patients were stratified by genotype at the associated SNP rs12901499. n is the number of patients studied. The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The horizontal lines in each plot represent the mean and the standard error of the mean. P-values were calculated using a Kruskal-Wallis test.

The above analysis did not provide evidence for a correlation between *SMAD3* expression and genotype at rs12901499 in any of the tissues studied. However this technique is vulnerable to the natural fluctuation in gene expression, which can decrease its sensitivity and accuracy (206). An alternative method for correlating genotype with expression levels that is not vulnerable to such fluctuation is AEI.

4.3.4 AEI Analysis in Joint Tissues. Since rs12901499 is not a transcript SNP it could not itself be used in the AEI analysis. Therefore patients who were heterozygous for rs12901499 and for the common *SMAD3* 3′UTR SNP rs8031440 were studied, enabling the indirect testing of allelic output that may correlate with rs12901499, since if this were occurring, AEI would be observed in a high proportion of these compound heterozygotes. rs8031440 was chosen because it has a reasonably high minor allele frequency (0.26; Table 4.1) ensuring that a large proportion of our patients will be heterozygous and therefore suitable for the AEI analysis. Individuals who were homozygous for rs12901499 but heterozygous for rs8031440 were also studied. By doing this it could be assessed whether *SMAD3* was subject to *cis*-eQTLs that were acting independently of the OA association signal marked by rs12901499.

Table 4.1 The two transcript SNPs used in the AEI analysis, the associated SNP, and their pair-wise D′ and r^2 values.

SNPs	Alleles	MAF[1]	Pairwise linkage disequilibrium relative to					
			rs12901499		rs8031440		rs1052488	
			r^2	D′	r^2	D′	r^2	D′
rs12901499	G/A	0.443	-	-	0.01	0.21	< 0.01	0.03
rs8031440	G/A	0.260	0.01	0.21	-	-	0.10	1.00
rs1052488	T/C	0.324	< 0.01	0.03	0.10	1.00	-	-

[1]Minor allele frequency in Europeans (HapMap CEU)

Fourteen OA patients were analysed (patient details can be found in Table A1); eight were patients in whom knee cartilage was studied, two were patients in whom knee fat pad was studied, two were patients in whom knee synovium was studied, and two were patients in whom knee meniscus was studied. AEI was performed as described in Chapter 2, section 2.11. All 8 of the cartilage samples demonstrated significant AEI ($P < 0.05$) and for each patient the AEI was in the same direction, with decreased expression of the G allele of rs8031440 (Figure 4.7 and Table 4.2).

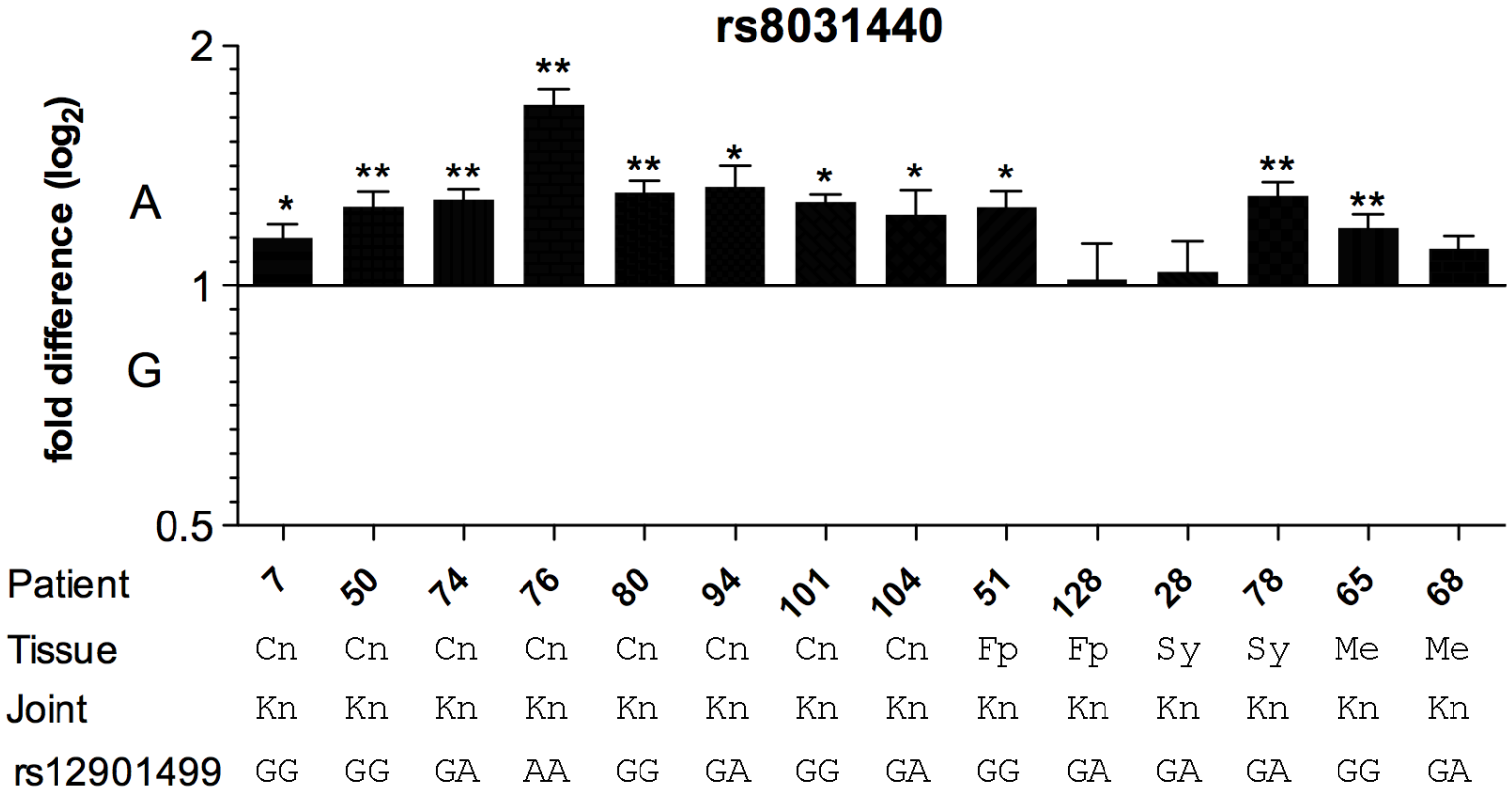

Figure 4.7 Allelic expression analysis of the 3′UTR transcript SNP rs8031440 using cDNA from knee joint tissues. Cn, cartilage; Fp, infrapatellar fat pad; Sy, synovium; Me, meniscus. The genotypes of the patients for the associated SNP rs12901499 are shown. The cDNA allelic ratio in each sample was compared to the DNA ratio using a Mann-Whitney U Test. The error bars represent the standard error of the mean. ** P <0.01 and * P < 0.05.

Table 4.2 The allelic ratios for the samples analysed for AEI at rs8031440. Allelic ratios are shown ± the standard error of the mean. P-values were calculated using a Mann-Whitney U Test. P-values ≤ 0.05 are highlighted in bold.

	rs8031440	
Patient	Allelic ratio (G/A)	*P* value
7	**1.15 ± 0.05**	**0.05**
50	**1.26 ± 0.06**	**<0.01**
74	**1.28 ± 0.04**	**<0.01**
76	**1.69 ± 0.08**	**<0.01**
80	**1.31 ± 0.05**	**<0.01**
94	**1.33 ± 0.09**	**0.02**
101	**1.27 ± 0.03**	**0.02**
104	**1.23 ± 0.09**	**0.05**
51	**1.25 ± 0.06**	**0.03**
128	1.02 ± 0.11	1.00
28	1.04 ± 0.09	0.90
78	**1.29 ± 0.05**	**<0.01**
65	**1.18 ± 0.05**	**<0.01**
68	1.11 ± 0.04	0.19

Significant AEI was also observed in a fat pad, a synovium and a meniscus sample; the observed AEI is not therefore cartilage-specific. Seven of the 14 patients were also heterozygous for rs12901499 whilst 7 were not. The most likely explanation for this AEI result is that the 3′UTR SNP rs8031440 is functional and mediating AEI or that it is in very high LD with the functional SNP. It is unlikely that the OA associated SNP rs12901499 is responsible (either directly or indirectly) for the AEI observed. However, since the rs12901499-rs8031440 compound heterozygotes demonstrated AEI, this possibility could not be excluded based on the rs8031440 AEI data alone. Therefore AEI analysis was performed using a second *SMAD3* 3′UTR SNP, rs1052488, which has an r^2 of 0.1 with rs8031440 (Table 4.1). Eight new patients were investigated (five cartilage and three fat pad; patient details can be found in Table A1), seven of who were compound heterozygous for rs1052488 and rs12901499 but homozygous for rs8031440, and one who was heterozygous for all three SNPs (Figure 4.8). None of the seven patients homozygous for rs8031440 demonstrated significant *SMAD3* AEI (all P-values > 0.05) but the patient that was heterozygous for this SNP did (patient 104, P = 0.03; Table 4.3).

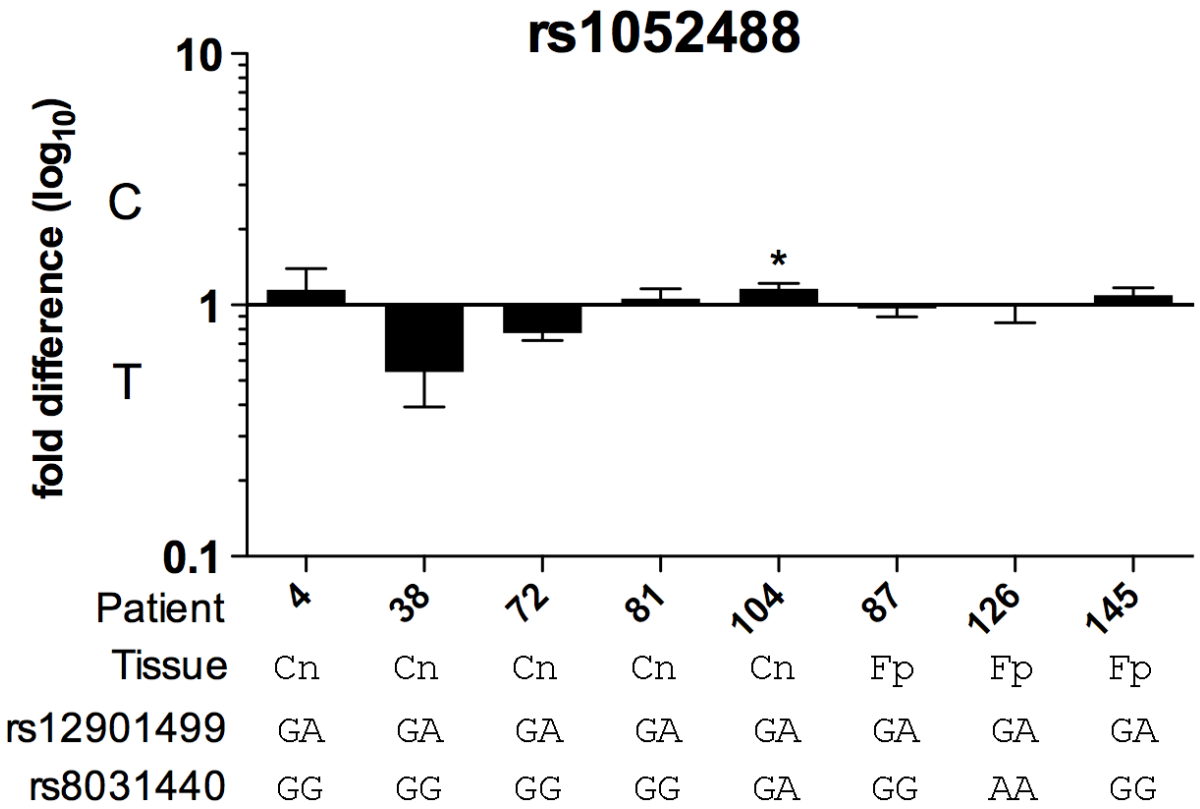

Figure 4.8 Allelic expression analysis of the 3′UTR SNP rs1052488 using cDNA from knee joint tissues. Cn, cartilage; Fp, infrapatellar fat pad. The genotypes of the patients for the transcript SNP rs8031440 and the associated SNP rs12901499 are shown. The cDNA allelic ratio in each sample was compared to the DNA ratio using a Mann-Whitney U Test. The error bars represent the standard error of the mean. * $P < 0.05$.

Table 4.3 The allelic ratios for the samples analysed for AEI at rs1052488**.** Allelic ratios are shown ± the standard error of the mean. P-values were calculated using a Mann-Whitney U Test. P-values ≤ 0.05 are highlighted in bold.

	rs1052488	
Patient	Allelic ratio (C/T)	*P* value
4	1.15 ± 0.25	0.84
38	0.54 ± 0.15	0.06
72	0.77 ± 0.05	0.1
81	1.06 ± 0.10	0.69
104	**1.16 ± 0.06**	**0.03**
87	0.97 ± 0.07	1.0
126	0.99 ± 0.20	1.0
145	1.09 ± 0.08	0.69

It can be concluded therefore that the OA associated SNP rs12901499 does not correlate with *SMAD3* AEI and that the AEI observed is mediated by rs8031440, or by a SNP in high LD with it, rather than by rs12901499, or a SNP in high LD with it. Having established that rs8031440 correlates with AEI, and is therefore highlighting a *SMAD3* eQTL, the *SMAD3* quantitative expression analysis was repeated, this time stratifying the data by genotype at rs8031440.

4.3.5 Quantitative Expression of SMAD3 in Cartilage, Fat Pad and Synovium Stratified by Genotype at rs8031440. A significant correlation between *SMAD3* expression in cartilage and rs8031440 genotype was observed (Figure 4.9), with the G allele of the SNP correlating with reduced *SMAD3* expression ($P = 0.005$). This G allele also correlated with reduced expression in the AEI analysis.

The OA patients were subsequently stratified into knee and hip cases (Figure 4.10) and a significant correlation between expression and genotype was observed for the knee cases ($P = 0.0004$). The P value for the hip cases was not significant ($P = 0.31$), however the number of hip cases was only 16, which may have limited the power to see an effect in this stratum. No significant correlation was observed between *SMAD3* expression and rs8031440 genotype in fat pad or synovium (Figure 4.11). This was surprising, as AEI of *SMAD3* had previously been observed in these two tissues (Figure 4.7, patients 51 and 78).

Figure 4.7 demonstrates a difference in the degree of AEI between tissues, with cartilage samples showing the largest average allelic imbalance (A/G ratio 1.31), followed by synovium and meniscus (average ratios of 1.18 and 1.15, respectively) with fat pad showing the smallest (ratio of 1.13). Therefore it is possible that the smaller number of samples used for analysing overall *SMAD3* expression, together with the smaller allelic imbalances observed for fat pad and synovium, restricted the sensitivity of the quantitative expression assay to detect *SMAD3* expression differences that correlated with rs8031440 genotype in these two tissues.

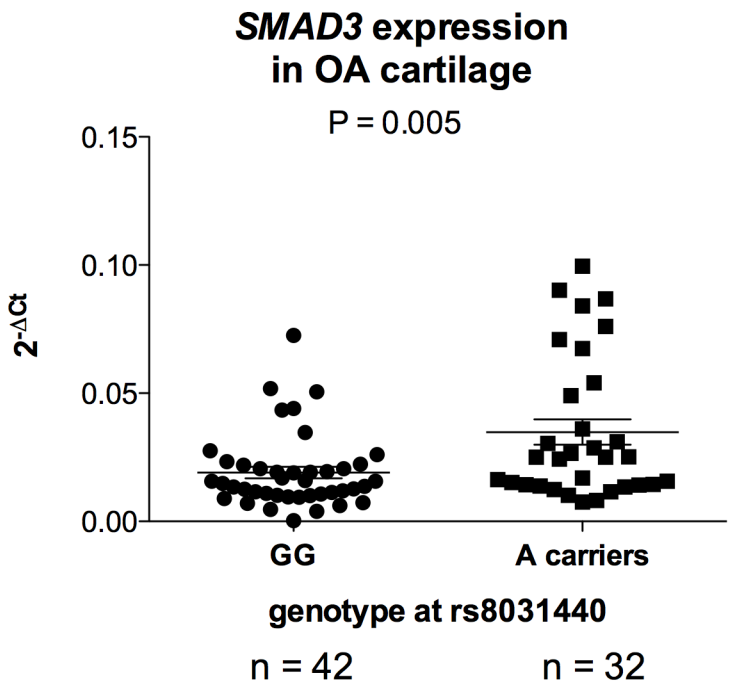

Figure 4.9 Columnar scatter plots of the quantitative gene expression of *SMAD3*, with patients stratified by genotype at the 3′UTR SNP rs8031440. n is the number of patients studied. The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The horizontal lines in each plot represent the mean and the standard error of the mean. The P-value was calculated using a Mann-Whitney U test.

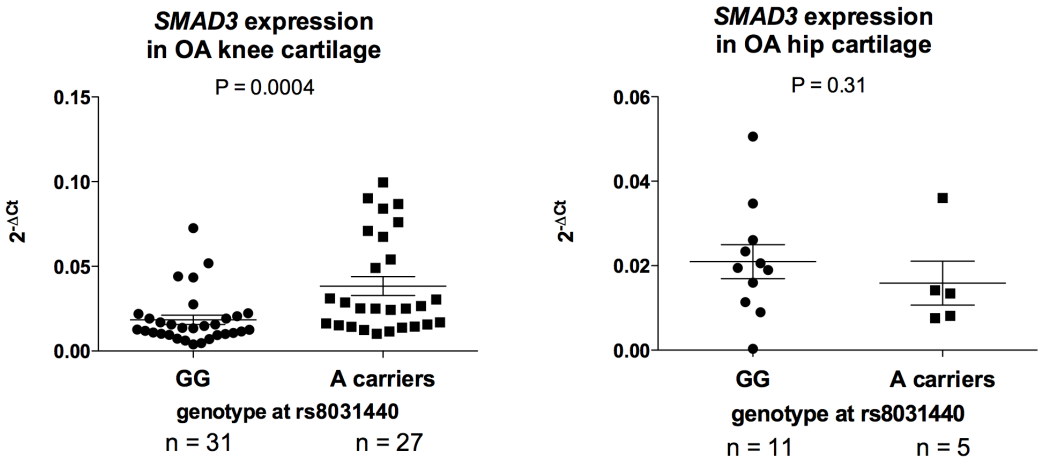

Figure 4.10 Columnar scatter plots of the quantitative gene expression of *SMAD3* in cDNA from OA hip and OA knee cartilage, with patients stratified by genotype at the 3′UTR SNP rs8031440. n is the number of patients studied. Due to the low frequency of AA homozygotes at rs8031440 in the hip cohort, the analysis was between GG homozygotes and A-allele carriers (GA and AA). The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The horizontal lines in each plot represent the mean and the standard error of the mean. The P-value was calculated using a Kruskal-Wallis test for the knees and a Mann-Whitney U test for the hips.

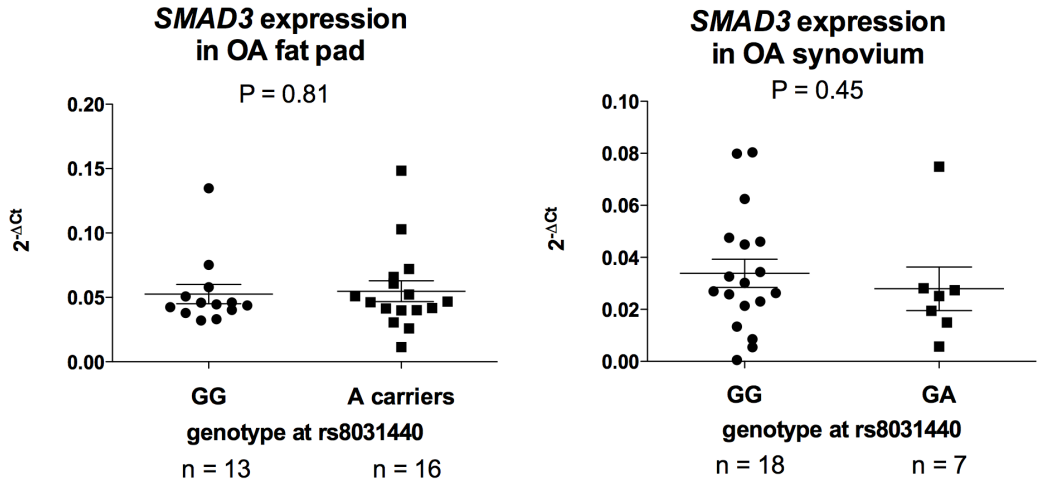

Figure 4.11 Columnar scatter plots of the quantitative expression of *SMAD3* in fat pad and synovium from OA knees. Patients were stratified by genotype at the 3′UTR SNP rs8031440. n is the number of patients studied. Due to the low frequency of AA homozygotes at rs8031440 for fat pad and synovium, the analysis was between GG homozygotes and A-allele carriers (GA and AA). The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The horizontal lines in each plot represent the mean and the standard error of the mean. P-values were calculated using a Mann-Whitney U test.

The above results confirm that rs8031440 or a SNP in high LD with it is modulating the expression of *SMAD3.* A database search (Chapter 2, section 2.9) of the *SMAD3* 3′UTR revealed that rs8031440 is in perfect LD (r^2 =1, D′ = 1) with five other 3′UTR SNPs: rs8025774, rs8031627, rs2278670, rs12595334 and rs3743342 (SNP details are listed in the appendix, Table A6). Subsequently these five SNPs were taken forward for functional analysis to identify which of them was modulating *SMAD3* expression.

4.3.6 Functional Analysis of Common SMAD3 3′UTR Variants. To directly assess whether rs8031440 or any of these five other SNPs were functional, fragments of the 3′UTR were cloned into a luciferase plasmid (Chapter 2, section 2.13.1-2), and reporter gene expression assays were performed using the SW1353 human chondrosarcoma cell line and the MG63 human osteosarcoma cell line (Chapter 2, section 2.13.3-7). Three fragments were amplified: amplicon A, containing rs8025774; amplicon B, containing rs8031440, rs8031627 and rs2278670; and amplicon C, containing rs12595334 and rs3743342 (Figure 4.12). For each amplicon both allelic forms were cloned, such that for amplicon A, the C and T alleles of rs8025774 were cloned, for amplicon B the naturally occurring GGC and AAT haplotypes for rs8031440-rs8031627-rs2278670 were cloned, and for amplicon C the naturally occurring CC and TT haplotypes for rs12595334-rs3743342 were cloned.

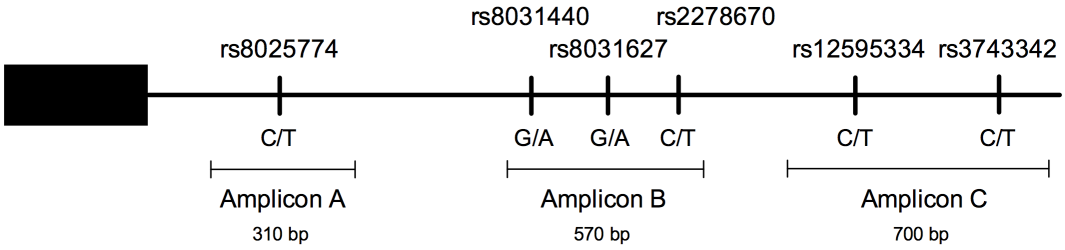

Figure 4.12 Diagram of the *SMAD3* 3′UTR showing the three amplicons used for cloning into a luciferase plasmid. The position of the six SNPs in perfect LD with one another and their alleles are also shown.

No significant differences in luciferase gene expression between the C and T alleles of rs8025774 or between the GGC and AAT haplotypes of rs8031440-rs8031627-rs2278670 were observed in either SW1353 (Figure 4.13) or MG63 (Figure 4.14) cells.

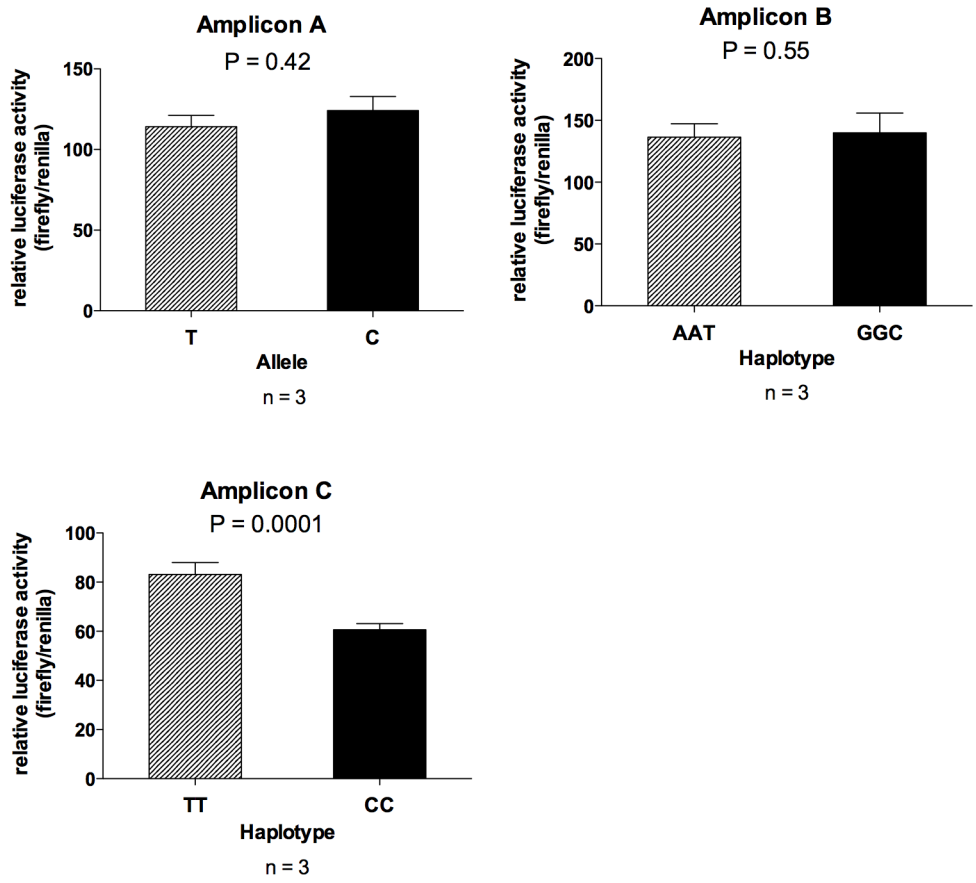

Figure 4.13 Functional analysis of the three amplicons (A B and C) containing the six *SMAD3* 3′UTR SNPs in SW1353 cells. Data are the fold expression relative to Renilla activity and are shown as the standard error of the mean from three biological repeats each with six technical replicates. P-values were calculated using a Mann-Whitney U Test.

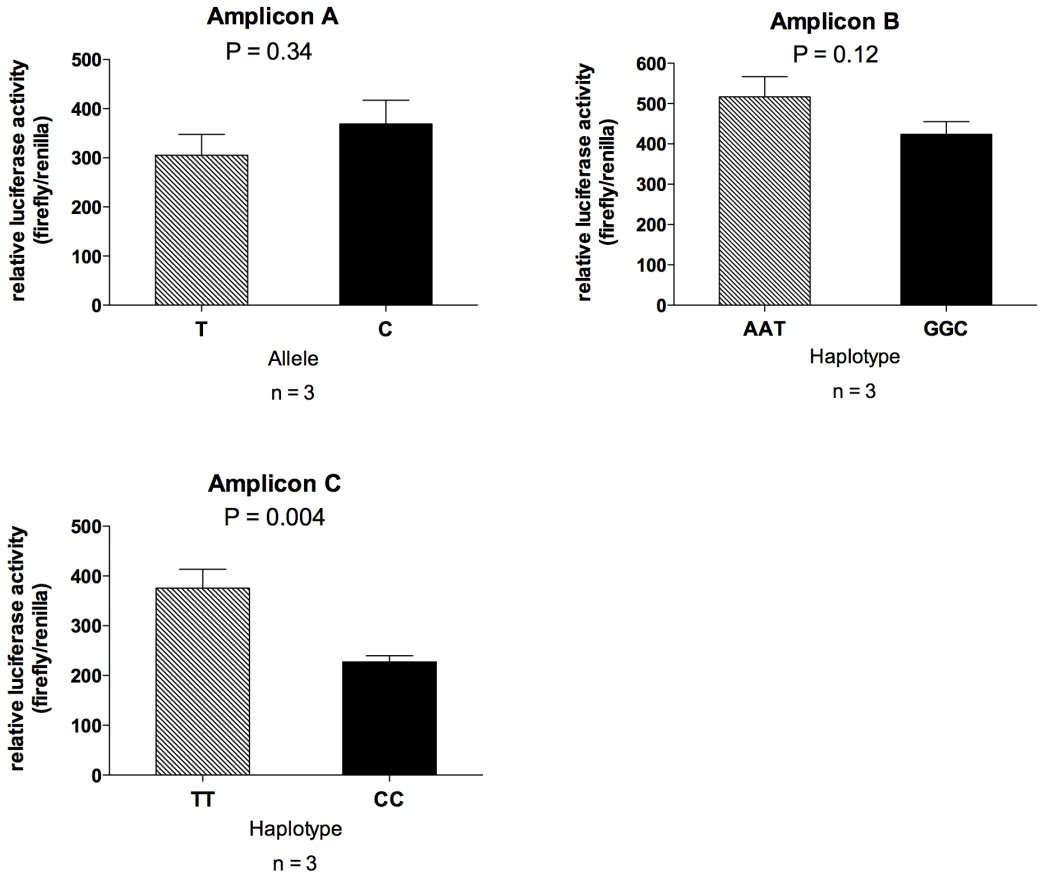

Figure 4.14 Functional analysis of the three amplicons (A, B and C) containing the six *SMAD3* 3′UTR SNPs in MG63 cells. Data are the fold expression relative to Renilla activity and are shown as the standard error of the mean from three biological repeats each with six technical replicates. P-values were calculated using a Mann-Whitney U Test.

However a significant difference between the TT and CC haplotypes of rs12595334-rs3743342 (amplicon C) in both cell lines was observed (P = 0.0001 for SW1353; P = 0.004 for MG63) indicating that one, or both, of these SNPs is a functional variant influencing gene expression. Using site-directed mutagenesis (Chapter 2, section 2.13.8) each allele of these two SNPs was mutated in turn to create the alternative rs12595334-rs3743342 haplotypes CT and TC. The luciferase activity between these and the naturally occurring TT and CC haplotypes was then analysed. In SW1353 cells, mutating rs3743342 from a C to a T had no effect (P = 0.97, Figure 4.15) whereas mutating rs12595334 from a C to a T caused a significant increase in luciferase activity (P = 0.0003, Figure 4.15). In MG63 cells, mutating rs3743342 from a C to a T did have an effect (unlike in SW1353), leading to increased activity (P < 0.0001, Figure 4.15). Mutating rs12595334 from a C to a T also caused a significant increase in luciferase activity in the MG63 cells (P < 0.0001, Figure 4.15), replicating the effect observed in SW1353. Since the natural CC haplotype of rs12595334-rs3743342 is in LD with the

low expressing G allele of rs8031440, this luciferase data supports the patient tissue expression data. Both rs12595334 and rs3743342 are functional, with rs3743342 appearing to have a more cell-restricted effect.

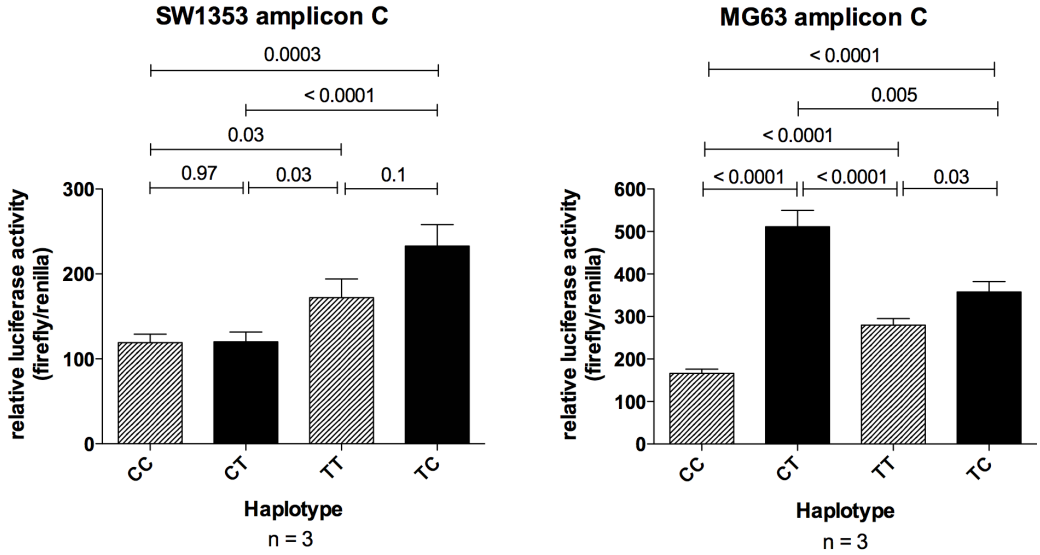

Figure 4.15 Functional analysis of amplicon C containing the four possible haplotypes (CC, CT, TT, and TC) for rs12595334 and rs3743342 in SW1353 and MG63 cells. Data are the fold expression relative to Renilla activity and are shown as the standard error of the mean from three biological repeats each with six technical replicates. P-values were calculated using a Mann-Whitney U Test.

4.3.7 Database Search of Potential miRNA Binding Sites in the SMAD3 3′UTR.

3′UTRs can be the target site for gene expression regulation by microRNAs (207). For example miR-140 has been shown to suppress the TGF-β pathway through targeting *SMAD3* (208). Therefore a database search of potential microRNA binding sites encompassing rs12595334 and rs3743342 (Chapter 2, section 2.14) was performed. However, no miRNAs were predicted to bind at either allelic form of the SNPs.

4.3.8 Genetic Association of rs12529334 and rs3743342 with OA. Finally the eQTL marked by rs8031440, and potentially mediated by rs12595334 and rs3743342, was assessed for association to OA. To do this, the arcOGEN GWAS dataset, comprising 5,804 OA cases that had undergone hip or knee joint replacement and 11,009 population controls, was utilised (135). Neither rs8031440, rs12595334 or rs3743342 are on the Illumina 610 Quad array used by arcOGEN, nor are any of the other *SMAD3* 3′UTRs SNPs that are in perfect LD with them. The SNP rs7166081, which is located downstream of *SMAD3*, is however on the array and is also in perfect LD with

rs8031440, rs12595334 and rs3743342. The association data generated for rs7166081 was therefore used as a direct proxy. This revealed that the *SMAD3* eQTL showed marginal evidence of associated with OA (P = 0.027; Table 4.4). Stratification of the arcOGEN dataset by sex, by site of OA (hip or knee), and by sex combined with site did not particularly enhance the association signal, with the most significant P-value being 0.001 (OR = 0.915) for male knee cases. In all instances the C allele of rs7166081, which is equivalent to the G allele of rs8031440 that correlates with decreased *SMAD3* expression, was less common in OA cases.

Table 4.4 Association analysis of rs7166081 with OA using the arcOGEN data.

Stratum	Cases	Controls	C-allele frequency		P	Odds ratio[1]
			cases	controls	value	
All	5804	11009	0.220	0.235	0.027	0.921 (0.857-0.991)
Hips	3039	11009	0.214	0.233	0.002	0.896 (0.837-0.960)
Knees	2164	11009	0.221	0.233	0.077	0.932 (0.862-1.008)
Females	3363	5515	0.217	0.235	0.030	0.904 (0.826-0.991)
Males	2441	5494	0.211	0.232	0.020	0.883 (0.794-0.980)
Female hip	1769	5515	0.222	0.235	0.183	0.931 (0.838-1.034)
Female knee	1217	5515	0.215	0.232	0.014	0.903 (0.832-0.980)
Male hip	1270	5494	0.220	0.232	0.232	0.931 (0.828-1.047)
Male knee	947	5494	0.218	0.233	0.001	0.915 (0.867-0.965)

[1](95% confidence intervals)

4.4 Discussion

The aim of this chapter was to assess whether the alleles of the OA associated *SMAD3* SNP rs12901499 correlated with differences in expression of the gene, using joint tissues from patients with primary OA, based on the hypothesis that this may be the mechanism through which the rs12901499 OA association is operating. There was no evidence of such a correlation, implying that this hypothesis is incorrect. It is possible however that the association marked by rs12901499 is operating during development and that an investigation of the joint tissues of young individuals may reveal a correlation between rs12901499 and *SMAD3* expression. The collection of a sufficiently large number of such tissues is however prohibitive, since young individuals do not routinely present for joint replacement surgery. In this study a significantly lower level

of *SMAD3* expression was detected in OA versus NOF cartilage. Whether this is a cause or a consequence of the OA, or of the NOF, disease process is not clear. Consistently lower allelic expression was seen for the G allele when AEI was studied using rs8031440 indicating that *SMAD3* was subject to a *cis*-acting eQTL. A significant difference in the degree of this AEI was observed between different joint tissues, implying that there are a range of modulating factors acting on this *SMAD3* eQTL, which may be tissue-specific. The 3′UTR SNPs rs12595334 and rs3743342 were then identified as potential mediators of this eQTL. Using the arcOGEN data it was observed that this eQTL demonstrated nominal association to OA. The allele that correlated with reduced expression of *SMAD3*, the G allele of rs8031440 (indirectly tested using the proxy SNP rs7166081), was less common in the arcOGEN cases versus the arcOGEN controls.

SMAD3 mediates TGF-β signalling, which as discussed in the introduction, has a pivotal role in joint development and maintenance (196, 198). The majority of the evidence to date indicates that decreased SMAD3 expression leads to an OA phenotype; SMAD3 knockout mice develop an OA-like phenotype and loss of SMAD3 enhances bone morphogenic protein signalling in the articular chondrocytes, leading to hypertrophy and OA-like changes (Valdes 2010). SMAD3 can form repressor complexes preventing the action of transcriptional activators that may be important in development or maintenance of articular cartilage (209). For example SMAD3 blocks RUNX2 function (210), which is required for chondrocyte maturation (211). If SMAD3 levels are increased then RUNX2 action will be blocked and chondrocyte maturation will be affected, which could have consequences for the production of the cartilage ECM. It is the balance of SMAD2/3 and SMAD1/5/8 that controls RUNX2 function and terminal differentiation. Dominant SMAD1/5/8 signalling triggers the articular chondrocytes to terminally differentiate leading to a chondrocyte autolytic phenotype characterised by degradation of its surrounding cartilage matrix (197). If SMAD3 levels are increased, the balance shifts away from SMAD1/5/8 signalling and deletion of SMAD1/5/8 transcription factors has been shown to cause defects in the formation of chondrocyte condensations and on chondrocyte maturation (212).

In conclusion, this chapter has demonstrated that *SMAD3* is subject to *cis*-acting regulatory polymorphism in joint tissues, and the 3′UTR SNPs rs12595334 and rs3743342 were identified as potential mediators of this effect.

Chapter 5: A *SMAD3 cis*-acting eQTL Identified in Osteoarthritis is also Operating in Aneurysms and Osteoarthritis Syndrome.

5.1 Introduction

Aortic aneurysm is a common condition and the mortality rate from dissections or ruptures is high (213-215). Aneurysms can be inherited in an autosomal dominant manner as a disorder known as familial thoracic aortic aneurysms and dissections (FTAAD; 216). FTAAD can be subdivided into syndromic and non-syndromic forms. Non-syndromic FTAAD can be caused by mutations in genes encoding the contractile proteins of the VSMCs such as α-actin 2 and myosin-11. Syndromic FTAAD includes systemic tissue connective disorders such as Marfan syndrome, which is caused by mutations in *FBN1,* and Loeys-Dietz syndrome, which is caused by mutations in *TGFBRI* and *TGFBRII*. All syndromic FTAADs are characterised by increased TGF-β signalling in the arterial wall (213). In 2011 a new syndromic form of TAAD was described named aortic aneurysms and osteoarthritis (AOS), characterised by arterial aneurysms and tortuosity (twisting of the blood vessels), mild craniofacial features, skeletal and cutaneous abnormalities and early onset OA (214). Genome-wide linkage analysis was performed using a SNP array, and a signal was identified on chromosome 15q. Fine mapping the 15q area mapped the locus to a 12.8Mb region containing 157 genes. *SMAD3* and *SMAD6* were chosen for sequence analysis from this region due to their roles in the TGF-β signalling pathway (214). No mutations were detected in *SMAD6* but three different mutations were discovered in *SMAD3;* two heterozygous missense mutations leading to amino acid substitutions and a deletion of two nucleotides in exon 6 leading to a frame shift and a premature stop codon in exon 7 (214). Sequencing data of the cDNA from patients carrying the deletion revealed normal *SMAD3* expression for the WT allele but very reduced expression for the mutated allele. When fibroblast cultures from the patients were treated with cycloheximide, which is an inhibitor of nonsense-mediated decay, there was increased expression from the mutant allele. This indicated that the abnormally truncated RNA, which is missing the Mad homology 2 (MH2) domain and the TGFβRI target site for SMAD phosphorylation, was subject to nonsense-mediated mRNA decay and that normal SMAD3 could not be formed by this allele (214). Both of the missense mutations affect highly conserved amino acids within the MH2 domain of SMAD3 and

the most likely effect of these is loss of function, with TGF-β signals not being propagated via SMAD3 (213). Although there is clear phenotypic overlap with syndromes such as Marfan and Loeys-Dietz the distinct difference between these syndromes and AOS is the invariable presence of early onset joint abnormalities including intervertebral disc degeneration and meniscal anomalies, in all cases with *SMAD3* mutations (214).

5.2 Aim

The aim of this chapter is to assess whether the SMAD3 *cis*-eQTL identified in joint tissues in Chapter 4 is also operating in tissues of relevance to AOS, thus potentially providing a pathological link between common OA and the syndrome. To achieve this, cells were studied that were isolated from AOS patients who harbour one of the missense mutations, a heterozygous mutation leading to the substitution of arginine for tryptophan at amino acid position 287 of SMAD3 (Arg287Trp).

5.3 Results

5.3.1 Expression of SMAD3 in Human VSMCs Derived from Aortic Tissue. cDNA from cultured VSMCs (Chapter 2, section 2.15) collected from four patients with AOS and an unrelated control (Figure 5.1; patient details in Table A1), was analysed for *SMAD3* expression by Real-Time qPCR (Chapter 2, section 2.6 - 7). All four patients and the control demonstrated expression of *SMAD3* (Figure 5.2). Patients were then genotyped for rs8031440 (Chapter 2, Section 2.8) and the heterozygotes were taken forward for AEI analysis to investigate the *SMAD3* 3′UTR eQTL's effect in aortic tissue.

5.3.2 AEI Analysis of SMAD3 in AOS. The activity of the *SMAD3* eQTL was analysed using cDNA from the VSMCs (Chapter 2, section 2.11). Two of the patients with AOS (C and D) and the control (E) were heterozygous for rs8031440 and were therefore eligible for AEI analysis using this SNP. The control, individual E (a male born in 1955), and patient C (a male born in 1979), both demonstrated a significant AEI ($P < 0.01$), whereas patient D (a male born in 1947) did not (Figure 5.3).

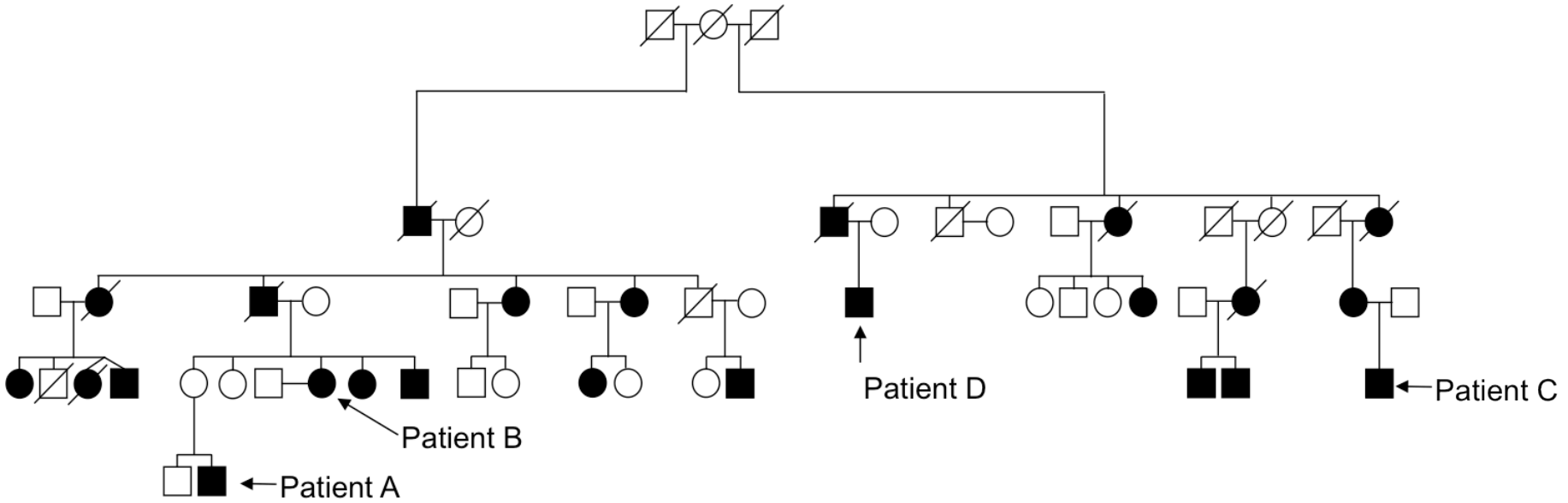

Figure 5.1 Simplified family tree of a family with AOS. Squares indicate males, circles represent females. Black symbols indicate the presence of the signs of AOS. Crossed out symbols represent deceased individuals. The four patients studied in this chapter are highlighted (Patients A-D). Figure courtesy of Aida M. Bertoli-Avella.

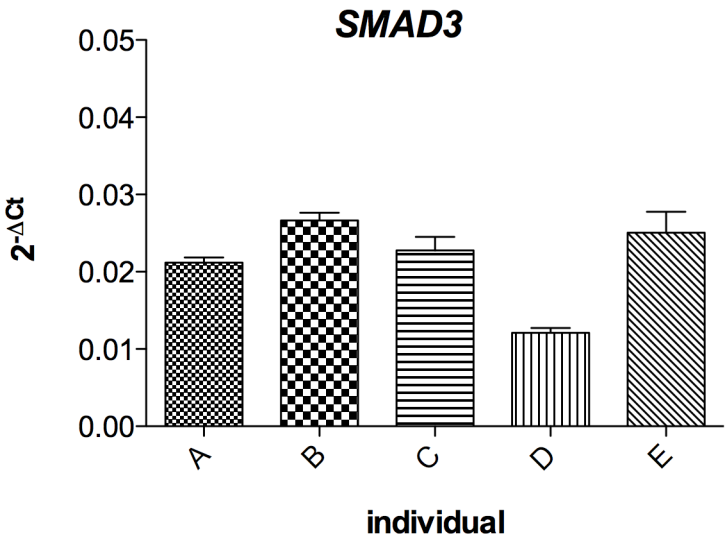

Figure 5.2 Column graph of the level of expression of *SMAD3* by Real-Time qPCR in vascular smooth muscle cells derived from aortic tissue from four patients (A, B, C, D) and an unrelated control (E). Reactions were carried out in triplicate and normalised against the mean Ct of *HPRT1*, *GAPDH* and *18S*. The error bars represent the standard error of the mean.

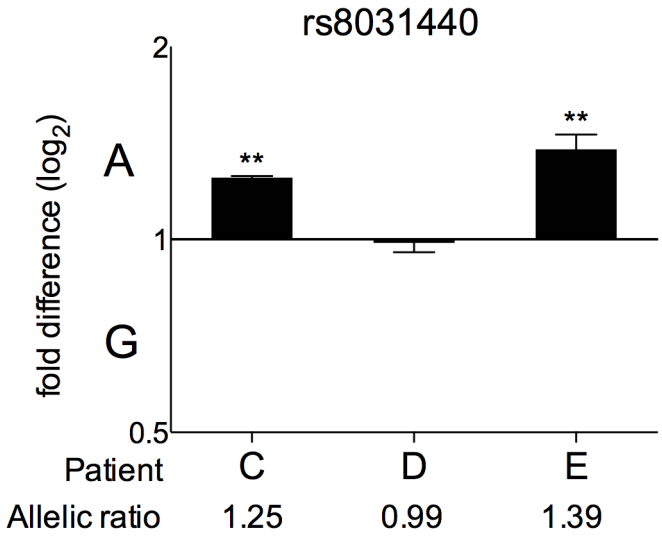

Figure 5.3 AEI analysis of the *SMAD3* 3′UTR SNP rs8031440 using cDNA from vascular smooth muscle cells. The cDNA allelic ratio in each sample was compared to the DNA ratio by a Mann-Whitney U Test. The error bars represent the standard error of the mean. ** P <0.01.

The AEI observed in the control and patient C was in the same direction as that observed in OA joint tissues, namely decreased expression of the G allele of rs8031440 (Chapter 4, Figure 4.7). The allelic ratios of 1.39 (control) and 1.25 (patient C) were also comparable to those seen in the OA patients (Table 4.2). Therefore it can be concluded that the *SMAD3* eQTL is also active in a tissue that is of direct relevance to the aneurysm component of AOS. Next it was assessed which of the alleles, the mutation or the WT, the low expressing allele of the eQTL was present on, by molecular haplotyping of the AEI samples.

5.3.3 Molecular Haplotyping of AOS Patients. Using cDNA as a template the molecular haplotypes for the Arg287Trp mutation and rs8031440 were assessed in patients C and D (Chapter 2, section 2.16). This revealed that in both patients the low-expressing G-allele of rs8031440 was on the WT allele (Figure 5.4).

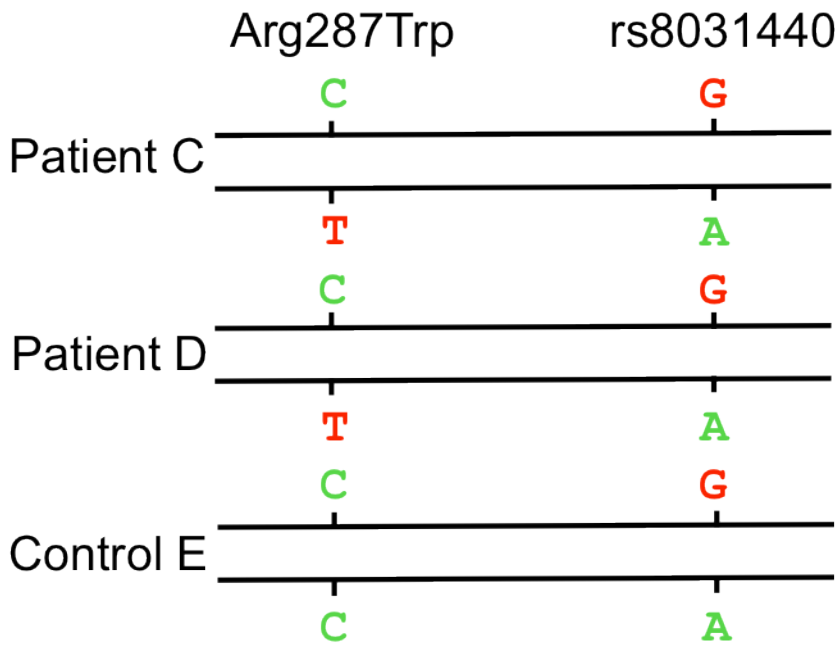

Figure 5.4 The haplotypes of the patients C and D and the control E, at the Arg287Trp mutation and the 3′UTR SNP rs8031440. Wild-type and high expressing alleles are highlighted green and mutant and low expressing alleles are highlighted red.

5.4 Discussion

As noted in the introduction, AOS arises from rare mutations of *SMAD3* (213-215). These mutations are heterozygous and appear to be loss-of-function. One mutation previously described involves a frame-shift mutation that results in nonsense-mediated mRNA decay leading to little or no detectable transcript from the mutant allele, whilst another leads to the substitution of arginine for tryptophan at position 287 of the gene. This Arg287Trp mutation creates a protein that is unable to form SMAD3 protein homomers or heteromers with SMAD4 (6), dysregulating TGF-β signalling. In this chapter it has been demonstrated that the *SMAD3* eQTL that was discovered in OA joint tissues is also operating in the VSMCs derived from the aortic tissue of an AOS patient, who has the Arg287Trp mutation. It is therefore likely to be of direct relevance to this syndrome. There is a high degree of both intra and interfamilial phenotypic variation in patients with AOS. Differences range from age of onset to presence of clinical features that segregate with AOS, such as congenital defects like bicuspid aortic valve, and vascular diseases elsewhere in the body such as intracranial aneurysms (216). This high level of variability implies that it is likely to be influenced by the genetic background on which the WT *SMAD3* allele is operating; if the WT allele happens to be on a low-expressing rather than a high-expressing copy of the gene, as is seen with patient C, then this may lead to an even greater reduction in functional SMAD3 protein and therefore a more severe phenotype. On the other hand, this eQTL could have variable relevance on different tissues, and could explain phenotypic variability determining, for example, a predominantly vascular skeletal phenotype in patients with *SMAD3* mutations. Comparing the inheritance of the eQTL between families in which AOS is segregating may therefore be informative in accounting for some of the phenotypic variability observed.

Patient C is the first cousin once removed of patient D and both have the heterozygous Arg287Trp mutation. The two patients have therefore inherited the G-allele of rs8031440 from a recent common ancestor. The fact that this G-allele shows reduced expression in patient C but not in patient D implies that its expression is likely to be modified by other *cis*-acting effects, which may be genetic or epigenetic.

In conclusion, *SMAD3* is subject to *cis*-acting regulatory polymorphism in the tissues of relevance to both primary osteoarthritis and the aneurysms-osteoarthritis

syndrome. Etiological insights from the common form of osteoarthritis are therefore informing a rare form of the disease and *vice versa.*

Chapter 6: Analysis of *DOT1L*, a Gene Associated with Hip OA.

6.1 Introduction

Disruptor of telomeric silencing 1-like (DOT1L) is a histone H3 methyltransferase (217). Histones are the protein core that forms the basis of the nucleosome, the most basic level of chromosome organisation (218). The nucleosome comprises 147bp of DNA wound round an octamer consisting of a histone H3/H4 heterotetramer and two H2A/H2B dimers (218-220). In order for differentiated cells to express genes in a tissue-specific manner, covalent modifications are utilised that allow only certain DNA sequences to be accessed by transcription factors and therefore specific genes to be expressed (221). Histone modifications include acetylation, phosphorylation, ubiquitylation and methylation (219). There are six lysine (K) residues in histones that are targeted for methylation (221). DOT1L is the only histone methyltransferase known to methylate K79 of histone H3 (H3K79; 220). Genome-wide studies have demonstrated that H3K79 methylation is enriched in gene coding regions (222), and in mammals DOT1L is found in a protein complex with RNA polymerase II (Barry 2010) suggesting that DOT1L methylation of H3K79 has a role in initiating gene transcription (219, 223). *DOT1L* also has a role in haematopoiesis; Knockout of *DOT1L* in human bone marrow cells has been shown to deplete haematopoietic stem cells, various progenitor cell populations, and terminally differentiated blood cell lineages (224). In addition to this, DOT1L and methylation of H3K79 are important in leukaemias that are caused by chromosomal translocations of mixed-lineage leukaemia (MLL) an H3K4 methyltransferase (225, 226). Interaction between DOT1L and MLL fusion proteins causes mistargeting of DOT1L to MLL target genes such as the HOX transcription factors. Aberrant hypermethylation of H3K79 leads to constitutive transcriptional activation of these HOX genes promoting uncontrolled proliferation of leukaemia cells (227).

A 2012 GWAS looking at the width of the joint space in the hips of 6523 participants from the Rotterdam study, identified a genome-wide significant signal in intron 1 of *DOT1L* ($P = 4.5x10^{-10}$; 132). JSW is considered a proxy for cartilage thickness and a

reduction in the JSW is indicative of cartilage degradation (Auleley 1998). This study was then replicated in three UK populations totalling 4,442 cases and the two studies were combined in a meta-analysis. The meta-analysis showed that the minor G-allele of rs12982744 is associated with a 12% reduced risk of hip OA (OR = 0.88, P = $1.5x10^{-4}$) and with an increased JSW of 0.09mm per copy of the G allele, implying that homozygote carriers of the G allele should have ~5% thicker cartilage than the control group (132). In a second GWAS totalling 9789 hip OA cases and 31873 controls it was discovered that in male subjects the C-allele of rs12982744 was associated with a 17% increased risk of hip OA (OR 1.17, P = 7x10-9; 113).

Having demonstrated that knockdown of *Dot1l* in the chondrogenic mouse cell line ATDC5 caused reduced synthesis of sulfated proteoglycan, reduced levels of type X collagen and aggrecan and increased expression of *Mmp9*, Castano Betancourt *et al.* hypothesised that *DOT1L* was the gene underlying the association signal by influencing chondrogenic differentiation. They also demonstrated that DOT1L physically interacts with TCF4 (132), a transcription factor interacting with β-catenin. As Wnt/β-catenin signalling is critical in the formation of cartilage and bone and in the development of the synovial joint (228) they investigated whether mRNA expression of Wnt target genes was affected in *Dot1l*$^{-}$ cells. They discovered that *Tcf1, Axin2* and *c-Myc* levels were increased in controls but not in *Dot1l*$^{-}$ cells, and there was an increase in osteocalcin, which is negatively regulated by Wnt/β-catenin signalling (132). Together these results suggest a role for DOT1L in chondrocyte differentiation through its interaction with the Wnt/β-catenin pathway.

6.2 Aim

The aim of this chapter is to investigate whether the OA associated SNP rs12982744 is influencing the expression of *DOT1L*.

6.3 Results

6.3.1 Quantitative Comparison of DOT1L Expression between Cartilage from OA and NOF Patients. *DOT1L* was analysed for expression by Real-Time qPCR (Chapter 2, section 2.6 -7) in cDNA from the hip cartilage of OA patients undergoing

THR and the hip cartilage of patients who had undergone surgery for NOF fracture (Figure 6.1). A significant increase in the expression of *DOT1L* in control cartilage compared to OA cartilage was seen (P < 0.0001). As the *DOT1L* signal was associated with hip OA, the data were then assessed for a difference in *DOT1L* expression between OA hip and OA knee cartilage.

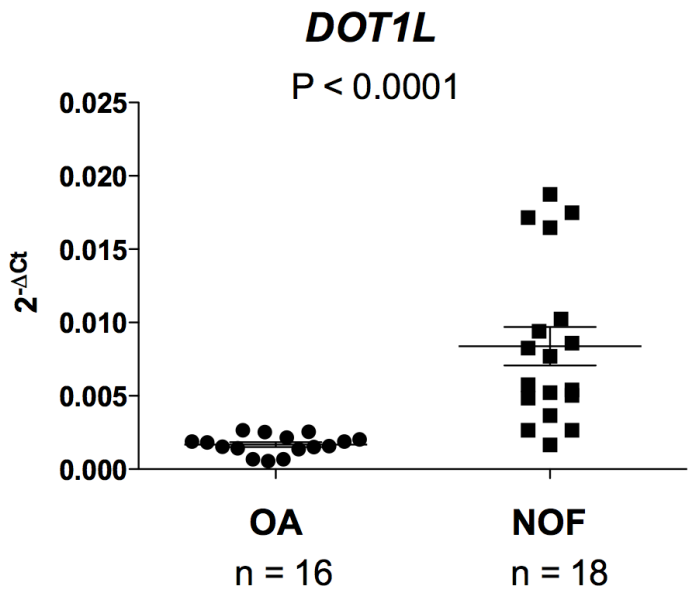

Figure 6.1 Columnar scatter plot of the quantitative expression of *DOT1lL* between cDNA from OA hip cartilage and cDNA from control hip cartilage. The control cartilage was collected from patients who had undergone surgery for a neck of femur (NOF) fracture. n is the number of patients studied. The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The horizontal lines in each plot represent the mean and the standard error of the mean. The P-value was calculated using a Mann-Whitney U test.

6.3.2 Quantitative Comparison of DOT1L Expression between Cartilage from OA Hips and Cartilage from OA Knees. *DOT1L* expression was then compared between knee cartilage from OA patients who had undergone TKR and hip cartilage from OA patients who had undergone THR (Figure 6.2). No significant difference in *DOT1L* expression was seen between cDNA from OA hip and cDNA from OA knee cartilage (P = 0.35). The data were then analysed for a correlation between *DOT1L* expression and genotype at rs12982744 to ascertain whether this signal could be modulating expression of *DOT1L* in cartilage from patients undergoing TJR for OA.

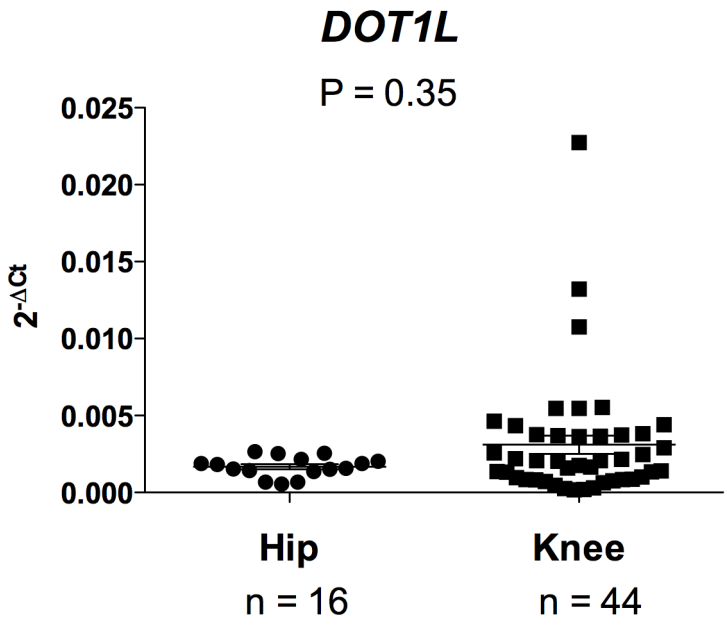

Figure 6.2 Columnar scatter plot of the quantitative expression of *DOT1L* between cDNA from OA hip cartilage and cDNA from OA knee cartilage. n is the number of patients studied. The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The horizontal lines in each plot represent the mean and the standard error of the mean. The P-value was calculated using a Mann-Whitney U test.

6.3.3 Quantitative Expression of DOT1L in Cartilage Stratified by Genotype at rs12982744. To investigate *DOT1L* for cartilage eQTLs, patients already analysed for overall expression were genotyped for rs12982744 by RFLP (Chapter 2, section 2.8) and the data were stratified by genotype at this SNP (Figure 6.3). No correlation was observed between the three genotypic groups for rs12982744 (P = 0.63). The analysis was then repeated after separating the hip and the knee cases (Figure 6.4). Again, no correlation was observed between *DOT1L* expression and the three different genotypic groups for rs12982744 for either hips or knees (P = 0.48 and 0.81 respectively).

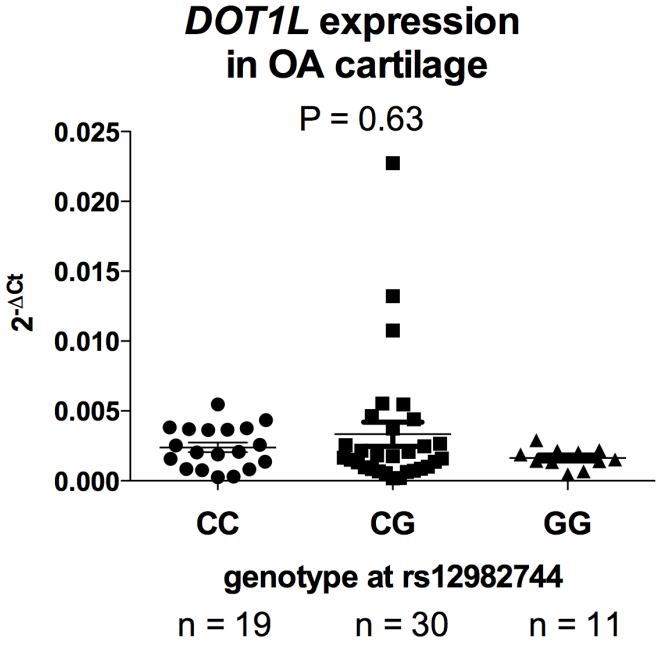

Figure 6.3 Columnar scatter plot of the quantitative expression of *DOT1L,* with patients stratified by genotype at the associated SNP rs12982744. n is the number of patients studied. The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The horizontal lines in each plot represent the mean and the standard error of the mean. The P-value was calculated using a Kruskal-Wallis Test.

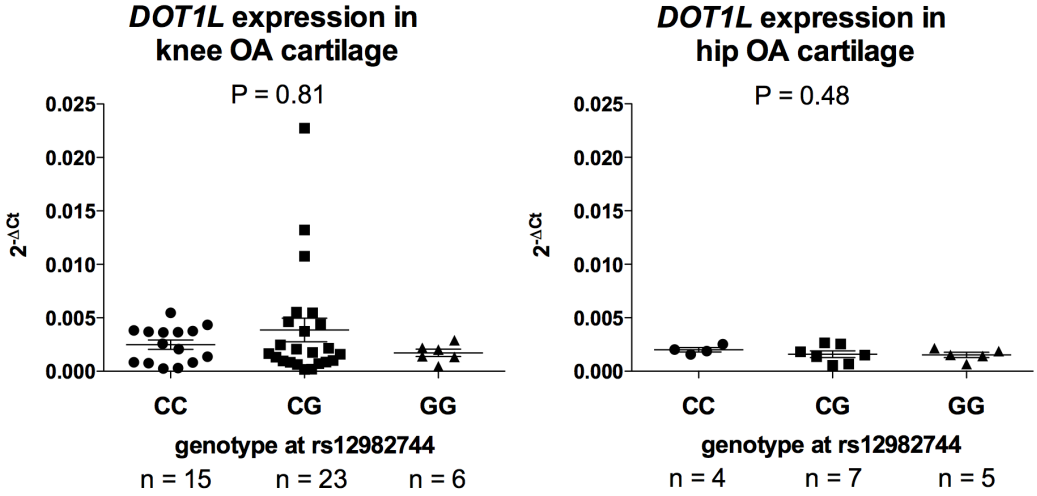

Figure 6.4 Columnar scatter plots of the quantitative expression of *DOT1L* in cDNA from OA knee (left panel) and OA hip cartilage (right panel), with patients stratified by genotype at the OA-associated SNP rs12982744. n is the number of patients studied. The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The horizontal lines in each plot represent the mean and the standard error of the mean. P-values were calculated using a Kruskal-Wallis test.

As in Chapter 4, despite the lack of correlation between *DOT1L* expression and genotype the samples were investigated for AEI analysis, as stratification of overall expression by genotype could be vulnerable to natural fluctuation in gene expression and as such may not pick up any difference in expression levels between the two alleles. As rs12982744 is an intronic SNP, and therefore cannot itself be interrogated for AEI, a database search was carried out (Chapter 2, section 2.9). This revealed four possible transcript SNPs within *DOT1L* that could be utilised for AEI, none of which were in high LD with rs12982744 (Table 6.1).

Table 6.1 The transcript SNPs identified for AEI and their pairwise D′ and r^2 values relative to the associated SNP rs12982744

Transcript SNP	Alleles	MAF[1]	Pairwise linkage disequilibrium relative to rs12982744	
			r^2	D′
rs3815308			0.462	0.889
rs2302061			0.084	1
rs740404			0.079	1
rs1558118			0.051	1

[1] Minor allele frequency in Europeans (HapMap CEU).

Taqman could not be used for AEI analysis as no ready-made genotyping assays were available for any of the transcript SNPs, and none of the custom genotyping assays passed Applied Biosystems internal quality control (QC). Two other methods were investigated as possible methods to analyse *DOT1L* for AEI: Sequenom and Pyrosequencing, however all assays designed for these two methods failed QC and so AEI could not be assessed.

6.4 Discussion

The aim of this chapter was to assess whether *DOT1L* expression correlated with genotype at rs12982744 using cartilage from patients who had undergone total joint replacement for primary OA. The results did not provide evidence for a correlation between the OA associated SNP rs12982744 and *DOT1L* expression. The lack of a correlation between rs12982744 and *DOT1L* expression implies that the association

marked by this SNP is operating by a route other than an effect on the expression of the gene. However *DOT1L* could not be successfully assessed for AEI so it cannot be conclusively said that rs12982744 is not regulating expression of *DOT1L*. Expression was very low in our cartilage samples, with Cts between 32-38, and it has been observed that *DOT1L* has a low abundance (229). In addition to this, this chapter demonstrated that *DOT1L* expression decreases with chondrogenesis and that there is a significant decrease in *DOT1L* expression in OA cartilage compared to NOF cartilage. It is therefore possible that it was difficult to assess AEI in *DOT1L* due to its low expression in mature cartilage, particularly in OA.

The association between the rs12982744 and cartilage thickness was also observed in controls without OA, indicating that the association with cartilage thickness is present before the onset of OA and implicating involvement of *DOT1L* in formation of the articular cartilage in skeletal development (132). In addition to this rs12982744 has been associated with height, further suggesting a role for *DOT1L* in skeletal development (230, 231). This indicates that it is possible that rs12982744 is modulating the expression of *DOT1L* at time points not studied in this chapter, for example during skeletogenesis or in adolescent tissues. In the future it would be beneficial to investigate overall expression and perform AEI analysis of *DOT1L* in the joint tissues of young individuals to see if a correlation between rs12982744 and *DOT1L* expression is revealed. If there had been more time to further investigate *DOT1L* AEI it may have been beneficial to select OA individuals that demonstrated high levels of overall *DOT1L* expression to validate the AEI assays, as low levels of expression will affect the QC of an assay.

Chapter 7: Genetic and Functional Analysis of *GNL3,* a Gene at an OA Susceptibility Locus on Chromosome 3p.

7.1 Introduction

Guanine nucleotide binding protein-like 3 or nucleostemin, encoded by *GNL3,* is a GTP-binding nucleolar protein that is expressed at high levels in embryonic and adult stem cells and tumour cells (232-234), but not in terminally differentiated cells (235). During differentiation *GNL3* levels are rapidly down-regulated immediately prior to exiting the cell cycle and siRNA-mediated depletion of *GNL3* has been shown to cause defects in the transition from metaphase to anaphase in cultured cells leading to cell cycle arrest (236-238). This indicates that reduction of *GNL3* has a role in cell cycle exit rather than being a consequence of differentiation (237-239). Knockout studies have also shown that *GNL3* is essential for cell proliferation and embryogenesis (238-240), and complete loss of function of *GNL3* causes early lethality of the embryo (233). In addition to depletion of *GNL3*, its overexpression also causes cell cycle arrest (237, 238). GNL3 exerts these effects on the cell cycle via its interactions with the tumour-suppressor p53. The unstable p53 protein is stabilised in order to activate transcription of genes that encode cell cycle proteins (238). In unstressed cells p53 is maintained at low levels by its inhibitor, mouse double minute-2 (MDM2), an E3 ubiquitin ligase. Knockdown of *GNL3* by siRNA causes cell stress leading to inhibition of MDM2 function and the subsequent stabilisation and activation of p53 (237). Overexpression of *GNL3* also leads to p53 activation and therefore cell cycle arrest, via direct binding of the excess GNL3 to MDM2 (237, 238).

p53 activation can also occur as a result of the disruption of ribosomal biogenesis (237). Ribosomal biogenesis is the process by which the cellular machinery required for protein synthesis is made (238). Ribosomal biogenesis includes the synthesis and processing of rRNA and the formation of ribosomal subunits and the mature ribosome (241). Disturbance of any step in the process can trigger ribosomal stress releasing free ribosomal proteins into the nucleoplasm, which directly bind MDM2 (242). GNL3 is a component of the nucleolar protein complex critical for rRNA processing, and reduction of endogenous GNL3 levels reduces the number of mature rRNAs, which may lead to the release of free ribosomal proteins (238).

GNL3 also binds and modulates telomeric repeat binding factor-1 (TRF1) a component of the telomere-capping complex, telosome (240). TRF1 prevents elongation of the telomeres, protecting continuously dividing cells from telomere degradation (233, 236, 240). Overexpression of GNL3 causes a reduction in TRF1 (233), which in turn leads to telomere degradation causing cellular senescence (243). Deletion of TRF1 also causes early embryonic lethality in mice, suggesting that it has an essential function in early embryogenesis (233, 236).

A large case-control GWAS was performed using 7410 unrelated OA patients and 11009 unrelated controls. The most promising signals were replicated in an independent set of 7473 cases and 42938 controls (135). Five of the loci showed GWS ($P = 5x10^{-8}$) and the most significant of these was on chromosome 3, identified by two SNPs in perfect LD with each other: rs11177 (allele A, $P = 1.25x10^{-10}$), a missense polymorphism within exon three of *GNL3* causing an Arginine-Glutamine (Arg-Gln) substitution, and rs6976 (allele T, $P = 7.24x10^{-11}$) situated in the 3′UTR of *GLT8D1,* both associated with hip and knee OA in patients who had undergone joint replacement surgery (135). The function of GNL3 in bone and cartilage is not known, with mice homozygous for deletion of *GNL3* showing very early embryonic lethality (233). Immunohistochemical staining showed strong nucleolar expression of GNL3 in primary cultured chondrocytes from patients with OA but GNL3 was barely detectable in cultured chondrocytes from controls, raising the possibility that this gene might by functionally important in the pathogenesis of OA (135).

7.2 Aim

The aim of this chapter is to investigate whether *GNL3* contributes functionally to the OA association signal marked by rs11177 and rs6976.

7.3 Results

7.3.1 Database Search of Amino Acid Prediction Websites. rs11177 is a missense polymorphism causing an Arg-Gln amino acid substitution, so a search of databases that predict the likelihood that an amino acid change will affect protein function was conducted (Chapter 2, section 2.17). All three databases predicted the substitution to be tolerated. According to PolyPhen, the tertiary protein structure had no homology to any known protein structures so the protein sequence was therefore sent for modelling, to

provide a clearer assessment of whether the substitution is likely to have a detrimental affect on the protein.

7.3.2 Modelling the GNL3 Protein Structure. The GNL3 protein was modelled by Jeremy Lakey, Professor of Structural Biochemistry, Institute for Cell and Molecular Biosciences, Newcastle University. Due to the lack of homology of the GNL3 protein with any other proteins, only the Guanosine Triphosphate (GTP)-binding domain could be modelled. Next to the GTP binding domain is an acidic domain, a coiled coil and a basic domain, which may have DNA binding capabilities. The rs11177 Arg-Gln change does not lie in any of the domains but lies in the linker region between the coiled coil and the basic domains. This, together with the database search of the prediction websites strongly suggests that the amino acid change is tolerated and does not affect the protein. Having ruled out the possibility that the association signal is marking this substitution in GNL3, the next step was to look at the possibility that it could be marking a transcriptional effect on the genes expression.

7.3.3 Quantitative Comparison of GNL3 Expression between Cartilage from OA and NOF Patients. Expression of *GNL3* was performed by Real-Time qPCR (Chapter 2, section 2.6 - 7) using hip cartilage cDNA from OA patients who had undergone THR and control patients who had undergone surgery for NOF fracture (Figure 7.1). No significant difference in expression was observed between OA and NOF cartilage (P = 0.76).

7.3.4 Quantitative Comparison of GNL3 Expression between Cartilage from OA Hips and Cartilage from OA Knees. *GNL3* expression in OA knee cartilage from patients who had undergone TKR was then compared to expression in OA hip cartilage from patients who had undergone THR (Figure 7.2). Again, no significant difference in *GNL3* expression was observed (P = 0.35).

7.3.5 Quantitative Expression of GNL3 in Cartilage Stratified by Genotype at the Associated SNP. To ascertain whether *GNL3* harboured a cartilage eQTL that correlated with the OA association signal, rs11177 was genotyped by RFLP analysis (Chapter 2, section 2.8) in all the OA patients analysed for overall *GNL3* expression and the expression data were then stratified by the genotype at this SNP (Figure 7.3).

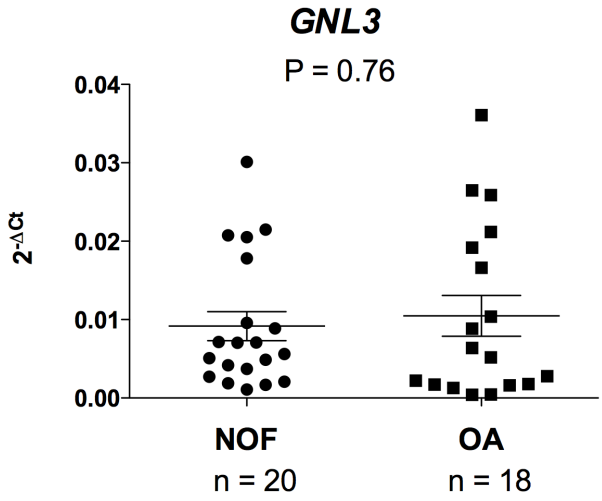

Figure 7.1 Columnar scatter plot of the quantitative expression of *GNL3* between cDNA from OA hip cartilage and cDNA from control hip cartilage. The control cartilage was collected from patients who had undergone surgery for a neck of femur fracture. n is the number of patients studied. The results were normalised against the mean Ct values of the *HPRT1*, *GAPDH* and *18S* housekeeping genes. The horizontal lines in each plot represent the mean and the standard error of the mean. The P-value was calculated using a Mann-Whitney U test.

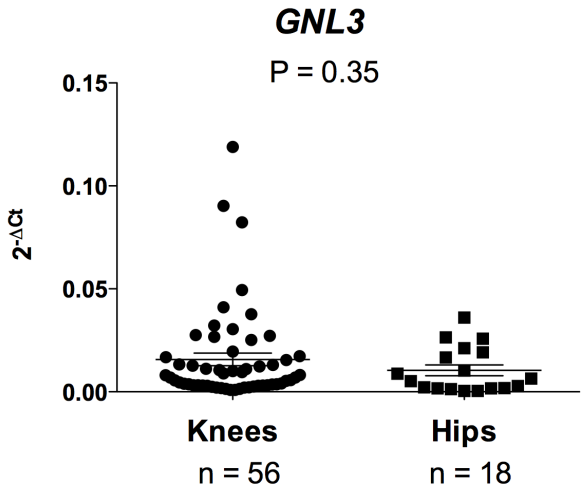

Figure 7.2 Columnar scatter plot of the quantitative expression of *GNL3* between cDNA from OA knee cartilage and OA hip cartilage. n is the number of patients studied. The results were normalised against the mean Ct values of the *HPRT1*, *GAPDH* and *18S* housekeeping genes. The horizontal lines in each plot represent the mean and the standard error of the mean. The P-value was calculated using a Mann-Whitney U test.

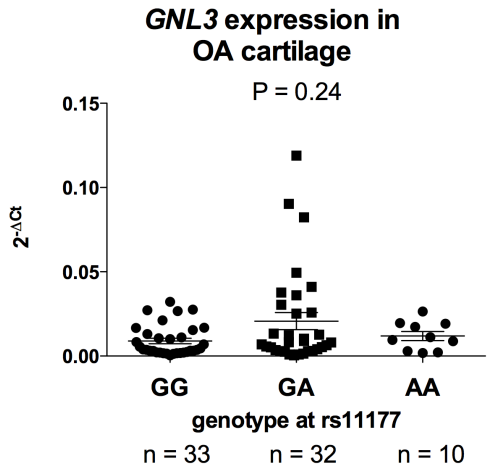

Figure 7.3 Columnar scatter plot of the quantitative expression of *GNL3* in cDNA from OA knee and hip cartilage, with patients stratified by genotype at the associated SNP rs11177. n is the number of patients studied. The results were normalised against the mean Ct values of the *HPRT1*, *GAPDH* and *18S* housekeeping genes. The horizontal lines in each plot represent the mean and the standard error of the mean. The P-value was calculated using a Kruskal-Wallis test.

No correlation was observed between *GNL3* expression and genotype at rs11177. The analysis was then repeated after separating the knee and hip data (Figure 7.4).

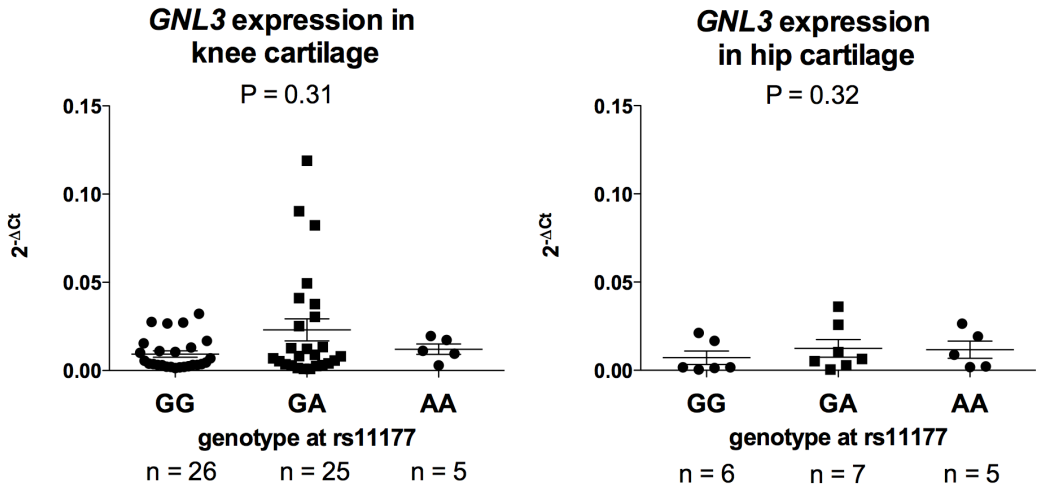

Figure 7.4 Columnar scatter plots of the quantitative expression of *GNL3* in cDNA from OA knee (left plot) and OA hip (right plot) cartilage, with patients stratified by genotype at the associated SNP rs11177. n is the number of patients studied. The horizontal lines in each plot represent the mean and the standard error of the mean. P-values were calculated using a Kruskal-Wallis test for the knees and a Mann-Whitney U test for the hips.

No correlation was seen between the expression and rs11177 genotype for either hips or knees (P = 0.32 and 0.31 respectively). However, this method is sensitive to natural variation in gene expression and can therefore be subject to false negatives, so the SNP was taken forward for AEI analysis.

7.3.6 AEI analysis of rs11177 in joint tissues. As rs11177 is a transcript SNP it could be directly interrogated for AEI. AEI analysis at rs11177 was carried out by Professor John Loughlin (Chapter 2, section 2.11). Of the 20 OA patients analysed (patient details are listed in Table A1), 11 showed significant AEI (Figure 7.5, Table 7.1). One of the 9 cartilage samples was from a NOF fracture (patient 182), and this also demonstrated significant AEI (P = 0.03), indicating that AEI is not dependent on disease status. In addition to the 20 joint tissues, blood samples from three controls not suffering from OA were also analysed. None of the three samples demonstrated AEI (patients 193, 194 and 195 in Table 7.1 and Figure 7.5). For all the joint tissues analysed the OA-associated A allele showed decreased expression.

Table 7.1 The allelic ratios for the samples analysed for AEI at rs11177. Allelic ratios are shown ± the standard error of the mean. P-values were calculated using a Mann-Whitney U Test. P-values ≤ 0.05 are highlighted in bold.

	rs11177	
Patient	Allelic ratio (A/G)	*P* value
27	0.75 ± 0.11	0.22
37	**0.73 ± 0.05**	**0.03**
39	**0.67 ± 0.06**	**0.03**
54	**0.74 ± 0.03**	**<0.01**
57	**0.82 ± 0.01**	**<0.01**
63	0.61 ± 0.18	0.15
131	**0.81 ± 0.03**	**0.02**
69	**0.70 ± 0.08**	**<0.01**
182	**0.76 ± 0.08**	**0.03**
49	**0.83 ± 0.03**	**0.03**
53	0.93 ± 0.02	0.22
58	0.90 ± 0.03	0.22
127	0.93 ± 0.02	0.06
136	0.75 ± 0.01	0.10
148	0.86 ± 0.04	0.15
67	0.88 ± 0.05	0.22
70	**0.85 ± 0.03**	**<0.01**
139	**0.66 ± 0.06**	**0.02**
60	0.87 ± 0.05	0.06
66	**0.81 ± 0.01**	**<0.01**
193	1.03 ± 0.03	0.69
194	0.97 ± 0.04	0.55
195	1.00 ± 0.06	0.84

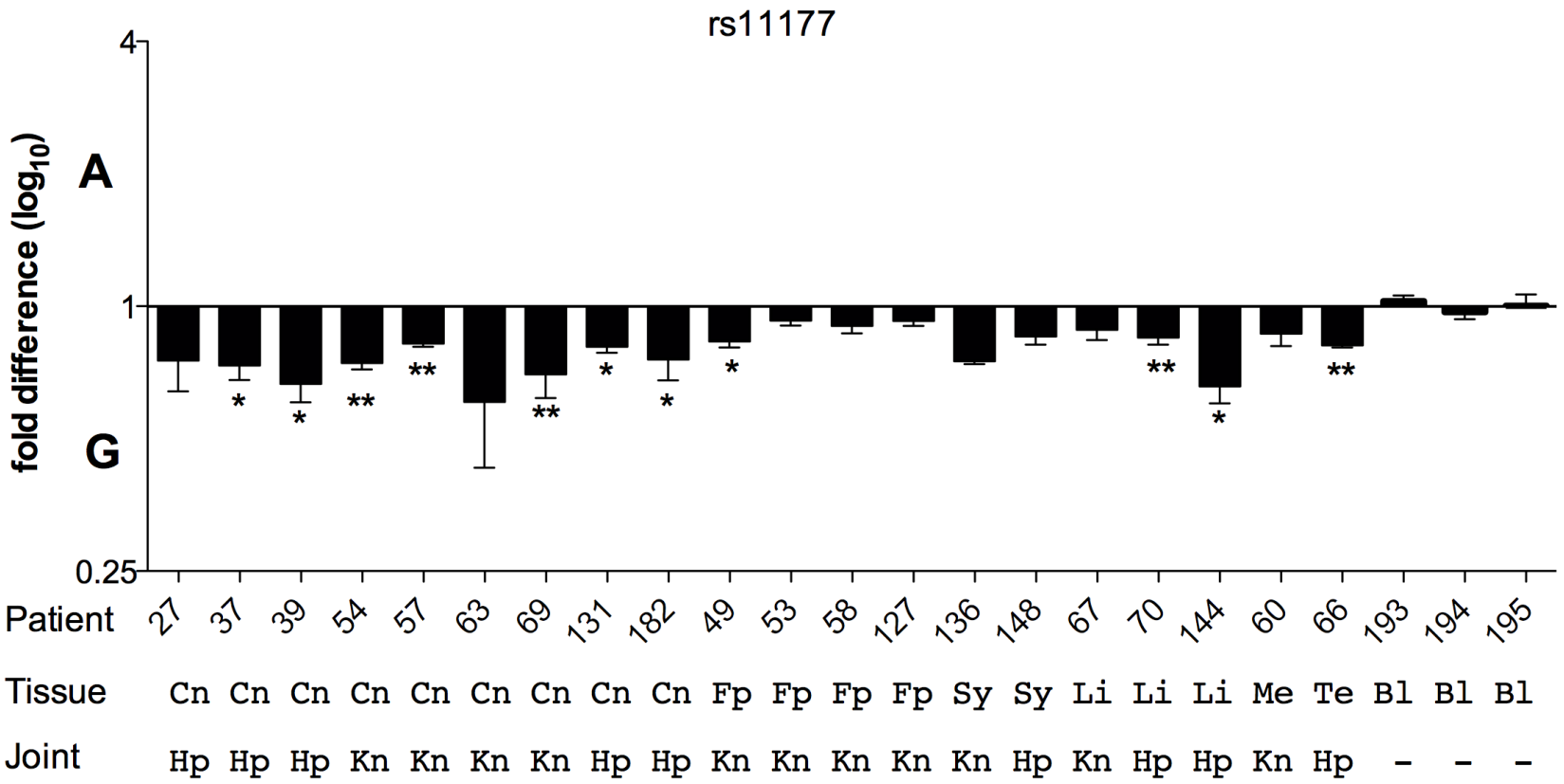

Figure 7.5 Allelic expression analysis of the transcript SNP rs11177 using cDNA from joint tissues and blood. Cn, cartilage; Fp, infrapatellar fat pad; Sy, synovium; Li, ligament; Me, meniscus; Te, tendon; Bl, blood. The cDNA allelic ratio in each sample was compared to the DNA ratio using a Mann-Whitney U Test. The error bars represent the standard error of the mean. ** P <0.01 and * P < 0.05.

7.3.7 siRNA-Mediated Knockdown of GNL3 in HACs. Since the AEI showed decreased expression of the OA-associated allele in joint tissues particularly cartilage, *GNL3* expression was knocked down in cultured HACs (Chapter 2, section 2.18; Figure 7.6). The knockdown of GNL3 protein was confirmed by western blotting (Chapter 2, section 2.19; Figure 7.6).

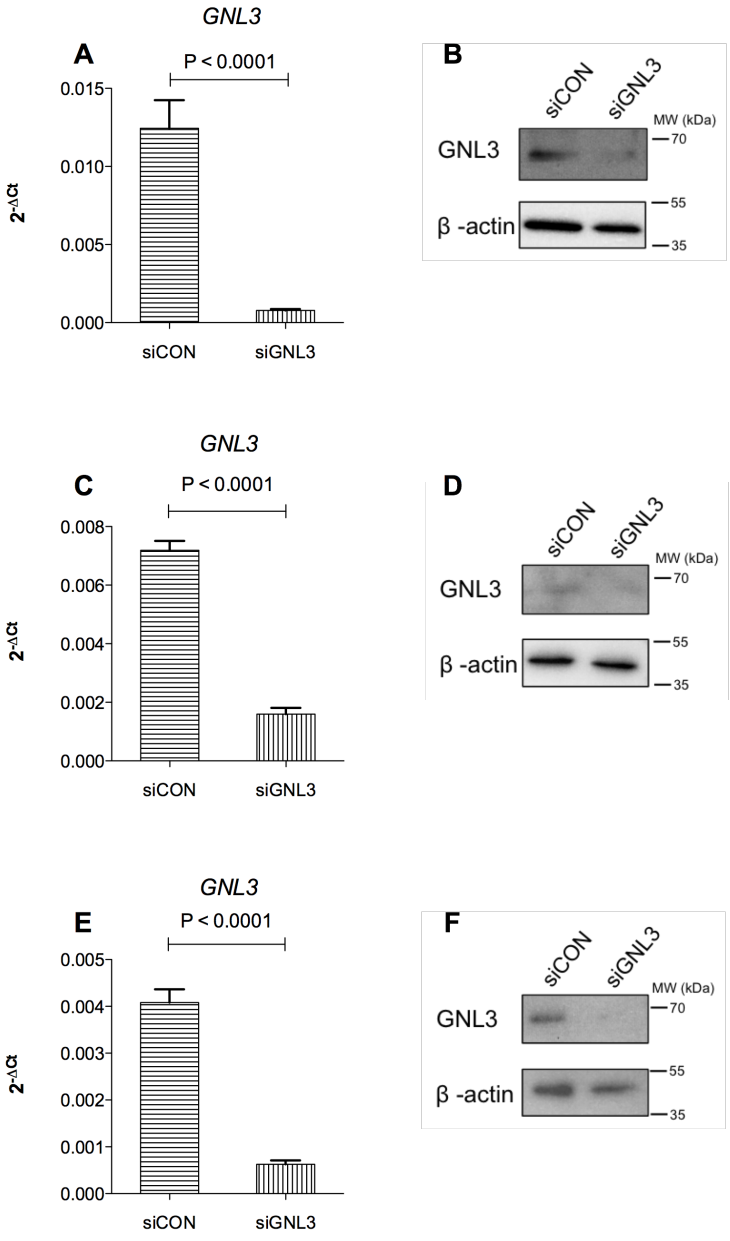

Figure 7.6 siRNA-mediated knockdown of *GNL3* in human articular chondrocytes. A, C and E, show Real-Time quantitative expression of *GNL3* in HACs from three donors (patients 196, 197 and 198 respectively; patient details can be found in the appendix Table A1) after treatment with an siRNA specific for *GNL3* (siGNL3) and a non-targeting control (siCON). The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The error bars represent the standard error of the mean. The P-values were calculated using two-tailed students t-tests. B, D and F show western blots of GNL3 and β-actin, the latter used as a loading control.

7.3.8 Real-Time Quantitative PCR Expression Analysis of a Panel of 12 Genes after GNL3 Knockdown in HACs. To investigate what effect *GNL3* knockdown had on a panel of genes important in cartilage anabolism and catabolism (Table 7.2), Real-Time qPCR (Chapter 2, section 2.6) was performed on cDNA extracted from the HACs (Chapter 2, section 2.18.4).

Table 7.2 The 12 genes analysed for expression following *GNL3* knockdown.

Gene	Gene Product	Role in cartilage
ACAN	Aggrecan	Anabolism
COL2A1	Type II collagen, α-1 chain	
SOX9	SRY-box 9 transcription factor	
MMP1	Matrix metalloproteinase 1	Catabolism
MMP9	Matrix metalloproteinase 9	
MMP13	Matrix metalloproteinase 13	
ADAMTS4	A disintegrin and metalloproteinase with thrombospondin motifs 4	
ADAMTS5	A disintegrin and metalloproteinase with thrombospondin motifs 5	
TIMP1	Tissue inhibitor of matrix metalloproteinase 1	Anti-catabolism
RUNX2	Runt-related transcription factor 2	Osteogenesis
COL1A1	Type I collagen, α-1 chain	
COL10A1	Type X collagen, α-1 chain	Hypertrophy

The results varied between the three patients analysed, with some genes showing a significant difference in expression in one patient but not others (Figures 7.7-7.9; Table 7.3), including, *ACAN* ($P = 0.0003$) in patient 196; *SOX9* ($P = 0.004$), *MMP9* ($P = 0.03$), *TIMP1* ($P = 0.01$), and *RUNX2* ($P = 0.04$), in patient 197; and *MMP13* ($P = 0.003$) in patient 198 (Figures 7. 8 - 7.10). Two genes showed a significant difference in expression in two patients but not the third, *COL2A1* in patients 196 and 198 ($P = 0.006$, 0.008 respectively), and *COL10A1* in patients 196 and 197 ($P = 0.04$, 0.02 respectively). *COL1A1* showed a significant difference in expression in all three patients ($P < 0.0001$, $P = 0.03$, $P = 0.003$), however the expression levels were inconsistent, showing an increase in expression in patient 197 and a decrease in expression in patients 196 and 198 (Figures 7.7 - 7.9). The only genes to show a consistent significant difference in expression after *GNL3* knockdown were *ADAMTS4,* which showed an increase in expression in all three patients ($P < 0.0001$), and *ADAMTS5* which showed a decrease in expression in all three patients ($P < 0.0001$)

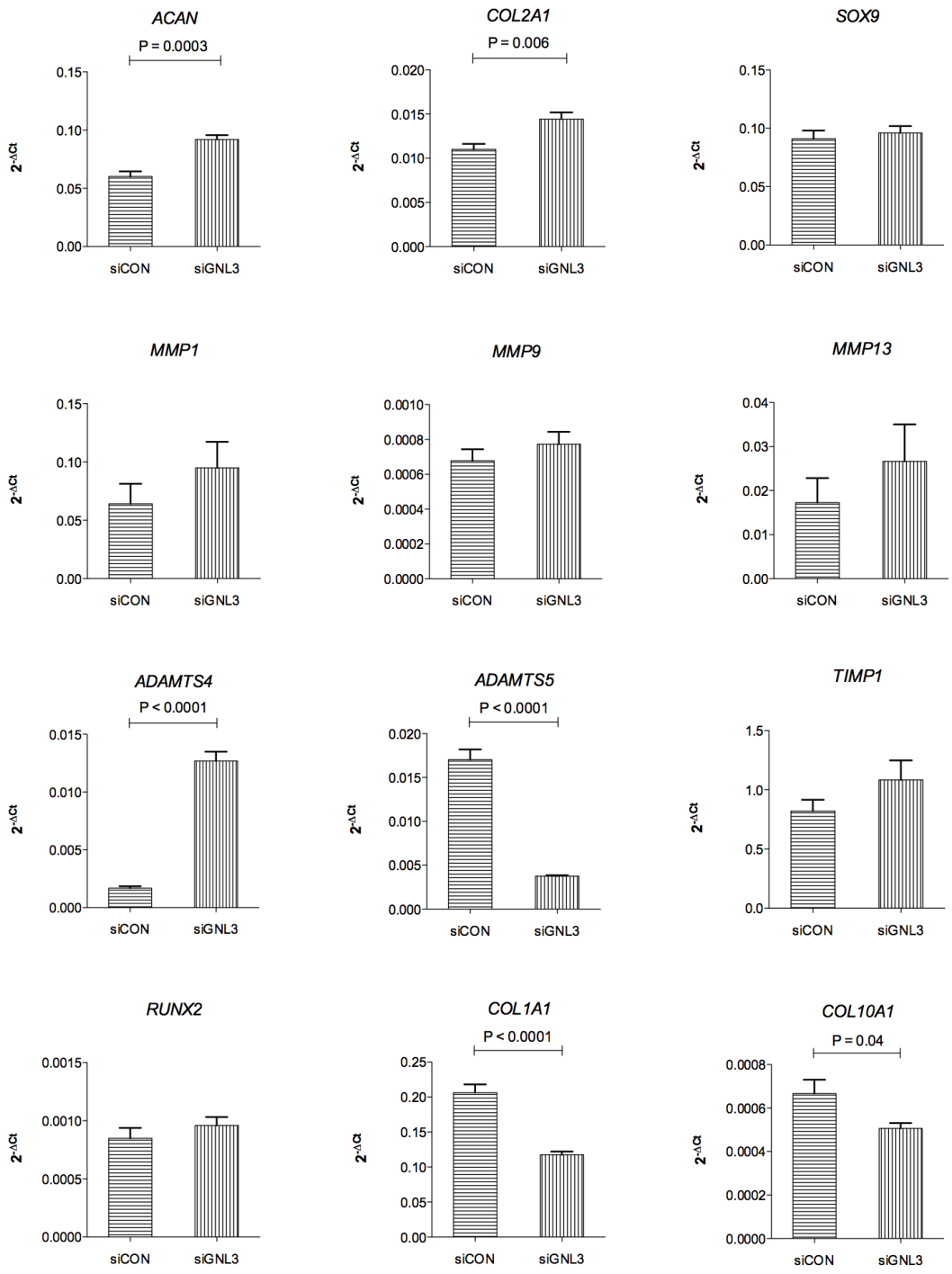

Figure 7.7 Real-Time quantitative expression of the 12 genes in human articular chondrocytes from patient 196 after treatment with siRNA specific for *GNL3* (siGNL3) and a non-targeting control (siCON). Each condition was performed with six technical replicates. The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The error bars represent the standard error of the mean. P values were calculated using two-tailed students t-tests.

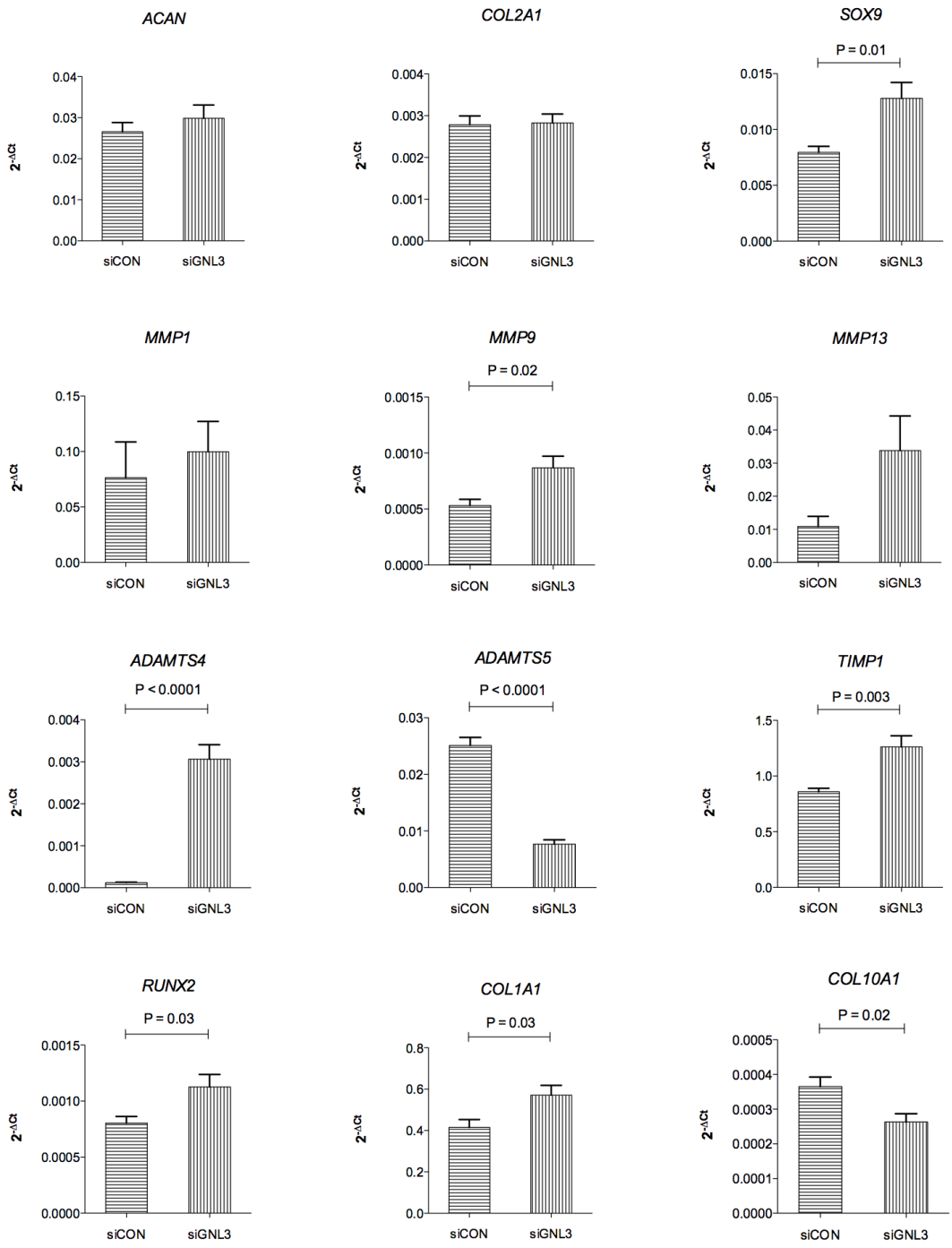

Figure 7.8 Real-Time quantitative expression of the 12 genes in human articular chondrocytes from patient 197 after treatment with siRNA specific for *GNL3* (siGNL3) and a non-targeting control (siCON). Each condition was performed with six technical replicates. The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The error bars represent the standard error of the mean. P values were calculated using two-tailed students t-tests.

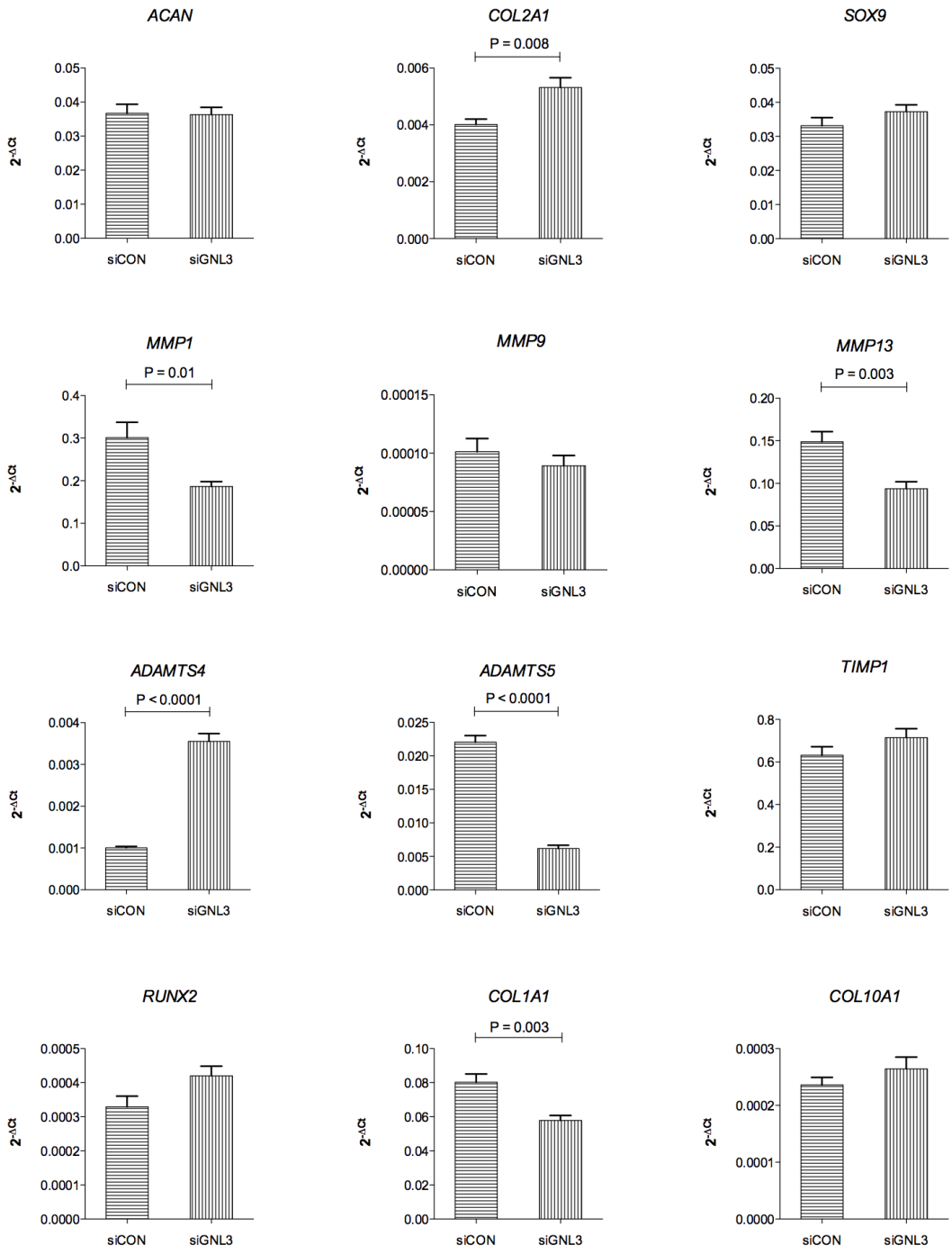

Figure 7.9 Real-Time quantitative expression of the 12 genes in human articular chondrocytes from patient 198 after treatment with siRNA specific for *GNL3* (siGNL3) and a non-targeting control (siCON). Each condition was performed with six technical replicates. The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The error bars represent the standard error of the mean. P values were calculated using two-tailed students t-tests.

Table 7.3 P- values of the 12 genes analysed for expression in all HAC patients studied following *GNL3* knockdown. The arrows indicate a significant increase or decrease in expression when *GNL3* is knocked down. The means of the biological replicates were also pooled to give the P-value across the three patients.

	Patient number							
Gene	196		197		198		Pooled	
ACAN	0.0003	↑	0.42		0.92		0.15	
COL2A1	0.006	↑	0.89		0.008	↑	0.31	
SOX9	0.59		0.01	↑	0.22		0.71	
MMP1	0.30		0.59		0.01	↓	0.56	
MMP9	0.35		0.02	↑	0.42		0.23	
MMP13	0.37		0.06		0.003	↓	0.68	
ADAMTS4	<0.0001	↑	<0.0001	↑	<0.0001	↑	<0.0001	
ADAMTS5	<0.0001	↓	<0.0001	↓	<0.0001	↓	<0.0001	
TIMP1	0.20		0.003	↑	0.18		0.01	
RUNX2	0.36		0.03	↑	0.06		0.11	
COL1A1	<0.0001	↓	0.03	↑	0.003	↓	0.83	
COL10A1	0.04	↓	0.02	↓	0.28		0.19	

7.3.9 siRNA-Mediated Knockdown of GNL3 in MSCs. As *GNL3* is expressed at high levels in both embryonic and adult stem cells (232, 233) siRNA-mediated knockdown of *GNL3* was also performed in mesenchymal stem cells (Chapter 2, section 2.18.2). Successful knockdown of *GNL3* mRNA was assessed using real-time quantitative PCR (Chapter 2, section 2.6) and successful knockdown of the GNL3 protein was confirmed by western blotting (Chapter 2, section 2.19, Figure 7.10).

7.3.10 Real-Time qPCR Expression Analysis of a Panel of 12 Genes after GNL3 Knockdown in MSCs. The cDNA extracted from the MSCs (Chapter 2, section 2.18.4) after siRNA-mediated *GNL3* knockdown was tested for differences in expression of the same 12 genes analysed in HAC cDNA. Again the results varied between the three patients analysed, with some genes showing a significant difference in expression in one patient but not others, such as *SOX9* ($P < 0.0001$) and *MMP13* ($P = 0.02$) for patient 199, and *TIMP1* ($P = 0.02$) for patient 201 (Figures 7.11-7.13: Table 7.4).

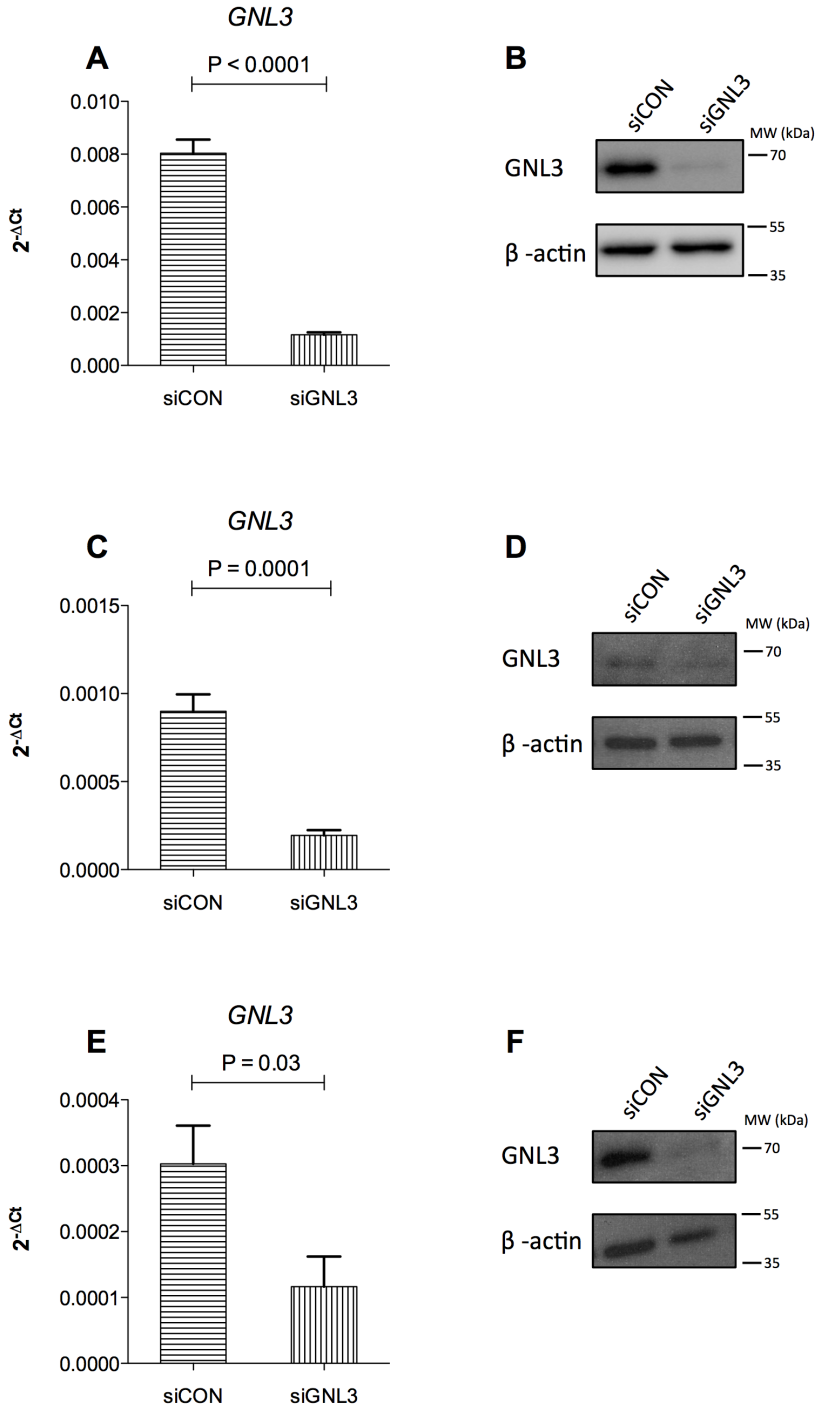

Figure 7.10 siRNA-mediated knockdown of *GNL3* in mesenchymal stem cells. A, C and E, show real time quantitative expression of *GNL3* in MSCs from three donors (patient 199, 200 and 201 respectively; patient details can be found in the appendix Table A1) after treatment with an siRNA specific for *GNL3* (siGNL3) and a non-targeting control (siCON). The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The error bars represent the standard error of the mean. The P values were calculated using two-tailed students t-tests. B, D and F show western blots of GNL3 and β-actin, the latter used as a loading control.

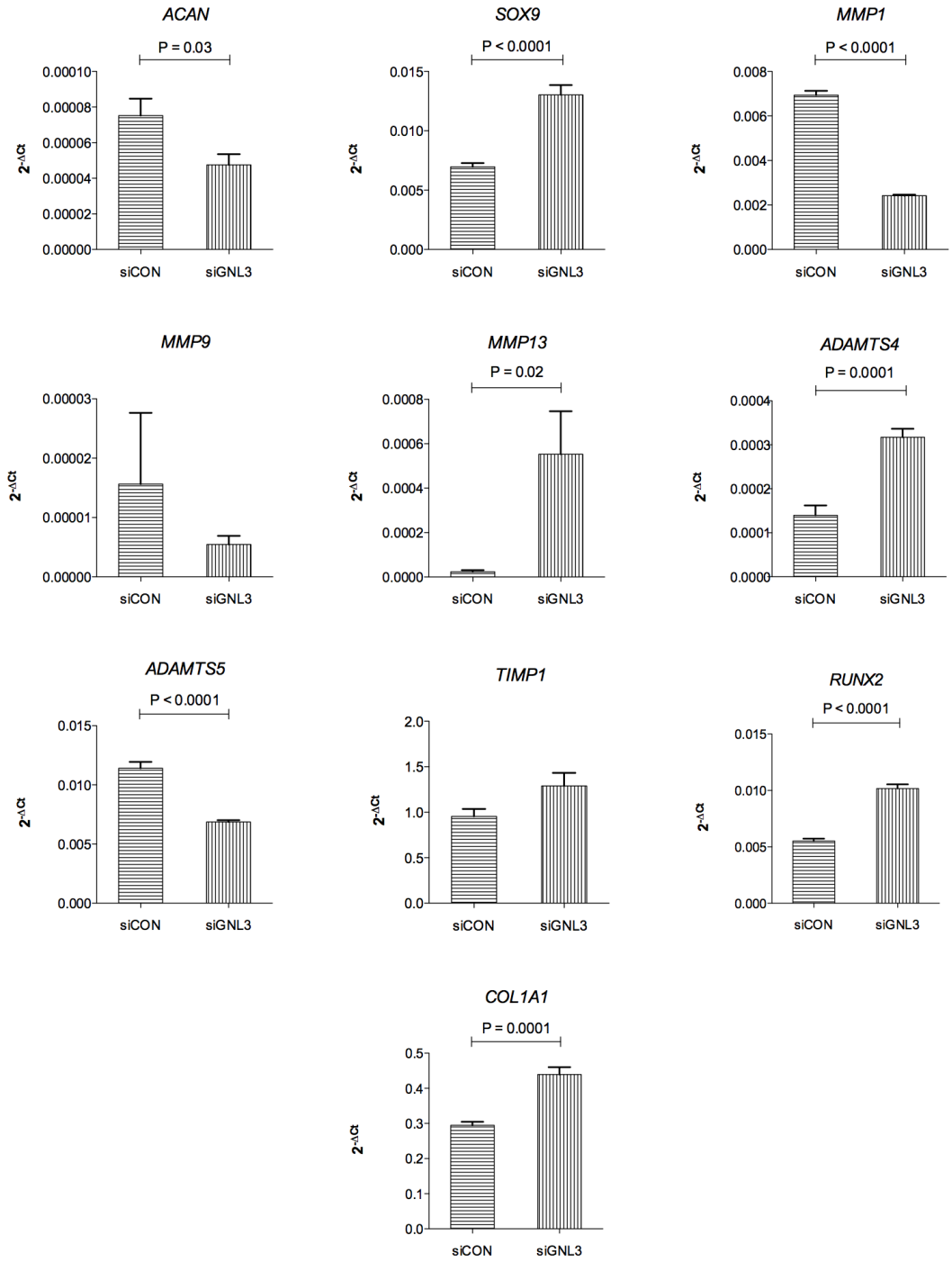

Figure 7.11 Real-Time quantitative expression of the 12 genes in mesenchymal stem cells from patient 199 after treatment with siRNA specific for *GNL3* (siGNL3) and a non-targeting control (siCON). Each condition was performed with six technical replicates. The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The error bars represent the standard error of the mean. P values were calculated using two-tailed students t-tests.

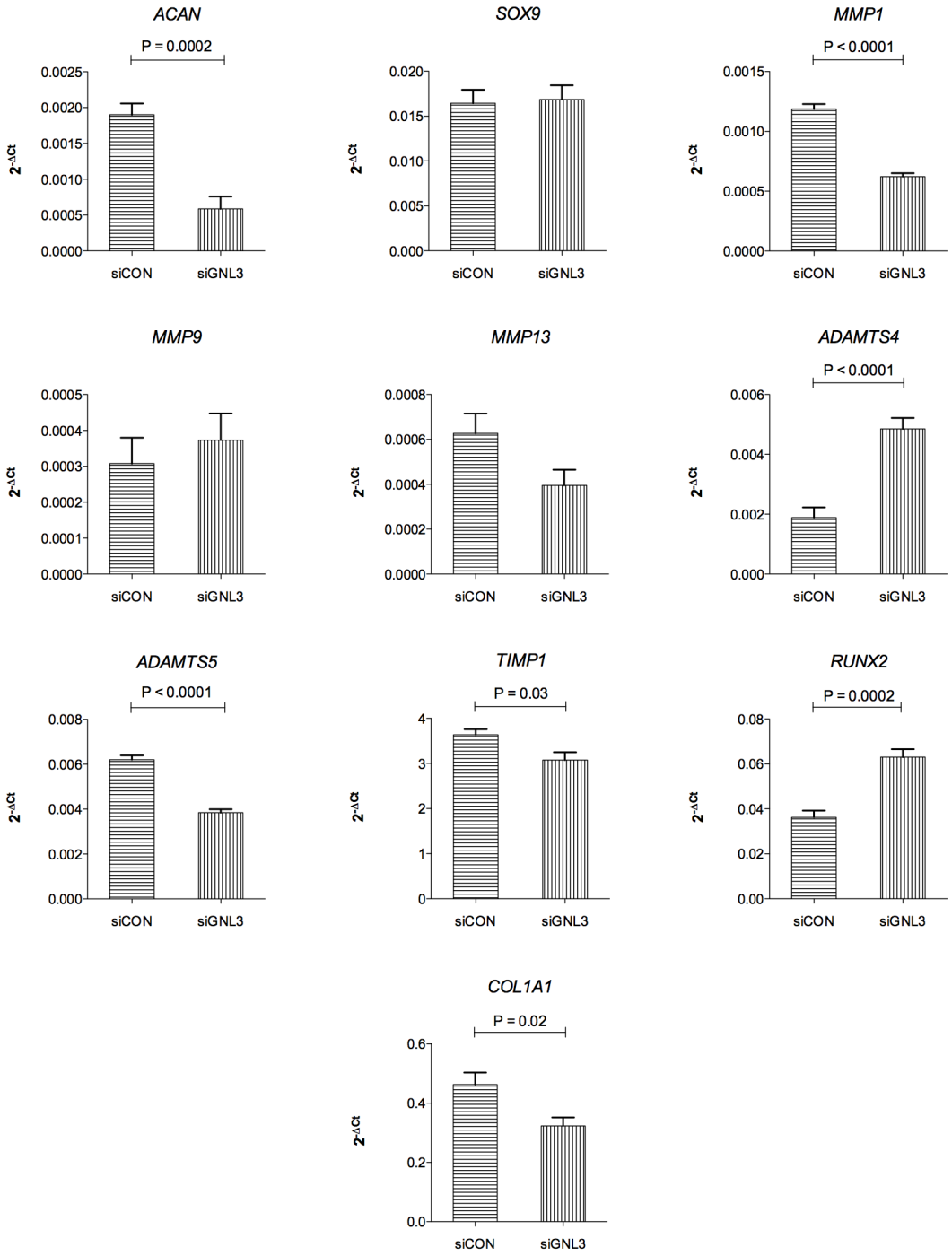

Figure 7.12 Real-Time quantitative expression of the 12 genes in mesenchymal stem cells from patient 200 after treatment with siRNA specific for *GNL3* (siGNL3) and a non-targeting control (siCON). Each condition was performed with six technical replicates. The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The error bars represent the standard error of the mean. P values were calculated using two-tailed students t-tests.

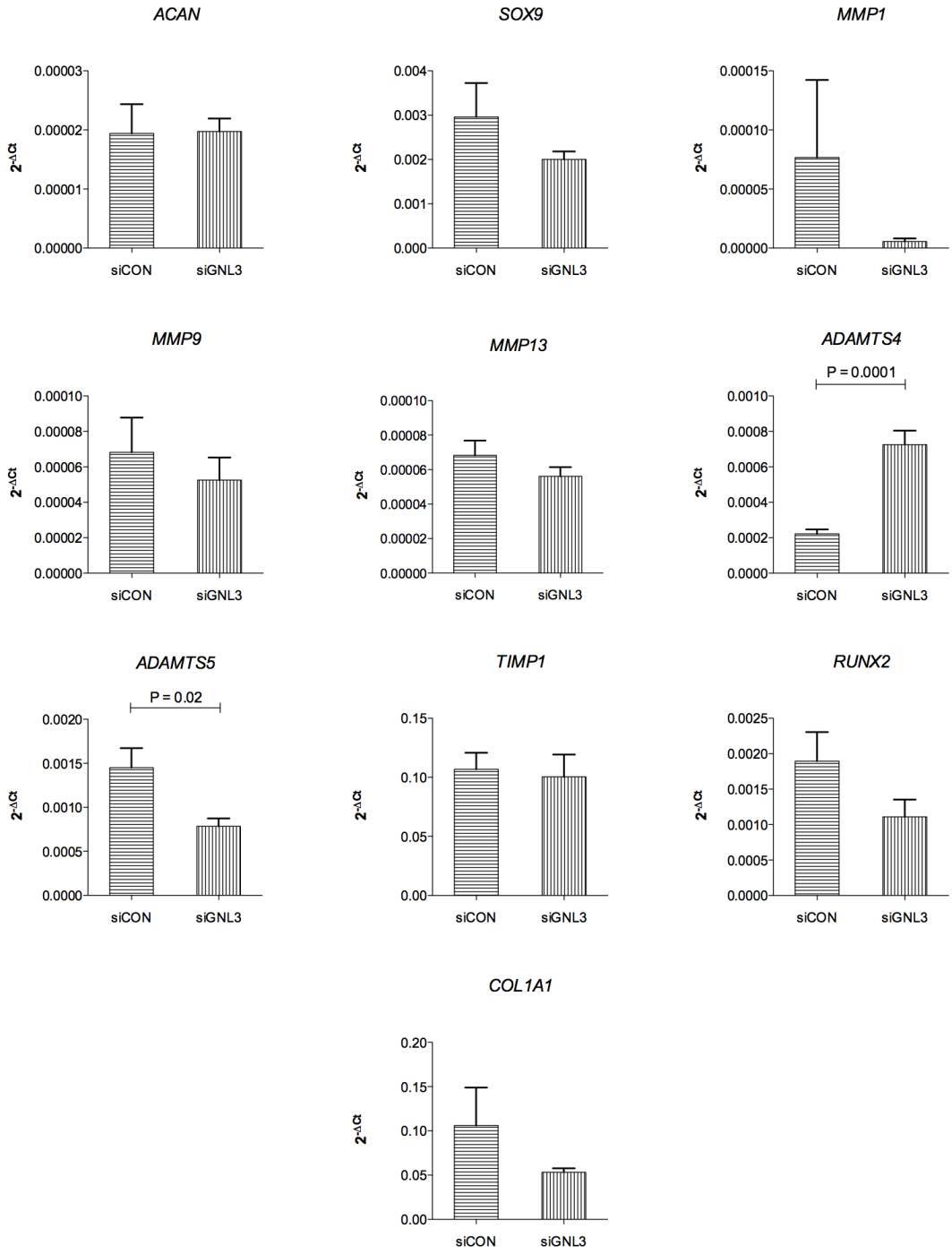

Figure 7.13 Real-Time quantitative expression of the 12 genes in mesenchymal stem cells from patient 201 after treatment with siRNA specific for *GNL3* (siGNL3) and a non-targeting control (siCON). Each condition was performed with six technical replicates. The results were normalised against the mean Ct values of the *HPRT1, GAPDH* and *18S* housekeeping genes. The error bars represent the standard error of the mean. P values were calculated using two-tailed students t-tests.

Table 7.4 P- values of the 12 genes analysed for expression in all MSC patients studied following *GNL3* knockdown. The arrows indicate a significant increase or decrease in expression when *GNL3* is knocked down. The means of the biological replicates were also pooled to give the P-value across the three patients. Genes that were not expressed in MSCs are marked as n/a.

	Patient number						
Gene	199		200		201		Pooled
ACAN	0.03	↓	0.0002	↓	0.95		0.06
COL2A1	n/a		n/a		n/a		n/a
SOX9	<0.0001	↑	0.85		0.25		0.41
MMP1	<0.0001	↓	<0.0001	↓	0.12		0.04
MMP9	0.49		0.56		0.52		0.87
MMP13	0.02	↑	0.07		0.25		0.45
ADAMTS4	0.0001	↑	<0.0001	↑	0.0001	↑	0.04
ADAMTS5	<0.0001	↓	<0.0001	↓	0.02	↓	0.04
TIMP1	0.07		0.03	↓	0.79		0.87
RUNX2	<0.0001	↑	0.0002	↑	0.13		0.20
COL1A1	0.0001	↑	0.02	↓	0.25		0.78
COL10A1	n/a		n/a		n/a		n/a

Patients 199 and 200, but not patient 201, showed a significant difference in expression in *ACAN* (P = 0.03, 0.0002 respectively), *MMP1* (P < 0.0001 for both), *RUNX2* (P < 0.0001, 0.0002 respectively), and *COL1A1* (P = 0.0001, 0.02 respectively), however the difference in *COL1A1* expression was inconsistent, showing an increase in expression after *GNL3* knockdown in patient 199 and a decrease in expression in patient 200 (Figures 7.11, 7.12). None of the three patients expressed either *COL2A1* or *COL10A1*. Interestingly, as was observed in HACs, the only genes to show a consistent significant difference in expression after knockdown of *GNL3* were *ADAMTS4,* which showed an increase in expression in all three patients ($P \leq 0.0001$), and *ADAMTS5* which showed a decrease in expression in all three patients ($P \leq 0.0001$ patients 199 and 200, P = 0.02 in patient 201; Figures 7.11 - 7.13).

7.4 Discussion

The aim of this chapter was to assess whether the association signal at chromosome 3p marked by a missense polymorphism, rs11177, correlated with expression of *GNL3* using joint tissues from patients who had undergone joint replacement for primary OA.

Mutations that effect GNL3 protein conformation can have an effect on the ability of its coiled coil domain to bind to MDM2, and mutations in the GTP-binding region of GNL3, which is responsible for its localisation in the nucleolus, causes it to be redistributed to the nucleoplasm (238), where is exists in a protein complex with p53, inhibiting its effects (235). The rs11177 SNP confers an Arg-Gln substitution at position 39 in the amino acid sequence, however a database search of the effect the amino acid change has on GNL3, and modelling of the GTP-binding region suggested that there was no effect on the protein and therefore should have no effect on its ability to interact with MDM2 or p53.

The results of AEI analysis demonstrated that there is consistently reduced expression of *GNL3* for the risk allele (A) indicating that *GNL3* is subject to a *cis*-acting eQTL mediated by rs11177 or a SNP in LD with it. Blood cDNA from three patients without any evidence of OA were also studied and showed no AEI however all the joint tissues studied did show AEI, implying that this may be a joint specific effect. In addition to this the NOF cartilage sample assessed did demonstrate AEI, and there was no difference in expression of *GNL3* between cartilage from OA and NOF patients, implying that this may not be OA-specific.

The role of *GNL3* in cellular homeostasis could make it important in OA pathogenesis. Maintaining a fine balance of *GNL3* levels is essential for cellular homeostasis and when *GNL3* levels become imbalanced p53 can be activated to halt cell growth and proliferation (238). If the number of stem cells and their decision to differentiate is disrupted either during embryonic development or in adults, premature senescence can occur (232). Therefore, if there is reduced expression of *GNL3*, as with the OA risk allele, then this could have an impact on the number of MSCs and subsequently disrupt differentiation into chondrocytes.

Of the 12 genes that were assessed for difference in expression following siRNA-mediated knockdown of *GNL3*, *ADAMTS4* and *5* demonstrated consistent results for all human articular chondrocyte (HAC) and mesenchymal stem cell (MSC) patients. Expression of *ADAMTS4* was significantly increased in all patients and expression of

ADAMTS5 was significantly decreased. ADAMTS 4 and 5 are members of the disintegrin and metalloproteinase with thrombospondin motifs family of aggrecanases, which have an important role in cartilage destruction in OA (244, 245). *ADAMTS4* has been shown to be overexpressed both at the mRNA level in OA chondrocytes and at the protein level by immunohistochemistry in OA cartilage (246), supporting the results demonstrated in this chapter, namely, the risk allele of rs11177 (A) correlates with decreased expressed of *GNL3* and when *GNL3* is knocked down in both chondrocytes and MSCs *ADAMTS4* expression is increased. However, *ADAMTS5* expression was significantly decreased after *GNL3* knockdown, and *ADAMTS5*-siRNA knockdown has been demonstrated to have a protective effect on cartilage degradation in human chondrocytes after aggrecan loss was significantly reduced (247). It is possible that the *ADAMTS5* expression levels observed in the knockdown experiments in this chapter occur as a response to increased *ADAMTS4* expression; Song *et al.* stated that the activity of *ADAMTS5* may make aggrecan more susceptible to cleavage by *ADAMTS4* (247), so *ADAMTS5* levels may be reduced in attempt to attenuate aggrecan degradation. Due to time constraints, the protein levels of ADAMTS4 and 5 were not assessed following *GNL3* knockdown. In the future it would be beneficial to carry out western blotting to ascertain whether the significant difference in mRNA expression levels of *ADAMTS4* and *5* was also demonstrated at the protein level, and to see if there is any affect on aggrecan levels. In addition, it would be worthwhile to differentiate the MSCs into chondrocytes following knockdown of *GNL3,* in order to see what effect reduced *GNL3* may have on the cells ability to form cartilage. Unfortunately this was not possible due to the short amount of time remaining.

As already mentioned, *GNL3* overexpression also has effects on the cell cycle (237, 238) so the next logical step would be to overexpress *GNL3* in HACs and differentiated MSCs and measure the expression of the same panel of 12 genes analysed for *GNL3* knockdown, to observe the effects of *GNL3* overexpression.

In conclusion, the data reported in this chapter demonstrates that *GNL3* is subject to a *cis*-acting eQTL potentially mediated by rs11177 and that *GNL3* has a role in both chondrocytes and MSCs. There are two other genes at the 3p locus, *GLT8D1* harbouring the most significant signal (rs6976, $P = 7.24x10^{-11}$) and a signal in *TP63*, (rs12107036, $P = 6.71x10^{-8}$). This chapter has highlighted the need for further molecular analysis of these genes at the 3p locus to identify the variant responsible for

the observed AEI, and for further functional analysis of *GNL3* in order to elucidate a possible role in the pathogenesis of OA.

Chapter 8: General Discussion

As more OA susceptibility loci are being identified it is becoming apparent that the majority of associated alleles are in non-coding regions and therefore contribute to risk by influencing gene expression (248). Recently, genetic research has shifted away from investigating protein structure altering polymorphisms, and most studies now focus on polymorphisms that regulate expression of genes, an excellent and well-characterised example of which is rs143383, a T-C transition in the 5´UTR of *GDF5* (249). AEI analysis of rs143383 revealed that the OA-associated T allele showed lower *GDF5* expression in all joint tissues (250, 251) and four *trans*-acting factors (Sp1, Sp3, P15 and DEAF-1) were later found to bind to the alleles of rs143383 and decrease expression of *GDF5* (252).

In this study, the expression patterns of genes at OA-associated loci have been analysed in joint tissues in order to identify potential eQTLs and regulatory variants.

8.1 Key Results

Chapter 3: Gene expression analysis of the osteoarthritis susceptibility locus mapping to chromosome 7q22

The aim of chapter 3 was to use expression analysis to assess whether the six genes at the chromosome 7q22 OA susceptibility locus are subject to *cis*-acting regulatory polymorphisms that could contribute to the association signal. The key results of the expression analysis were:

Five of the six genes at the locus, *PRKAR2B, HBP1, COG5, DUS4L* and *BCAP29,* are expressed at a significantly lower level in OA cartilage compared to control cartilage.

Reduced *HBP1* expression in cartilage correlates significantly with carriage of the OA associated alleles, particularly in knee cartilage, the joint most relevant to the association signal.

Reduced *HBP1* expression in OA synovium correlates with carriage of the OA associated allele.

These observed correlations were also seen at the allelic level.

Together the results from chapter 3 indicate that among the six genes at the locus, *HBP1* can be prioritised as the potential target of the OA susceptibility at chromosome 7q22, but there is still a need for further study to elucidate the causal variant.

One key observation from the arcOGEN study was that OA associations are often to a particular joint rather than to several different joints therefore stratification by joint is critical in the discovery of association signals and is essential to avoid false negative associations (119). The results from chapter 3 emphasise this, as the significant difference in *HBP1* expression was particularly pronounced in knee cartilage.

Chapter 4: Identification and analysis of a *SMAD3 cis*-acting eQTL operating in primary osteoarthritis

The aim of this chapter was to assess whether carriage of the alleles of the OA associated *SMAD3* SNP correlated with differences in expression. The key results from this chapter were:

There was no evidence of a correlation between the OA-associated SNP and *SMAD3* expression, implying that this is not the mechanism through which the association is operating.

Consistently lower allelic expression was seen for the OA-associated G allele when AEI was studied using a proxy SNP, indicating that *SMAD3* was subject to a *cis*-acting eQTL.

Two 3′UTR SNPs were identified as potential mediators of this eQTL.

This complex *cis*-regulation of *SMAD3* expression is also seen in the *GDF5* promoter where the transcriptional effect on rs143383 is dependent on a second SNP, rs143384, with decreased expression of the T allele of rs143383 only observed in individuals that are compound heterozygous for both SNPs (251). The A allele of a third SNP -41bp from the transcription start site, can neutralise the effects mediated by rs143383 (253). The results observed for *SMAD3,* like *GDF5,* indicate that expression of each allele is highly context specific and that effects on gene expression mediated by *cis*-acting regulatory sites are potentially modifiable by targeting other *cis* sites within the gene (253).

SNPs in the 3′UTR, such as the two functional SNPs identified in chapter 4, can affect microRNA (miRNA) binding sites. The study of miRNA in OA is an expanding field that has produced many interesting results, such as the two miRNAs,

miR-199a and miR-101.3, which regulate the *COX2* gene (254), and miR-101, which regulates *SOX9* (255). In particular, miR-140, which targets *SMAD3* (256), but was shown not to bind near the functional SNPs identified in chapter 4, has been identified as a critical miRNA in OA. This is due to its specific expression in cartilage, its important role in cartilage development, the lower levels observed in OA and the accelerated OA development in miR-140 knockout mice (257, 258). A recent study found opposing results, that levels of miR-140 in OA are increased (259). The difference in miR-140 levels in OA cartilage is unexplained but Swingler *et al.* used femoral heads and all other studies have used knees (260-262), once again highlighting the importance of stratification by joint in studies.

The OA-associated SNP in *SMAD3* did not show a correlation with expression in adult joint tissues but may show an effect in younger tissues. This highlights that future studies should not be confined to aged tissue samples but should also incorporate developmental tissues such as mesenchymal stem cells to identify possible eQTLs that are only operating during development.

Chapter 5: A *SMAD3 cis*-acting eQTL identified in osteoarthritis is also operating in aneurysms and osteoarthritis syndrome

In chapter 5 it was the aim to investigate the 3′UTR SNPs identified in chapter 4 as potential mediators of the *cis*-acting eQTL in OA joint tissues. The key result of this chapter was that the *SMAD3* eQTL is also operating in VSMCs derived from the aortic tissue of an AOS patient. Despite the small number of patient samples studied in this chapter, AEI was detected in VSMCs and is therefore likely to be of direct relevance to this syndrome.
Performing a comparison of the inheritance of the *SMAD3* eQTL identified in chapter 4 between AOS patients may establish whether the eQTL accounts for some of the phenotypic variability observed in AOS.

Chapter 6: Analysis of *DOT1L*, a gene associated with hip OA.

As previously mentioned in the introduction, heterogeneity in case ascertainment can make association studies difficult. The study of endophenotypes, which are measurable intermediate phenotypes generally closer to the role of the gene product than the disease status (263), often provide much greater power to localise and identify disease-related eQTLs than disease status alone and are increasingly being used as endpoints for analyses to uncover OA susceptibility (129). The aim of this chapter was to assess whether expression of *DOT1L*, identified in a study of endophenotypes specifically analysing joint-space narrowing in OA patients (132), and found to be associated with OA at genome-wide significance in a subsequent study (113), correlated with genotype at the OA associated SNP in OA joint tissues. The key result of this chapter was that there was no evidence of a correlation between the OA-associated SNP and expression of *DOT1L*.

Chapter 7: Genetic and functional analysis of *GNL3,* a gene at an OA susceptibility locus on chromosome 3p.

The aim of this chapter was to assess whether the association signal at chromosome 3p marked by a missense polymorphism, correlated with expression of *GNL3*. Following this, *GNL3* was knocked down in human cells and expression levels of a panel of genes were tested to see what affect the reduction in *GNL3* had. The key results of this chapter were:

AEI analysis demonstrated a consistently reduced expression of *GNL3* for the risk allele indicating that *GNL3* is subject to a *cis*-acting eQTL mediated by the OA-associated SNP, or a SNP in LD with it.

Of the 12 genes that were assessed for difference in expression following siRNA- mediated knockdown of *GNL3*, *ADAMTS4* demonstrated a consistent significant increase in expression for all patients, and *ADAMTS5* demonstrated a consistent significant decrease in expression for all patients.

The data reported in this chapter demonstrates that *GNL3* is subject to a *cis*-acting eQTL, correlating with genotype at the OA-associated SNP, and that *GNL3* has a role in both chondrocytes and MSCs.

There has been much interest in the role of *ADAMTS4* and *5* as therapeutic targets in OA (264), as the ADAMTSs are responsible for most of the aggrecan degradation in cartilage (265) and are 1000 times more efficient at cleaving aggrecan than any other enzyme (Gilbert 2007). Mice whose specific aggrecanase cleavage site in aggrecan was mutated to resist ADAMTS5 activity demonstrated diminished aggrecan loss and cartilage erosion in mice, and cartilage repair appeared to be stimulated (266). This has implications for ADAMTS5 as a therapeutic target for OA treatment however, due to the differences in expression and regulation of *ADAMTS5* in humans compared to mice, blocking ADAMTS5 action may not have the same protective effect in humans. A better understanding of human *ADAMTS4* and *5* regulation is required in order to further investigate their potential therapeutic use, and the effect *GNL3* knockdown has on *ADAMTS4* and *5* expression may give some insight into elucidating their regulation.

AEI at the OA-associated SNP consistently demonstrated reduced expression for the risk allele, however overall *GNL3* expression showed no correlation with carriage of this allele. This effect has also been demonstrated for *COL11A1* (267), and highlights that overall expression studies may not be sensitive enough to pick up slight differences in expression between the two alleles and should not be solely relied upon to indicate the presence of an eQTL.

Due to time constraints, western blotting to determine the protein levels of ADAMTS4 and 5 were not performed following *GNL3* knockdown to confirm that the expression differences were also observed at the protein level.

8.2 Future work

Overall, the data reported in chapter 3 clearly demonstrates the need for a comprehensive molecular analysis of potential *cis*-acting regulatory elements of *HBP1* in order to identify the actual functional variant(s) and to elucidate their haplotypic relationship to both the OA-associated SNPs. The fact that the results of the AEI analysis show an incomplete correlation between genotype suggests that there may be multiple *cis*-acting regulatory polymorphisms interacting to modulate the expression of *HBP1*. In the future, SNPs in *HBP1* in high LD with the OA-associated SNPs should be

subject to haplotype analysis, as was carried out for *SMAD3* SNPs in chapter 4, in order to identify the causal variant.

GPR22 was not expressed in any of the OA joint tissues investigated in chapter 3. This may be because it is expressed earlier in development and so it would be beneficial to assess whether there is a correlation between the OA-associated alleles and *GPR22* expression in a developmental tissue such as MSCs.
This approach would also be useful for further investigation of *SMAD3* and *DOT1L*. Since there was no evidence of a correlation between the OA-associated alleles and *SMAD3* or *DOT1L* expression in adult OA joint tissues it is possible that the associations are operating during development.

In addition to ascertaining whether the significant difference in mRNA expression levels of *ADAMTS4* and *5* was also demonstrated at the protein level, it would be informative to see if there is any effect on aggrecan protein levels. It would also be worthwhile to differentiate the MSCs into chondrocytes following knockdown of *GNL3,* in order to see what effect reduced *GNL3* may have on the cells ability to form cartilage.

GNL3 overexpression also has effects on the cell cycle (237, 238) so the next logical step would be to overexpress *GNL3* in HACs and differentiated MSCs and measure the expression of the same panel of 12 genes analysed for *GNL3* knockdown, to observe the effects of *GNL3* overexpression.

For both *HBP1* and *GNL3* further investigation is required to identify which of the potential SNPs is causing the observed AEI at these loci. The next step would be to identify SNPs in perfect LD with the *HBP1* and *GNL3* SNPs that could be regulating the eQTLs and carry out functional investigations such as were used in the investigation of the *SMAD3* eQTL (Chapter 4). If there are many potential SNPs the search can be narrowed to likely candidates with databases such as ENCODE and RegulomeDB, or if like 7q22 there is a region of high LD, trans-ethnic fine-mapping can be utilised to narrow the search for functional variants. Patterns of LD vary across ancestry groups due to divergent evolutionary and migratory histories (268). This difference in LD structure between ethnic groups can be exploited to fine-map functional variants. For example, in a study of cholesterol levels, six variants in high LD with one another in European and Asian cohorts were narrowed to one functional variant after an African American cohort was found to have no other SNPs in high LD with it (268).

Standard methods to investigate functionality of a SNP, such as luciferase reporter assays and electrophoretic mobility shift assays (EMSAs), as used to identify trans-acting factors for *GDF5* (252), only give us an idea of how the polymorphisms behave *in vitro*. As mentioned previously, with regard to polymorphisms in the *SMAD3* 3′UTR and the *GDF5* promoter, the cellular context of a polymorphism is paramount. This is demonstrated by the *SMAD3* luciferase reporter assays, which revealed that the 3′UTR SNPs behave differently in the two immortalised cell lines analysed. The next challenge in genetic research, after confirming functionality of variants *in vitro,* is performing allele editing on the whole genome to see what a variants actual effect on a human somatic cell is. This introduction of targeted genome sequence changes into living cells and organisms is a potential avenue for therapy of genetic diseases (269). One way to achieve genome editing is by using nucleases, composed of a customisable sequence-specific DNA-binding domain fused to a nuclease that cleaves DNA inducing double strand breaks (DSBs) into specific DNA sites. The DSBs are repaired by mechanisms that can be exploited to create sequence alterations at the cleavage site (270). Normal cellular repair of DSBs by non-homologous end joining (NHEJ) or homology-directed repair (HDR) can be exploited to introduce several types of genomic alterations including point mutations, deletions, insertions, duplications and translocations (270). NHEJ repair rapidly and efficiently ligates the two broken ends, but because it is an error prone pathway that often results in small insertions or deletions, it can be harnessed to knock out genes by creating a DSB in a coding exon and hence disrupting the open reading frame (271, 272). HDR is a more accurate repair pathway, and by providing donor DNA that shares homologous sequence to the target gene, the cellular repair machinery will use this template for HDR, permitting the transfer of sequence information from the donor to the gene of interest (272). The vast majority of targeted genome editing has been performed using zinc finger nucleases (ZFNs) but transcription activator-like effector nucleases (TALENs) have rapidly emerged as an alternative to ZFNs (270). TALENs have been successfully used to induce gene editing in C-C chemokine receptor 5 (*CCR5),* used by HIV to infect T-cells, in a mammalian cell line (273). TALENs have also been successfully transfected into human CD4+ T cells to edit IL2Rγ mutations, which cause severe combined immunodeficiency (SCID; 274). However the use of TALENS is limited by the need to engineer specific proteins for each target site (275). A more versatile system for genome editing is clustered, regularly interspaced, short palindromic repeats (CRISPRs) and CRISPR-associated (Cas) nucleases, which are essential components of nucleic acid-

based adaptive immune systems that are widespread in bacteria (275, 276). CRISPR systems consist of a series of short repeats separated by sequences that are incorporated from invading DNA elements. These CRISPR loci are transcribed and the transcripts are processed into a library of small CRISPR RNAS (crRNAs), each containing a guide sequence complementary to a fragment of incorporated foreign DNA (269, 276). Each crRNA hybridises with a second RNA, known as the trans-activating crRNA (tracrRNA), and these two RNAs complex with the one or more Cas proteins. The resulting ribonucleoprotein complex patrols the intracellular environment for targets that are complementary to the crRNA-guide sequence (276). In the simplest of these systems the ribonucleoprotein complex contains only one Cas protein, Cas9 (269), and the DNA-cleaving activity of CRISPR-Cas9 is programmed with a single guide RNA (sgRNA) that recognises specific DNA sequences (275). CRISPR RNA-guided nucleases are easier to use than TALENs or ZFNs as they rely on Watson-Crick base pairing therefore eliminating the need for protein engineering (276).

These approaches can be used to create entire human cell lines bearing specific single-nucleotide changes identified by large scale GWAS, ENCODE or other sequencing projects, potentially enabling the study of gene function and/or the modelling of disease causing mutations through the creation of a point mutation that is characteristic, for example, of a known disease predisposing allele such as the ones investigated here (270, 271). Additionally, with the use of CRISPRs, multiple sgRNAs can be introduced simultaneously in order to edit multiple genes (275). This has been demonstrated successfully in cynomolgus monkeys, where *Pparγ* and *Rag1* were simultaneous disrupted (277). This technique could be particularly useful for OA where the disease is caused by the carriage of multiple risk alleles in multiple genes.

8.3 Summary

Using a number of expression techniques, this study has:

Identified *HBP1* from six genes at the 7q22 locus as being the most likely gene harbouring the association.

Identified *cis*-acting eQTLs in *SMAD3* that are active in joint tissues and vascular cells.

Discovered a role for *GNL3* in chondrocytes and mesenchymal stem cells and a potential role in the regulation of *ADAMTS4* and *5*.

Despite the relative success of GWAS, there is a gap between the many associated sequence variants and our understanding of how most of these variants contribute to complex trait biology (278).

This is due in part to three main issues complicating the functional translation of GWAS signals:

1) SNPs, either individually or in combination explain only a small fraction of the genetic variance of most complex traits and the phenotypic effects caused by the trait-associated SNPs are subtle meaning that large numbers of independent studies are required to estimate the impact of each individual genetic variant (278); 2) Many signals identified are in loci containing a number of genes, often in a region of high LD, such as the 7q22 OA locus. This means that the signal could be marking one of many polymorphisms in any of the genes at the locus, making identification of the causal variant very difficult; 3) 74-95% of GWAS hits are in non-coding regions (279), preventing the direct interpretation of their functional effects (278).

Therefore the variants are likely to be involved in regulation of gene expression, which is dependent on cell type, developmental stage and environmental factors (278). SNPs that are associated with gene expression in disease-relevant tissues are more likely to be functionally relevant and the genes containing these SNPs are good candidates for the underlying pathological process (280, 281).

In addition to the issues surrounding further investigation of GWAS signals, associated loci are typically characterised by many common SNPs each of modest effect, which when considered together account for only a small proportion of the genetic variance of the trait (282). Low frequency and rare genetic variants not captured by GWAS may explain a substantial fraction of this missing genetic component of complex traits (278). Some of the missing heritability has also been attributed to imperfect tagging, for example, a weak GWAS signal might be indicating a variant that is much more strongly associated but which lies at some genomic distance (283).

Missing heritability, which is reflected in the small odds ratios and limited predictive value of these variants has raised questions about the ultimate applicability of GWAS findings to risk prediction and to potential therapies (283).

The ultimate goal of performing genetic and genomic studies of any human disease is to identify and understand its pathogenesis (284). Yet despite the identification of over 2000 associations for over 300 complex traits and diseases, only a negligible fraction of discovered GWAS loci have been followed up in experimental studies (283). There needs to be an increase in the number of investigations into the molecular mechanisms of confirmed associations (278), and data from such studies will improve our interpretation of GWAS findings and their relation to specific cell populations (284). The technique of combining eQTL data with GWAS data significantly facilitates identification of potential causal variants (285), and has been applied to studies of other complex diseases, for example, autism in which expression of the *SEMA5A* gene was shown to correlate with disease-associated SNPs (286). Studies such as this have demonstrated the practicality and efficiency of eQTLs as a screening tool for candidate regulatory variants, and use of this expression information should move us closer to developing a more comprehensive understanding of the genetic basis for complex traits (287), and these new insights about the pathogenesis of complex diseases will help in the identification of novel therapeutic strategies to treat human disease (284).

As discussed in section 8.2 there is a need for the SNPs identified in this thesis as modulating gene expression levels, to be further investigated using genome-editing techniques. For such techniques, specificity and delivery is paramount (278). The desired genome modification should occur with high frequency in the cell population, with no detectable off-target effects (278). Light inducible transcriptional effectors (LITEs) offer exciting possibilities for studying the function and regulation of mammalian genomes (288). LITEs consist of the light sensitive photoreceptor cryptochrome 2 (CRY2) fused to a customisable DNA-binding domain i.e TALENs or CRISPR/Cas9 systems. Illumination with blue light triggers a conformational change in CRY2 and subsequently recruits an effector to the target locus to mediate transcriptional modulation (288). Importantly, the construct allows spatially and temporally precise, reversible and non-invasive modulation of gene transcription (278). Multiple implicated GWAS signals, such as the many OA-associated SNPs, can potentially be disrupted in combination to examine their interaction or synergistic effects (278).

Overall, this thesis has demonstrated that OA tissues are subject to a range of regulatory elements that can influence disease, and that by combining GWAS and expression data the search for causal variants can be narrowed.

Chapter 9: Appendix

9.1 Tables

Table A1: The 201 patients used in this study.

Patient	Sex	Age at surgery (years)	Joint replaced at surgery	Tissue extracted
1	F	75	H	Cn
2	F	67	K	Cn
3	M	69	K	Cn
4	F	77	H	Cn
5	F	73	K	Cn
6	M	71	K	Cn
7	F	73	K	Cn
8	F	64	H	Cn
9	F	80	H	Cn
10	F	51	H	Cn
11	M	71	H	Cn
12	F	56	H	Cn
13	F	67	H	Cn
14	F	83	H	Cn
15	F	60	K	Cn
16	F	67	K	Cn
17	F	78	H	Cn
18	M	62	K	Cn
19	F	72	K	Cn
20	M	74	K	Cn
21	F	83	H	Cn
22	M	50	K	Cn
23	F	76	H	Cn
24	M	79	K	Cn
25	F	60	H	Cn
26	M	67	K	Cn
27	M	85	H	Cn
28	M	79	K	Sy
29	M	76	K	Cn
30	F	82	H	Cn
31	F	71	K	Cn, Sy
32	F	72	H	Cn
33	F	42	K	Cn
34	F	72	H	Cn
35	F	67	H	Cn
36	F	70	H	Cn
37	F	70	H	Cn
38	F	60	H	Cn
39	F	74	H	Cn
40	M	53	K	Fp, Sy
41	F	50	H	Cn
42	M	63	K	Cn, Fp
43	F	78	H	Cn
44	F	81	K	Cn

45	F	50	K	Cn
46	M	76	K	Cn
47	M	69	K	Cn, Fp
48	M	76	K	Cn, Fp
49	M	82	K	Cn, Fp
50	F	55	K	Cn, Sy
51	F	68	K	Fp
52	M	75	K	Sy
53	M	63	K	Fp
54	F	81	K	Cn, Sy
55	F	68	K	Cn
56	F	65	K	Fp
57	F	63	K	Cn
58	M	98	K	Fp
59	M	75	K	Sy
60	F	79	K	Me
61	F	69	K	Me
62	F	74	H	Fp, Li
63	M	57	K	Cn
64	M	85	K	Fp
65	F	64	K	Fp, Sy, Me
66	M	57	H	Cn, Te
67	F	78	K	Fp, Sy, Li
68	F	65	K	Sy, Me
69	M	72	K	Cn
70	M	65	H	Li
71	M	63	K	Fp
72	M	69	K	Cn, Sy
73	F	72	K	Sy
74	M	74	K	Cn, Fp, Sy
75	F	67	K	Cn
76	F	69	K	Cn, Fp
77	F	52	H	Cn
78	F	64	K	Sy
79	M	56	K	Sy
80	M	57	K	Cn
81	F	61	H	Cn
82	F	60	K	Cn, Fp
83	F	77	K	Cn
84	F	66	K	Cn
85	M	63	K	Cn
86	M	77	K	Cn
87	F	71	K	Fp
88	M	82	K	Cn
89	F	78	K	Cn, Sy
90	M	82	K	Cn
91	M	56	K	Cn, Sy
92	M	46	K	Cn
93	M	56	K	Cn, Sy
94	F	54	K	Cn, Sy
95	F	62	K	Fp
96	M	71	K	Cn
97	F	58	H	Cn
98	F	69	H	Cn
99	M	63	K	Cn
100	F	71	H	Cn
101	M	70	K	Cn

102	M	67	K	Cn, Sy
103	F	80	K	Cn
104	F	67	K	Cn, Fp
105	M	71	K	Fp
106	F	46	K	Cn
107	F	62	K	Cn
108	F	58	K	Cn
109	F	59	K	Cn
110	M	64	K	Cn
111	F	81	K	Cn
112	F	80	K	Cn
113	F	64	K	Cn
114	F	78	K	Cn
115	F	61	K	Cn
116	F	82	K	Cn
117	F	80	K	Cn
118	F	59	K	Cn
119	F	71	H	Cn
120	M	74	K	Cn
121	F	74	H	Cn
122	M	72	K	Cn
123	M	72	K	Cn
124	M	68	H	Cn
125	F	72	H	Cn
126	M	69	K	Fp
127	M	68	K	Fp
128	M	58	K	Fp, Li
129	M	75	K	Cn, Fp, Sy
130	F	60	K	Fp
131	M	76	H	Cn
132	M	71	K	Fp
133	F	73	K	Sy
134	M	66	K	Fp
135	F	58	K	Fp
136	F	69	K	Sy
137	F	82	K	Fp
138	F	81	K	Fp, Me
139	F	54	H	Li
140	M	68	H	Sy
141	M	72	K	Sy
142	M	85	H	Sy
143	F	72	H	Bn
144	F	70	K	Fp, Li
145	F	62	K	Fp
146	M	62	K	Fp
147	M	69	H	Sy
148	F	67	H	Sy
149	F	88	K	Fp
150	M	58	K	Fp
151	F	82	K	Fp, Sy
152	F	62	H	Cn
153	F	76	H	Cn
154	F	76	H	Cn
155	F	55	H	Cn
156	F	76	H	Cn
157	F	89	K	Fp
158	F	50	K	Fp

159	M	79	H	Cn
160	F	72	H	Cn
161	M	71	K	Ca
162	F	86	K	Fp
163	M	59	K	Me
164	F	71	K	Fp
165	F	64	K	Cn
166	M	86	K	Cn
167	F	85	H	Cn
168	M	83	H	Cn
169	F	69	H	Cn
170	F	71	H	Cn
171	F	69	H	Cn
172	F	81	H	Cn
173	F	72	H	Cn
174	F	84	H	Cn
175	M	78	H	Cn
176	F	94	H	Cn
177	F	84	H	Cn
178	F	84	H	Cn
179	F	68	H	Cn
180	F	52	H	Cn
181	F	80	H	Cn
182	F	86	H	Cn
183	F	52	H	Cn
184	F	73	H	Cn
185	F	89	H	Cn
186	F	82	H	Cn
187	F	82	H	Cn
188	F	91	H	Cn
189	M	79	H	Cn
190	M	84	H	Cn
191	F	82	H	Cn
192	F	91	H	Cn
193	F	42	/	Blood
194	F	25	/	Blood
195	F	28	/	Blood
196	F	68	K	HAC
197	M	54	K	HAC
198	M	65	K	HAC
199	F	22	/	MSC
200	M	26	/	MSC
201	F	24	/	MSC

F, female; M, male; K, knee; H, hip; Cn, articular cartilage; Fp, infrapatellar fat pad; Sy, synovium; Li, ligament; Me, meniscus; Bn, bone; HAC, Human articular chondrocytes; MSC, mesenchymal stem cells.

Table A2: Primers used to amplify cDNA for standard qualitative gene expression analysis. n/a denotes primers that do not amplify DNA fragments.

Gene	Forward primer 5′- 3′	Reverse primer 5′ - 3′	$MgCl_2$	Annealing temperature	Fragment size	
					cDNA	DNA
PRKAR2B	GATTCACAAGGCGTGCCTCAG	CCTTGATCAATTACATGCTCCC	2 mM	60°C	232bp	n/a
HBP1	AATACCCAAGATCATCTTGG	GAAGGCTGGTTCACTCTTCG	2 mM	60°C	152bp	570bp
COG5	GGAGATGCAAGTCAGGTGATTGG	ACAGAAGTGAGTAAGGGTTGCAC	2 mM	55°C	149bp	n/a
GPR22	GCATGCAGTTGAAGTATTGG	GAGACACTTGAAAGCTTAACGG	2 mM	60°C	238bp	n/a
DUS4L	AGCCAGTGCACTATGATTCC	AACACATGTAAGGAGTCCCG	2 mM	55°C	249bp	n/a
BCAP29	GTCTGGGGTAAAATTGCAAC	GTTTCATCTGTGTGTGTTCATAG	2 mM	55°C	172bp	n/a
HPRT1	CTGAACGTCTTGCTCGAGATG	TGCGACCTTGACCASTCTTTGG	2 mM	60°C	247bp	n/a

Table A3: Real time PCR primer and probe sequences for quantitative gene expression analysis.

Gene	Forward primer 5′-3′	Reverse primer 5′-3′	Probe 5′-3′
HPRT1	TGCTGAGGATTTGGAAAGGG	ACAGAGGGCTACAATGTGATG	AGGACTGAACGTCTTGCTCGAGATG
GAPDH	ACATCGCTCAGACACCATG	TGTAGTTGAGGTCAATGAAGGG	AAGGTCGGAGTCAACGGATTTGGTC
18S	CGAATGGCTCATTAAATCAGTTATGG	TATTAGCTCTAGAATTACCACAGTTATCC	TCCTTTGGTCGCTCGCTCCTCTCCC
PRKAR2B	TGATCAAGGTGACGATGGTG	TGTACATTAAGGCCAGTTCGC	TGTGATGGTGTTGGAAGATGTGTTGGT
HBP1	ACACCAGTAAGACACGAAAGG	ACAGTGCCAGACAGTTGAAG	CCAGGAATTGCACCATCCCAAATCA
COG5	TGGGATTGAGTCGTTGGAAG	CTGAAGTCTTGCTAGTTGTGC	CCCTGTAAAGCCCCAATTCTCGTCT
GPR22	GAAATCCATCAGCCAACTAAATCTC	CACATTCTTTTGGAGCAGTTGG	TCCCTGACAACAGTGAAGTAGCTGTG
DUS4L	TGGTAAAAGTCTGTGCCCC	TGGTAAATTCGCTGTCTCTGG	ACCAATGATTGTTGCCGCTGATTTTGT
BCAP29	TGGAACAAGGCTTTCCTTACC	TGTGTTCATAGGCATCAGGTC	TCATACCATTGAGAAGAGCTCCACCAG
SMAD3	ACCAGTTGACCCGAATGTG	CTTGTCAAGCCACTGCAAAG	TCTGTCTCCTGTACTCCGCTCCC
DOT1L	CCAAGAAGAACCAAACTGCAC	TCGGAGGTAGCTGGTAGAAC	CCTCACCCCAGGATGCCTACAGA
GNL3	AGCAGAAACTTGACAGGCAG	GAATTCTGTTTGCCCGACTTG	CCCAAACTCCTTTTCCATAGGTTCCACA
ACAN	TGTGGGACTGAAGTTCTTGG	AGCGAGTTGTCATGGTCTG	CTGGGTTTTCGTGACTCTGAGGGT
COL2A1	AACCAGATTGAGAGCATCCG	ACCTTCATGGCGTCCAAG	AGACCTGAAACTCTGCCACCCTG
SOX9	ACTTGCACAACGCCGAG	CTGGTACTTGTAATCCGGGTG	TCTGGAGACTTCTGAACGAGAGCGA
MMP1	AAGATGAAAGGTGGACCAACAATT	CCAAGAGAATGGCCGAGTTC	CAGAGAGTACAACTTACATCGTGTTGCGGCTC
MMP9	CGGCACTGAGGAATGATCTAAG	CGAACTTTGACAGCGACAAG	CCGCCACGAGGAACAAACTGTATCC
MMP13	AAATTATGGAGGAGATGCCCATT	TCCTTGGAGTGGTCAAGACCTAA	CTACAACTTGTTTCTTGTTGCTGCGCATGA
ADAMTS4	CTGGGTATGGCTGATGTGG	TGGCTTGGAGTTGTCATGG	TTCACTGCTGCTCATGAACTGGGT
ADAMTS5	CAAGTGCGGAGTATGTGGAG	GTCTTTGGCTTTGAACTGTCG	TTTATGTGGGTTGCCCCTTCAGGA
TIMP1	TTCTGCAATTCCGACCTCG	TCATAACGCTGGTATAAGGTGG	TTGACTTCTGGTGTCCCCACGAAC
RUNX2	AGCAAGGTTCAACGATCTGAG	TGAAGACGGTTATGGTCAAGG	CGGAGTGGACGAGGCAAGAGTTTC
COL1A1	CCCCTGGAAAGAATGGAGATG	TCCAAACCACTGAAACCTCTG	TTCCGGGCAATCCTCGAGCA
COL10A1	CAAGGCACCATCTCCAGG	TGGGCATTTGGTATCGTTCAG	ACTCCCAGCACGCAGAATCCAT

Table A4: Primers used to amplify fragments for RFLP.

Gene	SNP		Primers 5´ - 3´	$MgCl_2$	Annealing temperature	Fragment size
HBP1	rs7794598	F R	GCCTTCCCTCATCACCATTGGAAGGAGGAAACACCAGT GTGAGACTACTTCCTATGAGG	2 mM	55°C	136bp
COG5	rs3815148	F R	TGACCTGTTAGTTGGCATC TGTCTGAGGTAATTACAGG	2mM	59°C	308bp
DUS4L	rs711442	F R	GACTAGACTTCCCAAATAATTTTAATGTA GATCAGGATGTTTGAACCACTGGG	2 mM	55°C	225bp
DUS4L	rs4730250	F R	CAGCTGGTAAAAGTCTGTGCC CTTAGGACCCATAATTGCACTC	2 mM	55°C	378bp
SMAD3	rs12901499	F R	GAAAGCCACCACTCTATCGTTTC ATGGTGTTTTGTTCACTGTCAGGAGAGAAGAC	2 mM	55°C	185bp
SMAD3	rs8031440	F R	CTGAGTGAGACCCAGCAAC GAGCTAGATCATCTTATATC	2 mM	55°C	255bp
SMAD3	rs1052488	F R	GAGATGATGGGCTAAACAGG CTAGAGTATACATTACTTATGC	3 mM	55°C	141bp
DOT1L	rs12982744	F R	GTAGAACCCAGTGCTTCCGGC CTGCGTGATTTGACTGCGTG	2 mM	55°C	245bp
GNL3	rs11177	F R	CAGTGACTAGTGTTGTTACC GTTTCTCACTTGAATTCTGC	2 mM	60°C	297bp

Table A5: Restriction enzymes, buffers and reaction conditions for RFLP.

Gene	SNP	Restriction endonuclease	NEB buffer	Temperature	BSA	Digested fragments
HBP1	rs7794598	BspHI	4	37°C		94bp/42bp
COG5	rs3815148	MboII	4	37°C		205bp/103bp
DUS4L	rs4730250	DraII	4	37°C		285bp/103bp
DUS4L	rs711442	RsaI	4	37°C		167bp/44bp/28bp
SMAD3	rs12901499	BccI	1	37°C	✓	147bp/38bp
SMAD3	rs8031440	BstNI	2	60°C	✓	130bp/73bp
SMAD3	rs1052488	NsiI	3	37°C		121bp/19bp
DOT1L	rs12982744	MspI	4	37°C		144bp/101bp
GNL3	rs11177	AciI	3	37°C	✓	132bp/165bp

Table A6: The SNPs analysed in this study.

Gene	SNP	Location	Alleles (major/minor)	Allele Frequency*	
				Major	Minor
COG5	rs3815148	Intronic	A/C	0.783	0.217
DUS4L	rs4730250	Intronic	A/G	0.827	0.173
DUS4L	rs711442	3′UTR	C/T	0.531	0.469
HBP1	rs7794598	Exon 7	C/T	0.781	0.219
SMAD3	rs12901499	Intronic	G/A	0.531	0.469
SMAD3	rs8031440	3′UTR	G/A	0.770	0.230
SMAD3	rs1052488	3′UTR	T/C	0.739	0.261
SMAD3	rs8025774	3′UTR	C/T	0.770	0.230
SMAD3	rs8031627	3′UTR	G/A	0.770	0.230
SMAD3	rs2278670	3′UTR	C/T	0.770	0.230
SMAD3	rs12595334	3′UTR	C/T	0.770	0.230
SMAD3	rs3743342	3′UTR	C/T	0.770	0.230
DOT1L	rs12982744	Intronic	C/G	0.625	0.375
GNL3	rs11177	Exon 3	G/A	0.664	0.336

*HapMap CEU

Table A7: Primers used for AEI with the SNaPshot multiplex kit

Gene	SNP	Template	Forward primer 5′-3′	Reverse primer 5′-3′	MgCl2 concentration	Annealing temperature	Fragment size
HBP1	rs7794598	DNA	TCGAAGAGTGAACCAGCCTT	GTGAGACTACTTCCTATGAGG	2 mM	60°C	152bp
HBP1	rs7794598	cDNA	TCGAAGAGTGAACCAGCCTT	GAAGGCCAGGAATTGCACCATCC	2 mM	60°C	152bp
DUS4L	rs711442	DNA and cDNA	GGCCATGTTTGCTGGGATATGAGGAAACC	GATCAGGATGTTTGAACCACTGGG	3 mM	68.9°C	434bp

Table A8: Single base extension primer sequences.

Gene	SNP	Primer sequence 5′-3′
HBP1	rs7794598	GGAGGAAACACCAGTAAGACA
DUS4L	rs711442	CTGTGGTTTCACTGTGGGTCTAAAAGT

Table A9: Primers used to amplify DNA fragments for vector construction including restriction sites (underlined).

Gene	Encompassing SNPs	Forward primer 5′-3′	Reverse primer 5′-3′	Annealing temperature	Fragment size	Restriction enzymes
SMAD3	rs8025774	GGGGACTAGTCTTAAGTGAGCAGAACAGGTAG	GGGGAAGCTTCATCCCAAGTCTATCCAGCTCAC	70°C	310bp	SpeI, HIndIII
SMAD3	rs8031440 rs8031627 rs2278670	GGGGGAGCTCCATTAAATCAACTTTATCATATG	GGGGACGCGTCGTGGGATAGGTCATTTTATAC	65°C	570bp	SacI MluI
SMAD3	rs12595334 rs3743342	GGGGACTAGTGATCTATGGAAGTCGTGTC	GGGGAAGCTTGAAAGGGAATAAAAGTTGGCTC	65°C	700bp	SpeI, HindIII

Table A10: Restriction enzymes used for vector construction.

Gene	Restriction endonuclease	NEB buffer	Temperature	BSA
SMAD3	SpeI	4	37°C	✓
SMAD3	HindIII	2	37°C	
SMAD3	MluI	3	37°C	✓
SMAD3	SacI	1	37°C	✓

Table A11: Primers used to sequence vector constructs and haplotyping.

Gene	Vector		Forward primer 5′-3′		Reverse primer 5′-3′
SMAD3	pMIR report		AGGCGATTAAGTTGGGTA		ATTGCAACGATTTAGGTG
SMAD3	pCR4	M13	GTAAAACGACGGCCAG	T7	TAATACGACTCACTATAGGG

Table A12: Primers used for mutagenesis of *SMAD3*-luciferase vectors. Mutated alleles are shown underlined.

Gene	SNP	Change	Forward primer 5′-3′	Reverse primer 5′-3′
SMAD3	rs12595334	T→C	ATGTGATGAAATGA<u>C</u>ACGTTTGGCTGCAT	ATGCAGCCAAACGT<u>G</u>TCATTTCATCACAT
SMAD3	rs12595334	C→T	ATGTGATGAAATGA<u>T</u>ACGTTTGGCTGCAT	ATGCAGCCAAACGT<u>A</u>TCATTTCATCACAT
SMAD3	rs3743342	T→C	CATGACATCTTCAC<u>C</u>TTGCAGCTTGTGCT	AGCACAAGCTGCAA<u>G</u>GTGAAGATGTCATG
SMAD3	rs3743342	C→T	CATGACATCTTCAC<u>T</u>TTGCAGCTTGTGCT	AGCACAAGCTGCAA<u>A</u>GTGAAGATGTCATG

Table A13: Primers used for haplotyping of *SMAD3* cDNA from aneurysms-OA patients.

Forward primer 5′-3′	Reverse primer 5′-3′	Annealing temperature	$MgCl_2$ concentration	Fragment size
CCTGCAGCCAGTTACCTACTG	GCAGGGCCATGCAGACCTC	70°C	2 mM	1060bp

Table A14: siRNA sequences.

	Sense 5′-3′	Antisense 5′-3′
GNL3 siRNA	AAGCUGUACUGCCAAGAACdTdT	GUUCUUGGCAGUACAGCUUdTdT
Control siRNA	AAGCGCGCTTTGTAGGATTCdTdT	GAAUCCUACAAAGCGCGCUUdTdT

9.2 Abstracts for International Conferences

Dodd AW, Wreglesworth N, **Raine EV**, Gravani A, Loughlin J (2010) Allelic expression analysis of the genes within the chromosome 7q22 OA susceptibility locus reveals evidence for functional polymorphism in COG5. Presented at the 2010 OARSI international congress, Brussels, Belgium.

Riancho JA, García-Ibarbia C, Gravani A, **Raine EVA**, Rodríguez-Fontenla C, Soto-Hermida A, Rego-Perez I, Dodd AW, Gómez-Reino JJ, Zarrabeitia MT, Garcés CM, Carr A, Blanco F, González A, Loughlin J (2010) A multicenter study of the association of aromatase and estrogen receptor genes with hip and knee osteoarthritis. Presented at the 2010 OARSI international congress, Brussels, Belgium.

Raine EV, Reynard LN, Loughlin J (2012) Allelic expression analysis and association with OA of common variants in *SMAD3.* Presented at the 2012 OARSI international congress, Barcelona, Spain.

9.3 Publications

Allelic expression analysis of the osteoarthritis susceptibility locus that maps to chromosome 3p21 reveals cis-acting eQTLS at GNL3 and SPCS1.
Gee F, Clubbs C, **Raine EV**, Reynard LN, Loughlin J.
BMC Med Genet (2014).

Identification and analysis of a *SMAD3 cis*-acting eQTL operating in primary osteoarthritis and in the aneurysms and osteoarthritis syndrome.
Raine EV, Reynard LN, van de Laar IM, Bertoli-Avella AM, Loughlin J.
Osteoarthritis Cartilage. (2014).

Allelic expression analysis of the osteoarthritis susceptibility gene COL11A1 in human joint tissues.
Raine EV, Dodd AW, Reynard LN, Loughlin J.
BMC Musculoskelet Disord (2013).

Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study.
arcOGEN Consortium; arcOGEN Collaborators, Zeggini E, Panoutsopoulou K, Southam L, Rayner NW, Day-Williams AG, Lopes MC, Boraska V, Esko T, Evangelou E, Hoffman A, Houwing-Duistermaat JJ, Ingvarsson T, Jonsdottir I, Jonnson H, Kerkhof HJ, Kloppenburg M, Bos SD, Mangino M, Metrustry S, Slagboom PE, Thorleifsson G, **Raine EV**, Ratnayake M, Ricketts M, Beazley C, Blackburn H, Bumpstead S, Elliott KS, Hunt SE, Potter SC, Shin SY, Yadav VK, Zhai G, Sherburn K, Dixon K, Arden E, Aslam N, Battley PK, Carluke I, Doherty S, Gordon A, Joseph J, Keen R, Koller NC, Mitchell S, O'Neill F, Paling E, Reed MR, Rivadeneira F, Swift D, Walker K, Watkins B, Wheeler M, Birrell F, Ioannidis JP, Meulenbelt I, Metspalu A, Rai A, Salter D, Stefansson K, Stykarsdottir U, Uitterlinden AG, van Meurs JB, Chapman K, Deloukas P, Ollier WE, Wallis GA, Arden N, Carr A, Doherty M, McCaskie A, Willkinson JM, Ralston SH, Valdes AM, Spector TD, Loughlin J.
Lancet (2012).

Gene expression analysis reveals HBP1 as a key target for the osteoarthritis susceptibility locus that maps to chromosome 7q22.
Raine EV, Wreglesworth N, Dodd AW, Reynard LN, Loughlin J.
Ann Rheum Dis (2012).

Increased type II deiodinase protein in OA-affected cartilage and allelic imbalance of OA risk polymorphism rs225014 at DIO2 in human OA joint tissues.
Bos SD, Bovée JV, Duijnisveld BJ, **Raine EV**, van Dalen WJ, Ramos YF, van der Breggen R, Nelissen RG, Slagboom PE, Loughlin J, Meulenbelt I.
Ann Rheum Dis (2012).

Allelic expression analysis of the osteoarthritis susceptibility locus that maps to MICAL3.
Ratnayake M, Reynard LN, **Raine EV**, Santibanez-Koref M, Loughlin J.
BMC Med Genet (2012).

Common variations in estrogen-related genes are associated with severe large joint osteoarthritis: a multicenter genetic and functional study.

Riancho JA, García-Ibarbia C, Gravani A, **Raine EV**, Rodríguez-Fontenla C, Soto-Hermida A, Rego-Perez I, Dodd AW, Gómez-Reino JJ, Zarrabeitia MT, Garcés CM, Carr A, Blanco F, González A, Loughlin J.
Osteoarthritis Cartilage (2010).

Chapter 10: References

1. http://www.arthritisresearchuk.org/arthritis-information/data-and-statistics/osteoarthritis.aspx
2. http://www.hse.gov.uk/statistics/lfs/1112/swit1.htm
3. http://83.244.183.180/100pc/ibsda/icdgp/ccsex/a_carate_r_icdgp_c_ccsex_may12.html
4. http://njrcentre.org.uk
5. Goldring MB. Update on the biology of the chondrocyte and new approaches to treating cartilage diseases. *Best Pract Res Clin Rheumatol.* 2006 **20**(5):1003-1025.
6. Arthritis Research UK, Osteoarthritis in general practice 2013 (http://www.arthritisresearchuk.org/arthritis-information/data-and-statistics/osteoarthritis.aspx)
7. Martel-Pelletier J, Boileau C, Pelletier JP, Roughley PJ. Cartilage in normal and osteoarthritis conditions. *Best Pract Res Clin Rheumatol*. 2008 **22**(2):351-384.
8. Aigner T, McKenna L. Molecular pathology and pathobiology of osteoarthritic cartilage. *Cell Mol Life Sci*. 2002 **59**(1):5-18.
9. Ghosh P, Smith M. Osteoarthritis, genetic and molecular mechanisms. *Biogerontology.* 2002 **3**(1-2):85-88.
10. Herrero-Beaumont G, Roman-Blas JA, Castaneda S, Jimenez SA. Primary osteoarthritis no longer primary: three subsets with distinct etiological, clinical, and therapeutic characteristics. *Semin Arthritis Rheum*. 2009 **39**(2):71-80.
11. Arden N, Nevitt MC. Osteoarthritis:epidemiology. *Best Pract Res Clin Rheumatol*. 2006 **20**(1):3-25.
12. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis.* 1957 **16**(4):494-502.
13. Cole AA, Kuettner KE. Molecular basis for differences between human joints. *Cell Mol Life Sci*. 2002 **59**(1):19-26.
14. Peach CA, Carr AJ, Loughlin J. Recent advances in the genetic investigation of osteoarthritis. *Trends Mol Med*. 2005 **11**(4):186-191.

15. Kidd BL. Osteoarthritis and joint pain. *Pain*. 2006 **123**(1-2):6-9.
16. Juni P, Reichenbach S, Dieppe P 2006. Osteoarthritis: rational approach to treating the individual. *Best Pract Res Clin Rheumatol.* 2006 **20**(4):721-740.
17. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, *et al*. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis Rheum*. 1986 29(8):1039-1049.
18. Wenham CY, Conaghan PG. New horizons in osteoarthritis. *Age Ageing* 2013 **42**(3):272-278.
19. Keen HI, Wakefield RJ, Conaghan PG. A systematic review of ultrasonography in osteoarthritis. *Ann Rheum Dis*. 2009 **68**(5):611-619.
20. Umlauf D, Frank S, Pap T, Bertrand J, Cartilage biology, pathology, and repair. *Cell Mol Life Sci*. 2010 **67**(24):4197-4211.
21. Wieland HA, Michaelis M, Kirschbaum BJ, Rudolphi KA. *Nat Rev Drug Discov.* 2005 **4**(4):331-344.
22. McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM *et al*. OARSI guidelines for the non-surgical management of knee osteoarthritis. *Osteoarthritis Cartilage*. 2014 **22**(3):363-388.
23. Conaghan PG, Dickson J, Grant RL, Guideline Development Group. Care and management of osteoarthritis in adults: summary of NICE guidance. *BMJ*. 2008 **336**(7642):502-503.
24. Doherty M, Hawkey C, Goulder M, Gibb I, Hill N, Aspley S *et al*. A randomised controlled trial of ibuprofen, paracetamol or a combination tablet of ibuprofen/paracetamol in community-derived people with knee pain. *Ann Rheum Dis.* 2011 **70**(9):1534-1541.
25. Arroll B, Goodyear-Smith F. Corticosteroid injections for osteoarthritis of the knee: meta-analysis. *BMJ*. 2004 **328**(7444):869.
26. Grunke M, Schulze-Koops H. Successful treatment of inflammatory knee osteoarthritis with tumour necrosis factor blockade. *Ann Rheum Dis*. 2006 **65**(4):555-556.

27. Frakes EP, Risser RC, Ball TD, Hochberg MC, Wohlreich MM. Duloxetine added to oral non-steroidal anti-inflammatory drugs for treatment of knee osteoarthritis: results of a randomised, double-blind, placebo-controlled trial. *Curr Med Res Opin.* 2011 **27**(12):2361-2372.
28. Walsh DA, Bonnet CS, Turner EL, Wilson D, Situ M, McWilliams DF. Angiogenesis in the synovium and at the osteochondral junction in osteoarthritis. *Osteoarthritis Cartilage*. 2007 **15**(7):743-751.
29. Lane NE, Schnitzer Tj, Birbara CA, Mokhtarani M, Shelton DL, Smith MD *et al*. Tanezumab for the treatment of pain from osteoarthritis of the knee. *N Engl J Med*. 2010 **363**(16):1521-1531.
30. Fransen M, Agaliotis M, Nairn L, Votrubec M, Bridgett L, Su S *et al*. Glucosamine and chondroitin for knee osteoarthritis: a double-blind randomised placebo-controlled clinical trial evaluating single and combination regimens. *Ann Rheum Dis*. 2014. Epub ahead of print.
31. Ayral X. Injections in the treatment of osteoarthritis. *Best Prac Res Clin Rheumatol*. 2001 **15**(4):609-626.
32. Creamer P, Hochberg MC. Osteoarthritis. *Lancet*. 1997 **350**(9076):503-508.
33. Cooper C, Reginster JY, Chapurlat R, Christiansen C, Genant H, Bellamy N *et al.* Efficacy and safety of oral strontium ranelate for the treatment of knee osteoarthritis: rationale and design of randomised, double-bind, placebo-controlled trial. *Curr Med Red Opin*. 2012 **28**(2):231-239.
34. Reginster J, Badurski J, Bellamy N, Bensen W, Chapurlat R, Chevalier X *et al.* Efficacy and safety of strontium ranelate in the treatment of knee osteoarthritis: results of a double-bind, randomised placebo-controlled trial. *Ann Rheum Dis.* 2013 **72**(2):179-186.
35. Manicourt DH, Azria M, Mindeholm L, Thonar EJ, Devogelaer JP. Oral salmon calcitonin reduces Lequesne's algofunctional index scores and decreases urinary and serum levels of biomarkers of joint metabolism in knee osteoarthritis. *Arthritis Rheum*. 2006 **54**(10):3205-3211.
36. van baar ME, Dekker J, Oostendorp RA, Nijl D, Voorn TB, Bijlsma JW. Effectiveness of exercise in patients with osteoarthritis of hip or knee: nine months' follow up. *Ann Rhem Dis*. 2001 **60**(12):1123-1130.

37. Bennell K, Dobson F. Review: exercise interventions improve pain and function in people with knee osteoarthritis compared with no exercise. *Evid Based Nurs.* 2013. Epub ahead of print.
38. Onyekwelu I, Goldring MD, Hidaka C. Chondrogenesis, joint formation, and articular cartilage regeneration. *J Cell Biochem*. 2009 **107**(3):383-392.
39. Marijnissen AC, van Roermund PM, van Melklebeek J, Schenk W, Verboult AJ, Bijlsma JW *et al*. Clinical benefit of joint distraction in the treatment of severe osteoarthritis of the ankle: proof of concept in an open prospective study and in a randomised controlled study. *Arthritis Rheum*. 2002 **46**(11):2893-2902.
40. Ploegmakers JJ, van Roermund PM, van Melkebeek J, Lammens J, Bijlsma JW, Lafeber F *et al*. Prolonged clinical benefit from joint distraction in the treatment of ankle osteoarthritis. *Osteoarthritis Cartilage*. 2005 **13**(7):582-588.
41. Intema F, Thomas TP, Anderson DD, Elkins JM, Brown TD, Amendola A *et al.* Subchondral bone remodelling is related to clinical improvement after joint distraction in the treatment of ankle osteoarthritis. *Osteoarthritis Cartilage*. 2011 **19**(6):668-675.
42. Intema F, van Roermund PM, Marijnissen AC, Cotofana S, Eckstein F, Castelein RM *et al*. Tissue structure modification in knee osteoarthritis by use of joint distraction: an open 1-year pilot study. *Ann Rheum Dis*. 2011 **70**(8):1441-1446.
43. Wiegant K, van Roermund PM, Intema F, Cotofana S, Eckstein F, Mastbergen SC *et al.* Sustained clinical and structural benefit after joint distraction in the treatment of severe knee osteoarthritis. *Osteoarthritis Cartilage*. 2013 **21**(11):1660-1667.
44. Anakwe RE, Jenkins PJ, Moran M. Predicting dissatisfaction after total hip arthroplasty: a study of 850 patients. *J Arthroplasty*. 2011 **26**(2):209-213.
45. Archer CW, Dowthwaite GP, Francis-West P. Development of synovial joints. *Birth Defects Res C Embryo Today*. 2003 **69**(2):144-155.
46. Ralphs JR, Benjamin M. The joint capsule: structure, composition, ageing and disease. *J Anat*. 1994 **184**(Pt 3):503-509.
47. Knudson CB, Knudson W. Cartilage proteoglycans. *Semin Cell Dev Biol*. 2001 **12**(2):69-78.

48. Wong M, Carter DR. Articular cartilage functional histomorphology and mechanobiology: a research perspective. *Bone*. 2003 **33**(1):1-13.
49. Burr DB. Anatomy and physiology of the mineralised tissues: role in the pathogenesis of osteoarthritis. *Osteoarthritis Cartilage*. 2004 **12** Suppl A:S20-30.
50. Goldring MB, Goldring SR. Articular cartilage and subchondral bone in the pathogenesis of osteoarthritis. *Ann N Y Acad Sci*. 2010 1192:230-237.
51. Martel-Pelletier J, Pelletier JP. Is osteoarthritis a disease involving only cartilage or other articular tissues? *Eklem Hastalik Cerrahisi.* 2010 **21**(1):2-14.
52. Suri S, Walsh DA. Osteochondral alterations in osteoarthritis. *Bone*. 2012 **51**(2):204-211.
53. Scanzello CR, Goldring SR. The role of synovitis in osteoarthritis pathogenesis. *Bone*. 2012 **51**(2):249-257.
54. Wehrli FW. Structural and functional assessment of trabecular and cortical bone by micro magnetic resonance imaging. *J Magn Reson Imaging*. 2007 **25**(2):390-409.
55. Hui AY, McCarty WJ, Masuda K, Firestein GS, Sah RL. A systems biology approach to synovial joint lubrication in health. Injury and disease. *Wiley Interdiscip Rev Syst Biol Med*. 2012 **4**(1):15-37.
56. Bastiaansen-Jenniskens YM, Clockaerts S, Feijt C, Zuurmond AM, Stojanovic-Susulic V, Bridts C *et al*. Infrapatellar fat pad of patients with end-stage osteoarthritis inhibits catabolic mediators in cartilage. *Ann Rheum Dis*. 2012 **71**(2):288-294.
57. Clockaerts S, Bastiaansen-Jenniskens YM, Runhaar J, van Osch GJ, van Offel JF, Verhaar JA *et al*. The infrapatellar fat pad should be considered as an active osteoarthritis joint tissue: a narrative review. *Osteoarthritis Cartilage*. 2010 **18**(7):876-882.
58. Ushiyama T, Chano T, Inoue K, Matsusue Y. Cytokine production in the infrapatellar fat pad: another source of cytokines in knee synovial fluids. *Ann Rheum Dis*. 2003 **62**(2):108-112.
59. Lee JM, Fu FH. The meniscus: Basic science and clinical applications. *Oper Tech Orthop.* 2000 **10**(30):162-168.

60. Brindle T, Nyland J, Johnson DL. The meniscus: review of basic principles with application to surgery and rehabilitation. *J Athl Train*. 2001 **36**(2):160-169.
61. Benjamin M, Ralphs JR. Fibrocartilage in tendons and ligaments – an adaption to compressive load. *J Anat*. 1998 **193**(Pt 4):481-494.
62. Nordin M, Frankel VH. Basic biomechanics of the musculoskeletal system 3rd ed. Philadelphia: Lippincott Williams & Wilkins: 2001. pp 104-118.
63. Woo SL, Abramowitch SD, Kilger R, Liang R. Biomechanics of knee ligaments:injury, healing and repair. *J Biomech*. 2006 **39**(1):1-20.
64. Juneja SC, Veillette C. Defects in tendon, ligament and enthesis in response to genetic alterations in key proteoglycans and glycoproteins: a review. *Arthritis*. 2013. Epub ahead of print.
65. Michigami T. Current understanding on the molecular basis of chondrogenesis. *Clin Pediatr Endocrinol*. 2014 **23**(1):1-8.
66. DeLise AM, Fischer L, Tuan RS. Cellular interactions and signalling in cartilage development. *Osteoarthritis Cartilage*. 2000 **8**(5):309-334.
67. Hall Bk, Miyake T. All for one and one for all: condensations and the initiation of skeletal development. *Bioessays*. 2000 **22**(2):138-147.
68. Provot S, Schipani E. Molecular mechanisms of endochondral bone development. *Biochem Biophys Res Commun*. 2005 **328**(3):658-665.
69. Song B, Estrada KD, Lyons KM. Smad signalling in skeletal development and regeneration. *Cytokine Growth Factor Rev*. 2009 **20**(5-6):379-388.
70. Spater D, Hill TP, Gruber M, Hartmann C. Role of canonical Wnt-signalling in joint formation. *Eur Cell Mater*. 2006 **17**(12):71-80.
71. Long F, Zhang XM, Karp S, Yang Y, McMahon AP. Genetic manipulation of hedgehog signalling in the endochondral skeleton reveals a direct role in the regulation of chondrocyte proliferation. *Development*. 2001 **128**(24):5099-5108.
72. Mead TJ, Yutzey KE. Notch pathway regulation of chondrocyte differentiation and proliferation during appendicular and axial skeleton development. *Proc Natl Acad Sci USA*. 2009 **106**(34):14420-14425.
73. Horbelt D, Denkis A, Knaus P. Transforming growth factor β superfamily signalling: background matters. *Int J Biochem Cell Biol*. 2012 **44**(3):469-474.

74. Kitisin K, Saha T, Blake T, Golestaneh N, Deng M, Kim C *et al*. Tgf-Beta signalling in development. *Sci STKE*. 2007 **2007**(339):cm1.
75. Singh P, Schwarzbauer JE. Fibronectin and stem cell differentiation- lessons from chondrogenesis. *J Cell Sci*. 2012 **125**(Pt 16):3703-3712.
76. Studer D, Millan C, Ozturk E, Maniura-Weber K, Zenobi-Wong M. Molecular and biophysical mechanisms regulating hypertrophic differentiation in chondrocytes and mesenchymal stem cells. *Eur Cell Mater*. 2012 **24**(24):118-135.
77. Li Y, Xu L, Olsen BR. Lessons from genetic forms of osteoarthritis for the pathogenesis of the disease. *Osteoarthritis Cartilage*. 2007 **15**(10):1101-1105.
78. Rizkalla G, Reiner A, Bogoch E, Poole AR. Studies of the articular cartilage proteoglycan aggrecan in health and osteoarthritis. Evidence for molecular heterogeneity and extensive molecular changes in disease. *J Clin Invest*. 1992 **90**(6):2268-2277.
79. Lohmander LS, Ionescu M, Jugessur H, Poole AR. Changes in joint cartilage aggrecan after knee injury and in osteoarthritis. *Arthritis Rheum*. 1999 **42**(3):534-544.
80. Birkedal-Hansen H, Moore WG, Bodden MK, Windsor LJ, Birkedal-Hansen B, DeCarlo A *et al*. Matrix metalloproteinases: a review. *Crit Rev Oral Biol Med*. 1993 **4**(2):197-250.
81. Belcher C, Fawthrop F, Bunning R , Doherty M. Plasminogen activators and their inhibtors in synovial fluids from normal, osteoarthritis and rheumatoid arthritis knees. *Ann Rheum Dis*. 1996 **55**(4):230-236.
82. Langdon C, Kerr C, Hassen M, Hara T, Aresnault AL, Richards CD. Murine oncostatin M stimulates mouse synovial fibroblasts in vitro and induces inflammation and destruction in mouse joints in vivo. *Am J Pathol*. 2000 **157**(4):117-1196.
83. Nevitt MC, Zhang Y, Javaid MK, Neogi T, Curtis JR, Niu J *et al*. High systemic bone mineral density increases the risk of incident knee OA and joint space narrowing, but not radiographic progression of existing knee OA: the MOST study. *Ann Rheum Dis*. 2010 **69**(1):163-168.
84. Lee JY, Harvey WF, Price LL, Paulus JK, Dawson-Hughes B *et al*. Relationship of bone mineral density to progression of knee osteoarthritis. *Arthritis Rheum*. 2013 **65**(6):1541-1546.

85. Felson DT, Lawrence RC, Dieppe PA, Hirsch R, Helmick CG, Jordan JM et al. Osteoarthritis: new insights. Part 1: the disease and its risk factors. *Ann Intern Med*. 2000 **133**(8):635-646.

86. Goldring MB, Birkhead JR, Suen LF, Yamin R, Mizuno S, Glowacki J *et al*. Interleukin-1 beta-modulated gene expression in immortalized human chondrocytes. *J Clin Invest*. 1994 **94**(6):2307-2316.

87. Mengshol JA, Vincenti MP, Coon CI, Barchowsky A, Brinckerhoff CE. Interleukin-1 induction of collagenase 3 (matrix metalloproteinase 13) gene expression in chondrocytes requires p28, c-Jun N-terminal kinase, and nuclear factor kappaB: differential regulation of collagenase 1 and collagenase 3. *Arthritis Rheum*. 2000 **43**(3):801-811.

88. Tetlow LC, Adlam DJ, Woolley DE. Matrix metalloproteinase and proinflammatory cytokine production by chondrocytes of human osteoarthritic cartilage: associations with degenerative changes. *Arthritis Rheum*. 2001 **44**(3):585-594.

89. Park JY, Pillinger MH, Abramson SB. Prostaglandin E2 synthesis and secretion: the role of PGE2 synthases. *Clin Immunol*. 2006 **119**(3):229-240.

90. Distel E, Cadoudal T, Durant S, Poignard A, Chevalier X, Benelli C. The infrapatellar fat pad in knee osteoarthritis: an important source of interleukin-6 and its soluble receptor. *Arthritis Rheum*. 2009 **60**(11):3374-3377.

91. Iliopoulos D, Malizos KN, Tsezou A. Epigenetic regulation of leptin affects MMP-13 expression in osteoarthritic chondrocytes: possible molecular target for osteoarthritis therapeutic intervention. *Ann Rheum Dis.* 2007 **66**(12):1616-1621.

92. Pauli C, Grogan SP, Patil S, Otsuki S, Hasegawa A, Koziol J *et al*. Macroscopic and histopathologic analysis of human knee menisci in aging and osteoarthritis. *Osteoarthritis Cartilage*. 2011 **19**(90):1132-1141.

93. Costa CR, Morrison WB, Carrino JA. Medial meniscus extrusion on knee MRI: is extent associated with severity of degeneration or type of tear? *AJR Am J Roentgenol.* 2004 **183**(1):17-23.

94. van der Esch M, Steultjens M, Wieringa H, Dinant H, Dekker J. Structural joint changes, malalignment, and lacity in osteoarthritis of the knee. *Scand J Rheumatol*. 2005 **34**(4):298-301.

95. Chaganti RK, Lane NE. Risk factors for incident osteoarthritis of the hip and knee. *Curr Rev Musculoskelet Med*. 2011 **4**(3):99-104.
96. Suri P, Morgenroth DC, Hunter DJ. Epidemiology of osteoarthritis and associated comorbidities. *PM R*. 2012 **4**(5 Suppl):S10-19.
97. www.who.int
98. Abramson SB, Attur M. Developments in the scientific understanding of osteoarthritis. *Arthritis Res Ther*. 2009 **11**(3):227.
99. Knoop J, Steultjens MP, van der Leeden M, van der Esch M, Thorstensson CA, Roorda LD *et al*. Proprioception in knee osteoarthritis: a narrative review. *Osteoarthritis Cartilage*. 2011 **19**(4):381-388.
100. Simonet WS. Genetics of primary generalised osteoarthritis. *Mol Genet Metab*. 2002 **77**(1-2):31-4.
101. Richmond RS, Carlson CS, Register TC, Shanker G, Loeser RF. Functional estrogen receptors in adult articular cartilage: estrogen replacement therapy increases chondrocyte synthesis of proteoglycans and insulin-like growth factor binding protein 2. *Arthritis Rheum*. 2000 **43**(9):2081-2090.
102. Zhang Y, McAlindon TE, Hannan MT, Chaisson CE, Klein R, Wilson PW *et al*. Estrogen replacement therapy and worsening of radiographic knee osteoarthritis: the Framingham Study. *Arthritis Rheum.* 1998 **41**(10):1867-1873.
103. Nelson AE, Renner JB, Schwartz TA, Kraus VB, Helmick CG, Jordan JM. Differences in multijoint radiographic osteoarthritis phenotypes among African Americans and Caucasians: the Johnston County Osteoarthritis project. *Arthritis Rheum*. 2011 **63**(12):3843-3852.
104. Zhang Y, Xu L, Nevitt MC, Aliabadi P, Yu W, Qin M *et al*. Comparison of the prevalence of knee osteoarthritis between the elderly Chinese population in Beijing and whites in the United States: The Beijing Osteoarthritis Study. *Arthritis Rheum*. 2001 **44**(9):2065-2071.
105. Nevitt MC, Xu L, Zhang Y, Lui LY, Yu W, Lane NE *et al.* Very low prevalence of hip osteoarthritis among Chinese elderly in Beijing, China, compared with whites in the United States: the Beijing osteoarthritis study. *Arthritis Rheum*. 2002 **46**(7):1773-1779.

106. Sharma L, Song J, Felson DT, Cahue S, Shamiyeh E, Dunlop DD. The role of knee alignment in disease progression and functional decline in knee osteoarthritis. *JAMA*. 2001 **286**(2):188-195.

107. Sharma L, Eckstein F, Song J, Guermazi A, Prasad P, Kapoor D *et al*. Relationship of meniscal damage, meniscal extrusion, malalignment, and joint laxity to subsequent cartilage loss in osteoarthritic knees. *Arthritis Rheum*. 2008 **58**(6):1716-1726.

108. Sharma L, Lou C, Cahue S, Dunlop DD. The mechanism of the effect of obesity in knee osteoarthritis: the mediating role of malalignment. *Arthritis Rheum*. 2000 **43**(3):568-575.

109. Manek NJ, Hart D, Spector TD, MacGregor AJ. The association of body mass index and osteoarthritis of the knee joint: an examination of genetic and environmental influences. *Arthritis Rheum*. 2003 **48**(4):1024-1029.

110. Carman WJ, Sowers M, Hawthorne VM, Weissfeld LA. Obesity as a risk factor for osteoarthritis of the hand and wrist: a prospective study. *Am J Epidemiol.* 1994 **139**(2):119-129.

111. Gandhi R, Takahashi M, Rizek R, Dessouki O, Mahomed NN. Obesity-related adipokines and shoulder osteoarthritis. *J Rheumatol*. 2012 **39**(10):2046-2048.

112. Puranen J, Ala-Ketola L, Peltokallio P, Saarela J. Running and primary osteoarthritis of the hip. *Br Med J*. 1975 **2**(5968):424-425.

113. Evangelou E, Kerkhof HJ, Styrkarsdottir U, Ntzani E, Bos SD, Esko T *et al*. A meta-analysis of genome-wide association studies identifies novel variants associated with osteoarthritis of the hip. *Ann Rheum Dis*. 2013. Epub ahead of print.

114. Stecher RM, Hersh AH. Heberden's nodes: the mechanism of inheritance in hypertrophic arthritis of the fingers. *J Clin Invest*. 1944 **23**(5):699-704.

115. MacGregor AJ, Spector TD. Twins and the genetic architecture of osteoarthritis. *Rheumatology (Oxford)*. 1999 **38**(7):583-588.

116. Valdes AM, Spector TD. The genetic epidemiology of osteoarthritis. *Curr Opin Rheumatol.* 2010 **22**(2):139-143.

117. Hrdlickova B, Westra HJ Franke L Wijmenga C. Celiac disease: moving from genetic associations to causal variants. *Clin Genet*. 2011 **80**(3):203-313.

118. Van Meurs JB, Uitterlinden AG. Osteoarthritis year 2012 in review: genetics and genomics. *Osteoarthritis Cartilage*. 2012 **20**(12):1470-1476.

119. Reynard LN, Loughlin J. The genetics and functional analysis of primary osteoarthritis susceptibility. *Expert Rev Mol Med*. 2013 epub.
120. Meulenbelt I, Min JL, Bos S, Riyazi N, Houwing-Duistermaat JJ, van der Wijk HJ *et al*. Identification of DIO2 as a new susceptibility locus for symptomatic osteoarthritis. *Hum Mol Genet*. 2008 **17**(12):1867-1875.
121. Keen RW, Hart DJ, Lanchbury JS, Spector TD. Association of early osteoarthritis of the knee with a Taq I polymorphism of the vitamin D receptor gene. *Arthritis Rheum*. 1997 **40**(8):1444-1449.
122. Jin SY, Hong SJ, Yang HI, Park SD, Yoo MC Lee HJ et al. Estrogen receptor-alpha gene haplotype is associated with primary knee osteoarthritis in Korean population. *Arthritis Res Ther*. 2004 **6**(5):R415-421.
123. Smith AJ, Keen LJ, Billingham MJ, Perry MJ, Elson CJ, Kirwan JR *et al*. Extended haplotypes and linkage disequilibrium in the IL1R1-1L1A-IL1B-IL1RN gene cluster: association with knee osteoarthritis. *Genes Immun*. 2004 **5**(6):451-460.
124. Loughlin J, Dowling B, Mustafa Z, Chapman K. Association of the interleukin-1 gene cluster on chromosome 2q13 with knee osteoarthritis. *Arthritis Rheum*. 2002 **46**(6):1519-1527.
125. Meulenbelt I, Seymour AB, Nieuwland M, Huizinga TW, van Duijn CM, Slagboom PE. Association of the interleukin-1 gene cluster with radiographic signs of osteoarthritis of the hip. *Arthritis Rheum*. 2004 **50**(4):1179-1186.
126. Valdes AM, Loughlin J, van Oene M, Chapman K, Surdulescu GL, Doherty M *et a*l. Sex and ethnic differences in the association of ASPN, CALM1, COL2A1, COMP, and FRZB with genetic susceptibility to osteoarthritis of the knee. *Arthritis Rheum*. 2007 **56**(1):136-146.
127. Limer KL, Tosh K, Bujac SR, McConnell R, Doherty S, Nyberg F *et al*. Attempt to replicate published genetic associations in a large, well-defined osteoarthritis case-control population (the GOAL study). *Osteoarthritis Cartilage*. 2009 **17**(6):782-789.
128. Rodriguez-Fontenla C, Carr A, Gomez-Reino JJ, Tsezou A, Loughlin J, Gonzalez A. Association of a BMP5 microsatellite with knee osteoarthritis: case-control study. *Arthritis Res Ther*. 2012 **14**(6):R257.

129. Hochberg MC, Yerges-Armstrong L, Yau M, Mitchell BD. Genetic epidemiology of osteoarthritis: recent developments and future directions. *Curr Opin Rheumatol*. 2013 **25**(2):192-197.

130. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007 **39**(7):906-913.

131. Kerkhof HJ, Lories RJ, Meulenbelt I, Jonsdottir I, Valdes AM, Arp P *et al*. A genome-wide association study identifies an osteoarthritis susceptible locus on chromosome 7q22. *Arthritis Rheum*. 2010 **62**(2):499-510.

132. Castaño Betancourt MC, Cailotto F, Kerkhof HJ, Cornelis FM, Doherty SA, Hart DJ *et al*. Genome-wide association and functional studies identify the DOT1L gene to be involved in cartilage thickness and hip osteoarthritis. *Proc Nat Acad Sci USA*. 2012 **109**(21):8218-8223.

133. bEvangelou E, Valdes AM, Castano-Betancourt, Doherty M, Doherty S, Esko T *et al.* The DOT1L rs12982744 polymorphism is associated with osteoarthritis of the hip with genome wide-statistical significance in males. *Ann Rheum Dis*. 2013 **72**(7):1264-1265.

134. Day-Williams AG, Southam L, Panoutsopoulou K, Rayner NW, Esko T, Estrada K *et al*. A variant in MCF2L is associated with osteoarthritis. *Am J Hum Genet*. 2011 **89**(3):446-450.

135. arcOGEN Consortium. Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet*. 2012 **380**(9844):815-823.

136. Hou L, Zhao H. A review of post-GWAS prioritisation approaches. *Front Genet*. 2013 **9**(4):280.

137. Kindt AS, Navarro P, Semple CA, Haley CS. The genomic signature of trait-associated variants. *BMC Genomics*. 2013 **18**(14):108.

138. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS *et al*. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Nat Acad Sci USA*. 2009 **106**(23):9362-9367.

139. Hoffman MM, Ernst J, Wilder SP, Kundaje A, Harris RS, Libbrecht M *et al*. Integratice annotation of chromatin elements from ENCODE data. *Nucleic Acids Res*. 2013 **41**(2):827-841.

140. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet*. 2010 **6**(4):e1000214

141. Gilad Y, Rifkin SA, Pritchard JK. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends Genet.* 2008 **24**(8):408-415.

142. Majewski J, Ott J. Distribution and characterisation of regulatory elements in the human genome. *Genome Res*. 2002 **12**(12):1827-1836.

143. Kwan T, Benovoy D, Dias C, Gurd S, Provencher C, Neaulieu P *et al*. Genome-wide analysis of transcript isoform variation in humans. *Nat Genet*. 2008 **40**(2):225-231.

144. Veyrieras JB, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M *et al.* High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet*. 2008 **4**(10):e1000214.

145. Krawczak M, Thomas NS, Hundrieser B, Mort M, Wittig M, Hampe J *et al*. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum Mutat.* 2007 **28**(2):150-158.

146. Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. *Genes Dev*. 2003 **17**(4):419-437.

147. Coulombe-Huntington J, Lam KC, Dias C, Majewski J. Fine-scale variation and genetic determinants of alternative splicing across individuals. *PLoS Genet*. 2009 **5**(12):e1000766.

148. McCauley JL, Kenealy SJ, Margulies EH, Schnetz-Boutaud N, Gregory SG Hauser SL *et al*. SNPs in Multi-species Conserved Sequences (MCS) as useful markers in association studies: a practical approach. *BMC Genomics*. 2007 **6**(8):266.

149. Evangelou E, Valdes AM, Kerkhof HJ, Styrkarsdottir U, Zhu Y, Meulenbelt I *et al*. Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. *Ann Rheum Dis*. 2011 **70**(2):349-355.

150. Knutsen HK, Tasken K, Eskild W, Richards JS, Kurten RC, Torjesen PA *et al*. Characterisation of the 5´- flanking region of the gene for the cAMP-inducible protein kinase A subunit RIIbeta, in Sertoli cells. *Mol Cell Endocrinol*. 1997 **129**(1):101-114.

151. Zhang P, Smith-Nguyen EV, Keshwani MM, Deal MS, Komey AP, Taylor SS. Structure and allostery of the PKA RIIβ tetrameric holoenzyme. *Science*. 2012 **335**(6069):712-716.

152. Srivastava RK, Lee YN, Noguchi K, Park YG, Ellis MJ, Jeong JS *et al*. The RIIbeta regulatory subunit of protein kinase A binds to cAMP response element: an alternative cAMP signalling pathway. *Proc Natl Acad Sci USA*. 1998 **95**(12):6687-6692.

153. Cummings DE, Brandon EP, Planas JV, Motamed K, Idzerda RL, McKnight GS. Genetically lean mice result from targeted disruption of the RII beta subunit of protein kinase A. *Nature.* 1996 **382**(6592):622-626.

154. Planas JV, Cummings DE, Idzerda RL, McKnight GS. Mutation of the RIIbeta subunit of protein kinase A differentially affects lipolysis but not gene induction in white adipose tissue. *J Biol Chem*. 1999 **274**(51):36281-36287.

155. Brandon EP, Logue SF, Adams MR, Qi M, Sullivan SP, Matsumoto AM *et al.* Defective motor behaviour and neural gene expression in RIIbeta-protein kinase A mutant mice. *J Neurosci*. 1998 **18**(10):3639-3649.

156. Elliott MR, Tolnay M, Tsokos GC, Kammer GM. Protein kinase A regulatory subunit type II beta directly interacts with and suppresses CREB transcriptional activity in activated T cells. *J Immunol*. 2003 **171**(7):3636-3644.

157. Trevosian SG, Shih HH, Mendelson KG, Sheppard KA, Paulson KE, Yee AS. HBP1: a HMG box transcriptional repressor that is targeted by the retinoblastoma family. *Genes Dev* 1997. **11**(3):383-396.

158. Shih HH, Xiu M, Berasi SP, Sampson EM, Leiter A, Paulson KE *et al*. HMG box transcriptional repressor HBP1 maintains a proliferation barrier in differentiated liver. *Mol Cell Biol*. 2001 **21**(71):5723-5732.

159. Zhang X, Kim J, Ruthazer R, McDevitt MA, Wazer DE, Paulson KE *et al*. The HBP1 transcriptional repressor participates in RAS-induced premature senescence. *Mol Cell Biol.* 2006 **26**(22):8252-8266.

160. Escamilla-Powers JR, Daniel CJ, Farrell A, Taylor K, Zhang X, Byers S *et al*. The tumor suppressor protein HBP1 is a novel c-myc-binding protein that negatively regulators c-myc transcriptional activity. *J Biol Chem*. 2010 285(7):4847-4858.

161. Zhuma T, Tyrrell R, Sekkali B, Skavdis G, Saveliev A, Tolaini M *et al*. Human HMG box transcription factor HBP1: a role in hCD2 LCR function. *EMBO J.* 1999 **18**(22):6396-6406.

162. Shih HH, Trevosian SG, Yee AS. Regulation of differentiation by HBP1, a target of the retinoblastoma protein. *Mol Cell Biol*. 1998 **18**(8):4732-4743.

163. Lesage F, Hugnot JP, Amri EZ, Grimaldi P, Barhanin J, Lazdunski M. Expression cloning in K+ transport defective yeast and distribution of HBP1, a new putative HMG transcriptional regulator. *Nucleic Acids Res*. 1994 **22**(18):3685-3688.

164. Berasi SP, Xiu M, Yee AS, Paulson KE. HBP1 expression of the p47phoz gene: cell cycle regulation via the NADPH oxidase. *Mol Cell Biol*. 2004 **24**(7):3011-3024.

165. Ungar D, Oka T, Vasile E, Krieger M, Hughson FM. Subunit architecture of the conserved oligomeric Golgi complex. *J Biol Chem*. 2005 **280**(38):32729-32735.

166. Rymen D, Keldermans L, Race V, Regal L, Deconinck N, Dionisi C *et al*. COG5-CDG: expanding the clinical spectrum. *Orphanet J Rare Dis*. 2012 **10**(7):94.

167. Paesold-Burda P, Maag C, Troxler H, Foulquier F, Kleinert P, Schnabel S *et al.* Deficiency in COG5 causes a moderate form of congenital disorders of glycosylation. *Hum Mol Genet.* 2009 **18**(22):4350-4356.

168. Shestakova A, Zolov S, Lupashin V. COG complex-mediated recycling of Golgi glycosyltransferases is essential for normal protein glycosylation. *Traffic*. 2006 **7**(2):191-204.

169. Farkas RM, Giansanti MG, Gatti M, Fuller MT. The Drosophila homologue is required for cytokinesis, cell elongation, and assembly of specialised Golgi architecture during spermatogenesis. *Mol Biol Cell*. 2003 **14**(1):190-200.

170. Adams JW, Wang J, Davis JR, Liaw C, Gaidarov I, Gatlin J *et al*. Myocardial expression, signalling, and function of GPR22: a protective role for an orphan G protein-coupled receptor. *Am J Physiol Heart Circ Physiol*. 2008 **295**(2):H509-521.

171. O'Dowd BF, Nguyen T, Jung BP, Marchese A, Cheng R, Heng HH *et al*. Cloning and chromosomal mapping of four putative novel human G-protein-coupled receptor genes. *Gene.* 1997 **187**(1):75-81.

172. Yu F, Tanaka Y, Yamashita K, Suzuki T, Nakamura A, Hirano N *et al*. Molecular basis of dihydrouridine formation on tRNA. *Proc Natl Acad Sci USA*. 2011 **108**(49):19593-19598.

173. Rider LW, Ottosen MB, Gattis SG, Palfey BA. Mechanism of dihydrouridein synthase 2 from yeast and the importance of modifications for efficient tRNA reduction. *J Biol Chem*. 2009 **284**(16):10324-10333.

174. Reuter TY, Medhurst AL, Waisfisz Q, Zhi Y, Herterich S, Hoehn H *et al*. Yeast two-hybrid screens imply involvement of Fanconi anemia proteins in transcription regulation, cell signalling, oxidative metabolism, and cellular transport. *Exp Cell Res*. 2003 **289**(2):211-221.

175. Breckenridge DG, Nguyen M, Kuppig S, Reth M, Shore GC. The procaspase-8 isoform, procaspase-8L, recruited to the BAP31 complex at the endoplasmic reticulum. *Proc Natl Acad Sci USA*. 2002 **99**(7):4331-4336.

176. Kim KM, Adachi T, Nielsen PJ, Terashima M, Lamers MC, Kohler G *et al*. Two new proteins preferentially associated with membrane immunoglobulin D. *EMBO J*. 1994 **13**(16):3793-3800.

177. Adachi T, Schamel WW, Kim KM, Watanabe T, Becker B, Nielsen PJ *et al*. The specificity of association of the IgD molecule with the accessory proteins BAP31/BAP29 lies in the IgD transmembrane sequence. *EMBO J*. 1996 15(7):1534-1541.

178. Schamel WW, Kuppig S, Becker B, Gimborn K, Hauri HP, Reth M. A high-molecular-weight complex of membrane proteins BAP29/BAP31 is involved in the retention of membrane-bound IgD in the endoplasmic reticulum. *Proc Nat Acad Sci USA*. 2003 **100**(17):9861-9866.

179. Ladasky JJ, Boyle S, Seth M, Li H, Pentcheva T, Abe F *et al*. Bap31 enhances the endoplasmic reticulum export and quality control of human class I MHC molecules. *J Immunol*. 2006 **177**(9):6172-6181.

180. Neumann E, Riepl B, Knedla A, Lefevre S, Tarner IH, Grifka J *et al*. Cell culture and passaging alters gene expression pattern and proliferation rate in rheumatoid arthritis synovial fibroblasts. *Arthritis Res Ther*. 2010 **12**(3):R83.

181. Sampson EM, Haque ZK, Ku MC, Tevosian SG, Albanese C, Pestell RG *et al*. Negative regulation of the Wnt-beta-catenin pathway by the transcriptional repressor HBP1. *EMBO J*. 2001 **20**(16):4500-4511.

182. Kim J, Zhang X, Rieger-Christ KM, Summerhayes IC, Wazer DE, Paulson KE, Yee AS. Suppresion of Wnt signaling by the green tea compound (-)-epigallocatechin 3-gallate (EGCG) in invasive breast cancer cells. Requirement of the transcriptional repressor HBP1. *J Biol Chem*. 2006 **281**(16):10865-10875.

183. Li H, Wang W, Liu X, Paulson KE, Yee AS, Zhang X. Transcriptional factor HBP1 targets P16(INK4A), upregulating its expression and consequently is involved in Ras-induced premature senescence. *Oncogene*. 2010 **29**(36):5083-5094.

184. Luyten FP, Tylzanowski P, Lories RJ. Wnt signaling and osteoarthritis. *Bone*. 2009 **44**(4):522-527.

185. Loeser RF. Aging and osteoarthritis. *Curr Opin Rheumatol*. 2011 **23**(5):492-496.

186. Scott JL, Gabrielides C, Davidson RK, Swingler TE, Clark IM, Wallis GA *et al*. Superoxide dismutase downregulation in osteoarthritis progression and end-stage disease. *Ann Rheum Dis*. 2010 **69**(8):1502-10.

187. Yuasa T, Otani T, Koike T, Iwamoto M, Enomoto-Iwamoto M. Wnt/beta-catenin signaling stimulates matrix catabolic genes and activity in articular chondrocytes: its possible role in joint degeneration. *Lab Invest*. 2008 **88**(3):264-274.

188. Staines KA, MacRae VE, Farquharson C. Cartilage development and degeneration: a Wnt Wnt situation. *Cell Biochem Function*. 2012 **30**(8):633-642.

189. Enomoto-Iwamoto M, Kitagaki J, Koyama E, Tamamura Y, Wu C, Kanatani N *et al*. The Wnt antagonist Frab-1 regulates chondrocytes maturation and long bone development during limb skeletogenesis. *Dev Biol*. 2002 **251**(1):142-156.

190. Ryu JH, Kim SJ, Kim SH, Oh CD, Hwang SG, Chun CH *et al*. Regulation of the chondrocyte phenotype by beta-catenin. *Development*. 2002 **129**(23):5541-5550.

191. Price JS, Waters JG, Darrah C, Pennington C, Edwards DR Donell ST *et al*. The role of chondrocyte senescence in osteoarthritis. *Aging Cell*. 2002 **1**(1):57-65.

192. Henrotin Y, Kurz B, Aigner T. Oxygen and reactive oxygen species in cartilage degradation: friends or foes? *Osteoarthritis Cartilage*. 2005 **13**(8):643-654.

193. Babbitt CC, Silverman JS, Haygood R, Reininga JM, Rockman MV, Wray GA. Multiple functional variants in cis modulate PDYN expression. *Mol Biol Evol*. 2010 **27**(2):465-479.

194. Sellam J, Berenbaum F. The role of synovitis in pathophysiology and clinical symptoms of osteoarthritis. *Nat Rev Rheumatol*. 2010 **6**(11):625-35.

195. Nakao A, Imamura T, Souchelnytskyi S, Kawabata M, Ishisaki A, Oeda E *et al*. TGF-beta receptor-mediated signalling through Smad2, Smad3 and Smad4. *EMBO J*. 1997 **16**(17):5353-5362.

196. van der Kraan PM, Blaney Davidson EN, Blom A, van den Berg WB. TGF-beta signalling in chrondrocyte terminal differentiation and osteoarthritis: modulation and integration of signalling pathways through receptor-Smads. *Osteoarthritis Cartilage*. 2009 **17**(12):1539-1545.

197. van der Kraan PM, Blaney Davidson EN, van den Berg WB. A role for age-related changes in TGFβ signaling in aberrant chondrocyte differentiation and osteoarthritis. *Arthritis Res Ther*. 2010 **12**(1): 201.

198. Yang X, Chen L, Xu X, Li C, Huang C, Deng CX. TGF-beta/Smad3 signals repress chondrocyte hypertrophic differentiation and are required for maintaining articular cartilage. *J Cell Biol*. 2001 **153**(1):35-46.

199. Li T, Gao L, Sheu T, Sampson E, Flick L, Konttinen Y *et al.* Aberrant hypertrophy in Smad3-deficient murine chondrocytes is rescued by restoring transforming growth factor β-activated kinase 1/activating transcription factor 2 signalling. Arthritis Rheum. 2010 **62**(8):2259-3269.

200. Derynck R, Zhang Y, Feng XH. Smads: transcriptional activators of TGF-beta responses. *Cell*. 1998 **95**(6):737-740.

201. Heldin CH, Moustakas A. Role of Smads in TGFβ signalling. *Cell Tissue Res*. 2012 **347**(1):21-36.

202. Serra R, Johnson M, Filvaroff EH, LaBorde J, Sheehan DM, Derynck R *et al.* Expression of a truncated, kinase-defective TGF-beta type II receptor in mouse skeletal tissue promotes terminal chondrocyte differentiation and osteoarthritis. *J Cell Biol*. 1997 **139**(2):541-552.

203. Dabovic B, Chen Y, Colarossi C, Zambuto L, Obata H, Rifkin DB. Bone defects in latent TGF-beta binding protein (Ltbp)-3 null mice; a role for Ltbp in TGF-beta presentation. *J Endocrinol*. 2002 **175**(1):129-141.

204. Wu Q, Kim KO, Sampson ER, Chen D, Awad H, O-Brien T *et al*. Induction of an osteoarthritis-like phenotype and degradation of phosphorylation Smad3 by Smurf2 in transgenic mice. *Arthritis Rheum*. 2008 **58**(10):3132-3144.

205. Valdes AM, Spector TD, Tamm A, Kisand K, Doherty SA, Dennison EM *et al*. Genetic variation in the SMAD3 gene is associated with hip and knee osteoarthritis. *Arthritis Rheum.* 2010 **62**(8):2347-2352.

206. Johnson AD, Zhang Y, Papp AC, Pinsonneault JK, Lim JE, Saffen D *et al.* Polymorphisms affecting gene transcription and mRNA processing in pharmacogenetic candidate genes: detection through allelic expression imbalance in human target tissues. *Pharmacogenet Genomics* 2008, **18**(9):781-91.

207. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell.* 2004 **116**(2):281-97.

208. Pais H, Nicolas FE, Soond SM, Swingler TE, Clark IM, Chantry A *et al.* Analysing mRNA expression identifies Smad3 as a microRNA-140 target regulated only at protein level. *RNA*. 2010 **16**(3):489-494.

209. Massague J, Seoane J, Wotton D. Smad transcription factors. *Genes Dev*. 2005 **19**(23):2783-2810.

210. Alliston T, Choy L, Ducy P, Karsenty G, Derynck R. TGF-beta-induced repression of CBFA1 by Smad3 decreases cbfa1 and osteocalcin expression and inhibits osteoblast differentiation. *EMBO J*. 2001 **20**(9):2254-2272.

211. Iwamoto M, Kitagaki J, Gentili TC, Koyama E, Enomoto H, Komori T *et al*. Runx2 expression and action in chondrocytes are regulated by retinoid signalling and parathyroid hormone-related peptide (PTHrP). *Osteoarthritis Cartilage*. 2003 **11**(1):6-15.

212. Greenblatt MB, Shim JH, Glimcher LH. TAK1 mediates BMP signalling in cartilage. *Ann N Y Acad Sci*. 2010 1192:385-390

213. van de Laar IM, Oldenburg RA, Pals G, Roos-Hesselink JW, de Graaf BM, Verhagen JM *et al*. Mutations in SMAD3 cause a syndromic form of aortic

aneurysms and dissections with early-onset osteoarthritis. *Nat Genet*. 2011 **43**(2):121-126.

214. van de Laar IM, van der Linde D, Oei EH, Bos PK, Besserns JH, Bierma-Zeinstra SM *et al*. Phenotypic spectrum of the SMAD3-related aneurysms-osteoarthritis syndrome. *J Med Genet*. 2012 **49**(1):47-57.

215. van der Linde D, van de Laar IM, Bertoli-Avella AM, Oldenburg RA, Bekkers JA, Mattace-Raso FU et al. Aggressive cardiovascular phenotype of aneurysms-osteoarthritis syndrome caused by pathogenis SMAD3 variants. *J Am Coll Cardiol*. 2012 **60**(5):397-403.

216. Regalado ES, Guo DC, Villamizar C, Avidan N, Gilchrist D, McGillivray B *et al*. Exome sequencing identifies SMAD3 mutations as a cause of familial thoracic aortic aneurysm and dissection with intracranial and other arterial aneurysms. *Circ Res*. 2011 **109**(6):680-686.

217. Feng Q, Wang H, Ng HH, Erdjument-Bromage H, Tempst P, Struhl K *et al*. Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr Biol*. 2002 **12**(12):1052-1058.

218. Van Holde KE, Allen JR, Tatchell K, Weischet WO, Lohr D. DNA-histone interactions in nucleosomes. *Biophys J*. 1980 **32**(1):271-282.

219. Martin C, Zhang Y. The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol.* 2005 **6**(11):838-849.

220. Jones B, Su H, Bhat A, Lei H, Bakjo J, Hevi S *et al*. The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. *PLoS Genet*. 2008 **4**(9):e1000190.

221. Barry ER, Corry GN, Rasmussen TP. Targeting DOT1L action and interactions in leukemia: the role of DOT1L in transformation and development. *Expert Opin Ther Targets*. 2010 **14**(4):405-418.

222. Kouskouti A, Talianidis I. Histone modifications defining active genes persist after transcriptional and mitotic inactivation. *EMBO J*. 2005 **24**(2):347-357.

223. Kouzarides T. Chromatin modifications and their functions. *Cell*. 2007 **128**(4):693-705.

224. Nguyen AT, He J, Taranova O, Zhang Y. Essential role of DOT1L in maintaining normal adult hematopoiesis. *Cell Res*. 2011 **21**(9):1370-1373.

225. Ayton PM, Cleary ML. Molecular mechanisms of leukemogenesis mediated by MLL fusion proteins. *Oncogene*. 2011 **20**(40):5695-5707.

226. Hsieh JJ, Cheng EH, Korsmeyer SJ. Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression. *Cell.* 2003 **115**(3):293-303.

227. Nguyen AT, Zhang Y. The diverse functions of Dot1 and H3K79 methylation. *Genes Dev*. 2011 **25**(13):1345-1358.

228. Hartmann C, Tabin CJ. Wnt-14 plays a pivotal role in inducing synovial joint formation in the developing appendicular skeleton. *Cell.* 2001 **104**(3):341-351.

229. Mohan M, Herz HM, Takahashi YH, Lin C, Lai KC, Zhang Y *et al*. Linking H3K79 trimethylation signalling through a novel Dot1-containing complex (DotCom). *Genes Dev*. 2010 **24**(6):574-589.

230. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*. 2010 **467**(7317):832-838.

231. Sonio U, Bennett AJ, Millwood IY, Molitor J, O'Reilly PF, Timpson NJ *et al*. Genetic determinants of height growth assessed longitudinally from infancy to adulthood in the northern Finland birth cohort 1966. *PLoS Genet*. 2009 **5**(3):e1000409.

232. Tsai RY, McKay RD. A nucleolar mechanism controlling cell proliferation in stem cells and cancer cells. *Genes Dev*. 2002 **16**(23):2991-3003.

233. Zhu Q, Yasumoto H, Tsai RY. Nucleostemin delays cellular senescence and negatively regulates TRF1 protein stability. *Mol Cell Biol.* 2006. **26**(24):9279-9290.

234. Kafienah W, Mistry S, Williams C, Hollander AP. Nucleostemin is a marker of proliferating stromal stem cells in adult human bone marrow. *Stem Cells.* 2006 **24**(4):1113-1120.

235. Liu SJ, Cai ZW, Liu YJ, Dong MY, Sun LQ, Hu GF *et al.* Role of nucleostemin in growth regulation of gastric cancer, liver cancer and other malignancies. *World J Gastroenterol.* 2004 **10**(9):1246-1249.

236. Tsai RY. Nucleolar modulation of TRF1: a dynamic way to regulate telomere and cell cycle by nucleostemin and GNL3L. *Cell Cycle*. 2009 **8**(18):2912-2916.

237. Dai MS, Sun XX, Lu H. Aberrant expression of nucleostemin activates p53 and induces cell cycle arrest via inhibition of MDM2. *Mol Cell Biol*. 2008 **28**(13):4365-4376.

238. Lo D, Lu H. Nucleostemin: another nucleolar "Twister" of the p52-MDM2 loop. *Cell Cycle.* 2010 **9**(16):3227-3232.

239. Beekman C, Nichane M, De Clercq S, Maetens M, Floss T, Wurst W *et al*. Evolutionarily conserved role of nucleostemin: controlling proliferation of stem/progenitor cells during early vertebrate development. *Mol Cell Biol*. 2006 **26**(24):9291-9301.

240. Meng L, Hsu JK, Zhu Q, Lin T, Tsai RY. Nucleostemin inhibits TRF1 dimerization and shortens its dynamic association with the telomere. *J cell Sci*. 2011 **124**(Pt 21):3706-3714.

241. Zhang Y, Lu H. Signalling to p53: ribosomal proteins find their way. *Cancer Cell.* 2009 **16**(5):369-377.

242. Lindstrom MS, Deisenroth C, Zhang Y. Putting a finger on growth surveillance: insight into MDM2 zinc finger-ribosomal protein interactions. *Cell Cycle*. 2007 **6**(4):434-437.

243. Bodnar AG, Ouellette M, Frolkis M, Holt SE, Chiu CP, Morin GB *et al.* Extension of life-span by introduction of telomerase into normal human cells. *Science*. 1998 **279**(5349):349-352.

244. Kashiwagi M, Enghild JJ, Gendrom C, Hughes C, Caterson B, Itoh Y *et al*. Altered proteolytic activities of ADAMTS-4 expressed by C-terminal processing. *J Biol Chem*. 2004 **279**(11):10109-10119.

245. Fushimi K, Troeberg L, Nakamura H, Lim NH, Nagase H. Functional differences of the catalytic and non-catalytic domains in human ADAMTS-4 and ADAMTS-5 in aggrecanolytic activity. *J Biol Chem*. 2008 **283**(11):6706-6716.

246. Naito S, Shiomi T, Okada A, Kimura T, Chijiwa M, Fujita Y *et al*. Expression of ADAMTS4 (aggrecanase-1) in human osteoarthritic cartilage. *Pathol Int.* 2007 **57**(11):703-711.

247. Song RH, Tortorella MD, Malfait AM, Alston JT, Yang Z, Arner EC *et al*. Aggrecan degradation in human articular cartilage explants is mediated by both ADAMTS-4 and ADAMTS-5. *Arthritis Rheum*. 2007 **56**(2):575-585.

248. Montgomery SB, Dermitzakis ET. From expression QTLs to personalised transcriptomics. *Nat Rev Genet*. 2011 **12**(4):277-282.

249. Miyamoto Y, Mabuchi A, Shi D, Kubo T, Takatori Y, Saito S *et al*. A functional polymorphism in the 5′UTR of GDF5 is associated with susceptibility to osteoarthritis. *Nat Genet*. 2007 **39**(4):529-533.
250. Southam L, Rodriquez-Lopez J, Wilkins JM, Pombo-Suarez M, Snelling S, Gomez-Reino JJ *et al*. A 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*. 2007 **16**(18):2226-2232.
251. Egli R, Southam L, Wilkins JM, Lorenzen I, Pombo-Suarez M, Gonzalez A *et al*. Functional analysis of the osteoarthritis susceptibility-associated GDF5 regulatory polymorphism. *Arthritis Rheum*. 2009 **60**(7):2055-2064.
252. Syddall CM, Reynard LN, Young DA, Loughlin J. The identification of trans-acting factors that regulate the expression of GDF5 via the osteoarthritis susceptibility SNP rs143383. *PLoS Genet*. 2013 **9**(6):e1003557.
253. Dodd AW, Syddall CM, Loughlin J. A rare variant in the osteoarthritis-associated locus GDF5 is functional and reveals a site that can be manipulated to modulate GDF5 expression. *Eur J Hum Genet*. 2013 **21**(5):517-521.
254. Akhtar N, Haqqi TM. MicroRNA-199a* regulates the expression of cyclooxygenase-2 in human chondrocytes. *Ann Rheum Dis*. 2012 **71**(6):1073-1080.
255. Dai L, Zhang X, Hu X, Zhou C, Ao Y. Silencing of microRNA-101 prevents IL-1β-induced extracellular matrix degradation in chondrocytes. *Arthritis Res Ther*. 2012 **14**(6):R268
256. Pais H, Nicolas FE, Soond SM, Swingler TE, Clark IM, Chantry A *et al*. Analysing mRNA expression identifies Smad3 as a microRNA-140 target regulated only at protein level. *RNA* 2010 **16**(3):489-494.
257. Barter MJ, Bui C, Young DA. Epigenetic mechanism in cartilage and osteoarthritis: DNA methylation, histone modifications and microRNAs. *Osteoarthritis Cartilage*. 2012 **20**(5):339-349.
258. Goldring MB, Marcu KB. Epigenomic and microRNA-mediated regulation in cartilage development, homeostasis and osteoarthritis. *Trends Mol Med*. 2012 **18**(2):109-118.

259. Swingler TE, Wheeler G, Carmont V, Elliott HR, Barter MJ, Abu-Elmagd M *et al*. The expression and function of microRNAs in chondrogenesis and osteoarthritis. *Arthritis Rheum*. 2012 **64**(6):1909-1919.

260. Iliopoulos D, Malizos KN, Oikonomou O, Tsezou A. Integrative microRNA and proteomic approaches identify novel osteoarthritis genes and their collaborative metabolic and inflammatory networks. *PLoS One*. 2008 **3**(11):e3740.

261. Miyaki S, Nakasa T, Otsuki S, Grogan SP, Higashiyama R, Inoue A *et al*. MicroRNA-140 is expressed in differentiated human articular chondrocytes and modulates interleukin-1 responses. *Arthritis Rheum*. 2009 **60**(9):2723-2730.

262. Tardif G, Hum D, Pelletier JP, Duval N, Martel-Pelletier J. Regulation of the IGFBP-5 and MMP-13 genes by the microRNAs miR-140 and miR-27a in human osteoarthritis chondrocytes. *BMC Musculoskeletal Disord*. **30**(10):148.

263. van Meurs JB, Uitterlinden AG. Osteoarthritis year 2012 in review: genetics and genomics. *Osteoarthritis Cartilage*. 2012 **20**(12):1470-1476.

264. Bondeson J, Wainwright S, Hughes C, Caterson B. The regulation of the ADAMTS4 and ADAMTS5 aggrecanases in osteoarthritis: a review. *Clin Exp Rheumatol*. 2008 **26**(1):139-145.

265. Tortorella MD, Malfait AM, Deccico C, Arner E. The role of ADAM-TS4 (aggrecanase-1) and ADAM-TS5 (aggrecanase-2) in a model of cartilage degradation. *Osteoarthritis Cartilage*. 2001 **9**(6):539-552.

266. Little CB, Meeker CT, Golub SB, Lawlor KE, Farmer PJ, Smith SM *et al*. Blocking aggrecanase cleavage in the aggrecan interglobular domain abrogates cartilage erosion and promotes cartilage repair. *J Clin Invest*. 2007 **117**(6):1627-1636.

267. Raine EV, Dodd AW, Reynar LN, Loughlin J. Allelic expression analysis of the osteoarthritis susceptibility gene COL11A1 in human joint tissues. *BMC Musculoskeletal Disord*. 2013 **8**(14):85.

268. Wu Y, Waite LL, Jackson AU, Sheu WH, Buyske S, Absher D *et al*. Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. *PLoS Genet*. 2013 **9**(3):e1003379.

269. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*. 2014 **32**(4):347-355.
270. Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol*. 2013 **14**(1):49-55.
271. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet*. 2010 **11**(9):636-646.
272. Handel EM, Cathomen T. Zinc-finger nuclease based genome surgery: it's all about specificity. *Curr Gene Ther*. 2011 **11**(1):28-37.
273. Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF *et al*. A TALE nuclease architectire for efficient genome editing. *Nat Biotechnol*. 2011 **29**(2):143-148.
274. Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S *et al*. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*. 2005 **435**(7042):636-651.
275. Shen B, Zhang W, Zhang J, Zhou J, Wang J, Chen L *et al*. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods*. 2014 **11(**4):399-402.
276. Wilkinson R, Wiedenheft B. A CRISPR method for genome engineering. *F1000Prime Rep.* 2014 **2**(6):3.
277. Niu Y, Shen B, Chen Y, Wang J, Wang L, Kang Y *et al.* Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell*. 2014 **156**(4):835-843.
278. Paul DS, Soranzo N, Beck S. Functional interpretation of non-coding sequence variation: concepts and challenges. *Bioessays.* 2014 **36** (2):191-199.
279. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. *Genome Res*. 2012 **22** (9):1748-1759.
280. Zhong H, Yang X, Kaplan LM, Molony C, Schadt EE. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *Am J Hum Genet*. 2010 **86**(4):581-591.
281. Ryu D, Cho S, Kim H, Lee S, Kim W. GEPdb: a database for investigating the ternary association of genotype, gene expression and phenotype. *Bioinformatics*. 2014 Epub ahead of print.

282. Moutsianas L, Morris AP. Methodology for the analysis of rare genetic variation in genome-wide association and re-sequencing studies of complex human traits. *Brief Funct Genomics*. 2014 **13** (5):362-370.
283. Manolio TA. Bringing genome-wide association findings into clinical use. *Nat Rev Genet*. 2013 **14** (8):549-558.
284. Ombrello MJ, Sikora KA, Kastner DL. Genetics, genomics, and their relevance to pathology and therapy. *Best Pract Res Clin Rheumatol*. 2014 **28** (2):175-189.
285. Das SK, Sharma NK. Expression quantitative trait analyses to identify causal genetic variants for type 2 diabetes susceptibility. *World J Diabetes*. 2014 **5**(2):97-114.
286. Cheng Y, Quinn JF, Weiss LA. An eQTL mapping approach reveals that rare variants in the SEMA5A regulatory network impact autism risk. *Hum Mol Genet*. 2013 **22**(4):2960-2972.
287. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet*. 2010 **6** (4):e1000888.
288. Konermann S, Brigham MD, Trevino A, Hsu PD, Heidenreich M, Cong L *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature*. 2013 **500** (7463):472-476.