Functional Analysis of the OA Associated Variant rs143383

Catherine M. Syddall

Thesis submitted for the Doctor of Philosophy
Institute of Cellular Medicine

April 2013

Abstract

Osteoarthritis (OA) is a common, multifactorial musculoskeletal disease that is characterised by joint pain and reduced joint function. It is a polygenic disease and progresses as a result of the focal loss of the articular cartilage of the synovial joint. rs143383 is a C to T transition single nucleotide polymorphism (SNP) located in the 5′ untranslated region (UTR) of the growth differentiation factor 5 gene *GDF5*. The T allele of the SNP is associated with an increased risk of OA and of a number of other common musculoskeletal diseases. This susceptibility is mediated by the T allele producing less *GDF5* transcript relative to the C allele, a phenomenon known as differential allelic expression (DAE). The aim of my study was to identify *trans*-acting factors that bind to rs143383 and which regulate this *GDF5* DAE. Protein binding to the gene was investigated by two experimental approaches: 1) competition and supershift electrophoretic mobility shift assays (EMSAs) and; 2) an oligonucleotide pull down assay followed by quantitative mass spectrometry. Binding was then confirmed *in vivo* by chromatin immunoprecipitation (ChIP), and the functional effects of candidate proteins investigated by RNA interference (RNAi) and over expression. Using these approaches the *trans*-acting factors Sp1, Sp3, P15 and DEAF-1 were identified as interacting with the *GDF5* 5′UTR. Knockdown and over expression of the factors demonstrated that they were repressors of *GDF5* expression. Depletion of DEAF-1 modulated the DAE of *GDF5* and this differential allelic effect was confirmed following over expression, with the rs143383 T allele being repressed to a significantly greater extent than the rs143383 C allele. In combination, Sp1 and DEAF-1 had the greatest repressive activity. The genotype of a second *GDF5* 5′UTR SNP, rs143384, which is located downstream of rs143383, has been previously found to impact upon the DAE at rs143383. Thus protein binding to rs143384 was also investigated using EMSAs. In conclusion, I identified four *trans*-acting factors that modulate the expression of *GDF5* via the OA susceptibility locus rs143383.

Dedication

I would like to dedicate this thesis to my fiancé Cameron McGillivray and to my family for all the support and encouragement that they have given me over the past four years.

Acknowledgements

I'd like to acknowledge my supervisor Prof. John Loughlin for the fantastic support he has given to me during my PhD and thank him for his advice and encouragement, particularly whilst writing my thesis and applying to study medicine. I would also like to acknowledge my second supervisor Dr. David Young for his suggestions and advice and Prof. Tim Cawston, the coordinator of my PhD programme for his continued support.

Dr. Louise Reynard assisted with both the EMSA experiments in Chapter 4, and with the cloning of the Sp1, Sp3 and P15 EGFP expression vectors in Chapter 7. Shane Damsell was supervised by John Loughlin and I when performing the EMSA experiments in Chapter 9.

I'd like to thank Athanasia Gravani and Steven Woods for experimental advice and support. I'd also like to acknowledge members of the Osteoarthritis Genetics Group and members of the Musculoskeletal Research Group for their suggestions, advice and assistance and for making the lab a fun and friendly place to study.

Thank you to C. Garrison Fathman and Linda Yip for the kind donation of the DEAF-1 EGFP and empty EGFP expression vectors. Thank you to the research groups of Paul Albert and Jane E.Visvader for the kind donation of two DEAF-1 antibodies.

I'd like to acknowledge Achim Treumann and Karen Lowden (North East Proteome Analysis Facility (Newcastle)) for the mass spectrometry data requisition and for assistance with the design and analysis of this experiment.

Finally, I'd like to thank the Nuffield Foundation and the Oliver Bird Rheumatism Programme for funding this research.

Table of Contents

Figures and Tables

Abbreviations

3′UTR	Three Prime Untranslated Region
5′UTR	Five Prime Untranslated Region
APS	Ammonium persulfate
arcOGEN	Arthritis Research Campaign Osteoarthritis Genetics
BD	Brachydactyly
BP	Brachypodism
BMD	Bone mineral density
BMI	Body mass index
BMP	Bone morphogenetic protein
BSA	Bovine serum albumin
CDH	Congenital hip dysplasia
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
CDMP	Cartilage derived morphogenetic protein
COX	Cyclooxygenase
DAE	Differential allelic expression
DAPI	4',6-diamidino-2-phenylindole
DMEM	Dulbecco's Modified Eagles Medium
DMOAD	Disease modifying OA drugs
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
DZ	Dizygotic
DEAF-1	Nuclear Deformed Epidermal Auto-regulatory Factor 1
ECM	Extracellular matrix
EGFP	Enhanced green fluorescent protein
EMSA	Electrophoretic mobility shift assay
FBS	Fetal bovine serum
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GDF5	Growth Differentiation Factor 5
gDNA	Genomic DNA
GWAS	Genome wide association scan
HAC	Human articular chondrocyte
IL-1	Interleukin-1
KL	Kellgren and Lawrence

LDD	Lumbar disc disease
LLI	Leg Length Inequality
MMP	Matrix Metalloproteinase
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
MS	Mass Spectrometry
MZ	Monozygotic
NOF	Neck of femur fracture
NP40	Nonyl Phenoxypolyethoxylethanol
NSAID	Non Steroidal Anti Inflammatory
NTsiRNA	Non-targeting siRNA
OA	Osteoarthritis
OSM	Oncostatin M
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
RNA	Ribose Nucleic Acid
RNAi	RNA interference
RT-PCR	Reverse transcription PCR
SEM	Standard error of the mean
siRNA	Small interfering RNA
SNP	Single Nucleotide Polymorphism
Sp1	Specificity protein 1
Sp3	Specificity protein 3
TGFβ	Transforming Growth Factor Beta
TIMP	Tissue Inhibitor of Metalloproteinases
TMT	Tandem Mass Tag
WT	Wildtype
YY1	Yin Yang 1

Chapter 1: General Introduction

1.1 Arthritis

Arthritis is a term that describes the breakdown of the normal functioning of the joint and literally means painful inflammation and stiffness of the joint. This term encompasses a spectrum of musculoskeletal conditions in which the articular surfaces of the joint no longer have the ability to function properly, resulting in loss of joint stability and almost always pain. These conditions have variable presentations and patient outcomes. There are over 200 types of musculoskeletal conditions and typically these can be categorised into five groups; inflammatory arthritis such as rheumatoid, non inflammatory or degenerative arthritis such as osteoarthritis (OA), soft tissue musculoskeletal pain, back pain and connective tissue diseases such as lupus [1, 2].

1.1.1 Osteoarthritis and its Prevalence

OA is the most common musculoskeletal disorder in the world with 80% of individuals aged over 75 years in Western populations showing radiographic evidence of the disease [3]. An estimated 8 million people in the UK have OA and this high prevalence is ever increasing due to the ageing of the population and the increasing incidence of obesity [1]. Characterised by the progressive loss of articular cartilage, OA is a chronic, highly disabling disease that can cause a significant burden on individuals. There is substantial variation in the clinical presentation and outcomes of OA, owing to the large number of risk factors and pathologies associated with the disease. This means that it is not possible to treat the disease as a single disorder, instead, OA is considered as a heterogeneous group of disorders which may have differing aetiologies [4, 5]. In older patients, OA is one of the most common diagnoses in general practice medicine and is a leading cause of disability [6].

In addition to a functional disability and reduced quality of life, patients may also suffer financially as a result of loss of wages and increased medical costs. OA constitutes a large economic burden on society, amounting to 1-2% of gross national product [7]. This is a result of medical expenses (medication, physiotherapy and surgery) and of indirect costs such as time off work, loss of productivity whilst at work, and increased resources required for informal care giving [8].

1.1.2 Defining OA

Osteoarthritis can be characterised into primary OA, with cause unknown, or secondary OA, where the cause can be identified; for example, arising from a developmental disorder, trauma or metabolic syndrome [4]. Additionally, the number and distribution of the joint sites that are affected can be used to classify OA; OA can be joint specific, known as localised disease, with knees and hips tending to be the most commonly affected, or it can be more generalised, affecting a number of joints [9]. OA is usually defined by structural changes that are detailed radiographically; in end stage disease structural changes can include joint space narrowing, subchondral bone sclerosis, subchondral cyst and osteophyte formation [3].

Grading systems are available to define the severity of OA, with the Kellgren and Lawrence (KL) scale being most commonly used to grade disease severity on a scale of 0-4; a score of 0 signifies absence of OA changes and a score greater than 2 signifies definitive radiographic OA [10]. Magnetic Resonance Imaging (MRI) can also be used to examine joint structures and is suggested to be more sensitive and enables a more precise tracking of the disease in comparison with tracking joint space narrowing radiographically [11]. Symptomatic OA defines the presence of radiographic changes in addition to joint symptoms. The most common OA symptom is pain; patients can also experience joint stiffness, tenderness, functional limitation, variable inflammation and crepitus [12]. The disability associated with these symptoms can impact upon a patients quality of life and make everyday tasks difficult, which can lead to social isolation and depression [13]. Furthermore, evidence is now emerging that patients with OA have an increased risk of mortality [14].

The degree of pain, joint stiffness and loss of physical function that patients experience can be quantified using a numerical index such as the Western Ontario McMaster Osteoarthritis Index (WOMAC) [15]. However, presence of pain is not necessarily correlated with radiographic evidence of disease and research has shown that there is discordance between these features; a population-based study found that significant structural changes were asymptomatic in over 50% of subjects, whereas only 15% of patients experiencing knee pain had KL stage 2-4 OA changes. This further complicates the definition and diagnosis of OA, with a number of physicians wrongly diagnosing OA in patients without radiographic evidence of OA, or patients with radiographic changes being missed due to lack of symptoms [16].

1.1.3 OA Treatment

Therapeutic interventions for OA are limited; current management of the disease involves non pharmacological treatments such as exercise and lifestyle changes in addition to pharmacological treatments such as non steroidal anti-inflammatory drugs (NSAIDS) and analgesics. Although these treatments can reduce pain and decrease inflammation, they do not target the underlying cause of the disease and have little impact on slowing the progression to advanced OA. Clinical trials testing potential OA disease modifying drugs have had limited success, with arthroplasty remaining the basis for curative therapy [17, 18].

Acetaminophen is given initially as a first line oral analgesic followed by NSAIDs if required [17]. NSAIDs primarily target the enzyme cyclooxygenase (COX), an enzyme responsible for the formation of biological mediators of pain, the prostanoids. There are two COX isoforms, COX1 and COX2, and inhibitors against these enzymes have been used in the treatment of arthritis [19, 20]. Classical COX inhibitors are not selective and inhibit both isoforms of the enzyme. Although they demonstrated clinical efficacy, they were associated with adverse gastrointestinal (GI) side effects, prompting the production of the next generation of COX inhibitors which were specific for COX2 inhibition. The adverse side effects are a result of the widespread expression of COX1 compared to the limited expression of COX2 which is located at sites of inflammation and in the kidneys [19]. COX2 specific inhibitors (such as rofecoxib) became the preferred choice for treatment and lacked GI complications, however the association of rofecoxib with increased risk of myocardial infarction and stroke prompted its withdrawal from the market and raised concerns as to the safety of these drugs [21]. For patients who are contraindicated with COX inhibitors, opiod analgesics can be used. Topical NSAIDs can also be used, providing a local high dose of drug whilst minimising systemic side effects. Another option for OA pain, most commonly for use in the knee, is intra-articular injections of corticosteroids, although there are concerns that this treatment may enhance OA progression [22].

Surgery is usually considered in patients who have severe disease or in patients who are suffering from pain and disability, and who are showing no improvement following pharmacological intervention. The surgical options available include arthroscopy (which may involve debridement and lavage or small corrective surgery), osteotomy (which involves the re-alignment of the joint), joint fusion for the spine, hand and foot

and finally arthroplasty (which involves total joint replacement and is usually the option for advanced OA cases) [23].

Currently, the focus for treatment research is to identify disease modifying OA drugs (DMOADs). Trials examining the efficacy of substances such as glucosamine sulphate and chondroitin sulphate have concluded that these compounds have structure modifying properties [24]. A two year, randomised placebo controlled trial in the USA, the Glucosamine/Chondroitin Arthritis Intervention Trial (GAIT), has however provided evidence that these compounds do not significantly affect loss of cartilage or pain. These compounds had similar effects to those observed in subjects who were taking a COX2 inhibitor or a placebo [25].

Other approaches have focussed on targeting the pathophysiological changes that occur in OA, for example the targeting of catabolic enzymes such as matrix metalloproteinases (MMPs) or targeting factors that promote synovial hyperplasia such as the cytokine IL-1β. Trials testing MMP inhibitors have been complicated by painful musculoskeletal adverse effects such as tendonitis due to their broad protease inhibition. More selective inhibitors are under investigation to reduce the side effect risk [26]. An inhibitor of IL-1β, diacerein, has shown conflicting effects, with some studies reporting a positive anti-inflammatory effect, whilst others have reported no differences [27, 28].

Other options are to target bone remodelling or to try and promote cartilage repair. There is also interest in the use of mesenchymal stem cells for autologous transplantation. This involves the removal of stem cells from a patient, the differentiation of these cells *in vitro* (for example into chondrocytes) and then the transplantation of these cells back into the patient at the required site for new tissue growth [29]. Mesenchymal stem cells are found in the bone marrow and in many adult tissues and are capable of self renewal. Furthermore, these cells have the capacity to differentiate into a variety of cell lineages, including chondrocytes, adipocytes and osteoblasts [30, 31]. These cells therefore have the capacity to regenerate cartilage *in vivo* and thus in the future could be used for the treatment of OA. A recent study has further investigated the use of stem cells, by performing a high throughput screen to identify molecules which promote chondrocyte differentiation. Kartogenin was identified by this study and was found to induce chondrogenesis by regulating *CBFβ* and *RUNX1* transcription [32]. Furthermore, there is interest in generating new scaffolds for cartilage tissue engineering to promote cartilage repair.

In addition, growth factors such as TGFβ and members of the bone morphogenetic protein family are under consideration as potential OA therapeutics, although recent evidence suggests that the use of TGFβ superfamily members for bone tissue engineering has failed to duplicate in humans the promising results that were obtained in animal models [33].

1.2 Joint Structure and Function

Joints allowing free movement of the bone are known as diarthroses or synovial joints and comprise most of the joints of the body. The synovial joint is a complex organ that is composed of a number of tissue types surrounded by a fibrous capsule, as shown in Figure 1.1. These tissues include the articular cartilage, capsule, synovium, synovial fluid, tendon, ligament and subchondral bone. In addition, in joints such as the knee, the fat pad, menisci and patella are present. All of these tissues have specialised roles and are interdependent on each other, whereby damage or pathology in one tissue could adversely affect any one of these tissues. Together their composition allows the articulation and frictionless movement of joints, permitting pain free functioning. Damage to these tissues can alter the integrity of the joint, which may limit joint function, cause pain and could promote the onset of OA.

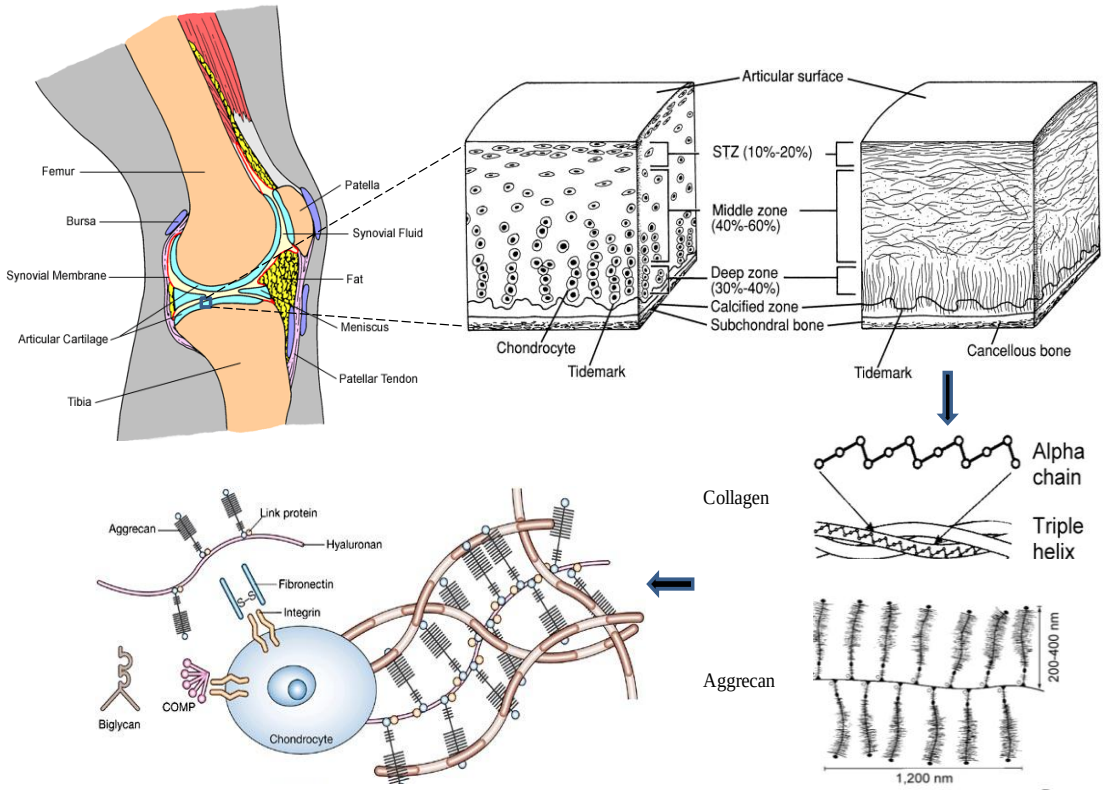


Figure 1.1
Structure of the synovial joint and composition of the extracellular matrix (ECM). The synovial joint is composed of a number of tissue types surrounded by a fibrous capsule. The articular capsule consists of the fibrous capsule and synovial membrane. The fibrous capsule surrounds a number of tissues including; articular cartilage, synovial fluid, meniscus, tendon, ligament, and the subchondral bone, covered by the periosteum. The articular cartilage consists of four different zones; the superficial zone (SZ), middle zone, deep zone and calcified zone. The triple helical structure of collagen is shown, as well as the structure of aggrecan, with its characteristic protein core, three globular domains and GAG chains. Finally, the composition of the ECM is shown; chondrocytes are surrounded by a matrix consisting mainly of collagen and aggrecan. Cartilage oligomeric matrix protein (COMP) and fibronectin (a high molecular weight glycoprotein) are also present within the ECM. The cartilage zones figure is taken from [34]. The collagen and aggrecan molecule figure is taken from [35].The ECM figure is from [36].

1.2.1 Articular Cartilage

Articular cartilage is the smooth coating covering the ends of skeletal bones, acting as a load bearing material, functioning to reduce friction and to help absorb shear and compressive stresses [37]. The articular cartilage contains a single unique cell type called the chondrocyte. These large and mature cells are able to function in a hypoxic environment that is avascular, alymphatic and aneuronal. Chondrocytes derive their nutrients from the surrounding synovial fluid and importantly, these cells are not inert; they synthesise, secrete and maintain their surrounding extracellular matrix (ECM).

Although chondrocytes make up only a small proportion of the cartilage (5%), their metabolism is responsible for the presence and maintenance of their surrounding matrix. The balance between the catabolism and anabolism of this matrix, mediated by the chondrocyte, is crucial for cartilage homeostasis [38, 39].

The ECM consists of proteoglycans (aggrecan being the most abundant) surrounded by a network of collagen (mainly type II) and water. Collagen is the most widely expressed protein within the body and is vital in providing cartilage with its structural properties. Collagen molecules have a characteristic triple helical structure and align to form fibrils that are stabilised by intramolecular crosslinking [39]. Aside from collagen type II, type IX and XI are also important within the ECM. Aggrecan constitutes the major non-fibrillar component of the ECM; it consists of a protein core, with three globular domains, G1, G2 and G3. G1 binds non covalently to hyaluronic acid, stabilised by a link protein. Between the G2 and G3 domains there are many glycosaminoglycan (GAG) chains including chondroitin sulphate and keratin sulphate [40]. The structures of collagen and aggrecan are shown in Figure 1.1 in addition to other components of the ECM.

Aggrecan polymers are maintained in an ordered state by their interaction via the link protein with collagen fibres. This network is surrounded by water which is drawn into the matrix by the negatively charged aggrecan polymers. The water provides a swelling pressure that is constrained by the collagen. Together these molecules provide cartilage with its structural properties, allowing it to resist compression and remain firm and intact.

Cartilage can be separated into four zones; the superficial, middle, deep and calcified. The superficial zone (representing 10-20% of cartilage thickness) is in contact with the synovial fluid and offers a smooth surface for articulation. This zone has the highest collagen content and is highly organised, enabling it to withstand shear forces. The chondrocytes within this zone are small and flat, primarily expressing proteins for lubrication and protection [41]. The middle zone (representing 40-60% of cartilage thickness) is less organised as a result of the lower collagen content and the chondrocytes are rounder, larger and more sparsely spread. The deep zone (representing 30% of thickness) has a high proteoglycan content. The chondrocytes in this zone are arranged in columns, parallel to the collagen fibres [39]. The tidemark separates the deep cartilage zone from the calcified cartilage which lies directly on the subchondral

bone [37]. Collagen fibrils extend from the calcified cartilage into the articular cartilage providing a strong link between the two zones. The calcified cartilage zone contains type X collagen, a marker of chondrocyte hypertrophy, and is a region of active matrix turnover.

Water provides cartilage with its resilient properties and is important in providing cartilage with nutrition and lubrication, constituting 65-80% of tissue composition [42]. The largest amount of water is found in the superficial zone near the surface of the cartilage; as depth increases, the concentration of water decreases. The zonal arrangement of the articular cartilage and the arrangement of collagen is shown in Figure 1.1.

1.2.2 Subchondral Bone

The subchondral bone is composed of two regions, the subchondral bone plate and the trabecular bone. The subchondral bone plate separates the calcified cartilage from the bone marrow spaces and consists of a thick dense layer of calcified tissue which supports cartilage, transmitting load into the trabecular bone beneath. The trabecular bone is a network of thin, calcified trabeculae consisting of plates and struts, orientated in different directions to form a supportive network with a primarily metabolic function. These two differences in function relate to structural differences; a large volume (80-90%) of the subchondral bone end plate is calcified, enabling a strong support structure for load, in comparison to a smaller volume (15-25%) of the trabecular bone [43].

Blood vessels from the subchondral bone extend into the calcified cartilage layer through channels in the subchondral bone plate. These channels serve as a source of nutrition for the chondrocytes located in the calcified zone. Both regions of the bone are composed of an organic matrix consisting primarily of collagen type I, but also non-collagenous proteins such as osteocalcin. In addition, both regions are mineralised with inorganic crystals of hydroxyapetite. The collagen in the end plate is arranged into sheets of parallel fibrils which extend into the lamellae of the trabecular bone. The calcified bone matrix is not inert, it contains cells capable of bone formation (osteoblasts) and bone resorption (osteoclasts). Most components of the bone matrix are synthesised and secreted by osteoblasts, osteoclasts are derived from osteoblast cells and are key for growth, bone remodelling and repair [44].

1.2.3 The Joint Capsule and the Synovial Membrane

The joint capsule is a dense, irregular connective tissue which physically holds adjacent bones together, providing support and structure. The synovial membrane lines the inside of the capsule and contains two cell types; the type A synovial cell, a macrophage protecting against infection, and the type B synovial cell, which secretes synovial fluid. These cells are responsible for maintaining homeostasis of the joint. Synovial fluid contains both hyaluronic acid and lubricin which reduce friction, absorb shock and facilitate the exchange of oxygen and nutrients with carbon dioxide and metabolic waste in chondrocytes [45]. Synovial fluid also contains mesenchymal stem cells which are present in preparation for repair of damage [46].

Tendons and ligaments are both connective tissues that attach alongside the capsule to the bone. Tendons are composed of predominately type I collagen fibres that connect muscle to bone and allow motion of the joint. Ligaments are composed of type I (90%) and type III (10%) collagen fibres in addition to a small amount of proteoglycans and connect bone to bone to help stabilise the joint [47]. The fat pad acts as a cushion to protect the joints and is also a rich source of proinflammatory adipokines and mesenchymal progenitor cells capable of differentiation [48]. Finally, the menisci are pads of fibrous cartilage that maintain stability of the joints and dissipate load [49].

All joint tissues, with the exception of articular cartilage, are innervated by nerves responsible for proprioception and nociception. Blood vessels penetrate the capsule and provide nutrition, whilst lymphatic vessels remove waste [50].

1.3 Joint Homeostasis

Joint homeostasis requires a fine balance between the anabolic and catabolic activity of chondrocytes. The activity of chondrocytes is influenced by biochemical mediators such as growth factors and cytokines, and also mechanical factors such as loading and stress. Inflammatory cytokines can mediate the catabolic activity of the cells, whilst anti-inflammatory cytokines can promote anabolic activity [51]. The most important growth factors found to be present within the joint during development include transforming growth factor β (TGFβ), bone morphogenetic proteins (BMPs), cartilage-derived morphogenetic proteins (CDMPs), connective tissue growth factor (CTGF), insulin-like growth factors (IGFs), hepatocyte growth factor (HGF) and fibroblast growth factor (FGF) [37].

Matrix metalloproteinases are the enzymes responsible for the breakdown of cartilage components. These proteases are divided into groups based on the substrates that they cleave and their location within the cell and include: collagenases (MMP 1, 8 and 13), membrane type MMP (MT-MMP), stromelysins (MMP 3, 10 and 11), gelatinases (MMP 9 and 2) and adamalysin (ADAM and the subgroup ADAMTS) [37]. Multiple members of the MMP family act by cleaving collagen whereas aggrecan is cleaved by both MMPs and aggrecanases (ADAMTS-4 and 5 in OA cartilage) [52]. These factors are important for cartilage homeostasis and mediate processes such as repair and remodelling of the joint. However, they have also been found to contribute to the pathology of joint disease [53, 54].

MMPs are potent catabolic enzymes and as such their activity must be regulated. MMPs are produced in a pro-enzyme form that requires the removal of an N-terminal fragment for activation. Pro-MMPs can be cleaved by a number of factors including activated MMPs, plasmin and reactive oxygen species (ROS) [55]. Further control of the MMPs is achieved through the synthesis of tissue inhibitors of metalloproteinases (TIMPs). TIMPs are synthesised in connective tissue cells and their expression is regulated by growth factors. They inhibit MMPs by two mechanisms, firstly by binding to them and blocking MMPs substrate cleavage and secondly by inhibiting the cleavage of pro-MMPs, and thus their activation [56].

In addition to changes to the cartilage, bone remodelling occurs in a coordinated sequence to replace old bone with new bone, enabling the shape, architecture or density of the skeleton to be changed. Bone formation occurs once bone resorption has taken place in an activation-resorption-formation sequence. This sequence ensures that in non disease states there is an equal amount of bone resorption and bone formation to maintain bone mass [43].

1.4 Development of the Joint

The epiphyses are the regions at the end of long bones, the diaphysis is the hollow portion in the middle of bone and the metaphysis is the transition zone between them (Figure 1.2). The epiphyses is derived from a different ossification centre to the metaphysis and diaphysis and during development these two regions are separated by a layer of cartilage, known as the epiphysis cartilage (growth plate). This region of cartilage matrix enables the growth of long bones until the end of the growth period. Bone development occurs through two distinct processes, intramembranous and

endochondral ossification. In the first, mesenchymal cells differentiate directly into osteoblasts which synthesise a woven bone matrix. Blood vessels are incorporated between the woven trabecular bone and form bone marrow. At a later stage, this woven bone is replaced with mature lamellar bone. This process does not involve the presence of cartilage [43].

Endochondral ossification is a two step process involving the differentiation of mesenchymal stem cells into chondrocytes, which create the cartilage matrix in a process known as chondrogenesis, followed by the formation of bone by osteoblasts. Chondrogenesis begins with the recruitment, migration and proliferation of mesenchymal stem cells. These cells aggregate into pre-cartilage condensations and start to produce the cartilaginous matrix for the developing limbs. This matrix consists mainly of collagen type II. In the centre of the cartilage model, chondrocytes proliferate and become hypertrophic at a region forming the primary ossification centre. The differentiation of chondrocytes is controlled by a number of factors; parathyroid hormone related peptide (PTHrP) is found in prehypertrophic chondrocytes and maintains these cells in a proliferative state. PTHrP expression and chondrocyte differentiation is also regulated by other factors such as Indian hedgehog (IHH), Fibroblast growth factor (FGF) and bone morphogenetic proteins (BMPs) [57]. The transcription factors Sox9 and Runx2 in addition to factors of the wnt signalling pathway control the differentiation of the chondrocytes within the chondrocyte lineage [58-60].

A ring of woven bone called the bone collar is formed at the periphery of the mid zone area, which following calcification is penetrated by blood vessels and osteoclasts which enter into the primary ossification centre. The blood vessels allow seeding of the hematopoietic bone marrow. The osteoclasts resorb calcified cartilage and allow osteoblasts to form woven bone. Secondary ossification centres begin to form at the epiphysis and form trabecular bone in a similar process. Between these two ossification centres, the epiphyseal cartilage remains until growth has finished in adulthood [43].

From the epiphysis to the metaphysis, the growth plate demonstrates the different stages of chondrocyte differentiation; firstly, the proliferative zone in which chondrocytes are synthesising cartilage matrix, secondly, the hypertophic chondrocytes which deposit a mineralised matrix, and lastly, after calcification of matrix the chondrocytes undergo

programmed cell death (apoptosis) allowing for the formation of bone by osteoblasts [43]. The process of endochondral ossification is shown in Figure 1.2.

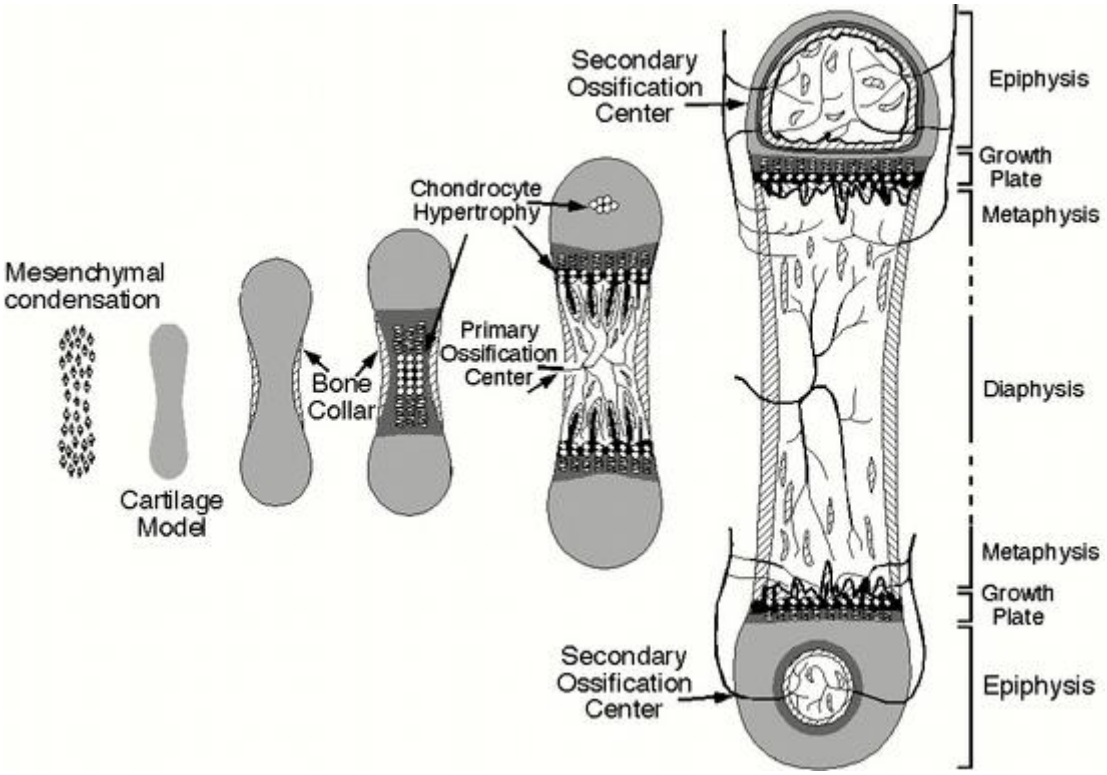


Figure 1.2
The process of endochondral ossification: Mesenchymal stem cells aggregate and differentiate into chondrocytes. The chondrocytes form the cartilage model by proliferating and secreting extracellular matrix (ECM) components. A ring of woven bone is formed called the bone collar at the periphery of the mid zone area following the secretion of osteoid by osteoblasts serving as a support for the new bone. Chondrocytes continue to proliferate, stop secreting ECM components and start to become hypertropic at the primary ossification centre which following calcification is penetrated by blood vessels. Chondrocytes undergo apoptosis, leaving cavities in the bone. Osteoblasts form the trabecular bone using the calcified matrix as a scaffold. Secondary ossification centres begin to form at the epiphysis (ends of long bones). The growth plate is epiphyseal cartilage that remains until growth has ended in adulthood. Figure taken from [43].

1.5 Pathology of OA

OA is not solely a disease of the articular cartilage, pathological changes can be seen in all tissues of the joint [61]. Although there are notable changes in the articular cartilage in the early phases of the disease, the bone, ligament and synovium also show evidence of damage. It is unclear as to where the initial pathological changes occur; some argue that the subchondral bone is the primary source of damage, although others argue that changes in the articular cartilage precede this [62]. As the disease progresses,

pathological changes occur in a number of tissues of the joint, including tendons, muscle and the nervous system, either as a direct effect or indirectly [63, 64]. The pathological changes that occur are the consequence of a number of different aetiological factors and are dependent on the stage of the disease.

1.5.1 Articular Cartilage

There are two mechanisms by which damage can occur to the articular cartilage, the first is the abnormal loading on normal cartilage, whereby an excessive load can damage the structure of the ECM. The second is the result of normal loading on abnormal cartilage and the risk factors associated with the development of abnormal cartilage are discussed further below [65]. In either case, the end result is structural damage to the cartilage and the inability to reform the complex structure of the ECM.

The pathological changes that occur in the articular cartilage during OA progression are striking. The normal smooth, elastic, layered cartilage texture following insult can become rough, fragmented and thinned or can be absent all together in the late stages of the disease. Macroscopically, cartilage changes in OA are observed as fibrillation (discontinuities in the cartilage surface), which can extend to become fissures (splits), chondromalacia (softening), erosion of cartilage and ulceration. Other histological changes include changes in cartilage thickness, the formation of cartilage clefts and a duplication of the tidemark between non calcified and calcified cartilage. Tidemark duplication and vascular invasion of vessels penetrating into the cartilage from the subchondral bone further decreases the thickness of the articular cartilage [39]. In areas where cartilage is absent, the bone is exposed and undergoes sclerosis, a process known as eburnation due to its shiny ivory like appearance [22].

Biochemically, the major components of the ECM are altered, changing the content, composition and structure of the matrix. The two key components of the ECM, collagen and aggrecan, are affected; cleavage of type II collagen and local loss of proteoglycans occurs initially at the cartilage surface resulting in an increased permeability of the matrix. Although there is an increase in the water content of the ECM, this increase in permeability results in a decrease in hydrostatic pressure, causing softening of the cartilage. The tensile strength of the cartilage is decreased and it becomes stiffer. A higher percentage of proteoglycans are present in the non aggregated form, unbound to hylauronate, and the chain length of proteoglycans is reduced. With disease progression,

there is a rapid loss of proteoglycan content relative to the amount of collagen. Furthermore, the organisation of collagen is affected [39].

Changes to the chondrocytes can also be observed; in the early stages of the disease, chondrocytes located in the superficial zone proliferate to form aggregates, believed to be replicative clones. As the disease progresses, the chondrocyte clusters increase in size and can be seen deeper into the cartilage zones. In addition, there are areas of decreased cellularity where chondrocytes have undergone cell death [66].

The reasons for these changes in the composition of articular cartilage are a result of chondrocyte activation; in normal cartilage, chondrocytes remain in a quiescent state, exhibiting a very low metabolic activity. In response to load or injury, chondrocytes become 'activated', characterised by proliferation and increased production of catabolic and anabolic factors. An increase in catabolic activity, fuelled by increases in pro-inflammatory cytokines and proteases, leads to a failure of the chondrocyte to maintain its environment. Levels of the collagenases MMP1 and MMP13 have been found to be increased in OA cartilage and are responsible for the degradation of type II collagen [67]. Furthermore, MMP3 and ADAMTS-4 degrade proteoglyans. There is an imbalance in the levels of MMPs compared with TIMPs in OA tissues resulting in destruction of the ECM and the release of matrix fragments [68]. Chondrocytes and synoviocytes respond by releasing growth factors in an attempt to initiate developmental pathways to remodel the matrix. The phenotypic changes in OA chondrocytes vary according to their location; chondrocytes at the superficial zone exhibit an osteoblastic phenotype, cells in the middle and deep zones however exhibit a hypertropic phenotype [69]. Overall, the chondrocytes are unable to maintain an adequate level of anabolic activity and cannot replicate the ordered structure of the matrix formed during development, leading to destruction and eventually complete loss of the articular cartilage [70].

1.5.2 Subchondral Bone

Pathological changes can be observed in the subchondral bone during the OA disease process. These changes include sclerosis, the development of osteophytes and the formation of bone cysts. Sclerosis of the trabecular bone can occur due to increased stress on the bone which causes remodelling, in turn increasing the trabecular bone volume. The increased volume is due to an increase in trabeculae number and thickness, reducing the space between trabeculae, thus altering the organisation of the structure

and the biomechanical properties of the bone. Furthermore, there is evidence of thickening of the subchondral bone plate [71]. Osteophytes arise at the joint margins, mainly located in non weight bearing areas of the joint, and form due to increased bone formation. It is still a debate as to whether osteophytes are damaging to the joint, some believe that they add to joint pathology, contributing to OA pain, whereas other researchers suggest that they help to stabilise the joint and redistribute load [72]. The absence of osteophytes in the hip joint of OA patients has been shown to increase disease progression [72]. Clinical studies have however positively correlated the extent of osteophyte formation with cartilage destruction [73]. In the advanced stages of the disease, subchondral bone cysts usually form in areas where cartilage has been completely lost. These are cavities within the bone and are formed due to the action of osteoclasts [74].

1.5.3 Synovium

OA is not classically known as an inflammatory disease, although mild synovial inflammation is common in the disease [75]. Signs and symptoms of inflammation such as joint pain, swelling and stiffness occur in early and advanced OA. Changes are confined to the synovium and include synovial hyperplasia, fibrosis and inflammatory cell infiltration. Inflammation is mediated by synovial mast cells and chondrocytes in addition to infiltrating macrophages, B and T cells. In contrast to rheumatoid arthritis, this inflammation is a secondary effect, a result of pathological changes within the joint [76].

The release of inflammatory mediators is usually in response to the release of degradation products such as proteoglycan fragments and type II collagen peptides into the synovium and the subarticular bone which elicit an inflammatory response. Synovioctes become activated, proliferate and secrete MMPs and cytokines which promote a catabolic response. Cytokines such as IL-1β, TNFα, oncostatin M (OSM) and interleukins 6, 17 and 18 can all affect the metabolic activities of chondrocytes, promoting an increase in the synthesis of degradative enzymes including MMPs and aggrecanases [37, 77]. IL-1β also induces other pro-inflammatory cytokines and chemokines and suppresses the expression of anabolic factors. Many of the catabolic factors synergise with each other promoting further catabolism, favouring catabolic changes over anabolic changes, resulting in significant and irreversible damage to the ECM [78].

1.5.4 Other tissues of the Joint

There are also changes in other tissues of the OA joint, although these have been less well studied in comparison to the aforementioned tissues. The meniscus can show degenerative changes such as horizontal cleavages, complex tears or destruction. In advanced stages of the disease, calcification within the menisci is common [79]. Thickening and fibrotic shortening of the capsule have also been reported in OA pathology and are associated with symptoms of pain [80]. Degenerative changes may occur at the site where the ligaments insert into bone and additionally muscle atrophy (mainly in the quadriceps) can occur [81, 82].

1.6 Risk Factors for OA

The major risk factor for OA is age, with prevalence of the disease rising with increasing age. Other important risk factors include genetics, obesity, gender, race/ethnicity, nutrition, bone metabolism in addition to joint specific risk factors including mechanical injury or trauma, bone density, muscle strength, leg length inequality (LLI), joint morphology/alignment, occupation and sporting activities [3]. These can be separated into those that promote the formation of abnormal cartilage, and thus predispose to the development of OA, and those that increase susceptibility to OA based on injury risk or trauma to the cartilage.

Whilst age is one of the primary risk factors for the development of OA, it is important to note that although changes in the ECM structure and function with increasing age are inevitable, the development of OA is not. Not all elderly people develop OA, it is therefore clear that there must be other important factors affecting susceptibility. With increasing age, the articular cartilage becomes softer and there is a decrease in the strength of the ECM, making it more susceptible to injury. There are changes in the key components of the ECM with increasing age, but these are distinct from the changes observed in OA pathology [65]. With increasing age, the size and number of aggrecan molecules is decreased, a result of decreased aggrecan synthesis or increased degradation, and this also leads to a decrease in the number of chondroitin sulphate chains [83]. Crosslinking of collagen molecules in both the bone and cartilage is increased with age [84, 85]. This is in contrast to the changes observed in OA cartilage, which as mentioned previously include increased levels of non aggregated aggrecan, and an alteration of the organisation of collagen [39].

With increasing age, there is also an increase in the accumulation of advanced glycation end products (AGEs) and their receptor RAGE. AGEs are the end products of a process

called non-enzymatic glycation. Their accumulation with age can affect the activity of the chondrocytes, and trigger the expression of cytokines and MMPs and the release of ROS [65, 86].

In addition to changes in the structure and composition of the ECM, the anabolic capacity of the chondrocyte is reduced with age, with chondrocytes being less responsive to growth factors [86]. This is known as senescence; senescent chondrocytes display aberrant expression of catabolic factors and are more responsive to cytokines [87]. Overall, the risk of OA is increased with ageing because the altered composition of the matrix renders the cartilage more sensitive to load/trauma and the ability of the chondrocytes to maintain cartilage homeostasis and repair following injury becomes compromised.

Female sex is associated with both higher prevalence and increased severity of OA [88]. The increased prevalence of OA at the time of the menopause led to the hypothesis that oestrogen levels may affect OA susceptibility. However, the results from clinical trials assessing the effect of oestrogen replacement therapy on the development of OA have been conflicting [89-91]. An MRI study investigating sex differences for prevelance for OA suggested that women may have thinner cartilage than men but it is not yet clear if cartilage loss is accelerated [92].

Loading can also affect the homeostatic balance within the ECM; chondrocytes respond to load by either up-regulating the synthesis of matrix components or increasing the levels of inflammatory mediators responsible for cartilage breakdown. Static compression has been found to cause degradation of matrix components; conversely, dynamic compression stimulates the synthesis of these components [93].

Obesity is a risk factor for the development of OA, with overweight or obese patients being 3 times more at risk of incident knee OA compared to normal weight individuals. Furthermore, risk rises with increasing body mass index (BMI) [94, 95]. Women who reduced their BMI by 2 or more units over a 10 year period decreased their risk of developing OA by 50% [96]. However, there is an increased incidence of obesity with hand OA; this association prompted further investigations as this suggests that the effect is not primarily a result of increased loading on the joint, but may be the result of metabolic or inflammatory factors [97]. Adipokines can be released from chondrocytes and directly affect cartilage homeostasis, acting locally to increase catabolic activity.

Inflammatory mediators increase the expression of adipokines such as leptin which then promotes further catabolic effects by increasing the expression of MMPs [98].

The effects of dietary factors on OA susceptibility have been conflicting. Increasing intake of vitamins C, D, E and K have all been tested and the results have been inconsistent [99]. Furthermore, bone mineral density (BMD) has been associated with OA risk, such that increased BMD increases OA susceptibility. However, this association is not understood; it is unclear if it is related to changes in bone remodelling or to increased load rather than a direct association of the BMD [100].

Joint specific risk factors, include occupation, injury and physical activity. Repeated use of the joint such as for squatting, an occupation involving kneeling or lifting or repeated physical activity can increase susceptibility to OA [101]. In addition, muscle strength may be a predisposing factor, although it is not clear as to whether muscle atrophy is a determinant of OA or a consequence of OA related disuse [102]. Structural properties of the joint can also affect susceptibility to OA. The alignment of the joint is a key example, whereby misalignment of the joint can result in abnormal load distribution and thus altered cartilage morphology [103]. In addition, LLI can affect susceptibility, such that a LLI of more than 2cm can alter the biomechanics of the lower extremities [104]. Lastly, joint morphology is also a predisposing factor, altered anatomy of the shape of the joint can affect load distribution and therefore render the joint susceptible to damage following loading [105].

Race and ethnic differences have been shown to impact upon OA prevalence. A cross sectional study observed racial differences in hip OA between Caucasians and African Americans in the Johnston County OA project, with an increased incidence of hip OA in African Americans [106]. In addition, a Chinese cohort was found to be less susceptible to hand and hip OA compared to a Caucasian population. However, knee OA was more prevalent among Chinese females compared to Caucasian females [107-109].

1.7 Genetics of OA

1.7.1 Family Aggregation Studies

Epidemiological studies have provided evidence that OA has a genetic component. The first of these studies, highlighting the familial clustering of OA, was in the 1940s and involved a family aggregation study. These studies investigate the genetic contribution

to a disease by looking at relative risks for a family member of an affected individual (for example a sibling) and comparing this risk to that of the general population. This first study identified that mothers and sisters of patients with Herberden's nodes (a sign of OA in the hands) were disproportionately affected by the same boney swelling [110] . Further studies have investigated the familial clustering of OA, confirming a genetic influence in hand OA and at other sites including the knee, hip and spine [111-117]. Family aggregation studies have provided a good basis for evaluating the genetic influence in OA risk, however they do not allow for the separation of clustering that is due to shared environmental factors from genetic factors. Factors such as diet and exercise may cluster in families, and it is therefore difficult to account for this within such studies [118].

1.7.2 Twin Studies

The investigation of disease clustering in monozygotic (MZ) and dizygotic (DZ) twins allows for the separation of environmental factors from genetic factors. MZ twins have identical genomes, meaning that any variation between pairs is a result of environmental influences, whereas DZ twins share on average about 50% of their genes, and so variation may be accounted for by environmental or genetic factors. Comparing the concordance of disease in these twin pairs therefore allows for the quantification of genetic and environmental contributions to disease [118]. The relative contribution of genetic and environmental factors in hand and knee OA was assessed in female MZ and DZ twins, identifying the influence of genetic factors as between 39-65% [119]. Furthermore, a longitudinal twin study has identified that prevalence of progression of knee OA was higher in MZ compared to DZ twins. The concordances for osteophyte formation and joint space narrowing were also higher in MZ compared to DZ twins, with heritability estimates of 69% and 80% respectively [120]. The influence of genetic factors in hip OA has also been investigated, with greater concordance in MZ compared to DZ female twins translating to heritability estimates of 58% for hip OA overall and 64% for joint space narrowing [121]. Furthermore, in a male twin pair study, the concordance levels of radiographic disease was higher in MZ compared to DZ twins [122]. Heritability of OA in the spine has also been investigated, with a study looking at disc degeneration in twins using MRI. Overall, the heritability was 74% at the lumbar spine and 73% at the cervical spine [123]. These studies have highlighted that OA is not a disease that is mediated by fully penetrant risk alleles. Conversely, polymorphisms within a number of genes are responsible for increasing an individual's risk of

developing OA. One allele may therefore only contribute to a fraction of the overall risk of OA, but in addition to other alleles and in combination with environmental factors, OA susceptibility is increased. Genetic factors account for at least half of the variation in susceptibility to disease. Furthermore, there is heterogeneity between different joint sites and between sexes, adding to the complex nature of the disease.

1.7.3 Candidate Gene Studies

Following the studies confirming that hereditary factors contribute to OA susceptibility, investigators then started to interrogate the molecular genetic nature of the disease, with the hope of identifying loci that harbour OA susceptibility. The first approach was to investigate candidate genes that are thought to be involved in OA pathology based on their expression, protein function or association with a related disease. Genes encoding structural components of the ECM such as *COL2A1* represented reasonable candidates given their function within the development of the joint, and considering that when mutated they are associated with developmental abnormalities such as chondrodysplasias [124]. Association and linkage studies within these genes have however been inconclusive or negative, suggesting that structural genes are unlikely to be conferring OA susceptibility, and instead are associated with rare monogenic disorders. Furthermore, genes encoding proteins that regulate bone density were predicted to affect OA susceptibility and the *VDR* gene encoding the Vitamin D receptor was investigated and produced mixed results [125-127].

Aside from structural components of the ECM, candidate genes involved in cartilage development and signalling have also been investigated. A variable number tandem repeat polymorphism in the Asporin *(ASPN)* gene was identified by a Japanese group and was found to inhibit chondrogenesis and to increase susceptibility to OA. *ASPN* was investigated as it is a member of the small leucine-rich proteoglycan (SLRP) family. SLRPs function as structural organisers, interacting with components of the ECM and modulating growth factors by binding to TGFβ [128]. The D14 allele (14 aspartic acid repeats) showed an increased frequency in cases and the D13 allele (13 aspartic acid repeats) showed an increased frequency in controls in an Asian cohort. However, this effect was not consistently observed in replication studies using samples from the UK, Greece, Spain and China. A meta analysis combining all five studies reported an association of the D14 allele in OA cases [129-132]. Significant heterogeneity was observed and it was evident that the effect of the D14 allele is much stronger in Asians in comparison with Europeans [133].

The same group also identified a functional polymorphism rs143383 within the growth differentiation factor 5 gene (*GDF5*), which to date still remains the most compelling candidate OA association signal. In this study, it was found to be associated with knee and hip OA in a Japanese population [134]. This protein was investigated as it is a member of the TGFβ superfamily and is involved in the formation of the joint and its mutation results in a number of human dysplasias. GDF5 is discussed further in Introduction section 1.9.

Although there have been significant positive findings using the candidate gene approach, this approach is limited to our understanding of the disease, and important genes conferring OA risk may be missed due to incomplete knowledge of the OA disease process. Research has thus moved away from the hypothesis-led candidate gene approach towards a hypothesis-free genome wide approach [135]. Regions of the genome that harbour OA susceptibility loci can be identified using genome-wide linkage scans and genome-wide association scans (GWAS).

1.7.4 Linkage Studies

The aim of the linkage studies is to pinpoint areas of the genome that co-segregate with a disease phenotype such as OA. In these studies, the genomes of affected sibling pairs (ASP) are compared in order to look for regions that overlap. The first OA genome wide linkage scan was performed by a group in Oxford using patients ascertained by total hip or knee replacement. Linkages to chromosomes 2 and 11 were identified concordant for hip OA [136, 137]. These results were then stratified according to sex and by joint type which uncovered additional linkages on chromosome 4, chromosome 6 and chromosome 16 [138] . As the linkage intervals in the genome scan were relatively large, these loci were then subjected to finer linkage mapping [139-142]. The linkage region on chromosome 6 encompassed two candidates, *COL9A1* and *BMP5*, however subsequent studies failed to provide an association for single nucleotide polymorphisms (SNPs) within these genes [141].

Further investigation of the chromosome 2 region revealed an association with a functional non synonymous SNP in the *FRZB* gene with female hip OA [143]. *FRZB* encodes secreted frizzled-related protein 3 (sFRP3) which is a soluble antagonist of wnt signalling, shown to possess chondrogenic activity *in vivo* [144]. The SNP resulted in an arginine to glycine substitution, resulting in the diminished ability of sFRP3 to antagonise wnt signalling. A study of differential allelic expression (DAE) of *FRZB*

revealed that only a small number of patients had allelic expression differences and these did not correlate with genotype at five selected SNPs [145]. A number of replication studies have been performed, with association to knee OA in women being reported and a meta analysis confirming a strong association with the *FRZB* G allele [146]. However, a number of studies have been unable to replicate this association, including a large scale meta analysis [147, 148] and this locus was not found to be associated with radiographic osteoarthritic outcomes [149]. More recent studies have reported mixed associations of *FRZB* with hip morphology [150, 151]. Although the meta analysis results did not confirm a role for *FRZB* in susceptibility to common OA, this may be a result of the different criteria used for the definition of OA cases in these studies. A standardised approach for grading OA severe cases and identifying the presence of mild OA in controls has been suggested to progress research at this locus [148]. Functional studies of *frzb* knockout mice have supported a role for this factor in the development of the joint and have implicated FRZB in OA pathology; the mice showed increased cartilage proteoglycan loss, increased cortical bone thickness with increased stiffness and an increase in bone formation following loading [152].

Following the publication of the UK genome linkage scan, further scans followed. The first of these was performed using 27 Finnish families and investigated distal interphalangeal joint OA [153]. The linkage regions mapped to chromosomes 2 and 4, with the region on chromosome 2 encompassing the IL-1 receptor and ligand cluster at 2q12-q13. This region on chromosome 2 was distinct from that identified in the Oxford study. Considering the role of IL-1 in cartilage homeostasis as a catabolic cytokine, further studies investigated this association. Some have provided supporting evidence of the association at this locus in the knee, hip and hand [153-159], however some studies in different ethnic populations have been negative [160, 161].

A third scan was performed on an Icelandic family who suffered with severe hip OA necessitating joint replacement [162]. The strongest evidence for linkage occurred at chromosome 16, which overlaps with the region identified in the Oxford study. Following further study, the *IL-4* gene, located within this region was found to predispose to hip OA in female Caucasians [163].

Further genome wide linkage scans have focussed on hand and spine OA with a number of loci on different chromosomes showing segregation with disease [164, 165]. In the

Framingham study, two regions that have been previously identified showed linkage, one on chromosome 7 [153] and the second on chromosome 11 [136, 142].

1.7.5 Genome Wide Association Scans

GWAS are now more widely used in comparison with linkage studies due to their ability to resolve with greater accuracy areas harbouring susceptibility loci. GWAS compare the SNPs within a cohort of disease cases with a cohort of controls. By exploiting linkage disequilibrium, the fact that genetic variance at one locus can predict with high probability genetic variance at another locus, means that the number of SNPs that need to be genotyped in a GWAS (100,000-500,000) is considerably less than the total number of SNPs in the human genome. Upon identification of a number of SNPs more prevalent in cases compared to controls, each polymorphism within this region is interrogated with the aim of identifying the associated SNP. The completion of the human genome sequencing project (HGSP) and data from the International Human HapMap project have made the GWAS approach possible [166, 167].

There have been three large scale OA GWAS performed by Dutch, Japanese and UK groups. The Dutch GWAS involved a Rotterdam cohort, followed by replication in Europeans and North Americans [168]. The overall numbers combined were 14938 cases and 39000 controls. A highly significant signal (P value less than 8×10^{-8}) was reported to a region of chromosome 7q22. An analysis of additional cohorts in a meta analysis increased the significance of the association [169]. The 7q22 locus is particularly relevant to knee OA and contains at least six known genes, *PRKAR2B, HBP1, COG5, GPR22, DUS4L* and *BCAP29.* None of these genes appeared to be obvious OA candidates, although *GPR22* was found to harbour an expression quantitative trait loci (eQTL). However this eQTL was discovered in transformed lymphoblast cells and so it's relevance to OA pathology is unclear. Further investigation of this locus consisted of expression studies performed in joint tissues from OA patients, in tissues from a mouse model of OA and in zebrafish embryos. In patient tissues all of the genes showed a universal pattern of expression except for *GPR22* which could not be detected in any of the joint tissues. There was a lower level of expression of the 5 genes in OA cartilage compared to control cartilage. Further analysis found that *HBP1* and *DUS4L* demonstrated DAE, with carriers of the associated allele demonstrating a significantly different *HBP1* DAE compared to non-carriers [170]. *GPR22* was also absent in normal mouse articular cartilage and synovium. However, *GPR22* positive chondrocytes were found in mice with papain induced arthritis and in

chondrocyte like cells present within osteophytes, taken from animals with instability induced OA [168].

The Japanese GWAS involved Japanese cases and controls, followed by replication in both Japanese and European populations [171]. Two SNPs located on chromosome 6 within the region of the human leukocyte antigen (HLA) locus were associated with knee OA with P values of less than 7×10^{-8}. Neither of the SNPs showed an association in the European cohort and this association failed to replicate in an independent Han Chinese cohort [172].

The UK GWAS, known as the arcOGEN study, initially involved cases and controls from the UK, with replication in samples from Europe and North America [173]. The combined numbers were 13768 cases and 53286 controls. No signal exceeded the genome wide significance threshold of P less than 5×10^{-8}, however a number of SNPs produced strong signals, one of which, located in *MICAL3* had a P value of 2.7×10^{-5} for knee and/or hip OA. For knee OA, the strongest signal was located in *C6orf130* and for hip OA the strongest signal was in *COL11A1*. Further analysis of *MICAL3*, *BCL2L13* and *BID*, the latter two genes being located within the GWAS signal, failed to observe an association between the OA associated SNP rs227783 and any changes in gene expression [174]. Further analysis into the association signal in *COL11A1* has been performed and is awaiting publication (Emma Raine, personal communication).

The next phase of the arcOGEN study included 7410 cases and 11009 controls from the UK and the most promising signals were subsequently replicated in 7473 cases and 42938 controls from samples from Iceland, Estonia, the Netherlands and the UK [175]. Five genome wide significant loci were identified as being associated to OA, with a further three loci just below the threshold. The strongest of these is within the *GNL3* gene. Other loci were found on chromosome 9 close to *ASTN2*, chromosome 6 between *FILIP1* and *SENP6* and on chromosome 12, within close proximity to *KLHDC5* and *PTHLH*. Another region within chromosome 12, close to *CHST11*, was also identified. Current research is interrogating each of these signals further by gene expression analysis and DAE analysis to establish the cause of these association signals. A genome wide expression profiling study also highlighted *CHST11*, and reported significantly higher levels of expression of this gene in OA cartilage compared to normal donor cartilage [176]. *CHST11* encodes a carbohydrate sulfotransferase enzyme that catalyses

the transfer of sulphate to chondroitin, a proteoglycan important in the cartilage ECM [177]. *CHST11* therefore represents an interesting susceptibility locus for OA.

Overall, GWAS has provided us with a few compelling OA association signals, and the signals most recently identified need to be investigated in order to identify which genes harbour the susceptibility, enabling the start of functional studies. The signals so far identified from the GWAS have had low odds ratios and this is a reflection of the polygenic nature of the disease, with a few loci that mediate a moderate impact on OA susceptibility. Future studies may focus on phenotypic stratification, which could enable the identification of new loci associated with different subsets of OA or with OA severity.

1.8 Bone Morphogenetic Proteins

BMPs are a group of multi-functional growth factors belonging to the transforming growth factor β (TGFβ) superfamily. BMPs are involved in the formation of cartilage and bone and were first discovered following the observation that injection of demineralised bone matrix minerals into muscle formed bone [178]. This BMP activity was mediated by a mixture of BMPs and TGFβ family members, two of which were identified subsequently; BMP3 was identified following purification and sequencing and BMP2 by cloning [179-181]. Further research has led to the identification of more than 20 BMPs with a diverse range of functions and subgroups: osteogenic proteins (Ops), cartilage derived morphogenetic proteins (CDMPs) and growth and differentiation factors (GDFs). In addition to having a vital role in the formation of cartilage and bone during the development of the limbs, they are involved in a number of non-osteogenic developmental processes such as cardiac, kidney and vascular development [182-185]. Most mature BMP molecules consist of two monomers that are linked covalently via a disulphide bond. Two molecules can be derived from the same BMP member, forming a homodimer, or from two different BMP molecules forming a heterodimer [186].

1.8.1 BMP Signalling

BMPs signal through serine/threonine kinase receptors, composed of two subtypes, type I and II. There are three type I receptors that BMPs bind to, type IA and IB BMP receptors and type IA activin receptor. There are also three type II receptors, the type II BMP receptor, and type II and IIB activin receptors. Activin receptors bind both BMPs and activins, whereas BMP receptors are specific to BMPs. The expression pattern of

these BMP receptors differs among cell types. After ligand binding, the receptors form a heterotetrameric activated complex which consists of two pairs of type I and two pairs of type II receptors [187]. Specifically, BMPs bind to BMP receptor II, promoting phosphorylation of receptor I, which in turn phosphorylates SMADs 1, 5 and 8. These then translocate to the nucleus in combination with SMAD4 (co-SMAD). This SMAD complex relays the BMP signal from the receptor to the nucleus and activates the transcription of target genes. Activated SMADs regulate the expression of transcription factors such as *RUNX2* [188]. In addition to the SMAD signalling cascade, BMP binding to the BMP receptor complex can stimulate SMAD independent pathways, resulting in the activation of ERK, P38α and JNK [189-191].

BMPs exert a diverse range of functions, but signalling is mediated through a limited number of receptors. There are several explanations for the ability of a small number of receptors to mediate versatile responses from a number of BMPs; receptor expression patterns in different cells, affinity of the different receptors for certain ligands and oligomerisation pattern of the receptor (i.e. homodimeric BMP6 has a high affinity for type I and low affinity for type II receptors and in contrast, heterodimeric BMP2/6 has a high affinity for both type I and II receptors) [192, 193]. The lateral mobility of receptors within the membrane has also been reported to determine which signalling cascade is activated, highlighting the complexity of this signalling pathway [194].

1.8.2 BMP Inhibitors

Considering the importance of BMPs during development it is not surprising that their activities are controlled by inhibitors, such as noggin, chordin and follistatin that bind to BMPs to prevent them binding to their receptors [195, 196]. The expression of these extracellular antagonists can be up-regulated in response to BMPs, providing a negative feedback mechanism in order to control their effects. In addition, there is regulation of the signalling pathways of BMPs; SMAD6 binds to BMP receptor I and prevents the activation of SMADs 1,5 and 8 [197]. Additionally, SMAD7 is an inhibitory SMAD. Tob interacts specifically with SMAD proteins and inhibits signalling [198]. Smurf1, an E3 ubiquitin ligase, regulates BMP signalling on a number of levels; it interacts with and mediates the degradation of SMAD 1 and 5, mediates Runx2 degradation and forms a complex with SMAD6 and targets the type I BMP receptors for degradation [199-201]. The importance of these inhibitors can be assessed following their over expression or depletion; noggin homozygous knock-out mice die at birth and have abnormalities in the ribs, vertebrae and limbs [202]. In contrast, the over expression of noggin in mature

mice osteoblasts causes osteoporosis [203]. Furthermore, noggin haploinsufficiency provided protection to articular cartilage against destruction in a mouse arthritis model, and over expression of the antagonist rendered cartilage more vulnerable to damage [204]. Mutations within the *NOG* gene in humans produce joint fusions, emphasising the vital role of the regulation of BMPs during the development of the joints [205]. Furthermore, over expression of SMAD6 delays the differentiation and maturation of chondrocytes [206]. Tob null mutant mice have increased BMP signalling and thus increased bone formation, whilst Smurf1 over expression in osteoblasts inhibits postnatal bone formation [207, 208].

1.9 GDF5

GDF5 was first discovered through its association with brachypodism mice; in 1952, a mutation was discovered in a number of Swiss albino mice causing growth abnormalities. Due to the decreased size of their paws compared to control mice, these mice were named brachypodism mice (bp) [209]. It was suggested that the abnormalities were a result of the defective formation of mesenchymal condensations. It was not demonstrated until 1994 that a frameshift mutation, leading to a premature stop codon and thus a functional null mutation within the *GDF5* gene, was responsible for this phenotype. These mice have shorter length long bones and smaller feet in comparison to control mice, a result of altered patterning of the segments in the digits. The length of metacarpals and metatarsals is also altered, with a disorganisation noted in the carpals and tarsals. However, the axial skeleton is unaffected. These defects are first noted at day 12 of gestation. The digit condensations are slow to initiate chondrogenesis and are malformed and the mesenchyme show reduced ability to form aggregates [210]. These findings are distinctly different from those observed in BMP5 mutant mice, which present with altered shape and size of the ears, sternum and ribs with the absence of any changes in limb bone length or the morphology of the digits [211]. These studies therefore highlight the distinct roles of different BMPs during the formation of the skeleton. Aside from the skeleton, *GDF5* transcripts were found to be present in a number of tissues in the adult mice, however the brachypod mice are fertile and do not appear to show any other abnormalities [210].

GDF5 is a member of the TGFβ superfamily, and structurally, is closely related to the BMPs [212]. GDF5 is classified within the growth differentiation factor family, which consists of 15 proteins. GDF5 is closely related to GDF6 and GDF7, sharing 74-92% of their mature signalling region in the C-terminus [210].

1.9.1 Examination of the *GDF5* gene

The human *GDF5* gene is located on chromosome 20 and consists of two coding exons. It is highly homologous to its murine equivalent, with the mature region of the protein differing by only one amino acid [213].

Figure 1.3 shows exon 1 and the flanking intronic sequence of the *GDF5* gene. The location of two polymorphisms, rs143383 and rs143384, are highlighted in addition to the ATG translation start site. There is no TATA binding sequence or CCAAT box upstream of exon 1.

Three mRNA transcripts for *GDF5* have been identified (UCSC genome browser, http://genome.ucsc.edu/); our research laboratory has identified the presence of only one of these transcripts within a number of relevant tissues corresponding to the shorter of the two *GDF5* transcripts, the *GDF5* O/S transcript has only been identified in testis tissue (unpublished data). RNA sequencing data identified the expression of *GDF5* within cartilage tissue from NOF patients (control cartilage tissue) and in OA cartilage tissue samples (Yaobo Xu, personal communication and unpublished observations). In this data *GDF5* is expressed at a higher level in OA than NOF and this data has confirmed the presence of only one primary transcript in these tissues, suggesting the absence of alternative splice variants.

-105 ACTGGAAAGGATTCAAAACTAGGGGGAAAAAAAAACTGGAGCACACAGGCAGCATTACGC

- 45 CATTCTTCCTTCTTGGAAAAATCCCTCAGCCTTATACAAGCCTCCTTCAAGCCCTCAGTC

- 16 AGTTGTGCAGGAGAAAGGGGGCGGTYGGCTTTCTCCTTTCAAGAACGAGTTATTTTCAGC

76 TGCTGACTGGAGACGGTGCACGTCTGGATACGAGAGCATTTCCACTATGGGACTGGATAC

136 AAACACACACCCGGCAGACTTCAAGAGTCTCAGACTGAGGAGAAAGCCTTTCCTTCTGCT

196 GCTACTGCTGCTGCCGCTGCTTTTGAAAGTCCACTCCTTTCATGGTTTTTCCTGCCAAAC

256 CAGAGGCACCTTYGCTGCTGCCGCTGTTCTCTTTGGTGTCATTCAGCGGCTGGCCAGAGG

316 ATGAGACTCCCCAAACTCCTCACTTTCTTGCTTTGGTACCTGGCTTGGCTGGACCTGGAA

376 TTCATCTGCACTGTGTTGGGTGCCCCTGACTTGGGCCAGAGACCCCAGGGGACCAGGCCA

436 GGATTGGCCAAAGCAGAGGCCAAGGAGAGGCCCCCCCTGGCCCGGAACGTCTTCAGGCCA

496 GGGGGTCACAGCTATGGTGGGGGGGCCACCAATGCCAATGCCAGGGCAAAGGGAGGCACC

556 GGGCAGACAGGAGGCCTGACACAGCCCAAGAAGGATGAACCCAAAAAGCTGCCCCCCAGA

616 CCGGGCGGCCCTGAACCCAAGCCAGGACACCCTCCCCAAACAAGGCAGGCTACAGCCCGG

676 ACTGTGACCCCAAAAGGACAGCTTCCCGGAGGCAAGGCACCCCCAAAAGCAGGATCTGTC

736 CCCAGCTCCTTCCTGCTGAAGAAGGCCAGGGAGCCCGGGCCCCCACGAGAGCCCAAGGAG

796 CCGTTTCGCCCACCCCCCATCACACCCCACGAGTACATGCTCTCGCTGTACAGGACGCTG

856 TCCGATGCTGACAGAAAGGGAGGCAACAGCAGCGTGAAGTTGGAGGCTGGCCTGGCCAAC

916 ACCATCACCAGCTTTATTGACAAAGGGCAAGGTGAGGGGGCGGGGTGGCAGGGGCACGG

Figure 1.3

The sequence of a section of the *GDF5* gene: Exon 1 (highlighted in red) and flanking sequence of the *GDF5* gene. Numbering is relative to the transcription start site. The two polymorphisms rs143383 (+41bp) and rs143384 (+268bp) are highlighted in green and blue respectively, Y, C/T. The ATG methionine start codon is highlighted in yellow and underlined.

1.9.2 Mode of Action of GDF5

As for most BMPs, *GDF5* is an intermolecular disulphide-bonded protein dimer, and features seven highly conserved cysteine residues in the carboxy-terminal region. Six of these conserved cysteines are present in the core of the monomer protein forming a cysteine knot, a seventh cysteine forms an intermolecular disulphide bond with the corresponding cysteine in the second monomer [214]. Once assembled, the large dimeric precursor protein (consisting of a 495 amino acid protein, preceding an 120 amino acid mature C terminal polypeptide) is then cleaved at an RXXR site (Arg, X, X,

Arg) by a protein from the pro-protein convertase family. The biologically active mature dimer is secreted and signals through its mature domain [215].

GDF5 binds to BMP receptors as described earlier for BMPs (section 1.8). GDF5 signals specifically through a heteromeric complex of BMP receptor II and BMP receptor IB (Alk 6). ActR-II may also serve as a type II receptor for GDF5 in some cell types [216]. GDF5 binds first to the type II receptor, which phosphorylates the type I receptor and subsequently transduces the signal by phosphorylating and thus activating signalling cascades such as SMAD and MAPK signalling pathways. GDF5 activates the transcription of genes involved in the formation of cartilage and bone, such as *COL2A1* (encoding collagen type II), *AGC1* (encoding aggrecan) and *SOX9* (encoding the transcription factor SOX-9) [217] (Figure 1.4).

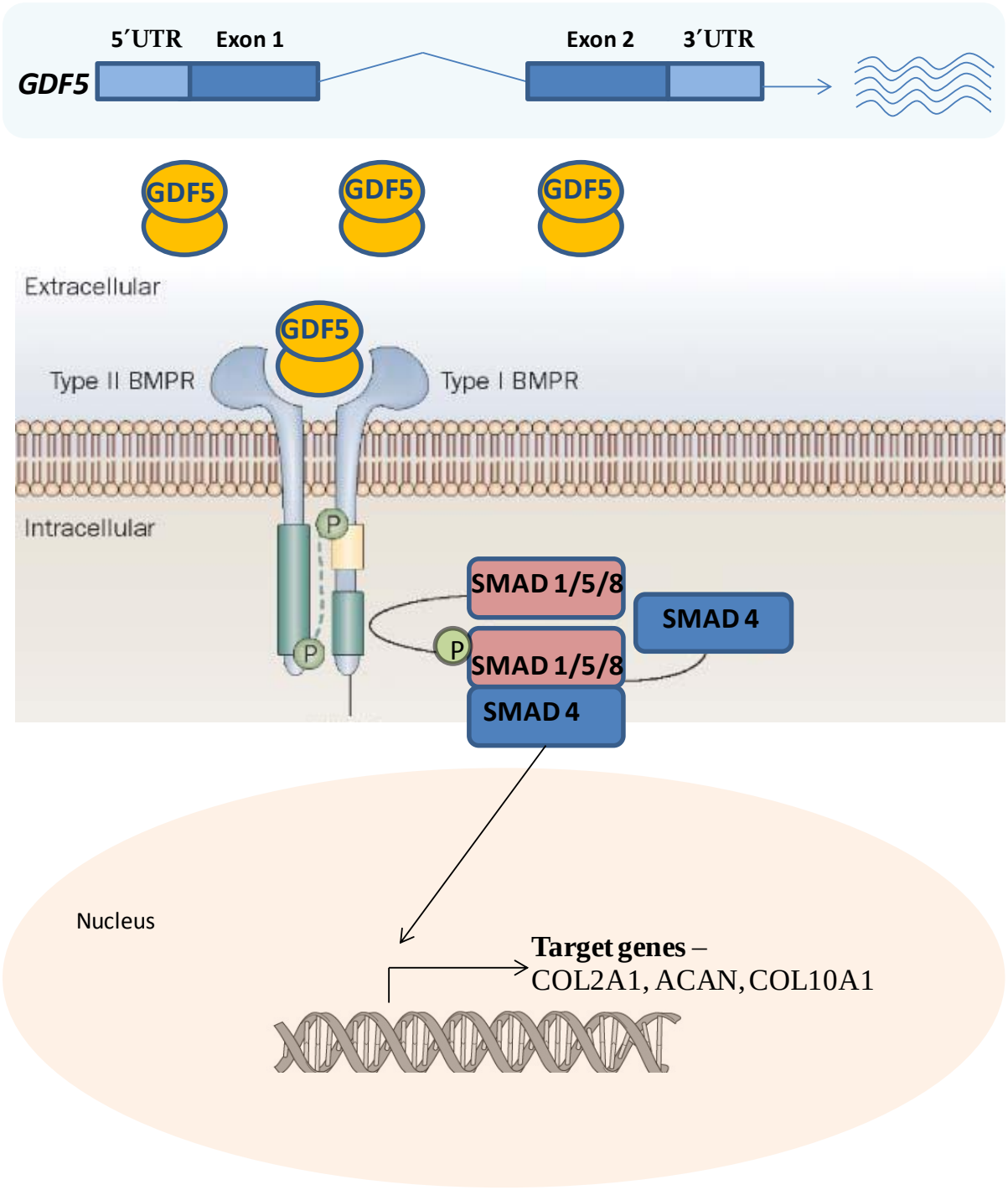


Figure 1.4

A schematic diagram of the GDF5 signalling cascade: The *GDF5* gene region is shown and contains two exons. GDF5 protein forms homodimers which bind to the BMP receptor II and BMP receptor IB. Binding promotes receptor phosphorylation which subsequently leads to the phosphorylation and thus activation of SMAD signalling. SMADs 1, 5 and 8 in combination with SMAD4 relay the signal to the nucleus where the transcription of target genes such as *COL2A1* (collagen type II), *ACAN* (aggrecan) and *COL10A1* (collagen type X) is activated. Figure modified from [218].

The binding profiles of GDF5 were found to be limited in comparison to BMP2, BMP4 and BMP7, suggesting that GDF5 has different biological functions compared to a number of the previously identified BMPs. Furthermore, the expression profile of the BMP receptor IB is more limited compared to that of BMP receptor IA, suggesting a more restrictive function of GDF5 during development [216].

1.9.3 Role of GDF5 during Development

GDF5 is essential for the formation of cartilaginous tissues during early limb development [210, 219, 220]. During joint specification GDF5 can be used as a marker for the site of joint formation due to its early presence within the joint space [212]. GDF5 is expressed in the condensing mesenchyme prior to joint development and later in development it can be detected in the early perichondral layer around the cartilage elements and in the joint interzone. The expression of *GDF5* is down-regulated as the joint cavitates [221]. GDF5 expression can be detected in early postnatal cartilage and has been detected in adult bovine and in healthy and osteoarthritic adult human articular cartilage [219, 220, 222].

GDF5 increases the numbers of mesenchymal cell condensations and mesenchymal aggregation, and increases cartilaginous nodules in mouse and rat embryonic limb bud mesenchymal cells [215, 221]. In chick limb bud, the over expression of GDF5 increased the size of the early cartilage condensation, increased chondrocyte proliferation in the later stages of skeletal development and, overall, increased the length of the skeletal elements [221]. Furthermore, using adult bone marrow derived MSCs it has been shown that treatment of cells *in vitro* with recombinant GDF5 promotes increased expression of collagen type II and sulphated glycosaminoglycans (GAGs), indicating enhanced chondrogenic differentiation. Additionally, there were increased levels of phosphorylated SMADs 1, 5 and 8 [223]. These gain of function studies have provided an insight into the key role of GDF5 during two distinct stages of development, first to control mesencyhmal cell recruitment and condensation and secondly in the joint interzone to control the growth of adjacent skeletal elements by controlling chondrocyte proliferation.

In addition to effects on chondrogenesis and synovial joint formation, GDF5 also affects endochondral ossification, with GDF5 deficiency in the bp mice resulting in shortened long bones, and the absence of GDF5 has been suggested to affect the duration of the

chondrocytes hypertrophic phase. Furthermore, GDF5 deficient bones are significantly weaker than control bones and more compliant when tested to failure in torsion [224].

GDF5, GDF6 and GDF7 have also been suggested to play a role in tendon and ligament formation [225]. GDF5 has been found to induce the osteogenic differentiation of human ligament cells through the activation of ERK1/2 and p38 MAPK. It is present at ossified sites, and treatment with recombinant GDF5 increases the expression of alkaline phosphatase and osteocalcin [226]. However, this effect on osteogenic factors appears to be more discrete when compared with BMP7 or BMP2. In the osteoblastic MC3T3-E1 cell line, in contrast to the positive effect of BMP2, GDF5 did not increase alkaline phosphatase activity [215].

In human mesenchymal stem cells GDF5 treatment was shown to enhance osteogenesis *in vitro* and *in vivo*, upregulating *COL1A1* (Collagen type 1A1), *ALPL* (alkaline phosphatase), and *OSTCN* (osteocalcin) mRNA [227]. Another study found that GDF5 stimulated early osteoblastic differentiation, enhancing expression of the homeobox proteins MSX1 and MSX2 in addition to Runx2 and osteocalcin [228]. GDF5 has also been shown to enhance angiogenesis in both chick chorioallantoic membrane and rabbit cornea assays [229].

1.9.4 Human Conditions Associated with GDF5

The importance of GDF5 during the development of the joints is evident from the phenotypes associated with mutation of *GDF5*, its receptors or within the antagonists controlling its function. Mutations within the *GDF5* gene have been associated with a spectrum of human conditions including Hunter-Thompson and Grebe type acromesomelic chondrodysplasias and DuPan syndrome, in addition to brachydactyly type C (BDC).

Hunter Thompson type acromesomelic chondrodysplasia is characterised by skeletal abnormalities that are restricted to the limbs and has a similar phenotype to the bp mutation. A frameshift mutation within the *GDF5* gene was found to be the cause of the phenotype and represented the first mutation found in a TGFβ family member causing a human disorder [230]. Grebe type acromesomelic chondrodysplasia (GCD) is a severe autosomal recessive condition characterised by limb shortening and appendicular bone dysmorphogenesis. Several mutations in *GDF5* have been associated with this condition; an insertion of a C nucleotide at position 297 in the coding sequence of *GDF5* was found in a cohort of patients, an insertion of a G at position 206 was found

in an affected boy and a novel four base insertion was found in an affected family. All three mutations caused a shift in the reading frame and premature termination resulting in total loss of function [231-233].

Du-Pan syndrome is an autosomal recessive trait that is characterised by fibular hypoplasia and complex brachydacyly. Affected individuals have reductions in the size of or absence of limb bones and appendicular bone dysmorphogenesis. Hunter-Thompson and Grebe type chondrodysplasias are phenotypically similar to Du-Pan syndrome and thus the *GDF5* gene was investigated as a possible source of the Du-Pan causative mutation. A mutation within the *GDF5* gene, causing a substitution in the mature domain of the protein was found. This mutation resulted in a conformational change in the protein structure and thus affected protein function. Individuals heterozygous for this mutation were unaffected [234].

The condition brachydactyly is characterised by a shortness of the fingers and toes. A number of mutations within *GDF5* have been associated with bracydactyly. Its association was first discovered through the investigation of a family with brachydactyly type A2 (BDA2), characterised by a shortening of the phalanges of the index finger and second toes. A genetic analysis revealed that these individuals had a mutation within the processing site of *GDF5*. The effect of this mutation, consisting of an arginine to glutamine substitution, was investigated in chicken micromass cultures. The mutation was found to interfere with the processing of the pro-protein, thus significantly reducing the amount of biologically active GDF5 [235]. Furthermore, a missense mutation within the coding region of *GDF5*, Arg388Cys, has been found to cause brachydactyly type A1 (BDA1). This mutation is semidominant, with patients who are heterozygous for this polymorphism being mildly affected, however those who are homozygous present with severe brachydactyly [236] .

Brachydactyly type C (BDC), characterised by shortening of the middle phalanges of the index, middle and little finger has been associated with mutations within *GDF5 [237, 238]*. A novel missense mutation has been identified within the pro-domain region of the gene resulting in the BDC phenotype, a mild shortening of the metacarpals was observed in heterozygous carriers demonstrating the semidominant nature of this mutation [239].

Mutations within the BMP receptor IB receptor are commonly associated with BDA2, but have been associated with BDC and a symphalangism-like phenotype [240, 241].

Furthermore, a mutation within the *GDF5* gene causing inhibition of the ligand-BMP receptor IB interaction resulted in a BDA2 phenotype [242]. A mutation in *GDF5* within the BMP receptor II interaction site was identified in patients who were suffering from multiple synostosis syndrome (SYNS), an autosomal dominant disorder characterised by progressive symphalangism, deafness and mild facial dysmorphism. This mutation caused impaired binding to the receptor but also resulted in resistance to Noggin, causing enhanced GDF5 action [243].

Mutations within the BMP inhibitor Noggin have previously been found in patients with SYNS and brachydacyly type B (BDB), characterised by terminal deficiency of the fingers and toes [244]. Activating mutations in GDF5 have now also been associated with SYNS. These mutations disrupt noggin binding, leading to noggin resistance and altered signalling, causing an increase in chondrogenic activity and an induction of cartilage formation in an *in vivo* chick model [242, 245, 246]. These studies demonstrate that both decreased and increased expression of *GDF5* can lead to abnormal joint development and thus regulation of the expression of this factor is vital.

1.9.5 Mouse Models of gdf5 Mutation

As described, mutations within *GDF5* have been reported to be the cause of a number of skeletal conditions in humans. In mice however, far fewer mutations have been found within *gdf5*. The loss of function bp mouse was described previously, and presents with abnormalities of the digits and reduced length of long bones. Transgenic mice have also been investigated, with the *gdf5* gene under the control of the type II collagen promoter. These mice exhibited chondrodysplasia, with expanded cartilage consisting of a reduced proliferative and increased hypertrophic zone of chondrocytes [247].

Further investigations sought to investigate in more depth the role of *gdf5* in development. The first of these studies used the chemical N-ethyl-N-nitrosourea (ENU) to induce mutations randomly in genomic DNA [248]. Following this treatment, mice were then screened for abnormal limb phenotypes transmitted as dominant traits and a mutant line that exhibited a brachypodism phenotype was identified. The mutation was within *gdf5* and was present in the BMP receptor I interaction domain. The protein dimerised and was secreted normally, but was unable to activate SMAD signalling, and acted in a dominant negative manner through hetero-dimerisation with wild type gdf5 *[248]* . These aforementioned studies have focused on the effects of loss of gdf5 activity

in mice and are therefore comparable with the conditions in humans associated with the loss of GDF5 activity. `

A second study examined the effect of *gdf5* hypomorphism in mice heterozygous for the bp mutation and examined susceptibility of these mice to developing an OA like phenotype [249]. These mice show no apparent skeletal abnormalities and grow normally, and the length of their long bones is comparable with the wild type (WT) mice. Additionally, there were no differences in BMD or bone area, although a significantly decreased trabecular area was observed in the long bones of the heterozygous mice compared to the WT mice. The effect of *gdf5* haploinsufficiency on the development of OA was assessed using four different mouse models: collagen induced arthritis (CIA), treadmill model, meniscus destabilisation model and papain-induced arthritis. Although no difference in cartilage damage between the WT and heterozygous mice was observed in the collagenase injected knees, the contralateral knees of the heterozygotes showed increases in synovial hyperplasia, capsule fibrosis and the number of osteophytes. These effects were thought to be secondary effects, a result of increased load of the unchallenged knee. Furthermore, in the treadmill model a significantly increased score for synovial hyperplasia was found in the heterozygous mice compared to the WT mice, and an increase in capsule fibrosis was also noted. No differences between WT and heterozygous mice were observed in the other two models tested. Gait analysis revealed that heterozygous mice had a decreased stride length compared to WT mice suggesting that joint stability may be mediating changes in load on the joint. A reduction in *gdf5* expression in these mice therefore appeared to mediate increased susceptibility to OA like changes in the CIA and running model, a result of altered gait and loading. Furthermore, changes in the subchondral bone and collagen fibre orientation were observed in heterozygous mice [249].

Other studies have examined the effect of *gdf5* mutation on the development of tendons and ligaments; bp mice show anterior dislocation, hypoplastic condyles and absence of intra-articular ligaments. These changes may be a result of excessive mesenchymal cell death that was observed in the future knee joint of these mice [250]. Furthermore, gdf5 deficiency was shown to alter the ultrastructure, mechanical properties and composition of the Achilles tendon compared with heterozygous littermates. The tendons in the deficient mice were structurally weaker, contained less collagen and had a smaller diameter. These findings suggested that joint dislocations that are associated with GDF5 deficiency may be due to changes in tendon and ligament laxity [251]. Mice deficient in

gdf5 have also been reported to show delayed fracture healing in the early phases of bone repair compared with WT mice, although long term fracture healing was unaffected [252].

1.9.6 Therapeutic Use of GDF5

Considering the vital role of GDF5 during the formation of the joint, its use therapeutically has been considered. Chondrocytes derived from both human adult healthy and osteoarthritic cartilage were responsive to GDF5; GDF5 stimulated the metabolic activity of chondrocytes, increasing proteoglycan synthesis in a cartilage explant model [222].

Models of animal injury have been used to investigate the use of GDF5 therapeutically. The first of these investigated the use of GDF5 following injury to tendons and ligaments; GDF5 treated tendon tissue specimens were thicker, with a higher strength and had increased collagen type II expression [253]. Additionally, another study examined the cellular and molecular response to GDF5 on Achilles tendon fibroblasts following recombinant GDF5 treatment; glycosaminoglycan (GAG), hydroxyproline (HYP) and total DNA content were increased. GDF5 treatment increased cell proliferation and ECM synthesis as well as the expression of a number of genes [254, 255]. The effect of GDF5 on tendon healing was enhanced when GDF5 was added alongside bone marrow derived stem cells [256].

GDF5 was also found to have early beneficial effects on flexor tendon repair in a rabbit model, with a higher histological scoring of collagen in the GDF5 treated animals [257]. Furthermore, the combined use of bFGF and GDF5 enhances the healing of medial collateral ligament injury by increasing proliferation and expression of *COL1A1* in rabbits [258].

Due to the success of these experiments, researchers are now moving on to identify new methods for delivery of growth factors to sites in need of repair. Dines et al. have developed a method of suture coating in order to provide the delivery vehicle for the growth factor and have had success in coating sutures with recombinant GDF5 [255]. Cell migration, cell proliferation and collagen synthesis were all stimulated on treatment with recombinant GDF5 in rat tendon fibroblasts [255]. Additionally, stimulation with GDF5 can modulate primary adipose derived stem cells on an electrospun scaffold that mimics collagen fibre bundles in tendon tissue; GDF5 up-regulated the expression of genes encoding the major tendon ECM proteins and a neotendon marker [259].

Aside from tendon and ligament healing properties, GDF5 expression has been demonstrated in non-degenerate and degenerate human inverterbral discs. Treatment of nucleus pulposus cells with GDF5 showed an increase in *ACAN* and *COL2A1* gene expression and increased production of GAGs [260]. A recombinant human GDF5-coated β-tricalcium phosphate graft has been shown to enhance bone formation in maxillary sinus lift augmentation procedures [261]. Finally, the use of GDF5 in spinal fusion models and in periodontal wound healing has also been investigated [262-265].

When considering the use of GDF5 therapeutically, the importance of controlled expression of GDF5 must be taken into account. GDF5 must be regulated both spatially and temporally for the correct development of the joint, with joint fusions being observed as a result of the injection of GDF5 into the joint space [266].

1.10 rs143383

Table 1.1 shows the population data available for the SNP. The T allele is the common allele in Caucasians and the two Asian populations shown, whilst CC homozygotes are uncommon in Caucasians and Asians. In the West African population however, the C allele only is present.

Group	**Genotype**			**Alleles**	
	CC	**CT**	**TT**	**C**	**T**
European	0.117	0.433	0.450	0.333	0.667
Han Chinese	0.047	0.349	0.605	0.221	0.779
Japanese	0.023	0.442	0.535	0.244	0.756
West African	1			1	

Table 1.1

Genotype frequencies of rs143383: Frequencies of the rs143383 CC, CT and TT genotypes in addition to the C and T allelic frequency within different population groups. Population frequency information obtained from dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/).

rs143383 is a C/T polymorphism within the 5´UTR of *GDF5*. As noted earlier, this polymorphism was first reported to be associated with OA in 2007 when it was demonstrated in a Japanese population to show significant association with hip OA. This association was reported as part of a study that involved the sequencing of *GDF5*

in 24 individuals with hip OA [134]. Fifty two polymorphisms were identified and 31 SNPs were chosen for further analysis based on minor allele frequencies of >0.10. The most significant association was observed for rs143383 ($p= 3.1 \times 10^{-11}$), with allelic frequencies of 83.6% (hip OA) and 74% (control) for the T allele. The odds ratio for the susceptibility allele was 1.79 and this effect was apparent following correction for confounding factors. This association was also replicated in a Japanese cohort with knee OA (718 knee OA patients and 861 controls), with an odds ratio of 1.3 ($p = 0.00021$), and subsequently in a Han Chinese cohort with knee OA (313 knee OA and 485 controls), with an odds ratio of 1.54 ($p = 0.00028$),. The promoter activity of *GDF5* was mapped to a 162 base pair region encompassing the polymorphism and it was therefore hypothesised that this SNP may affect transcription [134]. Further investigation demonstrated that the polymorphism was functional, with the susceptibility T allele showing a lower level of transcriptional activity compared to the C allele in three independent cell lines using luciferase reporter experiments. This study therefore suggested that a reduction in the expression of *GDF5* can increase susceptibility to OA [134].

A number of other groups replicated this association in different ethnic populations and further investigated the functional effect of the T allele. The association was first replicated at the genotype level, allele level and when carriers were combined (TT + CT) in a European cohort in 2007. This analysis was performed using 2487 cases with knee, hip and hand OA and 2018 controls, from both the UK and Spain [267]. In the T allele carrier analysis, the knee, hip, and knee and hip cases were all significant ($p<0.05$) with odds ratios greater than 1, whereas cases of hand OA were not significantly associated. The unstratified odds ratios were not as high as those observed in the Asian study (1.10 compared to 1.79). This suggests that the rs143383 SNP is not as major a susceptibility locus in Europeans, and that differences between ethnic groups may impact on the effect of the risk allele [267].

The functional effect of the SNP was further assessed in this study using *in vivo* allelic expression analysis. This analysis used RNA extracted from the cartilage of OA patients and enabled the quantification of the expression of the two alleles in heterozygotes. The two alleles were normalised to an assumed genomic DNA ratio of 1:1. It was discovered that the T allele showed a reduction in the expression of *GDF5* in comparison to the C allele. The average allelic difference observed in these patients was a 27% reduction in the expression of the T allele relative to the C allele. This alteration in gene expression

is found in patients with severe disease undergoing joint replacement surgery. Although only a small change is observed, over a prolonged period of time this reduced expression may render individuals more susceptible to developing OA [267].

A Greek study however reported no association of the rs143383 T allele with OA following the genotyping of 519 Greek Caucasian cases and controls ($p>0.05$) [268]. There are two possible explanations for this result: firstly, and as just touched on, this polymorphism may have differential effects in different populations and secondly it may be a result of the small sample size used. In this regard it is worth noting that within the original European study, alone the UK and Spanish cohorts did not detect significant associations between rs143383 genotype and OA, however when combined, these two cohorts showed a significant association. A meta analysis combining the data from the Asian and European studies with new data from the UK and the Netherlands provided strong evidence of this polymorphisms association with knee OA, implying the second explanation may be the reason for the lack of association in the Greek cohort ($p = 0.0004$ for the allele frequency model) [269].

rs143383 was not directly targeted by the GWAS, however, a recent study that directly targeted rs143383 has demonstrated an association of the T allele with knee OA with genome wide statistical significance ($p=6.2 \times 10^{-11}$): 6861 knee OA cases and 10103 controls of European descent and 718 cases and 1844 controls of Asian descent were combined and the T allele was shown to be associated with a 17% increased risk of knee OA (OR 1.17). There was no significant heterogeneity seen between studies [270].

The importance of this locus as a risk factor for knee OA has been reproducibly demonstrated, however, for hip and hand OA a significant association has not been demonstrated (statistical significance was borderline for OA of the hip and absent for hand OA) with a large amount of inter-study heterogeneity being observed [148]. It has been noted that extremely large sample sizes are needed in order to assess the association of common genetic variants with multifactorial diseases, which may only mediate modest effects, as with OA. Furthermore, as I previously discussed, a standardised approach for grading OA radiographically is needed as differences in the radiographic assessment of OA in these studies may introduce heterogeneity [148].

In order to assess whether rs143383 is associated with increased severity of OA of the knee, cases of knee OA were genotyped for rs143383 and their tibiofemoral K/L grade and patellofemoral grade was assessed following adjustments for age, gender and BMI.

The T risk allele is associated with a significantly higher tibiofemoral K/L grade suggesting that reduced expression of this chondroprotective factor influences the extent of radiographic damage. This effect could be a result of GDF5 influencing biological processes involved in joint damage, repair or the age of onset of the disease [271].

Following the discovery that the rs143383 T allele mediated a reduction in the expression of *GDF5* relative to the C allele *in vivo* in patient cartilage samples, our research group further assessed the extent of this differential allelic expression. OA is characterised by changes in a number of tissues of the joint, thus allelic expression analysis was performed using rs143383 on a variety of tissues taken from the knee and hip joints of OA patients. The effect of this polymorphism was found not just to be restricted to cartilage, but a reduced expression of the T allele was evident joint wide in the synovium, meniscus, ligament, tendon and fat pad of patients with OA [272].

1.10.1 rs143383 and Other Conditions

In addition to being associated with OA, the rs143383 T allele also correlates with reduced height in both males and females [273]. This study highlights that the functional effect of rs143383 is not only evident in adult life, predisposing individuals to OA, but its effect is also apparent in development, leading to variation in height. A reduction in the expression of *GDF5* may influence cartilage composition, limb proportions or joint angles, which may therefore affect stature in addition to susceptibility to OA. Furthermore, the CC homozygotes had a larger hip axis length and were at increased risk of developing non vertebral fractures in comparison to CT heterozygotes or TT homozygotes, suggesting that there may be differences in bone composition between genotype groups, although no effect on bone mineral density was reported [274].

The T allele of rs143383 has also been found to increase susceptibility to other musculoskeletal disorders including Achilles tendon pathology, congenital hip dysplasia and lumbar disc disease. Genotyping of 171 cases and 235 controls revealed that individuals with a TT genotype at this locus have twice the risk of developing Achilles tendon pathology, with an odds ratio of 2.24 [275]. This condition occurs as a result of excess mechanical loading or repetitive loading leading to damage of the Achilles tendon. Additionally, in *GDF5* deficient mice, tendons were found to contain less collagen and were much weaker [251]. CDH is a common congenital skeletal abnormality, characterised by abnormal positioning of the femoral head in the

acetabulum. [276]. CDH is a strong risk factor for the development of hip OA and has been shown to have a large genetic component. Some patients with BDC also present with dysplasia of the hip, hence the investigation of *GDF5* as a candidate gene for CDH. 960 cases and 622 controls of Han Chinese origin were genotyped for rs143383, revealing an association of the T allele with CDH in women, with an odds ratio of 1.4. This association was not significant in males. Furthermore, when stratified by severity, a significant increasing linear trend was observed when comparing the T allele frequency with worsening CDH severity [276].

Lumbar disc disease (LDD) is a major cause of back pain via the development of degenerate discs. LDD is known to have a genetic component and a number of loci have been tested for association using a candidate gene approach. An association with the T allele of rs143383 and LDD in women was reported, with an odds ratio of 1.72 following genotyping of five Northern European cohorts [277]. This increased risk of LDD development in T allele carriers is thought to be mediated either directly by affecting the disc itself or indirectly by affecting periarticular structures [277].

Overall, these studies suggest that GDF5 is crucial during development and the presence of the rs143383 polymorphism within *GDF5*, which affects gene function, can predispose individuals to a variety of skeletal abnormalities. They also highlight the pleiotropic nature of rs143383; the T allele has been associated with several different phenotypes, owing to the fact that *GDF5* has a number of functions and is expressed in different cell types at various stages of development.

1.11 Transcription

Transcription is the production of RNA in a DNA dependent manner, this process is carried out by RNA polymerases. Three different types of this enzyme exist in eukaryotes and act on different sets of genes. RNA polymerase II transcribes the genes encoding proteins in addition to some of the small nuclear RNAs involved in RNA splicing.

There are a number of proteins involved in transcriptional regulation, including proteins that associate with RNA polymerase II during transcriptional activation and these proteins are essential for the assembly of the stable transcriptional complex [278, 279]. RNA polymerases lack sequence specific recognition ability and thus these additional proteins are necessary to form an initiation complex. There are three classes of proteins known as general transcription factors that associate with RNA polymerase II to aid

transcriptional initiation: TATA binding protein associated factors (TAFs), mediator and upstream stimulatory activity (USA) derived cofactors [280]. Additionally, general cofactors and site specific transcription factors such as activators and repressors can bind and modulate the activity of the polymerase.

The core promoter of a gene is usually located just upstream of the transcription start site and genes can be characterised based on the presence of specific DNA elements within the core promoter (core promoter elements). An examples of such a core promoter element is the TATA box, an AT rich sequence located just upstream of the transcription start site which has a consensus sequence of TATA(A/T)A(A/T)(A/G) (genes can be classified as TATA-containing or TATA-less promoters) [281, 282]. In genes with promoters that lack the TATA sequence, other core promoter elements such as an initiator (Inr), containing a pyrimidine rich sequence can direct transcription initiation either alone or in conjunction with other core promoter elements.

Promoter analysis of the eukaryotic promoter database and the database of human transcriptional start sites has found that less than 22% of human genes contain a TATA box, 62% of which have an Inr, 24% of which have a downstream promoter element (DPE) and 12% have an upstream TFIIB recognition element (BRE^u). Among the TATA less promoters, 45% have an Inr, 25% a DPE and 28% have a BRE^u [283].

The TATA site is recognised by a subunit of TFIID, which binds first to DNA and is facilitated by TFIIA. TFIIA binds to TFIID and stimulates its activity by overcoming inhibitory factors; TFIID is associated with inhibitory factors which prevent its binding to DNA [284]. TFIIA also acts as a coactivator by linking DNA bound activators with the basal transcription complex. Following the binding of TFIID and TFIIA, TFIIB binds to TFIID. TFIIA and TFIIB help to stabilise promoter bound TFIID. TFIIB also binds to RNA polymerase II and so is essential in the formation of the basal complex. RNA polymerase II is pre-associated with TFIIF, both of which bind to the complex. Finally, the binding of TFIIE, TFIIH and TFIIJ finalises the pre initiation complex (PIC) [282]. TFIIH has helicase activity and is therefore involved in unwinding the double-stranded DNA ready for transcription creating an open complex. In addition, TFIIH has kinase activity and phosphorylates RNA polymerase II at the C terminal domain (CTD) initiating transcription. During transcription, TFIIF remains associated with RNA polymerase, whilst TFIID and TFIIA remain bound to the promoter enabling

repeated rounds of transcription [278]. This sequential binding of factors is known as the sequential assembly pathway.

Transcriptional activity resulting from the binding of general transcription factors alone is usually low and is increased following the binding of site specific factors to proximal promoter regions. The proximal promoter region is usually located close to the core promoter and factors binding to it can help recruit general transcription factors or stabilise the PIC. In addition, there are *cis* acting sequences which regulate gene expression from distal regions such as enhancers, silencers and insulators. Enhancer elements can further stimulate transcription by binding a multiprotein complex, known as the enhanceosome, which activates transcription and can recruit chromatin remodelling proteins. Proteins binding to enhancers can also promote looping out of the DNA in order to bring transcription factors binding at distal regions within close proximity of transcription factors at the core promoter or at another site. Silencing elements bind factors that decrease transcriptional activity by either inhibiting RNA polymerase activated transcription or by recruiting histone modifying complexes to create repressive chromatin structures [278]. In homo sapiens it has been estimated that there are between 200-300 general transcription factors that bind to the core promoter elements, including the subunits of RNA polymerase II. Additionally, there are 1,400 factors that bind to DNA in a sequence specific manner, these factors have either very broad (present in all or most tissues) or specific (present in one or two tissues) functions [285].

Alterations in gene expression as a result of the aberrant levels or altered structure or function of transcription factors underlies the pathology of a number of diseases, with genes encoding transcription factors accounting for more than 30% of the genes associated with malformation phenotypes [286]. For genes whose expression is highly regulated, such as those involved in development or cell cycle progression, aberrant expression of transcription factors could cause problems during development and could increase susceptibility to diseases such as cancer. The identity of the transcription factors binding to and regulating such important genes must therefore be studied to expand our understanding of gene regulation within both normal and disease states, and to provide the potential for the development of novel preventative or treatment strategies.

1.11.1 Differential Allelic Expression (DAE)

There are two broad categories of mutation/polymorphisms; in general, rare mutations that are highly deleterious cause mendelian diseases, whilst common polymorphisms are often associated with complex, multifactorial diseases. SNPs that alter gene expression levels are referred to as expression Quantitative Trait Loci (eQTLs) [287]. *Cis*-regulatory variation in humans is extensive, with around 40% of genes in an individual estimated to be heterozygous at functional *cis*-regulatory sites. The investigation of SNPs that are located within regulatory regions can be difficult compared with investigating SNPs in protein coding regions due to the large size of regulatory regions and the variability in their localisation from the transcriptional start site [288, 289]. Furthermore, the action of *cis*-regulatory elements has been shown to be context specific, so may be difficult to identify in a single tissue or cell type [290-292]. Although this research may be challenging, these sites need to be investigated further as it is likely that they are mediating individual's susceptibility to developing common complex disorders.

In order to assess the effect of a *cis*-regulatory polymorphism, the allelic output from the gene can be measured. The assumption here being that in the absence of a *cis*-regulatory polymorphism, both alleles will contribute an equal amount of transcript. Where a differential level of expression between the two alleles is seen, the gene is described as showing DAE (Figure 1.5).When DAE is present, an individual is likely to be heterozygous at a *cis*-regulatory site that is responsible for the transcription of the gene, for the processing of the mRNA (e.g. within a splice site), or for mRNA stability [293]. This is the case for the *GDF5* rs143383 SNP described previously, where there is a lower level of expression of the T allele relative to the C allele. The relative number of genomic DNA molecules for each allele is first determined and compared with the number of mRNA molecules. It is likely that polymorphisms present within the promoter region of a gene may alter expression of the gene by affecting DNA binding proteins with a role in transcription, whereas polymorphisms that are present within the 3′UTR of genes are more likely to affect transcript stability. It has been estimated that around 20-50% of genes show DAE dependent on the method used for detection [294, 295].

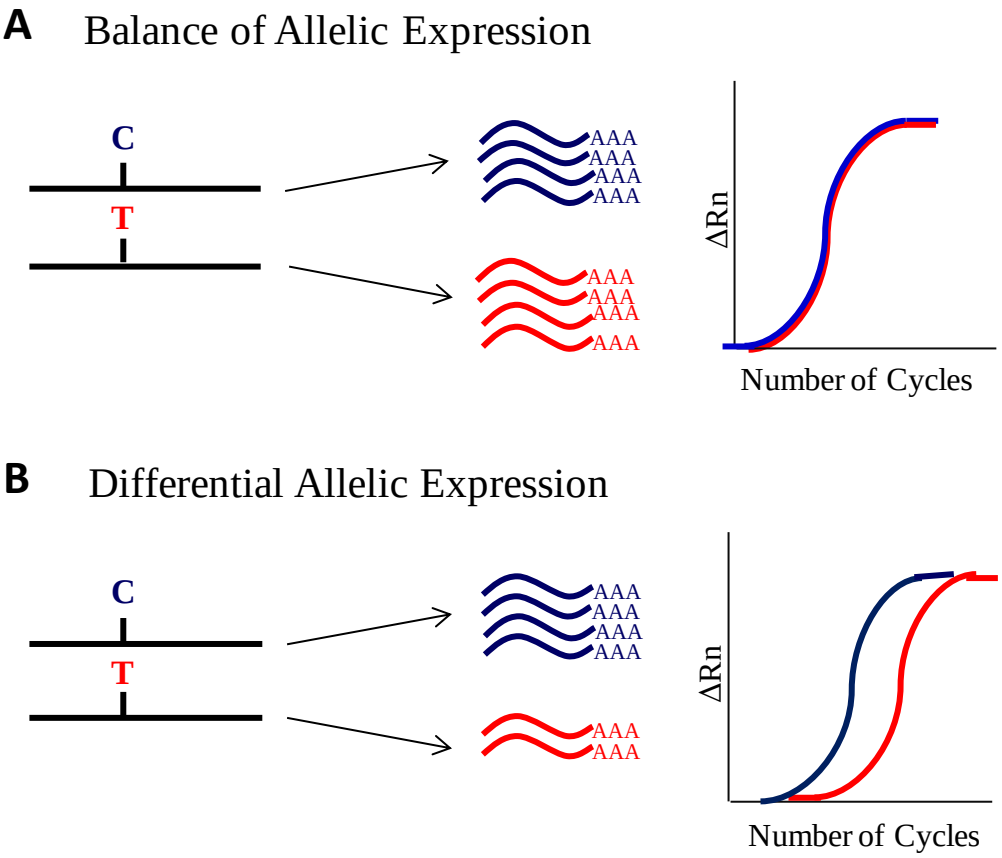


Figure 1.5

Differential allelic expression. A: A balance of allelic expression, with both the C and T allele expressing equal amounts of mRNA transcript. The two alleles have identical amplification curves assessed by real time PCR. B: Differential allelic expression, with one of the two alleles expressing a differential level of mRNA transcript; there is a decreased level of expression of the T allele relative to the C allele. The T allele is identified at a higher cycle threshold by real time PCR.

Summary

Osteoarthritis (OA) is a highly prevalent, chronic, complex disease which causes pain and significant dysfunction. OA is characterised primarily by a degeneration of articular cartilage, however a number of tissues within the synovial joint show disease related changes. Current therapeutic options are limited to pain relief and surgery.

OA occurs as a result of both genetic and non-genetic factors. OA is polygenic; many susceptibility alleles contribute to disease aetiology, each with varying individual effects. A number of polymorphisms have now been associated with OA through a variety of genetic approaches, with one of the most robust to date being rs143383 within the 5´UTR of *GDF5*. The T allele of this SNP produces less mRNA transcript in comparison with the C allele, and has been found to influence OA susceptibility across a range of ethnic groups.

GDF5 is an important growth factor that has been extensively characterised, revealing its many distinct functions within the body. GDF5 is essential in the joint during development with mutations causing a range of skeletal phenotypes in both mice and humans.

Functional analysis of OA associated loci will further enhance understanding of the OA disease process and could potentially provide targets for disease modifying drugs.

Aims

- To identify a cell line that expresses *GDF5*, is heterozygous for rs143383 and demonstrates *GDF5* DAE to provide a model system for *in vitro* manipulation.
- To identify factors binding differentially to the C and T alleles of rs143383.
- To modulate the expression of the identified factors to assess their regulation of *GDF5*.
- The overall goal of this research is to identify new targets which can be manipulated to normalise the expression of *GDF5* within T allele carriers and ultimately provide a novel therapeutic strategy to benefit OA patients.

Chapter 2: Materials and Methods

2.1 Cell Line Culture

For the purposes of this research, three different cell lines were used; the human liposarcoma cell line, SW872, initiated from a surgical specimen of a fibrosarcoma taken from a 36 year old male Caucasian, the human chondrosarcoma cell line, SW1353, taken from the humerus of a 72 year old female Caucasian, and finally the human osteoblast cell line, MG63, taken from a 14 year old male Caucasian.

2.1.1 Method

The cells were cultured in their respective culture medium (shown below) in vented T75cm^2 flasks (Corning, UK) at 37°C in a 5% (v/v) CO_2 atmosphere until they reached 70-80% confluency.

2.1.2 Medium

- SW872: Dulbeccos modified eagles medium: Hams F12 Glutamax medium in a 3:1 ratio (GIBCO, Invitrogen, Life Technologies, Paisley, UK) containing 5% (volume/volume (v/v)) fetal bovine serum (FBS), 100U/ml penicillin and 100µg/ml streptomycin (Sigma-Aldrich, Dorset, UK).
- SW1353: Dulbeccos modified eagles medium: Hams F12 medium in a 1:1 ratio (GIBCO) containing 10% (v/v) FBS, 100U/ml penicillin, 100µg/ml streptomycin and 2mM of L-glutamine (Sigma-Aldrich).
- MG63: Dulbeccos modified eagles medium (GIBCO) containing 10% (v/v) FBS, 100U/ml penicillin, 100µg/ml streptomycin and 2mM of L-glutamine (Sigma-Aldrich).

2.2 Human Articular Chondrocyte (HAC) Culture

2.2.1 Cartilage Sample Collection

Human articular cartilage samples were obtained from patients with osteoarthritis undergoing total hip or total knee replacement surgery at the Freeman Hospital, Newcastle upon Tyne. Ethical approval and informed consent were obtained prior to surgery. Prior to extraction, samples were stored in Hank's Balanced Salt Solution (HBSS) supplemented with Penicillin, Streptomycin and Nyastatin at 4°C.

2.2.2 Reagents:

- HBSS containing 200 IU/ml penicillin, 200µg/ml streptomycin and 40 IU/ml Nystatin
- DMEM culture medium containing 10% (v/v) FBS, 2mM L-glutamine, 200 IU/ml penicillin, 200µg/ml streptomycin and 40 IU/ml Nystatin.
- Hyaluronidase (1 mg/ml in phosphate-buffered saline (PBS), 5ml/g cartilage)
- Trypsin (2.5mg/ml in PBS, 5ml/g cartilage)
- Collagenase (2.5mg/ml in DMEM containing 10% FBS, 3ml/g cartilage)

2.2.3 Digestion and Culture

After the articular cartilage was removed with a scalpel and cut into small sections, articular chondrocytes were extracted from the ECM by the sequential addition of digestive enzymes. The cartilage was first incubated in Hyaluronidase (to digest hyaluronan) for 15 minutes at 37°C. Following washing in PBS, Trypsin (to digest aggrecan) was then added for 30 minutes at 37°C. Finally, following PBS washes, the cartilage was incubated with collagenase at 35.5°C overnight. Following digestion, the cartilage sample was passed through a 100µm cell strainer (BD Biosciences, Oxford, UK) and cells were pelleted using a desktop centrifuge at 130g for 5 minutes. The media containing collagenase was removed; the cell pellet was washed in PBS and centrifuged as before. Finally, the cell pellet was resuspended in culture media as described above. HACs were seeded at a density of 40,000 cells/cm² and when 80-90% confluent, cells were harvested for nuclear protein extraction.

2.3 Whole Protein Extraction

2.3.1 Reagents

Cell Lysis Buffer - 50mM Tris, 10% (v/v) glycerol, 50mM NaF, 1mM EGTA, 1mM EDTA, 10mM glycerol phosphate, 1% (v/v) Triton X-100, 1x complete inhibitor cocktail, 1µM microcystin-LR and 1mM Na_3VO_4

2.3.2 Method

Cells were harvested at the relevant time points, washed twice and scraped using an 18cm cell lifter (Corning) into ice cold PBS. The cell lysate was centrifuged at 4°C at 11,000g for 5 minutes using a microcentrifuge. The cell pellet was resuspended in 30µl of cell lysis buffer and left on ice for 20 minutes. Lysates were centrifuged again at 11,000g at 4°C for 10 minutes and protein supernatants containing whole cell protein were stored at -80°C.

2.4 Nuclear Protein Extraction

Two buffers were used for the preparation of nuclear extract:

2.4.1 Hypotonic Buffer

10mM HEPES, pH 7.6, 1.5mM $MgCl_2$, 10mM KCl, 1mM DTT, 10mM NaF, 1mM Na_3VO_4, 0.1% Tergitol (v/v), 1x complete protease inhibitor cocktail tablet per 50mls solution (Roche, Welwyn, UK).

2.4.2 High Salt Buffer

20mM HEPES, pH 7.9, 420mM NaCl, 20% glycerol (v/v), 1mM DTT, 10mM NaF, 1mM Na_3VO_4, 1x complete protease inhibitor cocktail tablet per 50mls of buffer (Roche).

2.4.3 Method

Cells were seeded on 500cm^2 cell culture dishes (Corning), when 70-80% confluent, the culture medium was removed and cells were washed in ice cold PBS. Cells were then scraped into 5mls of fresh PBS and centrifuged for 30 seconds at 10,000g at 4°C. The cell pellet was resuspended in 1ml of hypotonic buffer and incubated on ice for 15 minutes. Cells were centrifuged as before and the supernatant containing cytosolic proteins was collected, snap-frozen on dry ice, and stored at -80°C. The pellet was resuspended in ice cold hypotonic buffer supplemented with 0.25M sucrose (in order to fractionate the nuclei). Again, cells were centrifuged and the pellet was resuspended in 1ml high salt buffer. Following a 30 minute incubation on ice, cells were centrifuged for

the final time at 10,000g for 2 minutes at 4°C and the supernatant containing nuclear protein was snap-frozen on dry ice and stored at -80°C.

2.5 Protein Quantification

Protein quantification was carried out using a Bradford Assay: a standard curve was generated by diluting bovine serum albumin (BSA) stock solution (2mg/ml BSA protein standard (Pierce & Warriner, Chester, UK)) in the lysis buffer used for extraction. Extracted proteins were diluted in lysis buffer and 300µl of Bradford Ultra (Expedeon, UK) was added to each sample. Following a 5 minute incubation, protein absorbances were read at 595nm using a Tecan sunrise microplate absorbance reader (Tecan, Reading, UK).

2.6 Electrophoretic Mobility Shift Assay (EMSA)

2.6.1 Reagents

- Fluorescently labelled (5´DY682) oligonucleotides (Eurofins MWG Operon, Ebersberg, Germany)
- EMSA Annealing Buffer - 100mM Tris HCl, 500mM NaCl, 10mM EDTA
- 5xTBE – 445mM Tris, 445mM Boric Acid, 10mM EDTA pH8.
- 5% (weight/volume (w/v)) acrylamide gel, prepared using TEMED and ammonium persulfate (APS)
- Odyssey Infrared EMSA kit, containing all buffers and loading dye (LiCor Biosciences, Cambridge, UK).
- rs143383 Optimal Binding Reaction – 1x Binding Buffer, 25mM DTT, 2.5% Tween-20, 1µg/µl Poly (dI:dC), 100mM $MgCl_2$, 200fmol annealed oligonucleotide and 5µg nuclear extract.
- rs143384 Optimal Binding Reaction- 1x Binding Buffer, 25mM DTT, 2.5% Tween-20, 1µg/µl Poly (dI:dC), 200fmol annealed oligonucleotide and 5µg nuclear extract.
- 5x (1pmol), 10x (2pmol), 25x (5pmol), 50x (10pmol) unlabelled competitor oligonucleotide (Competitor sequences are listed in Table 1A and B Appendix).
- 2µg species matched control or candidate antibody (Table 2, Appendix).

2.6.2 Method

PROMO 3.0 (http://alggen.lsi.upc.es/), TESS (http://www.cbil.upenn.edu/cgi-bin/tess/tess), and TransFac (http://www.gene-regulation.com/pub/databases.html) online databases were used to predict protein binding to the C and T-alleles of *GDF5*.

Fluorescently labelled oligonucleotides were resuspended in H_20 (Sigma-Aldrich) to a concentration of 100pmol/µl, the oligonucleotides were heated to 95°C for 5 minutes and then allowed to cool to room temperature for annealing, in a solution containing EMSA annealing buffer. Following this incubation, the reaction was allowed to cool to room temperature for 2-3 hours to enable annealing of the probes. The native acrylamide gel was prepared the day before the EMSA reaction and left to set at 4°C overnight. The gel is pre-run for 30 minutes in 0.5x TBE buffer prior to the loading of samples in order to remove any traces of APS, to equilibriate ions within the running buffer and to ensure constant gel temperature. The binding reaction was prepared as described above for rs143383 or rs143384 and incubated at room temperature for 20 minutes in the dark. When using unlabelled competitor oligonucleotides or antibody to confirm the binding of candidates, these are added prior to the addition of the fluorescently labelled oligonucleotide. Orange G loading dye is then added (1x final concentration) and samples are loaded onto the gel. Electrophoresis is performed at 100V at 4°C in the dark for approximately 4 hours, or until the dye front has reached the end of the gel. Visualisation was carried out using the LiCor Odyssey Infrared Imager (LiCor Biosciences).

2.7 Oligonucleotide Pull Down Assay

A 212bp region encompassing rs143383 was amplified by the polymerase chain reaction (PCR) with a biotinylated 5´primer and unlabelled 3´primer (Sigma-Aldrich) (Table 3, Appendix). Two independent PCRs were performed, using homozygous C or T template DNA. 40pmol of PCR product was then coupled to 2mgs of Streptavidin Dynabeads® (Invitrogen) following manufacturers instructions. A sample containing no DNA was used as a control. Beads were then blocked in 500µl BC-100 buffer (20mM HEPES pH 7.9, 20% (v/v) glycerol, 100mM KCl, 0.2mM EDTA, 0.5mM PMSF, 0.5mM DTT) plus 5% (v/v) BSA, 0.1mg/ml salmon sperm DNA (Invitrogen) and 0.01% (v/v) Triton X-100 for 1 hour at 4°C. Previously extracted SW782 nuclear lysates were transferred to a tube for dialysis (Tube-O-dialyzer, VWR, UK), and dialyzed in a low salt buffer (20mM HEPES pH 7.9, 20% (v/v) glycerol, 0.1M KCl, 0.2mM EDTA, 0.5mM PMSF, 0.5mM DTT) for 4 hours at 4°C. The buffer was replaced and the lysates were dialysed for a further 16 hours at 4°C. Dialyzed lysates were then pre-cleared for 1 hour with 50µl Streptavidin Dynabeads® (Invitrogen). DNA-bead complexes were then resuspended in 1mg of the prepared SW872 protein extract and incubated for 2 hours at 4°C with rotation. Following this, complexes were

washed once with blocking buffer and six times with BC-100 buffer alone. Protein-DNA complexes were eluted from the beads following incubation at 95°C for 5 minutes and isolated following magnetic separation. The samples (CC, TT and no DNA) were loaded on to a 12% (w/v) acrylamide gel, and subjected to electrophoresis at 150V for 1 hour followed by coomassie blue staining.

2.7.1 Mass Spectrometry

Quantitative mass spectrometry was performed as previously described [296]. Briefly, the gel pieces were destained with 100% acetonitrile, reduced in DTT and cysteines were alkylated using idodacetamide. Proteins were then digested, whilst still in the gel, overnight using trypsin, washed, and the final volume was reduced by speedvac. The three samples (CC, TT and no DNA) were then labelled using a TMT isobaric mass tagging kit (Thermo Scientific, Surrey, UK). Labelled samples were mixed prior to off-gel fractionation of the peptides. High pressure liquid chromatography was performed on a Pepmap C18 reversed phase column (Dionex). Mass spectrometry was performed on an Orbitrap mass spectrometer. Quantitative analysis was carried out using the ProteinExplorer software, version 1.0 (Thermo Scientific) and the search engine MASCOT (Matrix Science Company) was used for identification of proteins.

2.7.2 Analysis

This experiment was performed three times, however the samples from experiments two and three (C, T and no DNA) were combined as a 6-plex and thus processed concurrently. In each case, a technical repeat of each experiment was conducted to differentiate background artefacts from protein binding differences.

The combined results from all three experiments were first sorted according to background, i.e. proteins identified that were absent in the no DNA control sample were ranked highest, proteins that were present in the background sample in one of the three experiments were ranked second highest, down to those which were present in the background in all three experiments being ranked lowest. We analysed the data provided for each protein and defined the most robust hits as those with medium to high confidence values (data with low confidence values were excluded), based on the identification of more than 2 unique peptide sequences from the protein and the coverage of peptides in the protein (i.e. more than two peptides located in different regions of the protein. The functions of the proteins identified were determined using DAVID pathway analysis functional annotation tool available at

http://david.abcc.ncifcrf.gov/. Proteins known to have a role in transcriptional activation or repression were then prioritised for further analysis.

2.8 Chromatin Immunoprecipitation (ChIP)

2.8.1 Crosslinking

ChIP was performed using the EZ Magna ChIP A kit (Millipore, UK), all subsequent reagents described were provided with the kit unless otherwise stated. SW872 cells were seeded at a density of $15x10^6$ on a $500cm^2$ square cell culture dish (Corning) in 80mls of SW872 cell media. Formaldehyde (Sigma) was added to crosslink cells at a final concentration of 1% and incubated at room temperature for 10 minutes. Glycine was added at a final concentration of 0.125M to quench un-reacted formaldehyde. Media was discarded and the dish was placed on ice. Cells were washed with 30mls PBS and then scraped into 30mls PBS containing 1x Protease Inhibitor Cocktail II. Cells were pelleted at 720g at 4°C for 8 minutes using a desktop centrifuge.

2.8.2 Lysis and Sonication

The cell pellet was then resuspended in 1.5mls sodium dodecyl sulphate (SDS) cell lysis buffer containing 1x protease inhibitor cocktail II and incubated on ice for 20 minutes. The lysate was divided into three 1.5ml Eppendorf tubes and sonication was carried out using a probe sonicator (Soniprep150, MSE, London, UK), on ice, at 8 microns for 10x10 second intervals with 20 second rest periods. Sonicated lysates were then centrifuged at 10,000g at 4°C for 10 minutes.

2.8.3 Immunoprecipitation

100µg of chromatin was diluted 1 in 10 with ChIP dilution buffer containing 1x protease inhibitor cocktail II and pre-cleared using 50µl magnetic protein A beads (Millipore). 10µg of primary antibody (Table 2, Appendix) was used for each immunoprecipitation, in addition to 40µl protein A magnetic beads. The samples were then incubated overnight at 4°C with rotation. The magnetic beads were pelleted using a magnetic separator (Millipore) and the supernatant removed. The pellet was then washed in low salt immune complex wash, high salt immune complex wash, lithium chloride complex wash and finally Tris-EDTA buffer, each for 5 minutes with rotation, with beads being separated after each wash and the supernatant discarded. Following the final wash, the beads were resuspended in 100µl ChIP elution buffer in addition to 1µl Proteinase K and incubated at 62°C for 2 hours with shaking for the reversal of crosslinks. Following an incubation of 95°C for 10 minutes, samples were cooled to

room temperature and the supernatant removed to a new tube following separation using the magnetic separator. DNA purification was carried out using the spin columns provided in the kit.

2.8.4 Polymerase Chain Reaction

DNA was amplified by PCR using primers located in the first exon of *GDF5* (Primer sequences are listed in Table 3, Appendix). 2µl of immunoprecipitated DNA was added to a 13µl PCR mastermix containing 0.5µM forward and reverse primers, PCR buffer (50mM KCL, 10mM Tris-HCL pH 8.3), 2mM $MgCl_2$, 0.2mM dNTPs, 0.08 units of AmpliTaq Gold DNA polymerase (Applied Biosystems, ABi, Life Technologies, UK) and H_20. The thermo-cycling conditions included a denaturation stage for 14 minutes at 94°C, followed by 32 cycles of 94°C for 30 seconds, annealing at 57°C for 30 seconds, and extension at 72°C for 30 seconds and a final step of 72°C for 5 minutes.

2.9 RNA Mediated Interference (RNAi)

SW872 cells were seeded at a density of 350,000 cells per well in a 6-well culture plate (Corning). After 24 hours, cells were transfected using Dharmacon ON-TARGET*plus* Smartpool siRNAs (100nM final concentration) targeted against candidates in addition to Dharmafect™ 4 lipid reagent (Dharmacon, Thermo Scientific, Surrey, UK). After the desired period of incubation (48 hours unless otherwise stated) the cells were harvested, nucleic acid and protein isolated and RNA reverse transcribed as described below. Depletion of mRNA expression was calculated in comparison to cells transfected with the ON-Target*plus* Non-Targeting Pool control siRNA (Dharmacon).

For the combination siRNA experiments examining protein expression (Chapter 8) the transfections were carried out as above using 100nM or 200nM final siRNA concentrations and 1x or 2x concentrations of Dharmafect™ 4 lipid reagent (Dharmacon). For the combination siRNA experiments examining gene expression, cells were seeded at a density of 10,000 cells per well in a 96-well culture plate, and transfected using the same final concentrations as above.

2.10 Nucleic Acid Extraction

Genomic DNA, total RNA and total protein were simultaneously extracted from SW872 cells following siRNA treatment using a spin column extraction kit according to the manufacturer's instructions (Nucleospin Triprep, Macherey-Nagel, supplied by Thermo Fisher, UK). Nucleic acids were quantified using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, USA). RNA was diluted to

250ng/µl and stored in diethylpyrocarbonate (DEPC)-treated H_2O (Invitrogen) at -80°C. DNA and protein were stored at -20°C.

2.11 Reverse Transcriptase Polymerase Chain Reaction (RT-PCR)

RNA was DNase treated prior to RT-PCR. RT-PCR was completed using the SuperScript First-Strand Synthesis System (Invitrogen); 1µg of RNA was DNase treated with 2 units of Turbo DNase (Ambion, Life Technologies, UK) and incubated at 37°C for 30 minutes. RNA was then mixed with 150ng random primers and 10mM dNTP mix (Invitrogen), incubated at 65°C for 5 minutes and placed on ice. 1x reaction buffer (50mM Tris-HCl (pH 8.3), 75mM KCl, 3mM $MgCl_2$; Invitrogen), 5mM $MgCl_2$ (Applied Biosystems), 10mM DTT (Invitrogen) and 40 units RNaseOUT inhibitor (Invitrogen) were added and incubated at 25°C for 1 minute. Superscript II enzyme (50 units) (Invitrogen) was then added and mixed by pipetting. The final 20 µl reaction was then incubated at 25°C for 10 minutes, 42°C for 50 minutes and finally 70°C for 15 minutes. Following this, 2 units of RNase H (New England Biolabs (NEB), Hitchin, UK) was added and a final incubation of 37°C for 20 minutes was performed. Complementary DNA (cDNA) was stored at -20°C prior to Real Time PCR.

For the combination siRNA experiments performed in 96-well culture dishes, 48 hours post transfection cells were lysed using the Cells-to-cDNA II cell lysis buffer (Ambion). Samples were heated to 75ºC for 15 minutes and stored at -80ºC prior to cDNA synthesis. The samples were then DNase treated (Qiagen) and incubated at 37ºC for 15 minutes and 75ºC for 5 minutes. 10ng random primers (Invitrogen) and 10mM dNTP mix (Invitrogen) were added to 8µl of cell lysate, which was then incubated for 5 minutes at 70ºC. Following this, 1x reaction buffer (50mM Tris-HCl (pH 8.3), 75mM KCl, 3mM $MgCl_2$; Invitrogen), 100 units MMLV reverse transcriptase (Invitrogen) and 10mM DTT (Invitrogen) was added to the reaction, followed by water to bring the final reaction volume to 20µl. The samples were then incubated at 37 °C for 50 minutes and 75 °C for 15 minutes. Finally, 30µl of water was added and the samples were then frozen at -20ºC prior to analysis by Real Time PCR.

2.12 Polymerase Chain Reaction

Following synthesis, the integrity of cDNA was investigated by PCR amplification of *HBP1*. The *HBP1* primers, located in different exons, are used to distinguish between cDNA and any residual genomic DNA (gDNA) contamination (primer sequences listed in Table 3, Appendix). A 15µl PCR reaction was performed, as described previously, with minor alterations; 0.5µl of cDNA was added to 14.5µl of master-mix, 35 PCR

cycles were performed and the annealing temperature was 60°C. Following the PCR, products were separated by electrophoresis through a 2% (w/v) agarose gel. Visualisation was carried out by ethidium bromide staining using a Syngene gel dock. Only cDNA samples containing the correct sized product were taken forward for gene expression analysis.

2.13 Gene Expression Analysis by Real Time PCR

Two different types of gene expression analysis by Real Time PCR were performed.

2.13.1 Taqman Real Time PCR

Gene expression analysis of *GDF5* and the expression of depleted candidates following RNAi was determined using a Taqman primer-probe based approach by Real Time PCR. The probe has a 5´ reporter dye and a 3´ quencher and when the polymerase enzyme cleaves the probe from DNA during extension, the reporter dye is released from the quencher and emits fluorescence.

Gene expression assays were purchased from either Applied Biosystems (Applied biosystems; *GDF5*) or Integrated DNA Technologies (IDT, Belgium). 10µl reactions were prepared containing 1x gene expression assay, 1x Taqman Gene Expression mastermix (containing AmpliTaq Gold® DNA Polymerase, Uracil-DNA Glycosylase, dNTPs with dUTP, Passive Reference 1 and optimized mix components) and cDNA (Undiluted for *GDF5*, diluted 1:20 for all other genes). The Prism 7900HT sequence detection system (Applied biosystems) was used for PCR cycling (50°C for 2 minutes, 95°C for 10 minutes and 40 cycles of 95°C for 15 seconds and 60°C for 1 minute) and detection. A fast mastermix, Taqman Fast Universal Mastermix (Applied biosystems), in addition to fast cycling conditions (95°C for 20 seconds followed by 40 cycles of 95°C for 1 second and 60°C for 20 seconds) were used for the expression analysis of *Sp1*, *Sp3*, *P15*, *DEAF-1*, and the house keeping genes. For analysis, gene expression was normalised to the house keeping gene *HPRT1* or *GAPDH* using the delta ct method ($2^{-(\text{ct test gene})-(\text{ct HPRT1})}$).

2.13.2 Allelic Expression Analysis

Differential Allelic Expression (DAE) analysis, to assess the expression of the C and T alleles of rs143383, was performed using a custom SNP genotyping assay (Applied biosystems) containing forward and reverse primers and allele specific probes (VIC or FAM labelled). A 10µl reaction was prepared with 1x DAE assay, 1x Taqman Universal Mastermix II, no UNG (Applied biosystems), and either cDNA diluted 1 in 5 or 20ng

DNA. Cycling conditions were 50°C for 2 minutes, 95°C for 10 minutes and 40 cycles of 92°C for 15 seconds and 60°C for 1 minute. Analysis was performed using the equation ($2^{\Lambda-ct}$) for FAM and VIC and the ratio between the two alleles calculated (FAM/VIC) for both cDNA and gDNA. The cDNA allelic ratio was normalised to the gDNA C/T allelic ratio (representing a 1:1 ratio) for each treatment group. The genotype of cell lines used for this research at rs143383 was also determined using this assay, with detection of fluorescence of just one reporter indicating a homozygous sample (CC or TT) and fluorescence of both FAM and VIC reporters indicating a heterozygote sample (C/T).

Sequences of all Real Time PCR primers and probes are detailed in Table 4, Appendix.

2.14 Immunoblotting

For the assessment of candidate protein depletion and over expression, total protein was isolated using the Triprep spin column method or whole protein extraction method respectively. Protein was then quantified as described, diluted to 10µg in 1x laemmli sample buffer (0.1M Tris-HCL, 0.35M SDS, 20% (v/v) glycerol, 0.01% (v/v) bromophenol blue, 5% (v/v) β-mercaptoethanol) and resolved on 10% (w/v) SDS polyacrylamide gels. Proteins were then transferred to a Immobilon-P polyvinylidene fluoride (PVDF) membrane (Merck Millipore, Nottingham, UK) by electroblotting in a Scie-Plas V20-SDB 20x20 semi-dry blotter (Scie-Plas, Cambridge, UK). After transfer, the membrane was blocked for non-specific protein binding by incubation in PBS containing 5% (v/v) Marvel milk and 0.02% (v/v) Tween-20 at room temperature for 30 minutes. The membrane was washed and incubated overnight with primary antibodies diluted in PBS/Tween-20 containing 5% (v/v) marvel at 4°C. Details of the antibodies are shown in Table 2, Appendix. Following antibody incubation, the membranes were washed in PBS/Tween-20, before incubation in secondary polyclonal goat anti-rabbit (1:2000) or polyclonal goat anti-mouse (1:10000) horseradish peroxidise (HRP) conjugated antibodies for 1 hour at room temperature (DAKO, UK). Detection was carried out using X-ray film (Kodak Film, Sigma-Aldrich) with ECL, ECL plus and ECL Advance detection reagents (GE Healthcare, Little Chalfont, UK). A monoclonal βActin antibody was used as a loading control (Abcam, Cambridge, UK).

2.15 Cloning

2.15.1 *GDF5* Luciferase Vectors

The *GDF5* promoter and part of the 5´UTR region spanning -97 to +305 (relative to the transcriptional start site of *GDF5*) was subcloned from the *GDF5* pGL3 plasmid [297] into the *Mlu/BglII* sites of the purified pGL3-Enhancer Vector (Promega, UK). The double digestion was performed using 15 units of *Mlu1* and 15 units of *BglII* restriction enzymes (New England Biolabs (NEB) UK), in addition to 1x NEB Buffer 3 (50mM Tris-HCl, 10mM $MgCl_2$, 100mM NaCl and 1mM DTT, pH 7.9) and 5µg of *GDF5* pGL3 plasmid for 3 hours at 37°C. The pGL3 enhancer vector was also digested in this manner. Following digestion, the PCR product, and pGL3 enhancer vector were purified using the QIAquick gel extraction kit (Qiagen, Crawley, UK) following manufacturers instructions and eluted in H_20. The digested promoter region was ligated into the digested pGL3 enhancer vector during a 12 hour incubation at 16°C, using a fragment: plasmid ratio of 5:1, in addition to T4 ligase (NEB) (400 units). Plasmid DNA was then transformed into MACH1 competent cells (Invitrogen) following manufacturers guidelines. Transformed cells were spread onto agar plates supplemented with ampicillin (100µg/ml) (Sigma-Aldrich) and incubated overnight at 37°C. Positive colonies were added to 3.5mls of LB broth (Sigma-Aldrich) supplemented with ampicillin (100µg/ml) and incubated at 37°C with shaking overnight. Glycerol stocks were prepared with the addition of 200µl of bacterial culture to 200µl of glycerol (Sigma-Aldrich) in a 1:1 ratio, and stored at -80^{o}C. Plasmid cultures were purified with reagents from the Plasmid Maxiprep Kit (Qiagen) using the following miniprep protocol; the culture was centrifuged at 10,000g for 3 minutes and the bacterial pellet was resuspended in 100µl P1 + RNase A solution (resuspension buffer), followed by the addition of 200µl P2 solution (lysis buffer) and the tubes were inverted to mix. After a 2 minute incubation, 150µl P3 solution (neutralization buffer) was added, mixed and centrifuged for 5 minutes at 17,900g. The supernantant was removed and added to 1ml of 100% ethanol. Following a 10 minute incubation at -80°C, the supernantant/ethanol mixture was centrifuged for 10 minutes at 17,900g. The DNA pellet was air dried for 10 minutes to remove residual ethanol, resuspended in 30µl H_20 (Sigma-Aldrich) and stored at -20°C. Successful insertion into the pGL3 enhancer vector and sequence verification was assessed following sequencing by Genome Enterprises (Norwich, UK).

2.15.2 Over expression Vectors

The *Sp1, Sp3* and *P15* open reading frames (ORF) were amplified from cDNA using gene specific primers containing either an *EcoR1* restriction or a *SacII* restriction site at the 5´ end of the primer (Table 5, Appendix). The PCR was performed using Titanium taq (1x final concentration; Clontech, Saint-Germaine-en-Laye, France), with the reaction containing 1x Titanium Taq Buffer (40mM Tricine-KOH, 16mM KCl, 3.5mM $MgCl_2$, 3.75µg/ml BSA), 0.2µM forward and reverse primers, 0.2mM dNTPs with cycling conditions of 95^{o}C for 1 minute followed by 40 cycles of 95^{o}C for 15 seconds, 63^{o}C for 1 minute and 68^{o}C for 1 minute 30 seconds and finally 68^{o}C for 7 minutes. After PCR amplification, the PCR product was purified and eluted in 30µl H_20 using the QIAquick PCR purification kit following the manufacturer's instructions (Qiagen). The digestion, purification, ligation and transformation reaction of the *Sp1, Sp3* and *P15* PCR products and EGFP-N1 Vector (Clontech) was as previously described for the *GDF5* pGL3 enhancer vector, with the following minor alterations: the restriction enzymes *EcoR1* and *SacII* were used, and the antibiotic kanamycin was used for resistance at a final concentration of 50µg/ml. The DEAF-1 EGFP-N1 and empty EGFP-N1 expression plasmids were kindly donated by C. Garrison Fathman and Linda Yip [298].

2.15.3 Site Directed Mutagenesis

To generate the four possible rs143383 and rs143384 haplotypes (C-C, C-T, T-C, T-T), the Agilent Quickchange II site directed mutagenesis kit (Agilent, Berkshire, UK) was used. A control reaction containing the pWhitescript control plasmid was performed following the manufacturer's instructions. The 50µl mutagenesis reactions contained 50ng of plasmid DNA, 125ng of forward primer, 125ng of reverse primer, 0.01mM dNTP mix, 1x reaction buffer and 2.5 units of *Pfu* Ultra high fidelity DNA polymerase. The cycling conditions consisted of 95^{o}C for 30 seconds followed by 18 cycles of denaturing at 95^{o}C for 30 seconds, annealing at 50^{o}C for 1 minute and extension at 68^{o}C for 7 minutes. The reaction was then placed on ice for 2 minutes. DNA was digested at 37°C for 1 hour following the addition of 10 units of *Dpn*I. The transformation was carried out following manufacturers instructions. Positive colonies were incubated, purified by miniprep and sequenced as described above.

2.15.4 Purification of Plasmid DNA Prior to Transfection by MaxiPrep

Prior to transfection, large quantities of purified plasmid DNA were prepared. First, 5µl of glycerol stock was added to 5mls of LB broth containing ampicillin (100µg/ml) or

Kanamycin (50µg/ml) and incubated at 37°C with shaking for 2-3 hours. This culture was then transferred to a large conical flask containing 250mls LB broth supplemented with antibiotics and incubated at 37°C with shaking overnight. The cultures were then centrifuged at 6,000g for 15 minutes at 4°C. All buffers described are part of the Qiagen Maxiprep Kit (Qiagen). Pellets were resuspended in 10mls of P1 + RNase, followed by 10mls P2, incubated for 5 minutes, and then 10mls chilled P3 and incubated for 20 minutes on ice. The lysate was then centrifuged at 20,000g for 30 minutes at 4^{o}C, and the supernatant transferred to a new tube and centrifuged under the same conditions for a further 15 minutes. A QIAGEN-tip 500 gravity column was equilibrated with 10mls of Buffer QBT, following this the supernatant was applied to the column. The column was subsequently washed twice with 30mls of Buffer QC, and plasmid DNA eluted using 15mls of Buffer QF into a fresh centrifuge tube. DNA was precipitated following the addition of 10.5mls of isopropanol and the sample was then centrifuged at 15,000g for 30 minutes at 4^{o}C. Once the residual isopropanol had evaporated, the pellet was resuspended in 500µl of H_2O. 1250µl (2.5x volume) of 100% ethanol was added and after mixing was stored at -80^{o}C for 15-20 minutes. Following a final centrifugation of 17,900g for 10 minutes at room temperature, the pellet containing DNA was air dried for 5-10 minutes prior to resuspension in 300µl of H_2O (Sigma-Aldrich) and quantified using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies).

2.16 Transfection of Cell Lines

SW1353 cells were seeded at a density of 17,500 cells per well in a 48-well cell culture plate (Costar, UK) and cultured for 48 hours prior to transfection. Cells were transfected with 2µg of plasmid DNA (containing 1µg of *GDF5* pGL3 enhancer vector and several combinations of either 1µg empty pEGFP-N1 vector, 500ng Empty EGFP-N1 and one of the transcription factor expression plasmids, or 500ng each of two transcription factor expression plasmids) in addition to 15ng of pTK-RL Renilla using ExGen500 *in vitro* transfection reagent (Fermentas, York, UK). Four wells were transfected per condition and a total of three individual experiments were performed. After 24 hours, cells were lysed as described in section 2.17.

For the examination of protein over expression, 250,000 SW1353 cells per well were seeded in 6-well culture dishes and transfected with plasmid vectors and ExGen500 as described for the 48-well cell culture plate, but the relative amounts of each were increased according to the final culture volume. After 24 hours, protein was extracted as described in section 2.3.

For the combination over expression and depletion experiments (Chapter 8), SW1353 cells were seeded at a density of 17,500 cells per well in a 48-well cell culture plate (Costar) and 350,000 cells per well in a 6-well cell culture plate (Costar). Cells were cultured for 48 hours prior to transfection. Cells were transfected with siRNA (100nM final concentration) or 1µg plasmid DNA or both in combination using Dharmafect^{TM} Duo transfection reagent (Dharmacon, Thermo Scientific). After 48 hours, cells in the 48-well plate were lysed as described in section 2.17 and protein was extracted from cells in the 6-well culture plate as described in section 2.3.

2.16.1 DEAF-1 EGFP Transfections

For the over expression of DEAF-1 EGFP prior to ChIP, SW872 cells were seeded at $15x10^6$ on a $500cm^2$ plate and cultured for 24 hours. Cells were transfected with DEAF-1 EGFP plasmid DNA using Fugene HD transfection reagent (Promega) in a 1:3 ratio. After 24 hours the cells were taken forward for ChIP.

For the over expression of DEAF-1 EGFP prior to immunoblotting, SW872 cells were seeded at 250,000 cells per well in a 6-well culture dish. After 24 hours, cells were untransfected, transfected with DEAF-1 EGFP or transfected with a control plasmid H11 using Fugene HD transfection reagent (Promega) in a 1:3 ratio. After 24 hours protein was extracted as previously described in section 2.3.

For over expression of DEAF-1 EGFP followed by DEAF-1 siRNA treatment, cells were seeded at a density of 250,000 per well in 6-well culture dishes. After 24 hours, cells were transfected with DEAF-1 EGFP using Fugene HD transfection reagent (Promega) in a 1:3 ratio. After 6 hours the transfection media was removed and cells were transfected with siRNA as previously described in section 2.9. Cells were then incubated for 48 hours prior to protein extraction.

2.17 Luciferase Activity Reading

Transfected cells were lysed and luciferase and renilla activity measured using the Dual Luciferase Assay system (Promega) with the MicroLumat Plus LB96V luminometer (Berthold Technologies UK, Harpenden, UK).

2.18 Immunofluorescence

For the verification of DEAF-1 depletion by immunofluorescence, cells were seeded at a density of 10,000 cells per well of a chamber slide (Nagel Nunc International, USA). After 24 hours, cells were either transfected with DEAF-1 siRNA (Dharmacon) (100nM) or untreated. After 48 hours, the media was removed and cells were washed in

PBS and fixed with 4% (w/v) paraformaldehyde in PBS. Permeabilisation buffer was then added (0.5% BSA, 0.2% Fish skin gelatine, 0.5% Triton X-100) and cells were incubated in each of the anti-DEAF-1 antibodies overnight at 4°C. Cells were then washed and incubated in either secondary Red fluorescent AlexoFluor 594nm or Green fluorescent AlexoFluor 488nm goat anti-rabbit IgG antibodies (Invitrogen molecular probes, Invitrogen) for 1 hour at room temperature, washed in PBS and mounted using vectashield with DAPI (4'6-diamidino-2-phenylindole)(Vector Laboratories, Burlingame, CA). Fluorescence was detected using a LEICA DMLB fluorescent microscope and a SPOT-RT camera.

To examine the over expression of EGFP-N1 vectors by immunofluorescence, SW1353 cells were seeded at a density of 10,000 cells/well in a chamber slide (Nagel Nunc International). After 48 hours, cells were transfected with 1µg of each candidate plasmid vector using ExGen500. 24 hours post transfection, cells were washed in PBS and fixed with 4% (w/v) Paraformaldehyde in PBS for 10 minutes, washed again in PBS and mounted and detected as above.

2.19 Immunoprecipitation

Cells were seeded at a density of 2.5 million in a 10cm^2 cell culture dish (Cellstar, Greiner Bio-One, Germany). After 24 hours cells were transfected with 5µg DEAF-1 EGFP plasmid using Fugene HD transfection reagent (Promega). After 24 hours, whole protein was extracted as previously described in section 2.3. For the immunoprecipitation, 10µg of each DEAF-1 antibody or the IgG control antibody was added to 500µg of cell lysate and the sample was incubated at 4ºC with rotation overnight. 70µl of protein A magnetic beads (Millipore) were added to the samples and incubated at 4ºC for four hours with rotation. The beads were separated using a magnet and washed with lysis buffer twice, and PBS once. The beads were then resuspended in 2x laemmli buffer and incubated at 100ºC for 5 minutes. The beads were once again separated using the magnet, and the supernatant collected. Immunoblotting was carried out as described in section 2.14 using the anti-EGFP antibody.

2.20 Statistics

Pairwise analysis p-values were calculated using the Students two tailed t-test. *p<0.05 **p<0.01 ***p<0.001. For the comparison of multiple treatments, statistical significance was calculated using an ANOVA. *p<0.05 **p<0.01 ***p<0.001.

Chapter 3: Establishing the Conditions for Testing DAE

3.1 Introduction

The rs143383 SNP is functional; the T allele produces less *GDF5* transcript in comparison to the C allele both *in vitro* and *in vivo* [134, 267]. Carriers of the T allele (TT or CT) are more susceptible to developing OA in comparison with C allele carriers, with odds ratios ranging from 1.12 in Caucasians to 1.79 in Asians [134, 267]. In order to investigate the reason for the differential expression of the two alleles, I first needed to design and validate an assay that would enable me to quantify the expression of the two alleles relative to one another. Secondly, I sought to identify a cell line that expressed *GDF5* and demonstrated a similar allelic expression imbalance to that observed previously in patients. Using this cell line and assay, my aim was to identify the causes of the allelic expression imbalance and investigate further how the DAE can be modulated.

3.2 Aims

- To design and characterise an assay to quantify DAE.
- To identify a cell line in which we can study the rs143383 DAE.

3.3 Results

3.3.1 Identification of a cell line for investigation of rs143383 DAE

The human cell lines that are available within our research group had been genotyped for rs143383 and rs143384 by Athanasia Gravani (research technician) and the expression of *GDF5* within each cell line has been investigated previously [299]. The genotypes of the cell lines that express *GDF5* are detailed in Table 3.1. The liposarcoma SW872 cell line is compound heterozygous for rs143383 and rs143384, and expresses *GDF5* and was thus used for the analysis of *GDF5* DAE.

Cell Line Name	Origin	rs143383 genotype	rs143384 genotype
CH8	articular chondrocyte	TT	TT
HeLa	cervix carcinoma	CT	CT
HOS-TE85	osteosarcoma	CC	CC
MDA-MB-231	breast adenocarcinoma	TT	TT
MG63	osteosarcoma	TT	TT
SaOS-2	osteosarcoma	TT	TT
SW1353	chondrosarcoma	CC	CC
SW872	liposarcoma	CT	CT
U937	histiocytic lymphoma	CC	CC

Table 3.1
The 9 cell line stocks available in my group for *GDF5* expression analysis. Modified from Reynard et. al 2011.

3.3.2 Design and validation of a rs143383 DAE assay

A real time PCR assay was designed with primers located in exon 1 of the *GDF5* gene encompassing rs143383 and with two differently labelled probes able to discriminate between the C (FAM label) and T alleles (VIC label) of the SNP, as described in Table 4, Appendix. This assay has been designed to enable the amplification of both DNA and cDNA.

I first assessed the ability of the assay to distinguish between C and T alleles of rs143383. I used DNA from the heterozygous cell line, SW872, the homozygous CC cell line, SW1353 and the homozygous TT cell line, CH8. Additionally, DNA from two patients was used, one of whom was homozygous CC and the other homozygous TT. Figure 3.1 shows the real time PCR amplification graphs. The CC CH8 cell line and the CC patient DNA were amplified efficiently using the real time PCR assay and the FAM reporter fluorescence, signifying the C allele probe detected the C allele DNA. Although there was a small amount of signal detection from the VIC reporter, this did not pass the cycle threshold (set automatically by the real time PCR machine) and was thus

undetected. Similarly, the TT SW1353 cell line and the TT patient DNA was detected using the assay by the VIC reporter fluorescence, signifying the T allele probe, detected the T allele DNA and again, although there was a small amount of detectable FAM reporter, this did not pass the cycle threshold. Both FAM and VIC reporter fluorescence were detected using the heterozygous SW872 cell line DNA, and both passed the cycle threshold, confirming that this assay is able to efficiently genotype the rs143383 SNP.

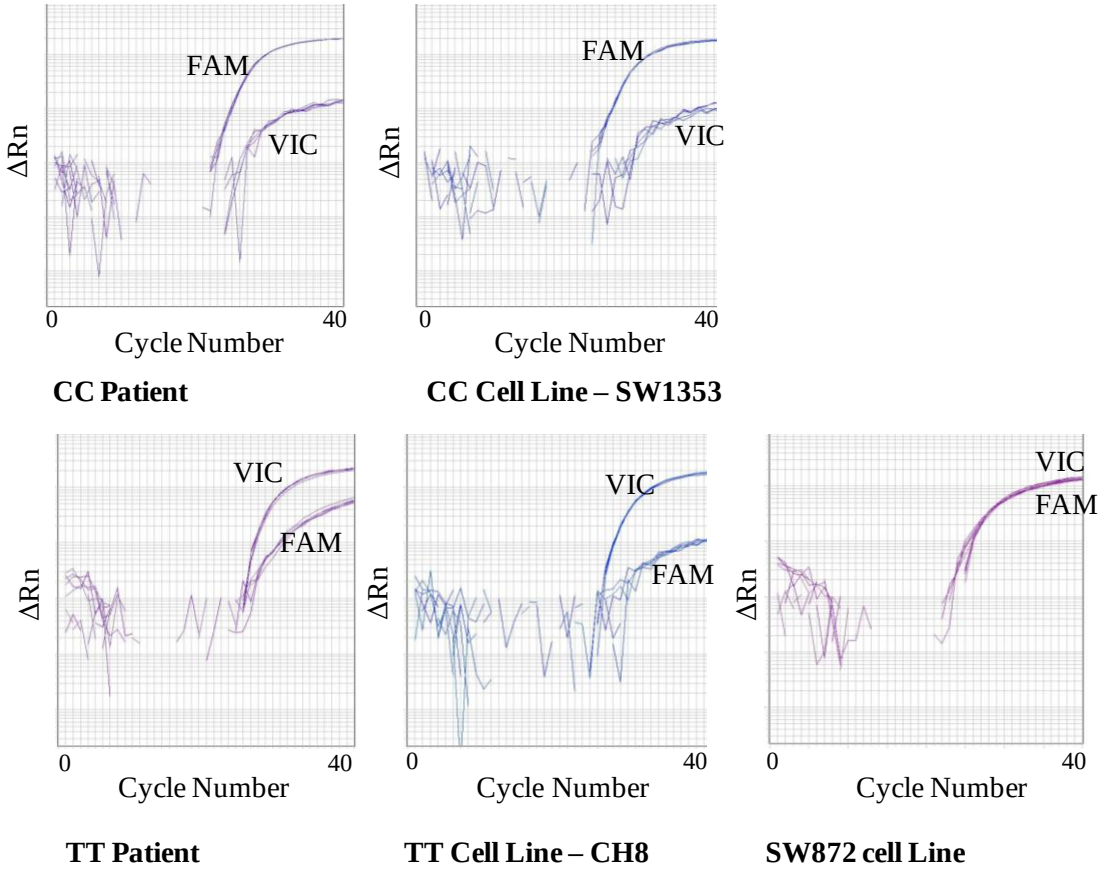


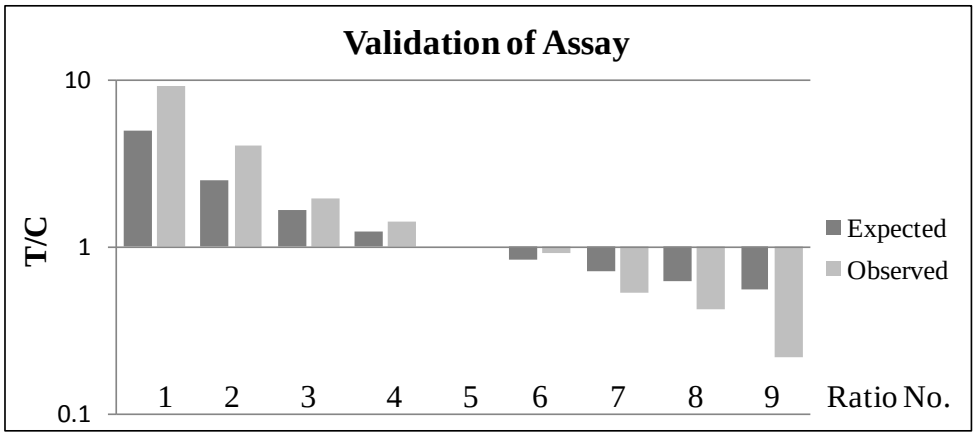
Figure 3.1
Real Time RT PCR amplification plots. Detection of both VIC (T) and FAM (C) labelled dyes are shown. Cycle number is shown on the x-axis and ΔRn (fluorescence emission intensity of the reporter dye/fluorescence emission intensity of the passive reference dye) on the y-axis. 20ng of DNA was used for each reaction, including patient DNA homozgous CC or TT at rs143383 and cell line DNA from SW1353 (CC), CH8 (TT) and SW872 (C/T).

I then wanted to assess if the assay was able to provide an accurate measurement of the amount of both alleles relative to one another. This was assessed by mixing CC and TT DNA of known concentration in various ratios. Figure 3.2A shows the 9 ratios used for the analysis, the expected T/C value and the observed T/C value. Figure 3.2B shows the expected and observed values relative to one another. The observed values are consistently higher in all DNA concentrations except for concentration number 6, where the observed value is lower than expected. In order to correct for the differences in the expected and observed values, the values were plotted against one another as shown in Figure 3.2C. The equation $y=mx + c$ was then used to correct for the observed values, with y being the corrected value, and substituting the graphical equation: corrected = (observed*0.4893) + 0.4931. All real time analysis using the DAE assay was corrected for in this manner.

A

Number	Ratio of C:T	Expected T/C	Observed T/C
1	0.2:1	5	9.32
2	0.4:1	2.5	4.02
3	0.6:1	1.67	1.96
4	0.8:1	1.25	1.42
5	1:1	1	1
6	1.2:1	0.83	0.92
7	1.4:1	0.71	0.53
8	1.6:1	0.625	0.41
9	1.8:1	0.56	0.21

B



C

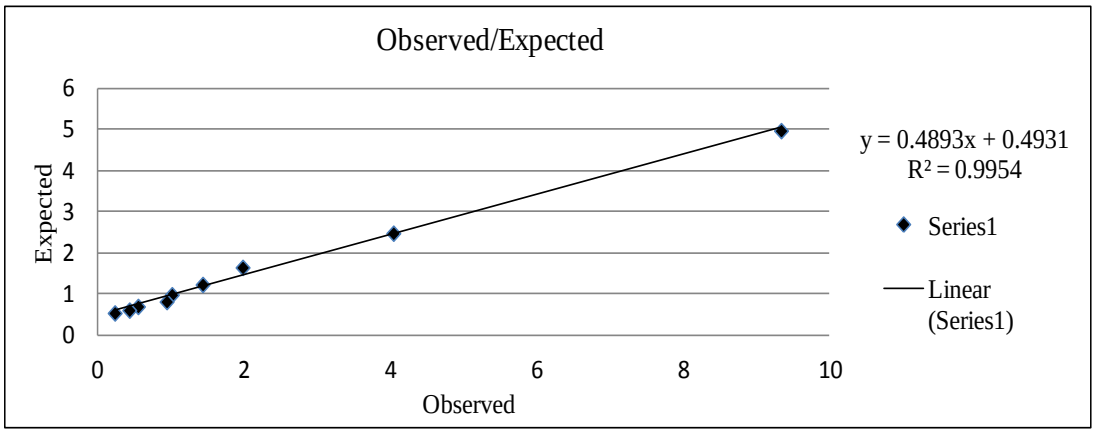


Figure 3.2

Validation of allelic expression assay. A. Table detailing the 9 ratios of TT and CC DNA, generated by mixing 20ng CC and TT DNA. The 3rd column shows the T/C expected ratio and the fourth column shows the amount of the T allele relative to C obtained following Real Time PCR using the assay. B. The expected and observed T/C ratios, plotted as ratios 1-9 on the x-axis against VIC (T allele)/ FAM (C allele). C. Observed/Expected T/C ratios: the x-axis shows the observed values, plotted against expected values on the y-axis. A linear treadline is shown. The equation generated was used to correct for the differences in observed/expected.

This assay was then used to assess the level of expression of both the rs143383 C and T alleles relative to one another in the heterozygous SW872 cell line as described in Materials and Methods section 2.13.2. The expression of the C allele relative to the T allele is shown in Figure 3.3. The C allele is expressed at a higher level in comparison to the T allele; this cell line is thus exhibiting DAE and as such is recapitulating the DAE that is observed in rs143383 heterozygous OA patients. In the SW872 cells, for every 1 T, there are 1.5 C transcripts. This cell line could therefore act as a model system for investigating DAE.

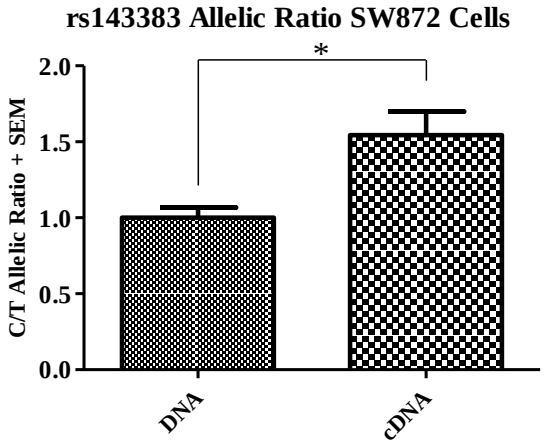


Figure 3.3
DAE of *GDF5* in SW872 cells assessed using rs143383. The C/T allelic ratio for genomic DNA (DNA) and complementary DNA (cDNA) are shown. Genomic DNA was normalised to 1.0 and then used to compare against the C/T allelic ratio obtained for cDNA. Error bars denote the standard error of the mean (SEM). * $p<0.05$, calculated using a Student's 2 tailed *t*-test. n=6.

3.4 Discussion

I have designed and validated an assay that can be used to quantitatively assess the expression of the rs143383 C and T alleles relative to one another. Furthermore, I have identified a cell line, SW872, that is heterozygous at rs143383, that expresses *GDF5* and which exhibits *GDF5* DAE, with the T allele showing a reduction in the expression of *GDF5* relative to the C allele. This cell line can now be used to dissect the mechanism behind the DAE mediated by rs143383.

Chapter 4: The Identification of Sp1 and Sp3 as *GDF5 Trans*-acting Factors by EMSA

4.1 Introduction

Previous functional research performed on rs143383 used EMSA analysis to investigate the binding of proteins to this polymorphic region. They reported a difference in mobility of the radioactively labelled C and T allele probes in three cell lines, MG63, CH8 and SW872, suggesting a difference in protein binding. Using competition EMSAs the T allele was found to be binding a unique complex that was unaffected by C allele competition. By searching online databases, a number of factors were identified that were predicted to bind in an allele specific manner to rs143383. The group focussed on three factors, the Myb protooncogene protein (c-Myb), early growth response protein 1 (EGR-1) and Deformed epidermal autoregulatory factor 1 (DEAF-1). Using consensus competitors for each factor, the binding of proteins to the *GDF5* C and T allele probes by EMSA was investigated. The DEAF-1 consensus competitor was able to compete the binding of the T allele protein complex but not the C allele protein complex. C-Myb and EGR-1 competitors did not compete binding of either complex. Furthermore, binding to a DEAF-1 consensus labelled probe was competed with unlabelled T allele competitor and to a lesser extent to with unlabelled C allele competitor.

The role of DEAF-1 was investigated further in cotransfection experiments using *GDF5* promoter constructs and a DEAF-1 over expression vector. Over expression of DEAF-1 repressed the expression of both the C and T allele luciferase constructs and in line with the stronger binding of DEAF-1 observed in the EMSA experiments, the T allele was significantly more repressed compared to the C allele. DEAF-1 was therefore identified as a *trans*-acting factor predicted to bind differentially to the two alleles of rs143383 [272].

The rs143383 polymorphism is located within the 5´UTR of the *GDF5* gene at +41bp relative to the transcription start site. There is little known with regards to transcription factor binding within this region and overall about the transcriptional regulation of *GDF5*. The transcription factor YY1 has been identified as a transcriptional activator binding upstream of the polymorphism at -41bp relative to the transcription start site and modulates the expression of *GDF5* [297, 300]. Due to the fact that a reduction in the expression of *GDF5* during development in the tissues of the joint could affect normal development of the joint and mediate the increased susceptibility to OA, knowledge of the factors that regulate this key growth factor could provide us with

potential methods for increasing *GDF5* expression for therapeutic benefit. Boosting the expression of *GDF5* above normal levels could be damaging; control of *GDF5* expression is vital for normal joint formation and has been highlighted by studies showing that mutations within *GDF5* that increase its activity or expression can result in fusions of the joint [242, 243]. In order to restore normal *GDF5* expression in T allele carriers, my approach is to identify the mediator of this DAE, providing us with a greater understanding of the regulation of *GDF5* and enabling the identification of novel therapeutic targets.

This chapter will therefore examine protein binding to the two rs143383 alleles using EMSAs and identify novel factors predicted to bind to the alleles using online database search tools. Chapter 6 will continue with the analysis of DEAF-1. The Sp1/Sp3 competition and antibody EMSAs were carried out with assistance from Dr. Louise Reynard.

4.2 Aims

- To examine *in vitro* protein complex binding to the C and T alleles using electrophorectic mobility shift assays (EMSAs).
- To identify *trans*-acting regulators of *GDF5* expression via the use of online bioinformatics databases.
- To interrogate binding of predicted factors through the use of EMSA competition assays and antibody supershifts.

4.3 Results

4.3.1 Extraction of Nuclear Protein

I used EMSAs to assess if there are any differences in protein binding to the C and T alleles of rs143383 in the SW872 cell line. Labelled EMSA probes and competitor sequences are shown in Table 1A, Appendix. Nuclear protein was extracted and the EMSA reaction performed as described in Materials and Methods sections 2.4 and 2.6 respectively.

Following the extraction of cytoplasmic and nuclear protein, the purity of each fraction was assessed following separation by gel electrophoresis and immunoblotting using two antibodies, anti lamin A/C and anti Glyceraldehyde 3-phosphate dehydrogenase (GAPDH). Lamins are proteins that are present in the nucleus of the cell and are important for cell structure and for transcription. Anti lamin A/C is able to identify both the A and C types of lamin. This antibody was used to assess the presence of nuclear protein in both the nuclear and cytoplasmic fractions. GAPDH is a key enzyme in glycolysis and is mainly present in the cytoplasm of the cell. The anti GAPDH antibody was therefore used to assess the presence of cytoplasmic proteins in both fractions. The presence of GAPDH and lamin A/C in the cytoplasmic and nuclear fractions following extraction is shown in Figure 4.1. There is a small level of GAPDH in the nuclear protein fraction (Nuc.), however there is a much larger level of GAPDH in the cytoplasmic fraction (Cyt.). Conversely, there is a large level of lamin A/C present in the nuclear fraction and a very small level present in the cytoplasmic fraction. I considered the purity of the nuclear fraction was sufficient for use.

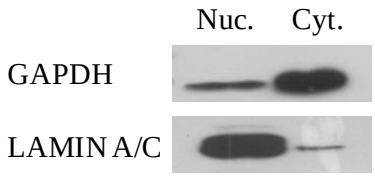


Figure 4.1

Optimisation of nuclear protein extraction. Cytoplasmic and nuclear proteins were isolated from SW872 cells. The purity of the two extracts (Nuclear – Nuc. and Cytoplasmic – Cyt.) was investigated following gel electrophoresis and western blotting using an anti-GAPDH antibody, an indicator of cytoplasmic protein content, and an anti-Lamin A/C antibody, an indicator of nuclear protein content.

4.3.2 Optimisation of the Electrophoretic Mobility Shift Assay (EMSA)

An optimisation EMSA was performed using 200fmol of C or T allele probe incubated with binding buffer, DTT and Poly (dI:dC) alone as shown in Figure 4.2A (lanes 1 and 2), or with the addition of 5µg of nuclear extract (lanes 3 and 4). Different components were also added to a reaction containing 5µg nuclear extract to assess optimal binding conditions, these included salmon sperm (lanes 5 and 6), glycerol (lanes 7 and 8), NP40 (lanes 9 and 10), KCl (lanes 11 and 12), $MgCl_2$ (lanes 13 and 14) and EDTA (lane 15). The five protein complexes binding to the two probes are highlighted by arrows. Most of the conditions show a similar pattern of binding, however, binding of complex two increases the presence of $MgCl_2$, whilst complex four appears to bind most abundantly in the presence of KCl. There does not appear to be any proteins binding solely to one allele, although there are intensity differences in some complexes between the two probes.

Binding to the C allele probe was then assessed for a second time using $MgCl_2$, KCl and EDTA and combinations of each of these (Figure 4.2B). Excess amounts of unlabelled C allele competitor were added to assess if the binding of these complexes was specific to the probe sequence. Complex binding to the probe was weak in the presence of both KCl and EDTA, whereas in the presence of $MgCl_2$ there was an abundance of protein binding. Furthermore binding of the two complexes indicated by the arrows was reduced upon addition of increasing amounts of unlabelled C allele competitor, confirming that the binding of these complexes is sequence specific. Of the different combination conditions tested, binding in the presence of $MgCl_2$ alone appeared to be optimal, thus subsequent EMSAs were performed in the presence of $MgCl_2$.

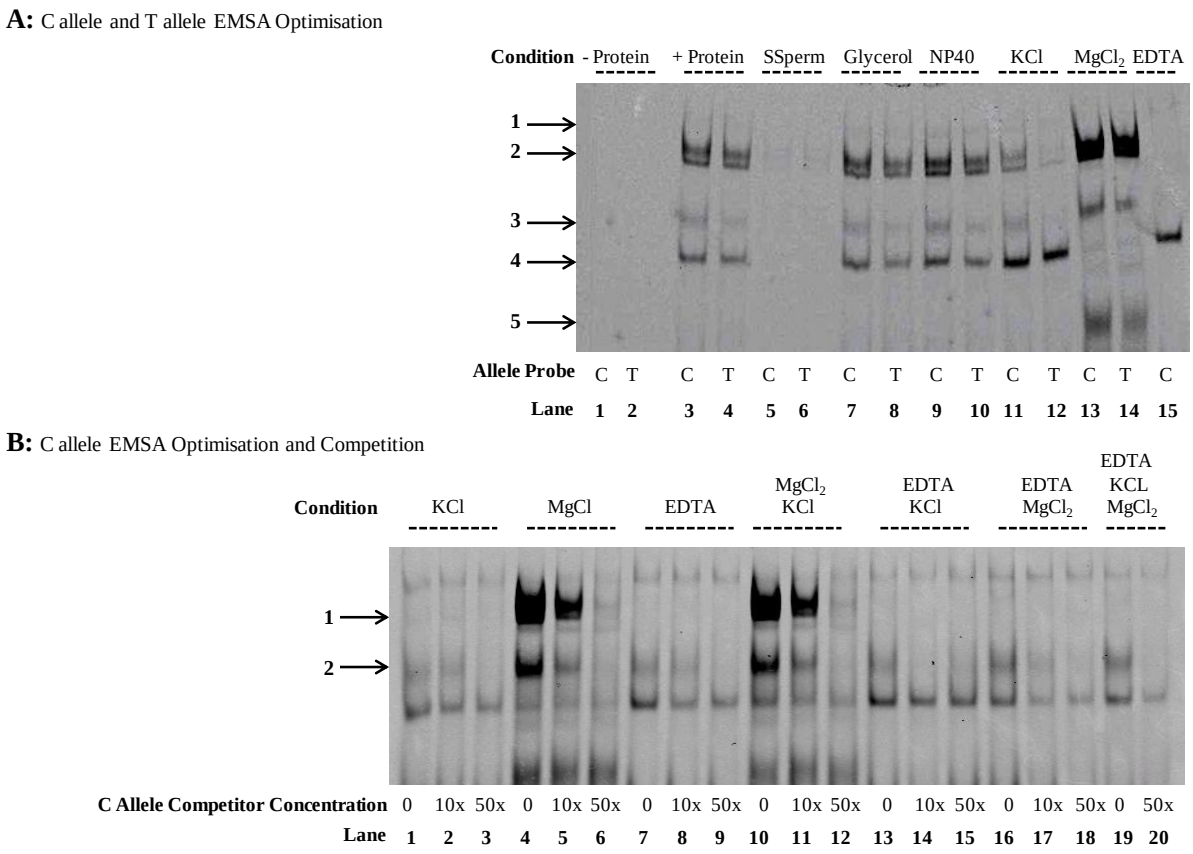


Figure 4.2

Optimisation of binding conditions of proteins to the C and T allele EMSA probes. **(A)** EMSA analysis optimising the conditions for the binding of proteins using C and T allele probes for rs143383 and nuclear extract from SW872 in lanes 3-16. Lanes 1 and 2 contain C and T allele probes but no nuclear extract. Lanes 5-16 contain salmon sperm DNA (SSperm), glycerol, NP40, potassium chloride (KCl), magnesium chloride ($MgCl_2$) and in lane 16, EDTA. The arrows indicate protein complexes binding to the probes. **(B)** EMSA analysis using the C allele probe and nuclear extract from SW872 cells. Increasing concentrations of unlabelled C allele competitor were added to the EMSA reaction in each condition of potassium chloride (KCl), magnesium chloride ($MgCl_2$), EDTA and combinations of $MgCl_2$ and KCl, EDTA and KCl, EDTA and $MgCl_2$, EDTA and KCl and $MgCl_2$. The arrows indicate specific binding complexes.

4.3.3 Differential Protein Binding to the rs143383 C and T alleles

Following optimisation of binding, unlabelled C and T allele competitors were used to identify any differences in binding affinity to the two probes (Figure 4.3). Both of the specific complexes binding to the C allele probe (marked by arrows in Figure 4.3) were outcompeted with excess unlabelled C and T allele competitor, and *vice versa* for the T allele probe. Higher concentrations of C allele unlabelled competitor were required to compete binding of the two complexes to the T allele probe and complex binding was competed from the C allele probe at a lower concentration of T allele competitor compared to C allele competitor. These results suggest the two protein complexes bind more avidly to the T allele, compared to the C allele.

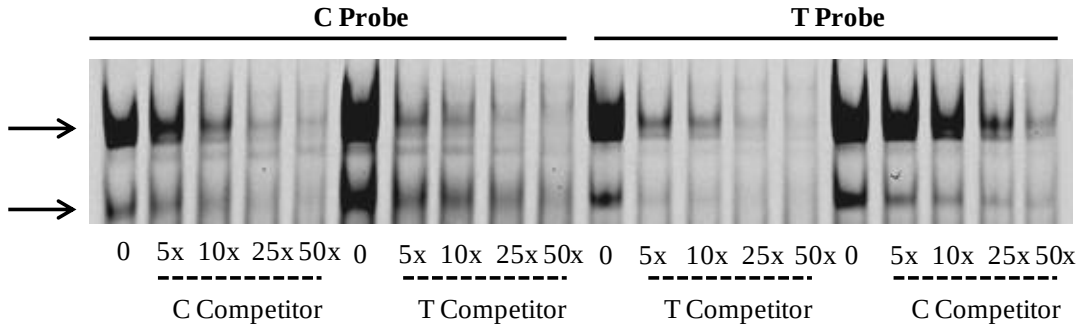
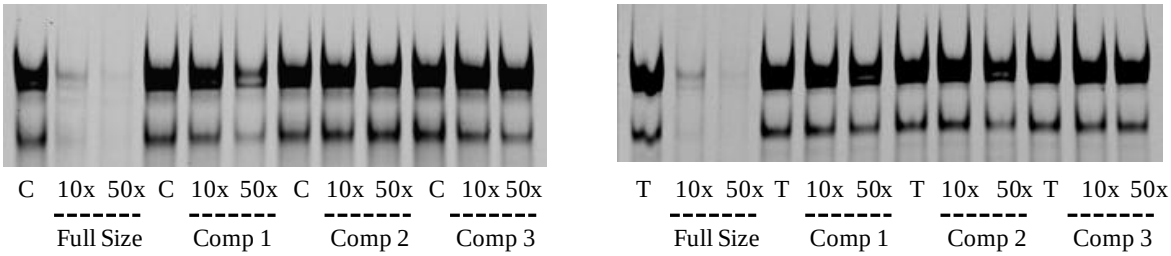


Figure 4.3

EMSA analysis using C and T allele probes and nuclear extract from SW872. Increasing concentrations of unlabelled C and T allele competitor were added to the EMSA reaction containing the C and the T allele probes, with the arrows indicating the specific complexes binding to the probes.

The region of binding of the two complexes to *GDF5* was then assessed by the addition of smaller competitor sequences. Figure 4.4 shows protein binding to the C and T allele full sized probes, and competition with three competitors encompassing different regions of the full sized probe. The competition assay suggests that protein binding is strongest to the full sized probe, suggesting this whole region is necessary for the binding of complexes 1 and 2. There is a small degree of competition from the C and T allele probes using the 50x concentration of competitor 1 and also from the T allele probe using competitor 2. Competitor 3 did not compete binding to either allele, suggesting that the region upstream of the polymorphism may be more important in comparison to the downstream region for complex 1 and 2 binding.



C Allele Competitors **T Allele Competitors**

Full Size - GAGAAAGGGGGCGGT**[C/T}**GGCTTTCTCC
Comp 1 - GAGAAAGGGGGCGGT**[C/T]**GG
Comp 2 - GGCGGT**[C/T]**GGCTTT
Comp 3 - GGT**[C/T]**GGCTTTCTCC

Figure 4.4
EMSA analysis investigating the region of complex binding to the *GDF5* C and T allele probes. The EMSA reaction contained the C or T allele labelled probes and nuclear extract from SW872. Increasing concentrations of the unlabelled competitors of varying sizes (full length competitor, and three shorter competitors) covering different regions were added to the EMSA reaction. The sequences of each of the competitors are shown below the EMSA, with the rs143383 polymorphism highlighted in bold and underlined.

4.3.4 The Identification of Sp1 and Sp3 binding to GDF5 by EMSA

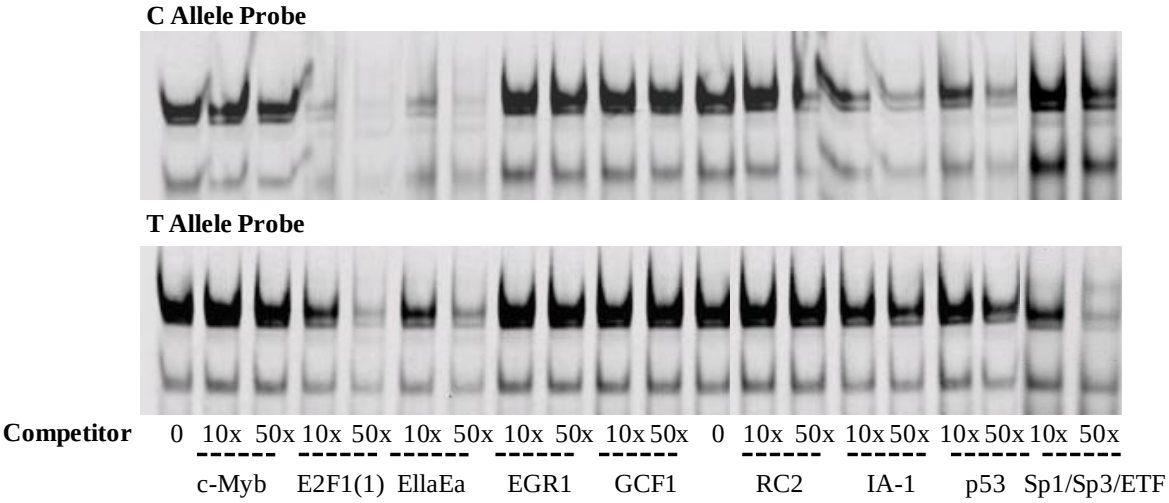
Following the discovery that two protein complexes bind to the *GDF5* probe and that their binding is modulated by the two alleles of rs143383, three online databases were used to identify proteins predicted to bind at this site as described in Materials and Methods section 2.6. Table 4.1 lists the transcription factors predicted to bind, the consensus sequence of the factors and the database used for identification. Increasing concentrations of consensus sequence competitors for each of these factors were added to the EMSA reaction and the results of this are shown in Figures 4.5A and 4.5B. Both of the complexes binding to the C and T allele probes are competed with increasing concentrations of E2F1, EllaEa, IA-1, p53, Sp1/Sp3/ETF (Sp1) and GABP competitors. Binding to the C allele probe is competed with the KLF16 competitor (Figure 4.5B).

Factor	Consensus Sequence (5′-3′)	Database
Sp1/Sp3/ETF	GGGGGCGGGG	Promo 3.0/Tess/TransFac
E2F1 (1) E2F1 (2) E2F1 (3) E2F2	AGGGCGG TCTTTCCCGCCTA TAGCCCGCGAAA ATTTTTCCCGCCT	Promo 3.0
EllaEa	GAGGGCG	Promo 3.0/Tess
EGR1	TGTGGGCGGGAGC	Promo 3.0
GCF1	AGCGCGGGCCG	Promo 3.0
RC2	GGTTTA	Promo 3.0
IA-1	TGTAAGGGGGCGA	Promo 3.0
p53	GGGCGGT	Promo 3.0
NF-1	GCCAA	Promo 3.0/Tess
CTF	GCGTTTGG	Promo 3.0
DRF1.1	GGCGGTGACT	Promo 3.0
KLF16	GGGGCGGTG	Promo 3.0
c-Myb	GGCGGTTG	Tess
GABP (1) GABP (2) GABP (3)	GGGGGGTT GGCACTTCCGGT TCGGGTGTT	Tess
CP2	ATTGG	Tess

Table 4.1

Consensus sequence of proteins predicted to bind to rs143383: The *trans*-acting factors predicted to bind to the sequence encompassing rs143383, their consensus sequence and the database used for identification.

A – Consensus Competitor Competition



B – Consensus Competitor Competition 2

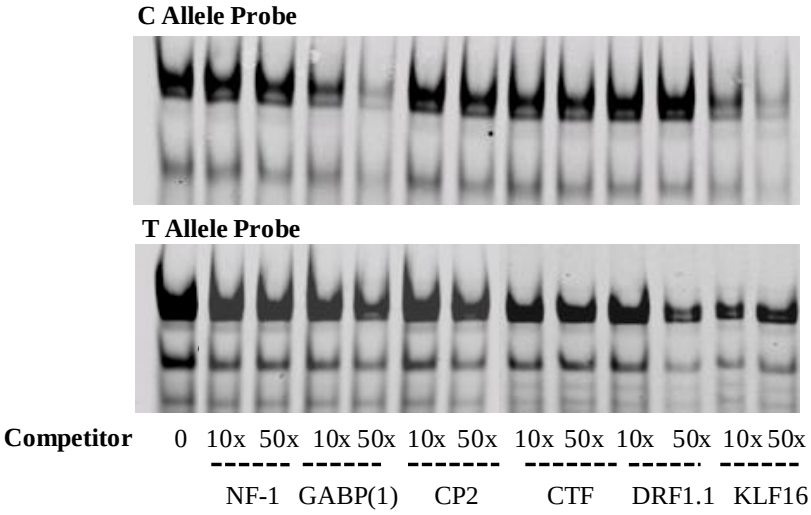


Figure 4.5

EMSA competition analysis. **(A)** and **(B)** The addition of increasing concentrations of unlabelled consensus competitors that were predicted to bind to *GDF5;* c-Myb, E2F1, EllaEa, EGR, GCF1, RC2, IA-1, Sp1/Sp3/ETF, NF-1, GABP, CP2, CTF, DRF1.1, KLF16.

Further assessment of Sp1/Sp3/ETF, E2F1, EllaEa, IA-1 and P53 was carried out, involving the addition of a greater range of concentrations of the consensus competitors. The Sp1/Sp3/ETF competitor outcompeted binding to both the C and T allele probes in a competitor concentration dependent manner (Figure 4.6A). Binding of the lower mobility complex to the T allele probe was outcompeted in a concentration dependant manner on the addition of increasing concentrations of E2F1 and EllaEa competitors, however binding of the higher mobility complex only slightly decreased at the 50x competitor concentration (Figure 4.6B). The IA-1 and p53 competitors competed protein complex binding to only a limited extent and did not do so in a competitor concentration dependent manner (Figure 4.6B). E2F1 is part of a family of E2F transcription factors and has three alternative consensus sequence binding sites. All three were tested whilst the binding of E2F2 was also assessed. Binding to the T allele probe was not affected by addition of the two of the three alternative E2F1 consensus competitors; the addition of increasing E2F2 consensus competitor decreased the binding of both complexes to the probe (Figure 4.6C). GABP also has three known consensus sequences; the two additional consensus competitors did not compete binding of either protein complex to the T allele probe (Figure 4.6C).

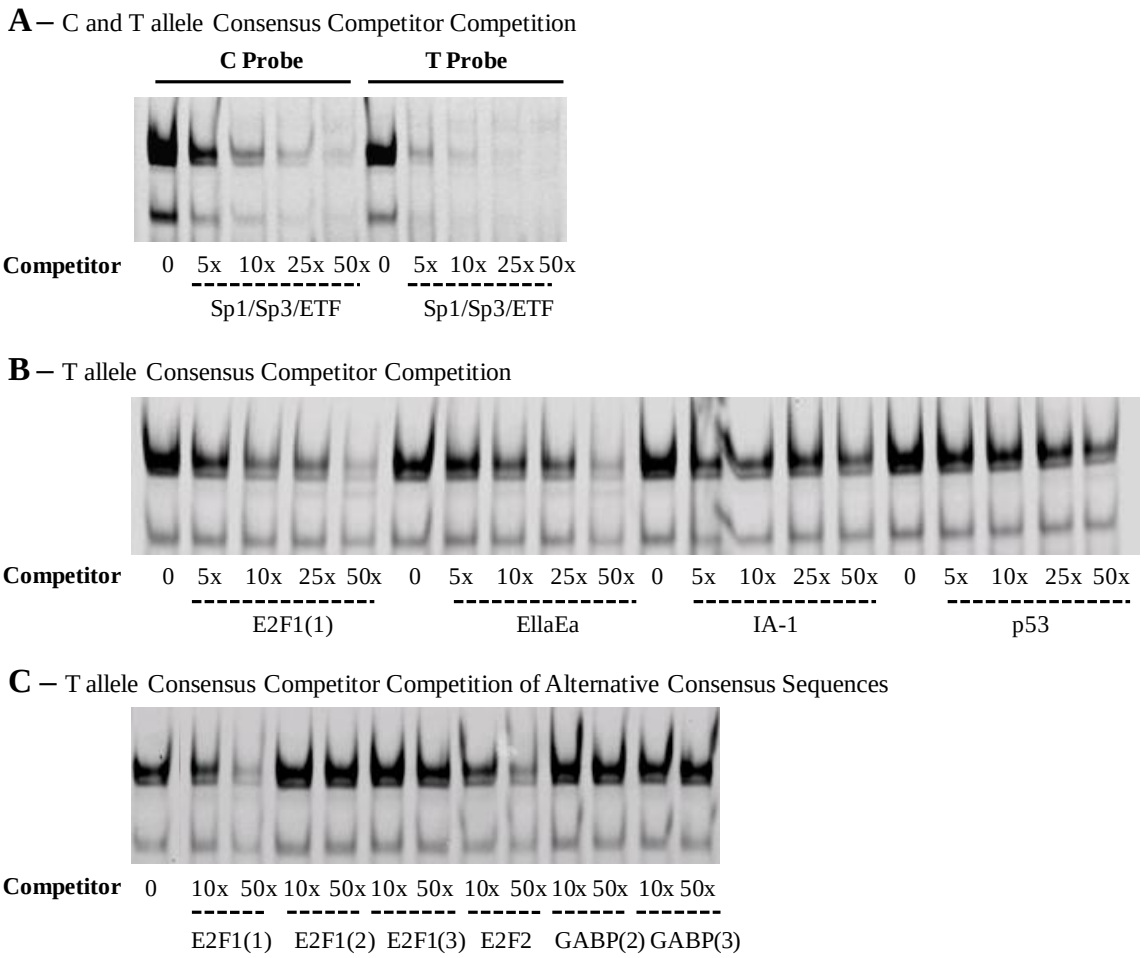


Figure 4.6

EMSA analysis using additional concentrations of competitors and alternative consensus sequences. **(A)** The addition of increasing concentrations of the Sp1/Sp3/ETF unlabelled consensus competitor to the C and T allele probes. **(B)** The addition of increasing concentrations of E2F1, EllaEa, IA-1 and p53 to the EMSA reaction containing the T allele probe. **(C)** The addition of increasing concentrations of three E2F1 consensus competitors, an E2F2 competitor and two alternative GABP competitors.

Addition of the Sp1/Sp3/ETF consensus competitor consistently competed binding to both the C and T allele *GDF5* probes in a concentration dependent manner, thus the binding of Sp1 and Sp3 was investigated further by the addition of antibodies targeting the transcription factors. The binding of E2F1 and E2F2 was further assessed using an antibody that recognises E2F1 and E2F2. I did not take the analysis of EllaEa any further due to the unavailability of an antibody targeting the protein. Additionally, the binding IA-1, p53 and GABP was not investigated further because the addition of increasing concentrations of consensus competitors did not compete binding to the *GDF5* probes in a concentration dependent manner. The binding of KLF16 was assessed further by antibody supershift and an antibody against EGR was used as an additional species matched control.

The antibody targeting Sp1 resulted in a supershift of two proteins from the lower mobility complex and addition of an antibody targeting Sp3 supershifted both the higher mobility complex, and a protein from within the lower mobility complex (arrows in Figure 4.7A). These results confirm the binding of Sp1 and Sp3 to the *GDF5* sequence *in vitro* in SW872 cells. Binding was also demonstrated by EMSA supershift in two additional human cell lines, the chondrosarcoma cell line SW1353, and the osteosarcoma cell line MG63, and also in human articular chondrocytes (Figure 4.7B and C). No supershifts were observed on the addition of antibodies targeting E2F1/E2F2, EGR and KLF16 (Figure 4.8).

Finally, the binding of the histone deacetylases HDAC1 and HDAC2 was investigated as these enzymes have previously been reported to form a repressive complex with Sp1 and Sp3. The addition of antibodies targeting the two enzymes did not however affect protein binding to either C or T allele probes (Figure 4.8).

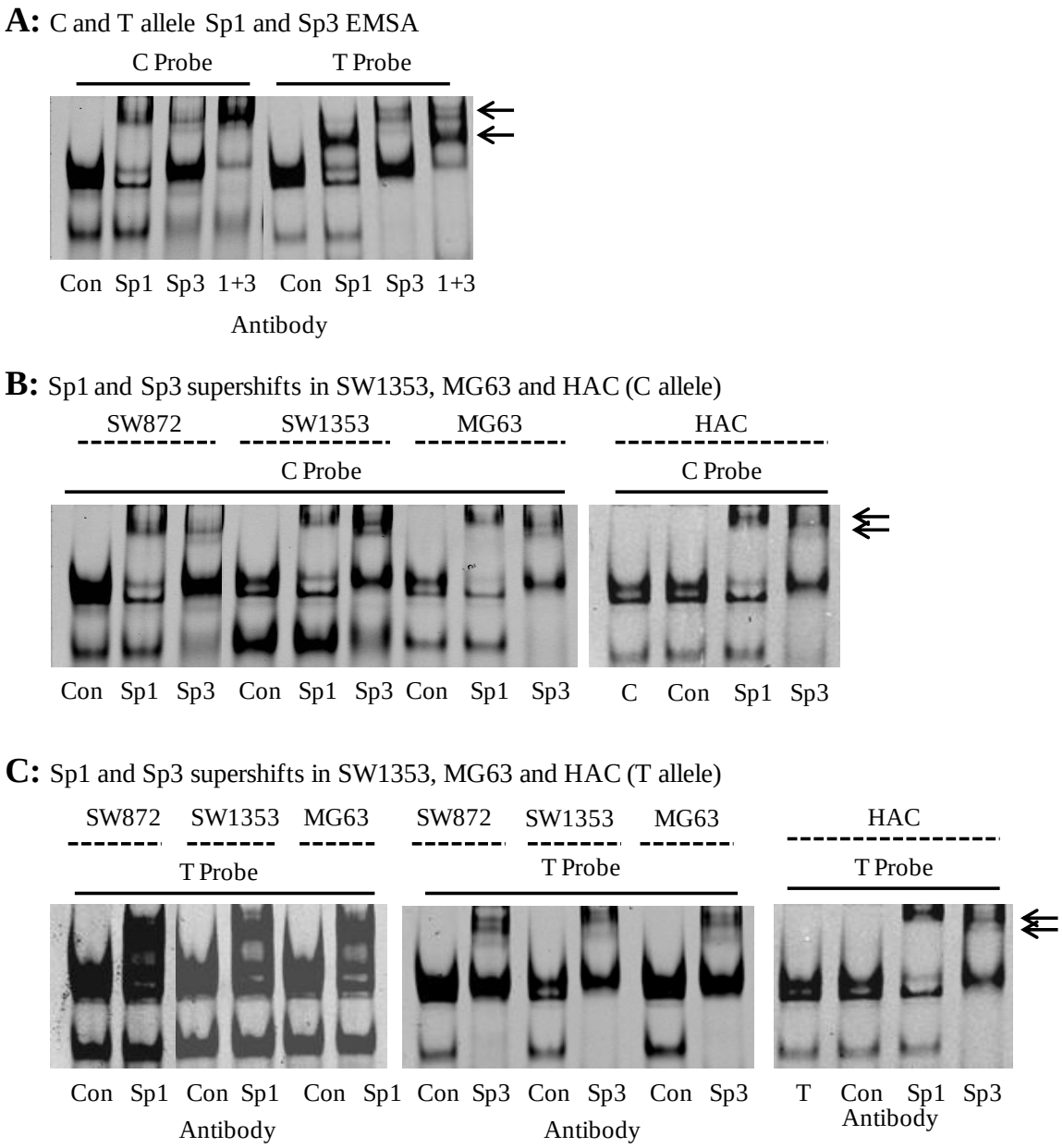


Figure 4.7

EMSA antibody analysis. **(A)** Supershift experiment demonstrating the effect of adding antibodies targeting Sp1, Sp3 or Sp1 and Sp3 together (1+3), compared to the IgG rabbit antibody control (Con) using the C or T allele probe. The arrows indicate the supershifted complexes. **(B)** Supershift experiment demonstrating the effect of adding antibodies targeting Sp1 and Sp3 to the EMSA reaction containing the C allele probe, compared to the IgG rabbit antibody control (Con). Nuclear extracts from SW872, SW1353 and MG63 cell lines and from human articular chondrocytes (HAC) were used. The arrows indicate supershifted complexes. **(C)** Supershift experiment demonstrating the effect of adding antibodies targeting Sp1 and Sp3 to the EMSA reaction containing the T allele probe, compared to the IgG rabbit antibody control (Con). Nuclear extracts from SW872, SW1353 and MG63 cell lines and from human articular chondrocytes (HAC) were used. The arrows indicate supershifted complexes.

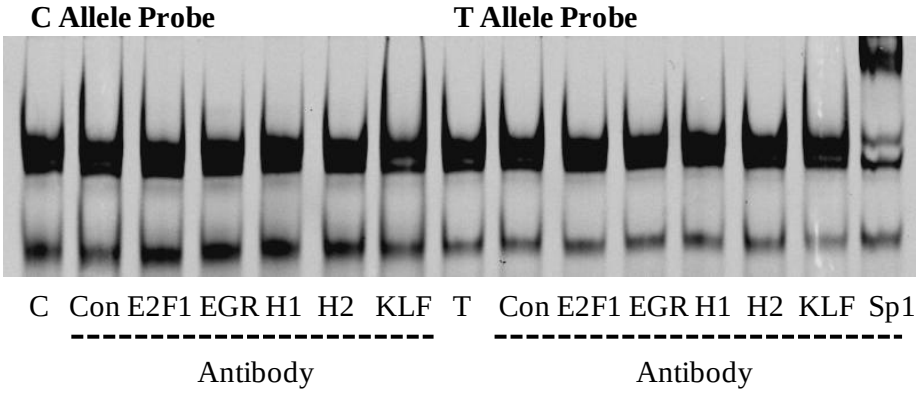


Figure 4.8

Supershift experiment: demonstration of the affect of adding antibodies targeting E2F1, EGR, HDAC1 (H1), HDAC2 (H2), KLF16 (KLF) using the C and T allele probes, compared to the IgG rabbit antibody control (Con) and C or T probe alone.

4.4 Discussion

Two complexes binding specifically to the C and T alleles of *GDF5* were identified by EMSA. These two complexes were found to bind differentially to the two alleles, with both complexes binding more avidly to the T allele. Considering a lower level of T allele transcript is observed, it is likely that the protein complexes binding more avidly to the T allele are repressing *GDF5* expression. A search of bioinformatics databases and the use of EMSA supershifts confirmed the binding of Sp1 and Sp3 to the C and T allele *GDF5* probes *in vitro* in the SW872 cells. I then confirmed the binding of Sp1 and Sp3 to *GDF5 in vitro* in SW1353 chondrosarcoma cells, in MG63 osteosarcoma cells and in human articular chondrocytes.

Sp1 and Sp3 are well characterised transcription factors and are members of a family of transcription factors, known as Sp1 like proteins or Krüppel-like factors (KLFs). Members of this family contain the same characteristic DNA binding domain, and bind to GC rich elements (GC boxes – GGGGCGGGG), in addition to GT boxes (GGGTGTGGC), within genes. Sp1 like proteins and KLFs are present in a number of species and so far 21 human Sp1/KLF like proteins have been identified. Sp1-Sp6 are highly related to Sp1 and form a subgroup of the family [301].

Sp1 was identified in the early 1980s and was one of the first transcription factors to be purified and characterised in mammalian cells [302, 303]. Sp1 was found to bind to the GC rich region in the simian virus 40 (SV40) promoter through its characteristic DNA binding domain [302]. Sp3 was initially identified as binding to the GT box in the promoter of the T-cell antigen receptor alpha gene [304]. Both Sp1 and Sp3 are ubiquitously expressed and regulate the expression of a wide variety of genes and are involved in a number of cellular processes [305]. Furthermore, there are estimated to be over 12,000 Sp1/Sp3 binding sites in the human genome [306].

Sp1 has been described as one of the most potent transcriptional activators, although it has been reported to repress gene expression when binding as part of a repressive complex [307, 308]. Many of the Sp proteins interact with co activators and co repressors via activation and repression domains. There are a variety of proteins reported to bind with Sp1 to modulate gene transcription, including activators such as p300 and CBP and other transcription factors such as YY1 and E2F [309, 310]. A co-factor for Sp1, CFSP is a large protein complex that has been purified and is required for Sp1 to mediate transcriptional activation [311]. Sp3 can act as an activator or a

repressor and its repression has been reported to be a result of competition with Sp1 for occupancy of the GC boxes. Sp3 was found to repress Sp1 mediated activation in a linear dose-dependent manner, and Sp3 with a mutated DNA binding domain did not affect Sp1 activation, confirming that it represses Sp1 activation through competition for binding to the DNA [312]. Sp3 contains an inhibitory domain and mutations within this domain change its repressive function to a strong *trans*-activator of promoters [313].

Sp1 and Sp3 have been reported to interact with proteins in the basal transcription machinery [314]. In addition, Sp1 with Sp3 can recruit the repressor complex HDAC1, HDAC2, RbAp48 to repress gene transcription [308]. Sp1 has also been reported to recruit chromatin remodelling proteins such as the SWI/SNF complex to modulate the expression of genes via altering chromatin structure [315]. Furthermore, a role for Sp1 in methylation of the promoters of genes has been proposed; Sp1 sites are critical for the maintenance of methylation free CpG islands in the Adenine phospho-ribosyltransferase (APRT) gene [316], Sp1 elements protect CpG islands from de novo methylation [316], and methylation of CpG sites prevents the binding of Sp1 [317].

When bound to gene promoters, Sp1 can form a tetramer, serving as a docking site for the binding of other proteins to synergistically affect the expression of genes. Sp3 cannot form homo-multimers and has been reported to have differential transcriptional effects depending on the number of Sp sites within promoters; Sp3 activates the expression of promoters with a single GC binding site, but responds weakly to or inhibits those with multiple binding sites [318].

Further evidence of the distinct roles of Sp1 and Sp3 is evident when comparing the knockout phenotypes in mice. Sp1 knockout mice display a number of abnormalities and are growth retarded confirming a key role for Sp1 during development, with embryo lethality at around day 11 of gestation. Many target genes of Sp1 however are unaffected in these mice and methylation free islands are maintained [319], furthermore Sp1 *null* embryonic stem cells have normal growth characteristics and survival rates. Considering the large number of genes that Sp1 is reported to regulate and its involvement in a number of cellular processes, it has been suggested that this phenotype may be due to the presence of Sp3 [319]. Sp3 deficient mice are growth retarded but develop until birth with no gross abnormalities. After birth, the mice die quickly with respiratory failure. Sp3 knockout mice also have a pronounced defect in late tooth

formation. Furthermore, Sp3 is required for proper skeletal ossification; both endochondral and intra-membranous ossification are impaired in Sp3 null embryos. This is due to the reduced expression of the osteoblastic specific marker gene osteocalcin [320] . It was concluded therefore that Sp1 and Sp3 have similar, and potentially redundant, functions during early development but exert distinct and highly specific functions in later developmental stages [321].

It is clear that both Sp1 and Sp3 have a wide variety of functions and are involved in many processes. Further analysis will aim to establish a role for these two ubiquitous transcription factors in the transcriptional control of *GDF5*. It appears from previous studies that Sp1 and Sp3 can form a repressive complex alongside other factors, and Sp1 can form homodimers enabling the binding of other proteins within the transcriptional complex. The next chapter will therefore focus on identifying additional novel factors binding to *GDF5* using a different methodology.

Chapter 5: The Identification by Mass Spectrometry of P15 as a *GDF5 Trans*-acting Factor

5.1 Introduction

In addition to the identification of *trans*-acting factors using online prediction software tools and EMSAs, I wanted to find an alternative method to identify differential protein binding to the *GDF5* gene. An oligonucleotide pull down assay using two regions of DNA, one for each rs143383 allele and magnetic beads to enable isolation was explored. I identified a method in the literature, whereby the research group had combined an oligonucleotide pull down assay with tandem mass spectrometry to enable quantitative assessment of protein binding to DNA. Their aim was to identify binding of factors to *mtk*, a gene important for immunity in Drosophila. They compared the binding of proteins to a wild-type region of DNA with binding to a mutated region of DNA and identified a transcription factor that regulates the expression of this gene [322].

Quantification of proteins within different samples can be carried out simultaneously using tandem mass tags (TMTs). TMTs are a set of identical tags that contain a reactive group (which binds to peptides on the amino terminus), a reporter ion (which following tandem mass spectrometry provides information on the abundance of the peptide) and a mass normalisation group (which balances mass differences between tags to ensure the same overall mass) [323]. Using an oligonucleotide pull down assay, in addition to TMT labelling and tandem mass spectrometry, I was therefore able to identify proteins binding to each rs143383 allele, and quantitatively assess if there were any differences in binding affinity for the two alleles.

Achim Treumann and Karen Lowden assisted in the design, data requisition and analysis of the TMT mass spectrometry experiment.

5.2 Aims

- To isolate proteins binding to the *GDF5* gene using an oligonucleotide pull down assay.
- To quantitatively assess differences in protein binding between the rs143383 C and T alleles using tandem mass tag labelling and tandem mass spectrometry.
- To validate proteins identified using RNAi.

5.3 Results

5.3.1 Optimisation of the PCR

A 212bp region of patient DNA (CC or TT genotype at rs143383) was amplified using a 5´ biotinylated primer and unmodified 3´primer. The PCR was first optimised by testing seven annealing temperatures ranging from 50-64.7°C and $MgCl_2$ at a final concentration of 1mM and 2mM. The 212bp DNA region is amplified optimally at 2mM $MgCl_2$ and a temperature of 54.2°C was chosen for the annealing (Figure 5.2). Following optimisation of the PCR, the concentration of DNA required for the pull down assay was determined. The DNA was quantified and the CC and TT DNA samples diluted to equal concentrations. 4µg of each DNA sample was used for a serial dilution and compared by agarose gel electrophoresis to verify that equal proportions of C and T allele DNA were present. The binding capacity of 1mg of magnetic beads is 20pmol DNA, thus an excess concentration of 30pmol of biotinylated template was added to the reaction in order to ensure the saturation of the streptavidin coated magnetic beads and minimise non specific protein binding to the beads.

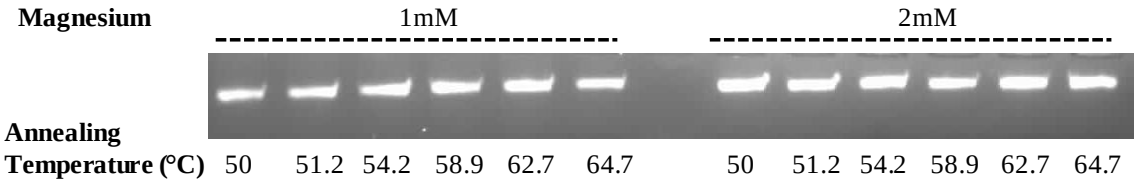


Figure 5.1

Optimisation of PCR reactions. These were performed using either 1mM or 2mM $MgCl_2$ and annealing temperatures ranging from 50-64.7°C. Following the PCR, the PCR products were analysed by agarose gel electrophoresis. The 212bp PCR product is shown.

5.3.2 Optimisation of the Binding Conditions

The purity of nuclear extract using the nuclear extraction protocol described in Materials and Methods section 2.4 has been previously discussed in Chapter 4.3.1. Following the extraction of nuclear protein, I assessed if the salt concentration of the nuclear lysate impacted upon protein binding to the *GDF5* DNA region. Following extraction, the nuclear protein buffer contained a high level of salt, I dialysed half of this sample into a low salt buffer as described in Materials and Methods section 2.7 and compared protein binding in the two buffers by completing concurrent pull downs using

the two extracts. Samples that contained no DNA were used as a control. The forward and reverse PCR oligonucleotides contain an *EcoR1* restriction enzyme site enabling the isolation of protein/DNA complexes from the magnetic beads following an *EcoR1* digestion. Protein samples and the isolated bead samples were separately electrophoresed and detected by coomasssie staining (Figure 5.2). Lanes 2-5 show binding of proteins in low salt conditions. Lanes 2 and 3 do not contain DNA and no protein is detected in these samples. Lanes 4 and 5 contain proteins that bound to DNA, lane 4 showing protein present bound to the beads, and lane 5 showing proteins that were separated from the beads following the digestion. Lanes 6-9 show binding of proteins in high salt conditions. Lanes 8 and 9 clearly have fewer bound proteins than lanes 4 and 5, demonstrating that there are a greater number of proteins binding under the low salt binding conditions. In subsequent pull down assays therefore, the proteins were dialysed into a low salt buffer prior to incubation with DNA to maximise protein binding. Additionally, there appears to be a greater number of proteins present in the bead samples (lanes 4 and 8) in comparison to isolated proteins following the *EcoR1* digest (lanes 5 and 9). This could be because a greater number of proteins are binding to the beads non specifically compared to the number binding to the DNA or because the *EcoR1* digest is not efficiently releasing the DNA/protein complexes from the beads. The latter appears more likely considering the similarity in the pattern of binding between the two lanes. In subsequent pull down assays therefore, I isolated proteins from the beads by heating to 95°C and using a magnetic separator. Although this new approach resulted in proteins that bind non specifically to the beads being present within the sample for analysis, the use of a beads only (no DNA) control sample enabled me to identify non specific binding proteins.

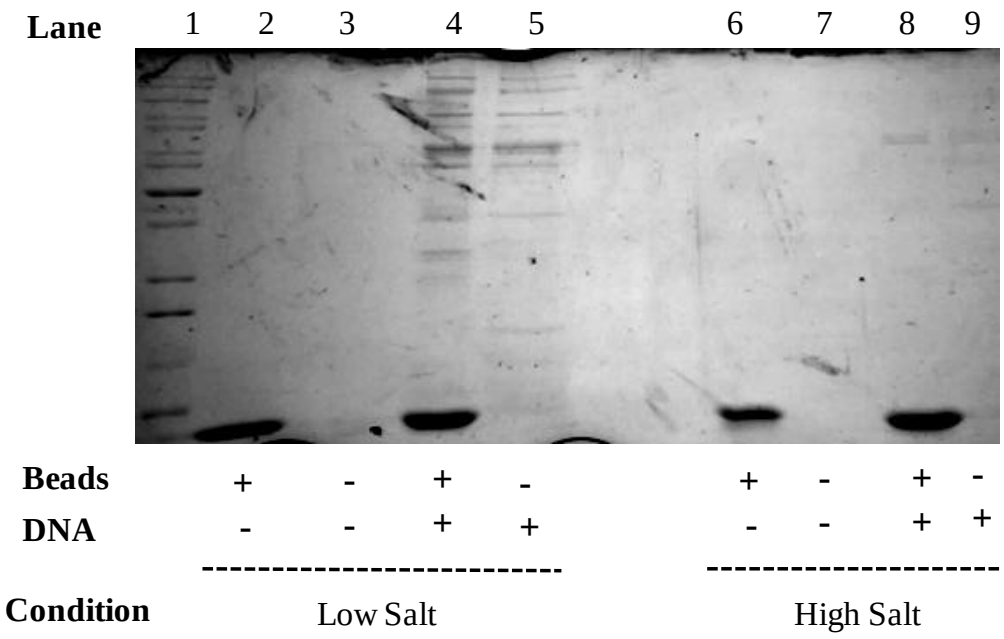


Figure 5.2

Optimisation of the binding of proteins to a region of *GDF5* DNA. Lane 1 contains a protein sizing ladder. Lanes 2-5 contain proteins that bound to the DNA under low salt conditions. Lanes 2 and 3 are control samples with no DNA. Lane 4 contains proteins present in the sample containing beads and lane 5 contains proteins isolated that were bound to DNA. Lanes 6-9 contain proteins bound under high salt conditions, lanes 6 and 7 have no DNA, lane 8 contains proteins present in the sample containing beads and lane 9 contains proteins bound to DNA.

5.3.3 Gels for Quantification

The three repeat oligonucleotide pull down assay gels that were taken forward for analysis by quantitative tandem mass spectrometry are shown in Figure 5.3. The pattern of protein binding to the two alleles in the three experiments is similar.

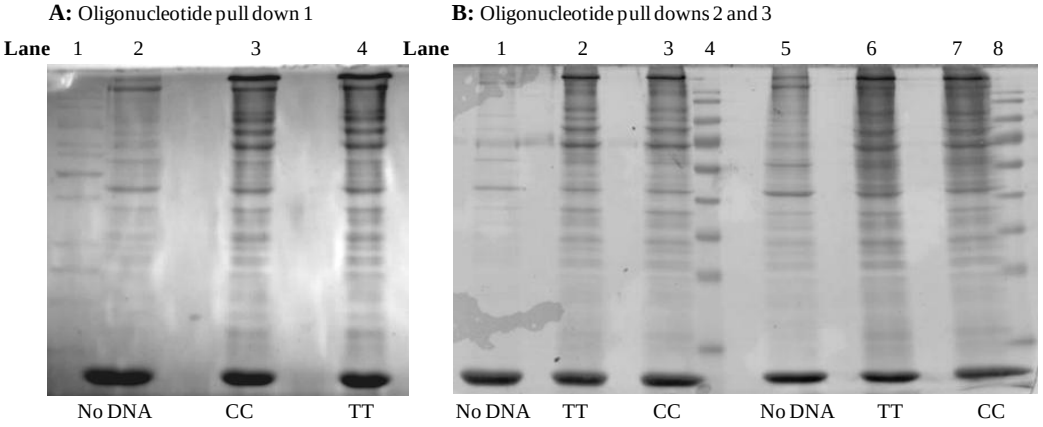


Figure 5.3

The three oligonucleotide repeat experiments taken forward for analysis by tandem mass spectrometry. **(A)** Oligonucleotide pull down repeat 1: Lane 2 has no DNA, lanes 3 and 4 contain proteins binding to the C and T alleles respectively. Lane 1 contains a protein sizing ladder **(B)** Oligonucleotide pull downs 2 and 3. Lane 1 and lane 5 have no DNA. Lanes 2 and 6 and lanes 3 and 7 contain proteins binding to the T and C alleles respectively. Lanes 4 and 8 contain a protein sizing ladder.

5.3.4 Mass Spectrometry data analysis

The data was analysed by DAVID pathway analysis functional annotation tool available at http://david.abcc.ncifcrf.gov/. The functional classifications of the proteins identified in each mass spectrometry experiment were listed according to a calculated P-value, which represents the enrichment of the proteins identified in the functional categories. For example, common functions of proteins that showed high enrichment common to all three experiments was acetylation, nuclear proteins, ribonucleoproteins, phosphoproteins, RNA binding proteins and mRNA processing proteins. Proteins with a known role in transcription, including activators, repressors, coactivators, proteins associated with RNA polymerase II and proteins involved in transcriptional regulation were also identified.

In order to compare the data obtained from the three repeat experiments, I first removed proteins from the analysis that were identified in the control (no DNA) sample in each of the three experiments. I then sorted the results according to their relative abundance in the C allele DNA sample compared with the T allele DNA sample (C/T ratio). The

results obtained in each of the three experiments are shown in Figure 5.4. In experiment one, in total, 352 proteins were identified, with C/T ratios ranging from 8 to 0.126. In experiment two, 88 proteins were identified, with C/T ratios ranging from 6.2 to 0.384. In experiment three, 127 proteins were identified with C/T ratios ranging from 8.9 to 0.164. The number of proteins identified in experiments two and three is similar, whilst far more proteins were identified in experiment one. This may be a result of the time periods between experiments; experiments two and three were performed within weeks of one another and analysed concurrently by tandem mass spectrometry, whilst the first experiment was performed and subject to tandem mass spectrometry a few months prior to this which could result in experimental variability. Aside from protein number, the spread of C/T values is similar in all three experiments.

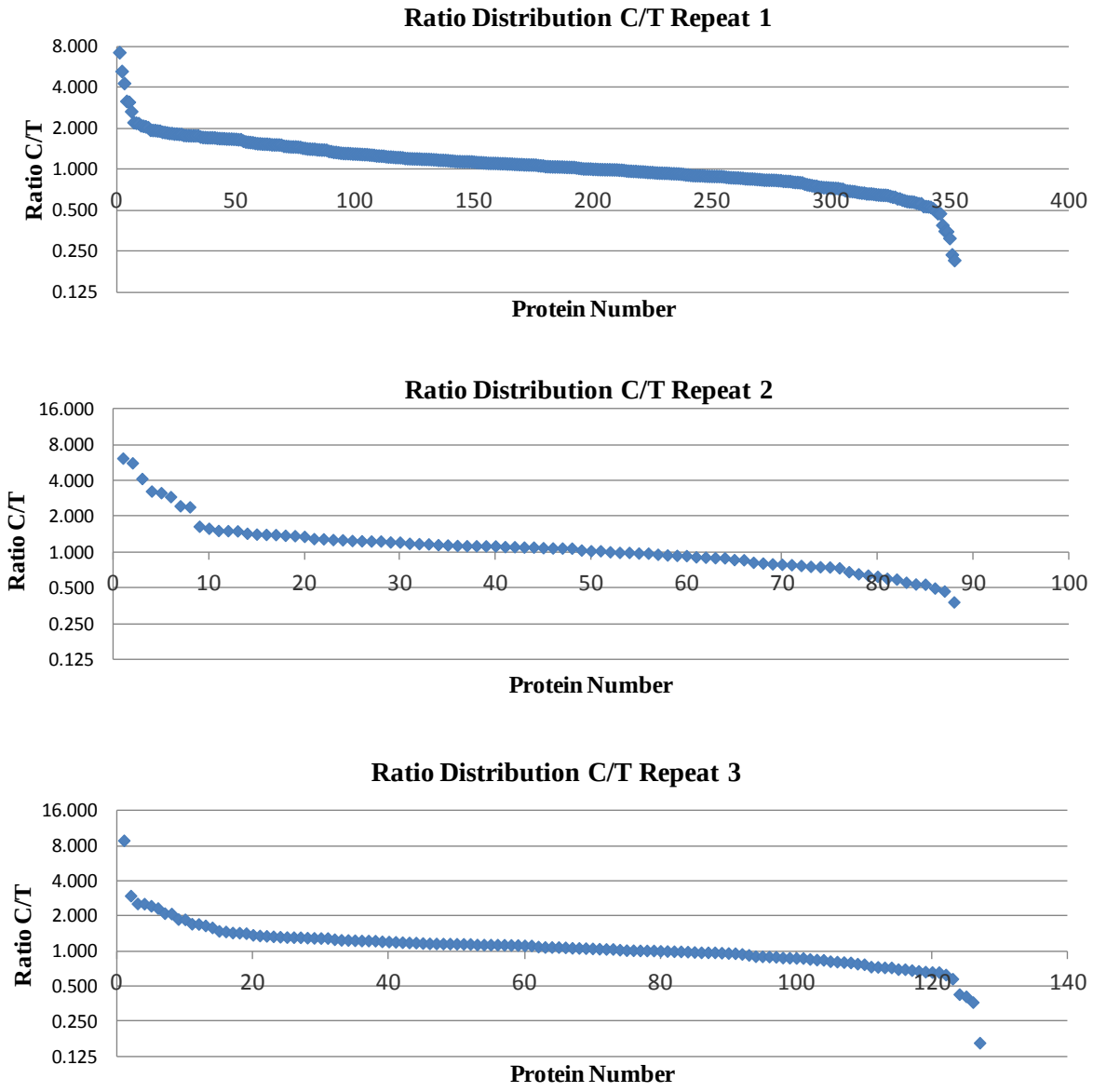


Figure 5.4

Comparison of the three repeat mass spectrometry experiments. Proteins were sorted according to their abundance in the C allele DNA sample in comparison with the T allele DNA sample (Ratio C/T). The proteins were then numbered, with protein 1 having the largest C/T ratio, and the highest number protein having the smallest C/T ratio.

Each experiment was first analysed independently and for each I identified proteins with a known role in transcription. Table 5.1 shows the transcription associated proteins that were identified in all three or in two of the three repeat experiments. The next stage of analysis involved sorting results based on C/T ratio to obtain a comparison of the abundance of proteins present within each sample. The results of each repeat were colour coded and combined, then sorted alphabetically. Any proteins that were present in the background sample were highlighted bold. Proteins present in all three experiments and absent in the background samples were ranked highest. Proteins present in all three experiments, and present in the background in one of the experiments were ranked second highest, continuing downwards in a tiered analysis, shown in Document 1, Appendix. The unmodified mass spectrometry data can be found on the excel file on the supplementary disc attached to this thesis (Document 2, Appendix).

Protein Identified	Identified by Repeat Number
RNA binding motif protein 14	All 3
Activated RNA polymerase II transcriptional coactivator P15	All 3
Poly (ADP-ribose) polymerase 1	All 3
Similar to RNA binding motif protein 39	All 3
Thyroid hormone receptor associated protein 3	All 3
Ubiquitin-like with PHD and ring finger domains 1	All 3
YLP motif containing 1	2 and 3
Non-POU domain containing, octamer binding	2 and 3
Replication factor C	2 and 3
Splicing factor proline/glutamine-rich	1 and 2
Hairy and enhancer of split 7	1 and 3
Upstream binding transcription factor	1 and 3
RuvB-like 1	1 and 2
DEAH (Asp-Glu-Ala-His) box polypeptide 9	1 and 2

Table 5.1
The transcription related proteins that were identified in two or more of the quantitative mass spectrometry experiments.

Following this analysis, there were three factors that were present in the top three groups that had a role in transcription and all these proteins were present in higher abundance in the T allele DNA sample compared to the C allele DNA sample. These factors were Poly(ADP-Ribose) polymerase I (PARP1), POLR2H and Activated RNA polymerase II transcriptional coactivator P15 (P15). After studying the published research into each of these factors, I prioritised P15 for further analysis, due to its dual function as a transcriptional activator and repressor. PARP1 on the other hand is predominantly known for its involvement in DNA repair and POLR2H is a component of the general transcription machinery.

5.3.5 Further analysis of P15

In order to verify if P15 is involved in *GDF5* transcription and if its activity is affected by genotype at rs143383, I first depleted the expression of *P15* by using siRNA. The optimal conditions for target protein knockdown using siRNA within this cell line were determined during my Master of Research project and the conditions used are described in Materials and Methods section 2.9 and the siRNA details are given in Table 6, Appendix. The depletion of *P15* mRNA was confirmed by real time RT-PCR (Figure 5.5A). The overall expression of *GDF5* was increased, with a significant fold change ($p<0.01$) observed upon *P15* knockdown (Figure 5.5B). I then used allele specific real time PCR to assess if P15 differentially affects the two alleles of rs143383, and as such could contribute to the DAE mediated by this SNP. Depletion of *P15* significantly ($p<0.05$) attenuated the DAE from a C/T allelic ratio of 1.5 in the control (NTsiRNA) to 1.26 (P15 siRNA) (Figure 5.5C).

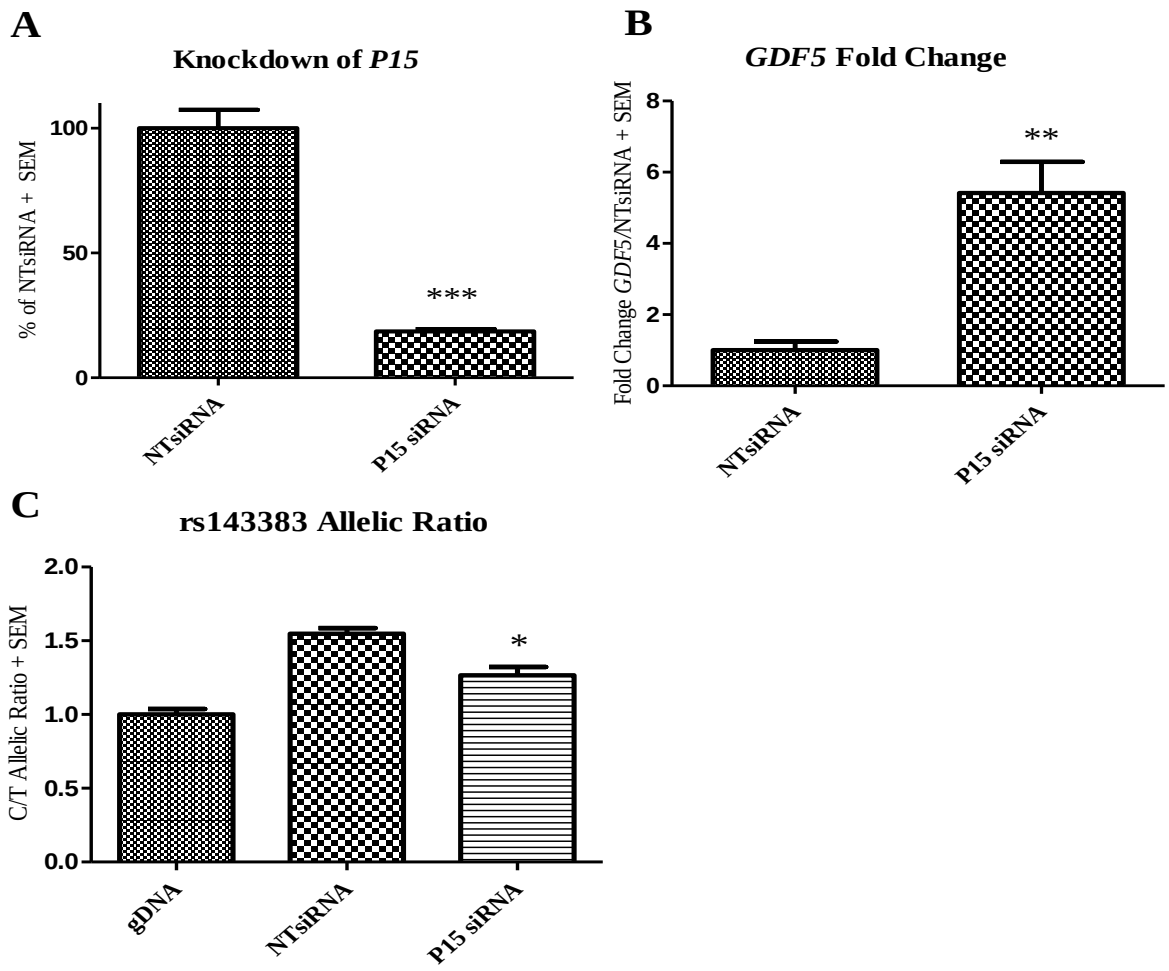


Figure 5.5

P15 depletion by RNAi **(A)** Expression of *P15* mRNA, shown as a percentage of the control non-targeting siRNA (NTsiRNA) treated cells following *P15* siRNA knockdown. Error bars denote the standard error of the mean (SEM). *** $p<0.001$, calculated relative to the NTsiRNA value using a Student's 2 tailed *t*-test. **(B)** Fold change in *GDF5* expression following *P15* siRNA knockdown and shown relative to the NTsiRNA control. Error bars denote SEM. ** $P<0.01$, calculated using a Student's 2 tailed *t*-test. **(C)** The rs143383 C/T allelic ratio is shown following P15 siRNA knockdown and compared against treatment with the NTsiRNA control. Allelic ratios were normalised to genomic DNA (gDNA). Error bars denote SEM. *$p<0.05$, calculated using a Student's 2 tailed *t*-test. Three replicates were performed.

5.3.6 EMSA analysis of P15

P15 does not have a known binding consensus sequence and I was therefore not able to use an EMSA to investigate competition for binding to the fluorescently labelled rs143383 C and T allele probes. On the addition of an antibody targeting P15 to the EMSA reaction, I observed a decrease in the two specific protein complexes binding to the two probes (Figure 5.6A). This effect was replicated using nuclear extracts from the chondrosarcoma cell line SW1353 and the osteosarcoma cell line MG63, and using nuclear extract from human articular chondrocytes (HACs; Figure 5.6B).

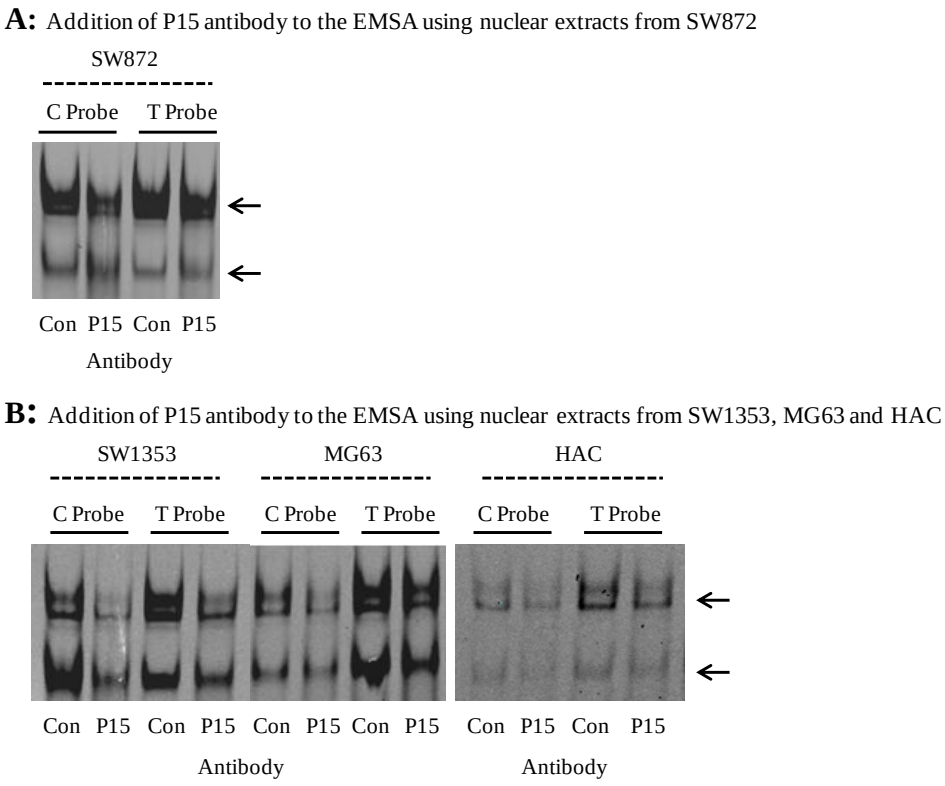


Figure 5.6

Demonstration of the effect of adding P15 antibody to the EMSA reaction compared to the IgG rabbit antibody control (Con). **(A)** The EMSA reaction contains C and T allele probes and SW872 nuclear extract. The arrows indicate the affected complexes. **(B)** The EMSA reaction contains C and T allele probes and extract from SW1353 and MG63 cells in addition to extract from human articular chondrocytes (HAC). The arrows indicate the affected complexes.

5.4 Discussion

I was able to isolate protein binding to the two alleles of rs143383 using an oligonucleotide pull down assay and identify and quantify the abundance of each of these proteins in the C and T allele DNA samples following tandem mass tag labelling and tandem mass spectrometry. After completing three repeat experiments, the data was analysed and candidates were prioritised based on their absence in the control (no DNA) sample, and their known functional involvement in transcription. P15 was identified as a transcription factor that was more abundant in the T allele DNA sample in comparison with the C allele DNA sample.

There are differences in the number of proteins identified as binding to *GDF5* in the three repeat experiments, and slight variability in the spread of C/T ratios observed. Unavoidable variability amongst the experiments may account for these differences, in particular the difference in number of proteins identified in the first repeat compared to the final two repeats may be a result of the time period between these experiments and potential variability in the tandem mass spectrometry process. This variability was not considered important since only proteins that were identified in all three or in two of the three experiments, and which showed consistent values in each repeat experiment, were considered further.

Depletion of *P15* mRNA increased the expression of *GDF5* and significantly attenuated the DAE. I was unable to investigate the binding of P15 to the rs143383 C and T allele probes using competition EMSAs. However, on the addition of an antibody targeting P15 to the EMSA reaction I observed a decrease in the binding of the two specific binding complexes, suggesting that P15 may be present within the complex binding to the two *GDF5* probes. Observing a decrease in the intensity of complexes binding to the EMSA probes on the addition of an antibody, rather than a supershifted complex, implies that the antibody is binding to the region of the protein that binds to the DNA, thereby leading to a reduction in the amount of protein-DNA complex observed [324]. This effect was also observed using nuclear extract from two further cell lines, SW1353 and MG63 and when using nuclear protein extracted from human articular chondrocytes.

P15 is also known as positive cofactor 4 (PC4). Positive cofactors (PC1-PC6) and negative cofactors (NC1, NC2) are a class of general cofactors that positively or negatively regulate transcription by mediating interactions between the basal

transcription machinery and gene specific transcription factors. These cofactors were identified following the fractionation of HeLa cell nuclear extracts to identify proteins stimulating activator dependent transcription [325]. Furthermore, they have a dual role in transcription; in the presence of activators, the positive cofactors increase promoter activity, in the absence of activators however, these cofactors are capable of repressing basal transcription [325]. They are distinct from other general transcription factors in that they are dispensable for basal-level transcription and distinct from activators as most do not bind to DNA or DNA binding is not sequence dependent.

P15 is a 15kDa protein and like other positive cofactors, when activators are absent, P15 can function as a repressor protein, repressing basal transcription by stabilizing a minimally transcriptionally active PIC. In the presence of activators however, P15 can interact with components of the basal transcriptional machinery and stimulate transcription; a large (up to 90 fold) induction of transcription was mediated by P15 in the presence of activators. Compared to the other positive cofactors, P15 mediates the greatest effect on transcriptional activation and the gene that encodes P15 *(SUB1)* is highly conserved between different species [326].

P15 is abundantly expressed and performs a diverse range of functions. Upon *P15* knockdown, 128 genes were up regulated and 49 down regulated. Furthermore, P15 has been reported to interact with a number of proteins containing different activation domains, including the acidic activation domain of VP16, the proline-rich activation domain of CTF, and the glutamine-rich activation domain of Sp1 [326]. P15 can bind to components of the basal transcription machinery including TFIIA, TFIIH and RNA polymerase II and serves as a bridging molecule between the PIC and coactivators. Furthermore P15 can enhance DNA binding of some activators, or enhance PIC formation or RNA polymerase II elongation. Recombinant P15 protein has been shown to enhance AP-2, NF-κB and Sp1 mediated gene activation [327, 328].

P15 has also been reported to associate with chromatin and can efficiently pull down all the core histones and preferentially binds to histones H3 and H2B. In addition, P15 induces chromatin condensation, suggesting a role in gene repression via chromatin remodelling. Knocking down expression of *P15* was found to alter chromatin organisation, leading to alteration in the higher ordered chromatin structure [329].

P15 knockout mice show embryonic lethality at the pre-implantation stage, indicating the vital role of P15 in mouse embryonic development. P15 embryos were smaller and

disorganised at E6.5 when gastrulation starts, indicating that the major cause of lethality was due to impaired cellular proliferation. Heterozygous P15 mice however did not exhibit developmental problems, suggesting that reduced amounts of P15 may be sufficient for normal development [330].

P15 appears to be a valid *trans*-acting factor candidate that may be binding differentially to the alleles of rs143383 and which may modulate the DAE at this locus. The role of P15, in addition to Sp1 and Sp3 is further assessed in Chapter 7.

Aside from P15, many other factors were identified that are binding differentially to the two rs143383 alleles. PARP1 in particular may be worthy of further investigation, but was not taken further in the scope of this thesis. PARP1 catalyses poly(ADP-ribosyl)ation, a process that involves the transfer of ADP ribose from nicotinamide adenine dinucleotide (NAD+ donor protein) to an acceptor protein [331]. Although primarily known for its role in DNA repair, PARP1 is also involved in nucleic acid metabolism and in the maintenance of chromatin structure, where it modifies target proteins by poly(ADP-ribosyl)ation. In addition to these functions, PARP1 also modifies transcription components such as PC3, histones, components of the basal transcriptional machinery such as TBP, RNA polymerase II, p53, YY1, E2F and Sp1 [332, 333]. Poly(ADP-ribosyl)lation of DNA binding proteins such as P53, YY1 and Sp1 prevents their binding to DNA. Furthermore, PARP1 can also mediate repressive effects on transcription through incorportation into chromatin, promoting the formation of highly condensed chromatin, through incorporation into a corepressor complex or by inhibiting ligand dependent transcription [334-336]. PARP1 can thereby act as a switch, either mediating repression or converting transcriptionally silent genes into an active state by dissociating from a co repressor complex and recruiting chromatin modifying enzymes and activators. Although it is unlikely considering the known functions of PARP1 that it is binding differentially to the two rs143383 alleles, it could be interacting in some way with factors that are binding to the DNA and potentially mediating a repressive effect on *GDF5* transcription.

Chapter 6: Investigating DEAF-1 binding to *GDF5* using five DEAF-1 antibodies

6.1 Introduction

Chapter 4 examined protein binding to the two rs143383 alleles using EMSAs and identified the binding of Sp1 and Sp3 to the alleles. This Chapter will continue with the analysis of DEAF-1.

Following the findings by Egli et al. that the *GDF5* rs143383 differential allelic expression is apparent in all tissues of the joint, this group sought to further characterise the function of rs143383 by identifying *trans*-acting factors that are binding to this locus [272]. As described in Chapter 4 (4.1) they first assessed binding to the two alleles by EMSA and found that the C and T allele probes had differing mobilities, suggesting differences in protein binding. The T allele was found to bind a unique protein that was unresponsive to increasing concentrations of C allele competitor.

Using three online prediction sites Egli et al. identified several transcription factors that were predicted to bind to rs143383 in an allele specific manner and three were chosen for further analysis; the Myb protooncogene protein (c-Myb), early growth response protein 1 (EGR1) and deformed epidermal auto-regulatory factor 1 (DEAF-1). On the addition of consensus competitor sequences against each of these factors, only the DEAF-1 competitor competed binding to the *GDF5* probe. Specifically, complex binding to the T allele probe was competed [272]. The role of DEAF-1 was therefore assessed further by Egli et al. using luciferase experiments. A DEAF-1 expression vector was co-transfected into cells in addition to *GDF5* C and T allele promoter luciferase constructs. DEAF-1 repressed the expression of the C and T allele *GDF5* luciferase constructs, but in accordance with the EMSA competition experiments, DEAF-1 repressed the T allele more significantly than the C allele [272]. This study highlighted DEAF-1 as a repressive factor that appears to be differentially modulating the expression of *GDF5*.

DEAF-1 was first discovered and identified as a nonhomeodomain protein in Drosophila Melanogaster that interacts, in addition to other factors, with the deformed protein at the *HOM-C* gene cluster [337]. Precise expression and interaction of transcription factors is required for the regulation of 'selector' or *HOM-C/Hox* genes that determine the morphological characteristics of an organism. In drosophila, such genes are encoded within the *HOM-C* gene cluster and govern the anterior to posterior

patterning of developing fly embryos [338]. In vertebrates there are four *Hox* gene clusters that show homology to the *HOM-C* cluster in drosophila. Mutations within these genes result in substitution of one body part for another. All of the proteins encoded within this gene cluster share a region of homology in their DNA binding domain known as the homeodomain, and have similar binding specificities [339, 340]. Deformed is a homeodomain protein that is expressed in drosophila and is required for the segmental expression patterns established during development [341]. Specificity in binding to target genes is achieved through binding with co-factors, such as DEAF-1, that enhance the binding of homeodomain proteins [342].

Mutations within DEAF-1 that improved binding *in vitro* increased the expression of deformed target genes. DEAF-1 was shown to bind to the human *HOXD4* promoter, a mammalian homolog gene of *deformed*, suggesting that a mammalian homolog of DEAF-1 may exist and may modulate the expression of *Hox* genes [343].

A mammalian homolog of DEAF-1 was then discovered by its affinity for the retinoic acid response element of target genes. Retinoic acid receptors (RARs) are nuclear transcription factors that bind, in addition to the retinoid-X-receptors (RXRs), to the retinoic acid response elements within target genes and mediate developmental effects such as cellular differentiation and embryogenesis [344]. DEAF-1 was identified as a transcription factor that interacts with RAR/RXR dimers and activates the transcription of the human *proenkephalin* promoter, a gene that is expressed in many reproductive and neuroendocrine tissues. DEAF-1 was found to be expressed in many of these rapidly dividing tissues and is expressed at high levels in the foetus, suggesting an important role for this protein during development [345].

This protein has been extensively characterised, the gene is found on human chromosome 11, and a number of different molecular weight proteins have been found in several tissues such as testis (72kDa), muscle (45kDa), brain (60kDa) and heart (47KDa). All tissues show the presence of *DEAF-1* mRNA, DEAF-1 protein is however thought to be unstable in most tissues and is mainly found to be expressed in rapidly dividing cells [345].

DEAF-1 knockout mice display skeletal abnormalities including rib cage defects, with a large proportion of the animals suffering from defective neural tube closure that causes death shortly after birth [346].

During my Master of Research project, I used siRNA to deplete the expression of *DEAF-1* and assessed the impact of *DEAF-1* depletion on the overall expression of *GDF5*. Upon *DEAF-1* knockdown, I observed a significant increase in the overall expression of *GDF5*. These experiments further highlighted the importance of DEAF-1 in modulating the expression of *GDF5*. The core consensus sequence for DEAF-1 is TCGG, the *GDF5* sequence surrounding rs143383 (shown here in bold) is GGT**C**GGCT, and therefore contains the DEAF-1 consensus sequence (underscored). The T allele changes the sequence to GGT**T**GGCT and therefore no longer contains the DEAF-1 core consensus sequence. This single base change may result in the differential binding of this *trans*-acting factor, and thus could account for differential *GDF5* expression.

In this chapter I sought to confirm the binding of DEAF-1 to *GDF5*, firstly *in vitro* using EMSAs and secondly *in vivo* using chromatin immunoprecipitation (ChIP). Furthermore, I have tested five different DEAF-1 antibodies and characterised each to assess their binding specificity.

6.2 Aims

- To assess the binding of DEAF-1 to *GDF5 in vitro* using EMSAs.
- To assess the binding of DEAF-1 to *GDF5 in vivo* using ChIP.
- To identify a specific antibody that targets DEAF-1.

6.3 Results

6.3.1 Investigating DEAF-1 binding by EMSA

DEAF-1 was not identified as a protein binding to the rs143383 sequence by the database search that I performed in Chapter 4 (using the databases Promo 3.0, TESS and TransFac) however Egli et al. had identified DEAF-1 using three databases, two of which are distinct from those that I used (TESS, TFsearch and MatInspector) [272]. The core consensus sequence for DEAF-1 is TCGG, which encompasses rs143383. Following the identification of the complexes binding to the two rs143383 alleles in Chapter 4 by EMSA (Figure 4.6, Chapter 4), I used competition EMSAs to assess if DEAF-1 was binding to *GDF5 in vitro* using SW872 cell nuclear extract, see Materials and Methods section 2.6. Figure 6.1A shows the addition of three different consensus DEAF-1 competitors (the first identified by Promo 3.0 online software tool, the second by Huggenvik et al. [345] and the third by Egli et al.[272]) to the EMSA reaction containing the rs143383 T allele probe. The two specific complexes that were found to

bind to the C and T allele probes in Chapter 4 are shown and were not competed on the addition of increasing concentrations of any of the DEAF-1 consensus competitors (Figure 6.1A).

There is an abundance of Sp1 and Sp3 protein binding to the *GDF5* probe (identified previously in Chapter 4), thus I hypothesised that the binding of DEAF-1 may be masked by these two proteins. I therefore used DEAF-1 recombinant protein instead of the SW872 nuclear extract to further investigate DEAF-1 binding to the *GDF5* probes. Unlabelled competitors for the C and T alleles and for DEAF-1 were added to the reaction; however, no binding to the probe was observed (Figure 6.1B).

Poly (dI:dC) has been used previously in the EMSA reaction to reduce non specific protein binding [347]. In order to verify that the addition of Poly (dI:dC) was not preventing the binding of DEAF-1 to the probe, I investigated if removing it from the EMSA reaction would enable DEAF-1 binding to the probe to be detected. Furthermore, I tested two different concentrations of DEAF-1 recombinant protein and added the Sp1 antibody to supershift the abundant Sp1 complex to verify that the binding of DEAF-1 was not masked by Sp1. There was however no binding to the probe observed in any of these modifications (Figure 6.1C).

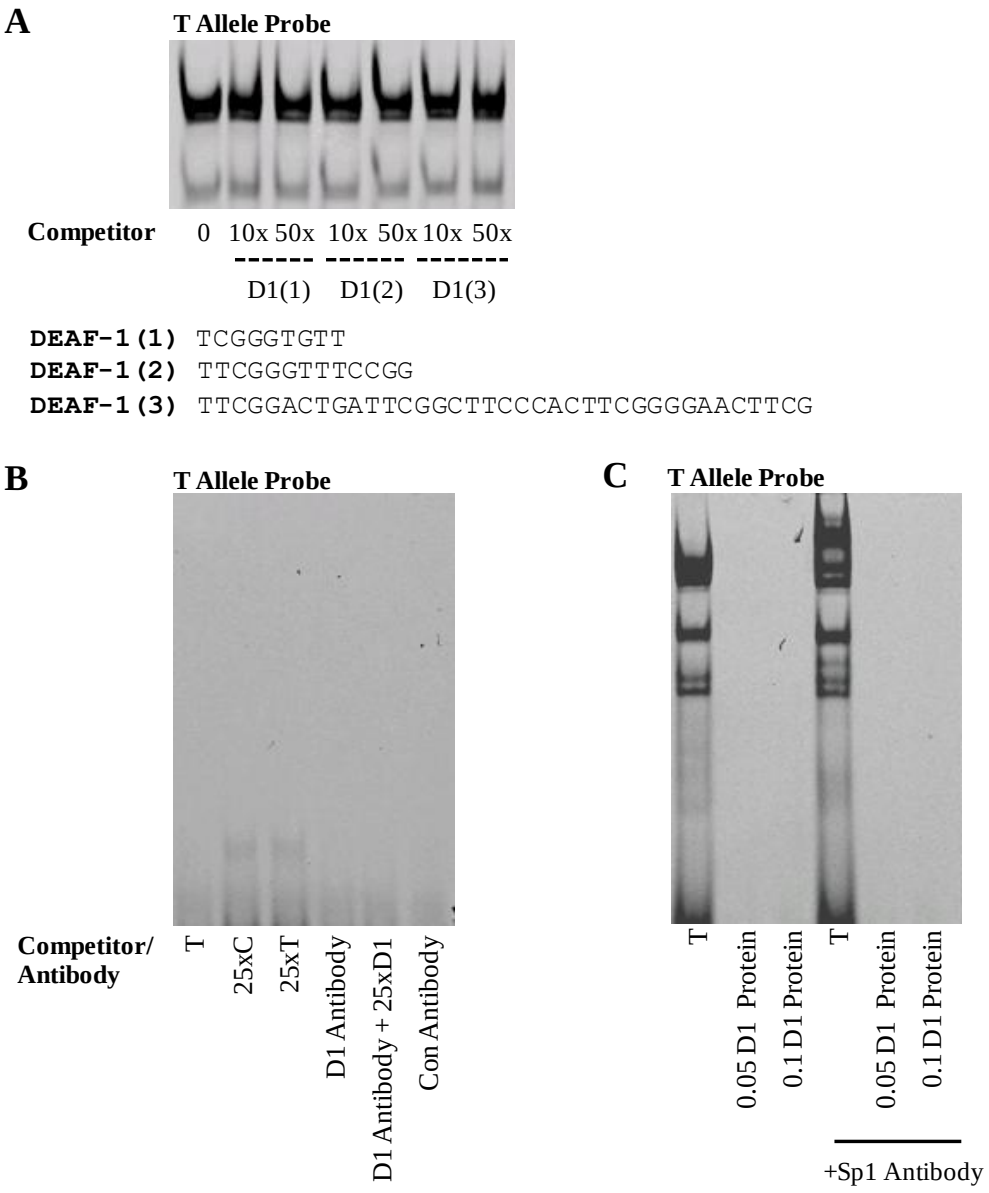


Figure 6.1

DEAF-1 competition EMSA analysis. **(A)** Increasing concentrations of three different unlabelled DEAF-1 competitors were added to the EMSA reaction containing the rs143383 T allele probe and SW872 nuclear extract. The sequences of each of these competitors are shown below. **(B)** EMSA analysis using the T allele probe and DEAF-1 recombinant protein. 25x C and T allele unlabelled competitors were added to the reaction. An antibody targeting DEAF-1 was added to the reaction (D1) both alone and in addition to 25x DEAF-1 competitor 1 and compared with the control IgG antibody (Con). **(C)** EMSA analysis using the T allele probe and both SW872 nuclear extract (T) and two different concentrations of DEAF-1 recombinant protein (0.05 and 0.1µg, containing no Poly (dI:dC)). The Sp1 antibody was also added to the reaction (+Sp1).

To further assess the binding of DEAF-1 to the *GDF5* EMSA probes, an antibody was added to the EMSA reaction (DEAF-1 antibody 4, Figure 6.2A, antibodies listed in Table 2, Appendix). A supershifted complex was observed on the addition of the DEAF-1 antibody. In addition to an antibody targeting DEAF-1, an antibody targeting Sp1 was added to the reaction to try and identify where the supershifted complex had shifted from. However, there was no apparent reduction in the intensities of any the complexes binding to the probe (Figure 6.2A). The supershift was confirmed using nuclear extract from human articular chondrocytes, with protein complexes binding to the C and T allele probes being less intense than those observed in the SW872 cells (Figure 6.2B). Further concentrations of the DEAF-1 antibody were added to the EMSA reaction containing SW872 nuclear extract; however, there was no increase in the intensity of the supershifted complex with increasing antibody concentration. A supershifted complex was also observed in the EMSA reaction containing the C allele on addition of the DEAF-1 antibody. The intensity of this supershifted complex was lower compared with the T allele supershifted complex (Figure 6.2C). An alternative DEAF-1 antibody (DEAF-1 antibody five) was then added to the EMSA reaction. This antibody did not affect the protein complexes binding to the *GDF5* T allele probe. Antibodies targeting Sp1 and Sp3 were used as positive control antibodies and were also added in addition to the DEAF-1 antibody (DEAF-1 antibody five), however complex binding to the T allele probe was unaffected (Figure 6.2D).

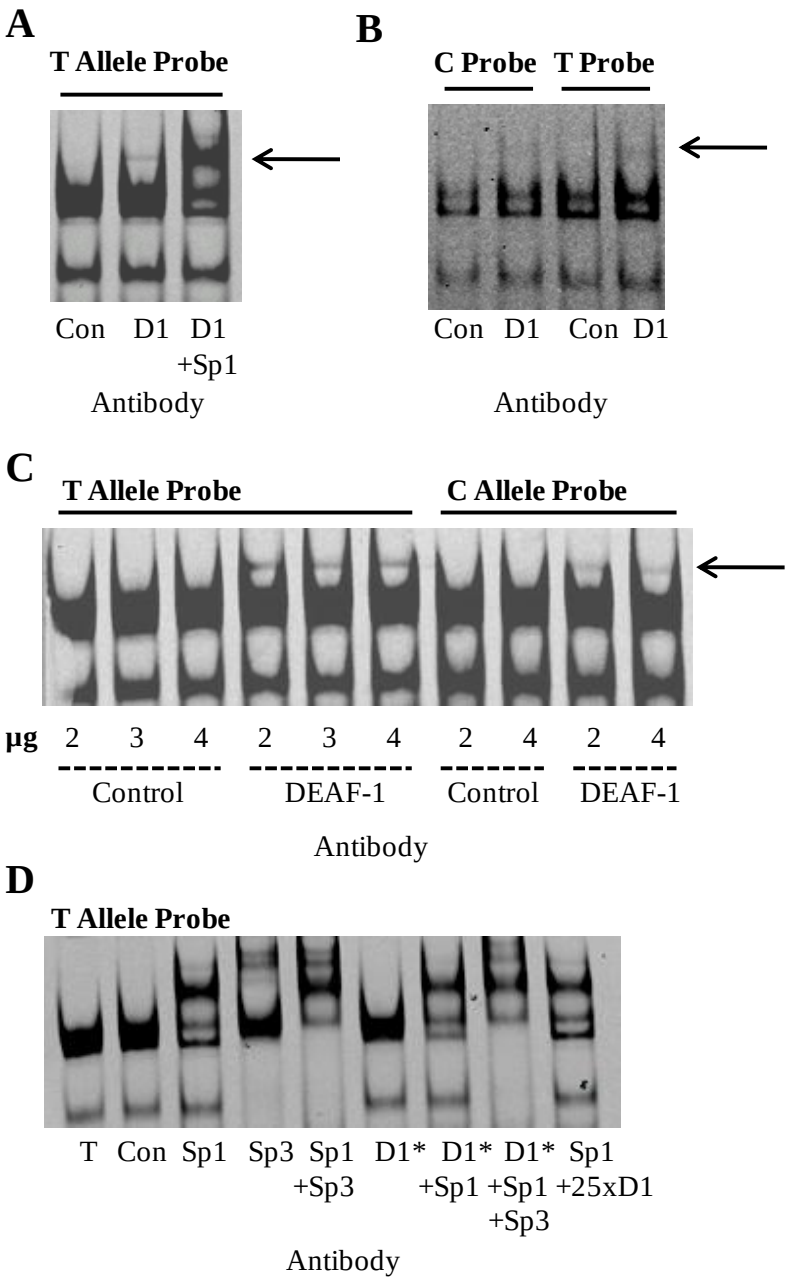


Figure 6.2

DEAF-1 antibody EMSA analysis. **(A)** Supershift experiment demonstrating the effect of adding an antibody targeting DEAF-1 alone or in addition to an antibody targeting Sp1 to the EMSA reaction containing SW872 nuclear extract. This EMSA reaction was performed using the T allele probe and was compared with the control IgG antibody (Con). The DEAF-1 supershifted complex is highlighted by an arrow. **(B)** Supershift experiment demonstrating the effect of adding an antibody targeting DEAF-1 to the EMSA reaction containing nuclear extract from human articular chondrocytes. This EMSA reaction was performed using the C and T allele probes and was compared with the control IgG antibody (Con). The DEAF-1 supershifted complex is highlighted by an arrow. **(C)** Supershift experiment demonstrating the effect of adding increasing concentrations of DEAF-1 antibody to an EMSA reaction containing the T or C allele probes. The effect was compared with the addition of increasing concentrations of the control IgG antibody (Con). The supershifted complex is highlighted by an arrow. **(C)**

Supershift experiment demonstrating the effect of adding antibodies to the EMSA reaction targeting Sp1 and Sp3 alone and in combination (Sp1 + Sp3), a different DEAF-1 antibody to that used previously (D1*), both alone and in combination with Sp1 and Sp1 in addition to Sp3. 25x DEAF-1 competitor 1 was added in addition to an antibody targeting Sp1 (Sp1 + 25xD1).

To summarise, protein binding to the *GDF5* C and T allele probes was not competed on the addition of three different DEAF-1 consensus competitor sequences. Furthermore, no binding to the probes was observed on the addition of DEAF-1 recombinant protein. However a supershifted complex was observed on the addition of one of the DEAF-1 antibodies to the EMSA reaction. I have therefore not been able to conclude from these EMSA experiments if DEAF-1 is binding to *GDF5 in vitro.*

6.3.2 Investigation of DEAF-1 binding by ChIP

6.3.2.1 Optimisation of DNA fragmentation by sonication

I next sought to assess the binding of DEAF-1 *in vivo to GDF5* by ChIP in SW872 cells, Materials and Methods section 2.8. In order to accurately identify the region in which a protein of interest binds to a gene, it is necessary to first sonicate genomic DNA into smaller pieces. Fragments of length 200-1000bp of DNA are recommended for ChIP (based on manufacturers instructions of the ChIP kit used, manufactured by Millipore). Figure 6.3A shows the DNA fragments produced following sonication using a probe sonicator (at amplitude five) for three different time periods. DNA was purified by phenol-chloroform extraction and electrophoresed through an agarose gel. A decrease in fragment size can be seen with increasing sonication time.

Sonication of SW872 cell DNA was then compared with sonication of DNA from JURKAT cells, an immortalized T-lymphocyte cell line; fragmentation of JURKAT DNA by sonication using a waterbath has been successfully performed previously in our laboratory and was therefore used for comparison of sonication conditions. I compared two different methods of sonication, the above probe sonicator and two sonicating waterbaths (Figure 6.3B). The probe sonicator (at amplitude five) produced SW872 DNA fragments within the desired size range at all of the time points studied. Both waterbaths fragmented SW872 cell DNA to the desired size range after 4x5 minute intervals (20 minutes total). JURKAT DNA however was not fragmented during the time periods studied using either waterbath (Figure 6.3B).

Further test sonications on SW872 DNA were carried out using the probe sonicator, assessing different time periods and amplitudes. Sonication for a total of 15 minutes

(30x10 second intervals) at amplitude five gave optimal lengths of DNA fragments, whilst sonicating at amplitude 10 for 15 minutes produced smaller fragments, of 100-400bp (Figure 6.3C).

From these optimisations, I concluded that a shorter sonication time of five minutes (10x10 second sonication intervals, with 20 seconds rest periods between sonications) would be optimal. A shorter time period also ensured that the DNA was kept at a low temperature to avoid degradation. Amplitude eight was chosen to ensure that there was sufficient fragmentation. Figure 6.3D shows DNA following sonication under these chosen conditions.

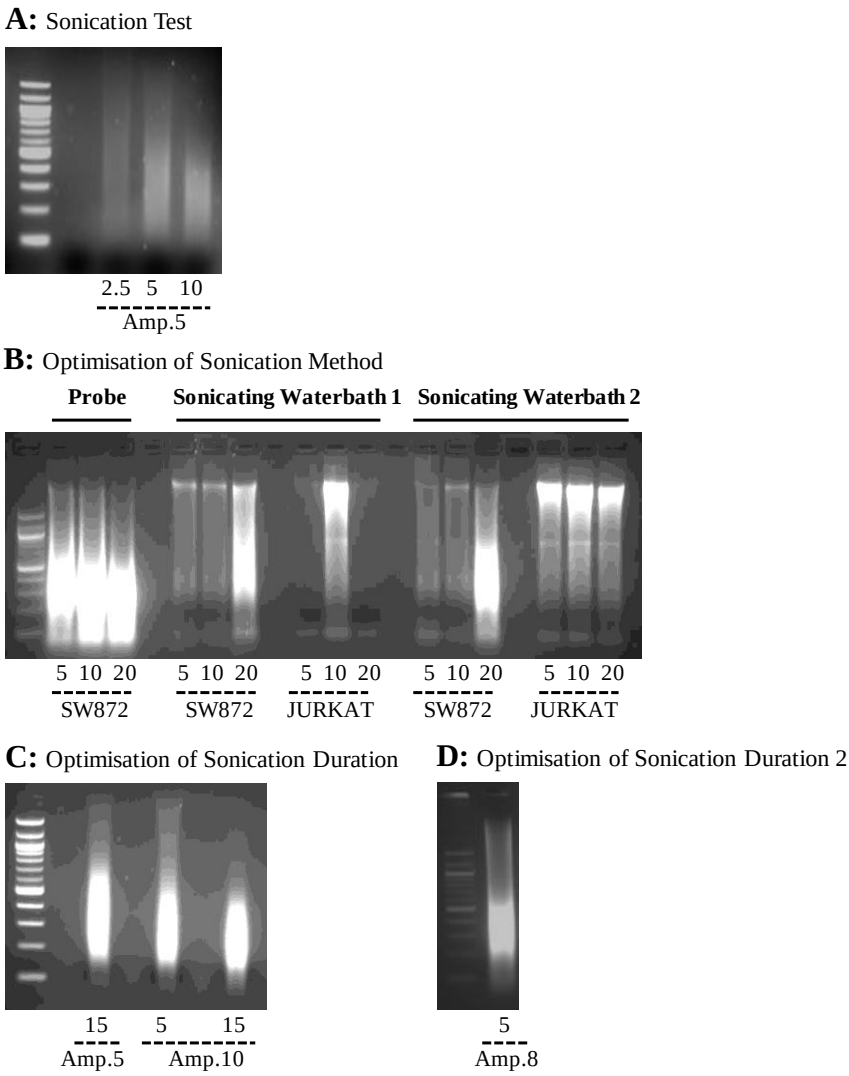


Figure 6.3

The optimisation of the conditions for fragmenting DNA by sonication. **(A)** DNA was sonicated using the probe sonicator at amplitude 5 for periods of 10 seconds, followed by 20 seconds rest (either 5 times (2.5 minutes total), 10 times (5 minutes total) or 20 times (10 minutes total)). The total sonication time (including rest periods) in minutes is shown beneath the image. DNA was purified by phenol-chloroform extraction. **(B)** Two different cell lines were used for the optimisation of the sonication method, JURKAT and SW872. The probe sonicator, and two sonicating water baths were tested. Using the probe, DNA was sonicated at amplitude 5 for periods of 10 seconds, followed by 20 seconds rest, for total periods of 5, 10 or 20 minutes. Using the two sonicating waterbaths, DNA was sonicated for periods of 5 minutes, followed by 5 minutes on ice, for a total of 5, 10 or 20 minutes. DNA was purified using the spin column method. **(C)** The probe sonicator was used to sonicate DNA, two different amplitudes were tested, amplitude 5 (Amp.5) and amplitude 10 (Amp.10) for 10 second periods, with 20 seconds rest for a total duration of 5 or 15 minutes. DNA was purified using the spin column method. **(D)** The probe sonicator was used to sonicate DNA at amplitude 8 (Amp.8) for 10 second periods, with 20 seconds rest for a total duration of 5 minutes. DNA was purified using the spin column method. A 100bp DNA sizing ladder is shown in the first lane of each image.

6.3.2.2 Assessment of five DEAF-1 antibodies

DEAF-1 immunoprecipitation was performed using five different DEAF-1 antibodies (Table 2, Appendix). Following each ChIP the DNA that was isolated was purified and subsequently analysed by PCR as described in Materials and Methods section 2.8.4. The region amplified included the *GDF5* exon 1 region (-315 to -167 relative to the transcription start site and encompassing rs143383). The enrichment of *GDF5* DNA was assessed following electrophoresis of the samples through an agarose gel. Anti-acetyl histone H3 antibody was used as a positive control for the ChIP protocol, with enrichment signifying *GDF5* is an actively transcribed gene, whilst a species matched IgG antibody was used as a negative control. Chromatin that has not been subject to immunoprecipitation was used to provide a comparison for the amount of DNA present in each sample and was referred to as the Input (10% of the pre-immunoprecipitated DNA was used). The first ChIP experiment was performed using DEAF-1 antibody 1 (Figure 6.4A). There was enrichment of *GDF5* DNA following ChIP with the Histone H3 positive antibody and a moderate amplification of DNA following ChIP with the IgG negative antibody, most likely an indication of the incomplete removal of unbound DNA during the washing stages of the protocol. There was an enrichment of *GDF5* DNA following ChIP with the anti-DEAF-1 antibody relative to the negative antibody. The *GAPDH* gene was then used to assess DEAF-1 binding to a negative control region; however, there was also an enrichment of *GAPDH* DNA following ChIP using the anti-DEAF-1 antibody. These results indicate that either DEAF-1 is binding to the *GADPH* gene in addition to the *GDF5* gene, or that DEAF-1 antibody 1 is binding non specifically.

DEAF-1 antibody 2 was next tested. There was however no enrichment of *GDF5* DNA following ChIP with this antibody (Figure 6.4B). DEAF-1 has been reported previously to bind to its own promoter to self regulate its expression [348]. In order to examine further the negative result obtained for antibody 2, I assessed by PCR the enrichment of a region of the *DEAF-1* promoter following ChIP with this antibody. DEAF-1 binding to this region however was also negative, with no enrichment of DEAF-1 DNA observed (Figure 6.4B). The negative binding to *GDF5* observed with antibody 2 could therefore not be verified. DEAF-1 antibody 3 was therefore next tested. The same result was obtained for this antibody as was seen for antibody 2; there was no enrichment of *GDF5* DNA or *DEAF-1* DNA following ChIP (Figure 6.4C). For both antibody 2 and 3 there was amplification of *GDF5* and *DEAF-1* DNA following ChIP with the positive control antibody, indicating the protocol was successful.

Finally, antibodies 4 and 5 were tested. As above, there was no enrichment of *GDF5* DNA following ChIP with these antibodies (Figure 6.5 A and B).

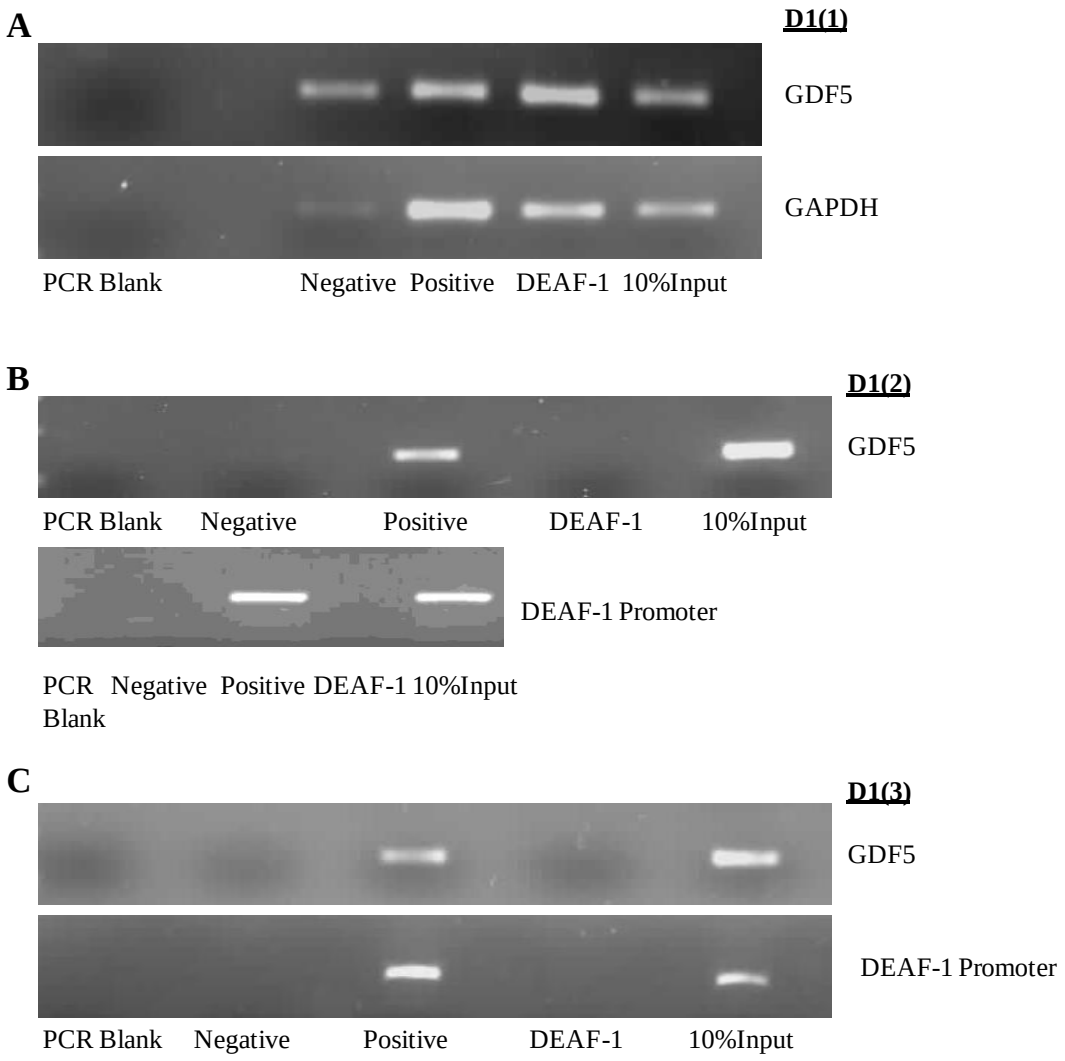


Figure 6.4

ChIP analysis of DEAF-1 using antibodies 1, 2 and 3. **(A)** Sheared genomic DNA was immunoprecipitated with DEAF-1 antibody 1 (D1(1)), rabbit polyclonal IgG control (negative) and anti-acetyl histone H3 control (positive) antibodies and then PCR amplified across either *GDF5* or *GAPDH*. **(B)** Sheared genomic DNA was immunoprecipitated with DEAF-1 antibody 2 (D1(2)), rabbit polyclonal IgG control (negative) and anti-acetyl histone H3 control (positive) antibodies and then PCR amplified across either *GDF5* or *DEAF-1*. **(C)** Sheared genomic DNA was immunoprecipitated with DEAF-1 antibody 3 (D1(3)), rabbit polyclonal IgG control (negative) and anti-acetyl histone H3 control (positive) antibodies and then PCR amplified across either *GDF5* or *DEAF-1*. The input represents 10% of the non-immunoprecipitated sheared genomic DNA. A PCR containing no DNA was used as a control (PCR Blank).

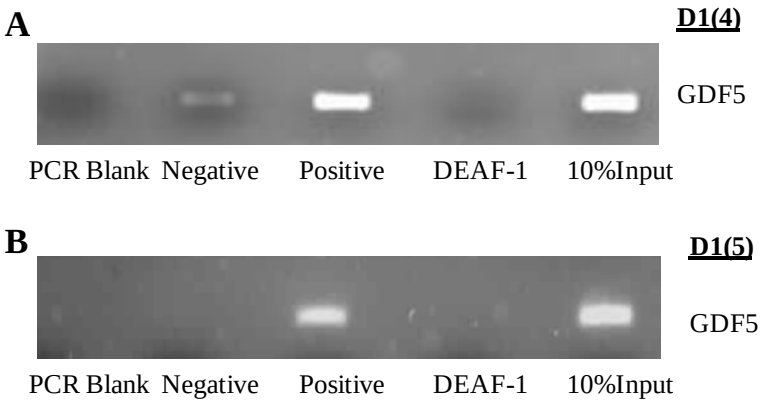


Figure 6.5

ChIP analysis of DEAF-1 using antibodies 4 and 5. **(A)** Sheared genomic DNA was immunoprecipitated with DEAF-1 antibody 4 (D1(4)), rabbit polyclonal IgG control (negative) and anti-acetyl histone H3 control (positive) antibodies and then PCR amplified across *GDF5*. **(B)** Sheared genomic DNA was immunoprecipitated with DEAF-1 antibody 5 (D1(5)), rabbit polyclonal IgG control (negative) and anti-acetyl histone H3 control (positive) antibodies and then PCR amplified across *GDF5*. The input represents 10% of the non-immunoprecipitated sheared genomic DNA. A PCR containing no DNA was used as a control (PCR Blank).

6.3.3 Characterisation of the DEAF-1 Antibodies

Four of the five antibodies targeting DEAF-1 yielded negative ChIP results suggesting that DEAF-1 does not bind to the region of *GDF5* examined. Antibodies 2 and 3 were however also tested for binding to the *DEAF-1* promoter and binding at this region was also negative. As DEAF-1 has previously been reported to bind to its own promoter, these negative results call into question the specificity of the antibodies. I therefore sought to characterise each of the five antibodies, assessing their specificities.

6.3.3.1 Knockdown of DEAF-1 and Immunofluorescence

The ability of the five antibodies to recognise endogenous DEAF-1 protein was first assessed by immunofluorescence (Figure 6.6; details of the antibodies used are provided in Table 2, Appendix). SW872 cells were untreated or treated with DEAF-1 siRNA (Details of DEAF-1 siRNA are given in Table 6, Appendix). Nuclei are stained blue with DAPI, as shown in the first row. The localisation of antibody binding is shown in the second row. The final row shows the merged DAPI and antibody images. Antibody 1 (D1(1)) appears to be binding to a protein that is abundant outside of the nucleus (Figure 6.6A). Furthermore, the amount of antibody binding does not appear to be reduced upon treatment with DEAF-1 siRNA when compared with the untreated cells. Antibody 2 is binding to a protein that is localised within the nucleus; the co localisation of DAPI with the antibody is shown in the merged image (Figure 6.6B). The amount of

antibody binding however does not appear to be reduced upon treatment with DEAF-1 siRNA in comparison with the untreated cells. Antibody 3 also appears to be binding to a protein that is located within the nucleus, but as with antibody 2, the amount of binding is not reduced upon treatment with DEAF-1 siRNA (Figure 6.6C). Antibody 4 is recognising a protein that is primarily located within the nucleus, with a small amount of binding in the cytoplasm. The amount of antibody binding appears to be reduced upon treatment with DEAF-1 siRNA; there is a lower intensity of immunofluorescence observed in the DEAF-1 siRNA treated cells when compared with the untreated control cells (Figure 6.6D). Lastly, antibody 5 also appears to be binding a protein that is localised within the nucleus, however the antibody binding is not reduced upon treatment with DEAF-1 siRNA when compared with the untreated cells (Figure 6.6E).

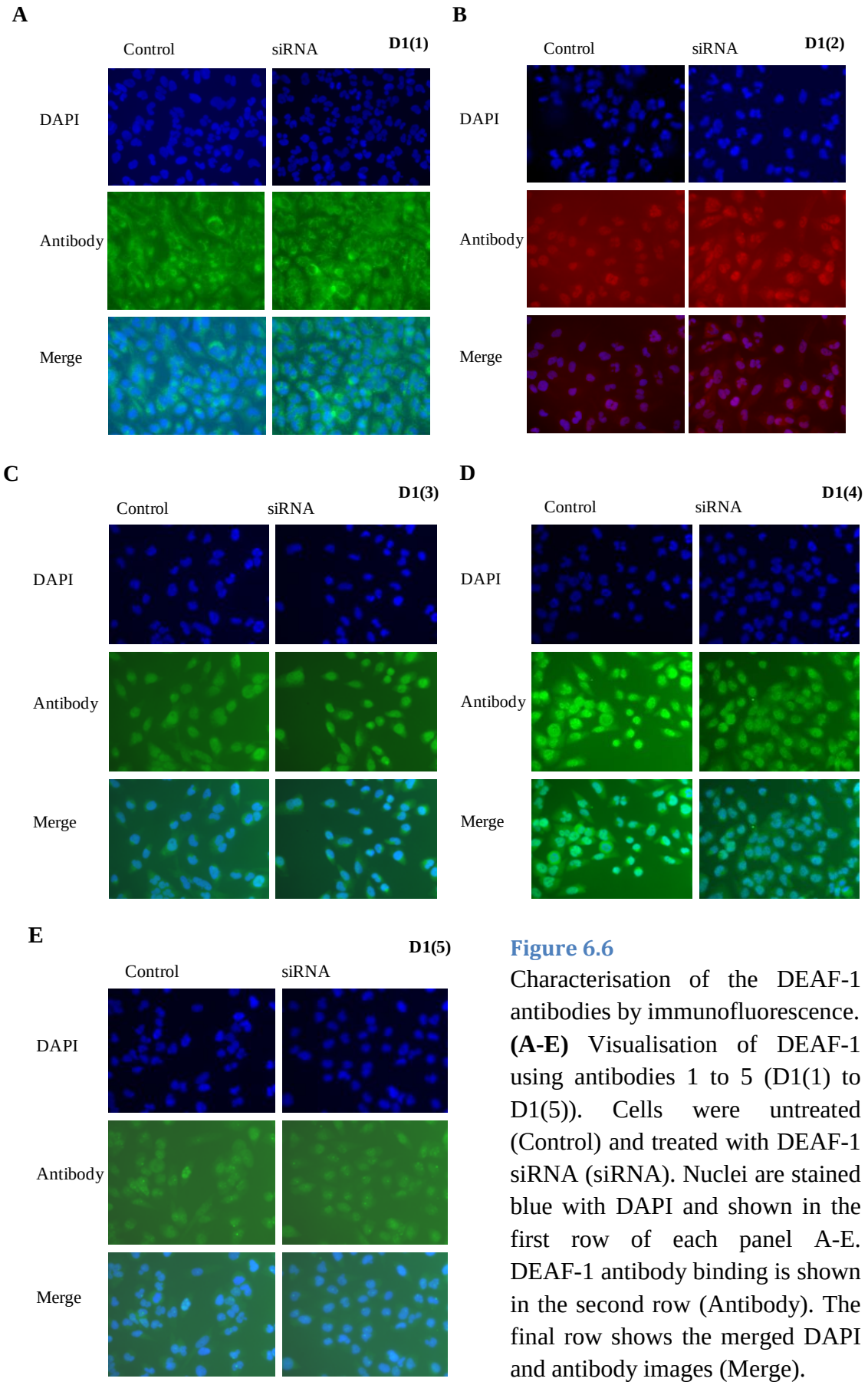


Figure 6.6

Characterisation of the DEAF-1 antibodies by immunofluorescence. **(A-E)** Visualisation of DEAF-1 using antibodies 1 to 5 (D1(1) to D1(5)). Cells were untreated (Control) and treated with DEAF-1 siRNA (siRNA). Nuclei are stained blue with DAPI and shown in the first row of each panel A-E. DEAF-1 antibody binding is shown in the second row (Antibody). The final row shows the merged DAPI and antibody images (Merge).

6.3.3.2 Addition of Recombinant Protein

Antibodies 2-5 are recognising a nuclear protein, and antibody 4 is binding to a protein that appears to be reduced upon treatment with DEAF-1 siRNA. However, when used for ChIP, these four antibodies yielded negative results. I therefore further tested the ability of each of these four antibodies to recognise DEAF-1 protein and DEAF-1 recombinant protein (tagged with glutathione S-transferase (GST), which is approximately 26kDa in size) by immunoblotting. DEAF-1 protein levels were assessed in SW872 cells following no treatment (cells), following treatment with a control non targeting siRNA (NTsiRNA) and following DEAF-1 siRNA treatment (D1 siRNA). Figure 6.7A is an immunoblot, probed for DEAF-1 using antibody 2. It is unclear from the immunoblot which protein is DEAF-1; there is no obvious reduction in the intensity of any of the proteins upon DEAF-1 siRNA treatment at either time point. The antibody appears to recognise DEAF-1 recombinant protein, as indicated by the arrow in Figure 6.7A. Using antibody 3, as with antibody 2 it is unclear which protein is DEAF-1 and there is no apparent reduction in intensity of the proteins upon DEAF-1 siRNA treatment. This antibody does not detect recombinant DEAF-1 protein. For antibody 4, there is a reduction in the intensity of one of the proteins upon treatment with DEAF-1 siRNA at 96 hours, indicated by the arrow in Figure 6.7C. Furthermore, this antibody appears to be able to detect recombinant DEAF-1 protein, as indicated by the second arrow. With DEAF-1 antibody 5, there is no apparent reduction in any of the protein upon DEAF-1 siRNA treatment at either time point. This antibody does not detect DEAF-1 recombinant protein.

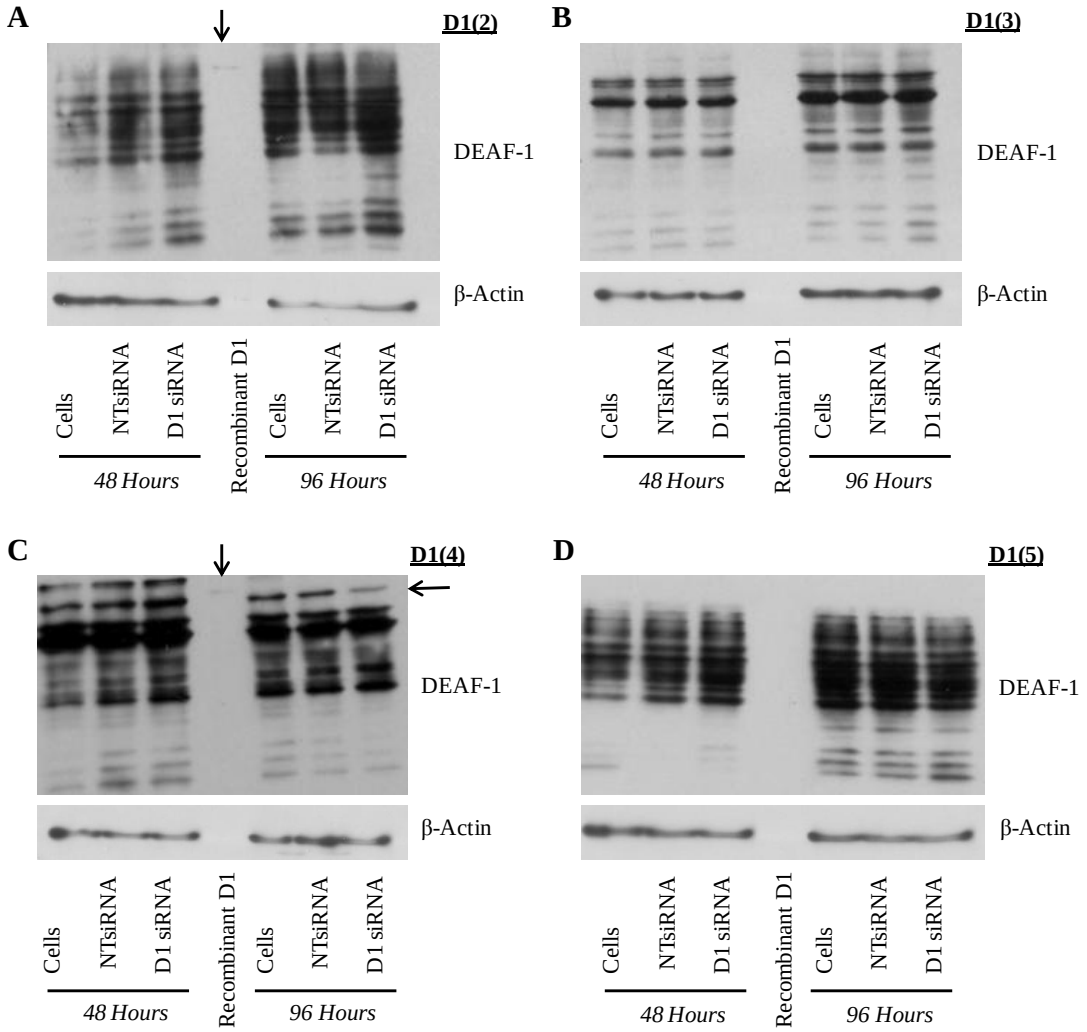


Figure 6.7

Characterisation of the DEAF-1 antibodies by immunoblotting. **(A-D)** Immunoblots demonstrating the effect of DEAF-1 siRNA treatment after two different incubation periods (48 and 96 hours). Protein extracted from cells treated with the NTsiRNA control was used for basal protein expression and cells are untreated protein samples. β-Actin was used as a loading control. GST tagged recombinant DEAF-1 protein (D1) was loaded as a positive control. Four of the DEAF-1 antibodies were used for immunoblotting (antibodies 2-5, D1(2) – D1(5)). Arrows indicate the identification of DEAF-1 recombinant protein (Recombinant D1).

6.3.3.3 Optimisation of the Over expression of DEAF-1

Although antibodies 2 and 4 appear to recognise DEAF-1 recombinant protein, it is still unclear if the antibodies are able to recognise endogenous protein within the cells. In order to determine if this uncertainty is a result of poor specificity of the antibodies, or alternatively if this could be a result of low levels of DEAF-1 protein within the SW872 cells, I obtained an enhanced green fluorescent protein (EGFP) tagged expression vector for DEAF-1 to over express DEAF-1 protein.

I first optimised the over expression of DEAF-1 EGFP within the SW872 cell line. I tested the transfection reagent (Fugene HD) at different transfection reagent: plasmid DNA ratios, different cell densities and for 24 and 48 hour transfection periods. The optimal conditions that resulted in the highest number of cells being transfected with the EGFP tagged plasmid are described in Materials and Methods section 2.16.1.

6.3.3.4 Over expression of DEAF-1

Following this optimisation, I assessed the ability antibodies 2-5 to detect over expressed DEAF-1. Figure 6.8 depicts the immunoblots following incubation with each of the four antibodies. The expression of DEAF-1 was assessed in SW872 nuclear and cytoplasmic protein extracts (N and C), in DEAF-1 EGFP transfected cells (D1), and in cells transfected with a control EGFP vector (H11). Following transfection of the cells with DEAF-1 EGFP and subsequent detection with each of the four antibodies (D1(2), D1(3), D1(4), D1(5)), it is clear that there is a protein being over expressed of approximately 130kDa in size. This protein is not present within the nuclear or cytoplasmic extracts or in the control EGFP vector treated cells. Panel B shows the protein levels of β-Actin, which is used as a loading control. Panel C shows the expression levels of EGFP. EGFP protein is detected using the EGFP antibody in the cells transfected with DEAF-1 EGFP (D1) and corresponds to the same molecular weight as the protein detected using the DEAF-1 antibodies. The control EGFP vector (H11) also shows an increase in the expression of EGFP protein, which is of lower molecular weight to the DEAF-1 EGFP protein. Collectively, these results suggest that DEAF-1 antibodies 2-5 can each recognise over expressed DEAF-1 protein.

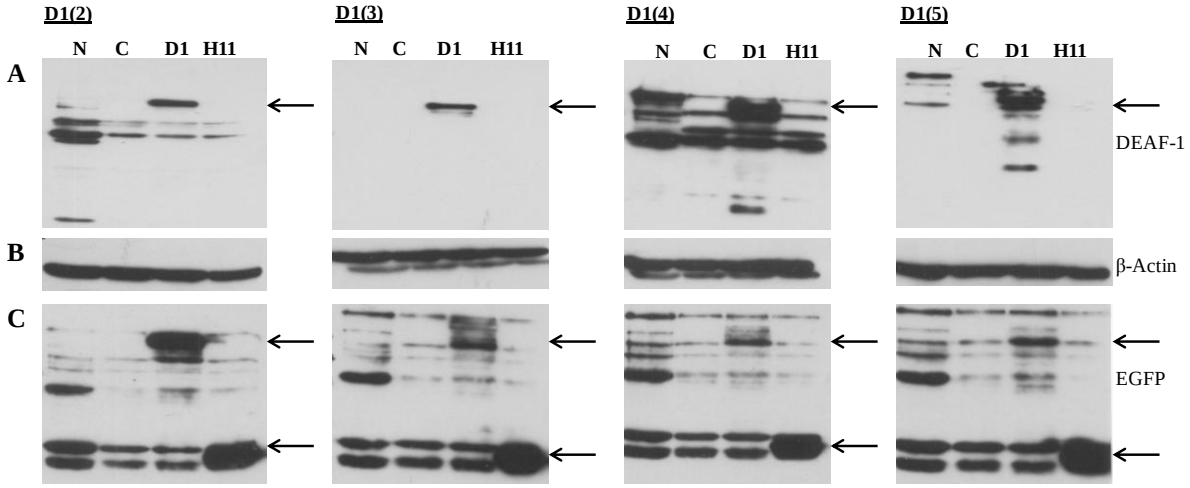


Figure 6.8

Characterisation of the DEAF-1 antibodies following DEAF-1 EGFP over expression. **(A)** Immunoblots examining the expression of DEAF-1 using four of the DEAF-1 antibodies (antibodies 2-5, D1(2)-(5)) in SW872 nuclear extract (N), SW872 cytoplasmic extract (C) and in SW872 nuclear extract following the over expression of either DEAF-1 EGFP (D1) or a control vector (H11). Arrows indicate DEAF-1 EGFP expression. **(B)** β-Actin expression, used as a loading control. **(C)** Detection of the protein using an antibody targeting EGFP. The upper arrow indicates DEAF-1 EGFP and the lower arrow indicates H11 EGFP.

I have been unable to confirm the protein levels of DEAF-1 following DEAF-1 siRNA treatment in the previous experiments (Figure 6.7), I therefore sought to assess if the DEAF-1 siRNA is able to successfully deplete DEAF-1 protein. To investigate this, I first over expressed DEAF-1 EGFP and then treated these cells with DEAF-1 siRNA. DEAF-1 EGFP over expression can be detected in the cells transfected with DEAF-1 EGFP (Figure 6.9). Cells that have been transfected with the DEAF-1 over expression vector and DEAF-1 siRNA have reduced levels of DEAF-1 EGFP protein, in comparison with cells treated with the non targeting siRNA control (NTsiRNA), suggesting that the DEAF-1 siRNA is able to successfully deplete DEAF-1 protein. β-Actin is used as a loading control. DEAF-1 EGFP expression was assessed using DEAF-1 antibody 3 (As Figure 6.8 showed that DEAF-1 antibodies 2-5 could all detect DEAF-1 EGFP).

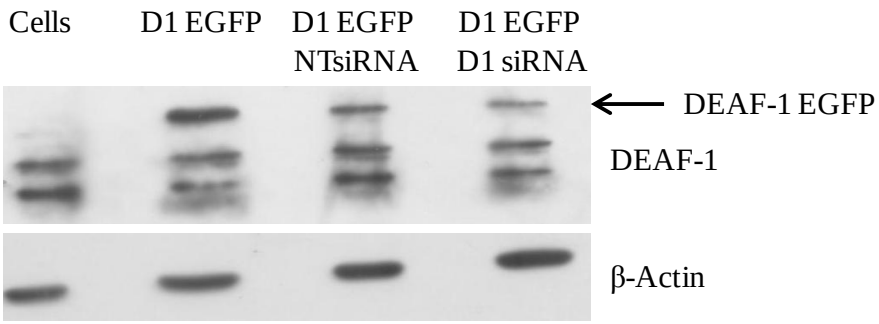


Figure 6.9
Examination of the effect of DEAF-1 siRNA treatment on DEAF-1 EGFP expression. Immunoblots demonstrating the effect of over expressing DEAF-1 EGFP (D1 EGFP) and the effect of concurrently depleting DEAF-1 expression using siRNA (DEAF-1 EGFP D1 siRNA). Protein extracted from cells that are over expressing DEAF-1 EGFP and have been treated with the NTsiRNA control (D1 EGFP NTsiRNA) were used for assessing basal protein expression. β-Actin was used as a loading control. Arrow indicates DEAF-1 EGFP expression.

The final test that I performed using DEAF-1 antibodies 3-5 was to assess their ability to immunoprecipitate DEAF-1 EGFP (antibody 2 was not used for this experiment because it was provided at a low concentration, making the volume required for the immunoprecipitation too high). Nuclear lysate from cells that had been transfected with DEAF-1 EGFP were used for the immunoprecipitation, as described in Materials and Methods section 2.19. Figure 6.10 shows the protein detected following immunoblotting with the anti-EGFP antibody. The first two lanes are controls, lane one containing no antibody and lane two containing a species matched IgG control; binding detected within lane two is background and as such the other antibodies were compared to this control sample. Immunoprecipitated protein obtained using anti-DEAF-1 antibodies 3, 4 and 5 is shown in lanes three, four and five respectively. EGFP transfected lysate is shown in lane six and represents a positive control enabling the migration size of DEAF-1 EGFP to be determined. Proteins of the same molecular weight as DEAF-1 EGFP are apparent following immunoprecipitation with DEAF-1 antibodies 4 and 5. These two antibodies can therefore be used to successfully immunoprecipitate over expressed DEAF-1 protein.

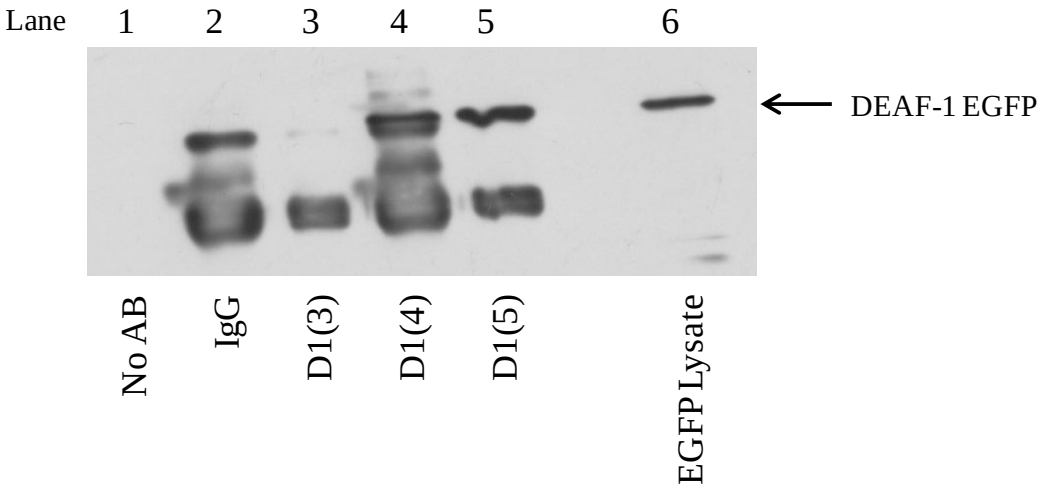


Figure 6.10

Immunoprecipitation of DEAF-1 EGFP. Immunoblot demonstrating the expression of DEAF-1 following immunoprecipitation using SW872 cell lysates that are over expressing DEAF-1 EGFP with three of the DEAF-1 antibodies (antibodies 3-5, D1(3), D1(4) and D1(5). EGFP cell lysates were either not treated with antibody (No AB), treated with a negative control rabbit IgG antibody or with one of the three DEAF-1 antibodies. SW872 lysate over expressing DEAF-1 EGFP was used as a positive control (EGFP lysate). Arrow indicates DEAF-1 EGFP.

6.3.3.5 DEAF-1 Over expression and ChIP

As reported above, I have previously used five antibodies targeting DEAF-1 for ChIP, four of which yielded negative results. As I have now demonstrated that DEAF-1 EGFP can successfully be over expressed and that two of the five antibodies can immunoprecipitate DEAF-1 EGFP, I next sought to assess if DEAF-1 binding to *GDF5* could be detected using ChIP following the over expression of DEAF-1. I used antibody 4 for the ChIP experiments as this antibody has been shown in this chapter to specifically bind to DEAF-1. Figure 6.11A shows the enrichment of *GDF5* DNA following ChIP and PCR at three different cycle numbers. *GDF5* DNA is enriched following immunoprecipitation with antibody 4 in comparison with the IgG negative control antibody (Neg). This enrichment is apparent at all three different cycle numbers. Figures 6.11B and C represent an additional two immunoprecipitation experiments performed. The second EGFP ChIP experiment was performed using both DEAF-1 antibody 4 and an anti-EGFP antibody. It is unclear as to whether there is enrichment of *GDF5* following immunoprecipitation using antibody 4 in comparison with the IgG negative antibody control (Neg). *GDF5* DNA is enriched following immunoprecipitation using the EGFP antibody in comparison with the antisera negative antibody control (Neg AS). *GDF5* DNA is also enriched following the third immunoprecipitation in comparison with the negative antibody control (Figure 6.11C).

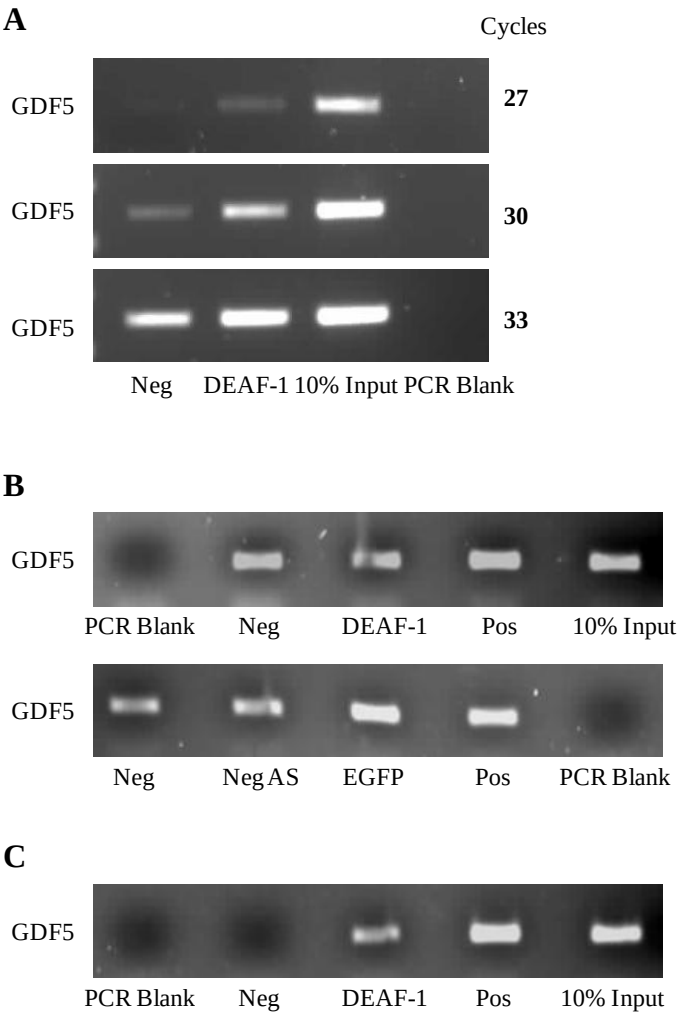


Figure 6.11

ChIP analysis of DEAF-1 in cells over expressing DEAF-1 EGFP. **(A)** Sheared genomic DNA was immunoprecipitated with DEAF-1 antibody 4 (DEAF-1) and rabbit polyclonal IgG control (Neg) antibody and then PCR amplified for either 27, 30 or 33 cycles across *GDF5* exon 1. **(B)** A second ChIP using DEAF-1 antibody 4 and an anti-EGFP antibody. Sheared genomic DNA was immunoprecipitated with antibody 4 (DEAF-1), rabbit polyclonal IgG control (Neg) and anti-acetyl histone H3 control (Pos) antibodies and then PCR amplified across *GDF5* exon 1 (upper gel image). Sheared genomic DNA was immunoprecipitated with EGFP, rabbit polyclonal IgG control (Neg), rabbit antisera (Neg AS) and anti-acetyl histone H3 control (Pos) antibodies and then PCR amplified across *GDF5* exon 1 (lower gel image). **(C)** A third ChIP using DEAF-1 antibody 4. Sheared genomic DNA was immunoprecipitated with DEAF-1 antibody four (DEAF-1), rabbit polyclonal IgG control (Neg) and anti-acetyl histone H3 control (Pos) antibodies and then PCR amplified across *GDF5* exon 1. In all experiments the input represents 10% of the non-immunoprecipitated sheared genomic DNA. A PCR containing no DNA was used as a control (Blank).

In addition to binding to the *GDF5* promoter I also assessed by PCR, using several different cycle numbers, DEAF-1 binding to its own promoter following over expression of DEAF-1 EGFP (Figure 6.12). There is enrichment of *DEAF-1* DNA following immunoprecipitation with DEAF-1 antibodies 4 and 5 and with the anti-EGFP antibody (apparent after $\geq$ **33** PCR cycles) in comparison with their respective controls. The IgG negative control antibody (Neg) is used as a comparison for antibody 4 and the antisera (Neg AS) is used as a negative control for both antibody 5 and the anti-EGFP antibody. The highest level of enrichment of *DEAF-1* DNA is achieved following immunoprecipitation with DEAF-1 antibody 5.

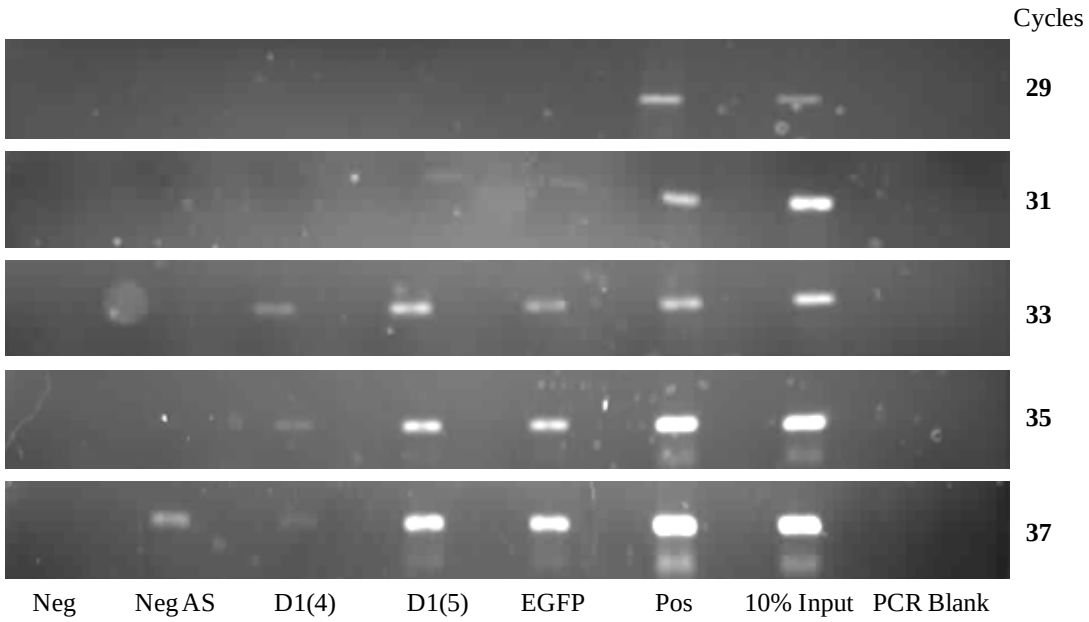


Figure 6.12
ChIP analysis of DEAF-1 at the *DEAF-1* promoter in cells over expressing DEAF-1 EGFP. Sheared genomic DNA was immunoprecipitated with DEAF-1 antibody 4 (D1(4)), DEAF-1 antibody 5 (D1(5)), EGFP antibody, rabbit polyclonal IgG control (Neg), rabbit antisera (Neg AS) and an anti-acetyl histone H3 control antibody (Pos). DNA was PCR amplified for either 29, 31, 33, 35 or 37 cycles across the *DEAF-1* promoter region. The input represents 10% of the non-immunoprecipitated sheared genomic DNA. A PCR containing no DNA was used as a control (PCR Blank).

6.4 Discussion

Following on from the functional studies performed by Egli et al., which identified DEAF-1 as a *trans*-acting factor that may be differentially regulating the expression of *GDF5*, and from the siRNA knockdown studies performed during my Master of Research project, I sought to further investigate the role of DEAF-1. I first investigated DEAF binding to *GDF5 in vitro* by EMSA. Although the addition of three different DEAF-1 competitors did not alter protein complex binding to the C and T allele probes, the addition of an antibody targeting DEAF-1 produced a supershifted complex. This supershifted complex was more intense in the T allele probe EMSA in comparison to the C allele probe EMSA. This supershift was confirmed using nuclear extract from human articular chondrocytes. A second DEAF-1 antibody however did not produce a supershifted protein complex.

The Sp1 and Sp3 complexes that were identified as binding to the C and T allele probes in Chapter 4 are abundant and could therefore be masking the binding of DEAF-1. To test this possibility, I investigated protein binding to the C and T allele probes using DEAF-1 recombinant protein. However, no complex binding was observed in the different conditions tested. It is therefore not possible from these EMSA experiments to determine if DEAF-1 is binding to *GDF5 in vitro.* Future experiments could use a range of DEAF-1 recombinant protein concentrations and test the addition of recombinant protein in addition to SW872 nuclear extract; the binding of DEAF-1 may be dependent upon the binding of other factors, and thus the use of DEAF-1 recombinant protein alone may prevent the binding of DEAF-1 to the probes. Furthermore, Sp1 and Sp3 could be depleted using siRNA, and the nuclear extract from these siRNA treated cells could be used to identify if there is any DEAF-1 binding masked by the abundant Sp1/Sp3 protein complexes.

The expression of DEAF-1 within the SW872 cell line may be low, therefore the binding of DEAF-1 may be difficult to detect by EMSA. Furthermore, the conditions used for the EMSA experiments may not be optimal for the binding of DEAF-1. A previous study has suggested that modest concentrations of the non specific competitor Poly (dI:dC) can displace DEAF-1 binding to its target gene. However, Poly (dI:dC) was removed from the DEAF-1 recombinant protein EMSA and no binding to the probes was observed. It is important to note that the binding of *trans*-acting factors to the 33bp probes *in vitro* may be different to the *in vivo* binding to the DNA. Following the EMSAs I therefore assessed DEAF-1 binding to *GDF5 in vivo* using ChIP. Five

different DEAF-1 antibodies were tested, one of which was positive and four of which were negative. For antibody 1, which produced a positive result, binding to the control gene *GAPDH* was also positive. For antibodies 2-5, binding to *GDF5* was negative. However, binding to the *DEAF-1* promoter was also negative; in the absence of a positive control region for DEAF-1 binding, I am unable therefore to use these results to confirm whether DEAF-1 is binding to *GDF5 in vivo.*

Following on from this negative ChIP result, I performed several experiments to deduce if each antibody was able to specifically recognise DEAF-1 protein. Using immunofluorescence I demonstrated that antibodies 2-5 were all recognising proteins that are primarily located in the nucleus. Antibody 1 however was binding to a protein located in the cytoplasm. DEAF-1 has been shown previously to be located in the nucleus, with a small level of expression also found in the cytoplasm [345]. I therefore concluded that antibody one was not binding specifically to DEAF-1. Using antibody 4, the immunofluorescence was reduced upon treatment with DEAF-1 siRNA, suggesting that this antibody may be recognising DEAF-1 protein. Using antibodies 2, 3 and 5, the immunofluorescence was not reduced upon treatment with DEAF-1 siRNA suggesting that either the siRNA is not able to deplete DEAF-1 protein or these antibodies are not specific. Antibodies 2 and 4 detected DEAF-1 recombinant protein by immunoblotting. Only antibody 4 was successful in detecting a decrease in protein following DEAF-1 siRNA treatment, 96 hours post transfection.

A number of DEAF-1 proteins of varying molecular mass have been detected [345]. In CV-1 (monkey kidney cell lines), JEG-3 cells (human placental choriocarcinoma cell line) and in rat testis a 72kDa protein was identified. In muscle, brain and heart tissue samples taken from rats lower molecular weight proteins of 59kDa have been detected [345]. Furthermore in HEK293 cells (human embryonic kidney cell line) DEAF-1 is 70kDa [298] .

DEAF-1 has not been detected previously in SW872 cells, it was therefore difficult to predict the size of the protein that is being expressed within these cells. In order to assess conclusively if any of the proteins on the immunoblot represent endogenous DEAF-1, the sample could be pre-absorbed with the peptide used for the generation of the antibody to investigate if any proteins disappear. I did not carry out this experiment due to the unavailability of the peptide sequence used for the generation of the antibodies.

Over expressed DEAF-1 protein (DEAF-1 EGFP) was detected using four of the five DEAF-1 antibodies (antibodies 2-5). Furthermore, following the over expression of DEAF-1, DEAF-1 siRNA depleted DEAF-1 EGFP. This result suggests that the DEAF-1 siRNA is working; thus the negative depletion results obtained with the endogenous DEAF-1siRNA immunoblots implies that antibodies 2, 3 and 5 are not able to detect endogenous DEAF-1 protein. Finally, DEAF-1 EGFP was successfully immunoprecipitated using DEAF-1 antibodies 4 and 5.

Taking into account these antibody characterisation results, I concluded that DEAF-1 antibody 4 was binding to DEAF-1, as this antibody recognised a nuclear protein that is depleted upon DEAF-1 siRNA treatment. Additionally, DEAF-1 recombinant protein and DEAF-1 EGFP were detected using this antibody by immunoblotting. Finally, this antibody was successfully used to immunoprecipitate DEAF-1 EGFP. DEAF-1 antibody 4 produced a supershifted band when added to the EMSA reaction suggesting DEAF-1 is binding *in vitro* to *GDF5*, but produced a negative result for *in vivo* binding assessed by ChIP. In light of the absence of a positive region for DEAF-1 binding in the ChIP experiments, the final section of this chapter assessed the *in vivo* binding of DEAF-1 using ChIP following the over expression of DEAF-1 EGFP. In three independent ChIP experiments, *GDF5* DNA was enriched following immunoprecipitation with this antibody compared with the IgG control antibody. Furthermore, the binding of DEAF-1 was confirmed at the *DEAF-1* promoter by ChIP using DEAF-1 antibodies 4, 5 and an EGFP antibody following DEAF-1 over expression.

Correspondence with other researchers who investigate DEAF-1 has confirmed the difficulties in detecting endogenous DEAF-1 within both cell lines and tissues (Michael Collard, Southern Illinois University and Linda Yip, Stanford University). These researchers have suggested that DEAF-1 protein is expressed at a low level in all cell types investigated and thus is difficult to detect. When over expressed, ChIP experiments suggest that DEAF-1 is binding to *GDF5 in vivo*, however in untreated cells DEAF-1 binding is negative. It is tempting to speculate that DEAF-1 is binding to *GDF5 in vivo* and this can be detected when increasing the amount of DEAF-1 within SW872 cells. However, it could be argued that when DEAF-1 protein is over expressed, an abundance of the protein within the cell may lead to non specific binding of DEAF-1. It is therefore not possible to state for certain that DEAF-1 is binding *in vivo* to *GDF5* under normal cellular conditions.

DEAF-1 has been identified as an important transcription factor in regulating the expression of the *5HT-1A receptor* [349-351]. The 5HT-1A receptor is an auto-receptor found in pre-synaptic neurons in the brain, responsible for regulating the levels of the neurotransmitter serotonin. Alterations in the levels of serotonin in the brain have been associated with depression and the auto-receptor is important for maintaining normal physiological levels of this neurotransmitter [352]. In DEAF-1 knockout mice, there was increased expression of the 5HT-1A autoreceptor and reduced raphe serotonin levels [353]

A polymorphism (C -1019 G) within the promoter of the *5HT-1A* receptor gene was found to reduce the repressor activity of DEAF-1 at this locus, promoting increased *5HT-1A* receptor gene expression and affecting response to antidepressant treatment. Increased expression of the *5HT-1A* receptor decreases the levels of serotonin in the synapse and thus increases susceptibility to depression [349, 354]. This effect however was found to be cell type specific, such that DEAF-1 was found to have the opposite effect at postsynaptic neuronal cells, where binding of DEAF-1 to the *5HT-1A* receptor promoter enhanced transcription [350]. This study highlights the importance of DEAF-1 in the regulation of target genes and the dual repressor and enhancer functions in different cell types reveals the complex nature of this transcription factors distinct functional roles. Furthermore, this study suggests that a single base change within the DNA binding sequence of DEAF-1 can alter its binding affinity and repressive activity. As mentioned previously, the core consensus sequence of DEAF-1 is TCGG, which resides directly over the *GDF5* rs143383 C allele; changing this base to a T allele may account for a change in repressive activity, thus could account for an alteration in the expression of *GDF5*.

Other than the *5HT-1A* receptor, other target genes of DEAF-1 include the *nuclear ribonuclearprotein (RNP) A2/B1* promoter (a potential early biomarker in lung cancer) and genes encoding peripheral tissue antigens in the pancreatic lymph nodes [298, 348]. Additionally, a study has shown that in patients with type 1 diabetes, lower expression of the canonical DEAF-1, and higher expression of an alternatively spliced variant lead to a reduction in the expression of peripheral tissues antigens, promoting diabetes disease pathogenesis [298]. This is due to the impaired formation of tolerance and the survival of auto-reactive T cells that are specific for the tissue antigens [298]. These studies highlight that DEAF-1 is an important *trans*-acting factor, regulating the expression of a number of genes.

The functional domains of DEAF-1 have been defined and include a SAND (Sp100, AIRE-1, NucP41/75 and DEAF-1) domain which contains the nuclear localisation signal and confers DNA binding activity and a MYND domain which mediates protein-protein interactions [343, 355, 356]. DEAF-1 has been suggested to form dimers with itself and has been reported to interact with Lim only protein 4 (LMO4) and members of the Clim protein family [357]. Furthermore, the C terminal region of DEAF-1 shares amino acid homology with the repressor domain of the transcription factor MTG8. MTG8 has been reported to interact with a number of repressor proteins such as N-CoR and Sin3A. This degree of similarity suggests that the C terminal region of DEAF-1 may also be involved in protein-protein interactions. Deletion experiments involving the removal of parts of the C terminal region of DEAF-1, resulting in a lower level of transcriptional repression have confirmed that this region may recruit co repressors [348].

In summary, DEAF-1 is expressed in many neuroendocrine and reproductive tissues and is expressed at high levels in the foetus, suggesting an important role during development. In addition to binding to and activating or repressing the transcription of target genes, it may also form complexes with a number of other proteins and modulate their function. Further experiments will be carried out in the following chapters of this thesis to study DEAF-1 binding to *GDF5* and its effect on the expression of this gene.

Chapter 7: The Investigation of the Functional Effects of Sp1, Sp3, P15 and DEAF-1 on *GDF5* Gene Expression

7.1 Introduction

Sp1 and Sp3 were predicted to bind to *GDF5* using online prediction tools. Using EMSAs I have confirmed the binding of these two factors to *GDF5 in vitro*. Furthermore, the binding of Sp1 and Sp3 is modulated by the genotype at rs143383, with both factors binding more avidly to the T allele in comparison with the C allele (Chapter 4). P15 was identified as binding to *GDF5* DNA *in vitro* using an oligonucleotide pull down assay and tandem mass spectrometry (Chapter 5). P15 was found to be more abundant in the T allele DNA sample in comparison with the C allele DNA sample. DEAF-1 has previously been suggested to bind to *GDF5* at rs143383, and on the addition of an antibody to the EMSA reaction I observed a supershifted complex. The experiments I have performed to identify *in vivo* binding of DEAF-1 to *GDF5* however have been inconclusive (Chapter 6).

All four of these factors are known to be involved in transcriptional regulation. Further work in this chapter will assess the extent to which each factor is modulating both the overall and allelic expression of *GDF5*.I will also assess whether manipulating the expression of these factors can alter the expression of *GDF5*.

The cloning of the Sp1, Sp3 and P15 expression constructs was carried out with assistance from Dr. Louise Reynard.

7.2 Aims

- To investigate the binding of Sp1, Sp3 and P15 to *GDF5 in vivo* in SW872 cells using chromatin immunoprecipitation (ChIP).
- To knockdown the expression of Sp1, Sp3, P15 and DEAF-1 using siRNA and examine the effect on the overall and the allelic expression of *GDF5*.
- To over express the four factors and examine the *GDF5* rs143383 C and T allele promoter activities using a luciferase reporter assay.

7.3 Results

7.3.1 The Investigation of Sp1, Sp3 and P15 binding to *GDF5 in vivo* using Chromatin Immunoprecipitation (ChIP)

The sonication conditions for ChIP were previously determined within the SW872 cell line in Chapter 6. Following immunoprecipitation using antibodies against Sp1, Sp3, P15 and a species matched IgG control, the DNA was purified and amplified by PCR as described in Materials and Methods section 2.8. The enrichment of *GDF5* DNA was assessed following electrophoresis of the samples through an agarose gel. Figure 7.1 represents four independent experiments. Anti-acetyl histone H3 antibody was used as a positive control for the ChIP protocol and a species matched IgG antibody used as a negative control. In all four experiments, there is enrichment of *GDF5* DNA following ChIP with the positive antibody and a moderate amplification of DNA following ChIP with the negative antibody, most likely an indication of the incomplete removal of unbound DNA during the washing stages of the protocol.

GDF5 DNA is enriched following ChIP with antibodies targeting Sp3 and P15 relative to the negative antibody in all four ChIP experiments, most notably for P15 in experiments two, three and four. *GDF5* DNA is enriched following ChIP with an antibody targeting Sp1, relative to the IgG control, in ChIP experiments one, three and four. In experiment two there is less *GDF5* DNA following ChIP with an antibody targeting Sp1 in comparison to the negative control antibody suggesting a problem may have occurred. The DNA present in the 10% input sample is for the most part comparable with that present following ChIP with the four antibodies, suggesting that at least 10% of the chromatin inputted into the immunoprecipitation is bound by these factors. Overall, these results suggest that this region of *GDF5* is enriched for Sp1, Sp3 and P15 binding.

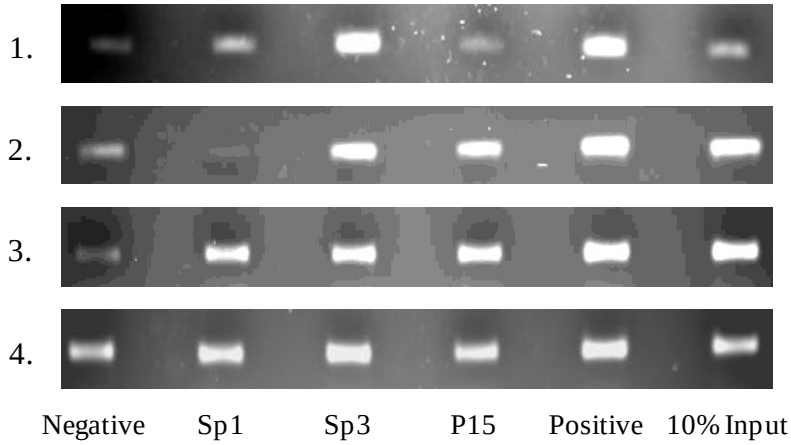


Figure 7.1
ChIP analysis of Sp1, Sp3 and P15. Sheared genomic DNA was immunoprecipitated with Sp1, Sp3, P15, rabbit polyclonal IgG (negative control) and anti-acetyl histone H3 (positive control) antibodies and then PCR amplified across exon 1 of *GDF5*. The input represents 10% of the non-immunoprecipitated sheared genomic DNA.

7.3.2 Investigation of *GDF5* expression following Sp1, Sp3, P15 and DEAF-1 depletion

Following the identification of the binding of Sp1, Sp3 and P15 *in vivo* by ChIP and the binding of DEAF-1 *in vitro* by EMSA (Chapter 6), I sought to assess if the four factors influence the expression of *GDF5*. Two techniques were used in order to assess the impact of each factor upon *GDF5* expression. The first, described in this section, assessed the effect of Sp1, Sp3, P15 and DEAF-1 depletion by RNAi in SW872 cells. The second, described in the following section, assessed the effect of over expressing the factors.

As described in Chapter 5, the optimal conditions for the knockdown of target proteins using siRNA were determined by me during my Master of Research project. The same conditions were used here and can be found in Materials and Methods section 2.9, whilst details of the siRNAs used can be found in Table 6, Appendix. The depletion of the mRNA for each gene was confirmed by real time RT-PCR. Significant decreases in the expression of *Sp1, Sp3, P15* and DEAF-1 were observed following siRNA treatment in comparison with the non-targeting siRNA (NTsiRNA) control ($p<0.05$) (Figure 7.2A). The depletion of target protein was confirmed by immunoblotting. Sp1, Sp3 and P15 protein levels were decreased relative to the NTsiRNA (Figure 7.2B). As discussed in Chapter 6, it was difficult to assess the depletion of DEAF-1 protein levels due to the low level of expression of DEAF-1 within the SW872 cell line (Figure 7.2B). I was however able to detect a decrease in the expression of over expressed EGFP tagged

DEAF-1 using immunofluorescence following DEAF-1 siRNA treatment suggesting that the siRNA is able to successfully deplete DEAF-1 protein (Chapter 6, Figure 6.9).

The overall expression of *GDF5* was increased following the depletion of each factor (Figure 7.2C). Small and non-significant fold changes were seen in *GDF5* expression following Sp1 (1.37 fold), Sp3 (1.8 fold) and P15 (1.38 fold) knockdown, relative to NTsiRNA. A significant fold change (4.6 fold, $p<0.001$) was observed upon DEAF-1 knockdown.

Allele specific real time PCR was then used to assess if the depletion of any of the four factors differentially affects the two alleles of rs143383, and as such could contribute to the DAE mediated by this SNP. DNA and RNA were extracted from the siRNA transfected cells and cDNA synthesis was performed as described in the Materials and Methods sections 2.10, 2.11 and 2.12. Depletion of Sp1 and Sp3 resulted in small and non-significant increases in the C/T allelic ratio (ratio of 2.1 in the control (NTsiRNA) to 2.7 (Sp1 siRNA) or 2.4 (Sp3 siRNA)) whilst P15 depletion did not alter the DAE (Figure 7.2D). DEAF-1 depletion increased the DAE from a C/T ratio of 2.1 in the control (NTsiRNA) to 4.7 (DEAF-1 siRNA) and this was highly significant ($p<0.001$, Figure 7.2D).

Overall, the data suggest that all four factors are involved in the transcriptional activity of *GDF5*, each repressing *GDF5* expression, with DEAF-1 also clearly contributing to *GDF5* DAE.

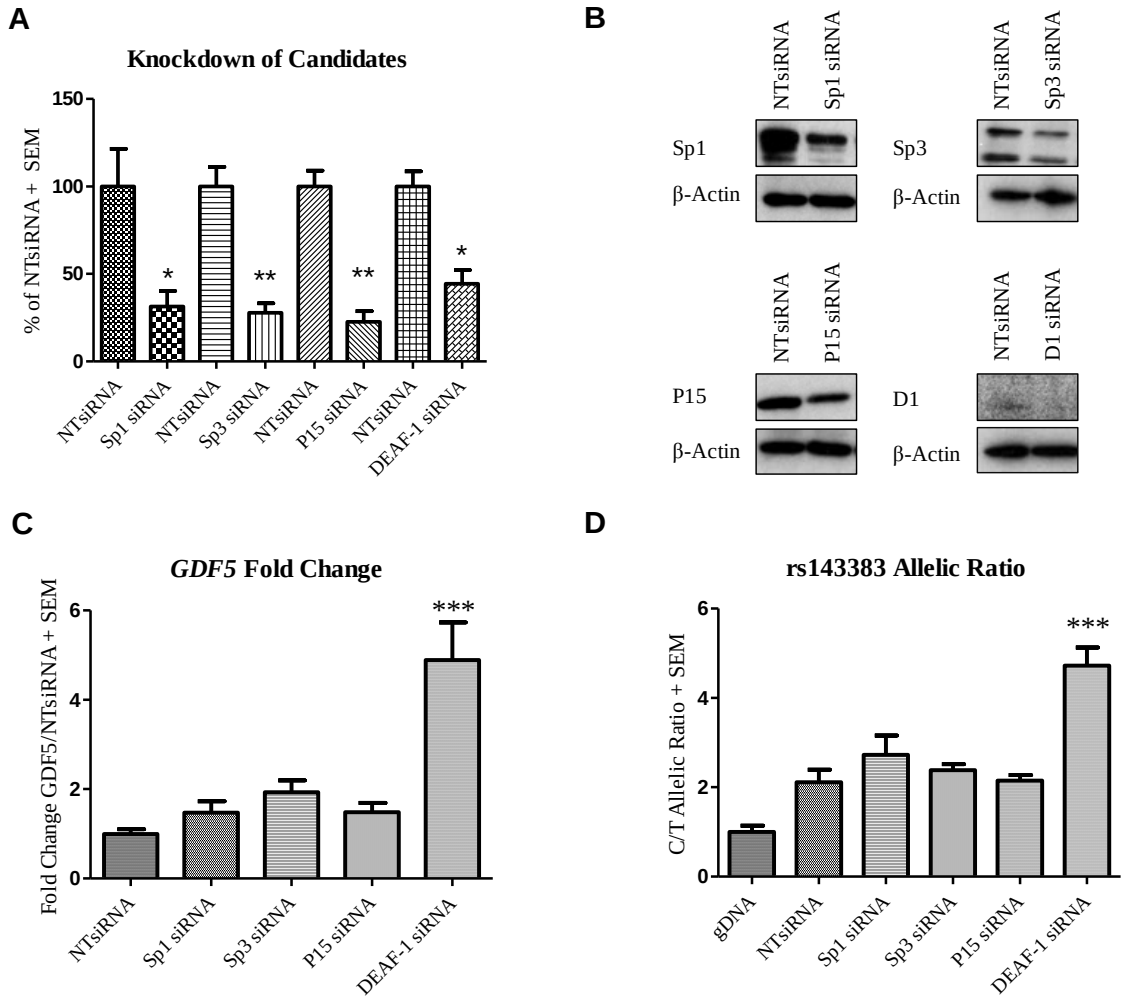


Figure 7.2

GDF5 expression following Sp1, Sp3, P15 and DEAF-1 depletion. (**A**) Expression levels of *Sp1, Sp3, P15* and *DEAF-1* mRNA are shown as a percentage of the control non-targeting siRNA (NTsiRNA) treated cells following *Sp1, Sp3, P15* and *DEAF-1* siRNA knockdown. Error bars denote the standard error of the mean (SEM). *p<0.05, calculated relative to the NTsiRNA value using a Student's 2 tailed *t*-test. (**B**) Immunoblots demonstrating Sp1, Sp3, P15 and DEAF-1 (D1) protein depletion following siRNA treatment. Protein extracted from cells treated with the NTsiRNA control was used for basal protein expression whilst β-Actin was used as a loading control. (**C**) Fold change in *GDF5* expression following *Sp1, Sp3, P15* and *DEAF-1* siRNA knockdown and shown relative to the NTsiRNA control. Error bars denote the SEM. ***p<0.001, calculated using a ANOVA. (**D**) The rs143383 C/T allelic ratio is shown following *Sp1, Sp3, P15* and *DEAF-1* siRNA knockdown and compared against treatment with the NTsiRNA control. Allelic ratios were normalised to genomic DNA (gDNA). Error bars denote the SEM. ***p<0.001, calculated using a ANOVA. Each siRNA experiment was performed three times and each with three technical repeats.

7.3.3 Investigation of *GDF5* expression following Sp1, Sp3, P15 and DEAF-1 over expression

The second technique employed to assess the impact of each candidate upon *GDF5* expression was their over expression. A region of *GDF5* containing the promoter and 5′UTR sequence encompassing rs143383 was cloned into a luciferase reporter plasmid as described in Materials and Methods section 2.15. Two constructs were generated, one containing the rs143383 C allele and the second containing the T allele by site directed mutagenesis (Materials and Methods section 2.15.3). Furthermore, I cloned the mRNA sequence of *Sp1, Sp3* and *P15* into an expression plasmid containing an enhanced green fluorescence protein (EGFP) tag. DEAF-1-EGFP was kindly donated by C. Garrison Fathman and Linda Yip. [298]. Details of the primers used to create inserts for cloning and those used for sequencing can be found in Table 5, Appendix. The over expression of plasmid DNA has been previously tested in the SW872 cell line in our laboratory by Andrew Dodd, PhD student, however the transfection efficiency was low. The chondrosarcoma cell line, SW1353 was used for the over expression experiments due to higher transfection efficiencies being observed.

I first optimised the transfection of the over expression plasmids into the SW1353 cells using the empty EGFP vector. I tested different cell confluencies and different transfection reagent: DNA ratios. I assessed the transfection efficiency by visualising the fluorescence 24 hours post transfection. The conditions chosen for transfection are described in Materials and Methods section 2.16. Over expression of Sp1, Sp3, P15 and DEAF-1 proteins was confirmed by immunoblotting and immunofluorescence (Figure 7.3A and B). Each EGFP tagged protein was localised to the nucleus following transfection as shown in Figure 7.3B.

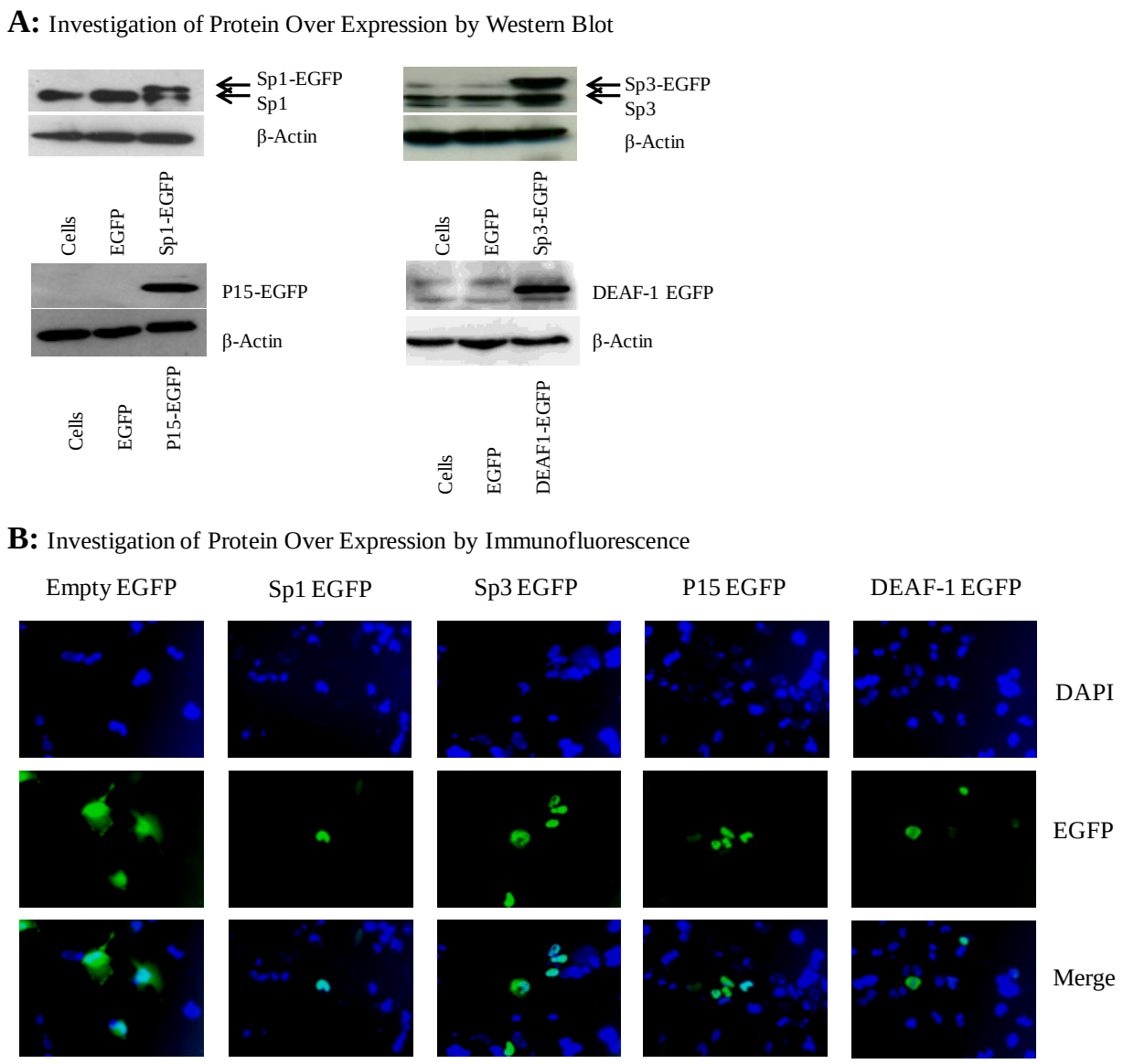


Figure 7.3

Over expression of Sp1, Sp3, P15 and DEAF-1. **(A)** Immunoblots showing Sp1 (Sp1-EGFP), Sp3 (Sp3-EGFP), P15 (P15-EGFP) and DEAF-1 (DEAF-1 EGFP) protein levels following over expression compared to the EGFP/pGL3 combination empty vector control (EGFP). Cells are untreated protein samples whilst β-Actin was used as a loading control. The arrows indicate basal protein and over expressed protein levels. **(B)** Immunofluorescence images following over expression. Nuclei are stained blue with DAPI and shown in the first row. The localisation of the EFGP fusion proteins (Empty EGFP, Sp1 EGFP, Sp3 EGFP, P15 EGFP and DEAF-1 EGFP) is shown in the second row (EGFP). The final row shows the merged DAPI and EGFP images (Merge).

Following transfection of the *GDF5* promoter luciferase constructs into SW1353 cells, I first assessed the effect that the C-T single nucleotide change had on luciferase activity. I observed that the presence of a T allele at rs143383 significantly reduced the luciferase activity in comparison with a C allele at rs143383 ($p<0.001$). The average C/T allelic

ratio was 1.2 (Figure 7.4). This confirms findings previously observed in CH8, SW872 and MG63 cell lines [272].

Over expression of Sp1 decreased the promoter activity of both C and T allele constructs, with a significant repressive effect on the T allele ($p<0.05$; Figure 4A), significantly increasing the C/T ratio to 1.38 ($p<0.01$). Over expression of Sp3 decreased the promoter activity of both the C and T allele constructs, and this effect was significant with the T allele construct ($p<0.001$; Figure 4B) significantly increasing the allelic ratio to 1.48 ($p<0.001$). P15 over expression decreased the promoter activity of both alleles, however, this repressive effect was not significant (Figure 4C). Finally, DEAF-1 over expression significantly repressed the promoter activity of both alleles (C and T alleles $p<0.001$; Figure 4D), but most notably repressed the T allele construct, decreasing its activity to near that of the empty control and significantly increasing the allelic ratio to 1.37 ($p<0.01$). These results confirm that all four factors repress *GDF5* expression, and that for Sp1, Sp3 and DEAF-1 this repression is greater for the T allele of rs143383.

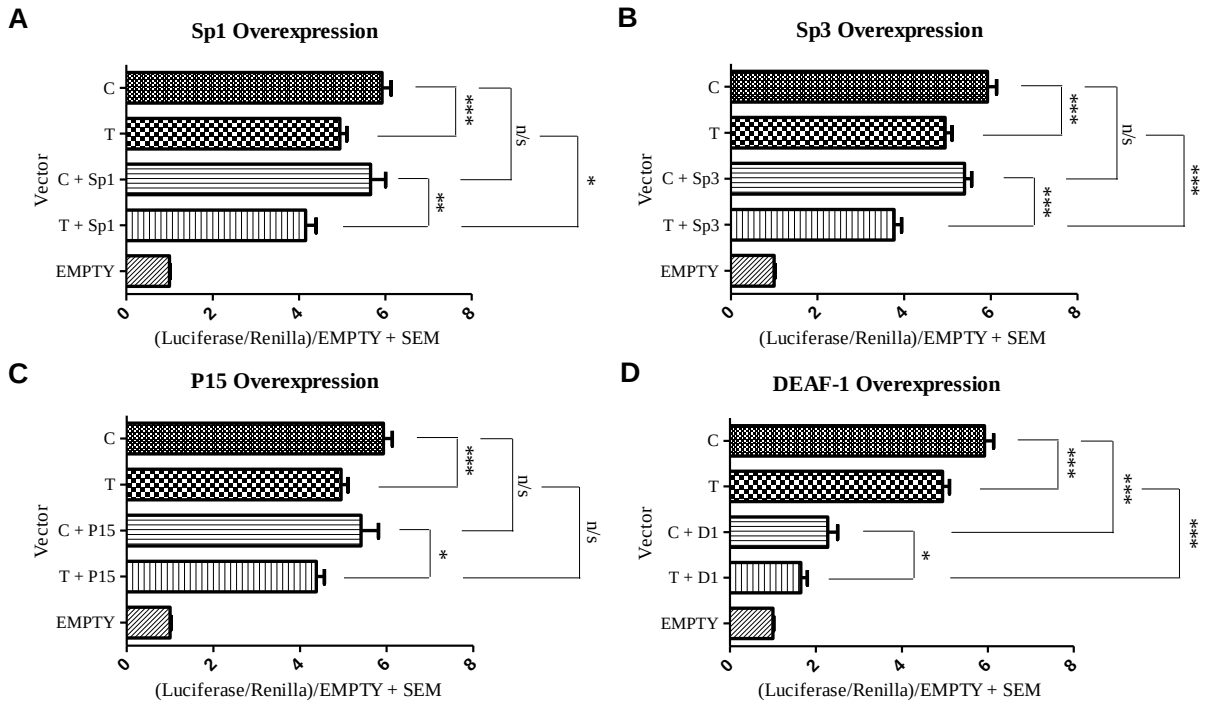


Figure 7.4

Over expression of the Sp1, Sp3, P15 and DEAF-1 (D1) proteins. **(A-D)** Promoter activity of the C and T *GDF5* luciferase vectors is shown relative to *Renilla*. Values are normalised to the luciferase levels of the EGFP/pGL3 empty vector (EMPTY). Promoter luciferase levels of both C and T allele vectors are shown in addition to the empty EGFP vector (C and T) and following over expression of **(A)** Sp1 (C+Sp1 and T+Sp1), **(B)** Sp3 (C+Sp3 and T+Sp3), **(C)** P15 (C+P15 and T+P15) and **(D)** DEAF-1 (C+D1 and T+D1). Error bars denote the standard error of the mean (SEM). *p<0.05, **p<0.01, ***p<0.001, n/s = not significant, calculated using a Student's 2 tailed *t*-test. Six replicate experiments were performed, each with four technical repeats.

7.4 Discussion

Following the identification of the binding of Sp1, Sp3 and P15 to *GDF5 in vitro*, I have confirmed the binding of these three factors *in vivo* in SW872 cells using ChIP. *GDF5* DNA was enriched following ChIP with antibodies targeting these three factors. In order to accurately assess the binding of each of these factors to *GDF5 in vivo*, the next stage would be to use Real Time PCR to quantitatively investigate the fold enrichment of *GDF5* over the negative control antibody. Using Real Time PCR would enable a more accurate assessment of the binding of each factor relative to one another. Furthermore, the identification of a region that is negative for the binding of each of these factors will confirm that the binding at this locus is specific.

The depletion of Sp1, Sp3 and P15 resulted in small, non significant increases in the overall expression of *GDF5*. Upon DEAF-1 knockdown, the overall expression of *GDF5* was significantly increased. The depletion of Sp1, Sp3 and DEAF-1 increased the DAE, with the DEAF-1 depletion significantly increasing the C/T allelic ratio. The depletion of P15 increased the expression of both the C and T alleles relative to the NTsiRNA but did not alter the C/T allelic ratio. In chapter 5 (Figure 5.6), I have previously depleted *P15* using siRNA and I observed an increase in the overall expression of *GDF5* and a decrease in the C/T allelic ratio from 1.5 in the control (NTsiRNA) to 1.26 (P15 siRNA). This previous experiment was based on data from one experiment, with three technical repeats. The experiment completed in this chapter is based on three independent experiments, each also with three technical repeats. As the differential repressive effect observed in Chapter 5 following P15 depletion has not been replicated here with a larger number of samples, I conclude from the depletion experiments that P15 is not repressing the C and T alleles differentially.

Overall, the siRNA knockdown experiments suggest that all four factors are repressing the expression of *GDF5*. Sp1, Sp3 and DEAF-1 are also modulating the DAE observed. DEAF-1 appears to be repressing the expression of *GDF5* most significantly. However, although the mRNA expression of *DEAF-1* was decreased following DEAF-1 siRNA treatment, I cannot confirm the depletion of DEAF-1 at the protein level. The characterisation studies that I performed in Chapter 6 did however suggest that the DEAF-1 siRNA is able to deplete DEAF-1 protein and I can therefore deduce from these experiments that DEAF-1 protein has been decreased.

The difference observed in the promoter activity between the C and T alleles is comparable with that reported previously [272], where the presence of a T allele at rs143383 was found to mediate a reduction in luciferase activity relative to a C allele. Here I examined the extent to which each factor is modulating the allelic expression of *GDF5* using over expression experiments. These confirmed that each factor is repressing *GDF5* expression, supporting the findings from my siRNA depletion studies. Furthermore, Sp1, Sp3 and DEAF-1 showed a greater repressive effect on the T allele promoter construct compared to the C allele promoter construct.

P15 over expression repressed the two alleles to a minor extent and its depletion led to a small and non significant increase in the expression of *GDF5*. P15 may therefore only be mediating a modest repressive effect and be binding as part of a larger repressive complex. When depleted, the other factors that form the complex are able to continue to differentially repress *GDF5* expression.

The over expression studies suggest that Sp1, Sp3 and DEAF-1 repress the T allele to a greater extent than the C allele. Therefore, upon depletion of these factors, a greater increase in the expression of the T allele, relative to the C allele, would be expected. This would cause the C/T allelic ratio to decrease. Conversely, I observed an increase in the C/T allelic ratio upon Sp1, Sp3 and DEAF-1 knockdown. For each factor, the expression of both alleles increased, however the C allele increased to a greater extent relative to the T allele. This may be a result of the incomplete depletion of the protein of each factor. If each factor represses the T allele to a greater extent and there is still protein present following siRNA knockdown, then any remaining protein would be expected to bind to the T allele more avidly, which could therefore explain the larger increase in C allele expression that was observed.

Overall, I have confirmed the binding of Sp1, Sp3 and P15 *in vivo* to *GDF5*. Furthermore, Sp1, Sp3, P15 and DEAF-1 have been identified as repressors of *GDF5*, with Sp1, Sp3 and DEAF-1 differentially modulating the expression of the two rs143383 *GDF5* alleles. The next chapter will investigate further the role of each factor using combination over expression and knockdown experiments.

Chapter 8: The Interaction of Sp1, Sp3, P15 and DEAF-1 at rs143383

8.1 Introduction

In Chapter 7, I investigated the effects that the *trans*-acting factors, Sp1, Sp3, P15 and DEAF-1 mediated on the overall and the allelic expression of *GDF5*. Using siRNA depletion and over expression experiments all four factors were shown to repress the expression of *GDF5*. Additionally, each factor was found to repress the T allele to a greater extent in comparison with the C allele. The degree to which each factor modulated the expression of *GDF5* varied, thus I wanted to explore further how these four factors may be interacting to mediate their repressive effects. In this chapter I have used combination knockdown and combination over expression experiments in order to establish how these factors are interacting.

As described in Chapters 6 and 7, DEAF-1 EGFP can be detected using the DEAF-1 antibody, however, I have not been able to detect endogenous DEAF-1 protein. Therefore the immunoblots for DEAF-1 will be shown following DEAF-1 depletion and over expression but depletion of the endogenous protein will not be discussed.

8.2 Aims

- To investigate the effect of depleting two factors simultaneously using siRNA on the overall and the allelic expression of *GDF5*.
- To investigate the effect of over expressing two factors simultaneously on *GDF5* C and T allele promoter activities using a luciferase reporter assay.
- To deplete each of the four factors, whilst concurrently over expressing each in different combinations to compare their individual and combined impacts on *GDF5* expression.

8.3 Results

8.3.1 Combination Knockdowns

I first optimised the transfection conditions for the depletion of two factors simultaneously; a final siRNA concentration of 100nM and 1x concentration of transfection reagent has been used previously for siRNA depletion experiments (Chapters 5, 6 and 7). In this chapter I have tested if cells could be transfected using a final siRNA concentration of 200nM and 2x concentration of transfection reagent. Figure 8.1 shows the protein levels of Sp1, Sp3, P15 and DEAF-1 following treatment with 100nM of target siRNA in addition to 100nM NTsiRNA and a concentration of 1x or 2x transfection reagent. Sp1, Sp3 and P15 protein levels were depleted following treatment with their respective siRNAs in comparison with the NTsiRNA control (200nM final concentration) (Figure 8.1 A-C). In each instance the addition of 2x concentration of the transfection reagent resulted in a greater depletion of the target protein.

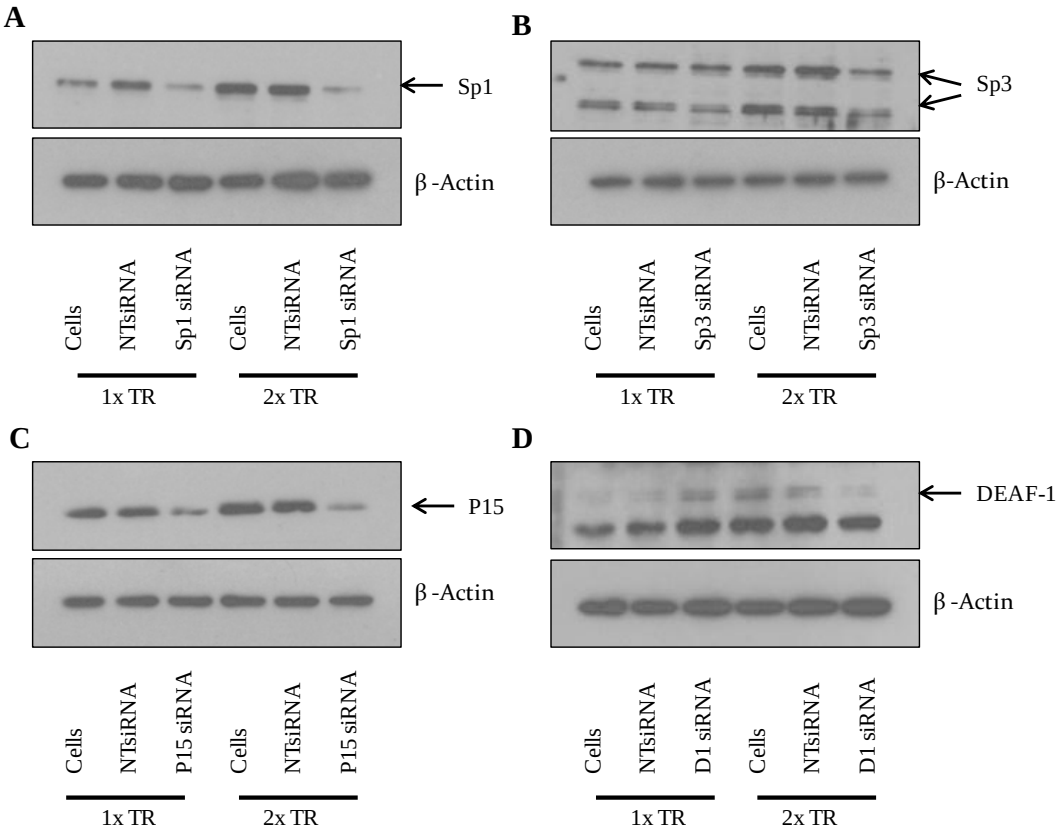


Figure 8.1

Optimisation of transfection reagent concentration for siRNA depletion. Immunoblots demonstrating Sp1, Sp3, P15 and DEAF-1 protein depletion following siRNA treatment. Cells are untreated, NTsiRNA represents treatment with a final concentration of 200nM NTsiRNA. Sp1, Sp3, P15 and DEAF-1 siRNA treatments represent cells treated with a 200nM final concentration of siRNA; 100nM of the target siRNA and 100nM NTsiRNA. Protein extracted from cells treated with the NTsiRNA control were used for basal protein expression whilst β-Actin was used as a loading control. Two different concentrations of transfection reagent were tested (1xTR and 2xTR).

I then assessed the effect of transfecting more than one siRNA at the same time and the effect of a 2x concentration of transfection reagent on the mRNA expression levels of *Sp1, Sp3, P15, DEAF-1, GDF5* and *GAPDH* (Figures 8.2 and 8.3). *Sp1* mRNA is depleted following treatment with Sp1 siRNA relative to the NTsiRNA control (Sp1 siRNA and NTsiRNA used at final concentration of 100nM) (Figure 8.2A). Additionally, *Sp1* mRNA is depleted following treatment with combinations of Sp1 siRNA with NTsiRNA, with Sp3 siRNA, with P15 siRNA and with DEAF-1 siRNA (both siRNA together representing a 200nM final siRNA concentration), relative to the 100nM NTsiRNA control but not relative to the 200nM NTsiRNA control. This reduction in *Sp1* gene expression was apparent when transfecting with both 1x and 2x transfection reagent concentrations. A reduction in *Sp1* gene expression is also apparent following transfection with 100nM NTsiRNA treated cells using 2x transfection reagent compared with using 1x transfection reagent.

The expression levels of *Sp3, P15* and *DEAF-1* follow similar patterns, each time with the target siRNA depleting expression of the target gene relative to the 100nM NTsiRNA control but not relative to the 200nM NTsiRNA control. Furthermore, 2x concentration of transfection reagent appears to deplete gene expression. These results suggest that adding a final siRNA concentration of 200nM and adding 2x concentration of transfection reagent is affecting gene expression in an inconsistent manner.

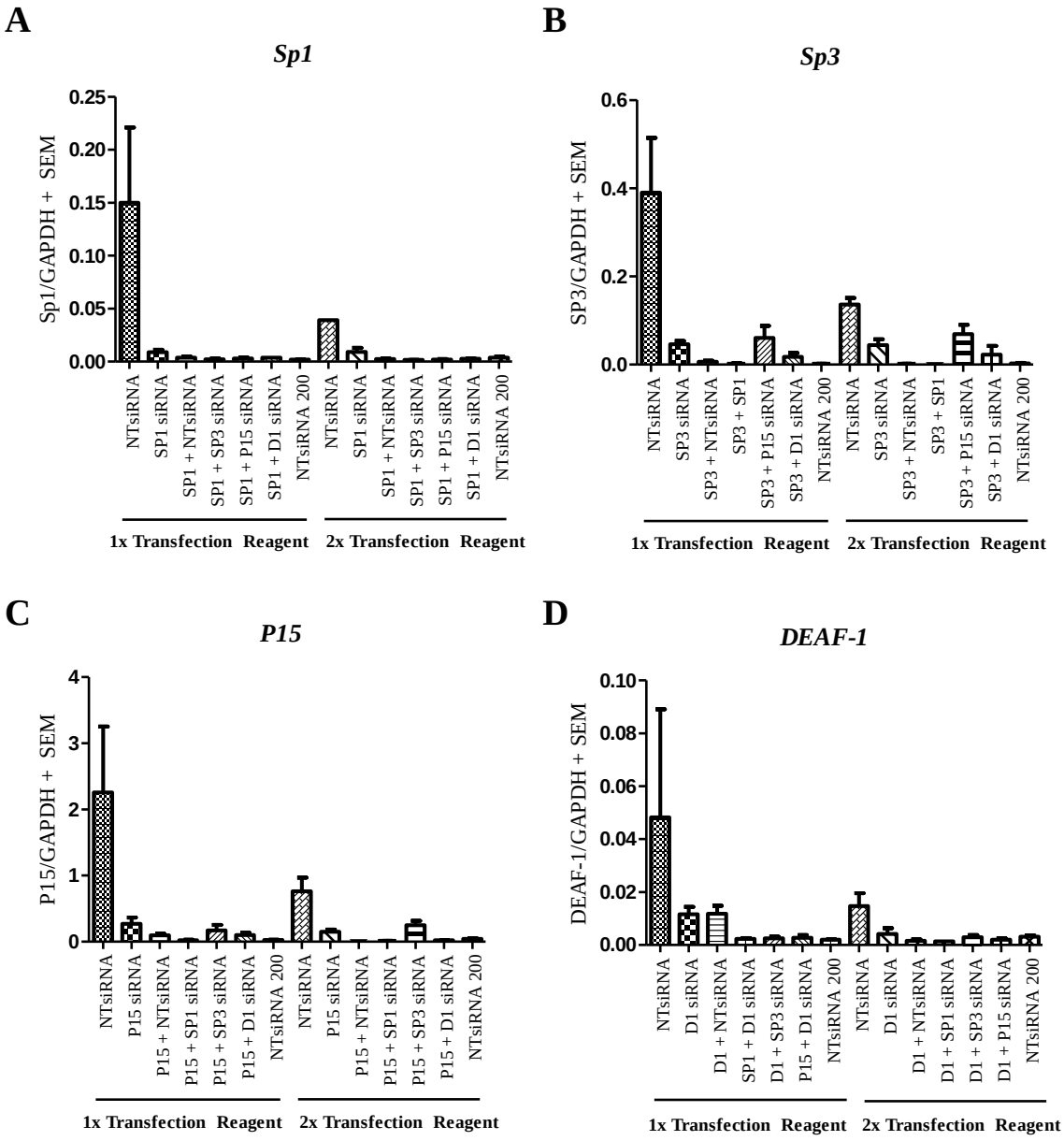


Figure 8.2

Expression of *Sp1*, *Sp3*, *P15* and *DEAF-1* following combination depletion. Expression levels of *Sp1*, *Sp3*, *P15* and *DEAF-1* mRNA are shown relative to the housekeeping gene *GAPDH*. Cells were treated with different combinations of siRNAs; Sp1 (SP1), Sp3 (SP3), P15 or DEAF-1 (D1) alone or in combination with the NTsiRNA control (SP1 + NTsiRNA, SP3 + NTsiRNA, P15 + NTsiRNA, D1 + NTsiRNA) or a second siRNA (**A:** SP1 + SP3 siRNA, SP1 + P15 siRNA, SP1 + D1 siRNA. **B:** SP3 + SP1 siRNA, SP3 + P15 siRNA, SP3 + D1 siRNA. **C:** P15 + SP1 siRNA, P15 + SP3 siRNA, P15 + D1 siRNA. **D:** D1 + SP1 siRNA, D1 + SP3 siRNA, D1 + P15 siRNA). Two different concentrations of the control NTsiRNA was transfected into cells, 100nM (NTsiRNA) and 200nM (NTsiRNA 200). 1x and 2x concentrations of transfection reagent were used. Error bars denote the standard error of the mean (SEM), the mean is calculated from three replicate experiments.

In addition to examining the expression of *Sp1, Sp3, P15* and *DEAF-1,* I also assessed the expression levels of *GDF5* and *GAPDH* following these same treatments (Figure 8.3). I have consistently observed an increase in the expression of *GDF5* following *Sp1, Sp3, P15* and *DEAF-1* depletion, shown previously in Chapter 7 (see Figure 7.2). Figure 8.3A depicts *GDF5* expression following different siRNA combinations in addition to 1x transfection reagent. For each treatment, *GDF5* expression is decreased relative to the 100nM NTsiRNA control. Following transfection with the NTsiRNA control using 2x concentration of transfection reagent, the expression of *GDF5* is decreased compared with using 1x concentration (Figure 8.3B). Sp1, Sp3 and P15 siRNA treatments increase the expression of *GDF5* relative to the 100nM NTsiRNA control using 2x concentration of transfection reagent. Additionally, Sp1 siRNA in combination with DEAF-1 siRNA results in a large increase in *GDF5* expression relative to the 100nM and 200nM NTsiRNA control.

These treatments are not producing consistent results, with decreases in the expression of genes being observed on the addition of a 200nM concentration of siRNA or a 2x concentration of transfection reagent. The expression levels of *GAPDH* were also examined and are shown in Figures 8.3C and D. The expression of *GAPDH* is very low in cells treated with 100nM of siRNAs targeting Sp1, Sp3, P15, DEAF-1 and a NTsiRNA control using 1x transfection reagent (Figure 8.3C). Cells treated with a 200nM final concentration of siRNA have increased expression levels of *GAPDH* but the expression levels are inconsistent. A similar pattern of expression is observed in the cells transfected using 2x transfection reagent (Figure 8.3D).

Overall, the results obtained from this optimisation are not reliable and suggest that changing the concentration of siRNA or transfection reagent will affect the gene expression. The expression levels of *Sp1, Sp3, P15* and *DEAF-1* were decreased following combination siRNA treatments, however the *GDF5* and *GAPDH* expression levels were not consistent following these treatments.

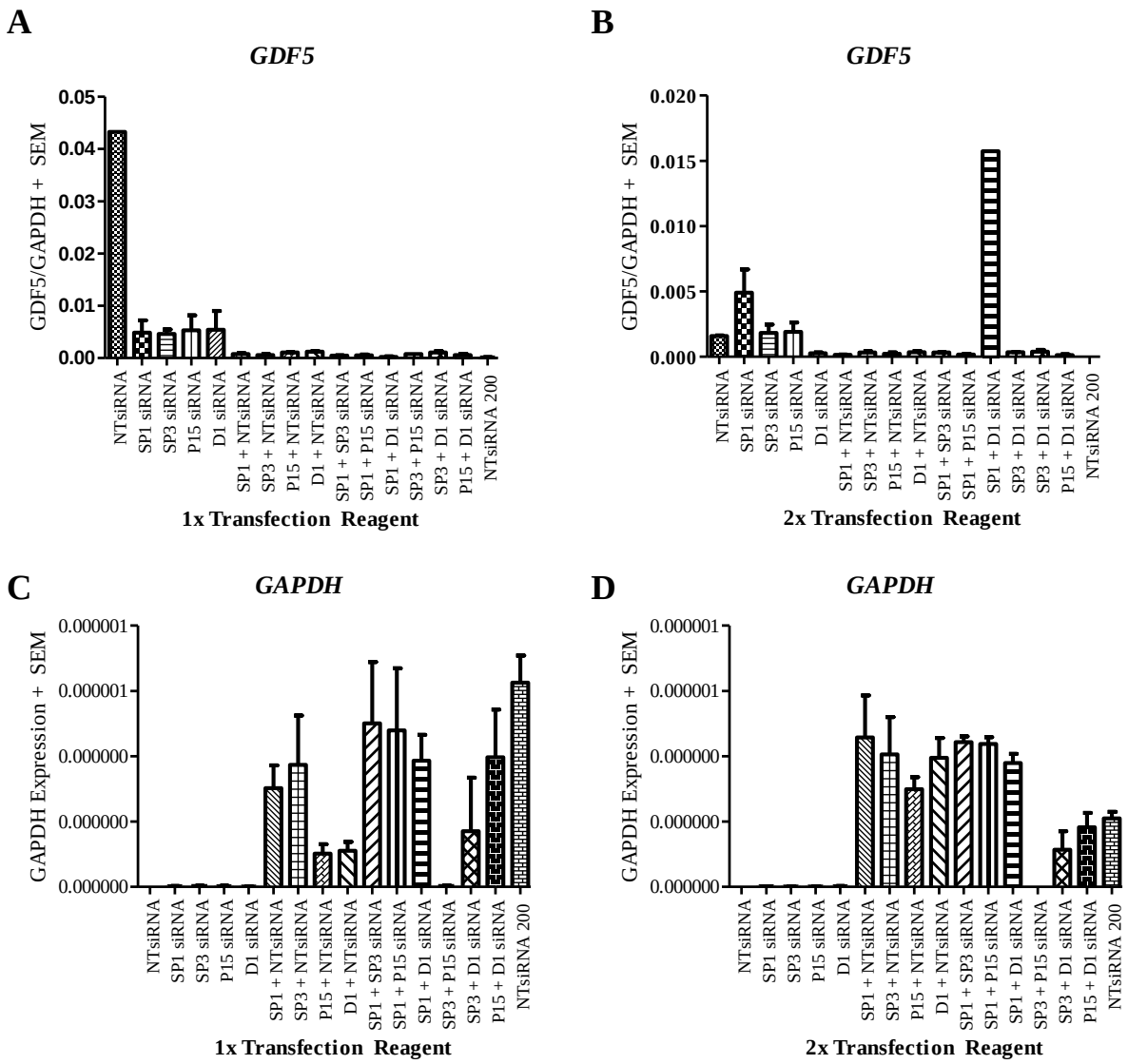


Figure 8.3

Expression of *GDF5* and *GAPDH* following combination depletion. Expression levels of *GDF5* are shown relative to the housekeeping gene *GAPDH* (Figure 8.3A and B). *GAPDH* expression is shown (Figure 8.3C and D). Cells were treated with different combinations of siRNAs; Sp1 (SP1), Sp3 (SP3), P15 or DEAF-1 (D1) alone or in combination with the NTsiRNA control (SP1 + NTsiRNA, SP3 + NTsiRNA, P15 + NTsiRNA, D1 + NTsiRNA) or a second siRNA (**A:** SP1 + SP3 siRNA, SP1 + P15 siRNA, SP1 + D1 siRNA. **B:** SP3 + SP1 siRNA, SP3 + P15 siRNA, SP3 + D1 siRNA. **C:** P15 + SP1 siRNA, P15 + SP3 siRNA, P15 + D1 siRNA. **D:** D1 + SP1 siRNA, D1 + SP3 siRNA, D1 + P15 siRNA). Two different concentrations of the control NTsiRNA was transfected into cells, 100nM (NTsiRNA) and 200nM (NTsiRNA 200). 1x and 2x concentrations of transfection reagent were used. Error bars denote the standard error of the mean (SEM), the mean is calculated from three replicate experiments.

Results from Figure 8.1 suggested that when transfecting 100nM of target siRNA in addition to 100nM of NTsiRNA using 1x or 2x concentrations of transfection reagent the target protein was successfully depleted. The gene expression analysis however seems to suggest that greater concentrations of siRNA and a 2x concentration of transfection reagent may be affecting the gene expression. I therefore further assessed the protein levels of Sp1, Sp3, P15 and DEAF-1 following treatment with either 200nM NTsiRNA or 100nM of each target siRNA (Sp1, Sp3, P15 and DEAF-1) in addition to 100nM NTsiRNA or combinations of two target siRNAs (Figure 8.4). Sp1 protein is decreased following treatment with Sp1 siRNA in addition to NTsiRNA in comparison with the NTsiRNA control. There is no protein present in the Sp1 and Sp3 siRNA combination treatment, indicating a problem has occurred. Sp1 was decreased to a minor extent following Sp1 and P15 siRNA treatment and was not depleted following Sp1 and DEAF-1 siRNA combination treatment (Figure 8.4A).

Sp3 expression is decreased to a minor extent following treatment with Sp3 siRNA, Sp3 siRNA in addition to P15 siRNA and Sp3 siRNA in addition to DEAF-1 siRNA in comparison with the NTsiRNA control. Sp3 + Sp1 siRNA represents the same sample as in A (Figure 8.4B). The protein levels of P15 are unaffected following treatment with the siRNAs relative to the NTsiRNA control (Figure 8.4C).

The results from this combination knockdown experiment are inconsistent and further optimisations would therefore be needed to progress this experiment. Due to time constraints, I was unable to optimise this combination knockdown further.

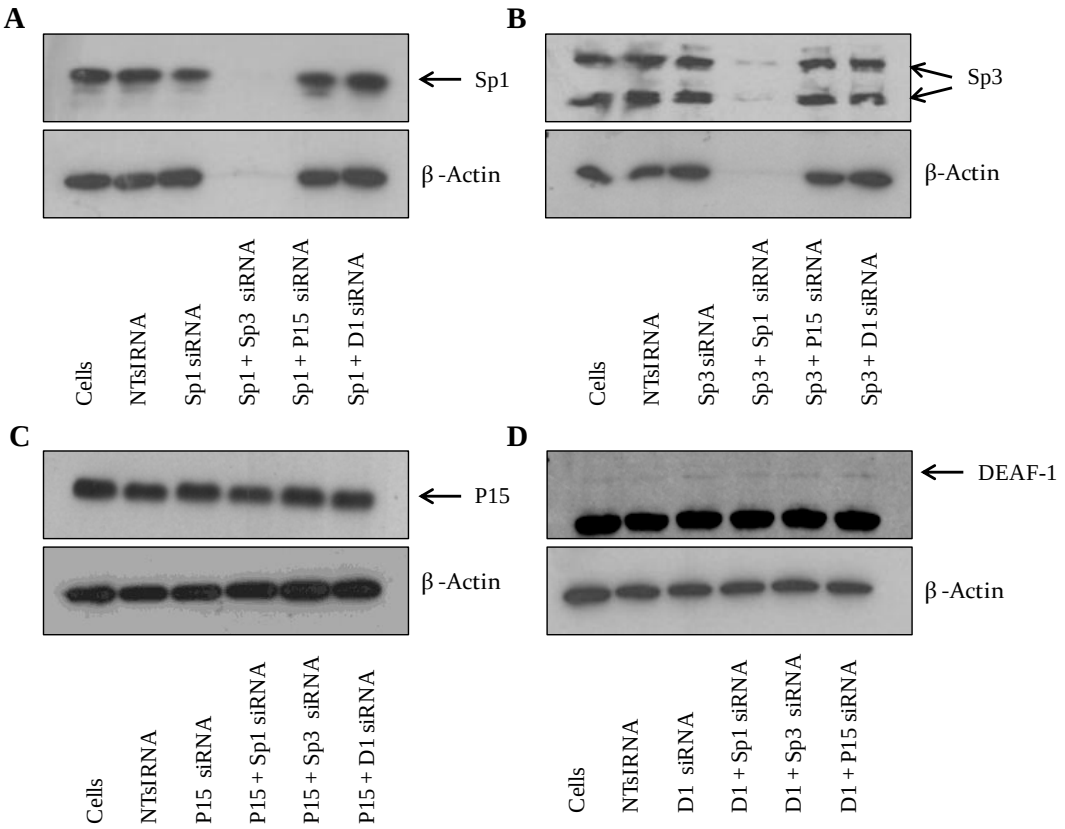


Figure 8.4

Immunoblots demonstrating Sp1, Sp3, P15 and DEAF-1 protein depletion following combination siRNA treatment. Protein extracted from cells treated with the NTsiRNA control were used for basal protein expression whilst β-Actin was used as a loading control. Cells are untreated (Cells) or transfected with a final concentration of 200nM NTsiRNA (NTsiRNA), 100nM of target siRNA plus 100nM of NTsiRNA (Sp1, Sp3, P15, DEAF-1 siRNA) or 100nM of target siRNA one and target siRNA two using a 2x final concentration of transfection reagent.

8.3.2 Combination Over expression

Following the single transfection of the *trans*-acting factor expression plasmids in Chapter 7, I assessed whether the repressive effects observed would be stronger if the factors were co-transfected and over expressed together. I first investigated if it was possible using the same experimental conditions to over express two of these factors simultaneously. The immunoblots following the combination over expressions are shown in Figure 8.5. 500ng of each plasmid was transfected into cells as with the single transfection experiments, detailed in Materials and Methods section 2.16. Each factor has been successfully over expressed alone, and in combination (Figure 8.5 A-D).

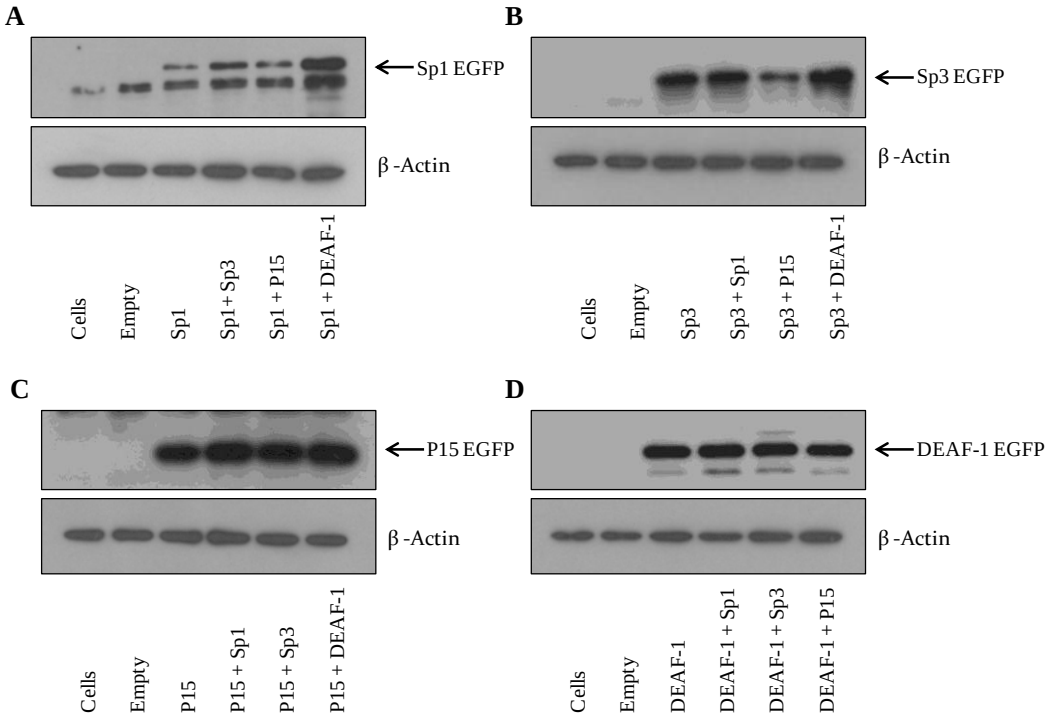


Figure 8.5

Combination over expression of the *trans*-acting factors. Immunoblots demonstrating Sp1, Sp3, P15 and DEAF-1 protein levels following over expression compared with the empty vector control (Empty). Cells are untreated or transfected with 500μg of empty, Sp1, Sp3, P15 or DEAF-1 EGFP vectors, or combinations of each. β-Actin was used as a loading control and the arrows indicate over expressed protein levels.

When Sp1 and Sp3 were jointly over expressed there was a significantly greater reduction in expression of both the rs143383 C and the T alleles relative to when they were over expressed alone (Figure 8.6A). Furthermore, the C/T allelic ratio significantly increased from 1.38 for the Sp1 over expression and 1.48 for the Sp3 over expression to 1.70 for the joint over expression ($p<0.001$ for the joint over expression versus Sp1 alone and $p<0.05$ for the joint over expression versus Sp3 alone; Table 8.1).

When Sp1 and P15 were jointly over expressed, there was a reduction in the expression of both the C and T alleles, relative to when they were over expressed alone, this repressive effect was significant for the T allele when compared to P15 over expression alone (Figure 8.6B, $p<0.001$ versus P15 alone). The C/T allelic ratio was not altered significantly for joint over expression ($p=0.79$ versus Sp1 alone and $p=0.06$ versus P15 alone; Table 8.1).

When Sp1 and DEAF-1 were jointly over expressed there was a reduction in expression of both the C and T alleles relative to when they were over expressed alone (Figure 8.6C). The C/T allelic ratios increased significantly from 1.38 for Sp1 and 1.37 for DEAF-1 to 1.55 for the joint over expression ($p<0.001$ versus the C/T EGFP empty vector ratio, Table 8.1). However, these C/T allelic ratio changes were not significant when compared with Sp1 or DEAF-1 over expression alone ($p = 0.1$).

When Sp3 and P15 were jointly over expressed there was an increase in the expression of the C allele when compared with Sp3 and P15 over expression alone. Conversely, the expression of the T allele was decreased when Sp3 and P15 were jointly over expressed when compared with Sp3 and P15 over expression alone (Figure 8.6B). Furthermore, the C/T allelic ratio significantly increased from 1.48 for the Sp3 over expression and 1.24 for P15 over expression to 1.59 for the joint over expression ($p=0.22$ versus Sp3 alone, $p<0.01$ versus P15 alone, Table 8.1).

When Sp3 and DEAF-1 were jointly over expressed, the expression of both C and T alleles was reduced relative to Sp3 over expression alone, but increased relative to DEAF-1 over expression alone (Figure 8.6C). The C/T allelic ratios also increased from 1.48 for Sp3 and 1.37 for DEAF-1 to 1.6 for the joint over expression, and this was a significant C/T difference compared to DEAF-1 over expression alone ($p = 0.01$, Table 8.1).

Finally, when P15 and DEAF-1 were over expressed the expression of both the C and T alleles was significantly reduced relative to P15 over expression alone ($p<0.001$, Figure 8.6B), but increased relative to DEAF-1 over expression alone. The C/T allelic ratio increased to 1.37 for joint over expression from 1.24 for P15 over expression and remained unchanged from 1.37 for DEAF-1 over expression alone ($p=0.45$ versus P15 alone, $p=0.99$ versus D1 alone, Table 8.1).

Overall these results suggest that Sp1 and DEAF-1 together are mediating the largest repressive effects on the expression of the C and T alleles. Sp3 appears to be mediating the largest differential effect; when over expressed in combination with Sp1, the C/T allelic ratio increases significantly to 1.7 due to the T allele being significantly more repressed than the C allele. This is also the case when Sp3 is over expressed in combination with P15 (increasing the C/T allelic ratio to 1.59) and DEAF-1 (increasing

the C/T allelic ratio to 1.6). P15 on the other hand does not appear to mediate any additional repressive effects when over expressed in combination with any of the other three factors compared to the over expression of each of these factors alone.

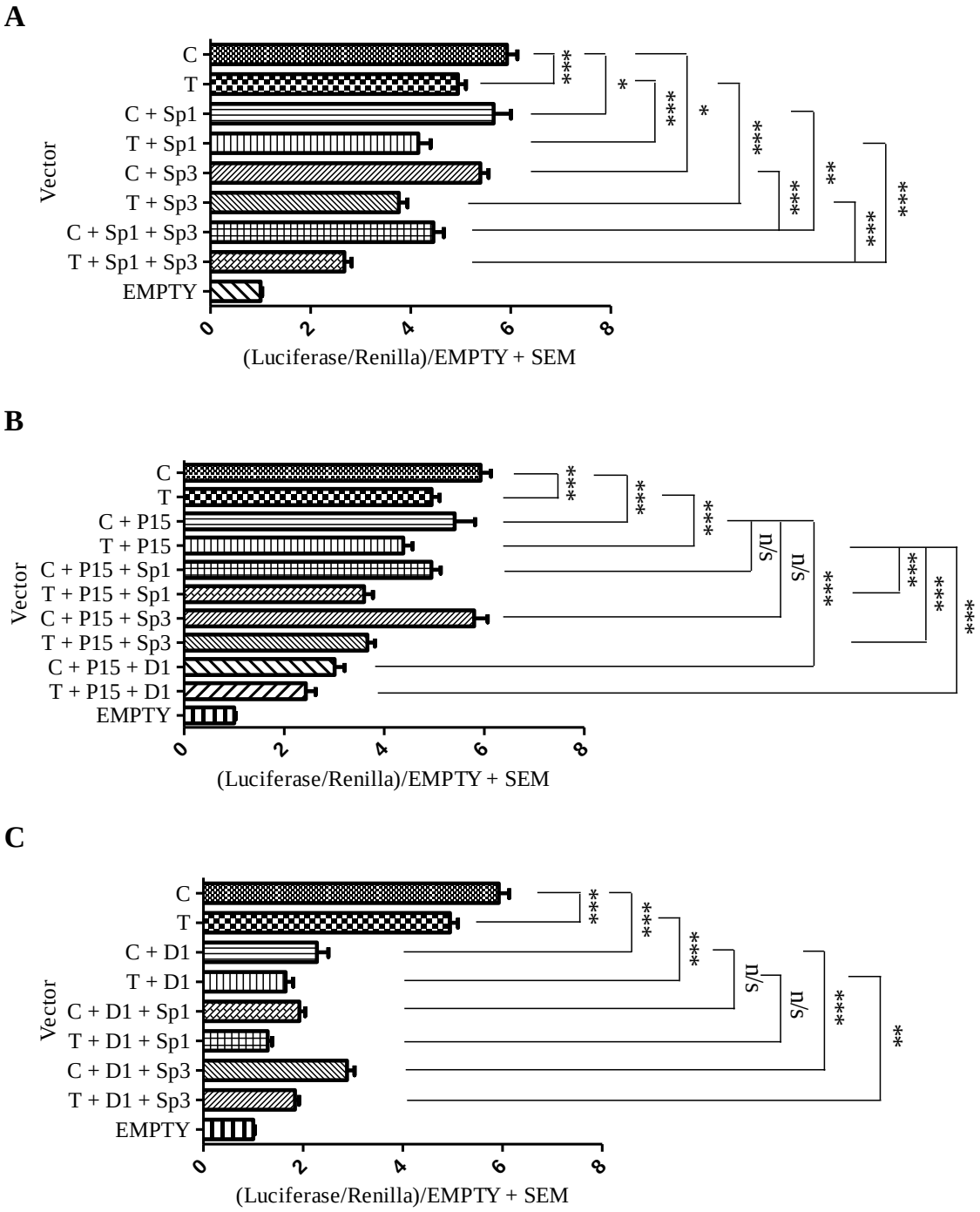


Figure 8.6

The effect of *trans*-acting factor over expression on *GDF5* promoter activity. Promoter activity of the C and T *GDF5* luciferase vectors is shown relative to *Renilla*. Values are normalised to the luciferase levels of the EGFP/pGL3 empty vector (EMPTY). (**A**) Promoter luciferase levels of both C and T allele vectors are shown in addition to the empty EGFP vector (C and T) and following over expression of Sp1 alone (C+Sp1 and T+Sp1), Sp3 alone (C+Sp3 and T+Sp3) and Sp1 and Sp3 in combination (C+Sp1+Sp3 and T+Sp1+Sp3). (**B**) Promoter luciferase levels of both C and T allele vectors are shown in addition to the empty EGFP vector (C and T) and following over expression of P15 alone (C+P15 and T+P15) and P15 in combination with Sp1 (C+P15+Sp1 and T+P15+Sp1), Sp3 (C+P15+Sp3 and T+P15+Sp3) and DEAF-1 (C+P15+DEAF-1 and T+P15+DEAF-1). **(C)** Promoter luciferase levels of both C and T allele vectors are shown in addition to the empty EGFP vector (C and T) and following over expression

of DEAF-1 alone (C+D1 and T+D1), and DEAF-1 in combination with Sp1 (C+D1+Sp1 and T+D1+Sp1) and in combination with Sp3 (C+D1+Sp3 and T+D1+Sp3). Error bars denote the standard error of the mean (SEM). *$p<0.05$ ***$p<0.001$ n/s = not significant, calculated using a Student's 2 tailed *t*-test. Three replicate experiments were performed, each with an n of 4.

Vector	C/T Ratio	P Value
C/T	1.2	
C/T + Sp1	1.38	0.005
C/T + Sp3	1.48	<0.0001
C/T + P15	1.24	0.56
C/T + D1	1.37	0.006
C/T + Sp1 + Sp3	1.70	<0.0001 (vs. C/T) <0.001 (vs. + Sp1) 0.03 (vs. + Sp3)
C/T + Sp1 + P15	1.4	0.003 (vs. C/T) 0.79 (vs. + Sp1) 0.06 (vs. + P15)
C/T + Sp1 + D1	1.55	<0.0001 (vs. C/T) 0.1 (vs. + Sp1) 0.1 (vs. + D1)
C/T + Sp3 + P15	1.59	<0.0001 (vs. C/T) 0.22 (vs. + Sp3) 0.0002 (vs. + P15)
C/T + Sp3 + D1	1.6	<0.0001 (vs. C/T) 0.2 (vs. + Sp3) 0.01 (vs. + D1)
C/T + P15 + D1	1.37	0.09 (vs. C/T) 0.45 (vs. + P15) 0.99 (vs. + D1)

Table 8.1

The C/T allelic ratios following over expression of the *trans*-acting factors. The promoter activities of the C and T *GDF5* luciferase vectors were compared to derive C/T ratios, which are shown for the empty vectors (C/T) and for when these vectors were co-transfected in combination with Sp1 (C/T + Sp1), Sp3 (C/T + Sp3), P15 (C/T + P15), DEAF-1 (C/T + D1), Sp1 and Sp3 (C/T + Sp1 + Sp3), Sp1 and P15 (C/T + Sp1 + P15), Sp1 and DEAF-1 (C/T + Sp1 + DEAF-1), Sp3 and P15 (C/T + Sp3 + P15), Sp3 and DEAF-1 (C/T + Sp3 + DEAF-1) and P15 and DEAF-1 (C/T + P15 + DEAF-1). P-values were calculated using a Student's 2 tailed *t*-test comparing the allelic ratios of each treatment group to either C/T, C/T + Sp1 (+Sp1), C/T + Sp3 (+Sp3), C/T + P15 (+P15) or C/T + D1 (+D1). Three replicate experiments were performed, each with an n of 4.

8.3.3 siRNA Knockdown and Over expression

In addition to depleting two factors in combination and over expressing two factors in combination, I wanted to optimise a technique that would enable the depletion of one factor and the over expression of another simultaneously. Dhamafect Duo is a transfection reagent that enables the transfection of siRNA and plasmid vectors concurrently. I first tested if this transfection reagent could be used to deplete protein using siRNA or over express protein by transfecting EGFP tagged over expression vectors into SW1353 cells. I tested this protocol using Sp3 siRNA and Sp3 EGFP, Materials and Methods 2.16.

The first four lanes of Figure 8.7A show the protein levels of Sp3 following no treatment (Cells), treatment with transfection reagent alone (DF Duo), NTsiRNA or Sp3 siRNA. Sp3 protein levels were depleted following treatment with Sp3 siRNA in comparison with the NTsiRNA control. Figure 8.7B depicts the protein levels of Sp3 following over expression with 1µg of Sp3 EGFP, 2µg of Sp3 EGFP or 1µg of Sp3 EGFP vector in addition to 1µg of empty EGFP vector. Sp3 protein was over expressed in all three treatments in comparison with the empty EGFP vector control. These results confirm that I am able to deplete or over express (up to 2µg) Sp3 protein using this transfection reagent. Lanes 5-9 of Figure 8.7A show Sp3 protein levels following the over expression of either the empty vector or Sp3 EGFP in addition to transfection with the NTsiRNA control or Sp3 siRNA. Sp3 protein was depleted following treatment with Sp3 siRNA in comparison with NTsiRNA when co transfected with 1µg of empty vector. Furthermore, Sp3 was over expressed when transfected into cells with NTsiRNA and this over expressed protein was depleted upon co-transfection with Sp3 siRNA. These results suggest this transfection reagent can be used to deplete and over express proteins concurrently.

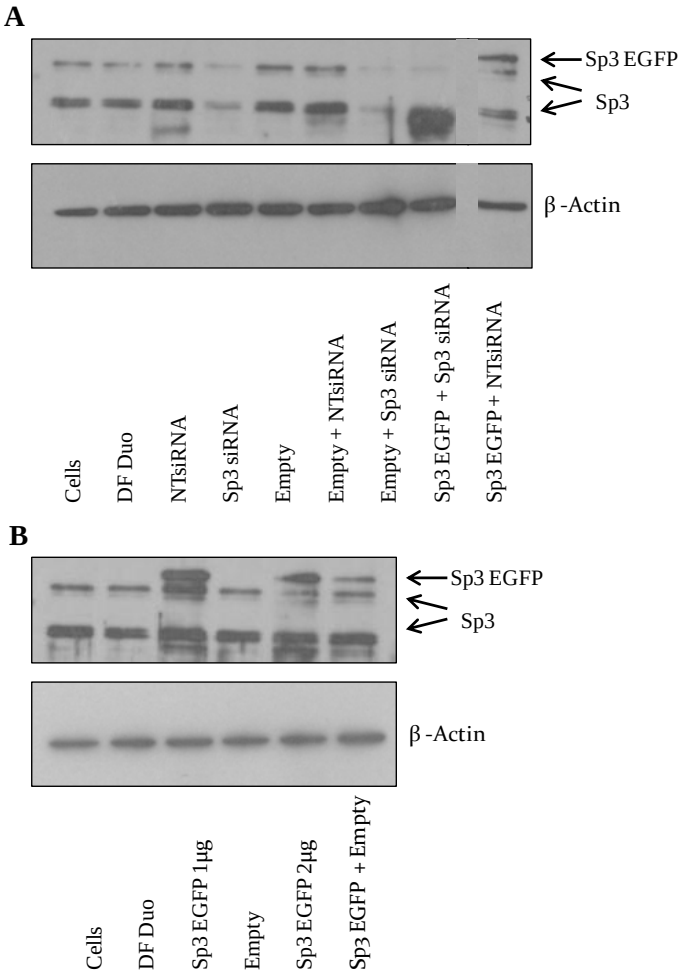


Figure 8.7

Immunoblots showing Sp3 protein levels following an optimisation experiment using Dharmafect Duo transfection reagent. Cells are untreated protein samples, DF Duo represents treatment of cells with transfection reagent alone. **(A)** Cells were transfected with NTsiRNA, Sp3 siRNA, empty vector, empty vector in addition to NTsiRNA (Empty + NTsiRNA) or Sp3 siRNA (Empty + Sp3 siRNA) or Sp3 EGFP vector in addition to Sp3 siRNA (Sp3 EGFP + Sp3 siRNA) or NTsiRNA (Sp3 EGFP + NTsiRNA). **(B)** Cells were transfected with either 1µg or 2µg of Sp3 EGFP vector (Sp3 EGFP 1µg or Sp3 EGFP 2µg), 1µg of empty vector (Empty) or 1µg of empty vector in addition to 1µg of Sp3 EGFP (Sp3 EGFP + Empty). β-Actin was used as a loading control. The arrows indicate endogenous and over expressed protein levels.

I then further optimised this experiment by testing different volumes of the Dharmafect Duo transfection reagent. Figure 8.8A shows Sp3 protein levels following the over expression of the Sp3 EGFP vector using either 4 or 6µl of transfection reagent. The expression level of Sp3 EGFP protein was not further increased by adding a greater volume of transfection reagent.

Next, I tested if DEAF-1 protein could be over expressed by transfecting 1µg of DEAF-1 EGFP vector in addition to 1µg empty vector using the Dharmafect Duo transfection reagent. DEAF-1 EGFP was successfully over expressed (Figure 8.8B).

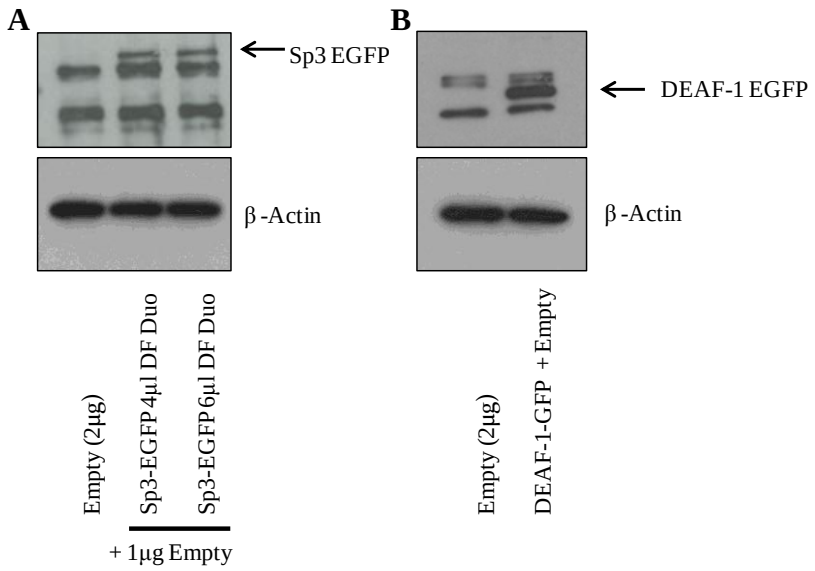


Figure 8.8
Immunoblots showing further optimisation of Sp3 over expression and the optimisation of DEAF-1 over expression using Dharmafect Duo transfection reagent. **(A)** Cells were transfected with 2µg of empty vector using 4µl of transfection reagent or 1µg of Sp3 EGFP in addition to 1µg of empty vector using two different volumes, 4µl and 6µl of transfection reagent. **(B)** Cells were transfected with either 2µg of empty vector or 1µg of DEAF-1 EGFP in addition to 1µg of empty vector. β-Actin was used as a loading control. The arrows indicate over expressed protein levels.

Following the successful over expression and concurrent knockdown of Sp3, I then assessed if Sp1, Sp3, P15 and DEAF-1 protein expression could be depleted following siRNA transfection, whilst concurrently over expressing either empty EGFP vector, Sp3 EGFP or DEAF-1 EGFP vectors (Figure 8.9A-D). Sp3 was depleted following treatment with Sp3 siRNA in comparison with the NTsiRNA control when concurrently transfected with the empty EGFP or DEAF-1 EGFP vectors (Figure 8.9A). Sp1 and P15 protein were depleted following treatment with Sp1 and P15 siRNA respectively when concurrently transfected with empty EGFP, Sp3 EGFP or DEAF-1 EGFP vectors

(Figure 8.9C and D). The successful over expression of Sp3 EGFP and DEAF-1 EGFP for this experiment is shown in Figures 8.10A and B.

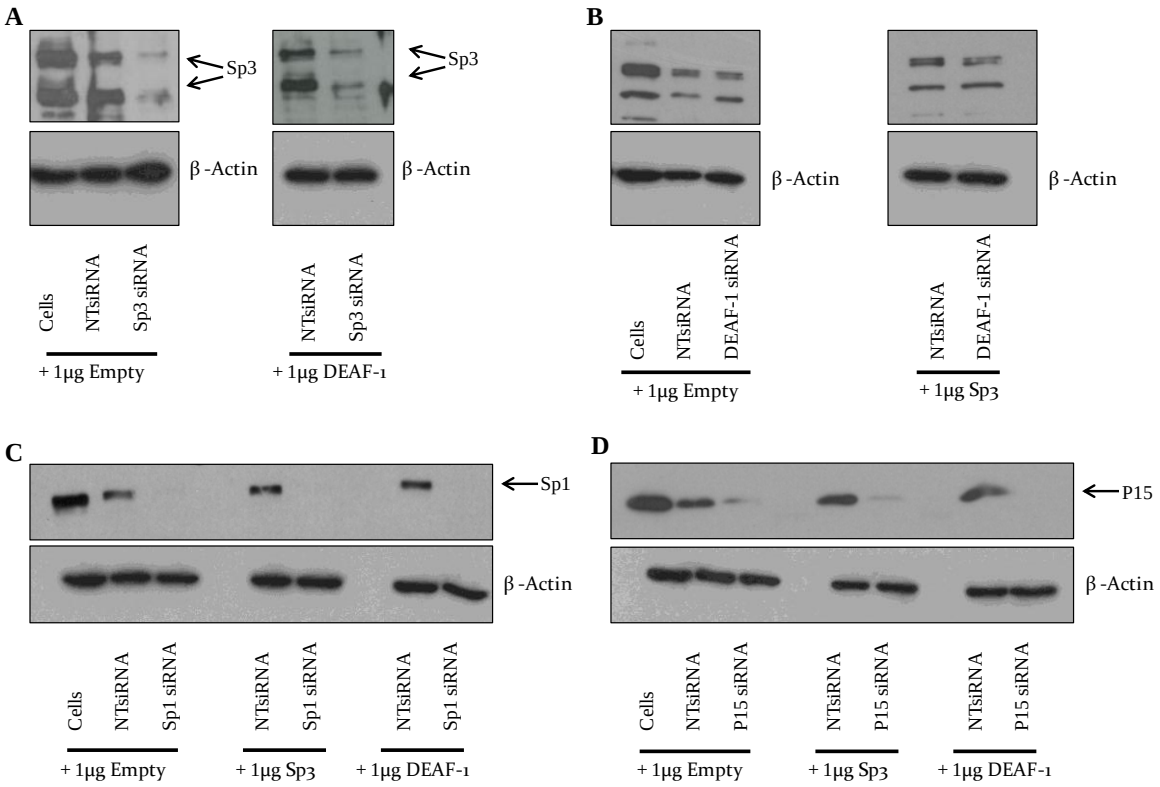


Figure 8.9

Concurrent over expression and depletion studies: the assessment of depletion. Immunoblots showing the depletion of Sp1, Sp3, P15 and DEAF-1 upon transfection with siRNA in addition to Sp3 EGFP and DEAF-1 EGFP over expression vectors using Dharmafect Duo transfection reagent. Cells are untreated protein samples **(A)** Sp3 protein levels following transfection with 1µg of empty vector or 1µg of DEAF-1 EGFP in addition to NTsiRNA or Sp3 siRNA. **(B)** DEAF-1 protein levels following transfection with 1µg of empty vector or 1µg of Sp3 EGFP in addition to NTsiRNA or DEAF-1 siRNA. **(C)** Sp1 protein levels following transfection with 1µg of empty vector, 1µg of Sp3 EGFP or 1µg DEAF-1 EGFP in addition to NTsiRNA or Sp1 siRNA. **(D)** P15 protein levels following transfection with 1µg of empty vector, 1µg of Sp3 EGFP or 1µg DEAF-1 EGFP in addition to NTsiRNA or P15 siRNA. β-Actin was used as a loading control. The arrows indicate the endogenous protein levels.

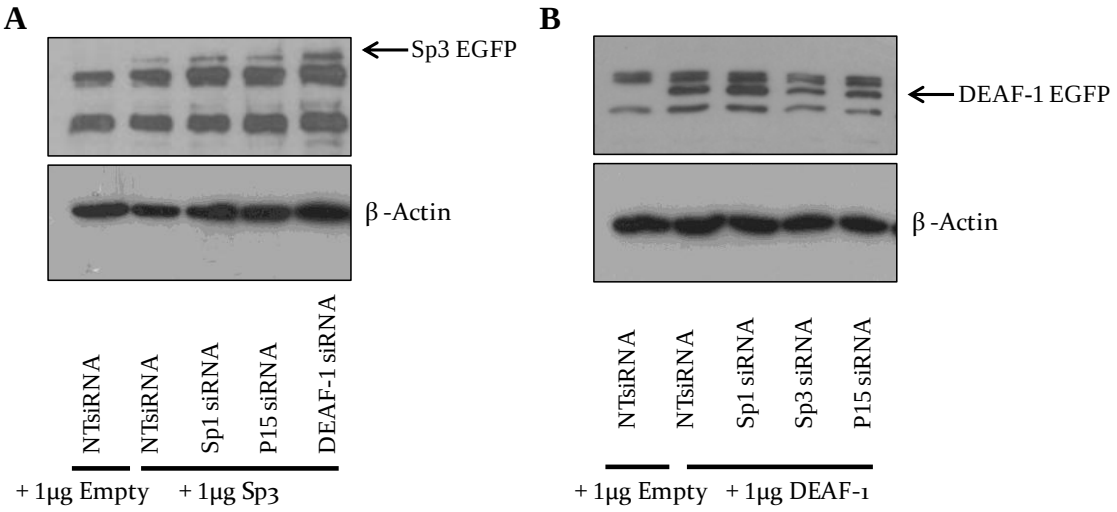


Figure 8.10
Concurrent over expression and depletion studies: the assessment of over expression. Immunoblots showing the over expression of Sp3 EGFP and DEAF-1 EGFP following transfection with siRNA in addition to Sp3 EGFP and DEAF-1 EGFP over expression vectors using Dharmafect Duo transfection reagent. **(A)** Sp3 protein levels following transfection with NTsiRNA, Sp1 siRNA, Sp3 siRNA or DEAF-1 siRNA in addition to 1µg of empty vector or 1µg of Sp3 EGFP. **(B)** DEAF-1 protein levels following transfection with NTsiRNA, Sp1 siRNA, Sp3 siRNA or DEAF-1 siRNA in addition to 1µg of empty vector or 1µg of Sp3 EGFP. β-Actin was used as a loading control. The arrows indicate the over expressed protein levels.

Following the successful concurrent transfection of siRNA and plasmid DNA using this transfection reagent, I wanted to assess if I could use this protocol to transfect cells with the *GDF5* C/T luciferase vectors, in addition to the *trans*-acting factor over expression vectors and siRNAs targeting each factor. This would enable the assessment of the effects of each factor on *GDF5* promoter activity. In Chapter 7 I had demonstrated that over expression of Sp3 alone consistently decreased the C and T allele promoter activity and had a more significant repressive effect on the T allele, decreasing the luciferase levels from 6.6 and 5.6 to 5.4 and 3.6 relative luciferase units respectively (Figure 7.4). For this new experiment, I observed that whilst over expressing Sp3, regardless of the siRNA transfected, the luciferase levels of both the C and T vectors were reduced to below 2, with each siRNA appearing to have similar repressive effects. Similarly, over expression of DEAF-1 alone (Figure 7.4) consistently decreased the C and T allele promoter activity and had a more significant repressive effect on the T allele, decreasing the luciferase levels from and 3.7 to 1.5 and 1.13 relative luciferase units respectively. As for Sp3 over expression, upon DEAF-1 over expression, regardless of the siRNA transfected, the luciferase levels of both the C and T vectors are reduced to below 1,

with each siRNA appearing to have similar repressive effects. Although a repressive effect was expected, this is greater than the effects seen previously and thus it appears to be a result of the co-transfection of the over expression vectors with the siRNA.

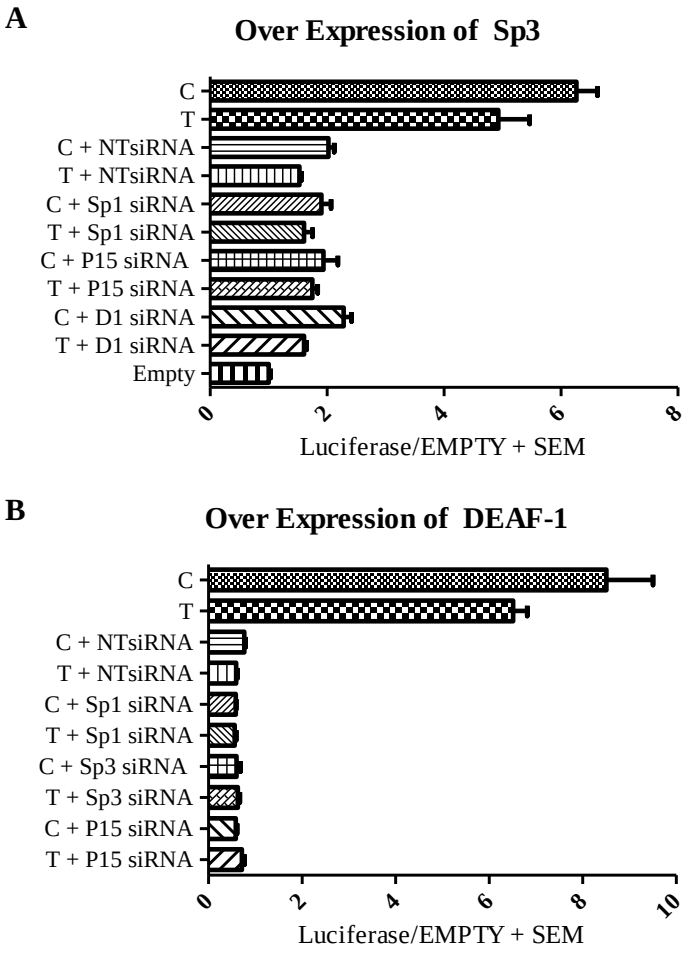


Figure 8.11

Over expression and concurrent siRNA knockdown. **(A)** Promoter activity of the C and T *GDF5* luciferase vectors is shown. Values are normalised to the luciferase levels of the EGFP/pGL3 empty vector (EMPTY). Promoter luciferase levels of both C and T allele vectors are shown in addition to the empty EGFP vector (C and T) and following over expression of Sp3 EGFP, in addition to a non targeting siRNA (+NTsiRNA), Sp1 siRNA (+Sp1), P15 siRNA (+P15) and DEAF-1 siRNA (+DEAF-1). **(B)** Promoter activity of the C and T *GDF5* luciferase vectors is shown. Values are normalised to the luciferase levels of the EGFP/pGL3 empty vector (EMPTY). Promoter luciferase levels of both C and T allele vectors are shown in addition to the empty EGFP vector (C and T) and following over expression of DEAF-1 EGFP, in addition to a non targeting siRNA (+NTsiRNA), Sp1 siRNA (+Sp1), Sp3 siRNA (+Sp3) and P15 siRNA (+P15). Error bars denote the standard error of the mean (SEM). This experiment is based on one experiment with four technical repeats.

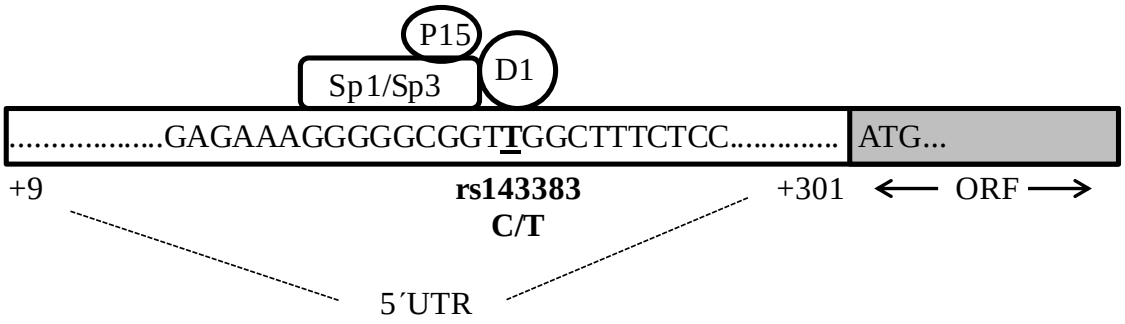


Figure 8.12

Proposed binding model of the four *trans*-acting factors to rs143383. A region (+9 to +301 relative to the transcription start site) of the *GDF5* 5′UTR is depicted, with the sequence immediately flanking rs143383 (T allele underlined) shown. We propose that DEAF-1 binds directly to rs143383 (at the TTGG site) and that Sp1 and Sp3 bind just upstream (to the Sp site GGGCGG), mediating a repressive effect through DEAF-1. P15 binds as part of the repressive multi-protein complex and may be serving as a linker between Sp1 and the general transcription machinery. ORF is the open reading frame of *GDF5* whilst ATG is the translation initiation codon.

8.4 Discussion

In Chapter 7 I used siRNA depletion and over expression studies to assess the individual effect that the four *trans*-acting factors Sp1, Sp3, P15 and DEAF-1 had on the expression of *GDF5*. These studies confirmed that all four factors act as repressor proteins and that each factor is differentially repressing the expression of the two rs143383 alleles. In this chapter I have further investigated the role of each factor using combination depletion and combination over expression experiments.

In the combination knockdown experiments, the protein levels of Sp1, Sp3 and P15 were all decreased following treatment with 100nM of their respective siRNAs. Using a 2x concentration of transfection reagent depleted the protein levels to a greater extent in comparison to using a 1x concentration. Although the protein levels of Sp1, Sp3 and P15 appeared unaffected by the increase in siRNA and transfection reagent, the gene expression levels were inconsistent. A concentration of 200nM NTsiRNA reduced the expression of *Sp1, Sp3, P15, DEAF-1* and *GDF5* in comparison with a concentration of 100nM. Furthermore the addition of 2x transfection reagent reduced the expression of *Sp1, Sp3, P15* and *DEAF-1*.

The protein levels following the transfection of two siRNAs in combination were assessed, none of the factors have been successfully depleted in comparison with the NTsiRNA when transfected in combination. Overall, it is difficult to draw accurate conclusions from the data considering the variability in gene expression and the poor depletion of protein following the transfection of more than one siRNA. Due to time constraints, the optimisation of this experiment could not be continued. Future experiments should test lower concentrations of each siRNA, for example 50nM and 25nM final concentrations to assess if protein levels of each factor could be depleted at these concentrations. If successful, this would enable the co-transfection of two or more siRNAs concurrently, without an overall increase in siRNA concentration compared with previous depletion experiments. The 2x concentration of transfection reagent may have been toxic to the cells; future experiments therefore should use a 1x concentration. If the protein of the four factors could be depleted using a final siRNA concentration of 25nM, this would also enable the concurrent transfection of up to four different siRNAs.

For the combination over expression experiments, I successfully over expressed two factors in combination using the same concentrations of each vector as used for the single transfection experiments reported on in Chapter 7. The over expression was

confirmed at the protein level, and the effect of the over expression upon *GDF5* C and T allele promoter vectors was assessed. When over expressed together, DEAF-1 and Sp1 mediated the greatest overall repressive effect whereas over expression of Sp1 and Sp3 together mediated the greatest differential allelic effect, repressing the T allele to a greater extent than the C allele. Over expression of P15 in combination with Sp1, Sp3 or DEAF-1 did not contribute any further significant repressive effects compared to over expression of the factors alone. Overall, these combination over expression experiments have enabled me to identify which factors appear to be mediating the greatest repressive effects, and has identified the combinations of factors that are repressing *GDF5* to the greatest extent. In order to fully explore the role of each factor, further studies could optimise the over expression of three factors in combination and assess the repressive effects when all four factors are over expressed concurrently.

Using Dharmafect Duo, I was able to successfully deplete protein following the transfection of an siRNA, and over express protein following the transfection of an over expression vector. I was also able to deplete the protein levels of Sp1, Sp3 and P15 whilst concurrently over expressing either the empty EGFP vector, Sp3 EGFP or DEAF-1 EGFP vectors. In principal, the over expression of Sp1 and P15 should be possible using this same technique, however further work should confirm that they can be over expressed when concurrently transfected with siRNAs.

Over expression of Sp3 and DEAF-1 in combination with a NTsiRNA reduced the luciferase levels of the *GDF5* C and T promoter vectors, beyond that of Sp3 and DEAF-1 over expression alone (without the siRNA; Chapter 7). There are several potential reasons for this and I have listed four of them below:

1. The transfection reagents used are different and therefore may have differing effects.
2. In the over expression experiments in Chapter 7, the cells were lysed 24 hours post transfection, however in the concurrent knockdown and over expression experiments reported in this chapter, the cells were lysed at 48 hours. I therefore lysed the cells for the luciferase experiment using the same time point (48 hours) which could have impacted upon the results.
3. The values in the previous over expression experiments were normalised to *renilla; renilla* was not used for this experiment because the amount of DNA

being transfected was already high and thus the transfection efficiencies have not been corrected for.

4. This effect may be caused by transfection of these vectors with the siRNA. To investigate this final reason further, I verified that the siRNAs used for this experiment did not target any regions within my *GDF5* promoter vector which could have resulted in the decrease in luciferase activity observed. The siRNA sequences however did not target any regions within this reporter vector.

Further experiments are therefore needed in order to optimise this protocol to enable the activity of the *GDF5* C and T allele luciferase constructs to be determined following the depletion and over expression of the four factors in different combinations. Optimisations such as cell number and DNA: transfection reagent ratio could be varied in order to first replicate what I have observed previously for the over expression experiments, to enable me to then assess the impact of each of the siRNAs. Due to time limitations I was unable to perform any further experiments.

Sp1 and Sp3, as previously described in Chapter 4 have been reported to interact with a number of proteins, including transcriptional activators and repressors. Sp1 and Sp3 have also been reported to interact with proteins in the basal transcription machinery; promoters that do not contain a TATA box commonly contain an Sp binding site and when bound, the Sp proteins have been reported to play a critical role in anchoring the basal transcriptional machinery to promote transcriptional initiation [358]. Sp1 facilitates the binding of TFIID through binding to TBP (TATA binding protein) associated factors (TAFs) which then recruit RNA polymerase II [359]. *GDF5* does not contain a TATA box. It appears likely therefore that in binding to the *GDF5* 5'UTR Sp1 and Sp3 may be mediating interactions with the basal transcriptional machinery to modulate transcription of the gene.

Sp1 and Sp3 have been shown *in vivo* to be associated with the GC boxes of a variety of promoters [360]. However, it is not known whether Sp1 and Sp3 bind simultaneously or separately to these sites. Some studies have reported that Sp1 and Sp3 do form a multi protein complex, however in some cell lines this has not been observed [361, 362]. Sp1 and Sp3 have been shown to co localise in the nucleus, suggesting interaction between these two proteins is possible [361].

Both Sp1 and Sp3 have been reported to repress the expression of genes previously, however when co expressed, it has been suggested that Sp3 represses gene expression by competing with Sp1 for occupancy of the GC binding site, Sp1 usually being associated with transcriptional activation [313, 362]. For example, Sp3 has been reported to repress Sp1 activation of the *COL2A1* gene in primary chondrocytes [363]. However, the depletion and over expression studies that I have performed suggest that both Sp1 and Sp3 are repressing *GDF5* expression.

As discussed in Chapter 5, P15 can also interact with a number of proteins to activate or repress gene expression [326, 329]. In addition, P15 can bind to components of the basal transcription machinery and serves as a bridge molecule between upstream transcription factors and the PIC [326]. One of the most important functions of P15 is considered to be the stabilisation of DNA-protein interactions [364]. Additionally, P15 directly interacts with the core histones H3 and H2B and induces chromatin condensation [329]. Depletion of P15 thus decompacts chromatin, upregulating the expression of a number of genes. Furthermore, Sp1 has been reported previously to interact with HDAC1 to repress gene expression [365]. Although I did not identify the binding of HDAC1 or HDAC2 to the *GDF5* probes by EMSA in Chapter 4, it is possible that P15 and Sp1 mediate their transcriptional repression by recruiting histone modifying proteins such as HDACs which may deacetylate histones and thus repress gene expression.

P15 and Sp1 have previously been reported to interact; P15 directly binds with Sp1 at the luteinizing hormone receptor (LHR). P15 is recruited by Sp1 and enhances Sp1 mediated gene activation, serving as a linker protein with the general transcription machinery [366]. Current research by this group is further investigating the linker function of P15.

As I described in Chapter 6, DEAF-1 has been previously reported to form dimers with itself and has been suggested to interact with other co repressor proteins through its C terminal region [348, 356]. DEAF-1 can also activate gene expression independently of promoter binding, perhaps through an interaction with the basal transcription machinery [345].

In a study examining the binding sequences of DEAF-1, over half of sequences DEAF-1 was shown to bind to contained two or more copies of the TCGG consensus sequence [345]. Within exon one of the *GDF5* gene, there are no additional DEAF-1 consensus

sequence sites (TCGG), there are however multiple TTGG sites. If DEAF-1 is able to bind more avidly to this sequence, it may be that multiple DEAF-1 proteins are able to bind to *GDF5* and mediate transcriptional activity. Furthermore, previous literature suggests that a single base change within the DEAF-1 consensus sequence can alter DEAF-1 transcriptional activity [349].

DEAF-1 has not been previously reported to interact with Sp1, Sp3 or P15. However, each of these proteins can interact with a wide variety of *trans*-acting factors, therefore an interaction is feasible. Both Sp1 and DEAF-1 can form multimers and P15, Sp1 and Sp3 have been reported to interact with the basal transcriptional machinery to modulate gene expression.

I have shown within this thesis that all four factors, Sp1, Sp3, P15 and DEAF-1 are regulating the expression of *GDF5*. Using my experimental data, the predicted binding regions for each protein and information from previous literature I have prepared a model for how I believe Sp1, Sp3, P15 and DEAF-1 are interacting relative to rs143383 (Figure 8.12). The core consensus site for DEAF-1 (TCGG) resides directly over the SNP, whereas the predicted Sp1/Sp3 binding site (underlined) <u>GGGGGCGG</u>T**T**GG is immediately upstream of rs143383 (highlighted bold). The repressive effect that we observed for P15 was subtler than that seen for DEAF-1, Sp1 or Sp3. When over expressed in combination, Sp1 and Sp3, Sp1 and DEAF-1 and Sp3 and DEAF-1 mediated additional repressive effects compared to when over expressed alone. I propose therefore that DEAF-1, Sp1 and Sp3 are forming a repressive complex that forms directly over rs143383 whilst P15 interacts with this complex and serves as a linker between Sp1 and the general transcription machinery. As I mentioned previously, this role has been proposed for P15 at the LHR promoter [366].

In our laboratory, we have very recently identified YY1 as a transcriptional activator that binds 80bp upstream of rs143383, within the *GDF5* promoter [297];YY1 and Sp1 have previously been shown to interact to modulate gene expression [367]. Therefore, it is possible that YY1 may indirectly interact with the complex at rs143383. Future experiments could be used to further investigate this possible interaction using co-immunoprecipitation assays or combined knockdown/over expression studies.

The rs143383 polymorphism is clearly modulating the expression of *GDF5*, and I propose that this effect is being modulated through the binding of the *trans*-acting

factors Sp1, Sp3, P15 and DEAF-1. Other factors have not been identified that may be binding within this complex, however it is not possible at this stage to rule out that there may be additional *trans*-acting factors binding within this complex.

Chapter 9: The Investigation of Protein Binding to rs143384

9.1 Introduction

The functional effects of the rs143383 polymorphism have been previously investigated [134, 267, 272]. Egli et. al. observed a reduction in the expression of the T allele relative to the C allele in OA patient tissue samples [272]. All of the joint tissues tested from nine of the ten heterozygous patients studied showed a reduction in the expression of the rs143383 T allele relative to the C allele. In the tenth patient however, there was a significantly decreased expression of the C allele relative to the T allele. This was the first time that this had been observed and implied the existence of at least one other *cis*-acting *GDF5* regulatory polymorphism. Egli et. al. sought to identify this polymorphism by genotyping a total of 14 polymorphisms from within a 5.5kb region encompassing *GDF5* in the ten patients. rs143384 was the only SNP for which the tenth patient had a unique genotype [272].

rs143384, like rs143383, is located within the 5´UTR of *GDF5*. It is downstream of rs143383 and is closer to the translation start codon. As for rs143383, rs143384 is a C/T transition SNP. The patient demonstrating the increased expression at rs143383 was CC homozygous at rs143384, whereas the other nine patients were CT heterozygous. There is relatively high linkage disequilibrium (LD) between rs143383 and rs143384, with a pairwise r^2 in Europeans of 0.75 (available from HapMAP). The nine patients carried CC and TT haplotypes (rs143383 - rs143384), whereas the tenth patient carried a TC and CC haplotype [272].

Further investigations into the role of rs143384, completed by Egli et.al., were carried out using luciferase reporter assays to investigate the impact of the rs143384 genotype on the allelic expression imbalance of *GDF5 [272]*. The promoter activity of the four possible haplotypes (CC, TT, TC and CT) was investigated in CH8, SW872 and MG63 cell lines. In all three cell lines, the OA associated T allele at rs143383 was only found to mediate decreased expression of *GDF5* relative to the C allele when it was on the background of a T allele at rs143384 (TT haplotype). When the rs143383 T allele is on the background of a C allele at rs143384 (TC haplotype), there was an increase in the expression of the rs143383 T allele relative to the C allele. This recapitulated what was observed in the tissue samples from the tenth patient [272].

Table 9.1 shows the population data available for rs143384. As for rs143383, the T allele is the common allele in Europeans and Asians. In West Africans the SNP is not polymorphic.

Group	**Genotype**			**Alleles**	
	CC	**CT**	**TT**	**C**	**T**
European	0.124	0.469	0.407	0.358	0.642
Han Chinese	0.070	0.349	0.581	0.244	0.756
Japanese	0.093	0.314	0.593	0.250	0.750
West African	1			1	

Table 9.1
Frequencies of the rs143384 CC, CT and TT genotypes in addition to the C and T allelic frequency within different population groups. Population frequency information obtained from dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/).

Following my earlier description of the identification of proteins that are binding to and modulating *GDF5* allelic expression at rs143383, I wanted to investigate further the role of rs143384. As noted above, the genotype at rs143384 impacts upon the DAE observed at rs143383 and thus there may be an interaction between the proteins binding to rs143383 and the proteins binding to rs143384. I therefore sought to characterise protein binding to rs143384 using a similar approach to that employed to investigate rs143383, using online prediction software tools followed by EMSAs. Shane Damsell, an undergraduate student completed the EMSAs within this chapter as part of a BSc project under supervision from Professor John Loughlin and I.

9.2 Aims

- To identify protein binding to rs143384.

9.3 Results

9.3.1 Optimisation of EMSA Binding Conditions

In order to assess if there are any differences in protein binding to the C and T alleles of rs143384 I used EMSAs. Nuclear protein was extracted from SW872 cells and the EMSA reaction performed as described in Materials and Methods sections 2.4 and 2.6 respectively. Labelled probes and competitor sequences are shown in Table 1B, Appendix.

An optimisation EMSA was performed using 200fmol of rs143384 C or T allele probe incubated with SW872 nuclear extract, binding buffer, DTT and Poly (dI:dC) as shown in Figure 9.1A (lanes 1, 2 and 15). Different components were also added to assess optimal binding conditions, these included salmon sperm (lanes 3 and 4), glycerol (lanes 5 and 6), NP40 (lanes 7 and 8), potassium chloride (KCl) (lanes 9 and 10), magnesium chloride ($MgCl_2$) (lanes 11and 12) and EDTA (lanes 13 and 14). Four protein complexes binding to the two probes are highlighted by arrows. Most of the conditions show a similar pattern of binding. However, binding of complex one appears to be T allele specific and binding of complex two increases in the presence of EDTA. Complex three appears to bind most abundantly in the +Protein sample (lanes 1, 2 and 15) and complex four appears to only bind in the presence of $MgCl_2$.

Binding to the C and T allele probes was then assessed for a second time using the following combinations of the additional components; $MgCl_2$ and glycerol, $MgCl_2$ and NP40, $MgCl_2$ and KCl and $MgCl_2$ and EDTA (Figure 9.1B). $MgCl_2$ was used in all the combinations to optimise the conditions for the binding of all four complexes but particularly complex four (which was binding in the presence of $MgCl_2$). Binding of complexes two and three was strongest in the presence of $MgCl_2$ and EDTA (lanes 13 and 14), however these conditions prevented the binding of complexes one and four. Complexes one, two and three bound in the presence of protein alone (+ Protein, lanes 3 and 4).

Optimal binding of complexes one, two and three was observed in the presence of protein alone and the binding of complexes two and three was enhanced in the presence of $MgCl_2$ and EDTA. Complex four bound only in the presence of $MgCl_2$, and was inhibited in the presence of EDTA. The final optimisation therefore assessed the specificity of binding to the two probes using these three conditions (protein alone

(+Protein), $MgCl_2$, and $MgCl_2$ and EDTA) in addition to increasing amounts of unlabelled C and T allele competitors (Figure 9.2 A and B). The complex binding specifically to the C allele probe is indicated and is reduced upon addition of increasing amounts of the unlabelled C allele competitor (Figure 9.2A). The lower molecular weight complex that binds in the presence of $MgCl_2$ is not specific as it is not competed by the C allele competitor. Complexes one and two (indicated by arrows) bind specifically to the T allele probe (Figure 9.2B) and are reduced on the addition of increasing concentrations of unlabelled T allele competitor. Complex one binding is inhibited in the presence of $MgCl_2$ and EDTA (lanes 11-15). Of the different conditions tested, binding in the presence of protein alone (+ Protein, lanes 1-5) appeared to be optimal, thus subsequent EMSAs were performed using nuclear protein extract with no added components.

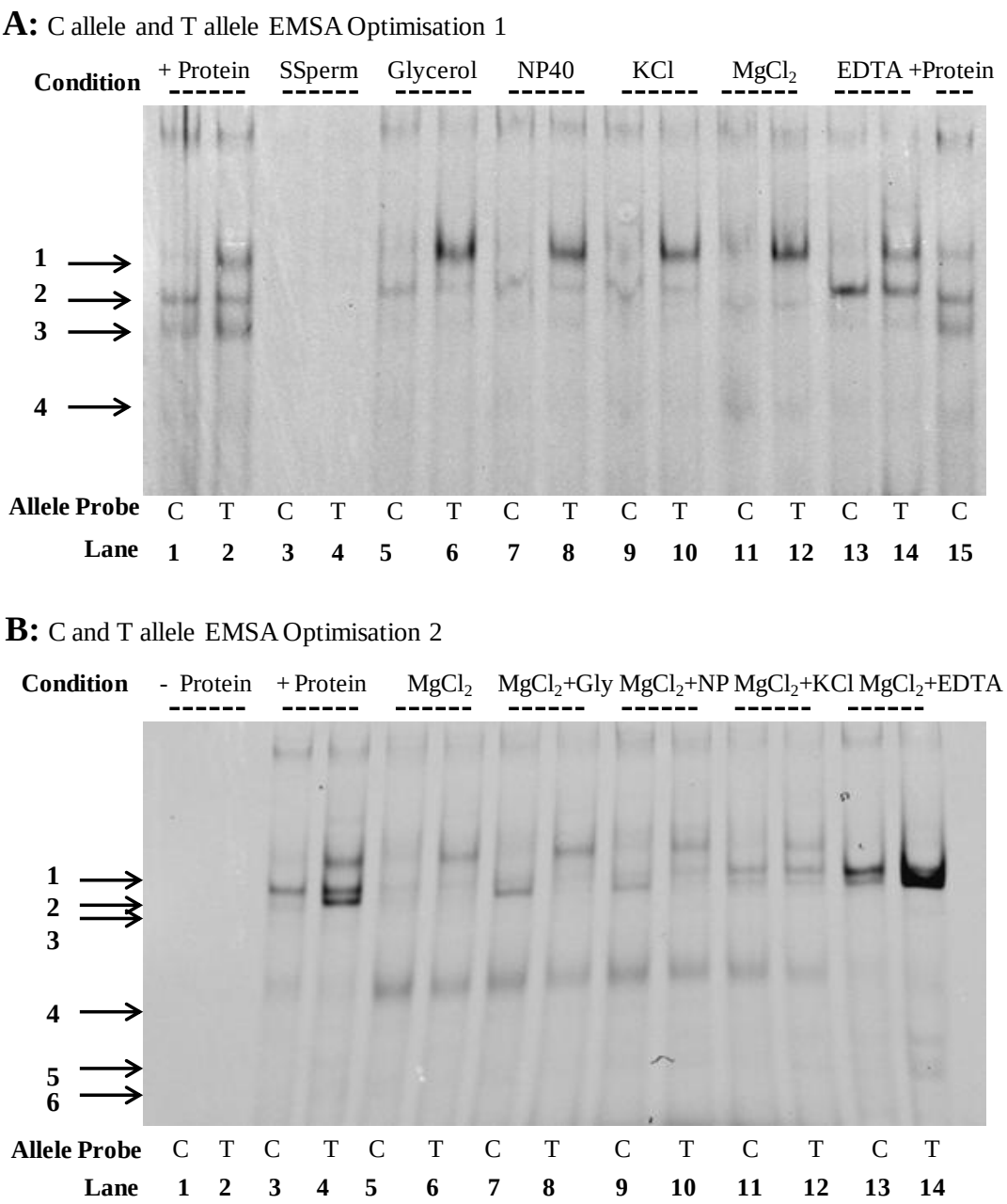


Figure 9.1

Optimisation of the binding conditions of proteins to the C and T allele EMSA probes. **(A)** EMSA Optimisation 1: EMSA analysis optimising the conditions for the binding of proteins using C and T allele probes for rs143384 and nuclear extract from SW872. Lanes 1, 2 and 15 contain C and T allele probes with no additional components. Lanes 3-14 contain C and T allele probes in addition to salmon sperm DNA (SSperm), glycerol, NP40, potassium chloride (KCl), magnesium chloride ($MgCl_2$) and EDTA. The arrows indicate protein complexes binding to the probes. **(B)** EMSA Optimisation 2: EMSA analysis using the C and T allele probes and nuclear extract from SW872 cells. Lanes 1 and 2 do not contain nuclear protein. Lanes 3-15 contain SW872 nuclear protein, with no additional components (Lanes 3 and 4), or in addition to $MgCl_2$ (lanes 5 and 6), $MgCl_2$ and glycerol ($MgCl_2$ and Gly, lanes 7 and 8), $MgCl_2$ and NP40 ($MgCl_2$ and NP, lanes 9 and 10), $MgCl_2$ and potassium chloride ($MgCl_2$ and KCl, lanes 11 and 12) and $MgCl_2$ and EDTA (lanes 13 and 14).

Following optimisation of binding, unlabelled C and T allele competitors were used to identify any differences in binding affinity to the two probes (Figure 9.2C). In comparison to the effect of adding increasing amounts of unlabelled C allele competitor to compete binding to the C allele probe (Figure 9.2A), complex binding to the C allele probe appears to be outcompeted at a lower concentration of unlabelled T allele competitor (Figure 9.2C). Furthermore, binding of complex one to the T allele probe appears to be competed at a lower concentration of unlabelled T allele competitor (Figure 9.2B), in comparison with unlabelled C allele competitor (Figure 9.2C). These results suggest that complex two binds more avidly to the T allele compared to the C allele, and verifies that complex one appears to be T allele specific.

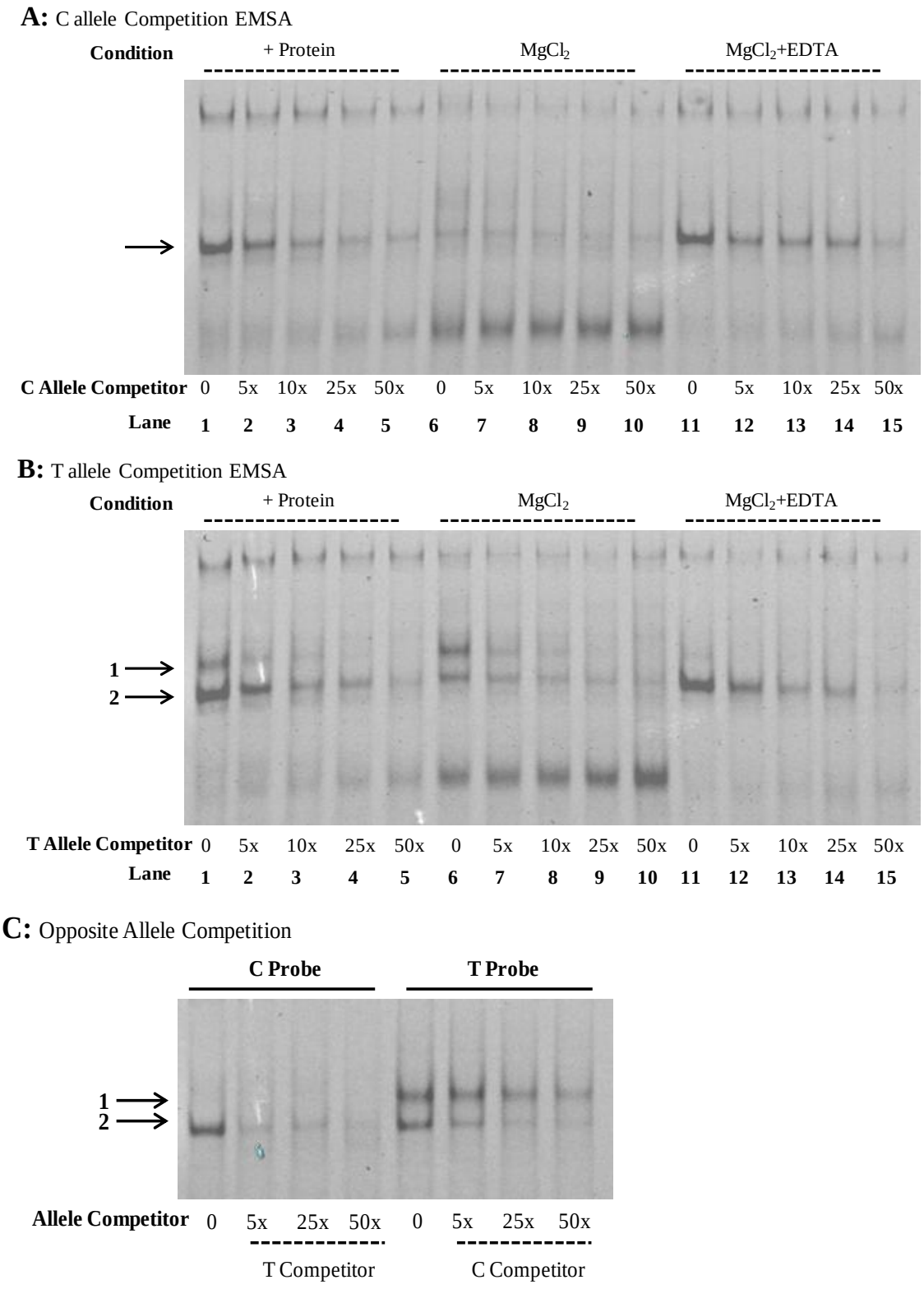


Figure 9.2

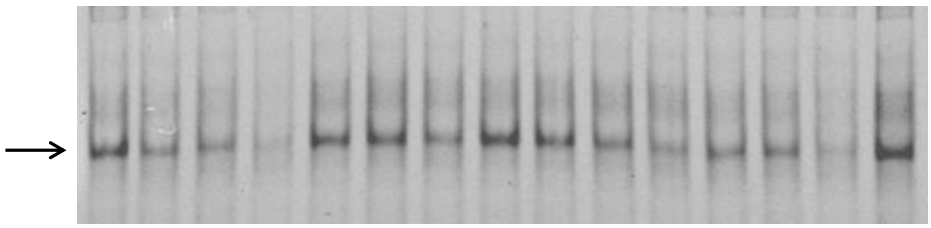
Further optimisation of binding conditions by EMSA analysis and the addition of unlabelled competitors. (**A**) Increasing concentrations of unlabelled C allele competitor were added to the EMSA reaction containing the C allele probe and SW872 nuclear extract. Three conditions (nuclear extract alone, $MgCl_2$, $MgCl_2$ plus EDTA) were assessed, with the arrow indicating the specific complex binding to the probe. (**B**) Increasing concentrations of unlabelled T allele competitor were added to the EMSA reaction containing the T allele probe and SW872 nuclear extract. Three conditions

(nuclear extract alone, $MgCl_2$, $MgCl_2$ plus EDTA) were assessed, with the arrows indicating the specific complexes binding to the probe. **(C)** Increasing concentrations of the C and T allele competitors were added to the EMSA reaction containing the C and T allele probes and SW872 nuclear extract. Binding to the C allele probe was competed with the T allele competitor and *vice versa*. The arrows indicate the specific complexes binding to the probes.

9.3.2 Investigation of Binding Region

The region of binding of the two complexes to *GDF5* was then assessed by the addition of smaller sized competitor sequences covering different regions of the *GDF5* probe to the EMSA reaction. Figure 9.3A shows protein binding to the C allele full sized probe and the effect of adding increasing concentrations of the full sized competitor and three smaller C allele competitors. The competition assay suggests that binding of the protein complex can be outcompeted using all four competitors, but competition is strongest when competing with the full sized competitor and competitor three (Comp 3). Figure 9.3B shows protein binding to the T allele full sized probe and the effect of adding increasing concentrations of the full sized competitor and three smaller T allele competitors. The competition assay suggests that for complex two, binding can be competed with all four competitors. Complex one binding however can only be competed using either the full sized competitor, or to a lesser extent using competitor one (Comp 1). These results suggest that the binding of complex two to the C and T allele probes may require the majority of the probe region, however the binding of the T specific complex (complex 1) requires the rs143384 polymorphic region and the region upstream.

A: C allele Small Sized Competitors

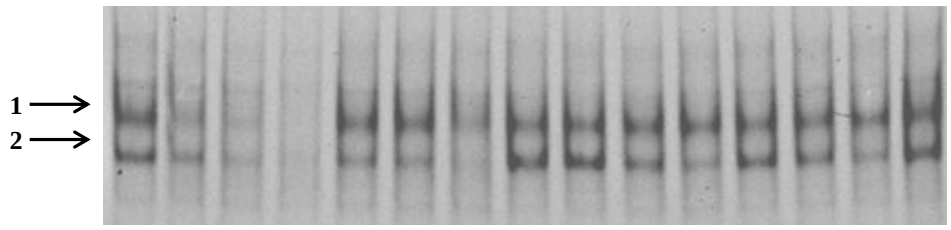


C Allele Competitors 0 5x 10x 50x 5x 10x 50x 0 5x 10x 50x 5x 10x 50x 0

Full Size Comp 1 Comp 2 Comp 3

Full Size - CAGAGGCACCTT**C**GCTGCTGCCGCTGTTCTCTT
Comp 1 - CAGAGGCACCTT**C**GC
Comp 2 - CTT**C**GCTGCTGCCGC
Comp 3 - TGCTGCCGCTGTTCTCTT

B: T allele Small Sized Competitors



T Allele Competitors 0 5x 10x 50x 5x 10x 50x 0 5x 10x 50x 5x 10x 50x 0

Full Size Comp 1 Comp 2 Comp 3

Full Size - CAGAGGCACCTT**T**GCTGCTGCCGCTGTTCTCTT
Comp 1 - CAGAGGCACCTT**T**GC
Comp 2 - CTT**T**GCTGCTGCCGC
Comp 3 - TGCTGCCGCTGTTCTCTT

Figure 9.3

EMSA analysis of the binding region. **(A)** Increasing concentrations (5x, 10x and 50x the probe concentration) of the C allele unlabelled competitors of varying sizes (full sized competitor, and three smaller competitors covering different regions: Comp 1, Comp 2 and Comp 3) were added to the EMSA reaction containing the C allele probe. The sequences of each of the competitors are shown below the EMSA, with the rs143384 polymorphism highlighted in bold and underlined. **(B)** Increasing concentrations (5x, 10x and 50x the probe concentration) of the T allele unlabelled competitors of varying sizes (full sized competitor, and three smaller competitors covering different regions: Comp 1, Comp 2 and Comp 3) were added to the EMSA reaction containing the T allele probe. The sequences of each of the competitors are shown below the EMSA, with the rs143384 polymorphism highlighted in bold and underlined.

9.3.3 Consensus Competition

Following the discovery that two protein complexes bind to the *GDF5* probe and that their binding is modulated by the two alleles of rs143384, online databases were used to identify proteins predicted to bind at this site, as described in Materials and Methods section 2.6. Table 9.2 lists the transcription factors predicted to bind.

Factor	Consensus Sequence (5′-3′)	Database
Pax-5	CGCTTTCG	Promo
HNF3-α	TGACCTTTGAC	Promo
LEF-1	CTTTGATC	Promo and Tess
HNF4-α2	TGACCTTTGAC	Promo
ADF-1	GCCGCTGC	Promo and Tess
TCF-4E	CTTTGCTC	Promo and Tess
MAC-1	TTTGCTC	Promo
AR AR (2)	TGTTCTGA CTGTTCT	Promo
GR GR (2)	TGTTCTG CATCTGTTCTTG	Tess
PR PR (2) PRβ	ACTGTTC GGATGTTCTG TGTTCTC	Tess
Pax-9a/b	CACCGT	Promo
NF1/L	CTGCAGCTGTTGT	Promo
LF-A1	TGCCCTG	Promo
CREM-tau	GCTGGTG	Promo
HoxA3	TTGTG	Promo
D.G.S	TTCCTTT	Promo
c-Myb	CAGTTGAGG	Promo
E2F1	TCCCGCC	Promo and Tess

Table 9.2

The *trans*-acting factors predicted to bind to the sequence encompassing rs143384, their consensus sequence and the database used for identification.

Increasing concentrations of consensus sequence competitors for each of these factors was added to the EMSA reaction (Figure 9.4A-C). Binding of complex two to the C and T allele probes was competed with increasing concentrations of Pax-5, HNF3-α (HNF-3), LEF-1, HNF4-α2 (HNF-4), ADF-1, TCF-4E and MAC-1 competitors, whilst binding of complex one to the T allele probe was competed with LEF-1, HNF4-α2 (HNF-4), ADF-1, TCF-4E and MAC-1.

AR, GR and PR each have two consensus binding sequences and both of these sequences were used for each factor in the competition experiment; however, the effect

on complex binding to the T allele for each alternative consensus competitor was not consistent (Figure 9.4B).

The random sequence into which the consensus competitor sequence is inserted for competition experiments was also examined to ensure that it is not competing for complex binding to the probes (Figure 9.4B and 9.5A). There was no competition seen on the addition of increasing concentrations of the random competitor, confirming that the competition for protein binding observed is specific to the consensus competitor sequences.

Binding of complex two to the T allele was also competed with increasing concentrations of Pax-9a/b and NF1/L, whilst the binding of complex one was competed with increasing concentrations of NF1/L and LF-A1 (Figures 9.4B and C).

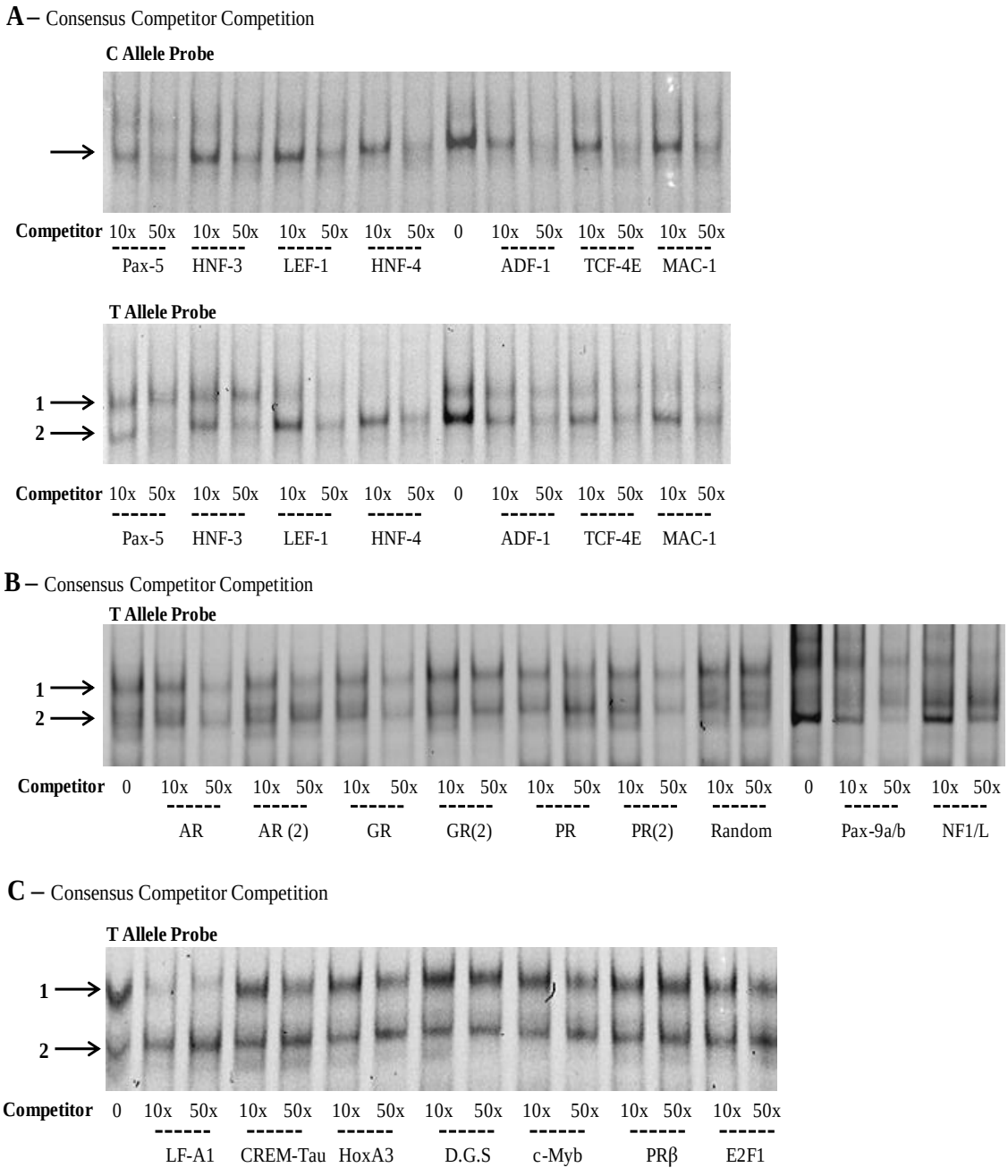


Figure 9.4

EMSA competition analysis. **(A)** The addition of increasing concentrations of unlabelled consensus competitors that were predicted to bind to *GDF5* to the EMSA reaction containing the C and T allele probes*;* Pax-5, HNF3-α (HNF-3), LEF-1, HNF4-α2 (HNF-4), ADF-1, TCF-4E, MAC-1. **(B)** and **(C)** Binding of the following *trans*-acting factors were also assessed for the T allele probe; AR, GR, PR, Pax-9a/b, NF1/L, LF-A1, CREM-Tau, HoxA3, DEF/GLO/SQ (D.G.S), c-Myb, PRβ, E2F1. AR, GR and PR each have an alternative consensus sequence, which was also assessed; AR(2), GR(2) and PR(2). The random sequence into which the consensus sequences of each *trans*-acting factor is inserted was also studied (Random).

A greater range of concentrations of LEF-1, HNF4-α2 (HNF-4), TCF-4E, MAC-1, ADF-1, Pax-5 and HNF3-α (HNF-3) consensus competitors was added to the EMSA and the effects are shown in Figure 9.5B and C. Increasing concentrations of LEF-1 and HNF4-α2 (HNF-4) competitors consistently affected both complex one and two binding to the *GDF5* probe, whilst the addition of increasing concentrations of TCF-4E and MAC-1 competed complex one binding but did not consistently compete the binding of complex two. Increasing concentrations of Pax-5 and HNF3-α (HNF-3) on the other hand consistently competed binding of complex two but their effects on the binding of complex one were not concentration dependent (Figure 9.5C).

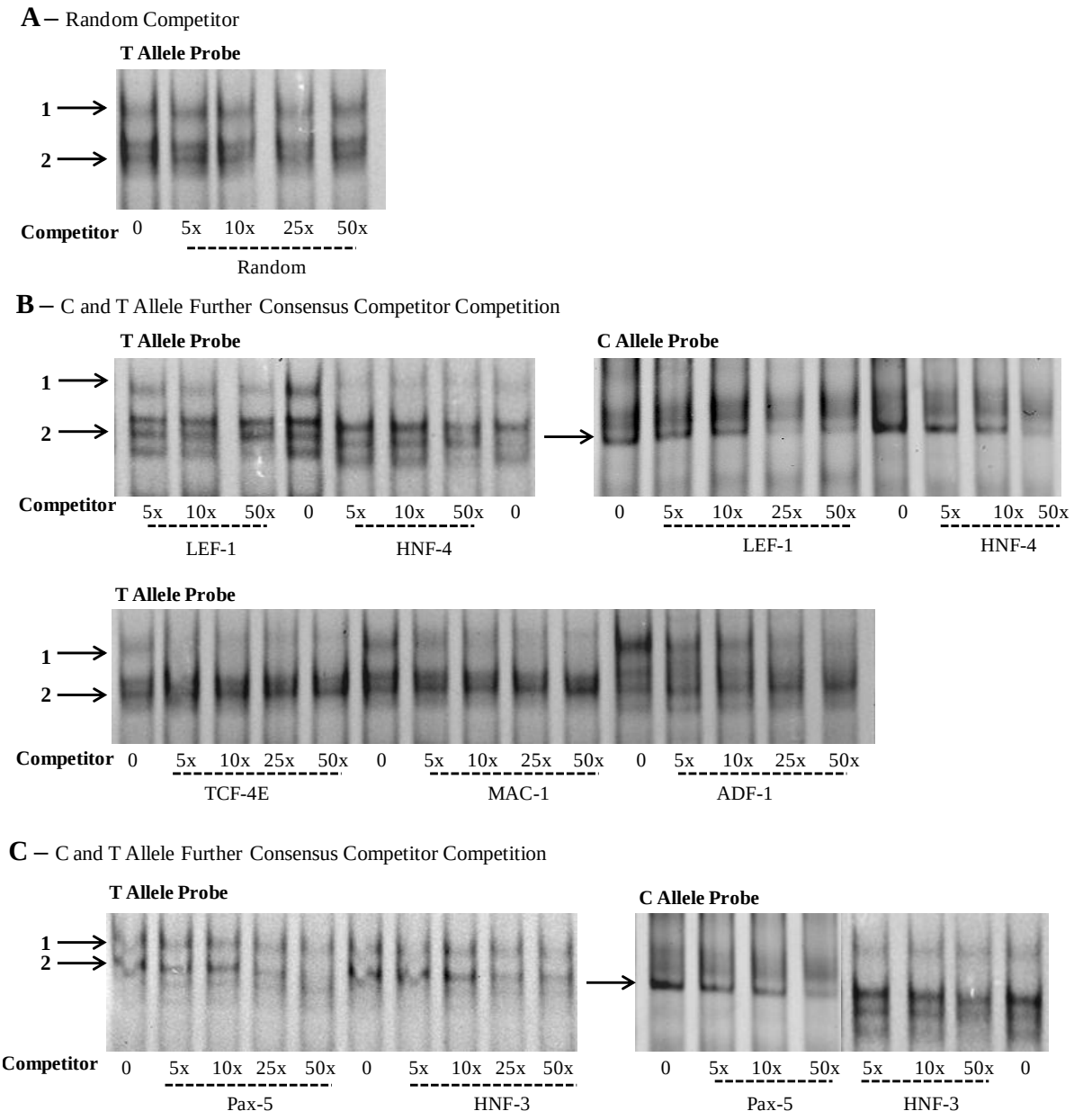


Figure 9.5

EMSA analysis using additional concentrations of competitors. **(A)** The addition of increasing concentrations of the Random unlabelled consensus competitor to the T allele probe. **(B)** The addition of increasing concentrations of LEF-1 and HNF4-α2 (HNF-4) to the EMSA reaction containing the C and T allele probes, and the addition of increasing concentrations of TCF-4E, MAC-1 and ADF-1 to the EMSA reaction containing the T allele probe. **(C)** The addition of increasing concentrations of Pax-5 and HNF3-α (HNF-3) to the EMSA reaction containing the C and T allele probes.

9.3.4 Addition of Antibodies

Based on the results from the competition EMSAs, the binding of LEF-1 and HNF4-α2 to the *GDF5* rs143384 probes was investigated further using antibodies targeting these two *trans*-acting factors. No supershifts were observed however (Figure 9.6).

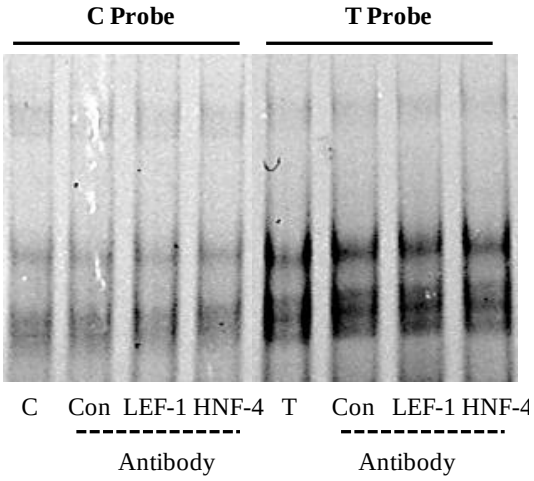


Figure 9.6

EMSA antibody analysis. Supershift experiment demonstrating the affect of adding antibodies targeting LEF-1 and HNF4-α2 (HNF-4) using the C and T allele probes, compared to the IgG rabbit antibody control (Con) and C or T probe alone.

9.4 Discussion

Protein binding to the rs143384 SNP was investigated by EMSA. One protein complex was observed as binding to the C and T alleles, whilst an additional protein complex was binding specifically to the T allele. On the addition of increasing amounts of unlabelled C and T allele competitor, the binding of the lower molecular weight complex was out-competed. Binding of this complex was competed at a lower concentration of T allele competitor compared with C allele competitor. The higher molecular weight complex binding to the T allele was not outcompeted using increasing concentrations of unlabelled C allele competitor. Using smaller sized competitor sequences, I refined the region of binding of the T specific complex to the region upstream of and encompassing the SNP. The binding of the lower molecular weight protein complex was not refined further as each of the smaller sized competitors produced a similar degree of competition.

By competition EMSA, I investigated factors predicted to bind to the probe. Increasing concentrations of Pax-5, HNF3-α, LEF-1, HNF4-α2, Pax9a/b and NF1/L competitors decreased the binding of complex two to the C and T allele probes. Pax-5, HNF3-α, LEF-1 and HNF4-α2 were all investigated at a greater range of competitor concentrations. The addition of increasing concentrations of LEF-1 and HNF4-α2 competitors also competed the binding of complex one. Pax-5, HNF3-α, LEF-1 and HNF4-α2 all share four common bases in their consensus sequences, CTTT. This sequence is identical to the region encompassing the rs143384 T allele probe, with the C allele probe containing the sequence CTTC. This preference for the T allele in the consensus sequence for the proteins could therefore explain the more avid binding of protein complex two to the T allele.

The binding of LEF-1 and HNF4-α2 was further investigated by supershift EMSA. These proteins did not however appear to be present within the complex binding to the *GDF5* probe. The antibodies chosen for this analysis were polyclonal, thus it is unlikely that the negative result is due to masking of the site of recognition. The negative result may be because the antibody is binding to the target protein with low affinity and a positive control would be needed in order to investigate this further.

Due to time constraints we did not investigate the binding of Pax-5 and HNF3-α further using antibodies. These two proteins remain valid candidates; paired box protein (Pax-5) is a member of the paired box (PAX) family of transcription factors and there are

examples in the literature of hepatocyte nuclear factor 3 (HNF3) interacting with Sp1 and Sp3 to regulate gene expression [368, 369].

The addition of increasing concentrations of TCF-4E and MAC-1, and to a lesser extent, ADF-1, consistently decreased the binding of complex one to the T allele. These three factors were predicted by the online prediction tool to bind to the T allele only. TCF-4E and MAC-1 share seven common bases in their consensus sites, TTTGCTC, and this sequence differs by only one base to the T allele rs143384 *GDF5* probe (TTTGCTG). These two proteins are worthy of further investigation, but as previously mentioned time constraints limited the number of antibodies that could be tested. TCF-4E is a member of the T cell factor family of DNA binding proteins; TCFs guide β-catenin to the promoter regions of specific target genes and are therefore involved in the Wnt developmental signalling cascade [370]. Furthermore, an interaction between TCF-4E and LEF-1 has been proposed [371]. MAC1 is involved in copper ion homeostasis in yeast, its role in humans has not yet been reported [372].

The identification of the factors binding at rs143384 will enhance our understanding of the interaction between this SNP and the OA associated rs143383 SNP. As we have already identified protein binding to rs143383, knowledge of the factors binding at rs143384 will enable us to identify the mechanism by which rs143383 and rs143384 interact, and will develop our understanding of the regulation of *GDF5* gene expression, and of the association of *GDF5* with OA.

Chapter 10: General Discussion

Perspective

The search for loci that are mediating susceptibility to common, complex diseases such as OA is ongoing. Through a variety of genetic analyses including candidate gene studies and GWAS, many SNPs have been associated with OA. However, the major challenge remains to dissect this susceptibility and to understand the molecular mechanisms that underlie these associations. To date, rs143383 remains the most robust association with OA and was identified following a candidate gene study [134]. Luciferase studies confirmed that rs143383 mediates DAE of *GDF5*, the T allele produces less transcript in comparison to the C allele [134]. This DAE was hypothesised to be mediated by the differential binding of a *trans*-acting factor(s) to the two alleles [272].

GDF5 expression is highly regulated, with precise spatial and temporal expression required for the normal development of the joint [373]. Thus, the identification of factors that are responsible for fine tuning the genes expression within the joint are of interest. There have been a small number of studies that have investigated the transcriptional regulation of *GDF5*. The first characterised the promoter region of the gene, assessing the regions required for promoter activity using luciferase reporter assays. Furthermore, using EMSAs, the investigators also identified the binding of the transcription factor YY1 [300]. A second study, performed in our laboratory, identified a rare polymorphism within the 5´UTR of *GDF5* close to the transcription start site [374]. Functional studies revealed that the A allele increased *GDF5* expression relative to the C allele, to such a degree that it was able to compensate for the reduced expression mediated by the T allele of rs143383. Interestingly, YY1 was then identified using EMSAs and siRNA depletion studies as an activator that is binding more avidly to the A allele, thus increasing the expression of *GDF5* [297]. Finally, a third group have identified the binding of SOX11 to a binding site within the 5´UTR of *GDF5* [375]. This study showed that SOX11 is an activator of *GDF5*, with SOX11 over expression increasing the expression of *GDF5*. Furthermore, using *in-situ* hybridisation, they found that SOX11 and GDF5 were co-localised in the joint interzone of developing mice [375].

Key Results

The main aim of my thesis was to identify the *trans*-acting factors that are mediating the DAE at rs143383. The key results from the thesis are highlighted below:

- The SW872 liposarcoma cell line is heterozygous for rs143383 and exhibits DAE in the same direction as seen in joint tissues, namely a relative reduction in the expression from the T allele.
- Sp1 and Sp3 were identified using EMSAs as *trans*-acting factors that are binding differentially to the two rs143383 alleles.
- P15 was identified using an oligonucleotide pull down assay and mass spectrometry as binding more avidly to the T allele.
- Five DEAF-1 antibodies have been investigated, one of which appears to bind specifically to the transcription factor. DEAF-1 binding to *GDF5* has been confirmed *in vitro* by EMSA.
- Sp1, Sp3 and P15 bind *in vivo* to *GDF5*, assessed by ChIP.
- Sp1, Sp3, P15 and DEAF-1 repress the expression of *GDF5*, confirmed by siRNA depletion and over expression studies. DEAF-1 represses *GDF5* most significantly and modulates the DAE.
- In combination, Sp1 and Sp3, Sp1 and DEAF-1, and Sp3 and DEAF-1 repress more avidly the expression of *GDF5*.
- Further investigation of a second *GDF5* 5′UTR SNP, rs143384, using EMSAs revealed differential protein binding to the C and T alleles of this SNP.

rs143383 has been reproducibly associated with OA and is the only susceptibility locus that has been functionally investigated to this degree. This type of functional work is required in order to understand the basis for the disease association and identify the factors mediating the functional effect. This will not only provide researchers in the field with a more detailed understanding of gene regulation within the OA disease state, but could also provide new targets for therapy. The modulation of *trans*-acting factors that are binding differentially to SNPs could represent a means to restore "normal" gene expression levels and alter disease progression.

Role of Sp1, Sp3, P15 and DEAF-1 in Joint Development

In Chapter 8, I proposed how Sp1, Sp3, P15 and DEAF-1 could be interacting, using both information known about each of these factors in addition to the experimental

evidence I have generated. All four factors are regulating the expression of *GDF5* and are therefore likely to be important for the normal formation of the synovial joint.

Sp1 and Sp3 regulate a large number of genes, some of which are vital for musculoskeletal development such as *SOX9*, *COL1A1* and *RUNX2*. A polymorphic Sp1 binding site within the *COL1A1* gene has been associated with increased incidence of osteoporosis in a number of populations [376, 377]. Sp1 was found to have increased binding affinity for the associated allele, thus affecting bone mineral density [376, 377]. This study also highlights that the activity of Sp1 can be affected by a single base pair change in its consensus sequence. Sp1 has also been found to regulate the expression of *SOX9* in addition to cyclic-AMP response element binding protein (CREB) [378]. Furthermore, Sp1 was found to activate *RUNX2* expression through an interaction with Ets-like factors during osteogenesis [379]. Sp3 has been found to repress the Sp1 mediated transactivation of the *COL2A1* gene in chondrocytes [380]. Sp1 and Sp3 knockout mice exhibit gross morphological defects, however, there does not appear to be any phenotypic overlap between these mice and the GDF5 null brachypodism mouse [319, 320]. This is not surprising considering the vast number of genes regulated by the Sp proteins.

P15 also regulates a large number of genes and cooperates with the basal transcriptional machinery to modulate gene expression. There have been no studies which have investigated the role of P15 in musculoskeletal development and therefore the regulation by P15 of joint specific genes is currently unknown. P15 knockout mice are lethal, highlighting the important role of this factor during development. However, heterozygous knockout mice display no overt phenotype, indicating there may be a threshold level of P15 that is required for normal development [330].

There are no reports of DEAF-1 modulating the expression of genes involved in the formation of the joint and although DEAF-1 knockout mice display skeletal abnormalities such as rib cage defects and abnormalities in their cervical vertebrae, there is no obvious overlap between the DEAF-1 knockout mouse and the GDF5 bp mouse. However, DEAF-1 is expressed at high levels during development and a large proportion of DEAF-1 knockout mice suffer from defective neural tube closure that causes death shortly after birth, indicating the importance of this factor for normal development [345, 346].

Further studies could assess the role of each of these factors further in the development of the joint using conditional knockout mice. Homozygous knockout mice for Sp1, Sp3, P15 and DEAF-1 are not viable, however, heterozygous knockouts for each of these factors are viable (although some DEAF-1 heterozygous mice suffer from excencephaly and die) [319, 321, 330, 346]. A conditional deletion of each of these factors in a joint tissue specific manner could provide more information as to the importance of each factor. Furthermore double, triple or even quadruple conditional knockouts could aid in understanding the combined effect of the four factors in the normal development of the joint. However, Sp1 and Sp3 compound heterozygous knockout mice are not viable, therefore joint depletion of these two factors could not be investigated simultaneously [381]. The viable single heterozygous knockout mice for each factor could be investigated further in ageing studies or could be challenged to discover if these mice are more susceptible to developing joint specific diseases.

The Impact of rs143383 Beyond OA

The relevance of the results within this thesis extend beyond OA, since the T allele of rs143383 has been associated with a number of other musculoskeletal phenotypes. These associations have been previously discussed in my Introduction, section 1.10.1 and include congenital hip dysplasia, lumbar disc degeneration and Achilles tendinopathy [275-277]. Furthermore, this SNP has been associated with variation in normal height, hip axis length, and fracture risk in women [273, 274]. Knowledge of the factors that are differentially regulating the expression of *GDF5* in T allele carriers could therefore be applied to these other diseases. As mentioned in the Introduction, section 1.9.6, GDF5 treatment has been shown to have therapeutic benefit in tendon and ligament injury animal models, in addition to being tested in spinal fusion models and for periodontal wound healing [253, 257, 264, 265]. Aside from treatment with recombinant GDF5 protein, the factors that have been identified that regulate *GDF5* expression could be targeted to increase *GDF5* and could therefore be of use in these repair models. The observation that rs143383 is associated with height indicates that the effect of the polymorphism is active during skeletal development in addition to in adult life. This therefore raises the question of when inhibitors targeting the *trans*-acting factors that regulate *GDF5* expression could be of therapeutic benefit; in individuals showing signs of OA, increasing *GDF5* expression is unlikely to repair tissue damage or slow disease progression if the susceptibility to OA is a result of the decreased expression of *GDF5* during development.

Transcription Factor Inhibitors as Therapy

Numerous studies have shown that variability in the expression of genes underlies the pathology of a number of diseases [382]. It is not therefore surprising that targeting transcription factors that regulate the expression of genes has been proposed as a potential therapeutic strategy. There are a number of examples within the literature of *trans*-acting factors being targeted in order to modulate gene expression [383].

Strategies that have been employed to inhibit transcription factor activity include:

- Triple-helix-forming oligonucleotides (TFOs) that bind to the target site of the transcription factor and prevent their binding to DNA.
- Antisense targeting of the transcription factor mRNA to prevent their expression.
- Use of mutant transcription factors which have a dominant negative effect.
- Targeting of transcription factor metabolism (i.e. inhibiting kinases that activate the factor).
- Preventing the transport of the transcription factor into the nucleus.
- Targeting transcription factors with aptamers that interfere with transcription factor binding.
- Transcription factor decoy approach, which involves the use of double stranded oligonucleotides to interfere with the recruitment of transcription factors to DNA.

All of these approaches have been used to target transcription factors and subsequently modulate the expression of genes [383]. One approach which has proven effective *in vivo* and is being considered for clinical application, is the inhibition of transcription factors with molecules that mimic the transcription factor binding site. This is known as the transcription factor decoy approach [384, 385]. Specifically, these molecules are double stranded DNA, designed to mimic the specific *cis*-element to which the *trans*-acting factor binds. On the addition of these molecules, *cis-trans* interactions are attenuated and thus the effect of the *trans*-acting factor is diminished. This approach has been used to target a number of transcription factors, including Sp1, GATA-1, NF- κB, AP-1 and E2F [383]. As an example, E2F decoy oligonucleotides have been shown in an *in vivo* model of rat carotid injury to inhibit the expression of a number of target genes including *c-myc*, and as a result inhibited smooth muscle proliferation and vascular lesion formation [386]. Furthermore, Sp1 regulates the expression of the

urokinase receptor (*uPAR*), a key molecule in tumour formation. Using transcription factor decoy, Sp1 binding to *uPAR* was inhibited, resulting in decreased *uPAR* expression and reduced breast cancer cell motility, highlighting a potential therapeutic benefit for this approach [387].

Potentially, the inhibition of Sp1, Sp3, P15 and DEAF-1 could be achieved using the decoy approach to modulate the expression of *GDF5*. Since each of these factors has been shown in Chapter 7 to repress the expression of *GDF5*, an increase in the transcription of *GDF5* would be expected in decoy treated cells. Certain aspects however would need to be considered. Firstly, the effect of inhibiting each factor independently or in different combinations should be assessed to determine which group of factors would need to be inhibited in order to normalise *GDF5* expression. Secondly, considering the variety of genes targeted by these transcription factors, the inhibitory effect should be limited to the joint, thus a targeted delivery approach would be needed. Lastly, important safety aspects would need to be considered, such as the stability of the decoy molecules *in vivo*, potential toxicities or effects on the immune response. ChIP seq could be utilised in order to identify the other regions of the genome that are regulated by each of these factors to predict potential off target effects.

In the cancer field, oestrogen receptor (ER) antagonists are an example of a therapy which has been trialled for the treatment of breast cancers, whose mechanism of action is to inhibit the transcriptional effects of the ER [388]. The genes regulated by the ER are associated with oncogenesis, regulating cell proliferation, survival, invasion and angiogenesis. Additionally, heat shock factor 1 (HSF1) is a transcription factor that can promote oncogenesis, thus a therapeutic role for HSF1 inhibitors is being investigated [389]. Aside from cancer, the inhibition of transcription factors has been investigated for the treatment of rheumatoid arthritis [390, 391]. Although small molecule inhibitors targeted against catabolic factors such as MMPs have been trialled for OA treatment, as yet, no inhibitors of transcription factors have been tested [18]. However, transcription factors which promote the expression of these catabolic factors are now being investigated and could be the targets of future therapeutics, with examples including NFAT1 and SAF1 [392, 393]. Additionally a novel DMOAD candidate has recently been identified, known as TD-198946, which increases the expression of the transcription factor Runx1. Increased expression of Runx1 promotes chondrogenic differentiation and suppresses hypertrophy, making this a promising new compound for OA therapy [394].

The Importance of GDF5 Methylation to rs143383 DAE

The extent of the rs143383 DAE varies between joint tissues of the same individual and also inter-individually, between individuals of the same haplotype, implying that DAE may be modulated epigenetically [272]. The epigenetic regulation of *GDF5* has been investigated within our laboratory [299]. The C alleles of both rs143383 and rs143384 form CpG dinucleotides that can be methylated. The *GDF5* promoter was found to be methylated in cell lines (CH8, SW1353, SaOS-2 and COLO205) and in OA patient tissues, and treatment with a demethylating agent 5-Aza-2′ deoxycytidine (AZA) increased *GDF5* expression. Treatment of the heterozygous SW872 cell line with AZA significantly increased the DAE between the C and T alleles from a C/T ratio of 1.1 to 2.24. There was no difference in the methylation of the C alleles of rs143383 and rs143384 following this treatment, however five different CpG sites within the *GDF5* promoter were found to be demethylated. This study suggests that in addition to genetic factors, epigenetic factors such as DNA methylation may contribute to the DAE of rs143383 [299]. One of the five sites that is demethylated is located 4bp upstream of rs143383. Current research in our laboratory is therefore investigating whether the binding of the four *trans*-acting factors that I have identified are affected when this site is methylated. If methylation of this site alters the binding of one of these four factors, this could account for the change in DAE observed upon demethylation of this site.

In Chapter 4 I discussed a role for Sp1 in methylation of the promoters of genes, whereby Sp1 sites are important for preventing the methylation of CpG islands within the *APRT* gene [316]. Furthermore, the methylation of CpG sites can prevent the binding of Sp1 and Sp3 [317, 395]. Methylation of the CpG site 4bp upstream of rs143383 could therefore affect Sp1/Sp3 binding, and the binding of these two factors, or all four identified factors, could have prevented the demethylation of the rs143383 CpG site.

Future Work

Sp1 and Sp3 were identified following database analysis and EMSAs whilst P15 was identified following a pull down experiment followed by mass spectrometry. There is no known consensus sequence for the binding of P15 and thus this factor is not present on the online prediction databases. P15 could therefore not have been predicted to bind using this type of analysis and I was not able to assess by competition EMSA whether this factor bound to the rs143383 probes. However, I was able to assess binding to the EMSA probes using an antibody targeting P15.

Sp1, Sp3 and DEAF-1 were not detected by the pull down experiment. There are a number of possible reasons for this.

- Firstly, the binding conditions (including the salt concentrations) used in the pull down assay were different to those used in the EMSA. As previously observed in Chapter 4 (Figure 4.1), altering the EMSA conditions can affect the binding of the complexes bound to the rs143383 probes.
- Secondly, the genomic DNA sequence used in the pull down assay was longer (212bp) than that used in the EMSA probes (33bp). I chose to use such a long sequence in order not to limit the capture of proteins that may bind over large DNA regions. It is possible however that by using such a long sequence, non specific proteins were captured that may have disrupted the binding of Sp1, Sp3 and DEAF-1.
- Thirdly, a non specific competitor was not used in the pull down assay, thus a greater number of proteins may have bound non specifically. In the EMSA reaction, the non specific competitor Poly (dI:dC) was used.

Overall, the length of the oligonucleotide probe and the absence of a non specific competitor was not considered an issue when designing the pull down experiment, as the primary goal was to identify proteins that bound differentially between the C and T alleles of rs143383. Therefore, any proteins binding non specifically or at a site distal to the polymorphism could be identified and removed from the analysis. However, the sensitivity of the pull down could have been attenuated by additional proteins binding. Perhaps the use of a shorter DNA sequence, the use of repeat concatamers of rs143383 and its immediate flanking sequence, or the use of a range of salt concentrations would have led to the concurrent identification of Sp1, Sp3 and DEAF-1 by the pull down approach. The EMSA and the pull down can therefore be considered complimentary approaches, with neither being infallible.

Following on from the identification of the binding of Sp1, Sp3, P15 and DEAF-1 in the SW872 liposarcoma cell line, I would extend this analysis by performing this work in a chondrogenic cell line, for example SW1353, and in primary human articular chondrocytes (HACs). The expression of all four factors in HACs and in tissues samples taken from OA and NOF patients could be determined using real time PCR. I performed the over expression studies in the chondrogenic cell line SW1353 (Chapter 7

and Chapter 8), due to difficulty with transfection efficiency I was unable to replicate these experiments in SW872 cells or HACs.

The binding of these four factors was investigated *in vivo* by ChIP using semi quantitative PCR. It would be helpful to optimise an assay for real time PCR to enable the quantitative analysis of DNA enrichment following ChIP. Additionally, it would be interesting to discover if the four factors were binding more avidly to the T allele, as predicted. This could be assessed using the DAE assay for *GDF5* to assess if more T allele DNA was present in the immunoprecipitated sample.

In Chapter 8, I used combination knockdown and over expression studies to examine the role of each of the four factors and to assess if these factors are interacting to mediate their repressive effects. I was able to over express two factors concurrently. To progress this research, I would look to optimise the over expression of three or four of the factors in combination. Additionally, I was unable to optimise the concurrent siRNA knockdown of each factor and their over expression in combination with siRNA knockdown experiments. These experiments need further optimisations and could then be used to assess the contribution of each factor.

Finally, in Chapter 9, I began to investigate protein binding to rs143384. I was not able to identify the protein complexes that were binding to the rs143384 EMSA probes during my PhD. Further competitors and antibodies could be investigated to try and identify these factors. Additionally, an oligonucletide pull down assay could be used, similar to that used for rs143383 to identify factors binding at this locus. Once these factors have been identified, co-immunoprecipitation experiments or yeast-2-hybrid assays could be utilised to assess if there is an interaction between the proteins binding at rs143383 and the proteins binding at rs143384. Furthermore, a technique such as chromatin conformation capture (3C) could be utilised to determine the spatial organisation of the *GDF5* genomic region [396]. For example, 3C could be used to investigate if the DNA is looping out, bringing rs143383 and rs143384 within close proximity, allowing for an interaction to occur between the proteins binding at these two loci. Knowledge of all of the factors binding to and regulating *GDF5* expression at this OA susceptibility locus would provide us with a better understanding of the genetic association and could potentially highlight novel targets for therapy.

Summary

Using a variety of techniques I have identified the differential binding to the two alleles of rs143383 of four proteins: Sp1, Sp3, P15 and DEAF-1. Depletion of all four factors increased the expression of *GDF5* whilst DEAF-1 depletion significantly modulated the DAE. Conversely, the over expression of all four factors repressed C and T allele expression, repressing the T allele more strongly. When over expressed together, DEAF-1 and Sp1 mediated the greatest overall repressive effect, whereas over expression of Sp1 and Sp3 together mediated the greatest differential allelic effect, repressing the T allele to a greater extent compared to the C allele. Overall, I have identified *trans*-acting factors that bind differentially to the alleles of rs143383 and which contribute to the DAE that is mediated by this important OA susceptibility locus.

A recently published study has adopted a similar approach to the oligonucleotide pull down used in Chapter 4 of this thesis, to investigate differential protein binding to 12 SNPs associated with type 1 diabetes [397]. The technique used by this group was called proteome-wide analysis of SNPs (PWAS) and involves the use of repeat concatamers for affinity purification combined with quantitative mass spectrometry to screen SNPs for differential protein binding. Several transcription factors were identified as binding differentially to four of the associated SNPs, thus this study represents an additional example of where the identity of factors binding differentially at susceptibility loci can assist in determining the molecular mechanisms responsible for the disease association.

GWAS are commonly being used to identify susceptibility loci that are contributing to the pathogenesis of complex diseases, thus there is now a need for functional analyses to be performed to understand the basis of these associations. Overall, this study offers a guide as to how to functionally dissect a common risk allele that mediates its effect by modulating the expression of its target gene.

Appendix

Probe/Competitor	Sequence (5´-3´)
C allele labelled probe	GAGAAAGGGGGCGGTCGGCTTTCTCC
T allele labelled probe	GAGAAAGGGGGCGGTTGGCTTTCTCC
Small Probe 1	GAGAAAGGGGGCGGTCGG
Small Probe 2	GGCGGTCGGCTTT
Small Probe 3	GGTCGGCTTTCTCC
Sp1/Sp3/ETF Competitor	AATTGGGGGGGCGGGGGTACGTAGCA
cmyb Competitor	AATTGGACCGGCGGTTGTACGTAGCA
E2F1 Competitor	AATTGGACCAGGGCGGGTACGTAGCA
EllaEa Competitor	AATTGGGAGGGCGTTAGTACGTAGCA
EGR1 Competitor	AATTGTGTGGGCGGGAGCACGTAGCA
GCF Competitor	AATTGGAAGCGCGGGCCGACGTAGCA
UHRF1Competitor	AATTGGACCTGGCATTGGACGTAGCA
RC2 Competitor	AATTGGACCTGGGGTTTAACGTAGCA
IA1 Competitor	ATGTAAGGGGGCGATAGTACGTAGCA
P53 Competitor	AATTGGACCTGGGCATGTACGTAGCA
NF1 Competitor	AATTGGACCTGGCTTTGGCCGTAGCA
GABP Competitor	AATTGGACAACCCCCCGTACGTAGCA
CP2 Competitor	AATTGGACCTGGCCCAATACGTAGCA
CTF Competitor	AATTGGACCTGCGTTTGGACGTAGCA
DRF1.1 Competitor	AATTGGACCGGCGGTGACTCGTAGCA
KLF16 Competitor	AATTGGAGGGGCGGTGGTACGTAGCA

Table 1A

The sequences of the rs143383 probes and of the competitor oligonucleotides used in the EMSA experiments. The forward primer sequences are shown. The consensus binding motif of the competitor proteins was identified using online prediction tools and is underlined. The flanking sequences were randomly generated.

Probe/Competitor	Sequence (5´-3´)
C allele labelled probe	CAGAGGCACCTTCGCTGCTGCCGCTGTTCTCTT
T allele labelled probe	CAGAGGCACCTTTGCTGCTGCCGCTGTTCTCTT
Small Probe 1	CAGAGGCACCTTCGC
Small Probe 2	CTTCGCTGCTGCCGC
Small Probe 3	TGCTGCCGCTGTTCTCTT
LEF-1 Competitor	AATTGGACCCTTTGCTCGTCTTAGTACGTAGCA
HNF4-α2 Competitor	AATTGGCACCTTTGCTCGTCTTAGTACGTAGCA
HNF3α Competitor	AATTGGCACCTTTGCTTACGTCTTAGTACGTAGCA
Pax-5 Competitor	AATTGGACGCTTTCGACGTCTTAGTACGTAGCA
TCF-4E Competitor	AATTGGACCCTTTGCTCGTCTTAGTACGTAGCA
MAC-1 Competitor	AATTGGACCTTTTGCTCGTCTTAGTACGTAGCA
ADF-1Competitor	AATTGGACCTGCTGCTGCTCTTAGTACGTAGCA
AR Competitor	AATTGGACCTGGCTTACGTCTTAGTGTTCTGAA
AR alt. Competitor	AATTGGACCTGGCTTACGTCTTACTGTTCTGCA
GR Competitor	AATTGGACCTGGCTTACGTCTTAGTGTTCTGCA
GR alt. Competitor	AATTGGACCTGGCTTACGTCCCGCTGTTCTCTA
GRα Competitor	AATTGGACCTGGCTTACGTCTTACTGTTCTGCA
PR Competitor	AATTGGACCTGGCTTACGTCTTGCTGTTCAGCA
PR alt. Competitor	AATTGGACCTGGCTTACGTCTCGCTGTTCTGCA
PRβ Competitor	AATTGGACCTGGCTTACGTCTTAGTGTTCTCCA
LFA-1 Competitor	AATTGGACCTGGCTTACGTGCCGCTACGTAGCA
Crem Tau1 Competitor	AATTGGACCTGGCGCTGCTGTTAGTACGTAGCA
HoxA3 Competitor	AATTGGACCTGGCTTACGTCTTACTGTTTAGCA
Def/Glo/Sq	AATTGGCACCTTTCTTACGTCTTAGTACGTAGCA
Pax-9a/b	AATTGGCACCTTCTTACGTCTTAGTACGTAGCA
E2F-1	AATTGGACCTGGCTTACGTGCCGCTACGTAGCA
NF1/L	AATTGGACCTGGCTTACCTGCCGCTGTTCTGCA
c-myb	AATTGGACCTGGCTTACGTCTTACTGTTCTCTA

Table 1B

The sequences of the rs143384 probes and of the competitor oligonucleotides used in the EMSA experiments. The forward primer sequences are shown. The consensus binding motif of the competitor proteins was identified using online prediction tools and is underlined. The flanking sequences were randomly generated.

Protein	Company	Catalogue No.	Species	Dilution (Immunoblot)
DEAF-1 (1)	Abcam	Ab50719	Rabbit	1:2000
DEAF-1 (2)	Abgent	AP2711b	Rabbit	1:100
DEAF-1 (3)	Abnova	H00010522-D01P	Rabbit	1:1000
DEAF-1 (4)	Kind donation from Paul Albert group	n/a	Rabbit	1:1000
DEAF-1 (5)	Kind donation from Jane E Visvader Group.	n/a	Rabbit	1:1000
β-Actin	Sigma	A5316	Mouse	1:2000
EGFP	Abcam	Ab290	Rabbit	1:2000
GAPDH	Millipore	MAB374	Mouse	1:40,000
Lamin A/C	Cell Signalling	2032	Rabbit	1:4000
Sp1	Santa-Cruz	sc-59 (PEP2)	Rabbit	1:200
Sp3	Santa-Cruz	Sc-644 (D-20)	Rabbit	1:200
P15	Santa-Cruz	Sc-48778 (H-114)	Rabbit	1:200
E2F1	Santa-Cruz	Sc193 (C-20)	Rabbit	n/a
HDAC1	Abcam	Ab7028	Rabbit	n/a
HDAC2	Abcam	Ab7029	Rabbit	n/a
EGR-1	Santa-Cruz	Sc-20689 (H-250)	Rabbit	n/a
KLF16	Santa-Cruz	Sc-131168	Rabbit	n/a
IgG Negative Control	Sigma-Aldrich	I5006	Rabbit	n/a
Positive ChIP (Anti-Acetyl Histone H3)	Millipore	06-599B	Rabbit	n/a
Negative ChIP Antibody (Rabbit IgG)	Millipore	PP64B	Rabbit	n/a
LEF-1	Santa-Cruz	Sc-28687	Rabbit	n/a
HNF-4	Santa-Cruz	Sc-8987	Rabbit	n/a

Table 2

Details of the antibodies used for the research.

Primer Name	Primers (5´ to 3´)
Oligonucleotide Pull Down *GDF5*	F: [biotin]CGTCGAATTCGCATTACGCCATTCTTCCTTC R: CGGGTGTGTGTTTGTATCCAG
ChIP *GDF5* (Exon 1)	F: CTTCAAGCCCTCAGTCAGTTG R: CTGGATACAAACACACACCCG
ChIP DEAF-1 Promoter	F: GACATTCGGCTCGTCTCGG R: AGCCGAGACCGAGTCCTGAAC
HBP1	F: TCGAAGAGTGAACCAGCCTT R: GAAGGCCAGGAATTGCACCATCC

Table 3

Details of the primers used for PCR reactions: Nucleotide sequences of the primers used for creating the 212 bp fragment used in the oligonucleotide pull down assay, of the primers used for PCR following ChIP, and the primers used for examining the integrity of cDNA following synthesis. F, Forward; R, Reverse.

Gene	Primers (5´ to 3´)	Probe (5´ to 3´)
GAPDH	F: GGCCATCCACAGTCTTCTG R: CAGCCTCAAGATCATCAGCAA	ATGACCACAGTCCATGCCATCACT
HPRT1	F: TGCTGAGGATTTGGAAAGGG R: ACAGAGGGCTACAATGTGATG	AGGACTGAACGTCTTGCTCGAGATG
GDF5	ABi Assay ID: Hs00167060_m1, part number: 4448892	
Sp1	F: TCAACTCTCCTCCATGCCA R: CAGGTGATCATGGAGCTCAG	ACCTGGATTCCTGAAGTACCCAATGC
Sp3	F: AGTTAGTCTAAGCACTGGTCAG R: GAAGAACCTGATCCTGAAGAGTG	ATCTGCAGGACTGTCAGCATTCTCTC
P15	F: GAAGCGATGCCTAAATCAAAGG R: AGACAGGTGAGACTTCGAGAG	CAACCTCACTGTCAGAATCACTGCCA
DEAF-1	F: GTACAGTCCCACCGAGTTTG R: GGATCTTAAACCCTCACGCT	ACCCTTGCAGTGCCTC
GDF5 DAE	F: AGTCAGTTGTGCAGGAGAAAGG R: TTCAAGAACGAGTTATTTTCAGCTGC	GGCGGTTGGCTTTCT (VIC) / GGCGGTCGGCTTTCT (FAM)

Table 4

The primers used for Real Time PCR. Nucleotide sequences of the primers and of the probes used for the real time RT-PCR assays measuring gene expression.

Primer Name	Primers (5´ to 3´)
Sp1 (*EcoR1)/(SacII)* (cloning)	F: 5´-GGGGGAATTCATGGATGAAATGACAGCTGTG-3´ R: 5´-GGGGCCGCGGGAAGCCATTGCCACTGATATT-3´
Sp3 (*EcoR1)/(SacII)* (cloning)	F: 5′-GGGGGAATTCATGACCGCTCCCGAAAAGCCC-3′ R: 5′-GGGGCCGCGGCTCCATTGTCTCATTTCCAGA-3′
P15 *(EcoR1)/(SacII)* (cloning)	F: 5′-GGGGGAATTCATGCCTAAATCAAAGGAACTT-3′ R: 5′-GGGGCCGCGGTCAGTTTTCTTACTGCATCATC-3′
EGFP-N1 Reverse (sequencing)	5′-CGTCCAGCTCGACCAGGAT-3′
Sp1 Internal (sequencing)	F: 5′- CACGATCAGCAGCTCTGGGTC -3′ R: 5′- AGCTGAGGCAATGGGTGTGAG -3′
Sp3 Internal (sequencing)	F: 5′-GGTCTGCCAGGAAATATTACG -3′ R: 5′-GCAAGCTACCCTCCGAAGTCT -3′

Table 5

The primers used to create the inserts for cloning in the over expression vectors and for sequencing of the vectors. Nucleotide sequences of the primers with the restriction enzyme sites underlined. F, Forward; R, Reverse.

siRNA Target Gene	Catalogue Number (Dharmacon)
Non-Targeting siRNA Control Pool	D-001810-10
Sp1	L-026959-00
Sp3	L-023096-00
P15 (SUB1)	L-009815-00
DEAF-1	L-020808-01

Table 6

Details of the siRNAs used for depletion studies.

Document 1

Tiered analysis of the mass spectrometry results. The results have been prioritised based on their identification in each repeat experiment and their presence in the background control sample. The 1st, 2nd and 3rd categories represent proteins identified by all three repeat experiments with C/T values that are all either below 1 (binding more to the T allele) or above 1 (binding more to the C allele). The categories that follow contain proteins identified by all three repeat experiments, with two C/T values that are either above or below 1 in each experiment, whilst the third C/T value shows the opposite. In the final two categories the protein was only identified in two of the three repeat experiments. In each category, the C/T values are given in brackets following the protein name, in experimental order. Proteins that were identified in the background control sample in addition to the C or T allele samples have the experiment C/T value underscored.

1st – In all 3 experiments, similar C/T ratios observed, not present in background

- cDNA FLJ57246 – probable Poly(A)-binding protein 1 (1.014, 1.108, 1.331).
- DNA topoisomerase II (fragment – more likely Beta) (1.019, 1.144, 1.888).
- DNA dependent protein kinase catalytic subunit (1.108, 1.000, 1.062).
- Heterogeneous nuclear ribonucleoprotein K (0.514, 0.940, 0.895)
- Heterogeneous nuclear ribonucleoprotein M (1.137, 1.518, 1.714).
- Histone H1.5 (1.462, 1.276, 1.235).
- HSP90 AA1 (fragment) (4.281, 1.660, 1.007).
- Myosin-9 (1.076, 1.019, 1.044).
- Poly [ADP-ribose] polymerase 1 (0.875, 0.384, 0.428).
- POLR2H (RNA poly II) (0.841, 0.870, 0.737).
- RNA binding motif 14 variant (fragment) (1.007, 1.596, 1.877).

2nd – In all 3 experiments, similar C/T ratios observed, present in background in 1 experiment

- Acidic ribosomal phosphoprotein P1 – binds more to T (0.693, 0.402, 0.743)
- Activated RNA polymerase II transcriptional coactivator p15 SUB1 (0.955, 0.474, 0.636)
- Antigen KI-67 (1.251, 1.175, 1.139)
- cDNA FLJ57877 – probable cleavage and polyadenylation specificity factor 7 (5.244, 1.252, 5.170).
- DNA topoisomerase II alpha (1.135, 1.109, 1.425).
- H/ACA ribonucleoprotein complex subunit 1 (1.297, 1.023, 1.069). Also subunit 2 (1.514, n/a, n/a), subunit 4 (0.576, 1.449, 1.162).
- Polypyrimidine tract-binding protein 1 (2.206, 1.135, 2.752).
- RPL32 (0.723, 0.568, 0.781).
- RSL1D1 (fragment) (0.765, 0.370, 0.726).
- Ribosomal protein L19 (1.682, 1.180, 1.021).

3rd – In all 3 experiments, similar C/T ratios observed, present in background in 2 experiments

- cDNA FLJ58772 – probable fragile x mental retardation syndrome-related protein 1 (1.271, 1.744, 1.668).
- Histone H2A (0.522, 0.966, 0.668).
- Myosin regulatory light chain 12A (0.966, 0.353, 0.727).
- Pre-mRNA-processing-splicing factor 8 (1.001, 1.154, 1.207).
- Protein RRP5 homolog (1.113, 1.034, 3.697).
- Myosin 10 (1.239, 1.049, 1.542).

4th – In all 3 experiments, similar C/T ratios observed, present in background in 3 experiments

5th –In 2 experiments a similar C/T ratio is observed, not present in background

- Plectin-1 (1.043, 1.116, 0.983).
- Thyroid hormone receptor-associated protein 3 (1.108, 0.988, 0.971).
- XRCC6 (1.001, 0.764, 0.781).
- XRCC5 (1.001, 1.086, 0.826).
- ATP dependent RNA helicase A (0.752, 0.944, 1.120).
- DNA ligase (0.351, 0.540, 1.163).
- DNA Topoisomerase 1 (0.648, 1.117, 1.242).
- Histone H2B type 1B (1.200, 0.644, 0.858).
- Uncharacterised - DKFZp781L0540 (1.181, 0.868, 0.984).
- Replication protein A 70kDa DNA binding protein (0.961, 0.986, 1.267).
- RNA binding motif 14 (0.687, 3.280, 0.978).
- Splicing factor U2AF (1.462, 0.911, 0.930).
- SWI/SNF related matrix associated dependent regulator of chromatin (0.833. 2.945, 0.861).

6th – In 2 experiments a similar C/T ratio is observed, present in background in 1 experiment

- Histone H1x (1.233, 0.923, 1.003).
- TATA-binding protein-associated factor 2N TAF15 (2.025, 0.261, 0.366).
- ATP binding cassette sub-family F member 2 (1.233, 0.501, 0.677).
- cDNA FLJ41699 – cytoskeleton associated protein 4 (0.533, 1.075, 0.978).
- Small nuclear ribonucleoprotein sm (2.925, 0.689, 0.581).
- rRNA methytranferase (0.882, 1.080, 1.231).
- pre-mRNA slicing factor ATP dependent RNA helicase DHX15 (0.472, 0.517, 1.320).
- DEAD box polypeptide 17 isoform p82 (0.994, 0.416, 1.355).
- Ewing sarcoma breakpoint region 1 (0.214, 0.290, 1.087).
- Exosome component 10 (0.573, 0.495, 1.019).
- Nucleolar and coiled body phosphoprotein 1 (0.831, 0.946, 1.136).
- Nucleolar transcription factor 1 (1.806, 0.504, 1.216).
- Periodic tryptophan protein 2 (0.595, 0.514, 1.210).
- Pre mRNA processing factor 19 (1.121, 1.310, 0.410).
- Uncharacterised - DKFZp667N107 (1.194, 0.570, 1.189).
- MSH2 (0.633, 0.299, 1.030).
- RPL32 (0.723, 0.568, 0.781).
- Ribosomal protein S8 (1.776, 1.149, 0.964)

- RuvB – like 1 (0.889, 1.283, 0.355).
- Transcription factor HES7 (0.996, 1.631, 0.992).
- WD repeat containing protein 3 (0.646, 0.281, 1.248).

7th – In 2 experiments a similar C/T ratio is observed, present in background in 2 experiments

- cDNA FLJ92633 – probable CCAAT box binding transcription factor CBF2 (0.994, 1.090, 1.140).
- ADP/ATP translocase 2 (1.707, 0.933, 1.053)
- SNARE protein Ykt6 (1.474, 1.254, 0.383).
- Ribosomal protein L14 (1.941, 1.310, 0.996).
- Exosome complex exonuclease MTR3 (1.033, 0.304, 1.039).
- Fusion isoform a variant (1.652, 0.973, 1.722).
- Gamma interferon inducible protein 16 – varied results, (n/a, 1.259, 1.141). BUT isoform CRA
- Myb binding domain 1A (1.287, 0.895, 1.248).
- Neuroblast differentiation associated protein AHNAK (0.887, 1.097, 1.086).
- Ribosomal protein L18 (1.388, 1.674, 0.467).
- RPS4X (1.837, 0.867, 1.046).
- Ribosomal protein L23 (1.382, 0.901, 0.675).
- THO complex subunit 4 (1.391, 0.992, 0.986)
- Tropomyosin isoform (0.948, 1.053, 0.619).
- Vimentin (0.463, 1.234, 1.294)

8th – In 2 experiments a similar C/T ratio is observed, present in background in 3 experiments

- Myosin light polypeptide 6 (2.159, 0.917, 0.557).
- RPL35A (1.097, 1.286, 0.628).

9th– Identified in 2 of the 3 experiments with a similar C/T ratio, not present in the background.

- BEN domain containing protein 3 (n/a, 0.543, 0.961).
- Poly(A)-binding protein 1 (n/a 1.108, 1.331).
- Scaffold attachment factor B (n/a, 1.088, 1.297).
- Zinc finger protein 638 (n/a, 1.012, 1.234).
- ATP dependent RNA helicase DDX3X (n/a, 6.206, 1.494).
- Histone H1.2 (n/a, 1.096, 1.298).
- Histone H2B type 1B C/E/F/G/I (n/a, 0.777, 0.826).
- Non POU domain containing octamer binding protein (n/a, 1.425, 2.455).
- PNKP (n/a, 1.382, 1.045).
- RPL36A (N/A, 0.561, 0.662).
- UBAP2L (N/A, 1.033, 1.195).
- RAD23 (N/A, 1.247, 3.001).
- Small nuclear ribonucleoprotein sm D2 (n/a, 0.630, 0.701).
- TCOF1 (n/a, 1.179, 1.079).

10th– Identified in 1 or 2 experiments, with a similar C/T ratio, present in the background of 1 or more experiments.

- CAPZB PROTEIN (n/a, 0.642, 0.850).
- cDNA FLJ27339 – TFIIH basal transcription factor complex helicase subunit (n/a, 0.758, 0.417). Also subunit xpb identified (n/a, 0.272, 0.814). And subunit p52 (1.026, n/a, n/a).
- cDNA FLJ33908 – DEAD/H box polypeptide 18 (n/a, 2.416, 1.283).
- cDNA FLJ38853 – lim domain and actin binding protein 1 (n/a, 0.446, 0.975).
- DNA topoisimerase 3 (n/a, 0.330, 0.882).
- Dolichyl-diphosphooligosaccharide proteinglycosyltransferase (5.663, 1.209).
- Heat shock protein beta 1 (1.093, 1.001).
- HSP 90-BETA (n/a, 0.294, 0.733).
- N-acetyl transferase 10 (n/a, 2.137, 1.446).
- Interleukin enhancer-binding factor 3 (n/a, 0.286, 0.733).
- Eukaryotic translation initiation factor 6 (1.672, 1.157).
- Heterochromatin protein-1 binding protein 3 (n/a, 0.515, 0.847).
- Hydroxysteroid (17 beta) dehydrogenase 4 variant (n/a, 0.303, 0.801).
- LIM domain 7 (n/a, 1.004, 2.121).
- Niban-like protein 1 (n/a, 1.524, 1.554).
- NOP2 (n/a, 0.587, 0.821).
- NOP 56 (n/a, 1.394, 1.665).
- Peroxiredoxin-1 (1.438, 1.069).
- PNAS-20 (n/a, 1.182, 1.138).
- PSIP1 (n/a, 3.600, 8.954).
- BCAP31 (n/a, 1.194, 1.117).
- Ribosomal protein L31 (n/a, 0.731, 0.679).
- Ribosomal Protein L7a (n/a, 0.692, 0.864).
- U3 small nucleolar RNA interacting protein 2 (n/a, 1.007, 2.812).

Document 2

Oligonucleotide pull down mass spectrometry results. Both the unmodified (sheet 1) and analysed mass spectrometry data (sheet 2) can be found on the excel file on the supplementary disc attached to this thesis.

Publications and Presentations

Conference attendances, Oral and Poster Presentations:

- OA Research Society International 2010 (Brussels).
- 8th International Conference on Bone Morphogenetic Proteins 2010 (Leuven) – Oral presentation and poster.
- Transcription and Disease Symposium 2011 (London).
- British Society of Matrix Biology (BSMB) 2011 (Newcastle) – Poster.
- OA Research Society International 2012 (Barcelona) – Oral presentation.
- 9th International Conference on Bone Morphogenetic Proteins 2012 (USA) – Oral presentation and poster.

Publications

- Dodd, A.W., Syddall, C.M., Loughlin, J. (2012) 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 doi: 10.1038/ejhg.2012.197.
- Syddall, C.M., Reynard, L.N., Young, D.A., Loughlin, J. The Identification of *Trans*-acting Factors That Regulate the Expression of *GDF5* via the Osteoarthritis Susceptibility SNP rs143383. Manuscript submitted.
- Reynard, L.N., Syddall, C.M., Xu, Y., Young, D.A., Loughlin, J. The allelic expression imbalance of the *GDF5* OA susceptibility SNP rs143383 is modulated by DNA methylation of the 5'UTR which affects binding of the SP1 and SP3 repressor proteins. Manuscript in preparation.

References

1. Arthritis Research UK. What is Arthritis? [Online] Available at: http://www.arthritisresearchuk.org/arthritis-information/conditions/arthritis/what-is-arthritis.aspx. [Accessed 7th August 2012]
2. Arthritis Care (2011). About Arthritis. [Online]. Available at: http://www.arthritiscare.org.uk/AboutArthritis [Accessed: 7th August 2012].
3. Arden, N. and Nevitt, M.C. (2006) Osteoarthritis: epidemiology. Best Pract Res Clin Rheumatol 20: 3-25.
4. Altman, R., Asch, E., Bloch, D., Bole, G., Borenstein, D., Brandt, K., Christy, W., Cooke, T. D., Greenwald, R., Hochberg, M., Howell, D., Kaplan, D., Koopman, W., Longley, S., Mankin, H., McShane, D. J., Medsger, T., Meenan, R., Mikkelsen, W., Moskowitz, R., Murphy, W., Rothschild, B., Segal, M., Sokoloff, L., Wolfe, F. (1986) Development of criteria for the classification and reporting of osteoarthritis: Classification of osteoarthritis of the knee. Arthritis Rheum 29: 1039-1049.
5. Solomon, L. (1984) Geographical and anatomical patterns of osteoarthritis. Br J Rheumatol. 23: 177-180.
6. McCormick, A., Fleming, D., Charlton, J. (1995) Morbidity statistics from general practices: fourth national study 1991-1992. Office of population censuses and surveys. Great Britain: Department of Health.
7. Reginster, J.Y. (2002) The prevelance and burden of arthritis. Rheumatology (Oxford). 41: 3-6.
8. Gupta, S., Hawker, G.A., Laporte, A., Croxford, R., Coyte, P.C. (2005) The economic burden of disabling hip and knee osteoarthritis (OA) from the perspective of individuals living with this condition. Rheumatology (Oxford) 44: 1531-1537.
9. Sharma, L., Kapoor, D., Issa, S. (2006) Epidemiology of osteoarthritis: an update. Curr Opin Rheumatol. 18: 147-156
10. Kellgren, J.H., Lawrence J.S. (1957) Radiographic assessment of osteoarthritis. Ann Rheum Dis. 16: 494-502.
11. Agnesi, F., Amrami, K., Frigo, C.A., Kaufman, K.R. (2009) Comparison of cartilage thickness with radiologic grade of knee osteoarthritis. Skeletal Radiol 37: 639-643.
12. Dieppe, P.A., Lohmander, L.S. (2005) Pathogenesis and management of pain in osteoarthritis. Lancet 365: 965-973.
13. Lin, E.H. (2008) Depression and Osteoarthritis. Am J Med 121: 16-19.
14. Nüesch, E., Dieppe, P., Reichenbach, S., Williams, S., Iff, S., Jüni, P. (2011) All cause and disease specific mortality in patients with knee or hip osteoarthritis: population based cohort study. BMJ 342: doi: 10.1136/bmj.d1165
15. Bellamy, N. (1982) Osteoarthritis - an evaluative index for clinical trials. MSc Dissertation, McMaster University: Hamilton, Ontario.
16. Hannan, M.T., Felson, D.T. Pincus, T. (2000) Analysis of the discordance between radiographic changes and knee pain in osteoarthritis of the knee. J Rheumatol 27: 1513-1517.
17. NICE. NICE clinical guidelines: CG59 Osteoarthritis: quick reference guide (2008) [Online] Available at http://guidance.nice.org.uk/CG59/QuickRefGuide/pdf/English [Accessed August 2012].

18. Burrage, P.S., Brinckerhoff, C.E. (2007) Molecular targets in osteoarthritis: metalloproteinases and their inhibitors. Curr Drug Targets 8: 293-303.
19. Vane, J.R., Botting, R.M. (1996) Mechanism of action of anti-inflammatory drugs. Int J Tissue React 102: 9-21.
20. Fung, H.B., Kirschenbaum, H.L. (1999) Selective cyclooxygenase-2 inhibitors for the treatment of arthritis. Clin Ther 21: 1131-1157.
21. Merck (2004). Merck announces voluntary worldwide withdrawal of VIOXX. [Press Release, online]. Whitehouse Station, N.J. Available online at www.vioxx.com/vioxx/documents/english/vioxx_press_release.pdf.
22. Wieland, H.A., Michaelis, M., Kirschbaum, B.J., Rudolphi, K.A. (2005) Osteoarthritis: an untreatable disease? Nat Rev Drug Discov 4: 331-344.
23. Ronn, K., Reischl, N., Gautier, E., Jacobi, M. (2011) Current surgical treatment of knee osteoarthritis. Arthritis 2011: doi: 10.1155/2011/454873.
24. Bruyere, O., Reginster, J. Y. (2007) Glucosamine and chondroitin sulfate as therapeutic agents for knee and hip osteoarthritis. Drugs Aging 24: 573-580.
25. Sawitzke, A.D., Shi, H., Finco, M.F., Dunlop, D.D, Harris, C.L., Singer, N.G., Bradley, J.D., Silver, D., Jackson, C.G. Lane, N.E., Oddis, C.V., Wolfe, F., Lisse, J., Furst, D.E., Bingham, C.O., Reda, D.J., Moskowitz, R.W., Williams, J.H, Clegg, D.O. (2010) Clinical efficacy and safety of glucosamine, chondroitin sulphate, their combination, celecoxib or placebo taken to treat osteoarthritis of the knee: 2-year results from GAIT. Ann Rheum Dis 69: 1459-1464.
26. Baragi, V.M., Baragi, V.M., Becher, G., Bendele, A.M., Biesinger, R., Bluhm, H., Boer, J., Deng, H., Dodd, R., Essers, M., Feuerstein, T., Gallagher, B.M Jr., Gege. C., Hochgürtel, M., Hofmann, M., Jaworski, A., Jin, L., Kiely, A., Korniski, B., Kroth, H., Nix, D., Nolte, B., Piecha, D., Powers, T.S., Richter, F., Schneider, M., Steeneck, C., Sucholeiki, I., Taveras, A., Timmermann, A., Van Veldhuizen, J., Weik, J., Wu, X., Xia, B. (2009) A new class of potent matrix metalloproteinase 13 inhibitors for potential treatment of osteoarthritis: Evidence of histologic and clinical efficacy without musculoskeletal toxicity in rat models. Arthritis Rheum 60: 2008-2018.
27. Alvarez-Soria, M.A., Herrero-Beaumont, G., Sanchez-Pernaute, O., Bellido, M., Largo, R. (2008) Diacerein has a weak effect on the catabolic pathway of human osteoarthritis synovial fibroblast—comparison to its effects on osteoarthritic chondrocytes. Rheumatol 47: 627-633.
28. Pelletier, J.P., Mineau, F., Fernandes, J.C., Duval, N., Martel-Pelletier,J. (1998) Diacerhein and rhein reduce the interleukin 1beta stimulated inducible nitric oxide synthesis level and activity while stimulating cyclooxygenase-2 synthesis in human osteoarthritic chondrocytes. J Rheumatol 25: 2417-2424.
29. Qi, Y., Feng, G., Yan, W. (2012) Mesenchymal stem cell-based treatment for cartilage defects in osteoarthritis. Mol Biol Rep 39: 5683-5689.
30. Chambord, P. (2010) Bone marrow mesenchymal stem cells: Historical overview and concepts. Hum Gene Ther 21: 1045-1056.
31. Pittenger, M.F., Mackay, A.M., Beck, S.C., Jaiswal, R.K., Douglas, R., Mosca, J.D., Moorman, M.A., Simonetti, D.W., Craig, S., Marshak, D.R. (1999) Multilineage potential of adult human mesenchymal stem cells. Science 284: 143-147.
32. Johnson, K., Zhu, S., Tremblay, M.S., Payette, J.N., Wang, Jianing, Bouchez, L.C., Meeusen, S., Althage, S., Cho, C.Y., Wu, X., Schultz, P.G. (2012) A Stem Cell-Based Approach to Cartilage Repair. Science 336: 717-721.
33. Feretti, C., Ripamonti, U. (2012) A critical appraisal of human bone tissue engineering. International Conference on Bone Morphogenetic Proteins, June 2012. Tahoe, USA.

34. Newman, A.P. (1998) Articular cartilage repair. Am J Sports Med 26: 309-324.
35. Jackson, D.W., Scheer, M.J., Simon, T.M. (2001) Cartilage Substitutes: Overview of basic science and treatment options. J Am Acad Orthop Surg 9: 37-52.
36. Chen, F.H., Rousche, K.T., Tuan, R.S. (2006) Technology Insight: adult stem cells in cartilage regeneration and tissue engineering. Nat Clin Pract Rheumatol 2: 373-382.
37. Martel-Pelletier, J., Boileau, C., Pelletier, J. P., Roughley, P. J. (2008) Cartilage in normal and osteoarthritis conditions. Best Pract Res Clin Rheumatol 22: 351-384.
38. Buckwalter, J.A., Mankin, H.J., Grodzinsky, A.J. (2005) Articular cartilage and osteoarthritis. Instr Course Lect 54: 465-480.
39. Pearle, A.D., Warren, R.F., Rodeo, S.A. (2005) Basic science of articular cartilage and osteoarthritis. Clin Sports Med 24: 1-12.
40. Knudson, C.B., Knudson, W. (2001) Cartilage proteoglycans. Semin Cell Dev Biol 12: 69-78.
41. Wong, M., Wuethrich, P., Eggli, P., Hunziker, E. (1996) Zone-specific cell biosynthetic activity in mature bovine articular cartilage: a new method using confocal microscopic stereology and quantitative autoradiography. J Orthop Res 14: 424-432.
42. Brocklehurst, R., Bayliss, M.T., Maroudas, A., Coysh, H.L., Freeman, M.A.R., Revell, P.A., Ali, S.Y. (1984) The composition of normal and osteoarthritis articular cartilage from human knee joints. J Bone Joint Surg Am 66: 95-107.
43. Baron, R. (2008) Anatomy and ultrastructure of bone - Histogenesis, growth and remodelling. In: Arnold, A. Diseases of bone and mineral metabolism. South Dartmouth, MA. Avaliable online at http://www.endotext.org/parathyroid/parathyroid1/parathyroid1.html. Accessed July 2012.
44. Buckwalter, J.A., Cooper, R.R. (1987) Bone structure and function. Instr Course Lect 36: 27-48.
45. Allan, D.A. (1998) Structure and physiology of joints and their relationship to repetitive strain injuries. Clin Orthop Relat Res 351: 32-38.
46. Jones, E.A., Crawford, A., English, A., Henshaw, K., Mundy, J., Corscadden, D., Chapman, T., Emery, P., Hatton, P., McGonagle, D. (2008) Synovial fluid mesenchymal stem cells in health and early osteoarthritis: Detection and functional evaluation at the single-cell level. Arthritis Rheum 58: 1731-1740.
47. Woo, S.Y., Fenwick, J.A., Kanamori, A., Gil, J.E., Chan Saw, S.S., Vogrin, T.M. (2001) Biomechanical considerations of joint function. In: Rubash, H.E. Osteoarthritis: Diagnosis and Medical/Surgical Management. 3rd Ed, Philadelphia: WB Saunders.
48. Klein-Wieringa, I.R., Kloppenburg, M., Bastiaansen-Jenniskens, Y. M., Yusuf, E., Kwekkeboom, J. C., El-Bannoudi, H., Nelissen, R. G. H. H., Zuurmond, A., Stojanovic-Susulic, V., Van Osch, G. J. V. M., Toes, R. E. M., Ioan-Facsinay, A. (2011) The infrapatellar fat pad of patients with osteoarthritis has an inflammatory phenotype. Ann Rheum Dis 70: 851-857.
49. Ghosh, P., Taylor, T.K. (1987) The knee joint meniscus. A fibrocartilage of some distinction. Clin Orthop Relat Res 224: 52-63.
50. Curtiss, P.H. (1964) Changes produced in the synovial membrane and synovial fluid by disease. J Bone Joint Surg Am 46: 873-888.
51. Goldring, M.B., Marcu, K.B. (2009) Cartilage homeostasis in health and rheumatic diseases. Arthritis Res Ther 11: doi:10.1186/ar2592.

52. Glasson, S.S., Askew, R., Sheppard, B., Carito, B., Blanchet, T., Ma, H.L., Flannery, C.R., Peluso, D., Kanki, K., Yang, Z., Majumdar, M.K., Morris, E.A. (2005) Deletion of active ADAMTS5 prevents cartilage degradation in a murine model of osteoarthritis. Nature 434: 644-648.
53. Shlopov, B.V., Lie, W.R., Mainardi, C.L., Cole, A.A., Chubinskaya, S., Hasty, K.A. (1997) Osteoarthritic lesions: involvement of three different collagenases. Arthritis Rheum 40: 2065-2074.
54. Billinghurst, R.C., Dahlberg, L., Ionescu, M., Reiner, A., Bourne, R., Rorabeck, C., Mitchell, P., Hambor, J., Diekmann, O., Tschesche, H., Chen, J., Van Wart, H., Poole, A.R. (1997) Enhanced cleavage of type II collagen by collagenases in osteoarthritic articular cartilage. J Clin Invest 99: 1534-1545.
55. Visse, R., Nagase, H. (2003) Matrix metalloproteinases and tissue inhibitors of metalloproteinases. Circulation Research 92: 827-839.
56. Brew, K., Dinakarpandian, D., Nagase, H. (2000) Tissue inhibitors of metalloproteinases: evolution, structure and function. Biochim Biophys Acta 1477: 267-283.
57. Yoon, B.S., Pogue, R., Ovchinnikov, D.A., Yoshii, I., Mishina, Y., Behringer, R.R., Lyons, K.M. (2006) BMPs regulate multiple aspects of growth-plate chondrogenesis through opposing actions on FGF pathways. 133: 4667-4678.
58. Bi, W., Deng, J.M., Zhang, Z., Behringer, R.R., de Crombrugghe, B. (1999) Sox9 is required for cartilage formation. Nat Genet 22: 85-89.
59. Dong, Y., Drissi, H., Chen, M., Chen, D., Zuscik, M.J., Schwarz, E.M., O'Keefe, R.J. (2005) Wnt-mediated regulation of chondrocyte maturation: Modulation by TGF-β. J Cell Biochem 95: 1057-1068.
60. Komori, T. (2005) Regulation of skeletal development by the Runx family of transcription factors. J Cell Biochem 95: 445-453.
61. Brandt, K.D., Radin, E. L., Dieppe, P. A., van de Putte, L. (2006) Yet more evidence that osteoarthritis is not a cartilage disease. Ann Rheum Dis 65: 1261-1264.
62. Bailey, A.J., Mansell, J.P. (1997) Do subchondral bone changes exacerbate or precede articular cartilage destruction in osteoarthritis of the elderly? Gerentology 43: 296-304.
63. Samuels, J., Krasnokutsky, S., Abramson, S.B. (2008) Osteoarthritis: a tale of three tissues. Bull NYU Hosp JT Dis 66: 244-250.
64. Felson, D.T., Neogi, T. Osteoarthritis: Is it a disease of cartilage or of bone? (2004) Arthritis Rheum 50: 341-344.
65. Goldring, M., Goldring, S. (2007) Osteoarthritis. J Cell Physiol 213: 626-634.
66. Blanco, F.J., Guitian, R., Vázquez-Martul, E., de Toro, F.J., Galdo, F. (1998) Osteoarthritis chondrocytes die by apoptosis. A possible pathway for osteoarthritis pathology. Arthritis Rheum 41: 284-289.
67. Martel-Pelletier, J., Boileau, C., Pelletier, J. P., Roughley, P. J. (2008) Cartilage in normal and osteoarthritis conditions. Best Pract Res Clin Rheumatol 22: 351-384.
68. Shingu, M., Nagai, Y., Isayama, T., Naono, T., Nobunaga, M., Nagai, Y. (1993) The effects of cytokines on metalloproteinase inhibitors (TIMP) and collagenase production by human chondrocytes and TIMP production by synovial cells and endothelial cells. Clin Exp Immunol 94: 145-149.
69. Fukui, N., Miyamoto, Y., Nakajima, M., Ikeda, Y., Hikita, A., Furukawa, H., Mitomi, H., Tanaka, N., Katsuragawa, Y., Yamamoto, S., Sawabe, M., Juji, T., Mori, T., Suzuki, R., Ikegawa, S. (2008) Zonal gene expression of chondrocytes in osteoarthritic cartilage. Arthritis Rheum 58: 3843-3853.

70. Aigner, T., Soder, S., Gebhard, P.M., McAlinden, A., Haag, J. (2007) Mechanisms of disease: role of chondrocytes in the pathogenesis of osteoarthritis--structure, chaos and senescence. Nat Clin Pract Rheumatol 3: 391-399.
71. Grynpas, M.D., Alpert, B., Katz, I., Lieberman, I., Pritzker, K.P. (1991) Subchondral bone in osteoarthritis. Calcif Tissue Int 49: 20-26.
72. Brandt, K.D. (1999) Osteophytes in osteoarthritis. Clinical Aspects. Osteoarthritis Cartilage 7: 334-335.
73. Van Osch, G.J., Van der Kraan, P.M., Van Valburg, A.A., Van de Berg, W.B. (1996) The relation between cartilage damage and osteophyte size in a murine model for osteoarthritis in the knee. Rheumatol Int 16: 115-119.
74. Ondrouch, A.S. (1963) Cyst formation in osteoarthritis. J Bone Joint Surg Br 45: 755-760.
75. Sharif, M., Saxne, T., Shepstone, L., Kirwan, J.R., Elson, C.J., Heinegard, D., Dieppe, P.A. (1995) Relationship between serum cartilage oligomeric matrix protein levels and disease progression in osteoarthritis of the knee joint. Br J Rheumatol 34: 306-310.
76. Pelletier, J.P., Pelletier, M. P., Abramson, S.B. (2001) Osteoarthritis, an inflammatory disease: potential implication for the selection of new therapeutic targets. Arthritis and Rheum 44: 1237-1247.
77. Fernandes, J.C., Martel-Pelletier, J., Pelletier, J.P. (2002) The role of cytokines in osteoarthritis pathophysiology. Biorheology 39: 237-246.
78. Cawston, T.E., Ellis, A.J., Humm, G., Lean, E., Ward, D., Curry, V. (1995) Interleukin-1 and oncostatin M in combination promote the release of collagen fragments from bovine nasal cartilage in culture. Biochem Biophys Res Commun 215: 377-385.
79. Noble, J., Hamblen, D.L. (1975) The pathology of the degenerate meniscus lesion. J Bone Joint Surg Br 57: 180-186.
80. Lloyd-Roberts, G.C. (1953) The role of capsular changes in osteoarthritis of the hip joint. J Bone Joint Surg Br 35: 627-642.
81. Hurley, M.V., Scott, D.L., Rees, J., Newham, D.J. (1997) Sensorimotor changes and functional performance in patients with knee osteoarthritis. Ann Rheum Dis 56: 641-648.
82. Tan, A.L., Toumi, H., Benjamin, M., Grainger, A.J., Tanner, S.F., Emery, P., McGonagle, D. (2006) Combined high-resolution magnetic resonance imaging and histological examination to explore the role of ligaments and tendons in the phenotypic expression of early hand osteoarthritis. Ann Rheum Dis 65:1267-1272.
83. Verbruggen, G., Cornelissen, M., Almqvist, K.F., Wang, L., Elewaut, D., Broddelez, C., de Ridder, L., Veys, E.M. (2000) Influence of aging on the synthesis and morphology of the aggrecans synthesized by differentiated human articular chondrocytes. Osteoarthritis Cartilage 8: 170-179.
84. Eyre, D.R., Dickson, I.R., Van Ness, K. (1988) Collagen cross-linking in human bone and articular cartilage. Age-related changes in the content of mature hydroxypyridinium residues. J Biochem 252: 495-500.
85. Fujii, K., Kuboki, Y., Sasaki, S. (1976) Aging of human bone and articular cartilage collagen: changes in the reducible cross-links and their precursors. Gerentology 22: 363-370.
86. Verzijl, N., DeGroot, J., Thorpe, S. R., Bank, R. A., Shaw, J. N., Lyons, T. J., Bijlsma, J. W., Lafeber, F. P., Baynes, J. W., TeKoppele, J. M. (2000) Effect of collagen turnover on the accumulation of advanced glycation end products. J Biol Chem 275: 39027-39031.

87. Forsyth, C.B., Cole, A., Murphy, G., Bienias, J.L., Im, H.J., Loeser, R.F.Jr. (2005) Increased matrix metalloproteinase-13 production with aging by human articular chondrocytes in response to catabolic stimuli. J Gerontol A Biol Sci Med Sci 60: 1118-1124.
88. Srikanth, V.K., Fryer, J.L., Zhai, G., Winzenberg, T.M., Hosmer, D., Jones, G. (2005) A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. Osteoarthritis Cartilage 13: 769-781.
89. Hanna, F.S., Wluka, A.E., Bell, R.J., Davis, S.R., Cicuttini, F.M. (2004) Osteoarthritis and the postmenopausal woman: Epidemiological, magnetic resonance imaging, and radiological findings. Semin Arthritis Rheum 34: 631-636.
90. Cirillo, D.J., Wallace, R.B., Wu, L, Yood, R.A. (2006) Effect of hormone therapy on risk of hip and knee joint replacement in the women's health initiative. Arthritis Rheum 54: 3194-3204.
91. Nevitt, M.C., Felson, D.T., Williams, E.N., Grady, D. (2001) The effect of estrogen plus progestin on knee symptoms and related disability in postmenopausal women: The heart and estrogen/progestin replacement study, a randomized, double-blind, placebo-controlled trial. Arthritis Rheum 44: 811-818.
92. Maleki-Fischbach, M., Jordan, J.M. (2010) New developments in osteoarthritis. Sex differences in magnetic resonance imaging-based biomarkers and in those of joint metabolism. Arthritis Res Ther 12: doi: 10.1186/ar3091.
93. Guilak, F., Fermor, B., Keefe, F.J., Kraus, V.B., Olson, S.A., Pisetsky, D.S., Setton, L.A., Weinberg, J.B. (2004) The role of biomechanics and inflammation in cartilage injury and repair. Clin Orthop Relat Res 423: 17-26.
94. Felson, D.T., Anderson, J.J., Naimark, A., Walker, A.M., Meenan, R.F. (1988) Obesity and knee osteoarthritis. The Framingham Study. Ann Intern Med 109: 18-24.
95. Blagojevic, M.J., C., Jeffery, A., Jordan, K.P. (2010) Risk factors for onset of osteoarthritis of the knee in older adults: a systematic review and meta-analysis. Osteoarthritis Cartilage 18: 24-33.
96. Felson, D.T., Zhang, Y., Anthony, J.M., Naimark, A., Anderson, J.J. (1992) Weight loss reduces the risk for symptomatic knee osteoarthritis in women. Ann Intern Med 116: 535-539.
97. Carman, W.J., Sowers, M., Hawthorne, V.M., Weissfeld, L.A. (1994) Obesity as a risk factor for osteoarthritis of the hand and wrist: a prospective study. Am J Epidemiol 139: 119-129.
98. Sandell, L. (2009) Obesity and Osteoarthritis: Is leptin the link? Arthritis Rheum 60: 2858-2860.
99. McAlindon, T., Felson, D.T. (1997) Nutrition: risk factors for osteoarthritis. Ann Rheum Dis 56: 397-400.
100. Nevitt, M.C., Zhang, Y., Javaid, M.K., Neogi, T., Curtis, J.R., Niu, J., McCulloch, C.E., Segal, N.A., Felson, D.T. (2010) 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 69: 163-168.
101. Amin, S., Goggins, J., Niu, J., Guermazi, A.L.I., Grigoryan, M., Hunter, D.J., Genant, H.K., Felson, D.T. (2008) Occupation-related squatting, kneeling, and heavy lifting and the knee joint: a magnetic resonance imaging-based study in men. J Rheumatol 35: 1645-1649.
102. Slemenda, C., Heilman, D.K., Brandt, K.D., Katz, B.P., Mazzuca, S.A., Braunstein, E.M., Byrd, D. (1998) Reduced quadriceps strength relative to body

weight: a risk factor for knee osteoarthritis in women? Arthritis Rheum 41: 1951-1959.
103. Sharma, L., Song, J., Dunlop, D., Felson, D., Lewis, C.E., Segal, N., Torner, J., Cooke, T.D., Hietpas, J., Lynch, J., Nevitt, M. (2010) Varus and valgus alignment and incident and progressive knee osteoarthritis. Ann Rheum Dis 69: 1940-1945.
104. Golightly, Y.M., Allen, K.D., Helmick, C.G., Schwartz, T.A., Renner, J.B., Jordan, J.M. (2010) Hazard of incident and progressive knee and hip radiographic osteoarthritis and chronic joint symptoms in individuals with and without limb length inequality. J Rheumatol 37: 2133-2140.
105. Gregory, J.S., Waarsing, J.H., Day, J., Pols, H.A., Reijman, M., Weinans, H., Aspden, R.M. (2007) Early identification of radiographic osteoarthritis of the hip using an active shape model to quantify changes in bone morphometric features: Can hip shape tell us anything about the progression of osteoarthritis? Arthritis Rheum 56: 3634-3643.
106. Nevitt, M.C., Lane, N.E., Scott, J.C., Hochberg, M.C., Pressman, A.R., Genant, H.K., Cummings, S.R. (1995) Radiographic osteoarthritis of the hip and bone mineral density. The study of osteoporotic fractures research group. Arthritis Rheum 38: 907-916.
107. Nevitt, M.C., Xu, L., Zhang, Y., Lui, L.Y., Yu, W., Lane, N.E., Qin, M., Hochberg, M.C., Cummings, S.R., Felson, D.T. (2002) 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 46: 1773-1779.
108. Zhang, Y., Xu, L., Nevitt, M.C., Aliabadi, P., Yu, W., Qin, M., Lui, L.Y., Felson, D.T. (2001) 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 44: 2065-2071.
109. Zhang, Y., Xu, L., Nevitt, M.C., Niu, J., Goggins, J.P., Aliabadi, P., Yu, W., Lui, L.Y., Felson, D.T. (2003) Lower prevalence of hand osteoarthritis among Chinese subjects in Beijing compared with white subjects in the United States: The Beijing Osteoarthritis Study. Arthritis Rheum 48: 1034-1040.
110. Stecher, R.M. (1941) HEBERDEN'S NODES: Heredity in Hypertrophic Arthritis of the Finger Joints. Am J Med Sci 201: 801-809.
111. Hirsch, R., Lethbridge-Cejku, M., Hanson, R., Scott, W.W. Jr., Reichle, R., Plato, C.C., Tobin, J.D., Hochberg, M.C. (1998) Familial aggregation of osteoarthritis: data from the Baltimore Longitudinal Study on Aging. Arthritis Rheum 41: 1227-1232.
112. Bijkerk, C., Houwing-Duistermaat, J.J., Valkenburg, H.A., Meulenbelt, I., Hofman, A., Breedveld, F.C., Pols, H.A.P., Van Duijn, C.M., Slagboom, P.E. (1999) Heritabilities of radiologic osteoarthritis in peripheral joints and of disc degeneration of the spine. Arthritis Rheum 42: 1729-1735.
113. Neame, R.L., Muir, K., Doherty, S., Doherty, M. (2004) Genetic risk of knee osteoarthritis: a sibling study. Ann Rheum Dis 63: 1022-1027.
114. Chitnavis, J., Sinsheimer, J. S., Clipsham, K., Loughlin, J., Sykes, B., Burge, P., D., Carr, A.J. (1997) Genetic influences in end-stage osteoarthritis - Sibling risks of hip and knee replacement for idiopathic osteoarthritis. J Bone Joint Surg Br 79: 660-664.
115. Lanyon, P., Muir, K., Doherty, S., Doherty, M. (2000) Assessment of a genetic contribution to osteoarthritis of the hip: sibling study. BMJ 321: 1179-1183.
116. Bukulmez, H., Matthews, A.L., Sullivan, C.M., Chen, C., Kraay, M.J., Elston, R.C., Moskowitz, R.W., Goldberg, V.M., Warman, M.L. (2006) Hip joint

replacement surgery for idiopathic osteoarthritis aggregates in families. Arthritis Res Ther 8: R25.
117. Riyazi, N., Meulenbelt, I., Kroon, H. M., Ronday, K.H., Hellio le Graverand, M.P., Rosendaal, F.R., Breedveld, F.C., Slagboom, P.E., Kloppenburg, M. (2005) Evidence for familial aggregation of hand, hip, and spine but not knee osteoarthritis in siblings with multiple joint involvement: the GARP study. Ann Rheum Dis 64: 438-443.
118. Spector, T.D., MacGregor, A.J. (2004) Risk factors for osteoarthritis: genetics. Osteoarthritis Cartilage 12: 39-44.
119. Spector, T.D., Cicuttini, J., Baker, J., Loughlin, J., Hart, D. (1996) Genetic influences on osteoarthritis in women: a twin study. BMJ 312: 940-943.
120. Zhai, G., Hart, D.J., Kato, B.S., MacGregor, A., Spector, T. D. (2007) Genetic influence on the progression of radiographic knee osteoarthritis: a longitudinal twin study. Osteoarthritis Cartilage 15: 222-225.
121. MacGregor, A.J., Antoniades, L., Matson, M., Andrew, T., Spector. T.D. (2000) The genetic contribution to radiographic hip osteoarthritis in women: Results of a classic twin study. Arthritis Rheum 43: 2410-2416.
122. Page, W.F., Hoaglund, F.T., Steinbach, L.S., Heath, A.C. (2003) Primary osteoarthritis of the hip in monozygotic and dizygotic male twins. Twin Res 6: 147-151.
123. Sambrook, P.N., MacGregor, A.J., Spector, T.D. (1999) Genetic influences on cervical and lumbar disc degeneration: A magnetic resonance imaging study in twins. Arthritis Rheum 42: 366-372.
124. Reginato, A.M., Olsen, B.R. (2002) The role of structural genes in the pathogenesis of osteoarthritic disorders. Arthritis Res 4: 337-345.
125. Baldwin, C.T., Cupples, L.A., Joost, O., Demissie, S., Chaisson, C., McAlindon, T., Myers, R.H., Felson, D. (2002) Absence of linkage or association for osteoarthritis with the vitamin D receptor/type II collagen locus: the Framingham Osteoarthritis Study. J Rheumatol 29: 161-165.
126. Loughlin, J. (2001) Genetic epidemiology of primary osteoarthritis. Curr Opin Rheumatol 13: 111-116.
127. Uitterlinden, A.G., Burger, H., van Dujin, C.M., Huang, Q., Hofman, A., Birkenhager, J.C., van Leeuwen, J.P.T.M., Pols, H.A.P. (2000) Adjacent genes for COL2A1 and the vitamin D receptor are associated with the separate features of radiographic osteoarthritis of the knee. Arthritis Rheum 43: 1456-1464.
128. Kizawa, H., Kou, I., Iida, A., Sudo, A., Miyamoto, Y., Fukuda, A., Mabuchi, A., Kotani, A., Kawakami, A., Yamamoto, S., Uchida, A., Nakamura, K., Notoya, K., Nakamura, Y., Ikegawa, S. (2005) An aspartic acid repeat polymorphism in asporin inhibits chondrogenesis and increases susceptibility to osteoarthritis. Nat Genet 37: 138-144.
129. Mustafa, Z., Dowling, B., Chapman, K., Sinsheimer, J.S., Carr, A., Loughlin, J. (2005) Investigating the aspartic acid (D) repeat polymorphism of asporin as a risk factor for osteoarthritis in a UK Caucasian population. Arthritis Rheum 52: 3502-3506.
130. Kaliakatsos, M., Tzetis, M., Kanavakis, E., Fytili, P., Chouliaras, G., Karachalios, T., Malizos, K., Tsezou, A. (2006) Asporin and knee osteoarthritis in patients of Greek Origin. Osteoarthritis Cartilage 14: 609-611.
131. Rodriguez-Lopez, J., Pombo-Suarez, M., Liz, M., Gomez-Reino, J.J., Gonzalez, A. (2006) Lack of association of a variable number of aspartic acid residues in the asporin gene with osteoarthritis susceptibility: case-control studies in Spanish Caucasians. Arthritis Res Ther 8: R55.

132. Jiang, Q., Shi, D., Yi, L., Ikegawa, S., Wang, Y., Nakamura, T., Qiao, D., Liu, C., Dai, J. (2006) Replication of the association of the aspartic acid-repeat polymorphism in the asporin gene with knee-osteoarthritis susceptibility in Han Chinese. J Hum Genet 51: 1068-1072.
133. Nakamura, T., Shi, D. Tzetis, M., Rodriguez-Lopez, J., Miyamoto, Y., Tsezou, A., Gonzalez, A., Jiang, Q., Kamatani, N., Loughlin, J., Ikegawa, S. (2007) Meta-analysis of association between the ASPN D-repeat and osteoarthritis. Hum Mol Genet 16: 1676-1681.
134. Miyamoto, Y., Mabuchi, A., Shi, D., Kubo, T., Takatori, Y., Saito, S., Fujioka, M., Sudo, A., Uchida, A., Yamamoto, S., Ozaki, K., Takigawa, M., Tanaka, T., Nakamura, Y., Jiang, Q., Ikegawa, S. (2007) A functional polymorphism in the 5′UTR of GDF5 is associated with susceptibility to osteoarthritis. Nat Genet 39: 529-533.
135. Loughlin, J. (2011) Genetic indicators and susceptibility to osteoarthritis. Br J Sports Med 45: 278-282.
136. Chapman, K., Mustafa, Z., Irven, C., Carr, A.J., Clipsham, K., Smith, A., Chitnavis, J., Sinsheimer, J.S., Bloomfield, V.A., McCartney, M., Cox, O., Cardon, L.R., Sykes, B., Loughlin,J. (1999) Osteoarthritis-susceptibility locus on chromosome 11q, detected by linkage. Am J Hum Genet 65: 167-174.
137. Loughlin, J., Mustafa, Z., Smith, A., Irven, C., Carr, A.J., Clipsham, K., Chitnavis, J., Bloomfield, V.A., McCartney, M., Cox, O., Sinsheimer, J.S., Sykes, B., Chapman, K.E. (2000) Linkage analysis of chromosome 2q in osteoarthritis. Rheumatology (Oxford) 39: 377-381.
138. Loughlin, J., Mustafa, Z., Irven, C., Smith, A., Carr, A.J., Sykes, B., Chapman, K. (1999) Stratification analysis of an osteoarthritis genome screen - suggestive linkage to chromosomes 4, 6, and 16. Am J Hum Genet 65: 1795-1798.
139. Loughlin, J., Dowling, B., Mustafa, Z., Southam, L., Chapman, K. (2002) Refined linkage mapping of a hip osteoarthritis susceptibility locus on chromosome 2q. Rheumatology (Oxford) 41: 955-956.
140. Forster, T., Chapman, K., Marcelline, L., Mustafa, Z., Southam, L., Loughlin, J. (2004) Finer Linkage mapping of primary osteoarthritis susceptibility loci on chromosomes 4 and 16 in families with affected women. Arthritis Rheum 50: 98-102.
141. Loughlin, J., Mustafa, Z., Dowling, B., Southam, L., Marcelline, L., Raina, S.S., Ala-Kokko, L., Chapman, K. (2002) Finer linkage mapping of a primary hip osteoarthritis susceptibility locus on chromosome 6. Eur J Hum Genet 9: 562-568.
142. Chapman, K., Mustafa, Z., Dowling, B., Southam, L., Carr, A.J., Loughlin, J. (2002) Finer linkage mapping of primary hip osteoarthritis susceptibility on chromosome 11q in a cohort of affected female sibling pairs. Arthritis Rheum 46: 1780-1783.
143. Loughlin, J., Dowling, B., Chapman, K., Marcelline, L., Mustafa, Z., Southam, L., Ferreira, A., Ciesielski, C., Carson, D. A., Corr, M. (2004) Functional variants within the secreted frizzled-related protein 3 gene are associated with hip osteoarthritis in females. Proc Natl Acad Sci USA 101: 9757-9762.
144. Hoang, B., Moos, M., Vukicevic, S., Luyten, F.P. (1996) Primary structure and tissue distribution of FRZB, a novel protein related to Drosophila frizzled, suggest a role in skeletal morphogenesis. J Biol Chem 271: 26131-26137.
145. Snelling S, Ferreira A, Loughlin J. (2006) Allelic expression analysis suggests that cis-acting polymorphism of FRZB expression does not contribute to osteoarthritis susceptibility. Osteoarthritis Cartilage 15: 90-92.

146. Valdes, A.M., Loughlin, J., Oene, M.V., Chapman, K., Surdulescu, G.L., Doherty, M., Spector, T.D. (2007) Sex and ethnic differences in the association of ASPN, CALM1, COL2A1, COMP, and FRZB with genetic susceptibility to osteoarthritis of the knee. Arthritis Rheum 56: 137-146.
147. Rodriguez-Lopez, J., Pombo-Suarez, M., Liz, M., Gomez-Reino, J.J., Gonzalez, A. (2007) Further evidence of the role of frizzled-related protein gene polymorphisms in osteoarthritis. Ann Rheum Dis 66: 1052-1055.
148. Evangelou, E., Chapman, K., Meulenbelt, I., Karassa, F.B., Loughlin, J., Carr, A., Doherty, M., Doherty, S., Gomez-Reino, J.J., Gonzalez, A., Halldorsson, B.V., Hauksson, V.B., Hofman, A., Hart, D.J., Ikegawa, S., Ingvarsson, T., Jiang, Q., Jonsdottir, I., Jonsson, H., Kerkhof, H.J., Kloppenburg, M., Lane, N.E., Li, J., Lories, R.J., van Meurs, J.B., Nakki, A., Nevitt, M.C., Rodriguez-Lopez, J., Shi, D., Slagboom, P.E., Stefansson, K., Tsezou, A., Wallis, G.A., Watson, C.M., Spector, T.D., Uitterlinden, A.G., Valdes, A.M., Ioannidis, J.P. (2009) Large-scale analysis of association between GDF5 and FRZB variants and osteoarthritis of the hip, knee, and hand. Arthritis Rheum 60: 1710-1721.
149. Kerkhof, J.M., Uitterlinden, A.G., Valdes, A.M., Hart, D.J., Rivadeneira, F., Jhamai, M., Hofman, A., Pols, H.A.P., Bierma-Zeinstra, S.M.A., Spector, T.D., van Meurs, J. B. (2008) Radiographic osteoarthritis at three joint sites and FRZB, LRP5, and LRP6 polymorphisms in two population-based cohorts. Osteoarthritis Cartilage 16: 1141-1149.
150. Waarsing, J.H., Kloppenburg, M., Slagboom, P.E., Kroon, H.M., Houwing-Duistermaat, J.J., Weinans, H., Meulenbelt, I. (2011) Osteoarthritis susceptibility genes influence the association between hip morphology and osteoarthritis. Arthritis Rheum. 63: 1349-1354.
151. Baker-LePain, J.C., Lynch, J. A., Parimi, N., McCulloch, C. E., Nevitt, M. C., Corr, M., Lane, N. E. (2012) Variant alleles of the Wnt antagonist FRZB are determinants of hip shape and modify the relationship between hip shape and osteoarthritis. Arthritis Rheum 64: 1457-1465.
152. Lories, R.J., Peeters, J., Bakker, A., Tylzanowski, P., Derese, I., Schrooten, J., Thomas, J. T. & Luyten, F. P. (2007) Articular cartilage and biomechanical properties of the long bones in Frzb-knockout mice. Arthritis Rheum 56: 4095-4103.
153. Leppavuori, J., Kujala, U., Kinnunen, J., Kaprio, J., Nissila, M., Heliavaara, M., Klinger, N., Partanen, J., Terwilliger, J. D., Peltonen, L. (1999) Genome scan for predisposing loci for distal interphalangeal joint osteoarthritis: evidence for a locus on 2q. Am J Hum Genet 65: 1060-1067.
154. Loughlin, J., Dowling, B., Mustafa, Z., Chapman, K. (2002) Association of the interleukin-1 gene cluster on chromosome 2q13 with knee osteoarthritis. Arthritis Rheum 46: 1519-1527.
155. Meulenbelt, I., Seymour, A. B., Nieuwland, M., Huizinga, T. W. J., Van Duijn, C. M., Slagboom, P. E. (2004) Association of the interleukin-1 gene cluster with radiographic signs of osteoarthritis of the hip. Arthritis Rheum 50: 1179-1186.
156. Näkki, A., Kouhia, S.T., Saarela, J., Harilainen, A., Tallroth, K., Videman, T., Battié, M.C., Kaprio, J., Peltonen, L., Kujala, U.M. (2010) Allelic variants of IL1R1 gene associate with severe hand osteoarthritis. (2010) BMC Med Genet 11: doi:10.1186/1471-2350-11-50.
157. Solovieva, S., Kämäräinen, O.P., Hirvonen, A., Hämäläinen, S., Laitala, M., Vehmas, T., Luoma, K., Näkki, A., Riihimäki, H., Ala-Kokko, L., Männikkö, M., Leino-Arjas, P. (2009) Association between interleukin 1 gene cluster polymorphisms and bilateral distal interphalangeal osteoarthritis. J Rheumatol 36: 1977-1986.

158. Moos, V., Rudwaleit, M., Herzog, V., Höhlig, K., Sieper, J., Müller, B. (2000) Association of genotypes affecting the expression of interleukin-1β or interleukin-1 receptor antagonist with osteoarthritis. Arthritis Rheum 43: 2417-2422
159. Smith, A.J.P., Keen, L. J., Billingham, M. J., Perry, M. J., Elson, C. J., Kirwan, J. R., Sims, J. E., Doherty, M., Spector, T. D., Bidwell, J. L. (2004) Extended haplotypes and linkage disequilibrium in the IL1R1-IL1A-IL1B-IL1RN gene cluster: association with knee osteoarthritis. Genes Immun 5: 451-460.
160. Stankovich, J., Sale, M.M., Cooley, H.M., Bahlo, M., Reilly, A., Dickinson, J.L., Jones, G. (2002) Investigation of chromosome 2q in osteoarthritis of the hand: no significant linkage in a Tasmanian population. Ann Rheum Dis 61: 1081-1084.
161. Jotanovic, Z., Etokebe, G., Mihelic, R., Kaarvatn, M., Mulac-Jericevic, B., Tijanic, T., Balen, S., Sestan, B., Dembic, Z. (2012) IL1B -511(G>A) and IL1RN (VNTR) allelic polymorphisms and susceptibility to knee osteoarthritis in Croatian population. Rheumatol Int 32: 2135-2141.
162. Ingvarsson, T., Stefánsson, S.E., Gulcher, J.R., Jónsson, H.H., Jónsson, H., Frigge, M.L., Pálsdóttir, E., Olafsdóttir, G., Jónsdóttir, T., Walters, G,B., Lohmander, L.S., Stefánsson, K. (2001) A large Icelandic family with early osteoarthritis of the hip associated with a susceptibility locus on chromosome 16p. Arthritis Rheum 44: 2548-2555.
163. Forster, T., Chapman, K., Loughlin, J. (2004) Common variants within the interleukin 4 receptor gene (IL4R) are associated with susceptibility to osteoarthritis. Hum Genet 114: 391-395.
164. Hunter, D.J., Demissie, S., Cupples, L. A., Aliabadi, P., Felson, D. T. (2004) A genome scan for joint-specific hand osteoarthritis susceptibility: The Framingham Study. Arthritis Rheum 50: 2489-2496.
165. Williams, F.M.K., Kato, B. S., Livshits, G., Sambrook, P. N., Spector, T. D., Macgregor, A. J. (2008) Lumbar disc disease shows linkage to chromosome 19 overlapping with a QTL for hand OA. Ann Rheum Dis 67: 117-119.
166. International HapMap Consortium. (2003) The International HapMap Project. Nature 426: 789-796.
167. Human Genome Sequencing. (2004) Finishing the euchromatic sequence of the human genome. Nature 431: 931-945.
168. Kerkhof, H.J.M., Lories, R.J.U., Meulenbelt, I., Jonsdottir, I., Valdes, A.M., Arp, P., Ingvarsson, T., Jhamai, M., Jonsson, H., Stolk, L., Thorleifsson, G., Zhai, G., Zhang, F., Zhu, Y., van der Breggen, R., Carr, A.J., Doherty, M., Doherty, S., Felson, D.T., Gonzalez, A., Halldorsson, B.V., Hart, D.J., Hauksson, V.B., Hofman, A., Ioannidis, J.P.A., Kloppenburg, M., Lane, N.E., Loughlin, J., Luyten, F.P., Nevitt, M.C., Parimi, N., Pols, H.A.P., Rivadeneira, F., Slagboom, E.P., Styrkársdóttir, U., Tsezou, A., van de Putte, T., Zmuda, J., Spector, T.D., Stefansson, K., Uitterlinden, A.G., van Meurs, J.B.J. (2010) A genome-wide association study identifies an osteoarthritis susceptibility locus on chromosome 7q22. Arthritis Rheum 62: 499-510.
169. Evangelou, E., Valdes, A.M., Kerkhof, H.J., Styrkarsdottir, U., Zhu, Y., Meulenbelt, I., Lories, R.J., Karassa, F.B., Tylzanowski, P., Bos, S.D.; arcOGEN Consortium, Akune, T., Arden, N.K., Carr, A., Chapman, K., Cupples, L.A., Dai, J., Deloukas, P., Doherty, M., Doherty, S., Engstrom, G., Gonzalez, A., Halldorsson, B.V., Hammond, C.L., Hart, D.J., Helgadottir, H., Hofman, A., Ikegawa, S., Ingvarsson, T., Jiang, Q., Jonsson, H., Kaprio, J., Kawaguchi, H., Kisand, K., Kloppenburg, M., Kujala, U.M., Lohmander, L.S., Loughlin, J., Luyten, F.P., Mabuchi, A., McCaskie, A., Nakajima, M., Nilsson,

P.M., Nishida, N., Ollier, W.E., Panoutsopoulou, K., van de Putte, T., Ralston, S.H., Rivadeneira, F., Saarela, J., Schulte-Merker, S., Shi, D., Slagboom, P.E., Sudo, A., Tamm, A., Tamm, A., Thorleifsson, G., Thorsteinsdottir, U., Tsezou, A., Wallis, G.A., Wilkinson, J.M., Yoshimura, N., Zeggini, E., Zhai, G., Zhang, F., Jonsdottir, I., Uitterlinden, A.G., Felson, D.T., van Meurs, J.B., Stefansson, K., Ioannidis, J.P., Spector, T.D.; Translation Research in Europe Applied Technologies for Osteoarthritis (TreatOA). (2011) Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. Ann Rheum Dis 70: 349-355.

170. Raine, E.V.A., Wreglesworth, N., Dodd, A.W., Reynard, L.N., Loughlin, J. (2012) Gene expression analysis reveals HBP1 as a key target for the osteoarthritis susceptibility locus that maps to chromosome 7q22. Ann Rheum Dis 71:2020-2027.
171. Nakajima, M., Takahashi, A., Kou, I., Rodriguez-Fontenla, C., Gomez-Reino, J.J., Furuichi, T., Dai, J., Sudo, A., Uchida, A., Fukui, N., Kubo, M., Kamatani, N., Tsunoda, T., Malizos, K.N., Tsezou, A., Gonzalez, A., Nakamura, Y., Ikegawa, S. (2010) New sequence variants in hla class ii/iii region associated with susceptibility to knee osteoarthritis identified by genome-wide association study. PLoS ONE 5: e9723. doi:10.1371/journal.pone.0009723.
172. Valdes, A.M., Styrkarsdottir, U., Doherty, M., Morris, D.L., Mangino, M., Tamm, A., Doherty, S.A., Kisand, K., Kerna, I., Tamm, A., Wheeler, M., Maciewicz, R.A., Zhang, W., Muir, K.R., Dennison, E.M., Hart, D.J., Metrustry, S., Jonsdottir, I., Jonsson, G.F., Jonsson, H., Ingvarsson, T., Cooper, C., Vyse, T.J., Spector, T.D., Stefansson, K., Arden, N.K. (2011) Large scale replication study of the association between hla class ii/btnl2 variants and osteoarthritis of the knee in european-descent populations. PLoS ONE 6: doi: 10.1371/journal.pone.0023371.
173. Panoutsopoulou, K., Southam, L., Elliott, K. S., Wrayner, N., Zhai, G., Beazley, C., Thorleifsson, G., Arden, N.K., Carr, A., Chapman, K., Deloukas, P., Doherty, M., McCaskie, A., Ollier, W.E.R., Ralston, S.H., Spector, T.D., Valdes, A.M., Wallis, G.A., Wilkinson, J.M., Arden, E., Battley, K., Blackburn, H., Blanco, F.J., Bumpstead, S., Cupples, L.A., Day-Williams, A.G., Dixon, K., Doherty, S.A., Esko, T., Evangelou, E., Felson, D., Gomez-Reino, J.J., Gonzalez, A., Gordon, A., Gwilliam, R., Halldorsson, B.V., Hauksson, V.B., Hofman, A., Hunt, S.E., Ioannidis, J.P.A., Ingvarsson, T., Jonsdottir, I., Jonsson, H., Keen, R., Kerkhof, H.J.M., Kloppenburg, M G., Koller, N., Lakenberg, N., Lane, N.E., Lee, A.T., Metspalu, A., Meulenbelt, I., Nevitt, M.C., O'Neill, F., Parimi, N., Potter, S.C., Rego-Perez, I., Riancho, J.A., Sherburn, K., Slagboom, P.E., Stefansson, K., Styrkarsdottir, U., Sumillera, M., Swift, D., Thorsteinsdottir, U., Tsezou, A., Uitterlinden, A.G., van Meurs, J.B.J., Watkins,B., Wheeler, M., Mitchell, S., Zhu, Y., Zmuda, J.M., arcOgen Consortium, Zeggini, E., Loughlin, J. (2011) Insights into the genetic architecture of osteoarthritis from stage 1 of the arcOGEN study. Ann Rheum Dis 70: 864-867.
174. Ratnayake, M., Reynard, L.N., Raine, E.V.A, Santibanez-Koref, M., Loughlin, J. (2012) Allelic expression analysis of the osteoarthritis susceptibility locus that maps to MICAL3. BMC Med Genet 13: doi: 10.1186/1471-2350-13-12.
175. arcOGEN Consortium: arcOGEN collaborators (2012) Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. Lancet 380: 815-823.
176. Karlsson, C., Dehne, T., Lindahl, A., Brittberg, M., Pruss, A., Sittinger, M.

, Ringe, J. (2010) Genome-wide expression profiling reveals new candidate genes associated with osteoarthritis. Osteoarthritis Cartilage 18: 581-592.
177. Habuchi, O. (2000) Diversity and functions of glycosaminoglycan sulfotransferases. Biochim Biophys Acta 1474: 115-127.
178. Urist, M.R. (1965) Bone: formation by autoinduction. Science 150: 893-899.
179. Wozney, J.M. (1992) The bone morphogenetic protein family and osteogenesis. Mol Reprod Dev 32: 160-167.
180. Luyten, F.P., Cunningham, N. S., Ma, S., Muthukumaran, N., Hammonds, R.G., Nevins, W.B., Woods, W.I., Reddi, A.H. (1989) Purification and partial amino acid sequence of osteogenin, a protein initiating bone differentiation. J Biol Chem 264: 13377-13380.
181. Wozney, J.M., Rosen, V., Celeste, A.J., Mitsock, L.M., Whitters, M.J., Kriz, R.W., Hewick, R.M., Wang, E.A. (1988) Novel regulators of bone formation: molecular clones and activities. Science 242: 1528-1534.
182. van Wijk, B., Moorman, A.F.M., van den Hoff, M.J.B. (2007) Role of bone morphogenetic proteins in cardiac differentiation. Cardio Vasc Res 74: 244-255.
183. Nakamura, J., Yanagita, M. (2012) BMP modulators in kidney disease. Discov Med 13: 57-63.
184. Lowery, J.W., de Caestecker, M.P. (2010) BMP signalling in vascular development and disease. Cytokine Growth Factor Rev 21: 287-298.
185. Dahn, R.D., Fallon, J.F. (1999) Limbiting outgrowth: BMPs as negative regulators in limb development. Bioessays 21: 721-725.
186. Aoki, H., Fujii, M., Imamura, T., Yagi, K., Takehara, K., Kato, M., Miyazono, K. (2001) Synergistic effects of different bone morphogenetic protein type I receptors on alkaline phosphatase induction. J Cell Science 114: 1483-1489.
187. Chen, D., Zhao, M., Mundy, G. (2004) Bone Morphogenetic Proteins. Growth Factors 22: 233-241.
188. Miyazono, K. (1999) Signal transduction by bone morphogenetic protein receptors: functional roles of Smad proteins. Bone 25: 91-93.
189. Guicheux, J., Lemonnier, J., Ghayor, C., Suzuki, A., Palmer, G., Caverzasio, J. (2003) TGF-β and BMP Signalling in Osteoblast Differentiation and Bone Formation. J Bone Miner Res 18: 2060-2068.
190. Massague, J. (2003) Intergration of Smad and MAPK pathways: a link and a linker revisted. Genes Dev 17: 2993-2997.
191. Hoffmann, A., Preobrazhenska, O., Wodarczyk, C., Medler, Y., Winkel, A., Shahab, S., Huylebroeck, D., Gross, G., Verschueren, K. (2005) Transforming growth factor-beta-activated kinase-1 (tak1), a map3k, interacts with smad proteins and interferes with osteogenesis in murine mesenchymal progenitors. J Biol Chem 280: 27271-27283.
192. Sieber, C., Kopf, J., Hiepen, C., Knaus, P. (2009) Recent advances in bmp receptor signalling. Cytokine Growth Factor Rev 20: 343-355.
193. Isaacs, M.J., Kawakami, Y., Allendorph, G.P., Yoon, B.H., Belmonte, J.C.I, Choe, S. (2010) Bone Morphogenetic Protein-2 and -6 Heterodimer Illustrates the Nature of Ligand-Receptor Assembly. Mol Endocrinol 24: 1469-1477.
194. Knaus, P. (2012) Subcellular Compartmentalization of BMP Signaling Pathways. In: International Conference on Bone Morphogenetic Proteins, June 2012. Tahoe, USA.
195. Cho, K.W., Blitz, I, L. (1998) BMPs, Smads and metalloproteases: extracellular and intracellular modes of negative regulation. Curr Opin Genet Dev 8: 443-449.
196. Piccolo, S., Sasai, Y., Lu, B., De Robertis, E. M. (1996) Dorsoventral patterning in Xenopus: inhibition of ventral signals by direct binding of chordin to BMP-4. Cell 86: 589-598.

197. Imamura, T., Takase, M., Nisihara, A., Oeda, E., Hanai, J., Kawabata, M., Miyazono, K. (1997) Smad6 inhibits signalling by the TGF beta superfamily. Nature 389: 622-626.
198. Yoshida, Y., Tanaka, S., Umemori, H., Minowa, O., Usui, M., Ikematsu, N., Hosoda, E., Imamura, T., Kuno, J., Yamashita, T., Miyazono, K., Noda, M., Noda, T., Yamamoto, T. (2000) Negative regulation of bmp/smad signaling by tob in osteoblasts. Cell 103: 1085-1097.
199. Zhu, H., Kavsak, P., Abdollah, S., Wrana, J., Thomsen, G.H.A. (1999) SMAD ubiquitin ligase targets BMP pathway and affects embryonic pattern formation. Nature 400: 687-693.
200. Zhao, M., Qiao, M., Oyajobi, B., Mundy, G.R., Chen, D. (2003) E3 ubiquitin ligase Smurf1 mediates core binding factor alpha1/Runx2 degradation and plays a specific role in osteoblast differentiation. J Biol Chem 278: 27939-27944.
201. Murakami, G., Watabe, T., Takaoka, K., Miyazono, K., Imamura, T. (2003) Cooperative inhibition of bone morphogenetic protein signalling by Smurf1 and inhibitory Smads. Mol Biol Cell 14: 2809-2817.
202. Brunet, L.J., McMahon, J.A., McMahon, A.P., Harland, R.M. (1998) Noggin, cartilage morphogenesis, and joint formation in the mammalian skeleton. Science 280: 1455-1457.
203. Wu, X.B., Li, Y., Schneider, A., Yu, W., Rajendren, G., Iqbal, J., Yamamoto, M., Alam, M., Brunet, L.J., Blair, H.C., Zaidi, M., Abe, E. (2003) Impaired osteoblastic differentiation, reduced bone formation and severe osteoporosis in noggin-overexpressing mice. J Clin Invest 112: 924-934.
204. Lories, R.J., Daans, M., Derese, I., Matthys, P., Kasran, A., Tylzanowski, P., Ceuppens, J.L., Luyten, F.P. (2006) Noggin haploinsufficiency differentially affects tissue responses in destructive and remodelling arthritis. Arthritis Rheum 54: 1736-1746.
205. Gong, Y., Krakow, D., Marcelino, J., Wilkin, D., Chitayat, D., Babul-Hirji, R., Hudgins, L., Cremers, C.W., Cremers, F.P., Brunner, H.G., Reinker, K., Rimoin, D.L., Cohn, D.H., Goodman, F.R., Reardon, W., Patton, M., Francomano, C.A., Warman, M.L. (1999) Heterozygous mutations in the gene encoding noggin affect human joint morphogenesis. Nat Genet 21: 302-304.
206. Horiki, M., Imamura, T., Okamoto, M., Hayashi, M., Murai, J., Myoui, A., Ochi, T., Miyazono, K., Yoshikawa, H., Tsunaki, N. (2004) Smad6/Smurf1 overexpression in cartilage delays chondrocyte hypertrophy and causes dwarfism with osteopenia. J Cell Biol 165: 433-445.
207. Yoshida, Y., Tanaka, S., Umemori, H., Minowa, O., Usui, M., Ikematsu, N., Hosoda, E., Imamura, T., Kuno, J., Yamashita, T., Miyazono, K., Noda, M., Noda, T., Yamamoto, T. (2000) Negative Regulation of the BMP/SMAD signalling by Tob in osteoblasts. Cell 103: 1085-1097.
208. Zhao, M., Qiao, M., Harris, S.E., Oyajobi, B., Mundy, G.R., Chen, D. (2004) Smurf1 inhibits osteoblast differentiation and bone formation in vitro and in vivo. J Biol Chem 279: 12854-12859.
209. Laudauer, W. (1952) Brachypodism, a recessive mutation of house mice. Hereditary 43: 293-298.
210. Storm, E.E., Huynh, T.V., Copeland, N.G., Jenkins, N.A., Kingsley, D.M., Lee, S.J. (1994) Limb alterations in brachypodism mice due to mutations in a new member of the TGF beta-superfamily. Nature 368: 639-643.
211. Kingsley, D.M., Bland, A.E., Grubber, J.M., Marker, P.C., Russell, L.B., Copeland, N.G., Jenkins, N.A. (1992) The mouse short ear skeletal morphogenesis locus is associated with defects in a bone morphogenetic member of the TGFβ superfamily. Cell 71: 399-410.

212. Edwards, C.J., Francis-West, P.H. (2001) Bone morphogenetic proteins in the development and healing of synovial joints. Semin Arthritis Rheum 31: 33-42.
213. Hotten, G., Neidhardt, H., Jacobowsky, B., Pohl, J. (1994) Cloning and expression of recombinant human growth/differentiation factor 5. Biochem Biophys Res Commun 204: 646-652.
214. Venkataraman, G., Sadidekharan, V., Cooney, C.L., Langer, R., Sasisekharan, R. (1995) Complex flexibility of the transforming growth factor Beta superfamily. Proc Natl Acad Sci 92: 5406-5410.
215. Hotten, G.C., Matsumoto, T., Kimura, M., Bechtold, R.F., Kron, R., Ohara, T., Tanaka, H., Satoh, Y., Okazaki, M., Shirai, T., Pan, H., Kawai, S., Pohl, J. S., Kudo, A. (1996) Recombinant human growth/differentiation factor 5 stimulates mesenchyme aggregation and chondrogenesis responsible for the skeletal development of limbs. Growth Factors 13: 65-74.
216. Nishitoh, H., Ichijo, H., Kimurai, M., Matsumotoi, T., Makishimai, F., Yamaguchi, A., Yamashita, H., Enomoto, S., Miyazono, K. (1996) Identification of type I and type II serine/threonine kinase receptors for growth/differentiation factor-5. J Biol Chem 271: 21345-21352.
217. Hatakeyama, Y., Tuan, R.S, Shum, L. (2004) Distinct functions of BMP4 and GDF5 in the regulation of chondrogenesis. J Cell Biochem 91: 1204-2017.
218. Shore, E.M. (2011) Fibrodysplasia ossificans progressiva: a human genetic disorder of extraskeletal bone formation, or—how does one tissue become another? Wiley Interdiscip Rev Dev Biol 1: 153-165.
219. Luyten, F.P. (1995) Cartilage-derived morphogenetic proteins. Key regulators in chondrocyte differentiation? Acta Orthop Scand Suppl 266: 51-54.
220. Chang, S.C., Hoang, B., Thomas, J.T., Vukicevic, S., Luyten, F.P., Ryba, N.J., Kozak, C.A., Reddi, A.H., Moos, M. (1994) Cartilage-derived morphogenetic proteins. New members of the transforming growth factor-beta superfamily predominantly expressed in long bones during human embryonic development. J Biol Chem 269: 28227-28234.
221. Francis-West, P.H., Abdelfattah, A., Chen, P., Allen, C., Parish, J., Ladher, R., Allen, S., MacPherson, S., Luyten, F.P., Archer, C.W. (1999) Mechanisms of GDF-5 action during skeletal development. Development 126: 1305-1315.
222. Erlacher, L., Ng, C.K., Ullrich, R., Krieger, S., Luyten, F.P. (1998) Presence of cartilage-derived morphogenetic proteins in articular cartilage and enhancement of matrix replacement in vitro. Arthritis Rheum 41: 263-273.
223. Coleman, C.M., Vaughan, E.E., Browe, D.C., Mooney, E., Howard, L., Barry, F.P. (2013) Growth differentiation factor-5 enhances in vitro mesenchymal stromal cell chondrogenesis and hypertrophy. Stem Cells Dev [epub ahead of print].
224. Mikic, B., Battaglia, T.C., Taylor, E.A., Clark, R.T. (2002) The effect of growth/differentiation factor-5 deficiency on femoral composition and mechanical behavior in mice. Bone 30: 733-737.
225. Wolfman, N.M., Hattersley, G., Cox, K., Celeste, A.J., Nelson, R., Yamaji, N., Dube, J.L., DiBlasio-Smith, E., Nove, J., Song, J.J., Wozney, J.M., Rosen, V. (1997) Ectopic induction of tendon and ligament in rats by growth and differentiation factors 5, 6, and 7, members of the TGF-beta gene family. J Clin Invest 100: 321-330.
226. Zhong, Z.M., Chen, J. T., Zhang, Y., Zha, D.S., Lin, Z.S., Zhao, C.Y., Xu, J.C., Li, T., Xu, Z.X. (2010) Growth/differentiation factor-5 induces osteogenic differentiation of human ligamentum flavum cells through activation of ERK1/2 and p38 MAPK. Cell Physiol Biochem 26: 179-186.

227. Cheng, X., Yang, T., Meng, W., Liu, H., Zhang, T., Shi, R. (2012) Overexpression of GDF5 through an adenovirus vector stimulates osteogenesis of human mesenchymal stem cells in vitro and in vivo. Cells Tissues Organs 196: 56-67.
228. Koch, F.P., Weinbach, C., Hustert, E., Al-Nawas, B., Wagner, W. (2012) GDF-5 and BMP-2 regulate bone cell differentiation by gene expression of MSX1, MSX2, Dlx5, and Runx2 and influence OCN gene expression in vitro. Int J Periodontics Restorative Dent 32: 285-293.
229. Yamashita, H., Shimizu, A., Kato, M., Nishitoh, H., Ichijo, H., Hanyu, A., Morita, I., Kimura, M., Makishima, F., Miyazono, K. (1997) Growth/Differentiation factor-5 induces angiogenesis in vivo. Exp Cell Res 235: 218-226.
230. Thomas, J.T., Lin, K., Nandedkar, M., Camargo, M., Cervenka, J., Luyten, F.P. (1996) A human chondrodysplasia due to a mutation in a TGF-beta superfamily member. Nat Genet 12: 315-317.
231. Stelzer, C., Winterpacht, A., Spranger, J., Zabel, B. (2003) Grebe dysplasia and the spectrum of CDMP1 mutations. Pediatr Pathol Mol Med 22: 77-85.
232. Faiyaz-Ul-Haque, M., Ahmad, W., Wahab, A., Haque, S., Azim, A.C., Zaidi, S.H.E., Teebi, A.S., Ahmad, M., Cohn, D.H., Siddique, T., Tsui, L.C. (2002) Frameshift mutation in the cartilage-derived morphogenetic protein 1 (CDMP1) gene and severe acromesomelic chondrodysplasia resembling Grebe-type chondrodysplasia. Am J Med Genet 111: 31-37.
233. Basit, S., Kamran-ul-Hassan Naqvi, S., Wasif, N., Ali, G., Ansar, M., Ahmad, W. (2008) A novel insertion mutation in the cartilage-derived morphogenetic protein-1 (CDMP1) gene underlies Grebe-type chondrodysplasia in a consanguineous Pakistani family. BMC Med Genet 9: doi: 10.1186/1471-2350-9-102.
234. Faiyaz-Ul-Haque, M., Ahmad, W., Zaidi, S.H.E., Haque, S., Teebi, A.S., Ahmad, M., Cohn, D.H., Tsui, L.C. (2002) Mutation in the cartilage-derived morphogenetic protein-1 (CDMP1) gene in a kindred affected with fibular hypoplasia and complex brachydactyly (DuPan syndrome). Clin Genet 61: 454-458.
235. Ploger, F., Seemann, P., Schmidt-von Kegler, M., Lehmann, K., Seidel, J., Kjaer, K.W., Pohl, J., Mundlos, S. (2008) Brachydactyly type A2 associated with a defect in proGDF5 processing. Hum Mol Genet 17: 1222-1233.
236. Byrnes, A.M., Racacho, L., Nikkel, S.M., Xiao, F., MacDonald, H., Underhill, T.M., Bulman, D.E. (2010) Mutations in GDF5 presenting as semidominant brachydactyly A1. Human Mutation 31: 1155-1162.
237. Polinkovsky, A., Robin, N. H., Thomas, J. T., Irons, M., Lynn, A., Goodman, F.R., Reardon, W., Kant, S.G., Brunner, H.G., van der Burgt, I., Chitayat, D., McGaughran, J., Donnai, D., Luyten, F.P., Warman, M.L. (1997) Mutations in CDMP1 cause autosomal dominant brachydactyly type C. Nat Genet 17: 18-19.
238. Yang, W., Cao, L., Liu, W., Jiang, L., Sun, M., Zhang, D., Wang, S., Lo, W.H., Luo, Y., Zhang, X. (2008) Novel point mutations in GDF5 associated with two distinct limb malformations in Chinese: brachydactyly type C and proximal symphalangism. J Hum Genet 53: 368-374.
239. Schwabe, G.C., Türkmen, S., Leschik, G., Palanduz, S., Stöver, B., Goecke, T.O., Mundlos, S. (2004) Brachydactyly type C caused by a homozygous missense mutation in the prodomain of CDMP1. Am J Med Genet 124: 356-363.
240. Lehmann, K., Seemann, P., Stricker, S., Sammar, M., Meyer, B., Suring, K., Majewski, F., Tinschert, S., Grzeschik, K.H., Muller, D., Knaus, P., Nurnberg,

P., Mundlos, S. (2003) Mutations in bone morphogenetic protein receptor 1B cause brachydactyly type A2. Proc Natl Acad Sci 100: 12277-12282.
241. Lehmann, K., Seemann, P., Boergermann, J., Morin, G., Reif, S., Knaus, P., Mundlos, S. (2006) A novel R486Q mutation in BMPR1B resulting in either a brachydactyly type C/symphalangism-like phenotype or brachydactyly type A2. Eur J Hum Genet 14: 1248-1254.
242. Seemann, P., Schwappacher, R., Kjaer, K.W., Krakow, D., Lehmann, K., Dawson, K., Stricker, S., Pohl, J., Ploger, F., Staub, E., Nickel, J., Sebald, W., Knaus, P., Mundlos, S. (2005) Activating and deactivating mutations in the receptor interaction site of GDF5 cause symphalangism or brachydactyly type A2. J Clin Invest 115: 2373-2381.
243. Schwaerzer, G.K., Hiepen, C., Schrewe, H., Nickel, J., Ploger, F., Sebald, W., Mueller, T., Knaus, P. (2012) New insights into the molecular mechanism of multiple synostoses syndrome (SYNS): Mutation within the GDF5 knuckle epitope causes noggin-resistance. J Bone Min Res 27: 429-442.
244. Lehmann, K., Seemann, P., Silan, F., Goecke, T.O., Irgang, S., Kjaer, K.W., Kjaergaard, S., Mahoney, M.J., Morlot, S., Reissner, C., Kerr, B., Wilkie, A.O.M., Mundlos, S. (2007) A new subtype of Brachydactyly type B caused by point mutations in the bone morphogenetic protein antagonist NOGGIN. Am J Hum Genet 81: 388-396.
245. Dawson, K., Seeman, P., Sebald, E., King, L., Edwards, M., Williams, J., Mundlos, S., Krakow, D. (2006) GDF5 Is a second locus for multiple-synostosis syndrome. Am J Hum Genet 78: 708-712.
246. Seemann, P., Brehm, A., Konig, J., Reissner, C., Stricker, S., Kuss, P., Haupt, J., Renninger, S., Nickel, J., Sebald, W., Groppe, J.C., Ploger, F., Pohl, J., Schmidt-von Kegler, M., Walther, M., Gassner, I., Rusu, C., Janecke, A.R., Dathe, K., Mundlos, S. (2009) Mutations in GDF5 reveal a key residue mediating bmp inhibition by NOGGIN. PLoS Genet 5: e1000747. doi: 10.1371/journal/pgen.1000747.
247. Tsumaki, N., Tanaka,K., Arikawa-Hirasawa, E., Nakase, T., Kimura,T., Thomas, J.T., Ochi,T., Luyten,F.P., Yamada, Y. (1999) Role of CDMP-1 in skeletal morphogenesis: promotion of mesenchymal cell recruitment and chondrocyte differentiation. J Cell Biol 144: 161-173.
248. Masuya, H., Nishida, K., Furuichi, T., Toki, H., Nishimura, G., Kawabata, H., Yokoyama, H., Yoshida, A., Tominaga, S., Nagano, J., Shimizu, A., Wakana, S., Gondo, Y., Noda, T., Shiroishi, T., Ikegawa, S. (2007) A novel dominant-negative mutation in Gdf5 generated by ENU mutagenesis impairs joint formation and causes osteoarthritis in mice. Hum Mol Genet 16: 2366-2375.
249. Daans, M., Luyten, F.P., Lories, R.J. (2011) GDF5 deficiency in mice is associated with instability-driven joint damage, gait and subchondral bone changes. Ann Rheum Dis 70: 208-213.
250. Harada, M., Takahara, M., Zhe, P., Otsuji, M., Iuchi, Y., Takagi, M., Ogino, T. (2007) Developmental failure of the intra-articular ligaments in mice with absence of growth differentiation factor 5. Osteoarthritis Cartilage 15: 468-474.
251. Mikic, B., Schalet, B.J., Clark, R.T., Gaschen, V., Hunziker, E.B. (2001) GDF-5 deficiency in mice alters the ultrastructure, mechanical properties and composition of the Achilles tendon. J Orthop Res 19: 365-371.
252. Coleman, C., Scheremeta, B., Boyce, A., Mauck, R., Tuan, R. (2011) Delayed fracture healing in growth differentiation factor 5-deficient mice: A Pilot Study. Clin Orthop Relat Res 469: 2915-2924.
253. Rickert, M., Wang, H., Wieloch, P., Lorenz, H., Steck, E., Sabo, D., Richter, W. (2005) Adenovirus-mediated gene transfer of growth and differentiation factor-5

into tenocytes and the healing rat Achilles tendon. Connect Tissue Res 46: 175-183.
254. Keller, T.C., Hogan, M.V., Kesturu, G., James, R., Balian, G., Chhabra, A.B. (2011) Growth/differentiation factor-5 modulates the synthesis and expression of extracellular matrix and cell-adhesion-related molecules of rat Achilles tendon fibroblasts. Connect Tissue Res 52: 353-364.
255. Dines, J.S., Cross, M.B., Dines, D., Pantazopoulos, C., Kim, H.J., Razzano, P., Grande, D. (2011) In vitro analysis of an rhGDF-5 suture coating process and the effects of rhGDF-5 on rat tendon fibroblasts. Growth Factors 29: doi: 10.3109/08977194.2010.526605.
256. Hayashi, M., Zhao, C., An, K.N., Amadio, P.C. (2011) The effects of growth and differentiation factor 5 on bone marrow stromal cell transplants in an in vitro tendon healing model. J Hand Surg Eur 36: 271-279.
257. Henn, R.F., Kuo, C.E., Kessler, M.W., Razzano, P., Grande, D.P., Wolfe, S.W. (2010) Augmentation of zone ii flexor tendon repair using Growth differentiation factor 5 in a rabbit model. J Hand Surg Am 35: 1825-1832.
258. Saiga, K., Furumatsu, T., Yoshida, A., Masuda, S., Takihira, S., Abe, N., Ozaki, T. (2010) Combined use of bFGF and GDF-5 enhances the healing of medial collateral ligament injury. Biochem Biophys Res Commun 402: 329-334.
259. James, R., Kumbar, S.G., Laurencin, C.T., Balian, G., Chhabra, A.B. (2011) Tendon tissue engineering: adipose-derived stem cell and GDF-5 mediated regeneration using electrospun matrix systems. Biomed Mater 6: 025021. doi: 10.1088/1748-6041/6/2/025011.
260. Le Maitre, C.L., Freemont, A.J., Hoyland, J.A. (2009) Expression of cartilage-derived morphogenetic protein in human intervertebral discs and its effect on matrix synthesis in degenerate human nucleus pulposus cells. Arthritis Res Ther 11: R137. doi:10.1186/ar2808.
261. Stavropoulos, A., Becker, J., Capsius, B., Açil, Y., Wagner, W., Terheyden, H. (2011) Histological evaluation of maxillary sinus floor augmentation with recombinant human growth and differentiation factor-5-coated β-tricalcium phosphate: results of a multicenter randomized clinical trial. J Clin Periodontol 38: 966-974.
262. Emerton, K.B., Drapeau, S.J., Prasad, H., Rohrer, M., Roffe, P., Hopper, K., Schoolfield, J., Jones, A., Cochran, D.L. (2011) Regeneration of periodontal tissues in non-human primates with rhgdf-5 and beta-tricalcium phosphate. J Dent Res 90: 1416-1421.
263. Stavropoulos, A., Windisch, P., Gera, I., Capsius, B., Sculean, A., Wikesjö, U.M.E. (2011) A phase IIa randomized controlled clinical and histological pilot study evaluating rhGDF-5/β-TCP for periodontal regeneration. J Clin Periodontol 38: 1044-1054.
264. Stavropoulos, A., Wikesjö, U.M.E. (2012) Growth and differentiation factors for periodontal regeneration: a review on factors with clinical testing. J Periodontal Res 47: 545-553.
265. Spiro, R.C., Thompson, A.Y., Poser, J.Y. (2001) Spinal fusion with recombinant human growth and differentiation factor-5 combined with a mineralized collagen matrix. Anat Rec 263: 388-395.
266. Merino, R., Macias, D., Gañan, Y., Economides, A.N., Wang, X., Wu, Q., Stahl, N., Sampath, K.T., Varona, P., Hurle, J.M. (1999) Expression and function of Gdf-5 during digit skeletogenesis in the embryonic chick leg bud. Dev Biol 206: 33-45.
267. Southam, L., Rodriguez-Lopez, J., Wilkins, J.M., Pombo-Suarez, M., Snelling, S., Gomez-Reino, J.J., Chapman, K., Gonzalez, A., Loughlin, J. (2007) An SNP

in the 5'-UTR of GDF5 is associated with osteoarthritis susceptibility in Europeans and with in vivo differences in allelic expression in articular cartilage. Hum Mol Genet 16: 2226-2232.
268. Tsezou, A., Satra, M., Oikonomou, P., Bargiotas, K., Malizos, K.N. (2008) The growth differentiation factor 5 (GDF5) core promoter polymorphism is not associated with knee osteoarthritis in the greek population. J Orthop Res 26: 136-140.
269. Chapman, K., Takahashi, A., Meulenbelt, I., Watson, C., Rodriguez-Lopez, J., Egli, R., Tsezou, A., Malizos, K.N., Kloppenburg, M., Shi, D., Southam, L., van der Breggen, R., Donn, R., Qin, J., Doherty, M., Slagboom, P.E., Wallis, G., Kamatani, N., Jiang, Q., Gonzalez, A., Loughlin, J., Ikegawa, S. (2008) A meta-analysis of European and Asian cohorts reveals a global role of a functional SNP in the 5' UTR of GDF5 with osteoarthritis susceptibility. Hum Mol Genet 17: 1497-1504.
270. Valdes, A.M., Evangelou, E,. Kerkhof, H.J.M., Tamm, A., Doherty, S.A., Kisand, K., Tamm, A., Kerna, I., Uitterlinden, A., Hofman, A., Rivadeneira, F., Cooper, C., Dennison, E.M., Zhang, W., Muir, K.R., Ioannidis, J.P.A., Wheeler, M., Maciewicz, R.A., van Meurs, J.B., Arden, N.K., Spector, T.D., Doherty, M. (2011) The GDF5 rs143383 polymorphism is associated with osteoarthritis of the knee with genome-wide statistical significance. Ann Rheum Dis 70: 873-875.
271. Valdes, A.M., Doherty, S., Muir, K.R., Zhang, W., Maciewicz, R.A., Wheeler, M., Arden, N., Cooper, C., Doherty, M. (2012) Genetic contribution to radiographic severity in osteoarthritis of the knee Genetic contribution to radiographic severity in osteoarthritis of the knee. Ann Rheum Dis 71: 1537-1540.
272. Egli, R., Southam, L., Wilkins, J.M., Lorenzen, I., Pombo-Suarez, M., Gonzalez, A., Carr, A., Chapman, K., Loughlin, J. (2009) Functional analysis of the GDF5 regulatory polymorphism that is associated with OA susceptibility. Arthritis Rhuem 60: 2055-2064.
273. Sanna, S., Jackson, A.U., Nagaraja, R., Willer, C.J., Chen, W.M., Bonnycastle, L.L., Shen, H., Timpson, N., Lettre, G., Usala, G., Chines, P.S., Stringham, H.M., Scott, L.J., Dei, M., Lai, S., Albai, G., Crisponi, L., Naitza, S., Doheny, K.F., Pugh, E.W., Ben-Shlomo, Y., Ebrahim, S., Lawlor, D.A., Bergman, R.N., Watanabe, R.M., Uda, M., Tuomilehto, J., Coresh, J., Hirschhorn, J.N., Shuldiner, A.R., Schlessinger, D., Collins, F.S., Davey Smith, G., Boerwinkle, E., Cao, A., Boehnke, M., Abecasis, G.R., Mohlke, K.L. (2008) Common variants in the GDF5-UQCC region are associated with variation in human height. Nat Genet 40: 198-203.
274. Vaes, R.B.A., Rivadeneira, F., Kerkhof, J.M., Hofman, A., Pols, H.A.P., Uitterlinden, A.G., van Meurs, J.B.J. (2009) Genetic variation in the GDF5 region is associated with osteoarthritis, height, hip axis length and fracture risk: the Rotterdam study. Ann Rheum Dis 68: 1754-1760.
275. Posthumus, M., Collins, M., Cook, J., Handley, C.J., Ribbans, W.J., Smith, R.K., Schwellnus, M.P., Raleigh, S.M. (2010) Components of the transforming growth factor-{beta} family and the pathogenesis of human Achilles tendon pathology--a genetic association study. Rheumatology (Oxford) 49: 2090-2097.
276. Dai, J., Shi, D., Zhu, P., Qin, J., Ni, H., Xu, Y., Yao, C., Zhu, L., Zhu, H., Zhao, B., Wei, J., Liu, B., Ikegawa, S., Jiang, Q., Ding, Y. (2008) Association of a single nucleotide polymorphism in growth differentiate factor 5 with congenital dysplasia of the hip: a case-control study. Arthritis Res Ther 10: R126. doi:10.1186/ar2540.

277. Williams, F.M.K., Popham, M., Hart, D.J., de Schepper, E., Bierma-Zeinstra, S., Hofman, A., Uitterlinden, A.G., Arden, N.K., Cooper, C., Spector, T.D., Valdes, A.M., van Meurs, J. (2011) GDF5 single-nucleotide polymorphism rs143383 is associated with lumbar disc degeneration in Northern European women. Arthritis Rheum 63: 708-712.
278. Latchman, D. (2007) Eukaryotic Transcription Factors. 5th ed. Burlington: Elsevier.
279. Brown, D.D. (1984) The role of stable complexes that repress and activate eukaryotic genes. Cell 37: 359-365.
280. Thomas, M.C., Cheng-Ming, C. (2006) The General Transcription Machinery and General Cofactors. Crit Rev Biochem Mol Biol 41: 105-178.
281. Mathis, D., Chambon, P. (1981) The SV40 early region TATA box is required for accurate in vitro initiation of transcription. Nature 290: 310-315.
282. Baumann, M., Pontiller, J., Ernst, W. (2010) Structure and basal transcription complex of RNA polymerase II core promoters in the mamallian genome:an overview. Mol.Biotechnol 45: 241-247.
283. Gershenzon, N.I., Ioshikhes, I.P. (2005) Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. Bioinformatics 21: 1295-1300.
284. Kuras, L., Struhl, K. (1999) Binding of TBP to promoters in vivo is stimulated by activators and requires Pol II holoenyme. Nature 399: 609-613.
285. Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., Luscombe, N.M. (2009) A census of human transcription factors: function, expression and evolution. Nat Rev Genet 10: 252-263.
286. Jimenez-Sanchez, G., Childs, B., Valle, D. (2001) Human disease genes. Nature 409: 853-855.
287. Stranger, B.E., Montgomery, S.B., Dimas, A.S., Parts, L., Stegle, O., Ingle, C.E., Sekowska, M., Smith, G.D., Evans, D., Gutierrez-Arcelus, M., Price, A., Raj, T., Nisbett, J., Nica, A.C., Beazley, C., Durbin, R., Deloukas, P., Dermitzakis, E.T. (2012) Patterns of cis regulatory variation in diverse human populations. PLoS Genet 8: e1002639. doi: 10.1371/journal.pgen.1002639.
288. Liu, J., Francke, U. (2006) Identification of cis-regulatory elements for MECP2 expression. Hum Mol Genet 15: 1769-1782.
289. Kleinjan, D.A., van Heyningen, V. (2005) Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet 76: 8-32.
290. Wilkins, J.M., Southam, L., Price, A.J., Mustafa, Z., Carr, A., Loughlin, J. (2007) Extreme context specificity in differential allelic expression. Hum Mol Genet 16: 537-546.
291. Koch, O., Kwiatkowski, D.P., Udalova, I.A. (2006) Context-specific functional effects of IFNGR1 promoter polymorphism. Hum Mol Genet 15: 1475-1481.
292. Qiao, L., MacLean, P.S., Schaack, J., Orlicky, D.J., Darimont, C., Pagliassotti, M., Friedman, J.E., Shao, J. (2005) C/EBPalpha regulates human adiponectin gene transcription through an intronic enhancer. Diabetes 54: 1744-1754.
293. Bray, N.J., Buckland, P.R., Owen, M.J., O'Donovan, M.C. (2003) Cis-acint variation in the expression of a high proportion of genes in human brain. Hum Genet 113: 149-153.
294. Lo, H.S., Wang, Z., Hu, Y., Yang, H.H., Gere, S., Buetow, K.H., Lee, M.P. (2003) Allelic variation in gene expression is common in the human genome. Genome Res 13: 1855-1862.
295. Cowles, C.R., Hirschhorn, J.N., Altshuler, D., Lander, E.S. (2002) Detection of regulatory variation in mouse genes. Nat Genet 32: 432-437.

296. Vivancos, P.D., Driscoll, S.P., Bulman, C.A., Ying, L., Emami, K., Treumann, A., Mauve, C., Noctor, G., Foyer, C.H. (2011) Perturbations of amino acid metabolism associated with glyphosate-dependent inhibition of shikimic acid metabolism affect cellular redox homeostasis and alter the abundance of proteins involved in photosynthesis and photorespiration. Plant Physiol 157: 256-268.
297. Dodd, A.W., Syddall, C.M., Loughlin, J. (2012) 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 doi: 10.1038/ejhg.2012.197.
298. Yip, L., Su, L., Sheng, D., Chang, P., Atkinson, M., Czesak, M., Albert, P.R., Collier, A.R., Turley, S.J., Fathman, C.G., Creusot, R.J. (2009) Deaf1 isoforms control the expression of genes encoding peripheral tissue antigens in the pancreatic lymph nodes during type 1 diabetes. Nat Immunol 10: 1026-1033.
299. Reynard, L.N., Bui, C. Canty-Laird, E.G., Young, D.A., Loughlin, J. (2011) Expression of the osteoarthritis-associated gene GDF5 is modulated epigenetically by DNA methylation. Hum Mol Genet 20: 3450-3460.
300. Sugiura, T., Hotten, G., Kawai, S. (1999) Minimal promoter components of the human Growth/differentiation factor-5 gene. Biochem Biophys Res Commun 263: 707-713.
301. Kaczynski, J., Cook, T., Urrutia, R. (2003) Sp1- and Kruppel-like transcription factors. Genome Biol 4: 206.
302. Dynan, W.S., Tjian, R. (1983) The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell 35: 79-87.
303. Kadonaga, J.T., Carner, K.R., Masiarz, F.R., Tjian, R. (1987) Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell. 51: 1079-1090.
304. Kingsley, C., Winoto, A. (1992) Cloning of GT box-binding proteins: a novel Sp1 multigene family regulating T-cell receptor gene expression. Mol Cell Biol 12: 4251-4261.
305. Black, A.R., Black, J.D., Azizkhan-Clifford, J. (2001) Sp1 and krüppel-like factor family of transcription factors in cell growth regulation and cancer. J Cell Physiol 188: 143-160.
306. Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., Wheeler, R., Wong, B., Drenkow, J., Yamanaka, M., Patel, S., Brubaker, S., Tammana, H., Helt, G., Struhl, K., Gingeras, T.R. (2004) Unbiased mapping of transcription factor binding sites along human Chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell 116: 499-509.
307. Suske, G. (1999) The Sp-family of transcription factors. Gene 238: 291-300.
308. Zhang, Y., Dufau, M.L. (2003) Repression of the luteinizing hormone receptor gene promoter by cross talk among EAR3/COUP-TFI, Sp1/Sp3, and TFIIB. Mol Cell Biol 23: 6958-6972.
309. Lee, J.S., Galvin, K.M., Shi, Y. (1993) Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. Proc Natl Acad Sci USA 90: 6145-6149.
310. Lin, S.Y., Black, A.R., Kostic, D., Pajovic, S., Hoover, C.N., Azizkhan, J.C. (1996) Cell cycle-regulated association of E2F1 and Sp1 is related to their functional interaction. Mol Cell Biol 16: 1668-1675.
311. Ryu, S., Tjian, R. (1999) Purification of transcription cofactor complex CRSP. Proc Natl Acad Sci USA 96: 7137-7142.
312. Hagen, G., Muller, S., Beato, M., Suske, G. (1994) Sp1-mediated transcriptional activation is repressed by Sp3. EMBO J 13: 3843-3851.

313. Majello, B., De Luca, P., Lania, L. (1997) Sp3 Is a bifunctional transcription regulator with modular independent activation and repression domains. J Biol Chem 272: 4021-4026.
314. Gill, G., Pascal, E., Tseng, Z.H., Tjian, R. (1994) A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the Drosophila TFIID complex and mediates transcriptional activation. Proc Natl Acad Sci USA 91: 192-196.
315. Lu, F., Zhou, J., Wiedmer, A., Madden, K., Yuan, Y., Lieberman, P.M. (2003) Chromatin remodeling of the kaposi's sarcoma-associated herpesvirus ORF50 promoter correlates with reactivation from latency. J Virol 77: 11425-11435.
316. Brandeis, M., Frank, D., Keshet, L., Siegfried, Z., Mendelsohn, M., Names, A., Temper, V., Razin, A., Cedar, H. (1994) Spl elements protect a CpG island from de novo methylation. Nature 371: 435-438.
317. Clark, S.J., Harrison, J., Molloy, P.L. (1997) Sp1 binding is inhibited by mCpmCpG methylation. Gene 195: 67-71.
318. Xu, H.G., Ren, W., Zou, L., Wang, Y., Jin, R., Zhou, G.P. Transcriptional control of human CD2AP expression: the role of Sp1 and Sp3. Mol Biol Rep 39: 1479-1486.
319. Marin, M., Karis, A., Visser, P., Grosveld, F., Philipsen, S. (1997) Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. Cell 89: 619-628.
320. Gollner, H., Dani, C., Phillips, B., Philipsen, S., Suske, G. (2001) Impaired ossification in mice lacking the transcription factor Sp3. Mechanisms Dev 106: 77-83.
321. Bouwman, P., Gollner, H., Elsasser, H.P., Eckhoff, G., Karis, A., Grosveld, F., Philipsen, S., Suske, G. (2000) Transcription factor Sp3 is essential for post-natal survival and late tooth development. EMBO J 19: 655-661.
322. Reed, D.E., Huang, X.M., Wohlschlegel, J.A., Levine, M.S., Senger, K. (2008) DEAF-1 regulates immunity gene expression in Drosophila. Proc Natl Acad Sci USA 105: 8351-8356.
323. Dayon, L., Hainard, A., Licker, V., Turck, N., Kuhn, K., Hochstrasser, D.F., Burkhard, P.R., Sanchez, J. (2008) Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. Anal Chem 80: 2921-2931.
324. Carey, M.F., Peterson, C.L., Smale, S.T. (2012) Experimental strategies for the identification of DNA-binding proteins. Cold Spring Harb Protoc 2012: 18-33.
325. Meisterernst, M., Roy, A.L., Lieu, H.M., Roeder, R.G. (1991) Activation of class II gene transcription by regulatory factors is potentiated by a novel activity. Cell 66: 981-993.
326. Ge, H., Roeder, R.G. (1994) Purification, cloning, and characterization of a human coactivator, PC4, that mediates transcriptional activation of class II genes. Cell 78: 513-523.
327. Ge, H., Si, Y., Wolffe, A.P. (1998) A novel transcriptional coactivator, p52, functionally interacts with the essential splicing factor ASF/SF2. Mol Cell 2: 751-759.
328. Kannan, P., Tainsky, M.A. (1999) Coactivator PC4 mediates ap-2 transcriptional activity and suppresses Ras-induced transformation dependent on AP-2 transcriptional interference. Mol Cell Biol 19: 899-908.
329. Das, C., Hizume, K., Batta, K., Kumar, B.R.P., Gadad, S.S., Ganguly, S., Lorain, S., Verreault, A., Sadhale, P.P., Takeyasu, K., Kundu, T.K. (2006) Transcriptional coactivator PC4, a chromatin-associated protein, induces chromatin condensation. Mol Cell Biol 26: 8303-8315.

330. Li, N. (2010) Critical Role of Transcription Cofactor PC4 in Mammals. PhD Thesis, Helmholtz-Zentrum Munich.
331. Lindahl, T., Satoh, M.S., Poirier, G.G., Klungland, A. (1995) Post-translational modification of poly(ADP-ribose) polymerase induced by DNA strand breaks. Trends Biochem Sci 20: 405-411.
332. D'Amours, D., Desnoyers, S., D'Silva, I., Poirier, G.G. (1999) Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions. Biochem J 342: 249-268.
333. Oei, S.L., Griesenbeck, J., Schweiger, M., Ziegler, M. (1998) Regulation of RNA polymerase II-dependent transcription by poly(ADP-ribosyl)ation of transcription factors. J Biol Chem 273: 31644-31647.
334. Kim, M.Y., Mauro, S., Gévry, N., Lis, J.T., Kraus, W.L. (2004) NAD+-dependent modulation of chromatin structure and transcription by nucleosome binding properties of PARP-1. Cell 119: 803-814.
335. Ju, B.G., Solum, D., Song, E.J., Lee, K.J., Rose, D.W., Glass, C.K., Rosenfeld, M.G. (2004) Activating the PARP-1 sensor component of the Groucho/ TLE1 corepressor complex mediates a CaMKinase II-dependent neurogenic gene activation pathway. Cell 119: 815-829.
336. Miyamoto, T., Kakizawa, T., Hashizume, K. (1999) Inhibition of Nuclear Receptor Signalling by Poly(ADP-Ribose) Polymerase. Mol Cell Biol 19: 2644-2649.
337. Pederson, J.A., LaFollette, J.W., Gross, C., Veraksa, A., McGinnis, W., Mahaffey, J.W. (2000) Regulation by homeoproteins: a comparison of Deformed-responsive elements. Genetics 156: 677-686.
338. McGinnis, W., Krumlauf, R. (1992) Homeobox genes and axial patterning. Cell 68: 283-302.
339. McGinnis, W., Hart, C.P., Gehring, W.J., Ruddle, F.H. (1984) Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of drosophila. Cell 38: 675-680.
340. Biggin, M.D., McGinnis, W. (1997) Regulation of segmentation and segmental identity by Drosophila homeoproteins: the role of DNA binding in functional activity and specificity. Development 124: 4425-4433.
341. Regulski, M., McGinnis, N., Chadwick, R., McGinnis, W. (1987) Developmental and molecular analysis of Deformed; a homeotic gene controlling Drosophila head development. Embo J 6: 767-777.
342. Mann, R.S. (1995) The specificity of homeotic gene function. Bioessays 17: 855-863.
343. Gross, C.T., McGinnis, W. (1996) DEAF-1, a novel protein that binds an essential region in a Deformed response element. Embo J 15: 1961-1970.
344. Gudas, L.J. (1994) Retinoids and vertebrate development. J Biol Chem 269: 15399-15402.
345. Huggenvik, J.I., Michelson, R.J., Collard, M.W., Ziemba, A.J., Gurley, P., Mowen, K.A. (1998) Characterization of a nuclear deformed epidermal autoregulatory factor-1 (DEAF-1)-related (NUDR) transcriptional regulator protein. Mol Endocrinol 12: 1619-1639.
346. Hahm, K., Sum, E.Y., Fujiwara, Y., Lindeman, G.J., Visvader, J.E., Orkin, S.H. (2004) Defective neural tube closure and anteroposterior patterning in mice lacking the LIM protein LMO4 or its interacting partner Deaf-1. Mol Cell Biol 24: 2074-2082.
347. Larouche, K., Bergeron, M.J., Leclerc, S, Guérin, S.L. (1996) Optimization of competitor poly(dI-dC).poly(dI-dC) levels is advised in DNA-protein interaction studies involving enriched nuclear proteins. Biotechniques 20: 439-444.

348. Michelson, R.J., Collard, M.W., Ziemba, A.J., Persinger, J., Bartholomew, B., Huggenvik, J.I. (1999) Nuclear DEAF-1-related (NUDR) protein contains a novel DNA binding domain and represses transcription of the heterogeneous nuclear ribonucleoprotein A2/B1 promoter. J Biol Chem 274: 30510-30519.
349. Lemonde, S., Turecki, G. Bakish, D. Du, L., Hrdina, P.D., Bown, C.D., Sequeira, A., Kushwaha, N., Morris, S.J., Basak, A., Ou, X.M., Albert, P.R. (2003) Impaired repression at a 5-hydroxytryptamine 1A receptor gene polymorphism associated with major depression and suicide. J Neurosci 23: 8788-8799.
350. Czesak, M., Lemonde, S., Peterson, E.A., Rogaeva, A., Albert, P.R. (2006) Cell-specific repressor or enhancer activities of Deaf-1 at a serotonin 1A receptor gene polymorphism. J Neurosci 26: 1864-1871.
351. Szewczyk, B., Albert, P.R., Burns, A.M., Czesak, M., Overholser, J.C., Jurjus, G.J., Meltzer, H.Y., Konick, L.C., Dieter, L., Herbst, N., May, W., Rajkowska, G., Stockmeier, C.A., Austin, M.C. (2009) Gender-specific decrease in NUDR and 5-HT1A receptor proteins in the prefrontal cortex of subjects with major depressive disorder. Int J Neuropsychopharmacol 12: 155-168.
352. Coppen, A.J., Doogan, D.P. (1988) Serotonin and its place in the pathogenesis of depression. J Clin Psychiatry 49: 4-11.
353. Czesak, M., Le Francois, B., Millar, A.M., Deria, M., Daigle, M., Visvader, J.E., Anisman, H., Albert, P.R. (2012) Increased Serotonin-1A (5-HT1A) autoreceptor expression and reduced raphe serotonin levels in Deformed Epidermal Autoregulatory Factor-1 (Deaf-1) gene knock-out mice.J Biol Chem 287: 6615-6627.
354. Lemonde, S., Du, L., Bakish, D., Hrdina, P., Albert, P.R. (2004) Association of the C(-1019)G 5-HT1A functional promoter polymorphism with antidepressant response. Int J Neuropsychopharmacol 7: 501-506.
355. Bottomley, M.J., Collard, M.W., Huggenvik, J.I., Liu, Z., Gibson, T.J., Sattler, M. (2001) The SAND domain structure defines a novel DNA-binding fold in transcriptional regulation. Nat Struct Biol 8: 626-633.
356. Jensik, P.J., Huggenvik, J.I., Collard, M.W. (2004) Identification of a nuclear export signal and protein interaction domains in deformed epidermal autoregulatory factor-1 (DEAF-1). J Biol Chem 279: 32692-32699.
357. Sugihara, T.M., Bach, I., Kioussi, C., Rosenfeld, M.G., Andersen, B. (1998) Mouse Deformed epidermal autoregulatory factor 1 recruits a LIM domain factor, LMO-4, and CLIM coregulators. Proc Natl Acad Sci USA 95: 15418-15423.
358. Azizkhan, J.C., Jensen, D.E., Pierce, A.J., Wade, M. (1993) Transcription from TATA-less promoters: dihydrofolate reductase as a model. Crit Rev Eukaryot Gene Expr 3: 229-254.
359. Pugh, B.F., Tjian, R. (1991) Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Genes Dev 5:1935-1945.
360. Stoner, M., Wormke, M., Saville, B., Samudio, I., Qin, C., Abdelrahim, M., Safe, S. (2004) Estrogen regulation of vascular endothelial growth factor gene expression in ZR-75 breast cancer cells through interaction of estrogen receptor [alpha] and SP proteins. Oncogene 23: 1052-1063.
361. Zhang, X., Li, Y., Dai, C., Yang, J., Mundel, P., Liu, Y. (2003) Sp1 and Sp3 transcription factors synergistically regulate HGF receptor gene expression in kidney. Am J Physiol Renal Physiol 284: 82-94.
362. Yu, B., Datta, P.K., Bagchi, S. (2003) Stability of the Sp3–DNA complex is promoter-specific: Sp3 efficiently competes with Sp1 for binding to promoters containing multiple Sp-sites. Nucleic Acids Res 31: 5368-5376.

363. Ghayor, C., Chadjichristos, C., Herrouin J.F., Ala-Kokko, L., Suske, G., Pujol, J.P., Galera, P. (2001) Sp3 represses the Sp1-mediated transactivation of the human COL2A1 gene in primary and de-differentiated chondrocytes. J Biol Chem 276: 36881-36895.
364. Conesa, C., Acker, J. (2010) Sub1/PC4 a chromatin associated protein with multiple functions in transcription. RNA Biol 7: 287-290.
365. Doetzlhofer, A., Rotheneder, H., Lagger, G., Koranda, M., Kurtev, V., Brosch, G., Wintersberger, E., Seiser, C. (1999) Histone Deacetylase 1 can repress transcription by binding to Sp1. Mol Cell Biol 19: 5504-5511.
366. Liao, M., Zhang, Y., Kang, J.H., Dufau, M.L. (2011) Coactivator function of Positive Cofactor 4 (PC4) in Sp1-directed luteinizing hormone receptor (LHR) gene transcription. J Biol Chem 286: 7681-7691.
367. Lescuyer, P., Martinez, P., Lunardi, J. (2002) YY1 and Sp1 activate transcription of the human NDUFS8 gene encoding the mitochondrial complex I TYKY subunit. Biochim Biophys Acta 1574: 164-174.
368. Bombail, V., Taylor, K., Gibson, G.G., Plant, N. (2004) Role of Sp1, C/EBP alpha, HNF3, and PXR in the basal- and xenobiotic-mediated regulation of the CYP3A4 gene. Drug Metab Dispos 32: 525-535.
369. Ceelie, H., Spaargaren-Van Riel, C.C., De Jong, M., Bertina, R.M., Vos, H.L. (2003) Functional characterization of transcription factor binding sites for HNF1-alpha, HNF3-beta (FOXA2), HNF4-alpha, Sp1 and Sp3 in the human prothrombin gene enhancer. J Thromb Haemost 1: 1688-1698.
370. Atcha, F.A., Munguia, J.E., Li, T.W.H., Hovanes, K., Waterman, M.L. (2003) A new beta-catenin-dependent activation domain in T cell factor. J Biol Chem 278: 16169-16175.
371. Hecht, A., Stemmler, M.P. (2003) Identification of a promoter-specific transcriptional activation domain at the C terminus of the Wnt effector protein T-cell factor 4. J Biol Chem 278: 3776-3785.
372. Jungmann, J., Reins, H.A., Lee, J., Romeo, A., Hassett, R., Kosman, D., Jentsch, S. (1993) MAC1, a nuclear regulatory protein related to Cu-dependent transcription factors is involved in Cu/Fe utilization and stress resistance in yeast. EMBO J 12: 5051-5056.
373. Merino, R., Macias, D., Gañan, Y., Economides, A.N., Wang, X., Wu, Q., Stahl, N., Sampath, K.T., Varona, P., Hurle, J.M. (1999) Expression and function of Gdf-5 during digit skeletogenesis in the embryonic chick leg bud. Dev Biol 206: 33-45.
374. Dodd, A.W., Rodriguez-Fontenla, C., Calaza, M., Carr, A., Gomez-Reino, J. J. Tsezou, A., Reynard, L. N., Gonzalez, A., Loughlin, J. (2011) Deep sequencing of GDF5 reveals the absence of rare variants at this important osteoarthritis susceptibility locus. Osteoarthritis Cartilage 19: 430-434.
375. Kan, A., Ikeda, T., Fukai, A., Nakagawa, T., Nakamura, K., Chung, U.I., Kawaguchi, H, Tabin, C.J. (2013) SOX11 contributes to the regulation of GDF5 in joint maintenance. BMC Dev Biol 13: doi: 10.1186/1471-213X-13-4.
376. Grant, S.F., Reid, D.M., Blake, G., Herd, R., Fogelman, I., Ralston, S.H. (1996) Reduced bone density and osteoporosis associated with a polymorphic Sp1 binding site in the collagen type I alpha 1 gene. Nat Genet 14: 203-205.
377. Mann, V., Hobson, E.E., Li, B., Stewart, T.L., Grant, S.F.A., Robins, S.P., Aspden, R.M., Ralston, S.H. (2001) A COL1A1 Sp1 binding site polymorphism predisposes to osteoporotic fracture by affecting bone density and quality. J Clin Invest 107: 899-907.

378. Piera-Velazquez, S., Hawkins, D.F., Whitecavage, M.K., Colter, D.C., Stokes, D.G., Jimenez, S.A. (2007) Regulation of the human SOX9 promoter by Sp1 and CREB. Exp Cell Res 313: 1069-1079.
379. Zhang, Y., Hassan, M.Q., Xie, R.L., Hawse, J.R., Spelsberg, T.C., Montecino, M., Stein, J.L., Lian, J.B., van Wijnen, A.J., Stein, G.S. (2009) Co-stimulation of the bone-related RUNX2 P1 promoter in mesenchymal cells by SP1 and ETS transcription factors at polymorphic purine-rich DNA sequences (Y-repeats). J Biol Chem 284: 3125-3135.
380. Ghayor, C., Chadjichristos, C., Herrouin, J.F., Ala-Kokko, L., Suske, G., Pujol, J.P., Galera, P. (2001) Sp3 represses the Sp1-mediated transactivation of the human COL2A1 gene in primary and de-differentiated chondrocytes. J Biol Chem 276: 36881-36895.
381. Krüger, I., Vollmer, M., Simmons, D., Elsässer, H.P., Philipsen, S., Suske, G. (2007) Sp1/Sp3 compound heterozygous mice are not viable: Impaired erythropoiesis and severe placental defects. Dev Dyn 236: 2235-2244.
382. Cookson, W., Liang, L., Abecasis, G., Moffatt, M., Lathrop, M. (2009) Mapping complex disease traits with global gene expression. Nat Rev Genet 10: 184-194.
383. Gambari, R. (2011) Recent patents on therapeutic applications of the transcription factor decoy approach. Expert Opin Ther Pat 21: 1755-1771.
384. Dzau, V.J. (2002) Transcription Factor Decoy. Circ Res 90: 1234-1236.
385. Mann, M.J., Dzau, V. J. (2000) Therapeutic applications of transcription factor decoy oligonucleotides. J Clin Invest 106: 1071-1075.
386. Morishita, R., Gibbons, G.H., Horiuchi, M., Ellison, K.E., Nakajima, M., Zhang, L., Yasufumi, K., Ogihara, T., Dzau, V.J. (1995) A gene therapy strategy using a transcription factor decoy of the E2F binding site inhibits smooth muscle proliferation in vivo. Proc Natl Acad Sci USA 92: 5855-5859.
387. Zannetti, A., Del Vecchio, S., Romanelli, A., Scala, S., Saviano, M., Cali, G., Stoppelli, M.P., Pedone, C., Salvatore, M. (2005) Inhibition of Sp1 activity by a decoy PNA-DNA chimera prevents urokinase receptor expression and migration of breast cancer cells. Biochem Phamacol 70: 1277-1287.
388. Osborne, C.K., Wakeling, A., Nicholson, R.I. (2004) Fulvestrant: an oestrogen receptor antagonist with a novel mechanism of action. Br J Cancer 90: S2-S6.
389. Whitesell, L., Lindquist, S. (2009) Inhibiting the transcription factor HSF1 as an anticancer strategy. Expert Opin Ther Targets 13: 469-478.
390. Tomita, T., Takano, H., Tomita, N., Morishita, R., Kaneko, M., Shi, K., Takahi, K., Nakase, T., Kaneda, Y., Yoshikawa, H., Ochi, T. (2000) Transcription factor decoy for NFkappaB inhibits cytokine and adhesion molecule expressions in synovial cells derived from rheumatoid arthritis. Rheumatology (Oxford) 39: 749-757.
391. Tomita, T., Kunugiza, Y., Tomita, N., Takano, H., Morishita, R., Kaneda, Y., Yoshikawa, H. (2006) E2F decoy oligodeoxynucleotide ameliorates cartilage invasion by infiltrating synovium derived from rheumatoid arthritis. Int J Mol Med 18: 257-265.
392. Ray, A., Ray, B.K. (2008) An inflammation-responsive transcription factor in the pathophysiology of osteoarthritis. Biorheology 45: 399-409.
393. Wang, J., Gardner, B.M., Lu, Q., Rodova, M., Woodbury, B.G., Yost, J.G., Roby, K.F., Pinson, D.M., Tawfik, O., Anderson, H.C. (2009) Transcription factor NFAT1 deficiency causes osteoarthritis through dysfunction of adult articular chondrocytes. J Pathol 219: 163-172.
394. Yano, F., Hojo, H., Ohba, S., Fukai, A., Hosaka, Y., Ikeda, T., Saito, T., Hirata, M., Chikuda, H., Takato, T., Kawaguchi, H., Chung, U. (2012) A novel disease-

modifying osteoarthritis drug candidate targeting Runx1. Ann Rheum Dis [epub ahead of print].
395. Zhu, W., Srinivasan, K., Dai, Z., Duan,W., Druhan,L.J., Ding, H., Yee, L., Villalona-Calero, M.A., Plass, C., Otterson, G.A. (2003) Methylation of adjacent CpG Sites affects Sp1/Sp3 binding and activity in the p21Cip1 promoter. Mol Cell Biol 23: 4065-4065.
396. de Wit, E., de Laat, W. (2012) A decade of 3C technologies: insights into nuclear organization. Genes Dev 26: 11-24.
397. Butter, F., Davison, L., Viturawong, T., Scheibe, M., Vermeulen, M., Todd, J.M., Mann, M. (2012) Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. PLoS Genet 8: e1002982. doi: 10.1371/journal.pgen.1002982.