Mechanisms of Osteoporosis Associated With Common Allelic Variants of the *COL1A1* Gene

By

HUILIN JIN (BDS, Msc)

The School of Molecular & Clinical Medicine

A thesis presented for the degree of Doctor of Philosophy at the University of Edinburgh

2007

To my family

"What makes the desert beautiful," says the little prince, "is that somewhere it hides a well."

Ce qui embellit le désert, dit le petit prince, c'est qu'il cache un puits quelque part...

Le Petit Prince (1943)

Declaration

I hereby declare that this thesis has been composed by myself and that it has not been accepted in any previous application for a Doctor of Philosophy degree. The work, of which is a record, except where specifically acknowledged, has been carried out entirely by myself. All sources of information have been specifically acknowledged by means of reference.

Huilin Jin

Mechanisms of Osteoporosis Associated With Common Allelic Variants of the *COL1A1* Gene

Huilin Jin
School of Molecular and Clinical Medicine
University of Edinburgh

ABSTRACT

Introduction: Osteoporosis is a common disease with a strong genetic component and the collagen type I alpha I gene (*COL1A1*) is an important candidate for susceptibility. Three single nucleotide polymorphisms (SNP) have been identified in the 5' flank of the *COLIA1* gene (-1997G/T; -1663IndelT and +1245G/T) which have been associated with osteoporosis in various populations. The mechanisms by which these SNP predispose to osteoporosis are unclear however.

Methods: I analyzed the effects of the individual SNPs and associated haplotypes on bone mineral density (BMD) and risk of osteoporotic fracture in women in association studies. I also studied associations between the SNPs, haplotypes and bone strength ex-vivo by biomechanical testing of bone cores obtained from femoral heads of patients undergoing surgery for hip fracture. The effects of the SNPs and haplotypes on gene expression were studied using promoter-reporter assays and gel shift assays were performed to investigate whether the SNPs were situated at transcription factor binding sites and whether the polymorphic variants affected DNA protein binding. In order to investigate the possible effects of the polymorphisms *in vivo,* gene targeting constructs were prepared containing two common haplotypes and these were transfected into ES cells.

Results: The individual polymorphisms were all associated with BMD, but the haplotypes defined by all three SNPs showed a stronger association with BMD, biomechanical strength of bone and hip fracture. Two haplotypes increased in frequency with age suggesting an effect on survival. However these haplotypes were particularly enriched in hip fracture patients. Biomechanical testing showed that all three SNPs were strongly associated with reduced bone strength, independently of BMD. Gel shift assays showed that the region surrounding the -1663insdelT polymorphism recognized the nuclear binding proteins Sp1, Nmp4 and Osterix and the -1663delT allele had greater binding affinities for Nmp4 and Osterix than -1663insT allele. The region surrounding the -1997 G/T polymorphism also recognized DNA binding proteins but the polymorphism did not affect DNA protein binding significantly. Reporter assays showed significant differences between the ability of different haplotypes to drive gene expression and constructs containing haplotype2 (G-delT-T) had the highest transcriptional activity ($p<0.001$). Transgenic constructs containing different 5' *COL1A1* haplotypes were completed and transformed into mouse embryonic stem cells by electroporation in an attempt to

generate a disease model of osteoporosis associated with common variants in the *COLIA1* gene

Conclusion: The studies suggest that haplotypes, rather than individual polymorphisms in the 5' flank of *COL1A1* predispose to osteoporosis, by affecting DNA protein interactions and gene expression.

Acknowledgements

I would like to express my sincere thanks to Professor Stuart H. Ralston, for his guidance and encouragement throughout the duration of this project. Without his numerous advice and careful review of the manuscript, the completion of this thesis would not have been possible. Many thanks to Dr. Robert van 't Hof for his advice and support.

Many thanks Dr. Tracy Stewart (for her advice and practical assistance on biomechanical tests), Dr. Julia Dorin and Fiona Kilanowski (for their practical assistance on gene targeting process), Dr. Omar Albagha (for his advice and support), Stuart Bear (for practical assistance and preparation of DNA samples), and Belinda Stephen (for her warm-hearted help). Thanks to Drs. Anna Daroszewska, Aymen Idris, Gavin Lucas and Helene Libouban for their advice, help discussion and support. Thanks to Euphemie Landao, Karen Alexander, Micaela Rios Petrakis, and Antonia Sophocleus for their "disturbance" and laughs. Thanks also to all the other members of the Bone group for making the lab such a nice place to work.

I would also like to thank Professor Benoit de Crombrugghe who kindly gave me the antibody to transcription factor Osterix, and Professor Joseph P. Bidwell who kindly gave me the antibody to transcription factor Nmp4.

Finally, thanks to my parents and Rui for their whole-hearted love and support in all ways.

This work was funded by an Oversea Research Student (ORS) Awards Scheme.

Publications from this thesis

PAPERS

JIN H and Ralston SH. Genetics of Osteoporosis. *Curr Rheumatol Rep.* 2005, **7(1)**: 66-70

Stewart TL, **Jin H**, McGuigan FE, Albagha OM, Garcia-Giralt N, Bassiti A, Grinberg D, Balcells S, Reid DM and Ralston SH. Haplotypes defined by promoter and intron 1 polymorphisms of the COLIA1 gene regulate bone mineral density in women. *J Clin Endocrinol Metab*. 2006, **91(9)**: 3575-83

ABSTRACTS & ORAL PRESENTATIONS

JIN H, Stewart TL, McGuigan FEA, Van 't Hof R, Garcia-Giralt N, Grinberg D, Balcells S, Reid DM and Ralston SH. Molecular mechanisms of osteoporosis associated with common allelic variants of the COL1A1 gene. Presented at the 2nd Inernational Conference on Osteoporosis and Bone Research, Chengdu, China, 2005.

JIN H, Stewart TL, Van 't Hof R, Aspden RM and Ralston SH. A rare haplotype in the 5' flank of the COLIA1 gene is enriched in hip fracture patients and reduces bone strength independent of BMD. Presented at the 34th European Symposium on Calcified Tissues, Copenhagen, Denmark, 2007. Published in *Calcified Tissue International* **80** (Suppl 1):s25, (Abstract).

AWARD

JIN H, Stewart TL, Van 't Hof R, Aspden RM and Ralston SH. A rare haplotype in the 5' flank of the COLIA1 gene is enriched in hip fracture patients and reduces bone strength independent of BMD. Presented and received Young Investigator Award at the 34th European Symposium on Calcified Tissues, Copenhagen, Denmark, 2007.

ABBREVIATIONS

1,25 $(OH)_2D_3$	1, 25 dihydroxyvitamin D3
129ola	An inbred mouse strain
3'	3 prime
5'	5 prime
χ^2	Chi-squared
µg	Microgram
µl	Microlitre
µM	Micromolar
A	Adenosine
ABI	Applied Biosystems
ANOVA	Analysis of variance
AP	Alkaline phosphatase
APOSS	Aberdeen prospective osteoporosis screening study
BMD	Bone mineral density
BMI	Body mass index
BMP2	Bone Morphogenic Protein 2
bp	Base pair
BSA	Bovine serum albumin
C	Cytidine
cDNA	Complementary DNA
cm	Centimetre
cM	Centimorgan
COL1A1	Collagen type I α 1
COL1A2	Collagen type I α 2
CSF-1	Colony Stimulating Factor 1
ddNTP	Dideoxynucleotide triphosphate
DNA	Deoxyribonucleic acid
DNase I	Deoxynuclease I
dNTP	Deoxynucleotide triphosphate
DTT	DL-dithiothreitol
DXA	Dual energy X-ray absorptiometry
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetra-acetic acid
EMSA	Electrophoretic mobility shift assay
ES cells	Embryonic stem cells
ERα	Estrogen receptor alpha
FN	Femoral neck
G	Guanosine
g	Grams
GENOMOS	Genetic Markers for Osteoporosis

GLM	General linear model
Hos TE85	A human osteoblast-like cell line
HRT	Hormone replacement therapy
HWE	Hardy-Weinberg equilibrium
IL	Interleukin
kb	Kilobase
LB	Luria Bertani
LD	Linkage disequilibrium
LRP5	Lipoprotein Receptor Related Protein 5
LS	Lumbar spine
M	Molar
mg	Milligram
MG63	A human osteoblast-like cell line
ml	Millilitre
mM	Millimolar
mRNA	Messenger RNA
MTHFR	Methylenetetrahydrofolate Reductase
NE	Nuclear extract
neo	Neomycin
OI	Osteogenesis imperfecta
OPG	Osteoprotegerin
p	Probability
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
pIC50	Inhibitory concentration at 50%
poly [d(I-C)]	Poly-deoxy-inosinic-deoxy-citidylic acid
PTH	Parathyroid hormone
QPCR	Quantitative PCR
QTL	Quantitative trait locus
RANK	Receptor for activation of nuclear factor kappa B
RANKL	Receptor for activation of nuclear factor kappa B ligand
RNA	Ribonucleoc acid
RNase	Ribonuclease enzymes
rpm	Revolutions per minute
RT	Reverse transcription
SD	Standard diviation
SDM	Site-direct mutagenesis
sec	Second
SEM	Standard error of the mean
SNP	Single nucleotide polymorphism
Sp1	Promoter-specific transcription factor-1
T	Thymidine
TBE	Tris-Boric acid-EDTA buffer
TDT	Transmission disequilibrium test

TGF-β1	Transforming Growth Factor Beta-1
TNF	Tumour necrosis factor
Tris-	Tris (Hydroxymethyl) aminoethane
U.K	United Kingdom
UV	Ultra violet
v	Volts
VDR	Vitamin D Receptor Gene
WHO	World health organization
y	year

CONTENTS

List of tables and figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

CHAPTER ONE

Introduction

1.1 Bone

1.1.1 Bone Composition

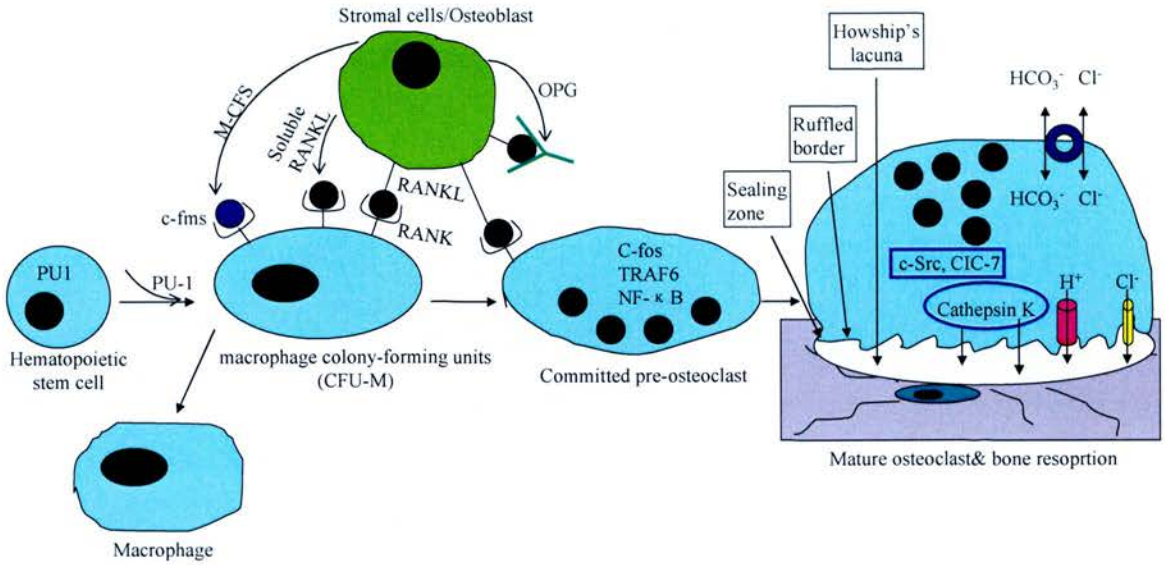
Bones provide structural support to the body; protect many internal organs from injury and act as an important reservoir for calcium and phosphate in the regulation of mineral homeostasis. The skeleton is made up of two different types of bone tissue called cortical bone and trabecular bone. The cortical bone is a dense layer of calcified tissue which covers the external surface of long bones. Within cortical bone there are neurovascular channels known as "Haversian canals", each one of which contains blood vessels and nerve fibres surrounded by concentric rings (lamellae) of bone matrix. Trabecular bone is metabolically more active than cortical bone and it is found mainly in the vertebrae, pelvis, flat bones and the end of long bones. It is made up of an interconnecting network of bone tissue interspaced with irregular cavities that contain red and yellow marrow, bone cells and other cell types. The most abundant protein in bone is type I collagen, a fibrillar protein composed of two identical alpha1 polypeptides and one alpha2 polypeptide bound to form a triple helix. Collagen comprises 90% of the protein in bone matrix and the fibres are organized in a preferential orientation to give a high density of collagen per unit of bone. The remaining 10% of bone matrix proteins consist of various non-collagenous proteins and glycoproteins which are thought to play an important role in bone mineralization and in regulating bone cell activity. The inorganic content of bone is composed mostly of calcium hydroxyapatite crystals ($Ca_{10}(PO_4)_6(OH)_2$) which confer properties of hardness and rigidity. The main cellular components of bone are osteoblasts (bone-forming cells), osteoclasts (bone-resorbing cells) and osteocytes which act as mechanosensors.

1.1.2 Osteoclasts and Bone Resorption

Osteoclasts are giant multinucleated cells derived from cells of the monocyte-macrophase lineage. Differentiation of osteoclasts from heamatopoietic precursor cells requires many key factors (Figure 1.1). The transcription factor PU-1 is necessary to commit the precursors into the myeloid lineage at early stage by binding to the promoter region of Csf1r (encoding M-CSF receptor) and positively regulating transcription (Zhang DE *et al*, 1994). Mice deficient in PU-1 display an osteopetrosis (*op*) phenotype similar to the *op/op* mice (Tondravi MM *et al*, 1997). Further differentiation of the osteoclast progenitor cells is controlled by macrophage colony stimulating factor (M-CSF). M-CSF binds to its receptor, c-fms, to simulate differentiation and proliferation of haematopoietic precursor cells to macrophage colony-forming units (CFU-M) which are the common precursors of macrophages and osteoclasts. At this stage, these factors are essential for stimulating differentiation and activation of mature osteoclasts including the Nuclear factor Kappa B (NF-κB, Franzoso G *et al*, 1997 & Iotsova V *et al*, 1997), the Receptor for Activation of NF-κB (RANK, Dougall WC *et al*, 1999 & Li J *et al*, 2000)), the RANK Ligand (RANKL, Kong YY *et al*, 1999), Tumor Necrosis Factor (TNF) receptor associated factor 6 (TRAF6, Lomaga MA *et al*, 1999 &Kobayashi T *et al*, 2003), Nuclear Factor of Activated T-cell, cytoplasmic, calcineurin-dependent 1 (NFATc1, Asagiri M *et al*, 2005), c-fos (Johnson RS *et al*, 1992 & Wang ZQ *et al*, 1992), and Fc receptor common γ subunit (FcRγ) / DNAX-activating protein (DAP) 12 (Koga T *et al*, 2004 & Mocasai A *et al*, 2004). The deficiency of any of these factors leads to the loss of multinucleated osteoclasts causing severe osteopetrosis despite a normal or increased number of monocyte/macrophages. For example, Koga T *et al* (2004) showed that mice lacking FcRγ and DAP 12 exhibited severe osteopetrosis due to impaired

osteoclast differentiation. FcRγ and DAP12 are required for the induction of NFATc1, which is the crucial step in the RANKL-induced osteoclast differentiation through activation of calcium signaling.

Figure 1.1 Mechanism of osteoclast differentiation and osteoclastic bone resorption.



The close contact between osteoclast precursors and stromal cells or osteoblasts is also important in osteoclastogenesis. The osteoblasts express RANKL, M-CSF and a decoy RANK receptor, osteoprotegerin (OPG). OPG blocks the interaction of RANK (expressed on the membrane of osteoclast) and RANKL by competitively binding to RANKL resulting in inhibition of osteoclast formation (Figure 1.1). It was found that mice lacking OPG have accelerated osteoclastogenesis and develop severe osteoporosis (Simonet *et al*, 1997). RANKL is also expressed in abundance by activated T lymphocytes, which can directly trigger osteoclastogenesis and are probably pivotal to the joint destruction in rheumatoid arthritis (Kong YY *et al*, 1999). Therefore, the balance between the expression of the stimulator of osteoclastogenesis (RANKL) and of the inhibitor (OPG), influences osteoclast formation and hence bone

resorption (Teitelbaum SL, 2000). In fact, most osteoclastogenic factors such as parathyroid hormone (PTH), tumour necrosis factor (TNF), and Interleukin-1 (IL-1) exert their effects mainly by regulating the expression of RANK, RANKL, OPG and M-CFS. It is clear that PTH stimulates osteoclastogenesis and the activity of mature osteoclasts by increasing M-CSF (Suda T *et al*, 1999) and RANKL (Yasuda H *et al*, 1998) and inhibiting OPG (Horwood NJ *et al*, 1998). TNF-α and IL-1 also stimulate bone resorption primarily by stimulating M-CSF production and by directly increasing RANKL expression (Walsh MC & Choi Y, 2003).

The mature osteoclasts exert their effects on bone resorption. Osteoclasts attach to the mineralized bone surface by forming a tight seal which creates an isolated extracellular microenvironment known as Howship's lacuna (Figure 1.1). A ruffled border with dense patches on each site (sealing zone) also forms along the plasma membrane facing the mineralized surface of bone (Figure 1.1). The importance of the ruffled border was demonstrated in mice with a targeted disruption of the c-Src proto-oncogene, which developed osteopetrosis due to non-functional osteoclasts (Soriano P *et al*, 1991). Subsequent experiments showed that Src-deficient mice have multinuclear osteoclasts which attach to bone via a normal sealing zone, but the osteoclasts lack a ruffle border (Boyce BF *et al*, 1992 & Lowe C *et al*, 1993).

Demineralization of bone matrix is initiated by acidification of the isolated extracellular microenvironment via a vacuolar H^+ -adenosine triphosphatase (H^+ -ATPase) and Cl^- channel in the cell's ruffled membrane. This acidic milieu first mobilizes bone matrix by digestion and dissolving the hydroxyapatite crystals. Then, the demineralized organic component of bone is degraded by collagenase or a lysosomal protease, Cathepsin K at low pH.

The matrix-degradation products are endocytosed by the osteoclast and transported to and released at the cell surface opposite to the ruffled border (Nesbitt & Horton, 1997). Therefore, bone resorption helps to maintain calcium and inorganic phosphate levels in the plasma and the concentration of hydroxyproline and N-terminal or C-terminal collagen peptide fragments in the serum or urine can be used as indirect measurements of bone resorption.

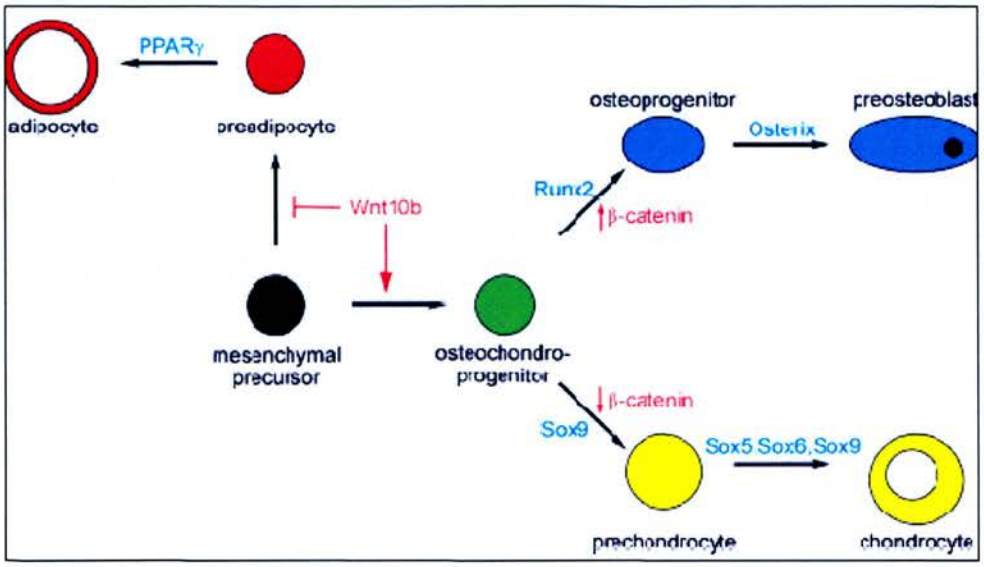
Several proteins are directly involved in the regulation of this resorption process, such as CLC-7 (*CLCN7*, Kornak U *et al*, 2001), Atp6i (*TCIRG1*, Li YP *et al*, 1999 & Frattini A *et al*, 2000) and Cathepsin K (Saftig P *et al*, 1998; Gowen M *et al*, 1999 & Kiviranta R *et al*, 2005). Mice deficient in these molecules have osteoclasts with no or very little bone resorbing activity, which indicates that these molecules are crucial for osteoclast function.

1.1.3 Osteoblasts and Bone Formation

Osteoblasts are cuboidal bone cells responsible for the production of the bone matrix and they are derived from mesenchymal stem cells. Undifferentiated mesenchymal stem cells are capable of differentiating into many cell types including chondrocytes, adipocytes, myoblasts, stromal cells and fibroblasts. As shown in figure 1.2, these precursors also undergo proliferation and differentiation into preosteoblasts and mature osteoblasts under the influence of local growth factors such as Wnt pathway proteins, and essential transcription factors such as runt-related transcription factor 2 (Runx2) and Osterix (Ducy P *et al*, 1997; Nakashima K *et al*, 2002).

The critical role of Runx2, also known as Cbfa1 (Core-binding factor a 1), in osteoblast differentiation was revealed by in *vivo* studies ofRunx2 null-mutant mice that had a cartilaginous skeleton with complete absence of osteoblasts (Komori, T *et al*, 1997 & Otto F *et al*, 1997). Runx2 regulates the expression of genes that are specific to the osteoblast phenotype such as osteocalcin and alkaline phosphatase (Ducy P *et al*, 1997). Thus, Runx2 is the earliest and most specific marker of osteogenesis. Osterix acts downstream of Runx2 during osteoblastic differentiation. Osx-null mice develop a perfectly patterned skeleton composed of normal cartilage, lacking osteoblasts and mineralized bone matrix (Nakashima K *et al*, 2002).

Figure 1.2 Essential factors for osteoblastogenesis.



High levels of canonical Wnt signaling (as determined by levels of β-catenin) with the presence of transcription factors Runx2 and Osterix are essential for osteoblastogenesis (adapted from Glass DA 2nd & Karsenty G, 2007).

Mature osteoblasts synthesize the various organic components of bone matrix, including the major product, type I collagen and a variety of noncollagenous proteins. Two of these noncollagenous proteins are useful serum markers of osteoblastic activity, and hence, bone formation in metabolic disease. One product is alkaline phosphatase (ALP), a metalloprotein essential for mineralization. ALP deficiency, as in the disease hypophosphatasia, results in defective mineralization of bone and teeth (Whyte MP, 1994). Osteocalcin is another noncollagenous protein and widely accepted as a clinical marker of bone turnover, as osteocalcin is incorporated into the bone matrix and released into the circulation during bone resorption. Osteocalcin is thought to inhibit the mineral deposition, as osteocalcin-deficient mice are shown to have increased BMD compared with normal mice (Ducy P *et al*, 1996). However, the precise role of osteocalcin remains unclear. Regarding the major synthetic product of osteoblasts, type I collagen, its degradation products are used as markers of bone

resorption, whereas levels of collagen propeptides are used as markers of bone formation.

Importantly, the osteoblast has receptors for parathyroid hormone (Karaplis AC & Goltzman D, 2000), prostaglandins, estrogen, vitamin D3 and several cytokines. Steroid and polypeptide hormones regulate the growth of osteoprogenitors and/or their progression to mature osteoblasts. PTH increases bone formation *in vivo* by increasing the number of the osteoprogenitor cells (Aubin JE & Heersche JNM, 2001), decreasing apoptosis of preosetoblasts and osetoblasts (Jilka RL *et al*, 1999), and increasing osteoblast proliferation (Canalis E *et al*, 1989), whereas *in vitro* experiments showed that PTH stops preosteoblasts from becoming mature osteoblasts and inhibits the production of collagen and other matrix proteins (Tintut Y *et al*, 1999).

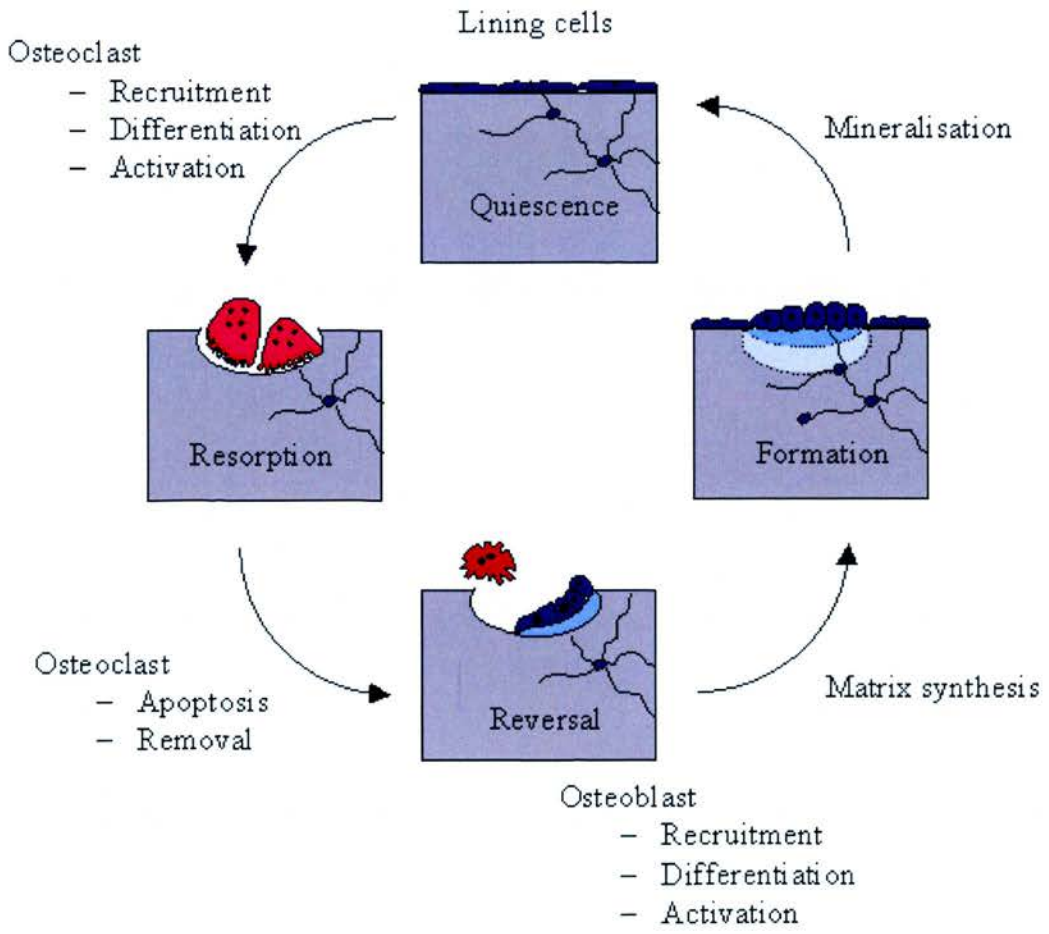
Osteoblasts are also a major source of the Colony Stimulating Factor 1 (CSF-1) and secrete some cytokines, such as RANKL and OPG, which regulate osteoclastogenesis.

Mature osteoblasts form new bone by secreting unmineralized bone matrix (called osteoid) which subsequently undergoes calcification (the osteoid maturation period) by deposition of hydroxyapatite crystals (Baron R, 1999). Osteoblasts do not function individually, but form a group of cells at the uncalcified bone matrix that they are producing. Behind the osteoblasts one or two layers of cells can usually be found: activated mesenchymal cells and preosteoblasts. Once osteoblasts become surrounded by bone matrix, some of these cells differentiate into osteocytes, whereas others differentiate into flat lining cells which cover the bone surface. Osteocytes are the most abundant cell type in bone and are thought to act as sensors of mechanical strain in bone (Ehrlich PJ & Lanyon LE, 2002).

1.1.4 Bone Remodelling and its regulation

Bone is constantly undergoing a process of bone removal and repair termed bone remodelling which is mediated by complex interactions between osteoclasts, osteoblasts and their precursors. This process is essential in maintaining the mechanical integrity of the skeleton by which damaged or weak bone is replaced by new bone. There are four sequential events involved in this bone remodelling process: quiescence, resorption, reversal, and formation which are shown in Figure 1.3 (Baron R, 1999).

Figure 1.3 The bone remodelling sequence.



Bone formation and bone resorption are tightly coupled, and the total amount of bone in the skeleton is determined by the balance between these two processes. During early stages of skeletal growth until peak bone mass is achieved in early adulthood, the rate of bone remodelling is high with bone formation exceeding resorption. Between the age of 20-40 peak bone mass is achieved and resorption equals formation. Following this, bone is removed more rapidly than new bone tissue is laid down, resulting in a steady decrement in bone mass with age and an increased risk for osteoporosis. This process occurs in both sexes, but men and women are affected differently. At the time of the menopause women lose bone more rapidly due to estrogen deficiency, which causes uncoupling of bone formation and resorption. After that, the loss slows down, but continues. In men the loss of bone mass is slower. But by age 65 or 70 men and women are losing bone at the same rate.

However, in diseases such as primary hyperparathyroidism, hyperthyroidism and Paget's disease, osteoclast-mediated bone resorption is activated with subsequent compensatory increased new bone formation, resulting in a disorganized mosaic of woven and lamellar bone, which is more susceptible to deformity or fracture than normal bone. Therefore, the balance between bone formation and resorption in the remodeling process is very important in maintaining skeletal integrity.

Bone remodeling is under the control of several systemic hormones, locally produced cytokines and growth factors. IL-6, a pleiotropic cytokine, is expressed and secreted by osteoblast, osteoclast and stromal cells in response to osteotropic hormones such as PTH, $1,25(OH)_2$-D_3, and IL-1. IL-6 has important effects on bone remodeling by stimulating bone resorption (Moonga BS *et al*, 2002) but also by promoting osteoblast generation in conditions of high bone turnover (Sims NA *et al*, 2004).

TGF-β inhibits osteoblast differentiation yet stimulates the proliferation of mesenchymal progenitors, thereby expanding the cell population that will differentiate into osteoblasts (Alliston TN & Derynck R, 2000). TGF-β also has a fundamental role in the control of bone resorption, because it directly enables osteoclast formation by antagonising inflammatory signalling (Fox SW *et al*, 2000a; Fox SW *et al*, 2003 & Lovibond AC *et al*, 2003), and indirectly suppresses osteoclast formation thorough its ability to regulate RANKL/OPG expression (Murakami *et al*, 1998; Thirunavukkarasu *et al*, 2001 & Quinn *et al*, 2001). During bone remodelling, low levels of TGF-β stimulate osteoclast formation. But with continuous release of TGF-β from bone matrix during the resorption process, TGF-β inhibits osteoclast activity and stimulates the proliferation of osteoblasts to initiate formation of new bone. Therefore, TGF-β acts as a central component coupling bone formation prior to resorption during bone remodeling.

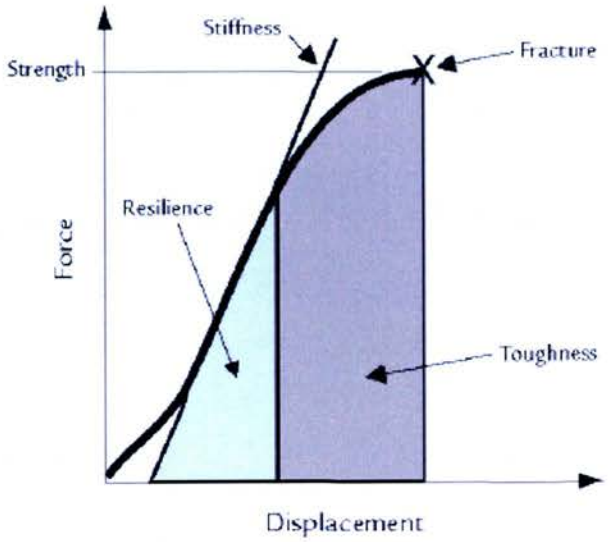
Mechanical stimuli are also thought to have an effect on initiation of bone remodeling and determining the site at which remodeling occurs. It has been proposed that not only the mechanical stimuli directly deform bone matrix and cells, but also the changes in interstitial fluid flow stimulate osteocytes, which occurs when a load is applied (Raisz LG, 1999).

1.1.5 Bone Strength

The strength of bone depends on its mass but also on the structure and material properties of the bone tissue. Currently, the most commonly used tool for diagnosing osteoporosis is measurement of BMD by dual-energy X-ray absorptiometry (DXA) at the spine, wrist, or proximal femur. However, BMD in itself is not sufficient to accurately predict fracture risk or monitor treatment effects of an anti-osteoporosis drug, as it does not measure variations in bone quality (structure and material properties) which may alter bone strength independently of its mass.

There are a number of biomechanical parameters used to describe bone quality. The key relationship between these parameters is the force-displacement curve resulting from a mechanical test of loading a bone specimen until it breaks (Turner CH & Burr DB, 1993; Turner CH, 2002; Figure 1.4). Each of these parameters reflects a different property of bone: ultimate force reflects the general integrity of the bone structure; stiffness is the primary property of bone and closely related to the mineralization of the bone; toughness is the amount of energy necessary to break the bone and determined by the collagen phase of bone matrix; and ultimate displacement is inversely related to the brittleness of the bone (Felsenberg D & Boonen S, 2005). Therefore, the material properties of bone play an important role in improving the overall bone strength. The most obvious clinical example of the mechanical effects of a collagen defect is osteogenesis imperfecta (OI). OI is a family of heritable disorders caused by mutations in the type I collagen genes (*COL1A1* & *COL1A2* genes) (Byers PH, 2001; Byers PH & Cole WG, 2002; Rowe DW, 1991). People with OI have a markedly increased risk of multi-fractures at a young age, which results from a combination of poor material properties of bone (abnormal quantity and structure of type I collagen), with low bone density and thin long bone cortices.

Figure 1.4 Force-displacement curve for bone tissue.



The height of the curve represents strength (the maximum load the material can sustain); the area under the curve is toughness (the energy required to break the material) and resilience (stored elastic energy); and the maximum slope of the curve is stiffness (initial reaction to a load, a factor important for bending and buckling). (adapted from Felsenberg D & Boonen S, 2005)

In addition to the bone mass and material properties, structural properties of bone also contribute to bone strength. In general, the size of bone appears to have an effect on overall fragility. Large, dense bones are less susceptible to osteoporosis and fractures than small bones. Smaller vertebral bone size has been found in patients with spine fracture who had the same area BMD as matched controls (Silva MJ & Gibson LJ, 1997). Whereas in patients with high bone mass (about 5SDs above normal) caused by mutations in the LPR5 gene, individuals are protected from osteoporotic fractures (Little RD *et al*, 2002).

In long bones, the differences in the extent of periosteal apposition and endocortical resorption during growth and aging contribute to differences in bone strength, because these differences in the long bone structure lead to variations in the diameter and cortical thickness of bone and to the distance of cortical mass from the neutral axis.

For example, mechanical loading preferentially increases periosteal bone formation in the region where mechanical stresses are highest to increase the bone mechanical strength. Moreover, the increased periosteal bone formation also could compensate by increasing bone rigidity when endosteal resoption is stimulated by the effect of aging (Turner CH, 2006).

Differences in bone dimensions also contribute to sex-based and race-based differences in bone strength. For instance, men and women generally have similar vertebral trabecular volumetric density and similar vertebral heights, but men have larger vertebral cross-sectional area which correlates with higher bone strength.

Therefore, bone strength is not only determined by its mass, but also bone quality such as material and structural properties. All these properties can be altered by the bone remodeling process to adapt to the prevailing loads. When the adaptation fails, bone becomes fragile and susceptible to fracture.

1.2 Osteoporosis

Osteoporosis is a complex disease characterized by loss of bone mass, micro-architectural deterioration of bone tissue, enhanced bone fragility and an increased fracture risk. Clinically, osteoporosis is recognized by the occurrence of low trauma fractures, most commonly at the hip, spine, and distal forearm. It is an important public health issue because about 50% of women and 20% of men older than 50 years in the white population will have a fragility fracture in their remaining lifetime (Department of Health and Human Services, 2004). The combined annual costs of all osteoporotic fractures have been estimated to be $20 billion in the USA and about $30 billion in the European Union (Cummings SR & Melton LJ, 2002). In the UK, the lifetime risk of symptomatic fracture for a 50 year old white woman has been estimated as 13% for the forearm, 11% for the vertebrae and 14% for the femoral neck. The annual cost to the NHS of managing osteoporotic fractures is £1.7 billion, with over 90% of this figure attributable to hip fracture (http://www.mrc.soton.ac.uk/). Osteoporotic fractures result from a combination of decreased BMD and deterioration in bone microarchitecture. There is a strong relationship between bone mass and fracture risk: for every standard deviation (SD) decrease in BMD, the risk of osteoporotic fracture increases by a factor of 1.5-2. The causes of osteoporosis include low peak bone mass, rapid bone loss at the menopause and accelerated bone loss due to diseases or drugs. Other factors such as diet (calcium intake and alcohol consumption), life style (smoking, physical activity and mechanical loading) also influence the risk of osteoporosis.

As described in section 1.1.4, bone mass increases during childhood and puberty and achieves its peak by the age of 20-30 years old. After that, bone mass remains static

between the ages of 20 to 45, but then bone loss occurs because the rate of bone turnover increases and the imbalance between bone formation and resorption widens. In addition, bone architectural change also occurs along with bone loss, which impairs bone strength further, such as fatigue damage, and loss of trabecular plates and connectivity.

1.3 Evidence for a Genetic Contribution to the Pathogenesis of Osteoporosis

Whilst osteoporosis is a multi-factorial disease with several environmental influences, genetic factors play an important role in its pathogenesis. Twin and family studies have shown the importance of genetics in determination of BMD and other osteoporosis-related phenotypes including biochemical markers of bone turnover, skeletal geometry, ultrasound properties of bone, body mass index (BMI), and muscle strength (Table 1.1).

The genetic contribution to the osteoporotic fracture has also been studied. Family history of hip fracture has been consistently shown to be a risk factor for fracture, independent of BMD in population-based studies. For example, Seeman E *et al* have shown that daughters of women with hip fractures have an increased risk for hip fractures (Seeman E *et al*, 1994). This was supported by a longitudinal study on 9516 white women aged 65 years or older, which found that a maternal history of hip fracture doubles the risk of hip fracture independently of BMD (Cummings SR *et al*, 1995). In keeping with this, the heritability of wrist fracture was estimated to be 25% and 54%, respectively, on the basis of family and twin studies and this was found to be largely independent of BMD (Deng HW *et al*, 2000a; Andrew T *et al*, 2005). This indicates that the predisposition may have been mediated through genetic influences on bone turnover, and bone geometry or other factors such as the risk of falling.

The study on a large cohort of 6021 Swedish twins has shown that the heritability of overall age-adjusted fracture was less than 20%, but it was greater for first hip fractures before the age of 69 years (0.68; 95% CI, 0.41-0.78) and between 69 and 79 years (0.47; 95% CI, 0.04-0.62). But the heritability of hip fracture dropped off rapidly with age to almost zero after 79 years old (0.03; 95% CI, 0.00-0.26)

(Michaelsson K *et al*, 2005). This suggested that genetic factors play an important role in the pathogenesis of osteoporotic fractures, but environmental factors become more important with age. This could explain why another elderly twin study from Finland showed little evidence to suggest that fractures were heritable (Kannus P *et al*, 1999).

Table 1.1 Estimated heritability (h^2) values for osteoporosis related phenotypes

Phenotype	h^2	Reference
BMD	0.5-0.85	Pocock NA *et al*, 1987; Smith DM *et al*, 1973 Krall EA & Dawson-Hughes B, 1993
Fracture (Colles's fracture)	0.25-0.54	Deng HW *et al*, 2000 Andrew T *et al*, 2005
Biochemical markers of bone turnover (in women)	0.59-0.75	Garnero P *et al*, 1996 Hunter D *et al*, 2001
Skeletal genometry (HAL)	0.62	Arden NK *et al*, 1996
Ultrasound properties of bone	0.53	Arden NK *et al*, 1996
BMI	0.63-0.82	Kaprio J *et al*, 1995 Carmichael CM & McGue M, 1995
Muscle strength	0.46	Arden NK & Spector TD, 1997
Age at menarche	0.37	Kaprio J *et al*, 1995 Snieder H *et al*, 1998
Age at natural menopause	0.63	
Age at surgical menopause	0.59	

Genetic factors are the main determinant of peak bone mass. Evidence from twin and family studies have indicated that between 50% - 85% of the variance in peak bone mass is genetically determined, depending on skeletal site and the age of the subjects studied (Smith DM *et al*, 1973; Pocock NA *et al*, 1987; Krall EA & Dawson-Hughes B, 1993; Gueguen R *et al*, 1995). It is clear that several genes regulate bone mass, each with a modest effect, and likely candidate genes include type I collagen genes (*COL1A1* & *COL1A2*), vitamin D receptor, *LRP5* gene and others, which will be described in the 'candidate gene' section (section1.4.3.4, page 39). Therefore, peak bone mass is determined by the combined effects of heritability and various

environmental factors including nutrition (especially calcium and caloric intake), physical activity and sex steroid status. However the overall contribution of genetic factors to age-related bone loss is less clear. One of the reasons for these difficulties is that non-genetic factors, such as changes in hormone levels, nutritional intake and co-existing diseases, may interact with genetic variations to eventually modulate bone loss.

In summary, it is currently believed that in the general population, genetic influences on BMD and the other phenotypes mentioned above are polygenic in nature and are mediated by the influence of several genetic variants each of modest effect size and their interaction with environmental factors. However, osteoporosis, fragility fractures or unusually high bone mass are features of several rare monogenic diseases including osteopetrosis, sclerosing bone dysplasias, osteogenesis imperfecta (OI), the osteoporosis-pseudoglioma syndrome (OPS) and osteoporotic syndromes associated with inactivating mutations of the estrogen receptor alpha and aromatase genes (Bilezikian JP *et al*, 1998; Janssens K & Van Hul W, 2002; Van Wesenbeeck L *et al*, 2003; Balemans W *et al*, 2005). These diseases are rare, but some of the subtle polymorphic variations in genes that are mutated in these diseases may also contribute to regulation of BMD in the general population. For example, the inactivating mutations and activating mutations in low density lipoprotein-related receptor-5 gene (*LRP5*) are responsible for OPS and autosomal dominant high bone mass, respectively. The *LRP5* gene polymorphisms have recently been shown to be associated with BMD, bone area and stature particularly in men (Ferrari SL *et al*, 2004a).

1.4 Approaches to Identify Osteoporosis Genes

1.4.1 Linkage Analysis

Linkage analysis can be used to identify regions of genome that might contain disease genes. The method follows the segregation of chromosomal regions marked by genetic variants in affected family associated with the disease or phenotype. Linkage analysis can be applied to both monogenic diseases with major gene effects and complex diseases with polygenic effects.

Linkage analysis is a well validated method to identify the responsible genes for rare monogenic bone diseases which are segregating according to Mendelian laws. From 1996 onward, a number of linkage studies suggested the presence of at least one important genetic locus for the regulation of bone mass on chromosome 11q12–13 because different types of monogenic bone conditions were assigned to this chromosomal region, including autosomal recessive osteoporosis pseudoglioma (OPPG) syndrome (Gong Y *et al*, 1996), the autosomal dominant high bone mass phenotype (Johnson ML *et al*, 1997) and autosomal dominant osteopetrosis type I (Van Hul E *et al*, 2002). Follow-up genetic studies revealed that mutations in the gene encoding low-density lipoprotein receptor-related protein 5 (*LRP5*), a coreceptor for Wnt molecules, are at the basis of all the above mentioned conditions (Gong Y *et al*, 2001; Boyden LM *et al*, 2002; Little RD *et al*, 2002). Following studies reported that various missense mutation of the *LRP5* gene also cause endosteal hyperostosis, Worth disease, van Buchem disease, autosomal dominant osteosclerosis, and autosomal dominant osteopetrosis type I (Boyden LM *et al*, 2004; Whyte MP *et al*, 2005; Kwee ML *et al*, 2005; Rickels MR *et al*, 2005; Balemans W *et al*, 2006).

Linkage analysis also has been used to detect chromosome regions which harbour genes regulating quantitative traits such as BMD in sib-pairs and families. The advantage of this method is that it has the potential to detect new genes and new molecular pathways that regulate bone metabolism. Some quantitative trait loci (QTL) regulating BMD have been found in genome-wide linkage mapping (Table 1.2), which gives valuable information about the likely chromosomal locations for genes involved in the regulation of BMD. However, linkage analysis can only identify large regions that are typically tens of cM containing hundreds of genes. So far it has been difficult to narrow down the region to map the specific gene contributing to the linkage peak at the majority of the QTL. This is because few QTL have been replicated between different genome searches in different studies, perhaps because of limited small sample size and limited statistical power to find genes or genetic variants having modest effects on BMD. There are only a few genome-wide linkage mapping studies for osteoporosis- related traits in large or extended multigenerational families (Styrkarsdottir *et al*, 2003; Kammerer *et al*, 2003).

Moreover, the involvement of the numerous environmental factors makes it more difficult to detect the small effects contributed by some biologically important genes to osteoporosis in the general population by standard linkage mapping approaches.

Table 1.2 Chromosome regions implicated by genome-wide linkage scans of BMD (adapted from Zmuda JM *et al*, 2006)

Region	Locus	Phenotype	LOD	Ethnicity	Comments
1p36		Femoral neck BMD	3.53	Caucasian (US)	42 families (N=254)
1p36		Whole body BMD	2.40	Caucasian (US)	1094 female twin pairs (N=2188)
1q		Spine BMD	4.3	Caucasian (US)	938 sisters
1q21-23		Spine BMD	3.64	Caucasian (US)	464 sister pairs (N=706)
2q23-24		Hip BMD	2.25	Caucasian (US)	7 pedigrees (N=149)
2q23-24		Forearm BMD	2.15	Chinese	96 nuclear families (N=218)
2q25		Femoral neck BMD	3.98	Mexican American	29 pedigrees (N=664)
3p21		Spine BMD	2.1-2.7	Caucasian (US)	1094 female twin pairs (N=2188)
4p		Forearm BMD	4.33	Mexican American	29 extended pedigrees (N=664)
4q31		Spine BMD	3.08	Caucasian (US)	53 extended pedigrees (N=630)
4q32		Wrist BMD	2.26	Caucasian (US)	53 extended pedigrees (N=630)
4q34		Hip BMD	2.95	Caucasian (US)	7 pedigrees (N=149)
5q33-35		Femoral neck BMD	2.23	Caucasian (US)	464 sister pairs (N=706)
6p11-12		Spine BMD	2.13	Caucasian (US)	464 sister pairs (N=706)
6p21		Femoral neck BMD	2.93	Caucasian (US)	330 extended pedigrees (N=1557)
8q24		Wards BMD	2.13	Caucasian (US)	330 extended pedigrees (N=1557)
10q26		Hip BMD	2.29	Caucasian (US)	53 extended pedigrees (N=630)
11pter-14		Whole body BMD	2.08	Caucasian (US)	1094 female twin pairs (N=2188)
11q12-13	LRP5	Spine BMD	5.74	Caucasian (US)	Single extended kindred
11q24		Spine BMD	2.08	Caucasian (US)	7 pedigrees (N=149)
12q23		Spine BMD	2.08	Caucasian (US)	330 extended pedigrees (N=1557)
12q24		Forearm BMD	2.53	Mexican American	29 extended pedigrees (N=664)
12q24		Spine BMD	2.96	Caucasian (US)	53 extended pedigrees (N=630)
13q		Femoral neck BMD	2.51	Mexican American	29 extended pedigrees (N=664)
13q14-22		Trachanter BMD	3.46	Mexican American	29 extended pedigrees (N=664)
13q33-34		Spine BMD	2.43	Caucasian (US)	53 extended pedigrees (N=630)
14q		Trachanter BMD	3.5	Caucasian (US)	774 sister pairs
14q31		Spine BMD	2.08	Caucasian (US)	330 extended pedigrees (N=1557
15q		Femoral neck BMD	4.3	Caucasian (US)	774 sister pairs
16p12-q23		Spine BMD	2.11	Caucasian (US)	1094 female twin pairs (N=2188)
20p12.3	BMP2	Spine+Hip BMD	4.93	Icelandic	207 extended pedigrees (N=1323)
21q		Trachanter BMD	3.14	Caucasian (US)	330 extended pedigrees (N=1557)
22q12-13		Spine BMD	2.13	Caucasian (US)	464 sister pairs (N=706)

So far only one candidate gene for BMD has emerged from a genome-wide linkage study and this is the bone morphogenetic protein 2 (*BMP2*) gene (Styrkarsdottir *et al*, 2003). An initial linkage study linked spine and hip BMD to chromosome 20p12. Subsequent work showed that *BMP2* contributed to part of the linkage peak due to an association between osteoporotic fracture and a serine - alanine amino acid change at codon 37 in the *BMP2* gene in Icelandic pedigrees and a Danish cohort of postmenopausal women. Further analysis showed that the Ser37Ala polymorphism was in LD with a polymorphism in a conserved *cis*-acting regulatory element in the 3'- untranslated region (UTRs) of *BMP2* gene (Fritz *et al*, 2004). It was suggested that this may cause subtle alterations in BMP2 protein level (Fritz *et al*, 2006). However, in a large population-based cohort study of 6353 men and women from the Rotterdam Study, there was no association of the *BMP2* polymorphisms Ser37Ala and Arg190Ser with various osteoporosis related phenotypes including BMD, bone loss, hip structure analysis and fracture risk (Medici M *et al*, 2006). Therefore, the *BMP2* gene can be considered the first osteoporosis gene identified by genome-wide linkage analysis, but more research will be needed to determine if it is associated with osteoporosis in other populations.

1.4.2 Genome-wide Association Analysis

Genome-wide association studies have been proposed as a potentially powerful strategy for detecting genes in complex diseases and may be a promising strategy for identifying osteoporosis susceptibility genes. This method involves mapping hundreds of thousands of single nucleotide polymorphisms (SNPs) throughout the whole genome in either a case-control study or a population based study to find variants that are associated with a disease-related phenotype. It has greater power to detect smaller effects than linkage analysis, but requires many more markers to be examined.

There are an estimated 10 million common SNPs in the human genome. It will be extremely expensive to screen all the common variants in a large sample. Fortunately evidence based on the international HapMap project (www.hapmap.org) suggests that a carefully chosen 300,000 to 500,000 SNPs should be sufficient to represent most common variants across the entire human genome, because recently it has been revealed that much of the genome may fall into segments of strong linkage disequilibrium (LD), where variants within these segments are strongly associated with each other (Reich *et al*, 2001; Daly *et al*, 2001; Gabriel *et al*, 2002). Therefore common variants are either genotyped directly or are assessed by analysis of an allele of single "tagging" SNP or a combination of alleles at several "tagging" SNPs. Another strategy for reducing the number of SNPs to be tested is based on direct association and the genotyping of all potentially functional SNPs across the genome (Peltonen & McMkusick, 2001; Bonstein & Risch, 2003). This method decreases the number of SNPs that need to be genotyped to 50,000 to 100,000, though additional efforts will be needed to validate all these SNPs (ie, non-synonomous, non-conservative coding changes). Very recently, the platforms for genotyping such large set of SNPs throughout the genome simultaneously have become available now. For

example, the new Affymetrix® Genome-Wide Human SNP Array 6.0 features more than 1.8 million markers for genetic variation, including more than 906,600 SNPs and more than 946,000 probes for the detection of copy number variation (CNV) (www.affymetrix.com). Similarly, the new Human1M BeadChip features over one million markers of SNPs and copy number variation (CNV) (www.illumina.com). These and the advances in the international HapMap project have made genome-wide association analysis feasible.

The first genome-wide association study was reported by Ozaki K *et al* (2002). In this study 92,788 gene-based SNPs were screened in a case (n=94) and control (n=658) population and 2 SNPs in the lymphotoxin-alpha (*LTA*) gene were found to be significantly associated with myocardial infarction. One SNP changed an amino-acid residue from threonine to asparagine (Thr26Asn) while another SNP in intron1 influenced the transcription level of *LTA*.

Recently, in a genome-wide association study conducted by Klein et al (2005), a polymorphism in the complement factor H gene (*CFH*) was found to be associated with age-related macular degeneration (AMD) among 116,204 SNPs genotyped. This polymorphism changed an amino-acid from tyrosine to histidine (T402H) in a region that binds heparin and C-reactive protein, while this gene is located on chromosome 1 in a region repeatedly linked to AMD in family-based studies.

Furthermore, a joint genome-wide association study has also been successfully performed by the Wellcome trust Case Control Consortium for seven common diseases using the Affymetrix GeneChip 500K Mapping Array Set. This study has examined ~2000 individuals for each of seven major diseases and a shared set of ~3000 controls. Case-control comparisons identified 25 independent association signals at $p<5x\ 10^{-7}$: 1 in bipolar, 1 in coronary artery disease, 9 in Crohn's disease, 3

in rheumatoid arthritis, 7 in type 1 diabetes and 3 in type 2 diabetes (Wellcome Trust Case Control Consortium, 2007). The following replication studies have confirmed that almost all of these signals reflect genuine susceptibility-effects (Zeggini E *et al*, 2007; Parkes M *et al*, 2007;Samani NJ *et al*, 2007; Frayling TM *et al*, 2007; Todd JA *et al*, 2007). The importance of appropriately large samples was confirmed by the modest effect sizes observed at most loci identified. This study thus represents a thorough validation of the genome-wide association approach.

So far genome-wide association analysis has not yet been completed for osteoporosis related traits. With the advances in high-throughput genotyping methods and the increases in the number of validated SNPs in the human genome, genome-wide association analysis will be cheaper, faster and more accessible in the near future. More gene regions are expected to be detected by this analysis with higher density SNP arrays and larger number of cases-controls.

It is also important to note, however, multifactorial genetic susceptibility may also often be due to the independent cumulative effects of a set of relatively low frequency variants at several different loci. Each of these will contribute only a small amount to the overall genetic determination of a trait, but the cumulative contribution of such variants may be quite substantial (Fearnhead NS *et al*, 2004). However, these "rare variants" will not be detected by association studies based on linkage disequilibrium with linked common SNPs, such as the genome-wide association analysis. The only productive approach to find such rare variants is to search for DNA variation in carefully chosen candidate genes in selected disease groups and compare variant frequencies, individually and in aggregate, between disease and control population.

1.4.3 Candidate Gene Association Analysis

It has become clear that genetic susceptibility to common complex diseases such as osteoporosis involves many genes, most of which have small effects. Candidate gene association studies are relatively easy to perform, have greater power to detect small allelic effects than linkage studies, and have been widely used to identify osteoporosis susceptibility genes. The classical candidate gene association study involves identifying polymorphisms of a particular candidate gene and relating alleles to BMD or osteoporotic fractures in a population based study or a case control study. Candidate genes include those that regulate calcium metabolism, those that are mutated in rare monogenic bone diseases or those that have been mutated, over-expressed or deleted in animal models and resulted in abnormal bone phenotypes.

However the major shortcoming of most association studies in the osteoporosis field is lack of power due to the small sample size. Other limitations and potential confounding factors include the genetic and population heterogeneity, multiple comparisons and population stratification. Family-based tests such as the transmission disequilibrium test (TDT) can help to avoid the false positive association due to the unrecognized population admixture (Risch N & Teng J 1998; Guo W & Fung WK, 2005), but it may be impractical to collect large families with late-onset diseases such as osteoporosis or fracture to carry out such studies.

Another approach is meta-analysis. Meta-analysis is a statistical technique for combining the findings across independent studies. It is a powerful tool for quantifying the true effect size of candidate genes on complex disease and addressing the unexpected diversity in genetic association studies, for example, by investigating heterogeneity (Fang Y *et al*, 2006; Uitterlinden A *et al*, 2006). Given the fact that candidate genes for osteoporosis have small effects, thousands of subjects are required

to achieve a satisfactory power. Although a single laboratory might not be able to obtain such numbers and many association studies with small sample size were not reproductive due to the low power, meta-analysis has the potential to overcome this problem. For example, the GENOMOS (Genetic Markers for Osteoporosis) project involves analyzing several candidate gene polymorphisms in relation to osteoporotic fractures and BMD in a large-scale study. The GENOMOS meta-analysis showed no associations between three polymorphisms of the estrogen receptor gene and BMD but a modest effect on fracture risk independent of BMD in 18,917 subjects (Ioannidis *et al*, 2004). Another GENOMOS meta-analysis of five common polymorphisms of vitamin D receptor gene in 26,242 subjects reported one observed association between Cdx2 polymorphism and the risk for vertebral fracture, but no effect on BMD (Uitterlinden A *et al*, 2006). The GENOMOS meta-analysis demonstrated that the Sp1 polymorphism of the type I collagen gene (*COL1A1*) was associated with reduced BMD and could predispose to incident vertebral fracture in women, independent of BMD, which is the largest study ever performed for the *COL1A1* Sp1 polymorphism with over 20,000 subjects (Ralston SH *et al*, 2006). In addition to being a very large study of genetics of osteoporosis, the GENOMOS meta-analysis also successfully avoids publication bias, selective reporting bias and between-study heterogeneity. Thus, such a meta-analysis approach might be able to identify the true effect size of individual genetic risk factors for complex disease.

It is important to remember that the gene polymorphisms associated with the disease might not have a causal role but instead be in LD with a nearby causal variant. Recent advances in the HapMap project provide a valuable resource for the design of candidate gene association studies which allows genotyping higher density of markers to improve the chances of detecting the true causal associations. These linked variants,

which tend to be inherited together on a region of a chromosome, are so called haplotypes. Once the common haplotype alleles within a candidate gene have been established, the causal variant can be located in the haplotype block within the gene. However, like mentioned previously, haplotype analysis has limitations in detecting the rare variants that might cause disease.

Numerous candidate genes have been associated with BMD or osteoporotic fracture (Table 1.3) including genes encoding hormones and receptors, cytokines and receptors, bone matrix components and degradative enzymes, osteoblast regulatory factors and osteoclast related genes. The most widely studied candidate genes such as collagen type I gene, the vitamin D receptor gene, the estrogen receptor genes and LRP5 are discussed below in details.

Table 1.3 Candidate genes for regulation of susceptibility to osteoporosis (adapted from Ralston SH & Crombrugghe B, 2006)

Hormones and receptors	**Bone matrix components & degradative**
Androgen receptor	Alpha HS2 glycoprotein
Aromatase	Collagen type 1 alpha I
Calcium sensing receptor	Osteocalcin
Calcitonin receptor	Proline lysine oxidase dehydrogenase type
Estrogen receptor alpha	Collagenase
Estrogen receptor beta	
Parathyroid hormone	
Vitamin D receptor	
Cytokines and receptors	**Osteoblast regulatory factors**
Interleukin-1 alpha and beta	Alox12 and Alox15
Interleukin-1 receptor antagonist	Bone morphogenic protein 2
Interleukin-6	Bone morphogenic protein 4
Tumour necrosis factor receptor 2	Core binding factor A1
Tumour necrosis factor	Insulin like growth factor 1
	Sclerostin
	Transforming growth factor beta
	Lipoprotein related receptor 5 and 6
	Peroxisome proliferators activated receptor gamma
Miscellaneous	**Osteoclast related genes**
Apolipoprotein E	Cathepin K
Methylene tetrahydrofolate reductase	Chloride channel 7
Klotho	Vacuolar proton pump a3 subunit
HLA class II	Osteoprotegerin
	Receptor activator of NF κ B
	Receptor activator of nuclear factor kappa B ligand

1.4.3.1 Type I Collagen Gene (*COL1A1*)

Type I collagen is the major protein of bone. It is a heterotrimer formed from two collagen α1(I) and one collagen α2(I) polypeptides that are encoded by the genes *COL1A1* and *COL1A2* respectively. Mutations within the exons of these genes are thought to be the major cause of osteogenesis imperfecta (OI), a hereditary disease characterized by increased bone fragility and premature osteoporosis (Row DW, 1991). These mutations lead to structural or quantitative defects of type I collagen. Many mutations affect the conserved glycine residues in either *COL1A1* or *COL1A2*, which results in a mixture of normal and abnormal collagen fibrils being produced. Such mutations result in excessive brittleness of bone. Mutations altering a splice site creating a premature stop code within *COL1A1* gene result in the reduced rate of collagen production, because the abnormal transcription products are destroyed due to nonsense mediated RNA decay, and only normal collagen chains are produced (Rauch F & Glorieux FH, 2004). There is a mouse model of osteogenesis imperfecta (OI) that was developed as a model of severe human OI and affected animals suffered from fragility fractures, and were shown to have abnormal orientation of collagen fibers and altered size of mineral crystals (Viguet-Carrin S *et al*, 2006).

In view of this, the collagen genes are strong candidate genes for osteoporosis. Grant *et al* described a common polymorphism affecting the transcription factor Sp1 binding site in the first intron of *COL1A1* gene that was associated with low bone mass and an increased risk of fracture (1996). Positive associations between the Sp1 polymorphism and BMD, fractures and postmenopausal bone loss were subsequently reported in several populations reviewed by Raslton SH & de Crombrugghe B (2006). Retrospective meta-analyses of published studies concluded that carriage of the T

allele is associated with reduced BMD at the lumbar spine and femoral neck and with vertebral fractures (Mann V *et al*, 2001, Efstathiadou Z et al, 2001, Mann V & Ralston SH, 2003),

Moreover, homozygote for the T allele of the Sp1 polymorphism was recently found to be associated with BMD and incident vertebral fractures in the GENOMOS study, which is the largest ever meta-analysis performed to study genetics of osteoporosis with over 20,000 subjects (Ralston SH *et al,* 2006). Although the GENOMOS study only detected a modest association between the Sp1 polymorphism, BMD and fractures, it demonstrated the importance of conducting studies that have adequate power to detect modest effects of common genetic variants on complex diseases. In addition, the GENOMOS study also found that the association between vertebral fracture and the Sp1 polymorphism was not fully explained on the basis of reduced BMD, which indicated that the Sp1 may act as a marker for bone quality. This is consistent with previous findings that the Sp1 predicts fractures independent of BMD and interacts with BMD to enhance fracture prediction (McGuigan EF *et al*, 2001). Furthermore, the Sp1 polymorphism has also been associated with other phenotypes relevant to osteoporosis, including femoral neck geometry (Qureshi AM *et al*, 2001), bone quality (Mann V *et al*, 2001), bone mineralization (Stewart TL *et al*, 2005) and the therapeutic response to etidronate therapy (Qureshi AM *et al*, 2002).

The frequency of the Sp1 polymorphism differs between ethnic groups. In Caucasian populations where osteoporosis is a common disease, the Sp1 polymorphism was found to be relatively common with an allele frequency of 22%, whereas it is rare in Africans and Asians, where the incidence of osteoporosis is lower than in Caucasian populations. It is not clear whether this relationship is incidental or one of cause and effect.

The mechanism by which the Sp1 polymorphism predisposes to osteoporosis has been investigated by *in vitro* functional studies. Mann V *et al* (2001) reported that the risk allele 'T' of the Sp1 polymorphism had increased affinity for Sp1 protein binding and was associated with elevated allele-specific transcription in heterozygotes, which abnormally increased production of collagen type I alpha 1 mRNA and α1 chain of collagen by osteoblasts. This abnormality led to an increased ratio of the α1 to α2 chains and reflected the presence of α1 homotrimer formation which is mechanically weaker than normal α1 (I)/ α2 (I) heterotrimers because of altered inter-molecular cross-linking (Misof K *et al*, 1997). The α1 homotrimer was observed in a mouse model of osteogenesis imperfecta with a mutation in the *COL1A2* gene which caused a conformational change in the α2 chain and prevented α2 chain associating with α1 chain. The homozygotes for this mutation had severe bone fragility where bones are composed completely of α1 homotrimers, whereas heterozygotes had a mild phenotype because the bones consist of a mixture of α1 homotrimers and normal α1 (I)/ α2 (I) heterotrimers (Chipman SD *et al*, 1993). In keeping with this, biomechanical testing of bone samples from heterozygotes for the Sp1 alleles showed reduced bone strength compared with homozygotes for 'G' allele and a slight reduction of mineralization of bone (Mann V *et al*, 2001). Taken together, these data suggest that the Sp1 polymorphism is a functional variant that has adverse effects on bone composition and mechanical strength.

Recently, two polymorphisms were identified in the promoter of the *COL1A1* gene that are in LD with the Sp1 polymorphism. These are a G/T polymorphism at position -1997 relative to the transcription start site (-1997G/T; rs1107946) and an insertion/deletion polymorphism at position -1663 (-1663indelT; rs2412298). These polymorphisms were found to interact with each other and with the Sp1

polymorphism to regulate BMD in a small study of Spanish postmenopausal women (García-Giralt N *et al*, 2002). The -1997G/T promoter polymorphism has been studied in relation to BMD in other populations and family based studied with mixed results although most of these studies have been of limited sample size (Stewart TL *et al*, 2006). Haplotype analysis has shown that an extended haplotype defined by the Sp1 polymorphism and two promoter polymorphisms may exert stronger effects on BMD than the individual polymorphism (García-Giralt N *et al*, 2005; Stewart TL *et al*, 2006). Subsequent work found these two promoter polymorphisms influence *COL1A1* transcription in a promoter-reporter assay and the -1663indelT polymorphism interacts with the transcription factor Nmp4 which plays a role in osteoblast differentiation by interacting with BMP/Smad signaling (García-Giralt N *et al*, 2005). Whilst a functional interaction has been reported between these two promoter polymorphisms in regulating *COL1A1* transcription (Garcı´a-Giralta N *et al*, 2005), future work will be required to establish them as functional variants and determine whether they interact with Sp1 polymorphism to regulate BMD and gene transcription.

1.4.3.2 Vitamin D Receptor Gene (*VDR*)

Vitamin D, through its biologically active metabolite 1, 25-Dihydroxyvitamin D3 [1,25-$(OH)_2D_3$], plays a crucial role in calcium homeostasis by interacting with its receptor to regulate bone cell growth and differentiation, intestinal calcium absorption and parathyroid hormone secretion. The vitamin D receptor (*VDR*) was the first candidate gene to be associated with bone-related phenotype (serum osteocalcin levels, Morrison NA & Eisman JA, 1992). Since then, a number of studies have focused on three polymorphisms in the 3' untranslated region (UTR) of the *VDR* gene recognized by restriction enzymes *Bsm*I, *Apa*I and *Taq*I, which were associated with

postmenopausal bone loss, circulating levels of osteocalcin and BMD in a twin study and a population-based study. A positive association has also been found between a haplotype defined by these three polymorphisms and osteoporotic fracture in one study (Uitterlinden A *et al*, 2001). But other studies have shown conflicting results of *VDR* alleles in relation to BMD or fracture (Ferrari SL & Rizzoli R, 2005). A retrospective meta-analysis of *Bsm*I polymorphism confirmed the association of spine BMD with relatively modest effects, but no association with femoral neck BMD (Thakkinstian A *et al*, 2004). Another meta-analysis of thirteen studies including 1632 fractures and 5203 controls showed no evidence of a relationship between the *VDR* *Bsm*I or *Taq*I polymorphism and fracture risk with any genetic model and significant between-study heterogeneity (Fang Y *et al*, 2006). It still remains unclear whether these are functional variants or not, although one study suggested that these polymorphisms may act as markers for RNA stability because they are in LD with a polymorphic polyA tract in the 3' UTR of *VDR* mRNA (Morrison NA *et al*, 1994).

Gross C *et al* (1996) described a polymorphism in exon 2 recognized by the *Fok*I restriction enzyme, which creates an alternative translational start site, resulting in the production of two isoforms of the VDR proteins differing in length by three amino acids. The *Fok*I polymorphism is not in LD with the three polymorphisms in the 3' UTR. The association of the *Fok*I polymorphism with BMD has been found in some populations (Arai H *et al*, 1997; Gennari L *et al*, 1999; Bandrés E *et al*, 2005 & Falchetti A *et al*, 2007), but in not others (Eccleshall TR *et al*, 1998; Langdahl BL *et al*, 2000). Functional studies have also yielded inconclusive results (Arai H *et al*, 1997; Jurutka PW *et al*, 2000; Colin EM *et al*, 2000 & Gross C *et al*, 1998).

Recently, another polymorphism has been identified in the promoter of the *VDR* gene, which is located in a binding site for the transcription factor Cdx-2 (Arai H *et al*,

2001). Functional differences showed that the "A" allelle had increased binding of the Cdx-2 protein and increased transcription activity of the *VDR* promoter compared with the "G" allele in the human colonic carcinoma cell line Caco-2. This polymorphism has been associated with BMD in a Japanese population (Arai H *et al*, 2001) and with fracture in other populations (Fang Y *et al*, 2003 & Casado-Díaz A *et al*, 2007). A large-scale haplotype study involving 6418 subjects of the Rotterdam Study has shown that certain haplotypes defined by polymorphisms in the promoter and 3' UTR were associated with an increased risk of fracture (Fang Y *et al*, 2005). This result was supported by functional studies that indicated the risk haplotypes resulting in a lower *VDR* mRNA level due to decreased transcription (associated with promoter haplotypes) and increased degradation (associated with 3' UTR haplotypes). However, the GENOMOS study involving 26,242 subjects, has shown no association of the *Fok*I or any polymorphism in the 3' UTR with BMD, but the Cdx2 polymorphism was associated with risk for vertebral fractures (Uitterlinden A *et al*, 2006). Therefore, the complex mechanism by which these polymorphisms predisposed to fracture or BMD remains unclear.

1.4.3.3.Estrogen Receptor Alpha Gene (*ESR1*)

Estrogen deficiency following the menopause is a major risk factor for osteoporosis and it exerts its effect on bone through interaction with the estrogen receptors. Estrogen receptor alpha, encoded by the *ESR1* gene, is a strong functional candidate in osteoporosis. A number of studies have looked for evidence of an association between *ESR1* alleles and BMD, mostly focussing on three polymorphisms: a TA repeat in the promoter and two polymorphisms in the first intron which are recognized

by *Xba*I and *Pvu*II restriction enzymes respectively. An association between the TA repeat in the *ESR1* promoter and BMD has been found in both Japanese and US populations. Associations of the *Xba*I and *Pvu*II polymorphisms with BMD have also been observed in postmenopausal and younger pre-menopausal Asian women. But inconsistent results have been reported possibly because of small sample size and population heterogeneity (reviewed by Williams & Spector, 2006). A meta-analysis of published studies including more than five thousand women from 22 eligible studies concluded that homozygotes for the *Xba*I XX genotype have significantly higher BMD and a reduced risk of fractures compared with other genotypes (Ioannidis JP *et al*, 2002). More recently, a prospective meta-analysis of data from over 18,000 subjects of eight European centres in the GENOMOS study, confirmed the association of the *Xba*I XX genotype with reduced risk of fractures. The effects on fracture seemed to be independent of BMD in this study, however, and no association was observed with BMD (Ioannidis JP *et al*, 2004). The mechanism responsible for the increased fracture risk remains unclear. The haplotype defined by the polymorphisms *Pvu*II and *Xba*I of *ESR1* alleles have been associated with other bone-related phenotypes including ultrasound properties of bone and bone loss (Albagha O *et al*, 2005). This suggested that the *ESR1* gene might have an effect on bone quality. Other evidence from *in vitro* functional studies showed that the *Xba*I & *Pvu*II polymorphisms may regulate reporter gene transcription (Herrington DM *et al*, 2002; Maruyama H *et al*, 2000). Interestingly, the *Pvu*II polymorphism is located in consensus recognition sites for the AP4 and Myb transcription factors, and both the *Xba*I and *Pvu*II polymorphisms are located in a region that is 70-80% conserved in the human, mouse and rat genomes. However, this does not exclude the possibility that these polymorphisms are in LD with the causal ones elsewhere in the *ESR1* gene.

1.4.3.4 Lipoprotein Receptor Related Protein 5 (*LRP5*)

The LRP5 and LRP6 proteins function as co-receptors for canonical Wnt signalling. Wnt signalling plays an important role in the development and maintenance of many organs and tissues. In particular, the canonical or Wnt/β-catenin signaling through LRP5 has been identified as a key pathway in the regulation of BMD and bone strength (Johnson ML *et al*, 2004). Early studies showed that two rare mendelian bone disorders, osteoporosis-pseudoglioma syndrome (OPPS, Gong Y *et al*, 1996) and high bone mass syndrome (HBM, Johnson ML *et al*, 1997) mapped to the same region of chromosome 11q12-13. Subsequent work using positional cloning identified two types of mutations in the *LRP5* gene as the cause for these two disorders (Gong Y *et al*, 2001 & Little RD *et al*, 2002). The OPPS syndrome is characterized by severe, early onset osteoporosis and congenital blindness due to vitreous opacity and is caused by loss-of-function mutations in *LRP5* due to different homozygous missense, nonsense, and frameshift mutations throughout the gene (Gong Y *et al*, 2001). Conversely, gain-of-function mutations in the *LRP5* gene lead to HBM, which is an asymptomatic autosomal dominant disorder characterized by abnormally high BMD. This was reported in association with a heterozygous missense mutation causing a substitution of valine for glycine at codon 171 (G171V) of *LRP5* within the first β-propeller domain of the molecule (Little RD *et al*, 2002). Since then, several other missense mutations have been identified as a cause of HBM, and all of these cluster in or around the first β-propeller domain of the *LRP5* gene (Van Wesenbeeck L *et al*, 2003).

In addition to studies in humans, transgenic mice expressing truncated or non-functional LRP5 protein have deficits in bone formation and develop the clinical features observed in OPPS patients (Kato M *et al*, 2002). Conversely, transgenic mice

that over-express wide type LRP5 or a gain-of-function mutation (G171V) have increased bone mass and strength (Babij P *et al*, 2003). These mice showed increased mineral apposition rate, a reduced rate of osteoblast apoptosis but unaffected bone resorption. Several functional studies have suggested that HBM-associated mutations (G171V, G171R, A214T, A214V, A242T, T253I, and D111Y) had lower Dickkopf-1 (Dkk1) binding affinity, and impaird Dkk1-mediated inhibition of Wnt-stimulated LRP5 signalling (Boyden LM *et al*, 2002 & Ai M *et al*, 2005). Taken together, these data indicate that the *LRP5* has an important role in determining bone mass, strength and function.

Moreover, a QTL for BMD in the general population was also mapped at 11q12-13, the *LRP5* locus (Koller DL *et al*, 1998; Carn G *et al*, & Livshits G *et al*, 2002). Several association studies and family-based study have been performed and found various common polymorphisms to be associated with BMD and osteoporotic fractures. Interestingly, many studies have reported strong associations with BMD in men (Ferrari SL *et al*, 2004a; Koh *et al*, 2004; Urano *et al*, 2004; Van Meurs *et al,* 2006). This effect might be explained by gender differences in mechanical loading. Recently the V667M and A1330V polymorphisms were specifically associated with LS peak bone mass in the physically active subgroup of young, Danish men from the Odense Androgen Study, which suggests a role for LRP5 as a mediator of load-induced bone formation (Brixen K *et al*, 2006). A subsequent study found that genetic variations in exons 10 and 18 of the *LRP5* gene modulate Wnt signaling and the relationship between physical activity and BMD in men, which suggests that Wnt-LRP5 may play a role in the adaptation of bone to mechanical load in humans, and may explain some gender-related differences in bone mass (Kiel DP *et al*, 2007).

The mechanism by which these polymorphisms affect LRP5 signalling has not been investigated, but an interaction between the *LRP5* A1330V polymorphism and a coding polymorphism of the *LRP6* (1062V) has been observed with additive effect on fracture susceptibility in the Rotterdam Study (Van Meurs *et al,* 2006). In conclusion, these data suggested that not only do rare mutations in the *LRP5* gene affect BMD in rare diseases; the subtle common polymorphisms also contribute to the genetic regulation of BMD in the normal population. This also makes other components of this pathway such as Wnts, Dkks, Frizzled, Kremen, and Sfrps, candidate genes for osteoporosis in further study,

1.4.3.5 Transforming Growth Factor Beta-1 (TGF-β1)

TGF-β1 is a multifunctional polypeptide produced by immune cells but it is also abundant in bone matrix. It has been suggested that TGF-β1 modulates proliferation and differentiated functions of osteoblasts and also plays a role in coupling bone resorption to bone formation (Noda M & Camilliere JJ, 1989; Bonewald LF & Mundy GR, 1990). Mutations in the *TGFBI* gene cause Camurati-Engelmann disease (CED), a condition associated with increased bone turnover and osteosclerosis affecting the diaphysis of the long bones that is inherited in an autosomal dominant manner (Kinoshita A *et al*, 2000; Janssens K *et al*, 2000).

Most causal mutations are located within the coding region of the latency-associated peptide (LAP). TGF-β1 is normally secreted as a complex composed of three proteins, which are the bioactive peptide TGF-β1, latency-associated peptide-β1 (LAP-β1), and latent TGF-β1 binding protein-1. TGF-β1 forms a complex with LAP-β1 noncovalently, which is called the small latent complex, and in this configuration

TGF-β1 is unable to bind to its receptors (Annes JP *et al*, 2003). Functional studies have shown that the CED-causing mutations impair binding of the inhibitory LAP to mature TGF-β1, resulting in activation of SMAD signaling (Janssens K *et al*, 2003). Polymorphisms of the *TGFBI* gene have been studied extensively in relation to osteoporosis, including polymorphisms in promoter; intron 4 & 5 and exon 1 & 5. These polymorphisms of *TGFBI* gene have been associated with BMD, osteoporotic fractures and biochemical markers of bone turnover in some populations (Langdahl BL *et al*, 1997; Yamada Y *et al*, 1998; Yamada Y *et al*, 1999; Hinke V *et al*, 2001; Keen RW *et al*, 2001; Langdahl BL *et al*, 2003;Dick IM *et al*, 2003) , but not in others (Ziv E *et al*, 2003; McGuigan FE *et al*, 2007). These conflicting results may be explained by the differences in the ethnic makeup of the populations studied and small sample size.

1.4.3.6 Bone Morphogenic Protein-2 (*BMP2*)

Bone morphogenetic protein-2 (BMP-2) is a growth factor located on chromosome 20p12 that belongs to the TGF-β superfamily. BMP-2 plays a critical role in early embryogenesis, skeletal development, and osteoblast differentiation (Thies RS *et al*, 1992; Kanzler B *et al*, 2000; Spinella-Jaegle S *et al*, 2001). Recent studies have also linked the *BMP2* gene to osteoporosis. Turgeman G *et al* (2002) reported that systemic administration of recombinant human *BMP2* abrogated bone loss in two mouse models of osteoporosis. Polymorphisms of the *BMP2* gene have also been associated with regulation of BMD and susceptibility to osteoporotic fractures. Styrkarsdottir U *et al* (2003) identified a locus for regulation of BMD on chromosome 20p12 by a linkage study in an Icelandic population, and subsequently identified a

missense polymorphism (Ser37Ala) located in exon 2 of the *BMP2* gene associated with osteoporotic fractures by positional cloning. This polymorphism (Ser37Ala) was overrepresented in Icelandic and Danish patients with osteoporosis (Styrkarsdottir U *et al* 2003). The association of the Ser37Ala polymorphism with lumbar spine BMD was also found by Reneland RH *et al* in an international sample of 805 women from Australia, New Zealand, United Kingdom, and Belgium (2005). But associations of the Ser37Ala or the Arg189Ser (located in exon3) polymorphism with osteoporosis were not observed in the Rotterdam Study (Medici M *et al*, 2006) or in healthy American whites (Ichikawa S *et al*, 2006). The mechanisms by which these polymorphisms affect BMP signalling and bone mass have not yet been investigated. A functional study suggested that one polymorphism (rs15705) disrupted a putative posttranscriptional regulatory motif within the highly conserved 3'-untranslated region of *BMP2* gene, altered BMP2 RNA decay *in vitro* and thus affect BMP2 protein expression level (Fritz DT *et al*, 2006).

1.4.3.7 Interleukin-6 (IL-6)

Interleukin-6 (IL-6) is a pleiotropic cytokine produced by osteoblasts, monocytes and T cells in bone. IL-6 plays an important role in post-menopausal bone loss by promotion of osteoclast differentiation and activation (Manolagas SC, 1998). Polymorphisms in the *IL-6* gene promoter region have been associated with *IL-6* gene expression levels (Ferrari SL *et al*, 2003; Fishman D *et al*, 1998; Terry CF *et al*, 2000), serum levels of IL-6 (Brull DJ *et al*, 2001; Fishman D *et al*, 1998) and C-reactive protein (CRP) (Ferrari SL *et al*, 2003), a marker of systemic IL-6 activity. Moreover, the *IL-6* polymorphisms have also been associated with low BMD, reduced ultrasound

properties of bone and increased risk of fractures (Feng D *et al*, 2003; Ferrari SL *et al*, 2003; Nordstrom A *et al*, 2004) in elderly women. In addition, the -174GG genotype was overrepresented in younger women and men with reduced peak BMD (Garnero P *et al*, 2002; Lorentzon M *et al*, 2000), and estrogen deficient women (Ferrari SL, *et al*, 2004b) Significant interactions were also found between the *IL-6* -174 genotypes and years since menopause, estrogen status, dietary calcium and vitamin D intake in women in the Offspring Cohort of the Framingham Heart Study, that comprises 1574 unrelated men and women (mean age 60 years) (Ferrari SL, *et al*, 2004b).

1.4.3.8 Methylenetetrahydrofolate Reductase (MTHFR)

MTHFR is an enzyme which catalyses the reduction of 5, 10-methylenetetrahydrofolate to 5-methylenetetrahydrofolate which is important in homocysteine metabolism. Mutations affecting this enzyme are one of the causes of homocysteinuria, a rare autosomal recessive disease characterized by excessive homocysteine in plasma and urine. Some studies have also shown a relationship between BMD or fracture incidence and homocysteine (Gjesdal CG *et al*, 2006; McLean RR *et al*, 2004). This makes *MTHFR* gene a candidate gene for osteoporosis. A common nonsynonymous C/T polymorphism (C677T) of the *MTHFR* gene results in an alanine-to-valine substitution, which causes a mild decrease in enzyme activity. The relationship between the C677T genotype and osteoporosis has been investigated in several populations, but the results are conflicting. For example, the association of C677T genotype with BMD was observed in postmenopausal Japanese women (Miyao M *et al*, 2000), postmenopausal women in Danish Osteoporosis Prevention Study (Abrahamsen B *et al,* 2003), but not in another Danish population (Jorgenssn

HL *et al*, 2002), Iranian women (Golbahar J *et al*, 2004), or postmenopausal British women (Baines M *et al*, 2007). A subsequent meta-analysis of published studies concerning the C667T genotype and BMD between the 2000-2006, found that individuals with the TT genotype showed a small but significantly reduced BMD compared to those with the TC/CC genotypes in women (Riancho JA *et al*, 2006). Only a few studies on the relationship between *MTHFR* genotypes and fractures have been published, and the results were mixed (Bathum L *et al*, 2004; Jorgensen HL *et al*, 2002; Villadsen MM *et al*, 2005).

1.4.4 Genetic Analysis of Mice

1.4.4.1 Genetic Linkage Studies in Mice

Genetic studies in mice including quantitative trait loci (QTL) mapping, transgenesis and gene inactivation ("knock out") have provided powerful tools for dissecting the genetics of osteoporosis and other diseases. Genetic mouse models have several advantages: mice are cheap and easy to manipulate; environmental factors can be controlled; large number of progeny may be generated and mouse genetic resource are quite extensive.

Genetic linkage studies in mice have succeeded in identifying several QTLs regulating osteoporosis-related traits. To perform QTL mapping, typically two different inbred strains of mice with low and high BMD are crossed to produce a first generation (F1). Then the F1 are inbred to generate a second generation (F2). The F2 mice have varying levels of BMD because of segregation of the alleles that regulate BMD in the F1 x F1 cross. All of the F2 progeny undergo phenotype assessment and then DNA samples are obtained for the genome-wide linkage analysis. The putative QTL region identified in linkage analysis is then narrowed by using a variety of breeding schemes, such as Interval Specific Congenic Recombinant (ISCR) strains which allows fine mapping of a QTL into a 1cM interval. This design is based on systematic marker-directed construction of a series of congenic strains (for example, Frankel *et al.* 1995; Morel *et al.* 1996; Yui *et al.* 1996), each is recombinant at a specific 1-cM sub-interval out of a series of small tandem intervals covering a chromosomal region, to which a QTL was previously mapped (Darvasi A, 1997). It has been shown that a specific and previously detected QTL of moderate or even small effect can be accurately mapped into a 1-cM interval in a program involving a

total of no more than 1000 individuals (Bennett B *et al*, 2002; Ehringer MA *et al*, 2002; Ruf C *et al*, 2004). This reduced interval size will facilitate identification of candidate genes, through bioinformatics, gene expression, and DNA sequencing strategies.

Functional studies are then required to demonstrate a candidate gene underlying the QTL of interest, such as gene expression analysis (RT-PCR, microarray), RNA interference and transgenic mice (knockouts, knockins). For example, recently the candidate gene *Alox15* was identified as a negative regulator of peak BMD by microarray analysis in inbred mouse strains. The mouse chromosome 11 region which harbours the *Alox15* gene, was strongly linked to peak BMD in a genome-wide linkage study (Klein RF *et al*, 1998). To identify genes that might regulate BMD in this region, Klein RF *et al* (2004) generated a DBA/2 (D2) background congenic mouse with an 82-megabase (Mb) region of chromosome 11 replaced by the corresponding region of the C57BL/6 (B6) genome. The congenic mice with the B6 chromosome 11 region had increased peak BMD and improved measures of femoral shaft strength relative to heterozygous or D2 littermates. After narrowing down the BMD QTL to a 31-Mb region on chromosome 11 by linkage analysis, the gene expression in B6 and D2 mice was analyzed by microarray analysis. The *Alox15* gene was identified as the only differentially expressed gene within this region, which was confirmed by RT-PCR analysis. The *Alox15* gene encodes 12/15-lipoxygenase (12/15-LO), an enzyme mediating endogenous ligands for the peroxisome proliferator activated receptor-γ (PPARγ) which is involved in a pathway that stimulates adipogenesis and inhibits osteoblastogenesis (Huang JT *et al*, 1999; Kuhn H *et al*, 2002; Lecka-Czernik B *et al*, 2002; Khan E & Abu-Amer Y 2003). Phenotyping of

the *Alox15* knockout mice showed that 12/15-LO-deficient mice had significantly higher whole body BMD and increased biomechanical indices of femoral shaft strength than that of mice homozygous for the 12/15-LO D2 allele. Moreover, an *in vivo* study showed that pharmacologic inhibitors of this enzyme improved bone density and strength in two rodent models of osteoporosis (Klein RF *et al*, 2004). A recent study has reported that polymorphic variation in a human homolog of this gene (*Alox12*) also regulates BMD in man (Ichikawa S *et al*, 2006). This association still needs to be replicated in other population and the molecular mechanism needs to be investigated by functional studies.

Many QTLs regulating BMD were reported for genome-wide linkage studies in mice, but so far only one gene (*Alox15*) has been identified by positional cloning in inbred strains in mice to date. As the linkage studies in human, the typical size of a QTL identified by linkage analysis is too large to detect the genes or gene variants with modest effects on osteoporosis-related traits. In addition, the genetic variants are essentially in LD with each other making identification of the causal variant difficult. Moreover, the most obvious drawback of this approach is that the QTLs regulating osteoporosis- related traits in mouse models are not always have the same effects in human.

1.4.4.2 Transgenic Mouse Models

Recently, the development of transgenic and gene-targeting technology combined with the sequencing of the mouse genome has dramatically increased our ability to create mouse models in which the functional significance of genes, gene variants, and quantitative trait loci (QTL) can be determined. A transgenic mouse is an organism that has had foreign DNA (transgene) introduced into one or more of its cells artificially, where the transgene is stably maintained, transmitted and expressed. The foreign DNA introduced can be a gene from humans or another species which is inserted into the mouse genome (transgenesis mouse model), or an endogenous gene which is disrupted (knockout mouse model) or mutated (knockin mouse model). This can be achieved in a number of ways: microinjection of DNA into the pronucleus, homologous recombination in embryonic stem cells, retroviral transfer and nuclear transfer.

Pronuclear injection of fertilized eggs is the most common and convenient way to make transgenic mice. Several hundreds copies of the transgene are injected directly into the male pronucleus of the recently fertilized host egg. These oocytes are subsequently transferred into the uterus of pseudo-pregnant recipient mice. The offspring is screened to confirm a copy of the transgene in every cell. These successful offspring are called founders. Since the integration of the transgene into the host genome is random, the transgene can integrate anywhere in the genome and multiple copies often integrate at a single locus in a head-to-tail fashion. In addition to insertional mutations, deletions and complex rearrangements of the host genome can occur. Therefore, only a small proportion of the mice born are transgenic and only a proportion of those express the gene well. The success rate of producing transgenic mice by this method is very low. However the microinjection is the predominant

transgenesis method. It has been often used to characterize the ability of a promoter to drive a direct tissue-specific gene expression. In this case, a reporter gene is placed under the control of the target promoter and the reporter gene product's activity is quantitatively measured. Commonly used reporter genes include the gene which encodes green fluorescent protein, luciferase gene and *lac Z* gene which encodes the protein beta-galactosidase. Another major use for transgenic mice produced by pronuclear injection of DNA is to examine the effects of over-expressing and misexpressing endogenous or foreign genes at specific times and locations in the animal.

Less random than other methods, embryonic stem cell transfer is based on homologous recombination. Embryonic stem cells (ESCs) are stem cells derived from the inner cell mass of a blastocyst, which is an early stage embryo. ESCs are pluripotent, meaning they are able to develop into almost any type of tissue when given sufficient and necessary stimulation for a specific cell type. When given no stimuli for differentiation, ESCs will continue to divide in vitro and each daughter cell will remain pluripotent. Embryonic stem cell transfer involves isolation of stem cells from a blastocyst, transfection of the transgene (by electroporation), selection of the stem cells in which the transgene has undergone homologous recombination at the desired locus and injection of the selected stem cells containing the desired DNA into an embryo of the blastocyst-stage. The resulting chimera will undergo further selective breeding for homozygous expression of the desired gene. This is also called gene-targeting.

However, homologous recombination is an extremely rare event. The most common strategy for selecting the rare cells with the homologous recombination is the use of genes conferring drug resistance and sensitivity. Typically, the positively selectable

gene, neo, which confers resistance to an antibiotic called G418 driven by a constitutive promoter (PGK) is inserted into the targeting construct so that neo is flanked by large stretches of DNA (2-7 kb) that exactly matches the genomic sequences surrounding the desired insertion point. A negative selection gene, such as thymidine kinase (TK) gene, is often placed outside of the homologous region of the targeting construct to select against ES cells that incorporate DNA by non-homologous recombination. Cells that express the TK gene are killed by the antibiotic Gancyclovir. Therefore, the stem cells that incorporate DNA by homologous recombination will survive in both G418 and Gancyclovir because only the homologous region of the targeting construct including neo is inserted into the genome while the flanking TK gene is outside of the homologous region and so it is not inserted. On the other hand, if the targeting construct is inserted randomly, both neo and TK will be inserted into the genome. That means the cells will resist the G418 but will be killed by Gancyclovir. After the selection, the neo inside the genome can be precisely excised using the Cre/loxP system.

The Cre/loxP system is a tool for precise excision of DNA. Cre is a 38kDa recombinase originally from the P1 bacteriophage, which mediates intramolecular and intermolecular site- specific recombination between loxP (locus of X-over P1) sites (Sauer B, 1993). The LoxP sequence is derived from the P1 bacteriophage, and comprises two 13bp inverted repeats separated by an 8bp asymmetric spacer region (Kuhn R & Torres RM, 2002). When cells that have loxP sites in their genome also express Cre, the enzyme Cre catalyzes the site- specific recombination in the 8bp asymmetric spacer region by binding loxP sites. That means the DNA sequence between two loxP sites is excised by the Cre and then ligated back together leaving one loxP site behind, when two loxP sites are in direct orientation. Otherwise, if two

loxP sites are in opposite orientation to each other, the DNA sequence between them is inverted.

The Cre/loxP system has the advantage of working in almost any type of cell. Therefore this precise removal of DNA by Cre/loxP system can be used to regulate the time specific / tissue specific expression of a transgene and the level of the expression. For example, to make tissue-specific conditional knockouts, the mouse strain having a target gene flanked by loxP sites in a direct orientation is mated with a mouse strain with Cre targeted to a specific tissue or cell type. Since the recombination only occurs in those cells expressing Cre recombinase, the target gene is only knocked out in those cells. The target gene can be a reporter transgene (Orban PC *et al*, 1992), or a functional endogenous gene (Kuhn R *et al*, 1995). The Cre/loxP system is also used to 'switch on' a transgene expression by cutting out an intervening stop sequence between the promoter and the coding region of the transgene (Lakso M *et al*, 1992). A more powerful approach to controlling gene expression has been developed: the expression of Cre is controlled by the tetracycline-regulated transcriptional system (Utomo ARH *et al*, 1999). In this case, the recombination only takes place when the tetracycline is administered to activate the Cre expression, which allows the transgene to be turned on or off at any time.

Therefore, compared to the traditional "gain-of-function" mutation created by microinjection of the gene into the one-celled zygote, gene-targeting based on homologous recombination in stem cells has been increasingly used to precisely modify the gene of interest. These manipulations include site-specific deletions, insertions, gene duplications, gene rearrangements and point mutations. In addition, transgenic technology also can be used for gene identification.

Many human genetic diseases can be modeled by introducing the same mutation into a mouse or other animal, especially in the major gene diseases. For example, a transgenic mouse model of osteogenesis imperfecta (OI) containing an internally deleted human *COL1A1* gene (encoded the alpha1 (I) chain of type I procollagen) developed a lethal phenotype with extensive fractures of ribs and long bones similar to the fractures seen in lethal variants of OI due to the biosynthesis of shortened pro alpha1 (I) chains and procollagen suicide (Khillan JK *et al*, 1991). Since then, this line of transgenic mice has been widely used to study the mechanism of OI *in vivo* such as the bone fragility at different sites and ages and mineral changes by means of ultrastructural, microanalytical, and spectroscopic studies (Cassella JP *et al*, 1994; Pereira RF *et al*, 1995), which has dramatically increased our understanding of the molecular mechanism of OI.

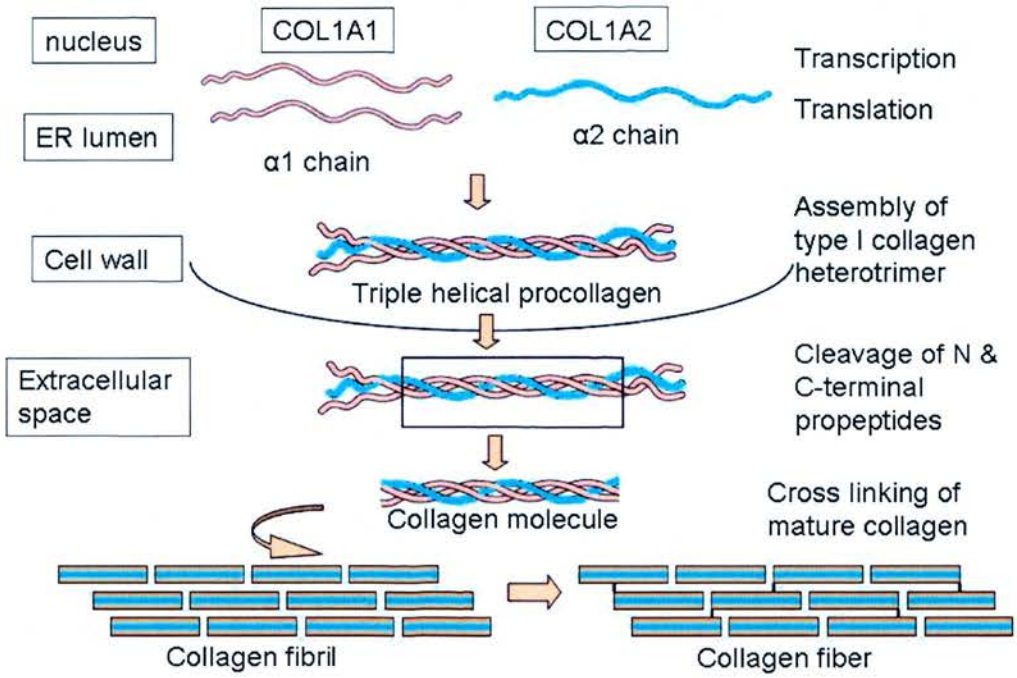
However, currently there is no such transgenic mouse model available for osteoporosis. Considering the polygenic background of osteoporosis, it is unlikely to establish a perfect mouse model for osteoporosis by introducing any of the human candidate gene variants with modest effect into the mouse genome. But genetically modified mice containing a potentially causal polymorphism or variant of a human candidate gene would be very helpful to study the molecular mechanism underlying associations and to test drugs that target to a specific molecular pathway. Although similar genetic manipulations can be performed in tissue culture, the interaction of transgenes with the complex components of an intact organism provides a much more reliable and complete picture of the candidate gene's function.

1.5 Type I collagen

Type I collagen is the major protein of bone accounting for 90% of the organic content bone (von der Mark 1999). The properties of type I collagen affect the mechanical properties of bone and provide a framework for mineralization. Type I collagen is synthesized in large amount by fibroblasts, osteoblasts and odontoblasts. It is a triple-helical, supercoiled molecules consisting of two identical α1(I) polypeptide chains and a structurally similar, but genetically different α2(I) polypeptide chain. Collagen α chains are characterized by a unique Gly-X-Y highly repetitive amino acid sequence (where X is frequently proline and Y is frequently hydroxyl-proline). The repetitive sequence allows three polypeptide α chains to fold into a right-handed triple helix of 1.5 nm diameter.

The biosynthesis of type I collagen is a complex process (Figure 1.5). The type I collagen molecule is assembled initially as a soluble precursor, type I procollagen. Individual α1(I) and α2(I) polypeptide chains (1464 and 1366 amino acids respectively) are synthesized on membrane-bound ribosomes. They then enter the lumen of the rough Endoplasmic Reticulum (ER) where selected proline and lysine residues are hydroxylated to form hydroxyproline and hydroxylysine respectively. Some of the hydroxylysine residues undergo further glycosylation. Two procollagen α1(I) polypeptide chains and one α2(I) polypeptide chain are assembled together to form triple helical molecules, type I procollagen, by hydrogen-bounds between the chains. This structure is stabilized by intra-chain disulphide bounds between the C-propeptides. Each procollagen molecule has three structurally distinct domains; an N-, C- propeptide domain, and a highly conserved central triple helix domain exhibiting a GLY-X-Y repetitive structure.

Figure 1.5 Posttranslational modifications and assembly of type I collagen



Type I collagen is a triple helical protein composed of three peptide chains derived from two distinct genes (*COL1A1* & *COL1A2*). Collagen precursor chains (α1 & α 2) are assembled in the ER lumen to form triple-helical procollagen molecules. After extensive posttranslational modification, procollagen is secreted from cell and converted to collagen in a peptide-cleaving reaction catalyzed by procollagen peptidase. Subsequently, the triple helical collagen molecules self assemble into collagen fibrils and these are crosslinked to produce mature collagen fibers.

These procollagens are then transported into the Golgi apparatus of the cell and secreted into the extracellular matrix in the form of soluble precursors. During the secretion, the N- and C- propeptides are removed by procollagen N and C proteinaes, which converts procollagen into collagen. Collagen molecules then self-assemble to form collagen fibrils. Fibrils are greatly strengthed by the formation of covalent intra- and inter-molecular cross-links between lysine residues of the collagen molecules. Collagen fibrils finally undergo intermolecular and interfibrillar crosslinks to produce mature collagen (collagen fibres), which are several micrometers in diameter.

Measurement of these bone-derived collagen cross-links in urine has proved to be a useful measure of bone resorption. The propeptides of collagen (P1NP & P1CP) which are cleaved off during collagen processing, are also used clinically as measurement of bone formation. Structural or quantitative defects in type I collagen cause the disease, osteogenesis imperfecta (OI), which is characterized by increased bone fragility and premature multiple fractures (Row DW, 1991).

In addition, changes in post-translational modification of collagen are also associated with bone diseases. Cartilage-associated protein (CRTAP) forms a complex with cyclophilin B and prolyl 3-hydroxylase 1, which is encoded by *LEPRE1,* and hydroxylates one residue in type I collagen, alpha1(I)Pro986. It has been reported that null mutations of CRTAP cause recessive osteogenesis imperfecta including type II & VII (Morello R *et al*, 2006). Moreover, null *LEPRE1* alleles cause a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta, where the *LEPRE1* mutations led to premature termination codons and minimal mRNA and protein. The affected collagen had minimal 3-hydroxylation of alpha1(I)Pro986 but excess lysyl hydroxylation and glycosylation along the collagen helix. The collagen

secretion was moderately delayed, but total collagen secretion was increased (Cabral WA *et al*, 2007).

The increased enzymatic collagen crosslinks may be associated with increased ultimate stress and stiffness (Viguet-Carrin S *et al*, 2006). Therefore, the collagen properties are an important determinant of bone strength and collagen abnormalities can cause different clinical disorders.

1.6 Type I Collagen Genes

The *COL1A1* and *COL1A2* genes code for the procollagen α1(I) and α2(I) polypeptide chains respectively. The *COL1A1* and *COL1A2* genes are located on chromosomes 17q21-q22 and 7q21.3-q22.1, respectively (Retief E *et al*, 1985). Human *COL1A1* is 18 kb in size and consists of 51 exons (Chu ML *et al*, 1984; Ramirez F *et al*, 1985), and *COL1A2* is 38 kb in size and consists of 52 exons (Ramirez F *et al*, 1985; de Wet W *et al*, 1987). Both genes are characterized by numerous short exons interrupted by longer introns. The central triple helix regions are encoded by 44 exons, which are all multiples of 9 bp in length corresponding to the repetitive Gly-X-Y triple amino acid sequences. Two exons at 5' end code for the N-terminal domain and four exons at the 3' end code for the C-terminal domain and 3' untranslated regions. Although the *COL1A1* and *COL1A2* genes are in different length (18kb and 40kb respectively), the transcribed mRNA molecules have a similar size. Sequence analysis of *COL1A1* and *COL1A2* showed significant homology among species, indicating the collagen genes are evolutionary conserved (Von der Mark, 1999).

Although these two genes are on different chromosomes, they are co-ordinately regulated so that the ratio of the steady-state levels of mRNA for *COL1A1* and *COL1A2* is 2:1. This ratio correlates with the fact that two α1(I) and one α2(I) poplypeptides are in the triple helix structure of collagen (Slack JK *et al*, 1993). It is still unclear how the transcription of collagen genes is co-ordinate in this way. The transcriptional regulation of the two genes is complex since collagen protein is synthesized in a tissue-specific manner and at widely differing levels during normal embryonic and postnatal development. Furthermore, the synthesis of type I collagen responds both positively and negatively to stimuli generated by tissue injury and repair, cytokines, hormones and is up-regulated in a variety of fibrotic disorders.

The expression of both *COL1A1* and *COL1A2* are regulated by the interactions of *trans*-acting enhancers and repressors with multiple *cis*-acting elements. A number of regulatory elements required for basal and inducible expression of *COL1A1* have been identified in the promoter and first intron of both genes with the positive and negative regulatory factors, as described below.

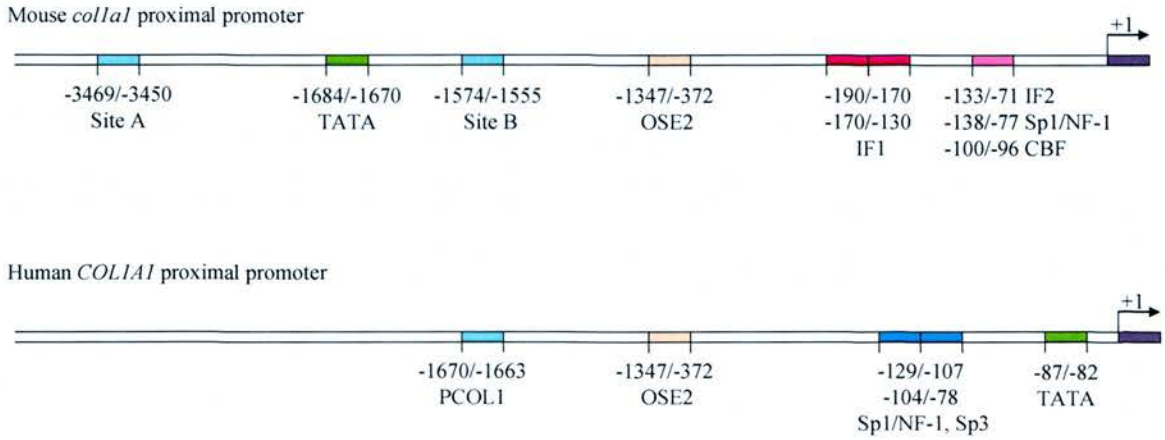
1.6.1 *COL1A1* Promoter

The regulatory regions within the promoter of the *COL1A1* gene contain some *cis*-acting elements which are responsible for the basal, tissue-specific and cytokine-regulated gene expression by interacting with different *trans*-acting factors. Several cytokine-responsive elements have been identified, such as a TGF-β response element, a TNF-α responsive element and an IFN-γ responsive element (Rossert J *et al*, 2000). Reporter assays have identified several *cis*-acting elements in the promoter of *COL1A1* gene which regulate the expression of reporter genes such as luciferase and lacZ genes *in vitro* or *in vivo*. For example, three elements located upstream of the proximal promoter (-800bp, -1700bp and >-2300bp) drive the expressions of reporter genes in skin fibroblasts, osteoblasts/ odontoblasts and tendon/ facial fibroblasts, respectively (Rossert J *et al*, 2000). These results are consistent with the observations that the sequences up to -2.3kb upstream of the transcriptional start site are important for *COL1A1* transcription in bone, whereas sequences between -3.5kb and -1.7kb are important for transcription in tendon (Bogdanovic Z *et al*, 1994). Subsequent work has defined an element containing a TATA motif at -1683 to -1670bp that enhances *Col1a1* transcription in mouse calvarial bone (Dodig M *et al*, 1996, Figure 1.6). Most of the transcription factors that bind to these elements are unknown, but some *in vitro*

experiments suggested that they induce tissue-specific expression by modifying the chromatin structure (Rossert J *et al*, 2000).

This proximal promoter of *COL1A1* is highly conserved between species. Several studies have shown that DNA sequences between -222bp and -80bp of proximal promoter of mouse *Col1a1* have very strong transcriptional activity which is reduced by the addition of more upstream sequences (Karsenty G & de Crombrugghe B, 1990). There are several *cis*-acting elements for transcription factors with positive and negative effects on transcription situated within this crucial region (Figure 1.6). The promoter TATA box is situated in this crucial region along with other four *cis*-acting elements, and binds the transcription factor TFIID to initiate the transcription. The other four *cis*-acting elements in the proximal promoter are involved in the binding of two inhibitory factors 1 and 2 (IF1 and IF2), and one enhancer (CCAAT bind factor, CBF, Saitta B *et al*, 2000). IF1 binds to two upstream elements located at -190 to -170bp and -170 to -130bp of mouse *Col1a1* promoter. This binding is only present in collagen-producing cells. IF2 binds to two GC-rich 12bp repeat sequences located in between -133 and -71bp and competes with the enhancer CBF. CBF inhibits IF2 binding by a stronger interaction with CCAAT motif (Karsenty G *et al*, 1991). Mutations in binding sites of IF1 or IF2 that caused inhibition of the binding of IF1 or IF2 increased the *Col1a1* promoter activity (Karsenty G, de Crombrugghe B, 1990 & 1991). Recently, CBF was reported to interact with CCAAT box located at -100 to -96bp of the human *COL1A1* promoter and the binding activity is 3-5 fold higher in SSc fibroblasts (Ghosh AK, 2002).

Figure 1.6 Schematic presentations of the *cis* elements in the mouse and human *COL1A1* proximal promoter



Cis-elements in the proximal promoter of mouse and human *COL1A1* gene are showing in colored boxes.

In the human *COL1A1* gene, reporter assays have shown that the short segment of the promoter between -220 and +40bp drives luciferase gene expression at high levels, (García-Giralt N *et al*, 2005; Slack JL *et al*, 1993). Inhibitory elements have also been mapped in the human *COL1A1* promoter between positions -2300 and -205, and the elements were subsequently mapped to a narrow region between -1284 and -254. However the precise elements within this region that mediate inhibition of transcription have not been fully identified yet.

Recently, a polyT region between -1670 and -1663 (PCOL1 site) was identified in the human *COL1A1* promoter homologous to the rat *Col1a1* promoter site B (-1574 to -1555) which binds nuclear matrix promoter 4 (nmp4) (García-Giralt N *et al.* 2002). A subsequent study has shown that Nmp4, the human homologue of the rat nmp4, is present in nuclear extracts of MG-63 human osteosarcoma cells and is able to bind to the PCOL1 site in the human *COL1A1* promoter (García-Giralt N *et al.* 2005). Nmp4 is a Cys_2His_2 zinc finger protein. The zinc finger motif is one of the most common DNA-binding motifs found in eukaryotic transcription factors, with the domain used

not only for protein-DNA interaction but also for protein-RNA and protein-protein interactions (Mackay JP & CrossleyM, 1998; Shastry BS, 1996). Despite the vast DNA recognition potential of Cys_2His_2 zinc finger proteins, the Nmp4 protein is one of the few transcription factors known to be associated with the minor groove of homopolymeric (dA·dT) sites (Torrungruang KT *et al.* 2002). The Nmp4 protein comprises four isoforms 11H, 13H, 21H and 28H, and binds the site A (-3469 to -3450) and B (-1574 to -1555) in rat *Col1a1* promoter (Thunyakitpisal P *et al*, 2001). This binding regulates gene expression by bending or looping DNA to alter interactions between other *trans*-acting proteins. The Nmp4 binding sites in the rat *Col1a1* promoter are within or proximal to the distinct *cis*-regulatory regions that mediate basal transcription in osteoblasts within bone and in cell culture (Torrungruang KT *et al.* 2002). Yet there is no clear evidence of *cis*-acting elements present near the Nmp4 site in human *COL1A1* promoter.

There is a potential Osterix binding site situated upstream of Nmp4 binding site in human *COL1A1* promoter (personal communication, Benoit de Crombrugghe), but there is no evidence yet about whether Osterix interacts with Nmp4. Osterix is one of the two essential transcription factors for osteoblast differentiation and regulates the expression of many important osteoblast genes including type I collagen genes (Nakashima K *et al*, 2002). Osterix is encoded by the *Sp7* gene and belongs to zinc-finger family (Sp) of proteins having 78% identity with Sp1 in the three zinc-finger DNA-binding domain, but there is little homology elsewhere.

Furthermore, Nmp4 inhibits the activation of osteoblast genes in response to parathyroid hormone (PTH), prostaglandin E_2 and BMP2. For PTH, the possible mechanism is that PTH enhances the DNA-binding activity of nmp4 to the rat *Col1a1* promoter (Alvarez M *et al.* 1998). This is in agreement with studies showing that PTH

inhibits osteoblast growth and collagen synthesis *in intro* (Bogdanovic Z *et al*, 2000; Kream BE *et al*, 1993). Recently, Nmp4 was found to be involved in the modulation of BMP2-induced osteoblast differentiation and was suggested to act as an inhibitor of BMP/Smad signalling. In osteoblastic MC3T3E1 cells, the over-expression of Nmp4 suppressed BMP2-induced expression of osteoblastic differentiation markers including type I collagen (Shen JZ *et al*, 2002). Therefore, the interaction between the Nmp4 and its binding site in human *COL1A1* promoter may alter promoter activity and contribute to gene transcription and expression.

Several other factors have been demonstrated to regulate *COL1A1* transcription. The Sp1 is ubiquitously expressed and regulates gene expression in all cell types. Sp1 protein belongs to the Sp family of transcription factors which have a highly conserved DNA-binding domain consisting of three zinc fingers close to the C-terminus and glutamine-rich amino acid domains at the N-terminal regions (Kadonaga JT *et al*, 1987). The zinc fingers preferably bind to the GC box, 5'-(G/T)GGGCGG(G/A)(G/A)(C/T)-3' and the glutamine rich regions are strong activation domains through which Sp1 forms protein-protein interactions with other *trans*-acting factors.

There is evidence that Sp1 and NF-1 interact with each other to regulate *COL1A1* gene transcription. NF-1 and Sp1 binding sites overlap in the mouse *Col1a1* proximal promoter (-138 to -77bp, Figure 1.6). In NIH/3T3 cells over-expression of NF-1 increased expression of a *Col1a1*-reporter gene and mutations in the NF-1 site led to reduced expression of the reporter gene, whereas over-expressing Sp1 inhibited both basal and NF-1 stimulated expression of the reporter gene (Nehls MC *et al*, 1991). In co-transfection studies using *Drosophila* Schneider L2 cells, both NF-1 and Sp1 stimulated gene expression and Sp1 was found to be a stronger transactivator than

NF-1. Co-expression of NF-1 with Sp1 resulted in inhibition of Sp1 activation by NF1 (Nehls MC *et al*, 1992). In the human *COL1A1* promoter, a short region in the proximal promoter (-129/-107 bp and -104/-78 bp) containing two tandem NF-1/Sp1 elements up-regulated transcriptional activity in SSc fibroblasts, and indeed, SSc fibroblasts contained 3.4-fold greater DNA binding activity recognizing these elements than normal cells (Hitraya EG *et al*, 1998).

There is also evidence that Sp1 and Sp3 interact in human *COL1A1* promoter. The Sp1 and Sp3 proteins bind GC- rich sequences and these interactions can be cooperative or Sp3 can decrease Sp1-dependent transactivation. Chen SJ *et al* (1998) reported the interactions of Sp1, Sp3 and CTF/NF-1 proteins present in human skin and lung fibroblasts nuclear extracts in the region between -129 to -107bp of human *COL1A1* promoter (Figure 1.6). It was concluded that although over-expressing Sp1 or Sp3 transactivated the *COL1A1* promoter, excess Sp3 can block the Sp1-stimulated COL1A1 promoter activity (Chen SJ *et al*, 1998).

Sp1 also directly interacts with TATA-binding protein factors (TAFs) and other nuclear cofactors which comprise the basal transcription factors. In the human *COL1A1* promoter, Sp1 has been demonstrated to play an important role in enchancing gene expression with a Sp1 site upstream of TATA box (-87 to -82bp, Figure 1.6), is obligatory for transciprional activation (Li L *et al*, 1995; Poppleton HM & Raghow R, 1997).

Runx2 (also called Cbfa1) is a runt-related osteoblast-specific transcription factor, which actives the expression of type I collagen genes in osteoblasts by binding to the OSE2 binding sites. There are two OSE2 sites (Figure 1.6) in the mouse *Col1a1* promoter (-1347 and -372bp) and one in the first exon (+12bp) of mouse *col1a2* gene, which are conserved in the mouse, rat and human (Kern B *et al*, 2001).

1.6.2 The First Intron of the *COL1A1* Gene

The *COL1A1* first intron is also important in regulating gene expression. The first evidence was obtained from the lethal phenotype of homozygous Mov 13 mice which contain the Moloney murine leukaemia retrovirus (MMLV) within the first intron of *Col1a1* (Schnieke A *et al*, 1983). Subsequent work showed that the retroviral insertion resulted in transcriptional inactivation of the gene in mouse embryo cell lines (Hartung S *et al*, 1986) but a normal amount of type I collagen present in dentin, which suggested that the first intron may have a tissue-specific effect on transcriptional regulation (Kratochwil K *et al*, 1989). In addition, it was shown that the first intron along with the integrated retrovirus was spliced out in osteoblasts and odontoblasts, which indicated that incorrect splicing could not account for the lethality of the mutation in homozygous mice and the different *cis*-acting elements affected gene regulation in tissue-specific manner.

Recently, mice with a large deletion in first intron of *Col1a1* gene and normal splicing have been generated to identify the function for the first intron in expression of α 1(I) collagen. Hormuzdi *et al* created a large deletion in first intron of *Col1a1* gene in mice using a gene-targeting approach and demonstrated an age-dependent reduction in expression of the mutated allele in lung and skeletal muscle (Hormuzdi SG *et al*, 1998). Later this group found that the expression of mutated allele increased substantially to a level near that of the wide-type allele when responding to physiological signals, which implies that reduced expression of the mutated allele is conditional and can be reversed by the function of elements elsewhere in the gene (Hormuzdi SG *et al*, 1999). More recently, Rahkonen O *et al* (2004) reported that mice homozygous for this deletion are predisposed to dissection and rupture of the aorta during their adult life, because the integrity of the aortic wall depends on an

adequate content of type I collagen and the first intron functioned in continued synthesis of collagen. Another study has shown that the mechanism of this dissection is decreased type I collagen synthesis and content in the aortic wall (Marjamaa J *et al*, 2006).

Assays of collagen transcription i*n vitro* also supported the above results in showing that deletion of the first intron results in a 2-3 fold reduction in *COL1A1* transcription (Bonstein P & McKay J, 1988). Several elements within the first intron have been implicated in transcriptional regulation. For example, there is evidence that the first intron of *COL1A1* contains negative regulatory elements which exert an inhibitory effect on transcription in reporter assays but only in the presence of upstream promoter sequences between -625 and -293 (Bornstein P *et al.* 1996). The region between +380 and +1440bp in first intron of *COL1A1* is involved in regulatory transcription of the collagen gene in scleroderma skin fibroblasts (Hitraya EG & Jimenez SA, 1996). Taken together, these findings suggested that the first intron of *COL1A1* gene can exert bidirectional effects on transcription which may be due to the interactions between different transcription factors bound to the promoter and first intron.

A series of transcription factors have been identified that regulates the *COL1A1* expression by interacting with the first intron, promoter and upstream regulatory elements of the gene. The protein AP1 belongs to a family of transcription factors comprised of c-Fos and c-Jun proteins. These modulate the target gene expression as homo- or heterodimers. An AP1 binding site was found located in a 29bp sequence between +292 to +670 bp of first intron of the human *COL1A1* gene and was found to be involved in transcriptional regulation of *COL1A1* expression (Liska DJ *et al*, 1990). Slack JL *et al* (1992) reported that transformation of rodent fibroblasts by oncogenic

Ras resulted in transcriptional inhibition of *Col1a1* gene and later they found that Ras-mediated inhibition of *COL1A1* transcription in fibroblasts was achieved through an AP1 binding site in first intron (Slack JL *et al*, 1995). As c-jun and c-fos are also activated by phorbol 12-myristate 13-acetate (PMA), the inhibition of *COL1A1* transcription by PMA may be through an AP1 response element. Estradiol has also been shown to suppress the Type I collagen synthesis in murine mesangial cells via stimulation of AP1 activity (Silbiger S *et al*, 1999).

There are several Sp1 binding sites in the first intron. Two tandem Sp1 binding sites are located in a 274bp region of first intron and a DNA-protein complex formed with this element inhibits the *COL1A1* promoter activity (reviewed by Ghosh AK, 2002). Transmission electron microscopic analysis of DNA-protein complexes showed that protein-protein interactions between Sp1 and other transcription factors bring the promoter and intron into physical contact with each other, with the resulting looping out of the intervening DNA. It is possible that this interaction creates an optimum environment for the binding of the transcriptional machinery (Vergeer WP *et al*, 2000). Further work will be needed to identify the transcription factors which interact with Sp1 and co-regulate *COL1A1* transcription and expression.

1.7 The Aims of the Study

The general aim of this study was to utilize some of the approaches described earlier to examine the roles of three polymorphisms in the 5' flank of the *COL1A1* gene in the pathogenesis of osteoporosis.

The specific aims of this study were:

1. To investigate the role of three polymorphisms and haplotypes defined by three polymorphisms in osteoporosis by means of association analysis in a cohort of Scottish premanopausal women, Scottish female patients with osteoporotic fractures and age-matched female controls..
2. To ascertain whether the promoter and intron1 polymorphisms have functional roles by studying DNA-protein binding and gene transcription.
3. To investigate the association between the three polymorphisms and haplotypes and the mechanical properties of bone *ex vivo*.
4. To further define the functional roles of three polymorphisms and haplotypes in bone *in vivo* by generating an animal model of osteoporosis mediated by the polymorphisms.

CHAPTER TWO

Materials and Methods

2.1 Materials

2.1.1 Suppliers

Sigma-Aldrich Limited, Poole, Dorset, UK

Sigma-Genosys, Ltd, London Road, Pampisford, Cambridge, UK

Invitrogen Ltd, Paisley, UK

Promega, Ltd, Southampton, UK

Qiagen Ltd, West Sussex, UK

GRI, Essex, UK

New England Biolab (UK) Ltd, Hitchin, Hertfordshire, UK

Stratagene, Amsterdam, Netherlands

Cambrex Bio Science Rockland Inc, Rockland, USA

Roche Diagnostics Ltd, Lewes, East Sussex, UK

Hybaid, Teddington, UK

ABgene, Surrey, UK

Anachem Ltd, Bedfordshire, UK

Sarstedt, Leicester, UK

Corning Ltd, Buckinghamshire, UK

Cheshire Scientific Ltd, South Wirral, UK

Fisher Scientific Ltd, Leicestershire, UK

Sciquip Ltd, Shropshire, UK

Applied Biosystems, CA, USA

USB Corporation, Cleveland Ohio, USA

Amersham Pharmacia Biotech UK Ltd, Buckinghamshire, UK

2.1.2 General Plasticware

Micro tubes 0.5ml, 1.5ml and 2ml	Sarstedt
Tissue culture plasticware	Corning
0.2ml Flat and domed cap tube	Abgene
Thermo-Fast 96, shirted plate	Abgene
Flat cap strip	Abgene
Culture Loops	Cheshire
Microbiological spreader	Cheshire

2.1.3 Appatatus

Horizontal electroporesis tank	Hybaid
MJ Research Thermocycler	GRI
Balance Fisherbrand PS-200	Fisher
Microcentrifuge eppendorf 5415D	Fisher
IKA Minishaker MS 2	Fisher
Syngene Bio imaging system	Fisher
Sigma Laboratory centrifuge 4-15, 4K15	Sciquip
Micro Pipette (2µl, 10 µl, 100 µl, 200 µl, 1000 µl)	Anachem

MEGABACE™ automated DNA sequencer	Amersham

2.1.4 Chemicals

Isopropyl alcohol	Sigma-Aldrich
Ethanol	Sigma-Aldrich
Formamide	Sigma-Aldrich

2.1.5 Media and Media Components

Alpha MEM	Invitrogen
PBS	Invitrogen
Glutamine	Invitrogen
Penicillin	Invitrogen
Trypsin/EDTA	Invitrogen
Yeast extract	Invitrogen
Carbenicillin	Invitrogen

2.1.6 Molecular Biology Products

Hinc II DNA ladder	Abgene
1 kb DNA ladder (for quantification)	New England Biolabs
1 kb DNA ladder	Promega
Bovine serum albumin	New England Biolabs
dNTPs	Promega
Qiagen plasmid mini kit	Qiagen
Qiagen Endofree plasmid Maxi kit	Qiagen
Qiagen PCR purification kit	Qiagen
QIAquick gel extraction kit	Qiagen
QIAEXII gel extraction kit	Qiagen
PRISM Bigdye™ Terminator cycle sequencing kit	Applied Biosystems
QuickChange™ Site-Directed Mutagenesis Kit	Stratagene
Restriction endomucleases and buffer	New England Biolabs
Taq PCR core kit	Qiagen
Seakem®GTG® Agarose	Cambrex
Ethidium Bromide	Sigma-Aldrich
Orange G loading dye	Sigma-Aldrich
PCR primers and oligos	Sigma-Genosys
Platinum Pfx DNA polymerase kit	Invitrogen
pBluescript plasmid DNA	Stratagene
Shrimp Alkaline Phosphatase	USB Corporation

T4 DNA Ligase and buffer	Roche
XL-1 blue supercompetent cells	Stratagene
ExoSAP IT™	USB Corporation
DYEnamic™ ET dye terminator sequencing kit	Amersham
pGL -3 luciferase vectors (control & basic)	Promega
pRL –TK vector	Promega
T4 DNA Polymerase, buffer and EDTA	NEB
Subcloning Efficiency™ DH5 α competent cells	Invitrogen
Dual Luciferase Reagent	Promega
Picogreen assay	Invitrogen
DNase I	Invitrogen
dNTP (10mM each)	Invitrogen
0.1 M DTT	Invitrogen
Nuclear Extraction Kit	Acti motif
Bicinchoninic Acid Protein Assay	Pierce
LightShift Chemiluminescent EMSA Kit	Pierce
6% (w/v) native polyacrylamide gel	Invitrogen
Anti-sp1	Promega
hrSP1, USF, anti-USF	Santa cruze
Platinum® Pfx DNA Polymerase	Invitrogen
pBlueScript SK (-) vector	Stratagene
QuickChange™ Site- Direct Mutagenesis Kit	Stratagene

Agarose powder	Bio-Rad
Effectene Transfection Reagent	Qiagen

2.1.7 Computer Softwares

All the sequence files were read by program "Chromas Pro 1.33" (Technelysium Pty Ltd). The haplotypes were constructed by program "PHASE" (Stephens M *et al*, 2001; Stephens M & Donnelly P, 2003). The probability threshold at which haplotype phases are called as "correct" is 90% for all the association analyses in this thesis.

The statistical analysis was performed with program "Minitab version 12" (Minitab Inc. USA). The quantification of the gel shift results were performed with program "Genetools" (Syngene Bio Imaging, UK).

2.2 Stock Solutions and Media

2.2.1 General Buffers

1 x TBE:

89 mM	Tris base
89 mM	Boric acid
2 mM	Disodium EDTA

2.2.2 Cloning Solutions and Medium

LB medium (1L):

10g	Bacto-tryptone
5g	Yeast extract
10g	NaCl

Adjusted to pH 7.0 with 5M NaOH. Volume to 1L with deionized water. Autoclaved and stored at room temperature.

LB agar plates:

To 1L of LB medium (prior to autoclaving) 15g of bacto-agar were added. Mixed and then autoclaved. When antibiotic selection was required carbenicillin was added to a final concentration of 50ug/ml. Approximately 30ml LB agar was poured into each 9cm Petri dish. Plates were allowed to set at room temperature and stored at 4°C.

NZY$^+$ broth:

1.0g	NZ amine (casein hydrolysate)
0.5g	Yeast extract
0.5g	Nacl
95.5ml	Deionized water

Deionized water added to a final volume of 100ml. Autoclaved and stored at 4°C. The following reagents were filter sterilized and added immediately prior to use:

2.5ml	$MgCl_2$ (1M)
2.5ml	$MgSO_4$ (1M)
4.0ML	20% (w/v) glucose

2.2.3 Other solutions

10X Annealing buffer:

1ml	Tris-Cl (1M, pH8)
2ml	NaCl (5M)
200µl	EDTA (0.5M)
6.8ml	ddH_2O

Acetone/DMF (ES cell DNA extraction buffer):

140ml	70% acetone
10ml	5% Dimethyl Formamide

2.3 Study Populations

2.3.1 APOSS Population

The subjects were taken from the Aberdeen Prospective Osteoporosis Screening Study (APOSS) population. Participants in the study were randomly selected from the community health index, which is a database of all patients registered with general practitioners in Scotland, focussing on those living within a 25 mile radius of Aberdeen. This was a population-based screening study for osteoporotic fracture risk, involving 5119 women aged 45-54 years who were initially assessed over a 4 year period between 1990 and 1994. At this assessment, participants were weighed on a set of balanced scales calibrated to 0.05 kg (SECA, Hamburg, Germany) and height was measured with a stadiometer (Holtain, Ltd., Crymych, UK). Participants also completed a risk factor questionnaire including questions on menopausal status, physical activity, smoking and use of hormone replacement therapy (HRT). Details of this population have been previously described (MacDonald HM *et al*, 2003). Women were classified as premenopausal if they were menstruating regularly, as perimenopausal if their menstruation was irregular and/or up to 6 months had elapsed since their last period; and as postmenopausal if their menstruation had ceased for 6 months or more. On the basis of this information we categorized subjects into five groups as follows: 1= premenopausal, no HRT; 2 = perimenopausal, no HRT; 3 = postmenopausal, no HRT; 4 = postmenopausal, previous HRT user; 5 = postmenopausal, current HRT user. Following this assessment, the BMD results were

disclosed to participants and their general practitioners and subjects with BMD values in the lowest quartile of the population were advised to consider taking HRT therapy (Torgerson DJ *et al*, 1997).It has been previously found that the uptake of HRT in this group (approximately 30%) is 6% greater than in women who had not undergone BMD screening (Torgerson DJ *et al*, 1997).

All participants were invited to undergo further assessment between 1997 and 1999 and a total of 3883 women of the original 5119 attended at this time (75.8%). The participants in the present study comprised 3270 women who consented to give blood samples for DNA extraction and were successfully genotyped for at least one of the polymorphisms studied.

The APOSS data collection and DNA extraction were finished within the Department of Medicine & Therapeutics, the University of Aberdeen prior to this project.

2.3.2 Fracture and Control Populations

This study included analysis of two case and two control groups, all exclusively from North East Scotland. The hip fracture group comprised 303 consecutive female patients who were admitted over a 2-year period to Aberdeen Royal Infirmary (ARI) with osteoporotic hip fracture resulting from low-energy trauma (typically a fall from a standing height or less). Patients were excluded who had sustained a hip fracture as a result of significant trauma (e.g. motor vehicle accidents, or falls from a ladder), and who were suspected to have sustained a pathological fracture (e.g. Paget's disease or malignancy). The control group for hip fracture was recruited from patients admitted the same hospital who had fallen but not sustained a hip fracture. These patients had mostly been admitted to the geriatric acute assessment ward at ARI.

The vertebral fracture group comprised 151 female vertebral fracture patients diagnosed clinically and radiographically in the north east Scotland. The control group comprised 143 age-matched healthy females selected from the same area who had responded to an advertisement for volunteers to take part in a study of treatment for osteoporosis. Underlying secondary causes of osteoporosis, such as corticosteroid therapy and endocrine disorders, were excluded in both fracture populations by medical history, physical examination, and laboratory investigations.

All the subjects involved in this study gave informed consent to being included and the study was also approved by the local ethics committee.

The data collection was finished within the Department of Medicine & Therapeutics, the University of Aberdeen prior to the start of this project.

2.3.3 Subjects for the Biomechanical Tests

Biomechanical studies were performed on trabecular bone samples from femoral heads obtained from 23 male and 71 female patients who were admitted to Aberdeen Royal Infirmary and underwent hip replacement surgery as a result of osteoporotic hip fracture. The age at fracture ranges from 58 to 101 years old with an average of 81.7 years. The study was approved by the local hospital ethical committee, and all patients gave informed consent to tissue samples being taken.

The data collection was finished by Dr. Tracy Stewart, in the University of Aberdeen prior to the start of this project.

2.3.4 Bone Density Measurements

Bone mineral density (BMD) was measured at the femoral neck (FN-BMD) and lumbar spine (L2-4) (LS-BMD) by dual energy x-ray absorptiometry using Norland XR26 and XR36 densitometers. The majority of women were scanned using an XR26 machine, but 357 women (11.5%) were scanned using a XR36 machine. There was a small difference (1.258%) in mean BMD value between the machines and a correction factor was used to convert the XR36 values to XR26-equivalent values. There were no significant differences in the proportion of women scanned on each machine across different genotypes.

The BMD measurements were performed in the Aberdeen Royal Cornhill hospital prior to the start of this project.

2.4 Genotyping Methods

2.4.1 DNA Extration

DNA had been extracted from blood prior to the start of the project for all the patients using the following protocol: whole blood was collected from subjects into sodium EDTA Vacutainer® tubes (Becton Dickinson, Dublin, Ireland) and DNA extraction was performed using a commercially prepared kit according the manufacturers instructions (Nucleon II DNA extraction kit, Scotlab, Cambridge, UK). The purity of the extracted DNA was determined from OD_{260}/OD_{280} ratio and the concentration was determined by reading the OD_{260} using a UV spectrophotometer (GeneQuant RNA/DNA Calculator, Pharmacia Biotech, Cambridge, UK). DNA was then stored at -20°C. The DNA extraction was performed within the Department of Medicine & Therapeutics, the University of Aberdeen prior to the start of this project.

DNA was extracted from bone in the biomechanical tests by Dr. Tracy Stewart using the method described by Steward TL & Mann V (2003). Fresh or freshly frozen bone that contains marrow was disrupted into about 1 cm^3 per piece using bone cutters and transferred to a clean 5-mL bijoux container. 1mL of DNA extraction buffer was added into each bijoux and mixed well until a slurry was obtained. Then 500µl aliquots of the slurry were transferred into an Eppendorf tube for each sample. One volume of Tri-saturated phenol was added into each tube followed by one volume of chloroform. Each sample was mixed well and centrifuged at 10,000g for 20 minutes

to separate the phases. The upper layer was then transferred to a fresh tube and one volume of ice-cold isopropanol and 0.1 volumes of 3M sodium acetate were added into the tube. Each sample was mixed well and standed for 15 minutes on ice before proceeding to the centrifuge step (10,000g for 20 minutes). The supernatant was discarded carefully. The DNA pellet was washed with 1.75 mL of ice-cold ethanol and centrifuged at 10,000g for 5 minutes. The wash was repeated once. Finally the DNA pellet was dissolved in 10-50µl of water and quantitated by the same method described above. DNA was then stored at -20°C.

2.4.2 Polymerase Chain Reaction (PCR)

Genotyping for -1997G/T and -1663In/delT was carried out by DNA sequencing of PCR-amplified fragments of genomic DNA whereas genotyping for the Sp1 polymorphism was carried out predominantly by TaqMan analysis as previously described (Uitterlinden AG *et al*, 1998). All PCR reactions were performed in 96 well plates in a 25-µl volume containing the following reagents:

2.5µl	10× PCR buffer (Qiagen, UK)
2.0µl	dNTP (10mM each)
5.0µl	Q-solution (Qiagen, UK)
2.5µl	primer mix (10µM each of forward and reverse primer)
0.625U	Taq DNA polymerase (Qiagen, UK)
50ng	genomic DNA
Xul	ddH_2O up to 25ul

The PCR reactions were carried out on a Gene Engine thermocycler (MJ Research Inc, USA). For the -1997G/T and -1663In/delT polymorphisms, we generated a 513bp fragment using the following primers: 5'-TCACTAACCCTCATACTACCAAGC-3' and 5'-AAGATTCCATTGCCTCCCCC-3'. The thermal cycling protocol for both fragments consisted of an initial incubation for 5 min at 94°C, followed by 31 cycles of 60 sec at 94°C, 30 sec at 61°C and 60 sec at 72°C followed by an extension step at 72°C for 5 min in the last cycle.

2.4.2 Sequencing Reactions

The PCR products were treated with ExoSAP IT™ according to the manufacturers instructions (USB Corporation, USA) and sequenced using the DYEnamic™ ET dye terminator cycle sequencing kit according to manufacturers instructions (Amersham Pharmacia Biotech UK Ltd, UK). The sequencing reactions were analyzed using a MEGABACE™ automated DNA sequencer (Amersham Pharmacia Biotech UK Ltd, UK). The 513bp fragment containing both -1997G/T and -1663In/delT polymorphisms was sequenced using the forward or reverse PCR primer or an internal reverse primer 5'-CCTTTAATTATAGCCCCTGCA-3'. A random sample of 5% of subjects were genotyped twice for each polymorphism to minimize genotyping errors, with the exception of the -1663In/delT polymorphism where all subjects who were heterozygous or homozygous for the -1663InT allele were genotyped twice.

2.5 Luciferase Reporter Assay

2.5.1 An Overview of the Technique

Luciferase reporter assay systems are currently one of the best non-toxic, rapid and sensitive methods to measure gene expression. The assay is based on the detection of luciferase activity driven by a promoter or enhancer cloned upstream of the luciferase gene. Luciferase activity correlates with transcription regulated by DNA regulatory elements cloned upstream of the reporter gene as well as responses to extracellular and intracellular signals. In the quantification of gene expression using *Firefly* luciferase (pGL3-control/ basic vector), a second *Renilla* luciferase vector (pRL-TK vector) is commonly used as internal control in combination with the primary *Firefly* luciferase vector to co-transfect mammalian cells. The Dual- Luciferase® Reporter (DLR) Assay System integrates the assays of both firefly and Renilla luciferase from the same sample by the sequential measurement of their cognate substrate luciferins, respectively.

2.5.2 Construction of *COL1A1*-Luciferase Reporter Vectors

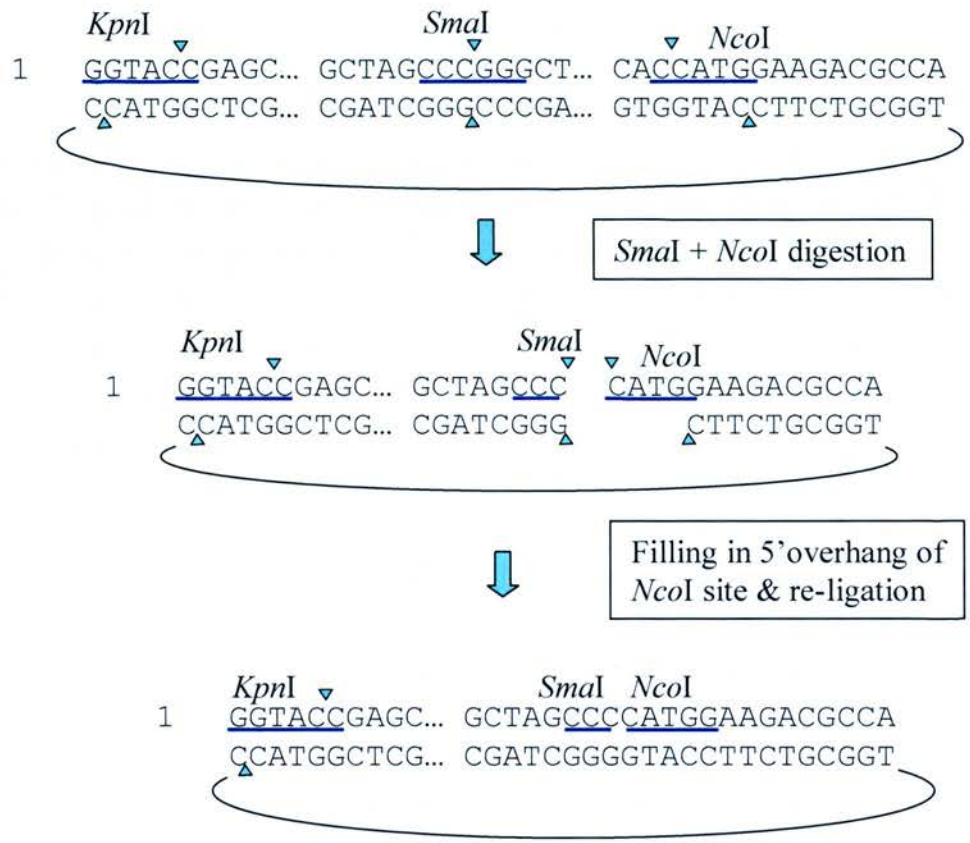
2.5.2.1 Generating 5' *COL1A1* DNA Fragment by Long-PCR

To create the reporter vector, a PCR was carried out to amplify 4kb of 5' human *COL1A1* sequence containing promoter, exon1, intron1 and part of exon2 with the following primers, *Kpn*I forward: TCACACGCCAGGTACCCAAA and *Kpn*I reverse: TCACGTCATCGCACAACACCT. The PCR was performed in a 0.2 ml PCR tube in a 25 µl volume containing 2.5 µl 10× PCR buffer (Qiagen Ltd, UK), 2.0 µl of dNTP (10mM each), 5.0 µl of Q-solution, 2.5 µl of primer mix (10µM each of forward and reverse primer), 0.5 µl Taq DNA polymerase (Qiagen Ltd, UK) and approximately 80ng human genomic DNA. The thermocycle protocol for this PCR consisted of an initial incubation for 5 min at 94°C, followed by 9 cycles of 15 sec at 94°C, 1 min at 56°C and 4 min at 68°C followed by 20 cycles of 94°C for 1 min, 56°C for 1 min and 68°C for 4 min with 10 sec increase per cycle. The PCR product was then digested with *Kpn*I restriction enzyme and purified with Qiagen PCR purification Kit. 1 µl of the purified *Kpn*I digested product was then loaded on 1% agarose gel along with 1kb DNA ladder (NEB, UK) for an approximate quantification.

2.5.2.2 Modification of the pGL3-basic Vector

To keep the translational reading frame of the luciferase gene intact, the pGL-3 basic vector (Promega, UK) was modified prior to the cloning of the 5' COL1A1 sequence. As shown in Figure 2.1, the pGL-3 basic vector was first double-digested by *Sma*I and *Nco*I to release a 57bp fragment from the Multiple Cloning Sites (MCS) region. The double-digestion was set up in a 0.2 ml PCR tube containing 3ug of pGL-3 basic plasmid, 30 U of each enzyme (NEB, UK), 3 µl of 10X NEbuffer4 , 1.5 µl of 20X BSA (NEB, UK) and deionized water up to 30 µl. The digest mix was incubated at 37°C for 3 hours and 65°C for 20 minutes. The digested products were then run on a 1% agarose gel and the 4.8kb band was extracted by Qiagen Gel Extraction Kit followed by quantification on the gel.

Figure 2.1 Diagram of modification of pGL3-basic vector.



As shown in Figure 2.1, to fill in the 5' overhang of the *Nco*I digest, the purified DNA (1ug) was incubated at 12°C with 3 units of T4 DNA polymerase (NEB, UK), 3 µl of 10X NEbuffer2, 100µM dNTPs, 1.5 µl 20X BSA and deionized water up to 50 µl. Stop reaction by adding 1 µl of 500mM EDTA into the 50 µl mix and heating to 75°C for 20 minutes. Then 2 µl of the DNA from the reaction was used in a self-liagtion with 1 unit T4 DNA ligase (Roche, UK), 1 µl of ligation buffer and 6 µl of deionized water. The ligation mix was incubated at 16°C for 16 hours. 1 µl of the ligation mix was added to 50 µl of Subcloning Efficiency™ DH5 alpha competent cells (Invitrogen, UK). The cells were incubated on ice for 30 minutes and followed by 20 seconds incubation at 37°C. 850 µl of pre- warm LB broth was added into the cells

and incubated at 37°C for 1 hour with constant shaking at 225rpm. 200 µl of the mixture was spread on a pre- warm LB agar plate and incubated overnight at 37°C. Two clones was picked out from the agar plate with yellow tips on the next day and incubated in 5ml LB broth with 80ug/ml carbenicillin at 37°C overnight with constant shaking at 225rpm. The plasmid DNA was purified from bacteria by Qiagen MiniPrep Kit according to the manual and quantified by BIO-TEK plate reader (Synergy HT, UK).

2.5.2.3 Generating *COL1A1*-luciferase Reporter Vectors

To take up the 4kb *Kpn*I digested COL1A1 5' fragment, 3ug of the modified pGL3-basic vector was digested with 30 units of *Kpn*I restriction enzyme (NEB, UK) with 3 µl 10X NEbuffer 1, 1.5 µl 20X BSA and deionized water up to 30 µl at 37°C for 2 hours. The digestion mix was purified with Qiagen PCR purification kit and then dephosphorelated with Shrimp Alkaline Phosphatase (SAP, USB corporation, UK). The SAP treatment was set up as follows:

3 µl	Shrimp Alkaline Phosphatase (1 U/µl)
4 µl	10X Shrimp Alkaline Phosphatase reaction buffer
30 µll	purified digested DNA (<3 ug)
3 µl	ddH_2O

The reaction was incubated at 37°C for 30 minutes and stopped at 65°C for 15 minutes. The DNA from the SAP treatment was used directly as vector in the ligation reaction. The following formula was used to calculate the amount of insert and vector in the ligation reaction:

[{Amount of vector (ng) * length of insert (kb)}/length of vector (kb)]* A

The molecular ratio of vector DNA to insert ('A' in the formula) is 1:3 for sticky ends when insert DNA and vector were approximately similar in length, otherwise ratio 1:1 or 1:2 was used.

In this study, the molar ratio of insert to vector was 3. Therefore, the reaction mix contained 100ng of 4.8 kb vector, 250ng of 4kb insert, 1µl of T4 DNA ligase (1U/ µl, Roche, UK), 1µl of 10X ligation buffer and water up to 10µl.

The ligation mix was incubated at 16°C for 16 hours. 2 μl of the ligation mix was added to 50 μl of Subcloning Efficiency™ DH5αcompetent cells (Invitrogen, UK). The transformation and the incubation of the agar plate were carried out as described above. Four clones were picked out to incubate in 5ml LB broth at 37°C overnight with constant shaking at 225rpm. After the same procedures for miniprep and quantification, the plasmid DNA was selected by a *Bsrg*I digest. The expected bands were 3546bp and 5248bp. The plasmid with correct digestion products was sequenced with primers listed in Appendix 1.2. The sequence reactions with these primers covered all the 4kb insertion and 100bp of luciferase gene.

2.5.2.4 Creating Eight Haplotypes by Site-Directed Mutagenesis

The other seven possible haplotypes defined by the three 5' *COL1A1* polymorphisms were introduced into reporter vectors by Site-Directed Mutagenesis reaction (SDM). The basic procedure utilizes a supercoiled double-stranded DNA vector with an insert of interest and two synthetic oligonucleotide primers containing the desired mutation (Figure 2.2). The primers, each complementary to opposite strands of the vector, are extended during temperature cycling by DNA polymerase to generate mutated plasmid containing staggered nicks. Following temperature cycling, the product is treated with *Dpn*I digest which is used to digest methylated and hemimethylated parental DNA template for selection of mutation-containing synthesized DNA. After the *Dpn*I selection, the nicked vector DNA containing the desired mutations is then transformed into competent cells.

QuikChange® II XL Site-Directed Mutagenesis kit (Strategen, UK) was used to create other seven reporter vectors containing all possible haplotypes. All the primers were designed complementary between 35-45bp with PAGE purification (listed in Table 2.1). The mutagenesis reaction was carried out in 0.2 µl PCR tubes. The reaction was prepared as indicated below:

5µl	10X reaction buffer
10-50ng	dsDNA template (plasmid DNA)
1.25µl	primer F (100ng/µl)
1.25µl	primer R (100ng/µl)

1µl	dNTP mix
3µl	QuikSolution Reagent

ddH_2O was added in to a final volume of 50µl. 1µl of *PfuUltra* HF DNA polymerase (2.5U/µl) was added last into the reaction. The thermal cycling protocol consisted of an initial incubation for 1 min at 95°C, followed by 18 cycles of 50 sec at 95°C, 50 sec at 60°C, and 9 min of 68°C followed by an extension step at 68°C for 7 min in the last cycle.

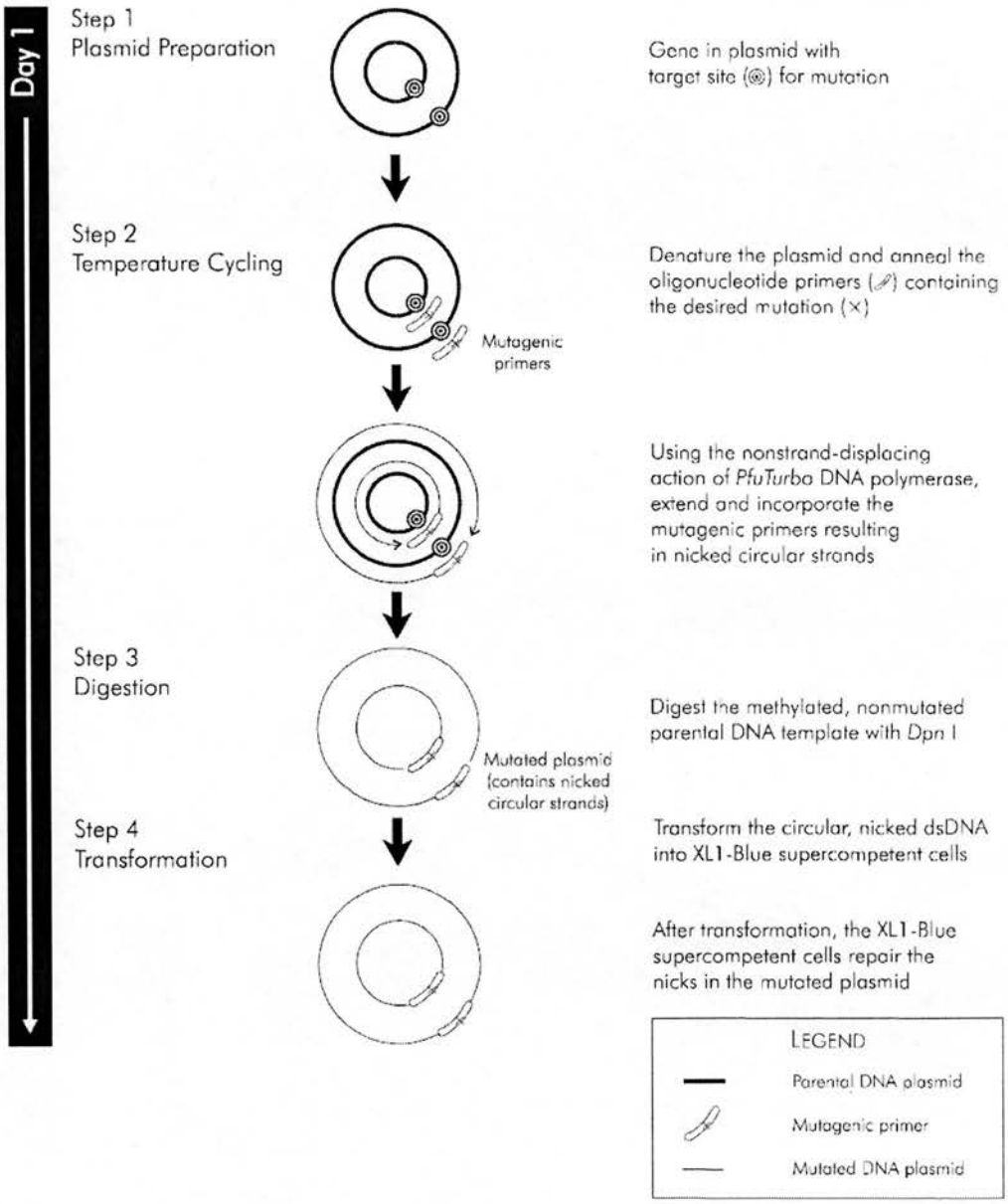
After 1 hour incubation with 1 µl *Dpn*I restriction enzyme (10U/ µl), 2 µl of mix were transformed into XL10-Gold ultracompetent cells according to the manufacturer instructions (Stratagene, UK). All the nicked vectors containing desired mutations were then purified from bacterial culture using Mini Prep Kit (Qiagen, UK) and sequenced to confirm the base changes.

Table 2.1 SDM primers used to create haplotypes in reporter vectors

Prime name	primer sequence (5'→ 3')
SP1G_SDMF	GCCCAGGGAATG***T***GGGCGGGATGAG
SP1G_SDMR	CTCATCCCGCCC***A***CATTCCCTGGGC
-1663delT -SDMF	CCCTCCCCTTCTGTTCC***TTTTTTT***CCCCTTTTGCCTTCGTTGC
-1663delT -SDMR	GCAACGAAGGCAAAGGGG***AAAAAAA***GGAACAGAAGGGGAGGG
-1997G_SDMF	CCCCTGTCGCCTATTAGGGAGGGG***G***TCTCTTGGAACTGACCCAC
-1997G_SDMR	GTGGGTCAGTTCCAAGAGA***C***CCCCTCCCTAATAGGCGACAGGG

Polymorphisms are in ***Bold Italics.*** F: forward primer; R: reverse primer

Figure 2.2 Overview of the QuikChange® site-direct mutagenesis method (from www.strategene.com)



2.5.3 Quantification of Reporter Vectors by Fluorometry

DNA concentration was measured using Quant-iT™ PicoGreen® dsDNA Assay Kit on a BIO-TEK *Synergy* HT multiwell plate reader. First, a standard curve was constructed using double-stranded λ DNA at known concentrations. DNA samples at known concentrations were analyzed in duplicate and their average concentrations calculated from the standard curve regression equation. Hoechst dye (100 µl at 20ug/ml) was added to each well of a black 96 well microtitre plate with clear bottom along 100 µl of either DNA at known concentrations or DNA standards of known concentrations. The plate was read after incubation at room temperature for 5 minutes. The fluorometer measured the relative each sample at 360 nm excitation and 460 nm emission wavelengths to create a concentration value. Each well was read twice and a mean value calculated.

2.5.4 Transfection of the Vectors into HOS TE85 Cells

All transiently transfection experiments were carried out using the human osteosarcoma cell lines, HOS-TE85. HOS TE85 and another human osteosarcoma cell line MG-63 are considered suitable models of osteoblastic phenotype, including the synthesis of type I collagen (Clover J & Gowen M, 1994).

Cells were cultured in 75cm^2 tissue culture flasks in alpha MEM medium (Invitrogen) supplemented with 10% fetal calf serum (FCS, Invitrogen), 50 IU/ml penicillin (Invitrogen) and 2 mM glutamine (Invitrogen). Medium was renewed every 3 days. Cells were incubated at 37°C in a humidified atmosphere of 5% CO_2.

Twenty- four hours prior to the transfection, cells were cultured in 24-well plates with the density of 4 x 10^4. On the day of transfection, cells with less than 50% confluence were transfected with pGL -3 control vector, pGL -3 basic vector and eight *COL1A1*-luciferase reporter vectors with eight different haplotypes, respectively using Magnet Assisted Transfection (IBA, Germany) according to the instructions. With this new technique nucleic acids are first associated with magnetic particles. After addition of transfection complexes to the cells, the culture plate is simply placed on top of a magnetic plate. The magnetic force rapidly draws the full nucleic acid dose towards the cells and delivers the nucleic acid leading to fast and efficient transfection.

2.5.5 Measurement of the Luciferase Activity

Twenty- four hours after transfection, cells were checked to be no more than 95% confluent. The growth medium was removed from the cultured cells and 2ml of phosphate buffered saline (PBS) was added to wash the surface of the 24-well plate. The plate was swirled briefly to remove detached cells and residual growth medium. The rinse solution was completely removed before applying passive lysis buffer (PLB, promega, UK). 80 µl of 1X PLB was added into each well in the 24-well plate to completely cover the cell monolayer. The plate was wrapped with foil paper and stored at -80°C for at least 24 hours prior to performing the DLR™ Assay.

Luciferase Assay Reagent II (LAR II, Promega, UK) was prepared by resuspending the provided lyophilized Luciferase Assay Substrate in 10ml of the supplied Luciferase Assay Buffer II. Once the substrates and buffer had been mixed, LAR II was aliquot and stored at -80°C. Frozen aliquot (80 µl / sample, 2.5ml) of LARII was thawed at room temperature before use. Stop & Glo® reagent was prepared by adding 1 volume of 50X Stop & Glo® Substrate (promega, UK) to 50 volumes of Stop & Glo® buffer. For each measurement (80 µl / sample) 50 µl of 50X Stop & Glo® Substrate was added to 2450 µl of Stop & Glo® buffer just before use. The frozen 24-well plates were thawed at room temperature before use. 40 µl of cell lysate from each sample was carefully transfered into a 96-well clear bottom Castor plate without bubbles. 40 µl of LAR II and Stop & Glo® reagents were added sequentially into the same sample and the each luciferase activity was read by BIO-TEK *Synergy* HT multiwell plate reader.

2.5.6 Detection of the *COL1A1*-luciferase Fused Protein Expression

To confirm expression of the *COL1A1*-luciferase fusion protein in transfected HOS TE85 cells, total RNA of the cells was extracted, and reverse transcribed into cDNA. PCR reaction was then carried out on this cDNA with one forward primer in *COL1A1* exon1 and the reverse primer in luciferase cDNA. The PCR product was loaded on gel to verify the length followed by DNA sequencing.

2.5.6.1 RNA Extraction

Twenty- four hours post- transfection, cells were collected for RNA extraction. The growth medium was removed from cultured cells and 330 µl of 1X cold PBS was added to wash the cells. The rinse solution was completely removed before applying 330 µl of TRrizol reagent to each well. The lysised cells were scraped off the plate and transferred to DEPC treated 1.5ml eppendorf. The lysate from the same transfection were collected in one eppendorf and mixed by pipetting up and down. The lysate was incubated for 5 minutes at the room temperature. 200 µl chloroform was added into each sample and mixed by shaking vigorously be hands for 15 seconds. The mixture was incubated for 3 minutes at room temperature and then centrifuged at 12000g for 15 minutes at 4°C. The aqueous phase (500-600 µl) was transferred carefully to the fresh DEPC treated 1.5ml eppendorf. 500 µl of isopropyl alcohol was added to each sample and mixed by inverting. The sample was incubated for 10 minutes at room temperature and then centrifuged at 12000g for 10 minutes at 4°C. The supernatant was removed and the pellet was washed with 1ml 75% ethanol for

each sample. The sample was mixed by vortexing and then centrifuged at 7500g for 5 minutes at 4°C. The supernatant was removed and the pellet was dry for 5 minutes on the bench. 20 µl of DEPC treated water was added to each sample and the RNA pellet was dissolved by heating at 60°C for 10 minutes. RNA was stored at -80°C.

2.5.6.2 RNA Purification and Reverse Transcription

The samples were treated with DNAse I to remove contaminated plasmid DNA before reverse transcription. The DNase I (Invitrogen, UK) clean reaction was set up as follows:

1 µl	DNase I (1U/µl)
1 µl	10X buffer
1 ug	RNA
ddH_2O up to 10 µl	

The reaction was incubated at room temperature for 15 minutes. 1 µl of EDTA (Invitrogen) was added to terminate the reaction. The mixture was incubated at 65°C for 10 minutes to inactive the DNase I.

The reaction products were then used for cDNA production by reverse transcription. The following components were added to a nuclease-free microcentrifuge tube:

1 µl	oligo $(dT)_{20}$ (50um, Invitrogen, UK)
1 µl	dNTP mix (10mM each, Invitrogen, UK)
11 µl	1ug clean RNA (from the DNase I reaction directly)

The mixture was heated to 65°C for 5 minutes and incubated on ice for at least 1 minute. The following components were added to the mixture:

4 µl 5X fist- strand buffer (Invitrogen, UK)

1 µl 0.1M DTT (Invitrogen, UK)

1 µl RNaseOut Recombinant RNase Inhibitor (40U/µl, Invitrogen, UK)

1 µl SuperScript III RT (200U/µl, Invitrogen, UK)

For the control reaction 2 µl of water was added to the reaction in stead of RNaseOut Recombinant RNase Inhibitor and SuperScript III RT. The reactions were mixed by pipetting up and down gently. The reaction was incubated at 50°C for 60 minutes and inactivated at 70°C for 15 minutes.

2.5.6.3 PCR Amplification of *COL1A1*-luciferase Fused cDNA

PCR primers were designed by Roche Universal Probe Library to optimize the amplification of *COL1A1*-luciferase fusion cDNA. After the optimization the following primers were used for amplification of a 242 bp fragment from *COL1A1*-luciferase fused cDNA: *Col1a1*_6F 5' ATGTTCAGCTTTGTGGACCTC 3' (binding from the start code of COL1A1 exon1) and *Col1a1*_5R 5' AGTTGCTCTCCAGCGGTTC 3' (binding to the luciferase cDNA). The PCR reaction was set up as following:

2.5 µl 10x PCR buffer (Qiagen, UK)

2 µl dNTP (10mM each)

5 µl	Q-solution (Qiagen, UK)
2.5 µl	primer mix (10µM each of forward and reverse primer)
0.625 µl	Taq DNA polymerase (Qiagen, UK)
5 µl	cDNA
4.875 µl	ddH_2O

The reaction was transferred to a PCR microplate and mixed by pipetting up and down. The thermal cycling protocol consisted of an initial incubation for 10 min at 95°C, followed by 35 cycles of 15 sec at 95°C, 30 sec at 60°C and 15 sec at 72°C.

This 242 bp qPCR product was then loaded on agarose gel to verify the length followed by DNA sequencing.

2.6 Electrophoretic mobility shift assay (EMSA)

2.6.1 An Overview of EMSA

Electrophoretic mobility shift assay (EMSA) is a rapid and simple technique used to detect the interaction of DNA binding proteins with their DNA recognition sequences. This method has been used widely to discover new proteins such as transcription factors or protein mixtures that are capable of binding to a specific DNA sequence. It is based on the observation that protein-DNA complexes migrate through a non-denaturing polyacrylamide or agarose gel more slowly than free DNA molecules. EMSA is carried out first by incubating the purified protein or a protein mixture such as a nuclear extract with a radioactive or fluorescence labeled DNA fragment containing the putative protein binding site(s). The reaction products are then analyzed on a nondenaturing polyacrylamide gel. The specificity of the DNA-binding protein for the putative binding site is established by competition experiments using DNA fragments or oligonucleotides containing a binding site for the protein of interest or other unrelated DNA sequences.

To detect the putative binding sites of transcriptional factor(s) in or around the -1997G/T and the -1663in/del T polymorphisms, several EMSAs were performed according to the manual of LightShift Chemiluminescent EMSA Kit (Pierce, UK). Sense oligonucleotides corresponding to "-1997G", "-1997T", "-1663delT" and "-1663inT" were synthesized and 3' end-labelled with biotin (invitrogen). Antisense oligonucleotides were synthesized (invitrogen) and annealed with sense biotin-

labelled oligonucleotides. The labelled oligonucleotides were incubated with cell nuclear extract (MG63 and HOS TE85), pure human recombinant Sp1 (Promega, UK) or Usf (SantaCruz, USA). Varying concentrations of unlabelled competitor oligonucleotides were added to assess relative binding affinity of transcription factor(s) to the "G", "T", "delT" and "insT" alleles. To identify the DNA-bound transcription factor(s), some antibodies (anti-Sp1, anti-Usf, anti-Osterix and anti-Nmp4) were added to the binding reaction. The reaction products were then electrophoresed on a 6% (w/v) native polyacrylamide gel (invitrogen). The DNA was then rapidly (30 minutes) transferred to a positive nylon membrane, UVcross- linked, probed with streptavidin-HRP conjugate and incubated with the substrate. The biotin-labelled DNA was then visualized with Gene GOME imager (Syngene Bio Imaging, UK). Band intensity was analysed with Gene Tools software

2.6.2 Nuclear Protein Extraction for EMSA

Nuclear extracts from MG63 and HOS TE85 cells were used for EMSA experiments. Cells were cultured in 75 cm^2 tissue culture flask. Once confluent, the cells were detached from the flask and the nuclear protein extractions were prepared by using nuclear extraction kit (Acti motif). The extracts were aliquoted and stored frozen at -80° C until use.

The concentration of the nuclear extract was determined with BSA as a standard protein by the bicinchoninic acid protein assay (Pierce, USA). To set up the standard curve, 10 µl of each concentration BSA (up to 20ug/ µl) was added into 200 µl of BCA (dye). Several dilutions of each nuclear protein (1 µl, 1.5 µl and 2 µl) were measured as expected concentration was about 3-5ug/ µl. All the samples were carried out duplicate.

2.6.3 Design and Formation of EMSA Oligonucleotide Probe

The oligonucleotide probes used in EMSA experiments were designed to span the two promoter polymorphisms and surrounding bases within the promoter. All the single stranded oligonucleotides were synthesized and HPLC purified (Invitrogen). All the forward single stranded oligonucleotides were biotin 5' end-labelled. Double stranded oligonucleotides were made from forward and reverse single stranded oligonucleotides (Table 2.2)

Table 2.2 Single stranded oligonucleotides used in EMSA probe formation

Oligonucleotide	DNA sequence
-1663insT (8T) F	Biotin-CGGCCAGCCGACGTGGCTCCCTCCCCTTCTGTTCC**TTTTTTTT**CCCCTTTGCC
-1663insT (8T) R	GGCAAAGGGG**AAAAAAAA**GGAACAGAAGGGGAGGGAGCCACGTCGGCTGGCCG
-1663delT (7T) F	Biotin-CGGCCAGCCGACGTGGCTCCCTCCCCTTCTGTTCC**TTTTTTT**CCCCTTTGCC
-1663delT (7T) R	GGCAAAGGGG**AAAAAAA**GGAACAGAAGGGGAGGGAGCCACGTCGGCTGGCCG
-1997G F	Biotin- CGCCTATTAGGGAGGGG**G**TCTCTTGGAACTGACCC
-1997G R	GGGTCAGTTCCAAGAGA**C**CCCCTCCCTAATAGGCG
-1997T F	Biotin- CGCCTATTAGGGAGGGG**T**TCTCTTGGAACTGACCC
-1997T R	GGGTCAGTTCCAAGAGA**A**CCCCTCCCTAATAGGCG
sp1T F	Biotin- AGGGAATG**T**GGGCGGGATGAGGGCCT
sp1T R	AGGCCCTCATCCCGCCC**A**CATTCCCT

Polymorphisms are in **Bold**, underlined. F: forward oligo; R: reverse oligo

To generate a double stranded oligonucleotide, the complementary reverse oligonucleotide strand was annealed to the biotin-labelled forward oligonucleotide strand in a microcentrifuge tube:

2µl 1µM biotin-labelled forward oligonucleotide strand

3µl 1µM reverse oligonucleotide strand

5µl	10X annealing buffer
40µl	Deionized water

The reagents were mixed well and centrifuged briefly before being incubated at 90°C for 1 minute. The reactions were allowed to cool slowly to room temperature on the bench (approximately 1 hour), resulting in annealing of complementary oligonucleotide strands. The double stranded biontin-labelled oligonucleotide probes were stored frozen at -20°C.

Varying concentrations of unlabelled double stranded competitors of the same sequence as the labelled oligonucleotides were made by incubating corresponding amount of unlabelled forward and reverse oligonucleotides with annealing buffer and deionized water at 90°C for 1minute in a microcentrifuge tube. The reactions were left to cool to room temperature and stored frozen at -20°C.

2.6.4 The EMSA Binding Reaction

The binding reaction was performed in a 0.5 ml microcentrifuge tube at room temperature. The reaction components were added in the order listed below. To overcome strong non-specific interactions, a short pre-incubation at room temperature was done before adding the biotin- labelled probe. The binding reactions were incubated at room temperature for 20minutes. 5 µl of 5X loading buffer was then added to each sample, pipetting up and down several times.

2 µl 10X binding buffer

1 µl (1 µg/µl) Poly (dI.dC) (only used in the binding reactions for -1663in/delT)

1 µl protein source (nuclear extracts or hrSp1 or USF)

1 µl biotin- labelled oligonucleotide probe

Deionized water to final volume of 20 µl

2.6.5 Competition Assays and Super-shift EMSA

Competition EMSAs were carried out to compare the binding affinities of the polymorphisc variants for relevant nuclear proteins. Unlabelled double stranded oligonucleotides of varying concentration were added to the binding reaction to compete with biotin- labelled probes. The greater the binding affinity of the competitor, the weaker the resultant gel shift band. In super-shift assays antibodies (Sp1, Usf, Osterix, Nmp4, Actin, NFATc1) were added to the binding reaction before adding the biotin- labelled probe. Each antibody bound specifically to its protein resulting in a retarded band on the gel shift.

2.6.6 Electroporesis of Binding Reactions

A pre-cast 4-6% native polyacrylamide gel (12 wells, Invitrogen) was pre-electrophoresed in 0.5X TBE buffer at 100V for 60minutes. Each well was flushed by the running buffer and then 20µl of each sample was loaded on the gel. The gel was run at 100V until the bromophenol blue dye migrated approximately 3/4 down the length of the gel.

To transfer the binding reactions to nylon membrane by electrophorese, the gel and nylon membrane (soaked in 0.5X TBE for at least 10minutes) were formed a sandwich with two wet sponges and blotting papers in a clean electrophoretic transfer unit according the manufacturer's instructions (BioRad). Transfer was carried out in cool 0.5X TEB at 380 mA for 60minutes.

When the transfer was complete, the membrane was cross-linked for 15minutes with the membrane face down on a transilluminator equipped with 312 nm bulbs.

2.6.7 Detecting Biotin- labelled DNA by Chemiluminescence

The cross-linked membrane was then incubated with 10 ml Blocking Buffer (Pierce) for 15 minutes at room temperature with gentle shaking on an orbital shaker. After that, the membrane was incubated with 10 ml Conjugate/ Blocking Buffer for 15 minutes, and then washed four times with 10 ml Washing Buffer according the manufacturer's instructions (Pierce). After washing, the membrane was transferred to a clean tray and incubated with 10 ml Substrate Equilibrium Buffer for 5 minutes with gentle shaking. The membrane was then incubated with 600 µl Substrate Working Solution in a clean tray for 5 minutes without shaking.

The membrane was then exposed to Gene GOME imager (Syngene Bio Imaging, UK) in a sealed exposure cassette. Continual exposure from 1 to 4 minutes was used to visualize the membrane. Band intensity was analysed with Gene Tools software (Syngene Bio Imaging, UK). After the exposure, membrane was wrapped with kitchen cling film and stored at 4°C.

2.7 Biomechanical Testing

All bone samples were stored at 4°C in a calcium phosphate-buffered 0.15M saline solution (containing 0.2 mM $CaCl2{\cdot}2H_2O$, 0.2 mM Na_2HPO_4, 0.01 mM $Na_4P_2O_7{\cdot}10H_2O$, and 0.4g of sodium azide/l) because this has been shown to preserve the structure and composition of the bone (Lees S 1988).

Cylindrical cores of bone were removed from the superior region of the femoral head using a 9-mm-diameter hollow drill bit as described previously (Li B & Aspden RM, 1997). This site was chosen to represent the region being the most highly loaded in vivo (Hodge WA *et al,* 1986). Articular cartilage and the subchondral bone plate were removed from the bone cores, and both ends were trimmed parallel. The resulting cores of trabecular bone, having a mean ± SD of height 9.3 ± 1.2 mm, were subjected to an unconstrained compression test using an Instron 5564 materials testing machine (Instron, High Wycombe, UK). The test was performed at a strain rate of 10% per minute. The apparent density of washed and defatted bone samples was calculated by dividing the total mass by the sample volume, calculated from calliper measurements of diameter and length. The stiffness and yield were determined from the characteristics of the stress-strain curve using standard techniques described previously (Li B & Aspden RM, 1997).

Genotyping was carried out by PCR and DNA sequencing as described in section 2.4.

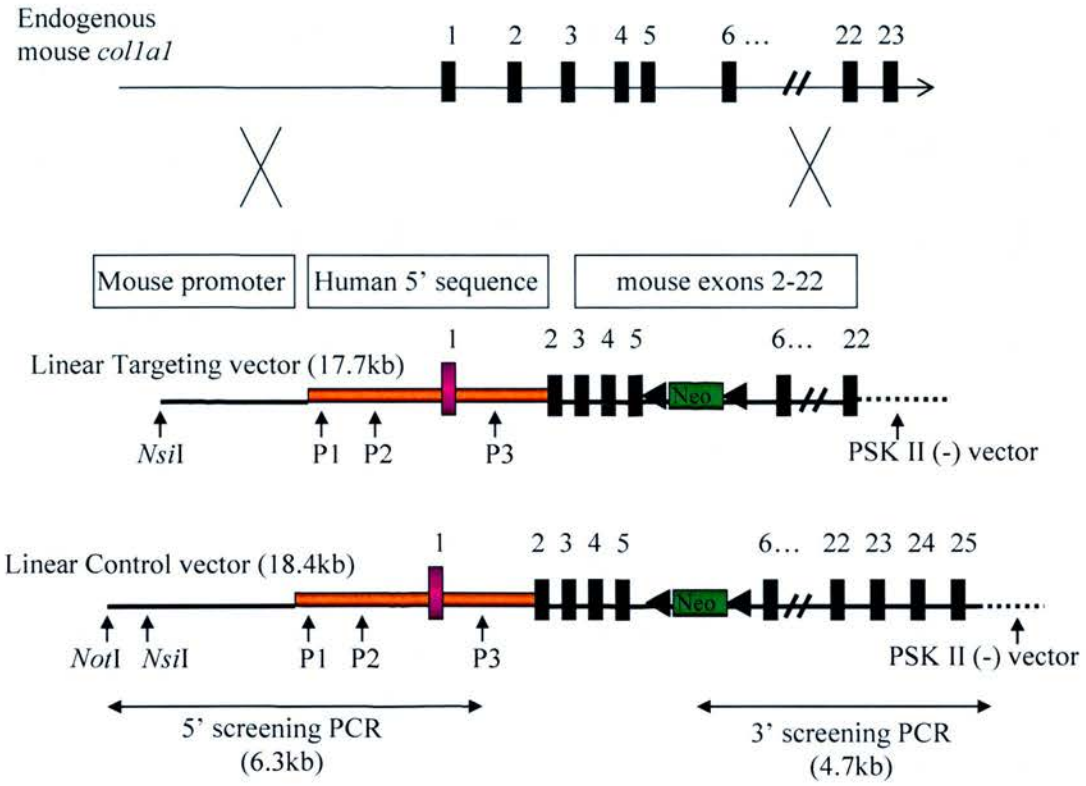
2.8 Generation of gene targeting constructs for polymorphic *COL1A1* alleles

2.8.1 Generation of Targeting Construct by Long PCR

The targeting construct contained 3.9 kb of the mouse *Col1a1* promoter, 3.6 kb of human *COL1A1* 5' sequence (2.3 kb promoter and exon1 & 1.3 kb intron1) used to replace the proximal mouse *Col1a1* promoter, 5.3 kbof mouse *Col1a1* (exons 2 to 22) and a selection marker (2 kb of Loxp-Neo-Loxp cassette), as shown in Figure 2.3.

The pBluescript SK II (-) plasmid (pSK II) was used as the background vector in all the cloning steps carried out in this expepriment.

Figure 2.3 Schematic presentation of endogenous mouse *Col1a1* gene, targeting vector and control vector.



This targeting event will replace the proximal promoter of the mouse *Col1a1* gene with the human *COL1A1* promoter, exon1 and intron1. Loxp-Neo-loxp cassette serves as a positive selection marker. A control vector with two extended homologous ends was generated to screen recombination events at both homologous ends using long-PCR reactions. The black boxes represent mouse *Col1a1* exons. The gold box represents human *COL1A1* promoter and intron1. The purple box represents human *COL1A1* exon1. P1, P2 & P3 represent three polymorphisms: -1997G/T, -1663ins/delT & sp1G/T, respectively. *Not*I and *Nsi*I site were used to linearize the targeting for the targeting experiments.

The human exon1 was mocked to generate an exon with the same amino acid sequence as mouse and keep the RNA splicing by inserting some mouse sequence into the 5' and the 3' end of intron1 (Figure 2.4). The full sequence of targeting vector is presented in appendix 1.

Figure 2.4 Sequence of the exon1 and RNA splicing sites in targeting vector

The targeting vector comprised the human promoter sequence (in black), and part of the human exon1 (tailored to the same length as mouse exon1). The RNA splicing sites (in orange boxes) of mouse *Col1a1* gene were introduced to the 5' and 3' of the human intron1 by PCR primers.
Letters in upper case: exon sequence; lower case: non-coding sequence; underlined: amino acid codons.

Only the DNA fragment of *Col1a1* exons (2-22) was generated by *Hinc*II and *Hind*III double restriction digestion of PCOL-Xho-Wt 10x plasmid (from Mann V). All the other four DNA fragments were amplified by long PCR amplification. All the PCR primers are listed in Table 2. 3 and the PCR conditions are listed in Table 2.4.

Table 2.3 PCR primers used in subcloning

Name of fragment	Name of primer	Primer sequence 5'→ 3'
Mouse col1a1 promoter	MoupromM13F	AACAGCTATGACCATGATTACGCCAAG
	Mouprom6kb***Spe*I**R	AGAAGG***ACTAGT*** GGTGGTGTGGGAAGGCACTG
Human COL1A1 promoter& exon1	Humprom ***Spe*I**F	AGGAAGTGG***ACTAGT***CACTAACCCTCA
	Humprom***Eco*RI**R	GTAGT***GAATT***CCCCAGAGTTTGGAACTTACTGTC CTCTTGGCCGTGCGTCAG
Human COL1A1 intron1	HumIntF7 ***Eco*RI**	TCAGTG***GAATTC*** GGGAGTGCAAGGATACTCTATA
	HumIntr7 ***Eco*RV**	GACTAC***GATATC***GCTCCTCATCAGCGCGGGTGA
Loxp-neo-loxp cassette	Loxp-***Afl*II**	TGACG***CTTAAG***GATCAATTCCGATCATA
	Loxp-***Mfe*I**	TGACGC***CAATTG***TCGATAAGCTTCGAGGGACCTA

F, forward primer; R, reverse primer; underlined, mismatched bases used to introduce a restriction site; ***Bold Italics***, introduced restriction site.
Sequence homologous to the end of mouse *Col1a1* exon1 is shown in blue.

All the PCR were performed in 0.2 ml PCR tube in a 50 µl volume containing the following reagents:

5 µl 10X Pfx amplification buffer (Invitrogen, UK)

5 µl dNTP mixture (10mM each)

5 µl 10 X PCRx Enhancer solution (Invitrogen, UK)

1 µl 50 mM MgSO4

2.5 µl primer mix (10 mM each)

0.8 µl Platinum® *Pfx* DNA polymerase (Invitrogen, UK)

100 ng of template DNA with deionized water to final volume of 50µl

The PCR reactions were carried out on a Gene Engine thermocycler (MJ Research Inc, USA). The PCR products were visualized on agarose gels directly.

Table 2.4 PCR conditions for subclonings

Name of fragment	Template	Thermal cycling protocol	Product size (bp)
Mouse COL1A1 promoter	P: mouse 6kb (from Bonstein)	Initial incubation: 94°C for 3min 10 cycles: 94°C for 15sec, 62°C for 30sec + 68°C for 3min 20 cycles: 94°C for 15sec, 62°C for 30sec 68°C for 3min (10sec increased per cycle)	3900
Human COL1A1 prmoter& exon1	P: 2300COL1WT (from Hobson E)	Initial incubation: 94°C for 3min 20 cycles: 94°C for 15sec, 60°C for 30sec + 68°C for 4.5min	2300
Human COL1A1 intron1	P: 2300G (from Hobson E)	Initial incubation: 94°C for 3min 9 cycles: 94°C for 15sec + 68°C for 2min 15 cycles: 94°C for 20sec + 68°C for 2min + 10sec increased per cycle Final extension: 68°C for 3min	1300
Loxp-neo-loxp cassette	P: Loxp vector (From Albagha O)	Initial incubation: 94°C for 3min 20 cycles: 94°C for 15sec, 60°C for 30sec + 68°C for 4.5min	2000

P: plasmid vector.

2.8.2 Subcloning and Site Direct Mutagenesis (SDM)

To subclone each targeting fragment (mouse *Col1a1* promoter, human *COL1A1* promoter & exon1, human *COL1A1* intron1 and mouse *Col1a1* exons 2-22) into pBlueScript® II SK (-) Phagemid vector (PSK (-), Stratagene, UK), restriction digest was required to make sticky or blunt ends on both the inserts (PCR products) and PSK (-) vector. The optimum incubation conditions for each restriction enzyme used in subcloning are listed in Table 2.5.

Table 2.5 Enzymes and incubating conditions for subcloning.

Name of subclone	Restriction enzyme	Restriction sequence (5'→ 3')	Optimum condition
Mouse COL1A1 promoter	***Eag*I**	C ∣ GGCCG GCCGG ∣ C	NEBuffer3, 37°C for 3 hours
	***Spe*I**	A ∣ CTAGT TGATC ∣ A	NEBuffer2, 37°C for 3 hours
Houman COL1A1 promoter & exon1	***Spe*I**	A ∣ CTAGT TGATC ∣ A	NEBuffer2, 37°C for 3 hours
	***EcoR*I**	G ∣ AATTC CTTAA ∣ G	NEbuffer *EcoR*I, 37°C for 2 hours
Human COL1A1 intron1	***EcoR*I**	G ∣ AATTC CTTAA ∣ G	NEbuffer *EcoR*I, 37°C for 2 hours
	***EcoR*V**	GAT ∣ ATC CTA ∣ TAG	NEBuffer3, 37°C for 3 hours
Mouse COL1A1 exons (2-22)*	***Hind*III**	A ∣ AGCTT TTCGA ∣ A	NEBuffer2, 37°C for 3 hours
	***Hinc*II**	GTY ∣ RAC CAR ∣ YTG	NEBuffer3, 37°C for 3 hours
Loxp-Neo-Loxp cassette	***Afl*II**	C ∣ TTAAG GAATT ∣ C	NEBuffer2, 37°C for 3 hours
	***Mfe*I**	C ∣ AATTG GTTAA ∣ C	NEBuffer4, 37°C for 3 hours

The cutting position inside a restriction site is indicated by symbol '□'.
*: made by Mann V.

All the digest reactions were performed in 50 µl reaction with the following reagents:

5 µl	10 X buffer (specific for the restriction enzyme, NEB, UK)
2.5 µl	20 X BSA (NEB, UK)
30 units	restriction enzyme
30 µl	PCR product (or up to 3 ug plasmid)
deionized water to final volume of 50 µl	

After enzyme digestion all the products were visualized by 1% agarose gel electrophoresis. Products of the correct size were extracted from agarose gel by a Gel extraction kit (QIAGEN, UK) according to the manual. One µl of the purified DNA was then loaded on 1% agarose gel along with 1 µl of 1kb DNA ladder (NEB, UK) for approximate quantification.

T4 DNA ligase (Roche, UK) was used to insert the DNA fragment into the PSK (-) plasmid vector in each subcloning. The amount of insert and vector in the ligation reaction were calculated by the same formula mentioned in section 2.5.2.3.

Therefore the ligation reaction mix was set up as follows:

1-2 µl	T4 DNA ligase (1 U/µl)
1 µl	10 X ligation buffer
80-100 ng	vector DNA
Inserted DNA with deionized water to final volume of 10 µl	

A control ligation was set up parallel each time with the same amount of vector, liagase, ligation buffer and deionized water to final volume of 10 µl. The ligation

reaction was performed in 0.2 ml PCR tube and incubated at 16°C for 16 hours on a Gene Engine thermocycler (MJ Research Inc, USA).

After 16 hours incubation, 1 to 5 µl of the ligation mix was transformed into 50 µl of XL1-Blue supercompetent Cells (Stratagene, UK) according to the manual. 200 to 250 µl of each transformation were spread on pre-warmed carbenicillin selective agar plates. Plates were incubated overnight at 37°C. The transformation reaction was stored at 4°C. Additional cells may be plated out the next day, if necessary.

Average 4 to 10 colonies were chosen from each transformed plate to incubate in 5 ml LB broth at 37°C overnight with constant shaking at 225rpm. Qiagen miniprep kit was used to purify plasmid DNA from the bacterial culture. The concentration of each plasmid DNA was determined by the absorbance ratio at 260/280 using a Bio-TEK plate reader (Synergy HT, UK). A 20 µl restriction digest was set up to check if the DNA was inserted into the correct position on the vector. The restriction enzymes and their digested products were listed in Table 2.6. The digest products were then loaded on agarose gel to check the sizes. The plasmids with correct digestion products are then sequenced with primer M13F and M13R which flank the multiple cloning sites (MCS) region of PSK II (-). All the primers used in sequencing reaction were listed in Appendix 1. For long term storage of plasmid glycerol stocks were made using 700 µl of transformed bacterial culture mixed with 200 µl 40% (v/v) sterile glycerol. This mix was stored at -80°C.

Table 2.6 Restriction enzymes used to check each ligation, and their products

Plasmid (bp)	Restriction enzyme	Expected bands on agarose gel (bp)
mouse COL1A1 promoter (6900)	*Eag*I	2222, 4713
human COL1A1 pormoter&exon1 (5300)	*Bsa*I	2778, 2001, 480
human COL1A1 intron1 (4300)	*Sac*I	4099, 199
Mouse COL1A1 exons 2-22 (8239)	*Kpn*I	4966, 3273

The desired base changes on the three polymorphic sites of 5' *COL1A1*, and the restriction sites *Hind*III and *EcoR*I were introduced into subclones by the QuickChange™ Site- Directed Mutagenesis Kit (Strategene, UK). Primers used to make base change on three polymorphic sites are listed in Table 2.1 (page 97). Primers used to introduce *Hind*III and *EcoR*I are listed in Table 2.7. All the primers were designed complementary between 35-45bp with PAGE purification. The mutagenesis reaction was carried out in 0.2ml PCR tubes. The reaction was prepared as indicated below:

5 µl	10X reaction buffer
10-50ng	dsDNA template (plasmid DNA)
1.25 µl	primer F (100 ng/µl)
1.25 µl	primer R (100 ng/µl)
1 µl	dNTP mix

ddH_2O was added in to a final volume of 50 µl. One µl of *PfuTurbo* DNA polymerase (2.5 U/µl) was added last into the reaction. The precise cycling parameters were dependent upon the type of mutation required and the plasmid length.

All the nicked vectors containing desired mutations were purified from bacterial cultures using Mini Prep Kit (Qiagen, UK) and then sequenced to confirm the base changes. The primers used in sequencing reaction are listed in Appendix 1.

Table 2.7 Primers used in site-directed mutagenesis

Prime name	primer sequence (5'→ 3')
*Hind*III_SDMF	CGCGAGCTCGAGAT***A*AGCTT**TGGTGGTTGGGTGATGG
*Hind*III_SDMR	CCATCACCCAACCACCA***AAGCTT*A**TCTCGAGCTCGCG
*EcoR*I_SDMF	CGACGGTATCGATAAGCTT***GAATTC***GCTCCTCATCAGCGCGG
*EcoR*I_SDMR	CCGCGCTGATGAGGAGC***GAATTC***AAGCTTATCGATACCGTCG

F, forward primer; R, reverse primer; ***Bold Italics***, restriction site introduced by SDM primer. Mutation introduced by SDM primer was indicated underlined.

2.8.3 Transfection of Embryonic Stem (ES) Cells

To introduce the targeting vectors into mouse ES cells by electroporation, the targeting vectors were first purified from a bacterial mega-culture (100ml) using Endofree Maxi Prep Kit (Qiagen, UK). The concentrations and the purity of each targeting vector were determined by Bio-TEK plate reader using the same method as described before.

The targeting vectors were then linearized by *Not*I restriction digest in the first targeting reaction because the *Not*I restriction enzyme only had one recognition site on the upstream of mouse *Col1a1* promoter. The *Not*I digest reaction was set up as follows:

200 µl	targeting plasmid (300 ug, ratio of 260/280~1.80)
30 µl	*Not*I (10 U/ µl, NEB, UK)
30 µl	10 X Buffer
3 µl	100 X BAS
37 µl	ddH2O

The above reagents were mixed well in a 0.5ml microcentrifuge tube and then aliquoted at 50 µl into 0.2 ml PCR tubes. The 0.2 ml PCR tubes were incubated at 37°C for 5 hours on Gene Engine thermocycler. To check if the NotI digestion was complete, 1 µl of the digest product from each tube was loaded on a 0.7% agarose gel along with 1kb DNA ladder. The digested products were purified using phenol chloroform extraction method. The completely digested products were sent to Julia Doran's lab (MRC Human Genetic Unit) to perform the first targeting experiment.

After the electroporation, the cells were incubated at 37°C overnight before adding G418 (175 μg/ml) for selection of recombinant colonies..

For the second targeting, the same amount of ES cells with 100ug DNA saved from the first targeting was electroporated following the same procedure. The G418 selection of the recombinant ES cells was carried out as the first time. The targeting procedure for both constructs was conducted in parallel.

For the third targeting, the constructs were prepared as described above using Qiagen Endofree Maxi Prep Kit and linearized by *Not*I and *Nsi*I restriction digest (Figure 2.3). The linear DNA was extracted by Elutrap Electro-seperation system (Schleicher & Schuell). 100ug of purified DNA was used in the third targeting by the same procedure as the first and the second targeting.

2.8.4 Detecting the Recombination Events by Long PCR

Two long PCR reactions were designed to screen the recombination events at 5' and 3'. ES cells survived a selective G418 medium after electroporation were lysed and the DNA was extracted. The concentration and purity of the extracted DNA were analyzed by Bio-TEK plate reader. One μl of DNA from each ES cell and 1 μl of 18.4kb control construct were used in PCR screening.

The PCR conditions for both 5' and 3' screening were optimized using 18.4kb control construct and/ or mouse 129/ola genomic DNA as template prior to the targeting. The formula (from GenOway) to calculate 1 equivalent copy of 18.4kb control vector to 150ng 129/ola genomic DNA is:

$$\{18400 \text{ bp} \times 150 \text{ ng}\} / 6 \times 10^9 \text{ bp} = 4.6 \times 10^{-4} \text{ ng}$$

The control vector was then diluted to 0.1 equivalent copy, 1 equivalent copy and 10 equivalent copies used as template to optimize both PCR screening conditions. 150ng of mouse 129/ola genomic DNA was used as control solo or mixed with control vector in PCR optimization. Both 5' and 3' screening PCR were carried out in 0.2ml PCR tube in a 25 μl volume containing the following reagents:

2.5 μl	10X PCR buffer (Qiagen)
2 μl	dNTP mixture (10mM each)
5 μl	5X Q solution (Qiagen)
2.5 μl	primer mixture (10nM each)
1 μl	DNA (extracted from each ES cells or control construct)
0.8 μl	Taq DNA Polymerase (12.5 U/ μl, Qiagen)

distilled H2O to final volume of 25 µl

The primers for 5' screening PCR (6.3kb) are: ACTTGGGAGCTGCGATGC and CTCATCCCGCCCCCATTC. The thermal cycling conditions consisted of an initial incubation of 2 min at 94°C, then 9 cycles of 30 sec at 94°C and 6 min 30 sec at 68°C, followed by 20 cycles of 20 sec at 94°C, 6 min 30 sec at 68°C with 10 seconds increased per cycle, and followed by an final extension of 3 min at 68°C.

The primers for 3' screening PCR (4.7kb) are: GGTGGGGGTGGGGTGGGATT and GGAGCACCACGTTCACCGGA. The thermal cycling conditions consisted of an initial incubation of 4 min at 94°C, then 9 cycles of 30 sec at 94°C, 1 min at 55°C and 5 min at 68°C, followed by 20 cycles of 30 sec at 94°C, 1 min at 55°C and 5 min at 68°C with 10 seconds increased per cycle, and followed by an final extension of 3 min at 68°C. The PCR reactions were carried out on a Gene Engine thermocycler (MJ Research Inc, USA). The PCR products were visualized on 1% agarose gel directly.

CHAPTER THREE

Association study of 5' *COL1A1* polymorphisms and BMD

Abstract

Osteoporosis is a complex disease with a strong genetic component. The *COL1A1* gene is a strong functional candidate gene for susceptibility. A polymorphism affecting an Sp1 binding site in the first intron of *COL1A1* gene has been established as a functional variant and associated with genetic regulation of BMD, osteoporotic fracture and other osteoporosis-related phenotypes such as postmenopausal bone loss, bone geometry, bone quality and bone mineralization. However, recently two promoter polymorphisms of *COL1A1* were identified and found to be in strong linkage disequilibrium (LD) with the Sp1 polymorphism in a small study of Spanish post-menopausal women.

To investigate the three polymorphisms in relation to BMD in women, an association study was performed in a large population-based cohort of 3270 Scottish women.

In this study, the polymorphisms were in strong linkage disequilibrium and three haplotypes accounted for over 95% of alleles at the *COLIA1* locus. Individual polymorphisms were each associated with BMD, but the most consistent associations were with haplotypes defined by all three polymorphisms. For example, homozygote carriers of haplotype 2 (-1997G /-1663delT / +1245T) had reduced BMD at baseline (p=0.007 for LS-BMD; p=0.008 for FN-BMD), whereas homozygotes for haplotype 3 (-1997T / -1663insT / +1245G) had increased BMD (p=0.007 for LS-BMD). In conclusion, haplotypes defined by polymorphisms in the 5' flank of the *COLIA1* regulate BMD in a bidirectional manner in the large cohort of Scottish perimenopausal women.

Introduction

Osteoporosis is a common disorder characterized by low bone mass, micro-architectural deterioration of bone tissue and enhanced bone fragility leading to an increased incidence of fracture. Genetic factors have been recognized to play an important role in both osteoporosis and its associated phenotypes including bone mineral density (BMD) and fracture risk. Many candidate gene polymorphisms have been implicated in the pathogenesis of osteoporosis but one of the most widely studied is a polymorphism that affects an Sp1 binding site in the first intron of the *COLIA1* gene at position +1245 relative to the transcription start site (rs1800012). The *COLIA1* Sp1 binding site polymorphism has been associated with bone density, fragility fractures and other osteoporosis-related phenotypes such as postmenopausal bone loss, bone geometry bone quality and bone mineralization (reviewed by Ralston SH & de Crombrugghe B, 2006). Functional analysis has shown that the osteoporosis associated T allele of the Sp1 polymorphism is associated with increased DNA-protein binding, increased transcription, abnormally increased production of the collagen type I alpha 1 mRNA and protein and bone mineralization (Stewart TL *et al*, 2005 & Mann V *et al*, 2003). Several retrospective meta-analyses of published studies concluded that carriage of the T allele is associated with reduced BMD at the lumbar spine and femoral neck and with vertebral fractures (Ralston SH *et al*, 2006b; Mann V *et al*, 2003; Efstathiadou Z *et al*, 2001 & Mann V *et al* 2003). However, two promoter polymorphisms of *COL1A1* (-1997G/T; rs1107946 and -1663indelT; rs2412298) were recently found to be in strong LD with the Sp1 polymorphism. These polymorphisms were found to interact with each other and the Sp1 polymorphism to regulate BMD in a small study of Spanish post-menopausal women

(García-Giralt N*et al*, 2002). Also they were found to regulate reporter gene transcription in a promoter-reporter assay (García-Giralt N *et al*, 2005). The -1997G/T promoter polymorphism has been studied in relation to BMD in other populations and family based studied with mixed results although most of these studies have been of limited sample size (Yamada Y *et al*, 2005 & Liu PY *et al*, 2004). In this study therefore, the effects of the promoter and Sp1 polymorphisms of the *COLIA1* gene on BMD was investigated in a large population-based cohort of 3270 predominantly postmenopausal Scottish women.

Methods

The subjects were taken from the Aberdeen Prospective Osteoporosis Screening Study (APOSS) population. This was a population-based screening study for osteoporotic fracture risk, involving 3883 women aged 45-54 years who were initially assessed over a 4 year period between 1990 and 1994 and further assessed between 1997 and 1999. Information was collected on weight, height, menopausal status, use of hormone replacement therapy (HRT), smoking, alcohol use and exercise at the second visit. Bone mineral density (BMD) was measured at the femoral neck (FN-BMD) and lumbar spine (L2-4) (LS-BMD) by dual energy x-ray absorptiometry. Rate of bone loss was calculated as yearly percentage of change in BMD.

The participants in the present study comprised 3270 women who consented to give blood samples for DNA extraction and were successfully genotyped for at least one of the polymorphisms studied by PCR-sequencing. The details of the APOSS population and the genotyping method are described in chapter 2.

Statistical analysis

Statistical analyses were performed using Minitab version 12. The chi-square test was used to test for Hardy-Weinberg equilibrium. I used the ANOVA general linear model (ANOVA-GLM) routine of Minitab to study differences in BMD according to genotype and haplotype, with adjustment for age, body mass index (BMI), smoking (current smoker or non-smoker), physical activity level (PAL) score, menopausal status (premenopausal *vs.* perimenopausal and postmenopausal) and HRT use (never used HRT, previous HRT user, current HRT user). At the baseline visit, BMD values were adjusted

only for age, BMI, HRT and menopausal status, since information on smoking and PAL was not available at this time point. When significant differences were detected by ANOVA-GLM across genotype groups, the means were compared between groups using Tukey's test. Multiple regression analysis and stepwise multiple regression analysis was used to evaluate associations between genotypes and haplotypes and BMD and to estimate the effect size of the genotype and haplotypes in relation to other predictors of BMD. For this analysis, continuous variables (age, BMI and PAL) were entered into the statistical model as covariates. For categorical variables, such as menopausal status, HRT use, and *COLIA1* genotype and haplotype, dummy variables were constructed for use in the multiple regression analysis. Although I performed multiple statistical tests of three polymorphisms (and haplotypes defined by these polymorphisms) in relation to BMD at baseline and follow up at two skeletal sites, these factors were interrelated and therefore not independent. In view of this I used a modified Bonferroni correction to adjust for the number of independent tests performed to take account of the fact that the polymorphisms were in linkage disequilibrium and the BMD measurements at the spine and hip at baseline and follow up were strongly correlated. To make this correction, I took an average of the D' values for all 3 polymorphisms (0.92) and subtracted this figure from 1 to get an indication of the number of independent polymorphisms tested [1+2*(1-0.92)]. According to this calculation, the number of independent polymorphisms tested was 1.16. I analysed two skeletal sites (FN-BMD and LS-BMD) but these were significantly correlated (r=0.668) and taking this correlation into account, I effectively analysed 1.33 independent sites [1+ (1-0.668)]. Similarly, baseline and follow up BMD were strongly correlated (r=0.883) and therefore the analysis of genotype in relation to

baseline and follow-up BMD can be considered as 1.12 independent tests [1+ (1-0.883)]. In total therefore, I conducted 3.61 independent tests (1.16+1.33+1.12). Applying a Bonferroni correction to our data using this figure gives an adjusted threshold for significance of p=0.014, equating to the conventional threshold for significance of p=0.05. The sample size gave approximately 99% power of detecting a difference of 0.15* BMD Z-score units between genotype groups assuming a population frequency of 0.20 for the rare allele.

*: z-scores are the number of standard deviations below the average for a person of the same age, race and gender. For example, a person with a z-score of -1 has 2.5 times the chance of a hip fracture compared to a woman with an average bone density. The 0.15 z-score was picked based on the mean ± SD BMD values of this population (see next page).

Results

Characteristics of the Population

The average ± SD age of the participants at the baseline visit was 48.4 ± 2.3 years (range 44-56 years) and was 54.8 ± 2.3 years at the follow up visit. The average time between baseline and follow-up visits was 6.3 ± 0.85 years and did not differ between genotype or haplotype groups (data not shown). At the baseline visit, 44.9% of participants were premenopausal and 16.9% were current users of HRT. At the follow up visit only 3.2% of participants were premenopausal and the proportion of current HRT users had roughly doubled to 37.5%.

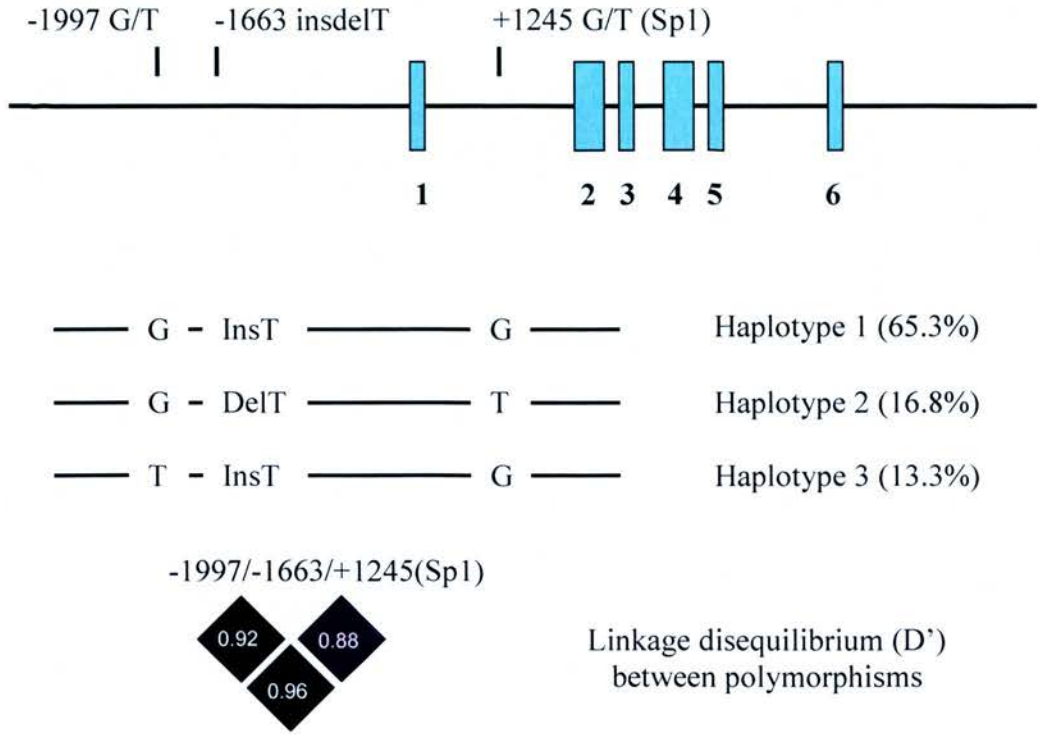
The average ± SD of the lumbar spine BMD at the baseline was 1.057 ± 0.329 g/cm^2 and was 1.010 ± 0.170 g/cm^2 at the follow up visit. For the femoral neck BMD, the baseline value was 0.881 ± 0.125 g/cm^2 and the follow up visit was 0.837 ± 0.122 g/cm^2.

The genotype rates were 98.77%, 88.87% and 82.87% for the Sp1 G/T, -1997 G/T and -1663ins/delT respectively. Genotype frequencies for all three polymorphisms were similar to those previously reported in other Caucasian populations (García-Giralt N *et al*, 2002 & Liu PY *et al*, 2004). All three polymorphisms were in strong linkage disequilibrium (LD) and D' values ranged from 0.88-0.96 (Figure 3.1). Genotypes for the -1997 G/T polymorphism and the Sp1 polymorphism were in Hardy-Weinberg equilibrium (HWE) ($p>0.05$) but genotypes for the -1663indelT polymorphism deviated significantly from HWE ($p=0.002$) due to over representation of homozygotes for the -1663InsT allele and under-representation of heterozygotes. In view of this I carefully scrutinized sequencing traces from all individuals who were heterozygous for -

1663indelT and all -1663delT homozygotes, but no errors were identified. I also re-sequenced from both directions (forward and reverse) for all the -1663indelT heterozygotes and all -1663delT homozygotes, no errors were identified. Thereby genotyping error was excluded as a cause of the HWE deviation.

Analysis of genotype data by the PHASE software program predicted the presence of all eight possible haplotypes in the APOSS population, but five of these were very rare, and three common haplotypes accounted for 95.4% of alleles at the 5' flank of the *COLIA1* locus (Figure 3.1). These were designated haplotype 1 (-1997G / -1663InsT / +1245G), which accounted for 65.3% of alleles; haplotype 2, which accounted for 16.8% of alleles (-1997G / -1663DelT / +1245T) and haplotype 3, which accounted for 13.3% of alleles (-1997T / 1663InsT / +1245G).

Figure 3.1 Common haplotypes in the promoter and intron 1 of *COLIA1*



Common haplotypes and their frequencies at the 5' flank of the *COLIA1* gene are shown in relation to the *COLIA1* gene structure. Shaded boxes represent exons. Linkage disequilibrium values (as D') between SNP are depicted in colored triangles at the bottom of the figure. Other haplotypes and their frequencies are shown in Table 5.4 page 202.

Association between Individual Polymorphisms and BMD

Relevant characteristics of the APOSS population at baseline and follow-up in relation to the individual polymorphisms are shown in Table 3.1. The BMD values shown are adjusted for age, body mass index, HRT use, and menopausal status at the baseline assessment and for age, body mass index, HRT use, menopausal status smoking and physical activity score at the follow-up assessment.

There was no difference in age, BMI, menopausal status or HRT use between the genotype groups at the baseline visit, but all three polymorphisms were associated with adjusted BMD at either the lumbar spine and/or femoral neck at baseline. For the -1997G/T polymorphism, there was a significant association with lumbar spine BMD (p=0.003) and comparison between genotypes showed that BMD values were significantly higher in homozygotes for the T allele.

For the -1663indelT, there was also a significant association with lumbar spine BMD values (p=0.006) and for the Sp1 polymorphism the p-value was nominally significant (p=0.018) but fell below the adjusted threshold for significance of p=0.012 when multiple testing was taken into account. For both the -1663indelT and Sp1 polymorphisms, comparison between genotypes showed that BMD values were significantly reduced in rare homozygotes. Nominally significant results were obtained for baseline femoral neck BMD in relation to the -1663indelT and Sp1 polymorphisms (p=0.024, p=0.048, respectively) but these did not reach the threshold for significance when multiple testing was considered. There was no significant association between -1997G/T and femoral neck BMD at baseline (p=0.196). When I performed multiple regression analysis, entering genotypes for all polymorphisms into the model assuming a recessive mode of

Table 3.1 APOSS subject characteristics and bone mineral density measurements according to *COLIA1* genotypes

Genotype	**-1997 G/G**	**-1997 G/T**	**-1997 T/T**	**P-value (ANOVA)**
N	2247 (74.6%)	699 (23.2%)	65 (2.1%)	(n=3011)
Age at baseline (yrs)	48.5 ± 2.4	48.5 ± 2.3	48.4 ± 2.3	0.885
BMI at baseline (Kg/m2)	25.4 ± 4.5	25.2 ± 4.2	25.3 ± 4.2	0.660
LS-BMD Baseline (g/cm^2)	1.053 ± 0.004	1.064 ± 0.006	1.111 ± 0.018 [a, b]	0.003
FN BMD Baseline (g/cm^2)	0.882 ± 0.003	0.890 ± 0.005	0.894 ± 0.014	0.196
Premenopausal / HRT Baseline	42.9% / 17.8%	46.6% / 16.5%	44.6% / 24.6%	0.24 / 0.25
LS-BMD Follow-up (g/cm^2)	1.037 ± 0.010	1.048 ± 0.011	1.109 ± 0.023 [a, b]	0.003
FN-BMD Follow up (g/cm^2)	0.844 ± 0.007	0.852 ± 0.008	0.878 ± 0.016	0.039
Premenopausal / HRT Follow-up	2.9% / 36.8%	2.9% / 34.3%	6.3% / 28.2%	0.28 / 0.45
Genotype	**-1663 Ins/Ins**	**-1663 Ins/Del**	**-1663 Del/Del**	
N	1967 (65.8%)	869 (29.0%)	151 (5.0%)	(n=2987)
Age (yrs)	48.4 ± 2.3	48.6 ± 2.5	48.9 ± 2.7	0.079
BMI (Kg/m2)	25.3 ± 4.4	25.3 ± 4.5	25.0 ± 4.3	0.610
LS-BMD Baseline (g/cm^2)	1.059 ± 0.004	1.054 ± 0.006	1.019 ± 0.019 [a, b]	0.006
FN BMD Baseline (g/cm^2)	0.885 ± 0.003	0.880 ± 0.004	0.860 ± 0.009 [a, c]	0.024
Premenopausal / HRT Baseline	44.1% / 17.5%	43.3% / 17.7%	39.3% / 20.0%	0.28 / 0.24
LS-BMD Follow-up (g/cm^2)	1.040 ± 0.010	1.037 ± 0.011	1.008 ± 0.017	0.117
FN-BMD Follow up (g/cm^2)	0.847 ± 0.007	0.842 ± 0.007	0.824 ± 0.013	0.074
Premenopausal / HRT Follow-up	3.0% / 39.6%	3.2% / 34.3%	0% / 38.7%	0.09 / 0.02
Genotype	**Sp1 G/G**	**Sp1 G/T**	**Sp1 T/T**	
N	2145 (66.5%)	963 (29.8%)	117 (3.6%)	(n=3225)
Age (yrs)	48.4 ± 2.3	48.5 ± 2.4	48.5 ± 2.3	0.178
BMI (Kg/m2)	25.3 ± 4.4	25.4 ± 4.6	24.8 ± 3.9	0.304
LS-BMD Baseline (g/cm^2)	1.060 ± 0.004	1.055 ± 0.006	1.020 ± 0.014 [a, c]	0.018
FN BMD Baseline (g/cm^2)	0.885 ± 0.003	0.885 ± 0.004	0.858 ± 0.011 [a, c]	0.048
Premenopausal / HRT Baseline	45.6% / 17.5%	44.8% / 17.7%	41.2% / 20.0%	0.28/0.24
LS-BMD Follow-up (g/cm^2)	1.040 ± 0.009	1.037 ± 0.010	0.997 ± 0.018 [c]	0.037
FN-BMD Follow up (g/cm^2)	0.849 ± 0.007	0.847 ± 0.007	0.823 ± 0.013	0.098
Premenopausal / HRT Follow-up	3.0% / 39.0%	3.5% / 33.4%	0% / 44.9%	0.038 / 0.003

BMD values are adjusted for age, BMI, menopausal status, and HRT use at baseline; and for age, BMI, menopausal status, HRT use smoking and physical activity score at follow-up. Values are numbers and percentages; least squares means ± SD for BMD and means ± SD for age and BMI. [a] $p<0.05$ from heterozygotes; [b] $p<0.01$ [c] $p<0.05$ from common homozygotes by Tukey's post-test .

inheritance, I found that homozygosity for the -1997T allele was an independent predictor of baseline lumbar spine BMD (p=0.002; 0.32% of the variance), along with BMI (p<0.001; 6.09% of the variance), age (p<0.001; 0.94% of the variance), menopausal status (p<0.001; 3.64% of the variance) and HRT use (p=0.003; 0.25% of the variance). Taken together, these variables explained 11.24% of the variance in lumbar spine BMD. When I performed a similar analysis for femoral neck BMD none of the individual genotypes was identified as an independent predictor of BMD (data not shown).

At the follow-up visit, there was a highly significant difference in the proportion of HRT users in relation to the -1663indelT (p=0.02) and Sp1 polymorphisms (p=0.003). For these polymorphisms, the proportion of HRT users in heterozygotes was less than in the other genotype groups and for the Sp1 polymorphism, there was an excess of HRT users among TT homozygotes. Genotype-specific associations with BMD at the follow-up visit for the -1997G/T polymorphism were similar to the baseline visit (p=0.003), except that the association with femoral neck BMD at follow-up was nominally significant (p=0.039). For the -1663indelT and Sp1 polymorphisms, the associations with BMD at follow-up were weaker than at baseline and nominal significance was observed only for the Sp1 polymorphism in relation to lumbar spine BMD (p=0.037). When I performed multiple regression analysis entering genotypes at all three polymorphisms into the regression model assuming a recessive mode of inheritance, I found that homozygosity for the -1997T allele was an independent predictor of follow-up lumbar spine BMD (p=0.001) accounting for 0.35% of the variance and of femoral neck BMD (p=0.027) accounting for 0.21% of the variance. Other non-genetic predictors of femoral neck and lumbar spine BMD at follow-up were similar to those at the baseline visit (data not shown).

I also analysed the relationship between individual polymorphisms and unadjusted BMD values at each visit by ANOVA and this yielded broadly similar results to the adjusted BMD values (data not shown).

Association between *COLIA1* 5' Haplotypes and BMD

Relevant characteristics of the APOSS population at baseline and follow-up in relation to the three commonest haplotypes are shown in Table 3.2. There was no significant association between any haplotype and age, BMI, menopausal status or HRT use at the baseline visit and there was no significant difference in adjusted lumbar spine BMD or femoral neck BMD values at baseline in relation to haplotype 1. There was a significant association between haplotype 2 and baseline BMD at the lumbar spine ($p=0.007$) and femoral neck ($p=0.008$) with reduced BMD values in homozygote carriers of haplotype 2. There was also a significant association between haplotype 3 and baseline BMD at the lumbar spine ($p=0.007$), but the association with baseline femoral neck BMD was not significant ($p=0.224$).

Multiple regression analysis showed that haplotypes 2 and 3 were significant predictors of lumbar spine BMD at baseline. Homozygosity for haplotype 2 accounted for 0.26% of the variance in lumbar spine BMD ($p=0.007$) and homozygosity for haplotype 3 accounted for 0.19% of the variance ($p=0.007$). Other predictors were BMI ($p<0.001$; 6.43% of the variance), menopausal status ($p<0.001$; 3.66% of the variance), age ($p<0.001$; 0.95% of the variance) and HRT use ($p=0.003$; 0.27% of the variance). Together these variables predicted 11.76% of the variance in baseline lumbar spine BMD. Homozygosity for haplotype 2 was also a significant predictor of baseline femoral neck BMD ($p=0.004$) accounting for 0.25% of the variance. Other predictors were BMI ($p<0.001$; 9.18% of the variance), age ($p<0.001$; 2.57% of the variance), menopausal status ($p<0.001$; 0.64% of the variance) and HRT use ($p=0.048$; 0.16% of the variance). Together these variables predicted 12.81% of the variance in baseline femoral neck BMD.

Table 3.2. Subject characteristics and bone mineral density measurements according to *COLIA1* haplotypes

Haplotype 1 (G-8-G)	**No copies**	**1 copy**	**2 copies**	**P-value (ANOVA)**
N	276 (9.5%)	1350 (46.7%)	1268 (43.7%)	(n=2890)
Age (yrs)	48.5 ± 2.5	48.6 ± 2.6	48.5 ± 2.5	0.651
BMI (Kg/m2)	25.0 ± 3.8	25.4 ± 4.6	25.4 ± 4.3	0.398
LS-BMD Baseline (g/cm^2)	1.052 ± 0.009	1.054 ± 0.005	1.058 ± 0.005	0.703
FN-BMD Baseline (g/cm^2)	0.879 ± 0.007	0.884 ± 0.004	0.884 ± 0.004	0.750
Premenopausal / HRT Baseline	50.5% / 20.4%	52.8% / 17.1%	50.8% / 18.1%	0.24 / 0.40
LS-BMD Follow-up (g/cm^2)	1.030 ± 0.014	1.042 ± 0.010	1.036 ± 0.010	0.508
FN-BMD Follow-up (g/cm^2)	0.833 ± 0.010	0.848 ± 0.007	0.844 ± 0.007	0.196
Premenopausal / HRT Follow-up	1.4% / 36.8%	3.3% / 36.9%	2.8% / 39.4%	0.136/0.391
Haplotype 2 (T-8-G)	**No copies**	**1 copy**	**2 copies**	
N	2180 (75.4%)	650 (22.4%)	60 (2.0%)	(n=2890)
Age (yrs)	48.5 ± 2.4	48.5 ± 2.4	48.5 ± 2.4	0.829
BMI (Kg/m2)	25.4 ± 4.5	25.3 ± 4.3	25.1 ± 4.1	0.846
LS-BMD Baseline (g/cm^2)	1.052 ± 0.004	1.063 ± 0.006	1.108 ± 0.019 [b]	0.007
FN-BMD Baseline (g/cm^2)	0.881 ± 0.003	0.889 ± 0.005	0.894 ± 0.015	0.224
Premenopausal / HRT Baseline	51.0% / 17.9%	54.1% / 17.1%	48.3% / 25.0%	0.34 / 0.30
LS-BMD Follow-up (g/cm^2)	1.034 ± 0.005	1.047 ± 0.008	1.103 ± 0.023 [a, b]	0.004
FN-BMD Follow-up (g/cm^2)	0.842 ± 0.007	0.851 ± 0.008	0.878 ± 0.017 [c]	0.030
Premenopausal / HRT Follow-up	3.0% / 37.5%	2.9% / 39.3%	0% / 40.7%	0.34 / 0.63
Haplotype 3 (G-7-T)	**No copies**	**1 copy**	**2 copies**	
N	2003 (69.3%)	786 (27.1%)	101 (3.4%)	(n=2890)
Age (yrs)	49.5 ± 2.4	48.7 ± 2.5	48.5 ± 2.4	0.162
BMI (Kg/m2)	25.3 ± 4.4	25.5 ± 4.6	25.0 ± 4.1	0.574
LS-BMD Baseline (g/cm^2)	1.060 ± 0.004	1.052 ± 0.006	1.014 ± 0.015 [a, b]	0.007
FN-BMD Baseline (g/cm^2)	0.887 ± 0.003	0.880 ± 0.004	0.852 ± 0.011 [a, b]	0.008
Premenopausal / HRT Baseline	52.3% / 17.5%	51.0% / 17.8%	48.3% / 23.3%	0.24 / 0.32
LS-BMD Follow-up (g/cm^2)	1.040 ± 0.005	1.036 ± 0.008	1.002 ± 0.023	0.100
FN-BMD Follow-up (g/cm^2)	0.847± 0.007	0.841 ± 0.008	0.816 ± 0.014 [c]	0.033
Premenopausal / HRT Follow-up	2.8% / 39.6%	3.1% / 33.0%	3.4% / 45.1%	0.20 / 0.002

BMD values are adjusted for age, BMI, menopausal status, and HRT use at baseline; and for age, BMI, menopausal status, HRT use smoking and physical activity score at follow-up. Values are numbers and percentages; least squares means ± SD for BMD and means ± SD for age and BMI. [a] $p<0.05$ from heterozygotes; [b] $p<0.01$ [c] $p<0.05$ from common homozygotes by Tukey's post-test.

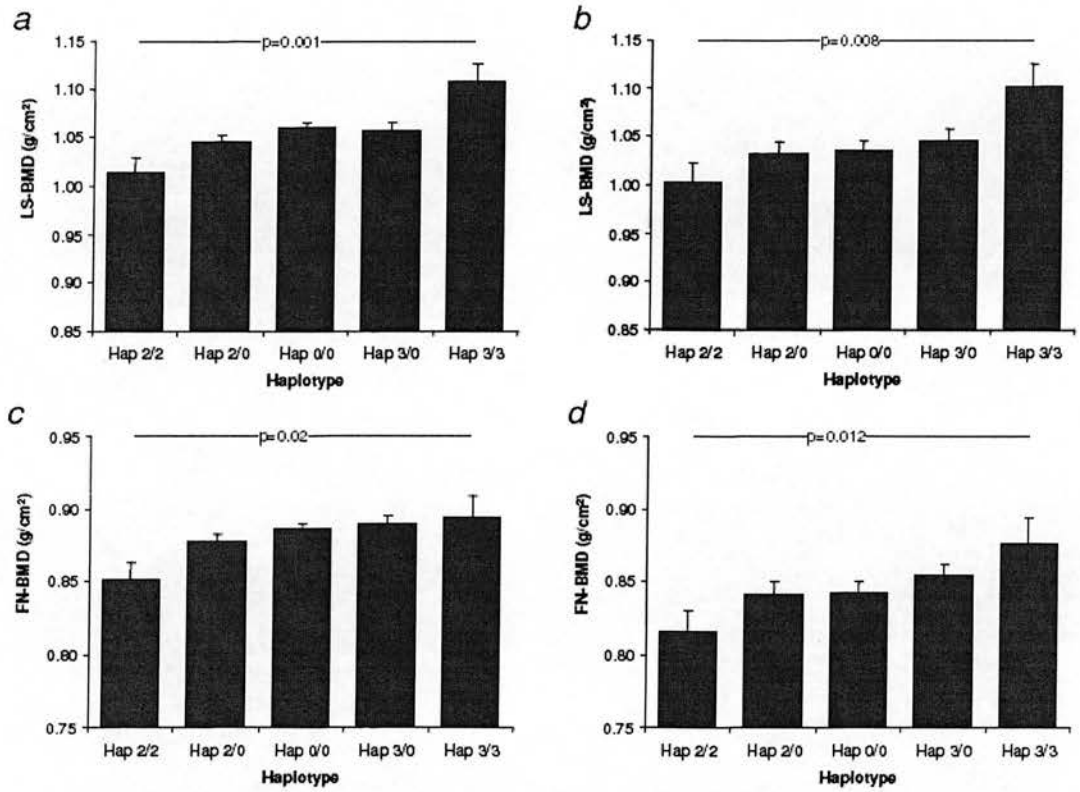
At the follow-up visit, there was no difference in age, BMI, smoking, or physical activity score in relation to haplotype (data not shown), but there was a significant difference in the proportion of HRT users in relation to carriers of haplotype 2, with a reduced number of HRT users in heterozygotes and an increased number of HRT users in haplotype 2 homozygotes (p=0.002). There was no significant difference in adjusted lumbar spine BMD or femoral neck BMD values at follow-up in relation to haplotype 1. There was no significant association between haplotype 2 and follow-up BMD at the lumbar spine, contrasting with the significant association at baseline. Moreover, the association between haplotype 2 and femoral neck BMD at follow-up was weaker than at baseline (p=0.033). For haplotype 3, the association with lumbar spine BMD at follow-up remained significant (p=0.004) with a similar effect size as at the baseline visit. Haplotype 3 was associated with femoral neck BMD at follow-up with a nominally significant pvalue of p=0.030.

Multiple regression analysis showed that haplotype 3 was a significant predictor of lumbar spine BMD at follow-up. Homozygosity for haplotype 3 accounted for 0.50% of the variance in BMD (p=0.004). Other predictors were BMI (p<0.001; 7.49% of the variance), age (p<0.001; 2.563% of the variance), menopausal status (p<0.001; 1.55% of the variance), and HRT use (p<0.001; 1.1% of the variance). Together these variables predicted 13.2% of the variance in follow-up lumbar spine BMD. For follow-up femoral neck BMD, multiple regression analysis showed that homozygosity for haplotype 2 was a significant predictor of BMD (p=0.021) accounting for 0.21% of the variance. Haplotype 3 was also a significant predictor of follow-up femoral neck BMD (p=0.05) accounting for 0.16% of the variance. Other predictors were BMI (p<0.001; 8.47% of the variance),

age ($p<0.001$; 2.12% of the variance), and menopausal status ($p<0.023$; 0.18% of the variance). Together these variables predicted 10.93% of the variance in follow-up femoral neck BMD. Since haplotype 2 and haplotype 3 had opposing effects on BMD, I stratified participants according to combinations of these two haplotypes. This showed a striking difference in BMD between the extreme genotypes at both baseline and follow-up as illustrated in Figure 3.2. Subjects who were homozygous for haplotype 3 had the highest BMD and homozygous for haplotype 2 had the lowest BMD. The difference between haplotypes was greatest for lumbar spine BMD at baseline ($p=0.001$) and follow-up ($p=0.008$), but was also evident and statistically significant for femoral neck BMD at baseline ($p=0.02$) and follow-up ($p=0.012$). The difference in BMD between the extreme haplotypes (homozygotes for haplotype 2, versus homozygotes for haplotype 3) was 0.094 g/cm^2 at the lumbar spine which is equivalent to about 0.58 BMD Z-score units and 0.042 g/cm^2 at the femoral neck which is equivalent to about 0.33 BMD Z-score units. At follow up, the difference in BMD between the extreme genotypes 2/2 and 3/3 was 0.010 g/cm^2 at the lumbar spine which is equivalent to about 0.81 BMD Z-score and 0.061g/cm^2 at femoral neck which is equivalent to about 0.50 BMD Z-score units.

I also analysed the relationship between individual haplotypes and unadjusted BMD at each visit by ANOVA and this yielded broadly similar results to those obtained with the adjusted BMD values (data not shown).

Figure 3.2 Association between *COLIA1* 5' haplotypes and BMD



The BMD values shown are adjusted for confounding factors by ANOVA-GLM in relation to carriage of haplotype 2 and 3 in the 5' flank of the *COLIA1* gene. Panels a and b show the results for lumbar spine BMD at baseline and follow up respectively, whereas panels c and d show the results for femoral neck BMD at baseline and follow-up respectively. Columns are least square means, and bars are standard deviation. 2/2 indicates homozygotes for haplotype 2; 2/0 indicates heterozygotes for haplotype 2 who were not carriers of haplotype 3; 3/0 indicates heterozygotes for haplotype 3 who were not carriers of haplotype 2; 3/3 indicates homozygotes for haplotype 3; 0/0 indicates subject who were not carriers of either haplotype 2 or haplotype 3.

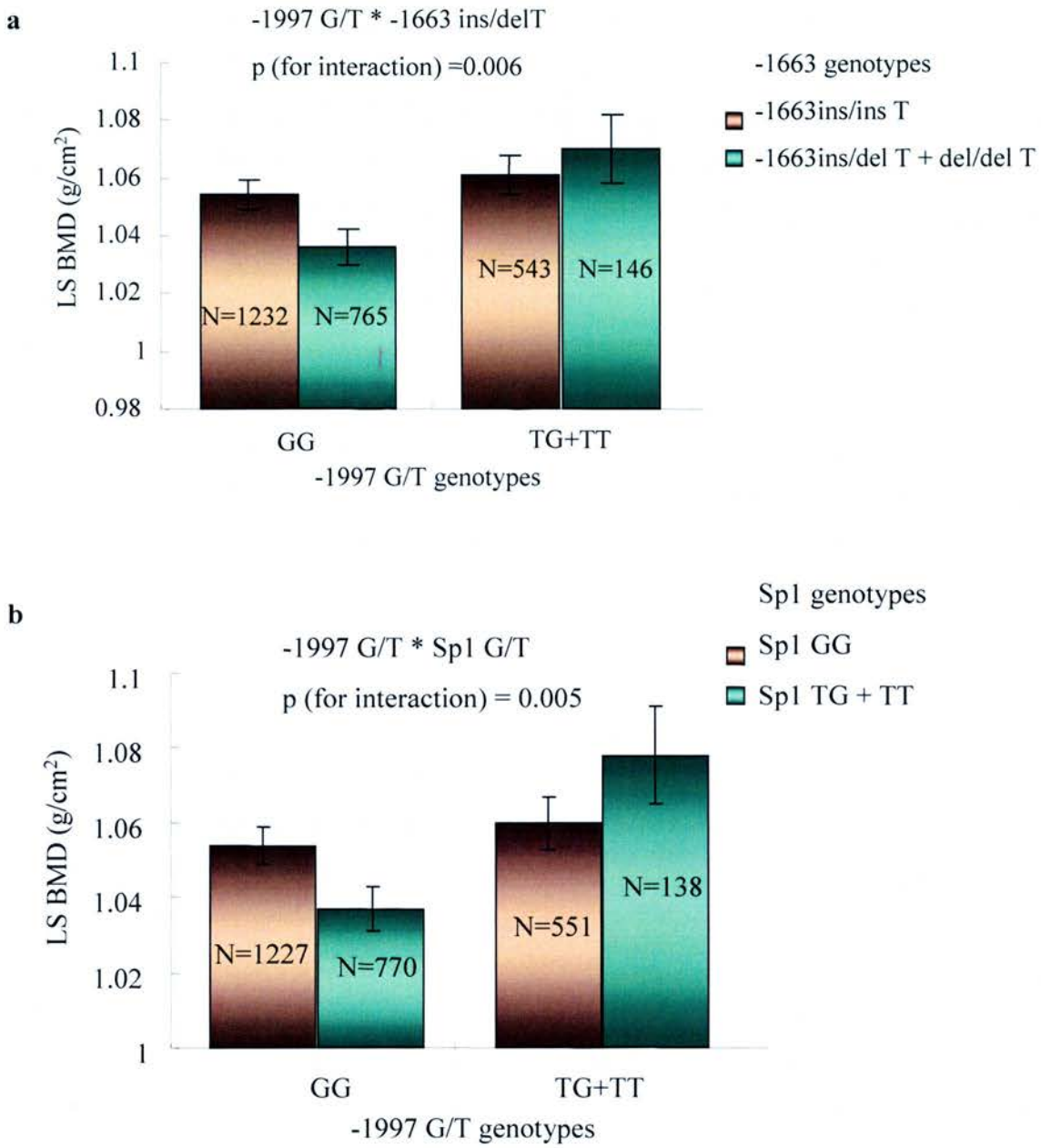
Genotype and Haplotype Associations with Bone Loss

There was no significant association between genotype or haplotype and bone loss at either the lumber spine or femoral neck in the APOSS population. Multiple regression analysis showed that the major predictors of lumbar spine bone loss were HRT use ($p<0.001$; 14.59% of the variance); menopausal status ($p<0.001$; 4.48% of the variance); BMI ($p<0.001$; 3.04% of the variance); and age ($p=0.01$; 0.19% of the variance). Together these variables accounted for 22.3% of the variance in bone loss. Similar analysis for femoral neck bone loss showed that the major predictors were HRT use ($p<0.001$; 7.85% of the variance); menopausal status ($p<0.001$; 2.17% of the variance); BMI ($p=0.01$; 0.4% of the variance). Together these variables accounted for 10.4% of the variance in femoral neck bone loss.

Pair-wise Interactions between Polymorphisms and BMD

Pair-wise interactions among the three polymorphisms on BMD were tested in this population. The method was developed by Bustamante M *et al* (2007). The interactions between polymorphisms were tested against BMD at lumbar spine and femoral neck by using program SPSS. The significant level is p value lower than 0.05. There were two interactions with significant *p* values are showed in Figure 3.3. The heterozygote and rare homozygote of -1663ins/delT can be divided into two groups with extreme mean BMD values according to the genotype at -1997. A similar interaction was observed in Sp1 genotypes. Significant associations between the interaction between -1997G/T and -1663ins/delT polymorphisms with lumbar spine BMD were observed at baseline (p=0.006) and follow-up (p=0.032) after adjusting for age, BMI, menopausal status and HRT use. This association was also observed with femoral neck BMD with a less significant level at only follow-up (p=0.05, data not shown). Significant associations for the interaction between -1997G/T and Sp1 G/T polymorphisms with lumbar spine BMD was detected at both baseline (p=0.005) and follow-up (p=0.02) after adjusting for age, BMI, menopausal status and HRT use. This association was also observed with femoral neck BMD with a less significant level only at follow-up level (p=0.046, data not shown). However, no association between the interaction between Sp1 G/T and -1663ins/delT polymorphisms with either lumbar spine BMD or femoral neck BMD was observed in this population.

Figure 3.3 Significant associations between interactions of polymorphisms and BMD



a. Interaction between the -1997G/T and -1663ins/delT polymorphisms is associated with adjusted lumbar spine BMD at baseline (p=0.006). **b**. Interaction between the -1997G/T and Sp1G/T polymorphisms is associated with adjusted lumbar spine BMD at baseline (p=0.005).
BMD values are adjusted for age, BMI, menopausal status, and HRT use. BMD values are showed as mean ± SEM. Genotypes are indicated either below the figure or by coloured code (right). The number of individuals per group (n) and the significance (p) are also shown.

Discussion

The *COLIA1* gene is a strong functional candidate for the genetic regulation of bone mass and susceptibility to fragility fractures, since it encodes the alpha 1 chain of type I collagen, the most abundant protein in bone. Most previous studies of *COLIA1* alleles in relation to osteoporosisrelated phenotypes have focused on the intronic Sp1 binding site polymorphism (Grant SFA *et al*, 1996). The promoter polymorphisms that I studied here were originally identified by Garcia-Giralt and colleagues on mutation screening of the proximal *COLIA1* promoter in 265 postmenopausal Spanish women (García-Giralt N *et al*, 2002). In that study, the -1997G/T polymorphism was the one most strongly associated with BMD and lumbar spine BMD values were reduced in TT homozygotes compared with the other genotype groups (García-Giralt N *et al*, 2002). Whilst the -1663InsT/DelT polymorphism was not significantly associated with BMD in Spanish women, regression analysis showed that alleles at the -1663 and Sp1 site interacted with alleles at the -1997 site to regulate lumbar spine BMD. The relation between the -1997G/T polymorphism and BMD has also been investigated in other cohorts. In a study of 308 postmenopausal Caucasian women from the USA, haplotypes defined by the -1997 and Sp1 polymorphisms were found to be associated with BMD such that higher values were found in carriers of the G allele at both polymorphic sites (Liu PY *et al,* 2004). Another family based study in Chinese subjects showed evidence of an association between the -1997G/T polymorphism and hip BMD using a quantitative TDT based approach (Zhang YY *et al*, 2005), whereas a population based study of unrelated Chinese women showed no association between -1997G/T alleles and BMD (Lau HH *et al*, 2005). A further study of 1100 Japanese postmenopausal women, reported a significant

association between -1997G/T alleles and hip BMD with lower BMD values in GG homozygotes (Yamada Y *et al*, 2005).

The present study of promoter and intron 1 polymorphisms at the *COLIA1* locus is the largest and most comprehensive to be conducted so far and is the first study to examine the role of all three polymorphisms and haplotypes in relation to BMD. In agreement with the previous study of Spanish women (García-Giralt N *et al*, 2002), I found that all three polymorphisms were in strong linkage disequilibrium and three common haplotypes accounted for more than 95% of alleles at the 5' flank of the *COLIA1* locus. These were; haplotype 1 (-1997G / -1666InsT / +1245G); haplotype 2 (1997G / -1663delT / +1245T) and haplotype 3 (-1997T/ -1663InsT / +1245G). Whilst the individual polymorphisms were associated with BMD to an extent, I observed stronger and more consistent associations with BMD in relation to the carriage of the *COLIA1* haplotypes and also observed significant differences in the strength of association by skeletal site and the time point at which participants were studied. For the -1997G/T polymorphism, homozygotes for the T allele had increased lumbar spine BMD at baseline and follow-up and had increased femoral neck BMD values at follow-up. These observations are in broad agreement with the findings reported by Yamada and colleagues, who also reported higher BMD values in relation to the -1997 T allele in Japanese postmenopausal women (Yamada Y *et al*, 2005). Our findings differ from the results reported in Spanish and US postmenopausal women however, where the -1997T allele was associated with reduced BMD (García-Giralt N *et al*, 2002 & Liu PY *et al,* 2004). The differences between these studies could be due to the fact that the polymorphisms studied here are in linkage disequilibrium with other functional polymorphisms at the *COLIA1* locus that regulate

BMD and that the patterns of LD differ in these different populations. Another more likely explanation is due to a more accurate estimate of effect size in the Yamada study (Yamada Y *et al*, 2005), and in this study, which respectively, were about 3 times and 10 times larger than the studies which reported an association between the -1997T allele and low BMD (García-Giralt N *et al*, 2002 & Liu PY *et al,* 2004).

For the -1663indelT and Sp1 polymorphism I observed a strong and highly significant association with lumbar spine BMD values at baseline, even after correction for multiple testing, but these associations were weaker at follow-up. The difference that I observed between baseline and follow up may have been influenced by differential use of HRT in the different genotype groups. This was most marked and statistically significant for the -1663indelT and Sp1 polymorphisms, where there was a highly significant difference across genotype groups in terms of HRT use at follow up, compared with no difference at baseline. This difference in HRT use may be accounted for by the fact that women in this cohort who had BMD values in the lowest quartile at baseline were advised to take HRT in order to protect from osteoporosis (Torgerson DJ *et al*, 1997). Since BMD values were significantly lower in homozygotes for the rare allele at the -1663indelT and Sp1 sites at baseline, this could have accounted for higher rates of HRT uptake at follow-up in these genotype groups.

The association between haplotypes and BMD were stronger and more consistent across skeletal sites than the associations with individual genotypes. Haplotypes 2 and 3 were both strongly associated with lumbar spine BMD at baseline and haplotype 2 was also strongly associated with femoral neck BMD at baseline. At follow-up, the association between haplotype 3 and lumbar spine BMD was similar to the baseline visit, whereas the

association with haplotype 2 was weaker than at baseline and was not statistically significant. These differences were most likely due to differences in uptake of HRT in carriers of haplotype 2, reflected by the significant difference in the proportion of HRT users in this haplotype group at the follow-up visit as compared with baseline. As was the case with the individual polymorphisms, I speculate that this difference may be increased use of HRT in homozygote carriers of haplotype 2.

The p-values for significance reported here should be interpreted in light of the fact that I studied associations between three polymorphisms and associated haplotypes in relation to BMD at two sites, measured on two occasions. However taking into account the LD between polymorphisms and the fact that the traits were correlated, many of the associations I observed remained below the adjusted threshold for significance (p=0.014) using a modified Bonferroni test to correct for the number of independent tests performed. The haplotype associations that I report here are of interest in relation to a recent study using promoter-reporter assays where it was shown that promoter constructs containing the -1997GDelT allele (corresponding to haplotype 2) had higher rates of transcription than the -1997T - 1663InsT allele (corresponding to haplotype 3) (García-Giralt N *et al,* 2005). The effects of the *COLIA1* Sp1 polymorphism on transcription were not investigated in the reporter studies mentioned above. These observations suggest that there is an association between high levels of *COLIA1* transcription and reduced BMD values, consistent with previous work which has suggested that the Sp1 T allele is associated with increased *COLIA1* transcription leading to an imbalance in the ratio of the collagen type I alpha 1 and collagen type I alpha 2 chains resulting in a subtle impairment of bone mineralization (Stewart TL *et al,* 2005). The present observations indicate that

the previously reported association between the *COLIA1* Sp1 "T" allele and increased allele specific transcription (Mann V *et al*, 2001) may actually be driven by an extended haplotype driven by both the intron 1 and promoter polymorphisms.

I also found significant associations between interactions of -1997G/T and -1663ins/delT (or Sp1 G/T) with BMD. The heterozygote and rare homozygote of -1663ins/delT (or Sp1 G/T) could be divided into two groups with extreme mean BMD values at lumbar spine and femoral neck, according to the genotype at -1997. The highest mean BMD value appeared in the group with -1997G/G homozygote and the heterozygote & rare homozygote of -1663ins/delT (or Sp1 G/T), whereas the lowest mean BMD value appeared in the group with -1997G/T & T/T and the heterozygote & rare homozygote of -1663ins/delT (or Sp1 G/T). This is the same as the results reported by García-Giralt N *et al,* (2002). The association between -1997G/T and Sp1 polymorphisms with lumbar spine BMD was also observed in a larger Spanish postmenopausal female population (p=0.01, N=719) by Bustamante M *et al*, (2007).

One of the limitations of this study is that genotypes for the -1663in/delT polymorphism deviated significantly from Hardy-Weinberg equilibrium. In view of this I carefully scrutinized sequencing traces from all individuals who were heterozygous for -1663indelT and all -1663delT homozygotes, but no errors were identified, thereby excluding genotyping error as a cause of the HWE deviation. Possible explanations include a greater proportion of failed reactions in the -1663indelT heterozygotes (Wittke-Thompson JK *et al*, 2005), since this polymorphism was technically the most difficult to

genotype and had the greatest number of dropouts; or the fact that the -1663indelT polymorphism may actually have been a composite of two polymorphisms within the poly-T tract on different haplotype backgrounds if the poly-T tract is viewed as "copy number variant". This has previously been documented as a cause of HWE deviation at other loci (Nancy Cox, personal communication) but I was unable to investigate this possibility in the present population since the participants were unrelated Another possible explaination is the population stratification, although it's not likely in this study since APOSS population were tested in other association studies and no HWE deviation was found. Therefore, the most possible explanation is the genotyping failure in this study. In future new methodologies are needed to investigate this polymorphisms, such as denaturing high-performance liquid chromatography (DHPLC).

It is also worth notice that the associations observed in this study might be caused by other polymorphisms which are in linkage disequilibrium with these 5' polymorphisms. I have checked the Hapmap database, however, no enough data suggests the existence of other candidate polymorphism.

In summary, this study demonstrates that haplotypes, defined by the promoter and intron 1 polymorphisms of the *COLIA1* gene, regulate BMD in postmenopausal women. One of the most interesting observations to emerge from the study was the much higher uptake of HRT in women who carried genotypes and haplotypes that were associated with low BMD at the baseline assessment. This example of a gene-environment interaction illustrates the challenges faced by investigators who seek to detect genetic variants that regulate BMD in populations such as this, where systematic BMD screening has been

followed by a specific therapeutic intervention. Whilst the mechanism by which *COLIA1* variants regulate BMD remains to be fully elucidated, current evidence is consistent with an effect of these different haplotypes on *COLIA1* transcription.

CHAPTER FOUR

Functional Analysis of 5'Polymorphisms by Gel Shift Assays and Reporter Assay

Abstract

The previous chapter showed that 5' flank *COL1A1* polymorphisms regulate BMD in a population of perimenopausal women and showed evidence of an interaction between the polymorphisms in regulating BMD. The underlying mechanisms of these associations are unclear. Previous studies showed that the +1245 Sp1 T allele had higher binding affinity for the transcription factor Sp1 than the G allele. A recent report also found that the -1663in/delT polymorphism is located within a binding site of Nmp4 proteins and that the -1663inT (8T) allele had greater binding affinity than the -1663delT (7T) allele. This study also showed that haplotypes defined by the promoter polymorphisms regulate reporter gene transcription *in vitro*.

In this chapter non-radioactive electrophoretic mobility shift assays (EMSAs) and super-shift assays were used to investigate the DNA-protein binding affinities of the regions surrounding both the -1997G/T and -1663indelT promoter polymorphisms. The results confirmed the binding of Nmp4 proteins and Osterix to the region surrounding the -1663in/delT polymorphism. The -1663delT allele (7T) consistently bound nuclear proteins with greater affinity than the -1663inT (8T) allele. The region surrounding the -1997G/T polymorphism contained a Sp1 binding site and the T allele had increased binding affinity than the G allele. However, supershift assays did not show evidence of Sp1 protein binding to DNA in osteoblast nuclear extracts.

I also studied the effect of the three polymorphisms on reporter gene transcription in osteoblast-like cells. Eight possible haplotypes defined by the three polymorphisms comprising 2300 bp upstream of the transcriptional start site to the beginning of exon 2 were generated and inserted into a modified pGL-3 basic vector. This resulted in a construct comprising the promoter, exon1 and intron1 of *COL1A1*, plus a small part of exon2, which drives expression of a fusion protein comprising 55 amino acids of *COL1A1* and *Firefly* luciferase. These reporter constructs were then co-transfected with the *Renilla* pRL-TK vector into HOS TE85 cells. For each sample, the *Firefly* luciferase activity was normalized by the *Renilla* luciferase activity and then compared with the mean value of the pGL3-basic vector. All reporter vectors had significantly higher luciferase activities than the pGL3-basic vector which lacks eukaryotic promoter and enhancer sequences ($p<0.001$). The pGL3-G7T vector had the highest gene transcription and differed from all the other vectors ($p<0.001$). The luciferase activity of the pGL3-T8G vector was the second highest and also

significantly differed from all the other vectors. The pGL3-T7G vector had the lowest luciferase activity. The luciferase activities of different haplotypes were not clearly related to the effects of any individual polymorphism. Therefore, these results suggested that the *COL1A1* 5' flank polymorphisms interact with each other to regulate reporter gene transcription.

Introduction

In the previous chapter, I showed that the 5' *COL1A1* polymorphisms were associated with BMD in a population based study. Previous studies have suggested that the +1245Sp1 polymorphism and other promoter polymorphisms may predispose to osteoporosis by affecting gene transcription. For example, Mann V *et al* reported that the Sp1 T allele within intron1 had significantly increased binding affinity for the transcription factor Sp1 compared with the G allele, and primary RNA transcripts derived from the Sp1T allele were found to be approximately three times more abundant than those from the G allele in heterozygotes ($p<0.05$, Mann V *et al*, 2001). A recent functional study showed that haplotypes defined by the two promoter polymorphisms regulated reporter gene transcription *in vitro* in osteoblast-like cells, and the -1663insT allele had increased binding affinity for nuclear proteins in osteoblasts. One of these proteins was identified by supershift assay as the nuclear matrix protein Nmp4 (also called cas interacting zinc finger protein, CIZ), a known inhibitor of BMP/Smad signalling (García-Giralt N *et al* 2005). Analysis of the sequence around the -1663insdelT polymorphism also revealed a potential Osterix binding site located upstream of the -1663indelT polymorphism in the human *COL1A1* promoter. This makes Osterix a strong candidate for DNA-protein binding interactions in the vicinity of -1663insdelT polymorphism. The region surrounding -1997G/T polymorphism is also predicted to contain nuclear protein binding sites such as a Sp1 binding site, but the nuclear proteins that bind to the -1997G/T region are still unclear.

In this chapter non-radioactive electrophoretic mobility shift assays (EMSAs) and super-shift assays were used to confirm the -1663indelT /Nmp4 binding and determine whether two promoter polymorphisms affected binding affinity for nuclear

proteins in osteoblasts. Reporter assays were used to investigate the transcriptional effect of the three polymorphisms on reporter gene expression in osteoblasts.

Methods

The binding affinities of nuclear proteins derived from human osteoblast cell lines (MG63 & HOS TE85) to the regions surrounding the two promoter polymorphisms were analysed by non-radioactive electrophoretic mobility shift assays (EMSAs) using the methods described in chapter 2.

The online program Alibaba 2.1 was used to predict the possible binding sites in all the oligonucleartides which flank the two promoter polymorphisms (http://www.gene-regulation.com/pub/programs/alibaba2/index.html). This type of program is based on the potential binding sequences so it often over-predicts. Only the transcription factors related to bone biology were analysed in this study.

The reporter vector was constructed by ligation of a PCR-amplified fragment of human DNA sequence containing the promoter, exon1, intron1 and part of exon2 (-2316 to +1723) of *COL1A1* (haplotype 1) into a modified pGL3-basic vector. The other seven possible haplotypes were introduced into reporter vectors by Site-Directed Mutagenesis (SDM) reactions. All the vectors were prepared by using the PureLink™ HiPure Plasmid DNA Purification Maxi prep Kit (Invitrogen). The concentration of each DNA was determined by Quant-iT™ PicoGreen® dsDNA Reagent (Invitrogen). The same amount of each reporter vector and two control vectors (pGL3-basic & pGL3-control) were co-transfected into the osteoblast-like cell line HOS TE85 cells with the pRL-TK vector by the Magnet Assisted Transfection (MATra) kit. Luciferase expression was measured by plate-reader (*Synergy HT*, Bio-Tek). The results were analysed by GLM-ANOVA (Minitab). All experiments were repeated four times under similar conditions.

To check the existence of the fusion protein, total cell RNA of the pGL-3 basic vector, pGL-3 T8G and pGL-3 G7T vectors was extracted and used as template to make

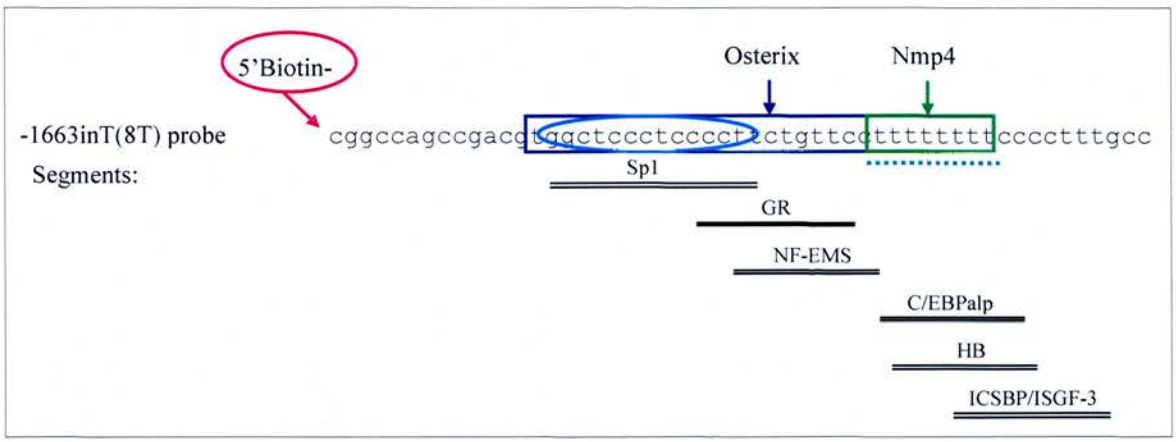
cDNA using reverse transcriptase twenty-four hours after the transfection. A 242bp fragment was then amplified by PCR with a forward primer in *COL1A1* exon 1 and a reverse primer in luciferase gene. The products were confirmed by sequencing reaction. Full details of the methodology are provided in chapter 2 *Material and methods*.

Results

Nmp4 and Osterix Bind to the Region Surrounding the -1663ins/delT Polymorphism

Several proteins were predicted to bind the DNA surrounding -1663ins/delT polymorphism as shown in Figure 4.1.

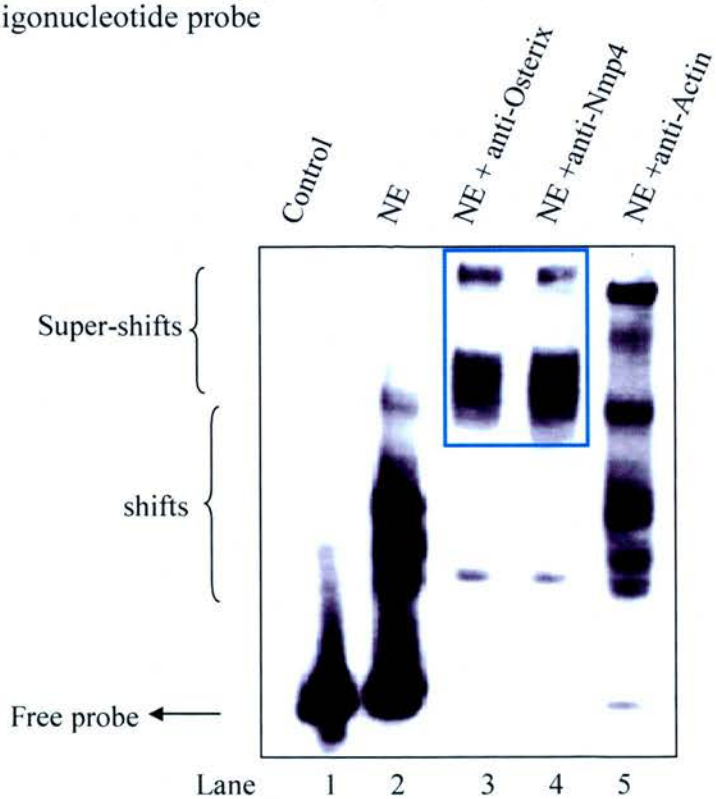
Figure 4.1 DNA-protein sequence of -1663insT probe used in EMSA and the potential binding motifs



The oligonucleotide probe used in EMSA was 5'-end labeled with biotin. Six potential binding motifs were predicted by online program Alibaba 2.1. Three proteins were investigated in this study (framed) were Sp1, Osterix and Nmp4. The other five binding sites for nuclear proteins (GR, NF-EMS, C/EBMPalp, HB, ICSBP & ISGF-3) were not investigated in this study. The -1663insT allele is indicated by dashed line.
GR: Glucocorticoid Receptor; NF-EMS: also called IRF-4, Interferon-regulatory factor 4; C/EBPalpha: CCAAT Enhancer Binding Protein alpha; HB: Hunchback transcription factor; ICSBP: Interferon Consensus sequence binding protein; ISGF-3: Transcription factor ISGF-3

When nuclear extracts (from MG63 & HOS TE85) were used as the protein source in EMSAs which contained biotin-labelled -1663ins/delT oligonucleotide probes, the DNA-protein complexes were seen as several distinct bands (shifts) on the gel (Figure 4.2; lane 2). In this study only Nmp4, Sp1 and Osterix were investigated due to their biological relation to bone metabolism. To confirm specific binding of Nmp4 and Osterix to the oligonucleotide probe, antibodies to Nmp4 and Osterix were added to the EMSA binding reaction resulting in super-shifts, which had an identical binding pattern (blue box, Figure 4.2). Addition of an antibody to actin showed a different pattern of binding. Although the intensity of the bands decreased, a similar pattern of bands was observed as compared with nuclear extract alone apart from a very high molecular weight band (lane 5, Figure 4.2).

Figure 4.2 EMSA demonstrating binding of Nmp4 and Osterix to the biotin-labelled -1663delT oligonucleotide probe



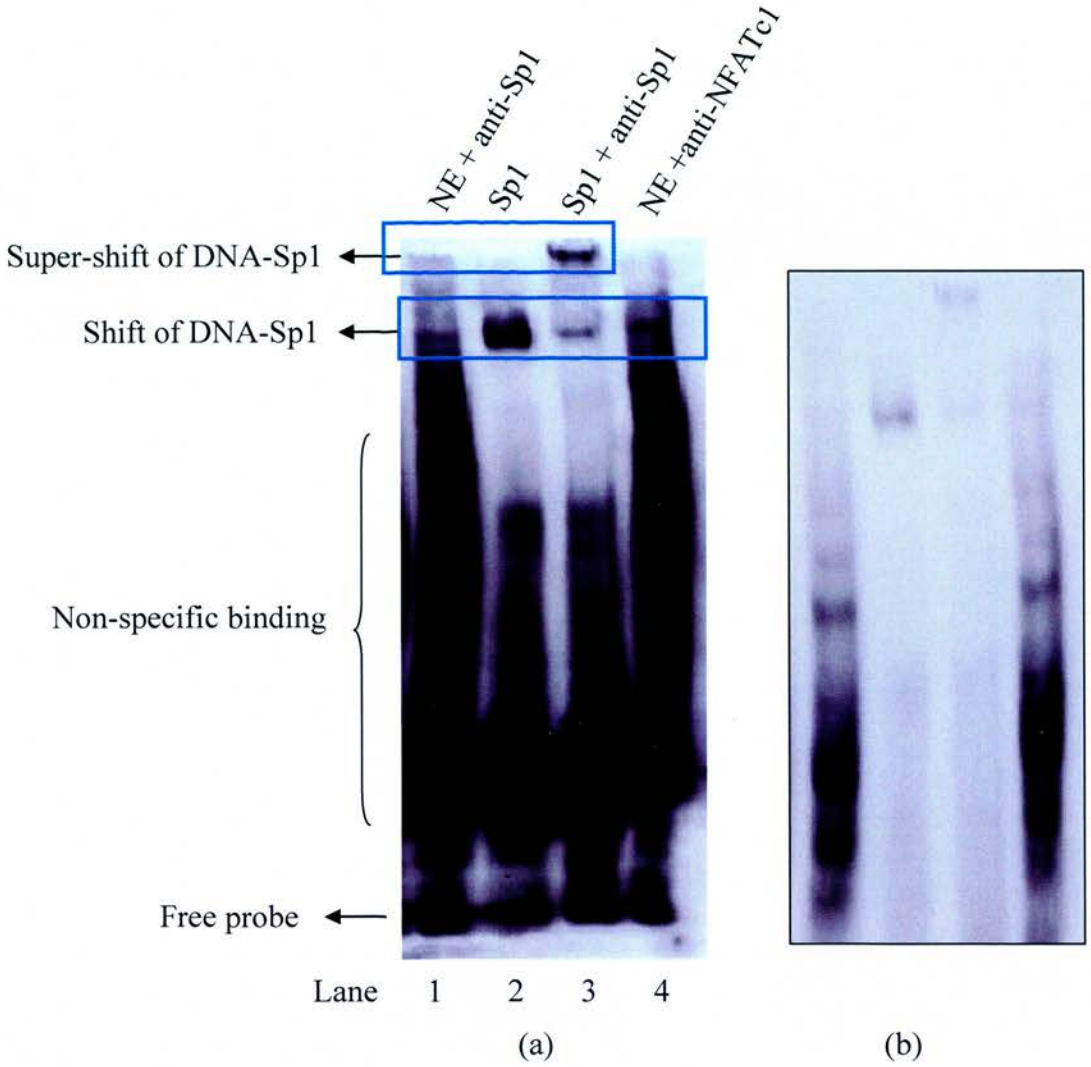
EMSA was performed on a 6% (w/v) polyacrylamide gel with biotin-labelled -1663delT (7T) probe (all lanes) and HOS TE85 cell nuclear extract (NE, lane 2-5). DNA-protein complexes were seen as shifts. Addition of antibodies for Nmp4 and Osterix further shifted the DNA-protein complexes (blue box, lane 3 & 4). Addition

of antibody for actin caused similar shifts as the NE (lane 2) apart from a very high molecular weight band (lane 5).

Sp1 Binds to the Region Surrounding -1663ins/delT Polymorphism

Since the DNA surrounding -1663insdelT polymorphism is predicted to contain an Sp1 binding site, I also investigated the effect of the SNP on Sp1 protein binding. When Sp1 protein was used in the binding reaction which contained a biotin-labeled -1663inT (8T) oligonucleotide probe, a single shift was seen on the gel (lane 2, figure 4.3). Addition of Sp1 antibody together with Sp1 protein caused a super-shift (lane 3). When Sp1 antibody was mixed with nuclear extract, the super-shift was very faint but still visible on gel (lane 1). EMSA with biotin-labeled -1663delT (7T) probe gave the same results (data not shown).

Figure 4.3 EMSA demonstrating binding of Sp1 to the biotin-labelled -1663insT oligonucleotide probe



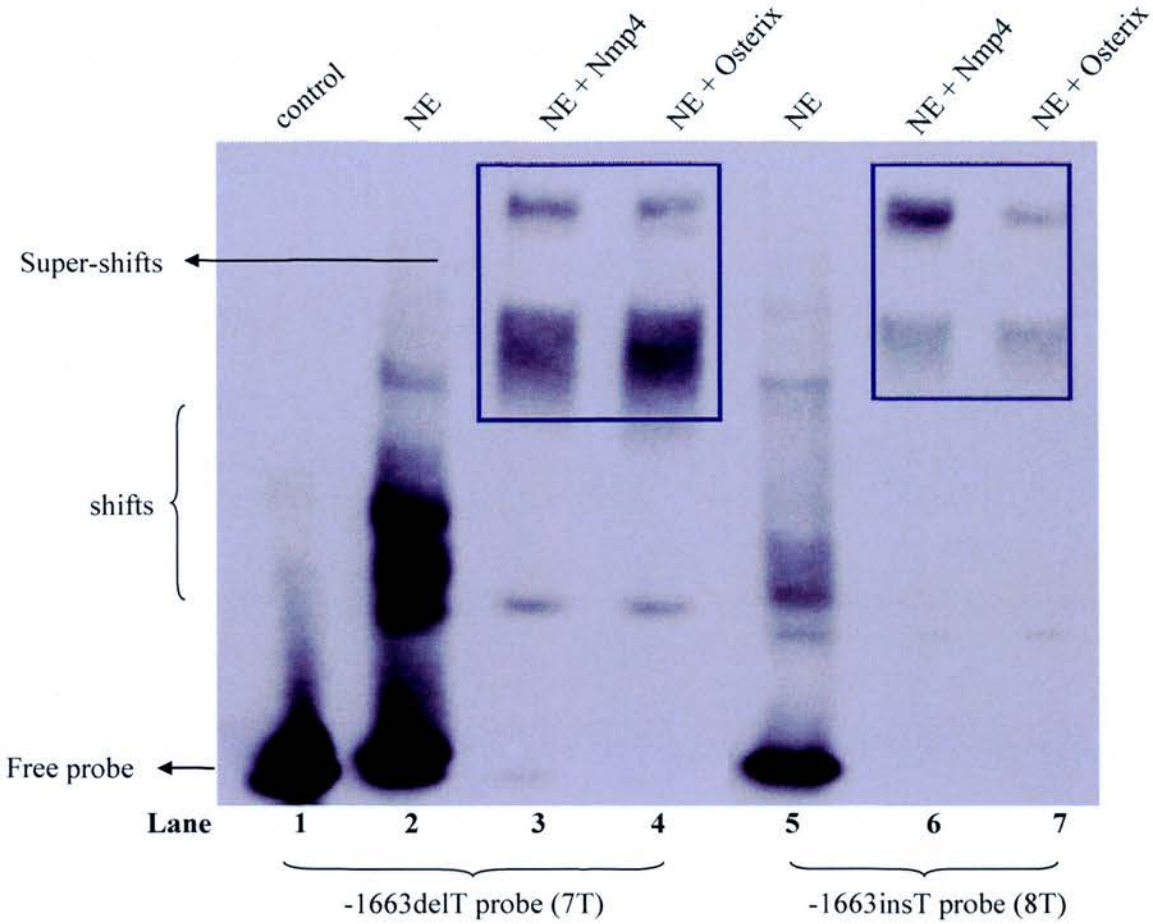
EMSA was performed with biotin-labeled -1663insT oligonucleotide probe (all lanes). The left panel (a) shows that when Sp1 protein was used as sole protein source, DNA-Sp1 binding complex was seen as shift (lane 2) and super-shift (lane 3) with addition of antibody to Sp1. When HOS TE85 nuclear extra was used as protein source, a very faint super-shift was seen on the gel (lane 1). Addition of antibody to NFATc1 to the nuclear extra did not alter the DNA-protein binding pattern (lane 4).
The right panle (b) is the same image but with a less intensity.

-1663delT (7T) Allele Has Greater Affinity for Nuclear Proteins than -1663insT (8T) Allele

The '7T' oligonucleotide consistently bound nuclear proteins with greater affinity than the '8T' oligonucleotide when adding the same amount of nuclear protein into

the binding reactions with the same amount of probes. This is illustrated on Figure 4.4: the bands had greater intensity with the biotin-labeled '7T' probe (lane 1-4) compared to those with the biotin-labeled '8T' probe (lane 5-7).

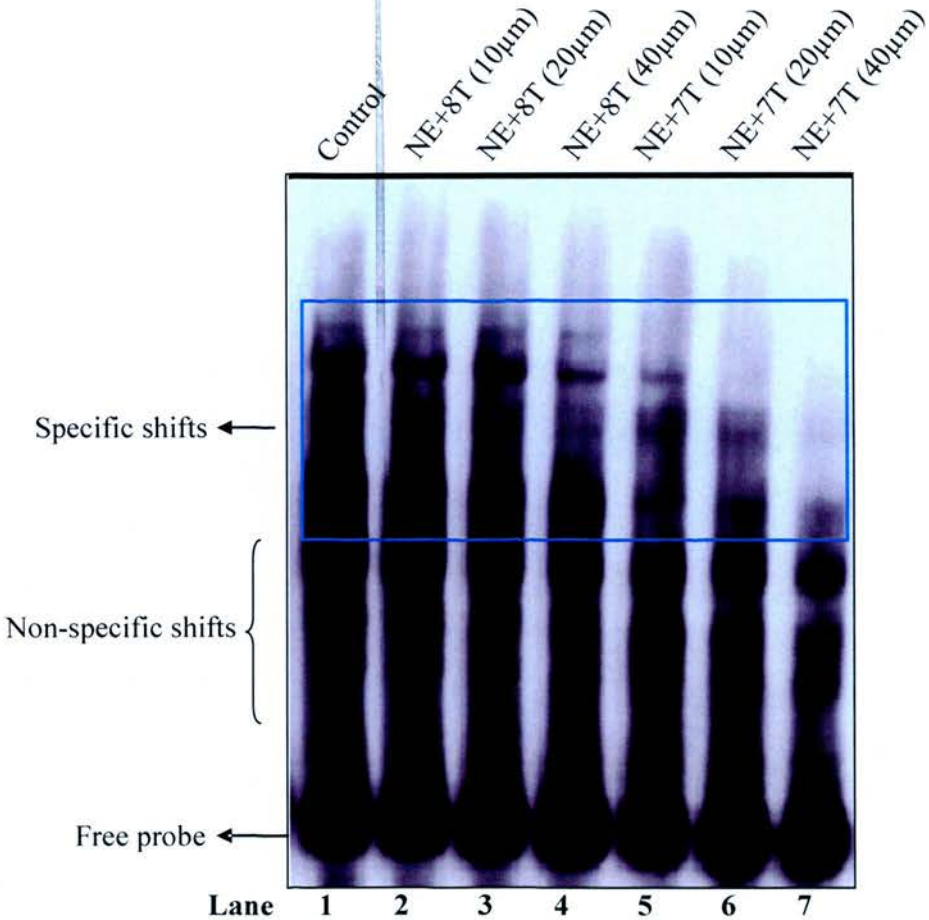
Figure 4.4.



EMSA was performed with biotin-labeled -1663delT (7T) oligonucleotide probe (lane 1-4) and biotin-labeled -1663inT (8T) oligonucleotide probe (lane 5-7). Both probe generated the same binding pattern. Upon incubation with nuclear proteins, the 7T oligonucleotide consistently produced retention bands of higher intensity than the 8T oligonucleotide, suggestive of a higher binding affinity for some of them.
NE, nuclear extracts. Experiments were repeated in triplicate under similar conditions.

To compare the difference in binding affinity of the polymorphic variants for nuclear proteins, competition assays were performed. The biotin-labeled '7T' oligonucleotide probe was incubated with nuclear extracts from HOS TE85 cells and with varying concentrations of either '7T' or '8T' unlabeled competitor oligonucleotide probes. The unlabeled '7T' probe bound nuclear proteins with greater affinity than did the unlabeled '8T' competitor resulting in weaker intensity of DNA-protein bands with '7T' competitor for each competitor concentrations (Figure 4.5).

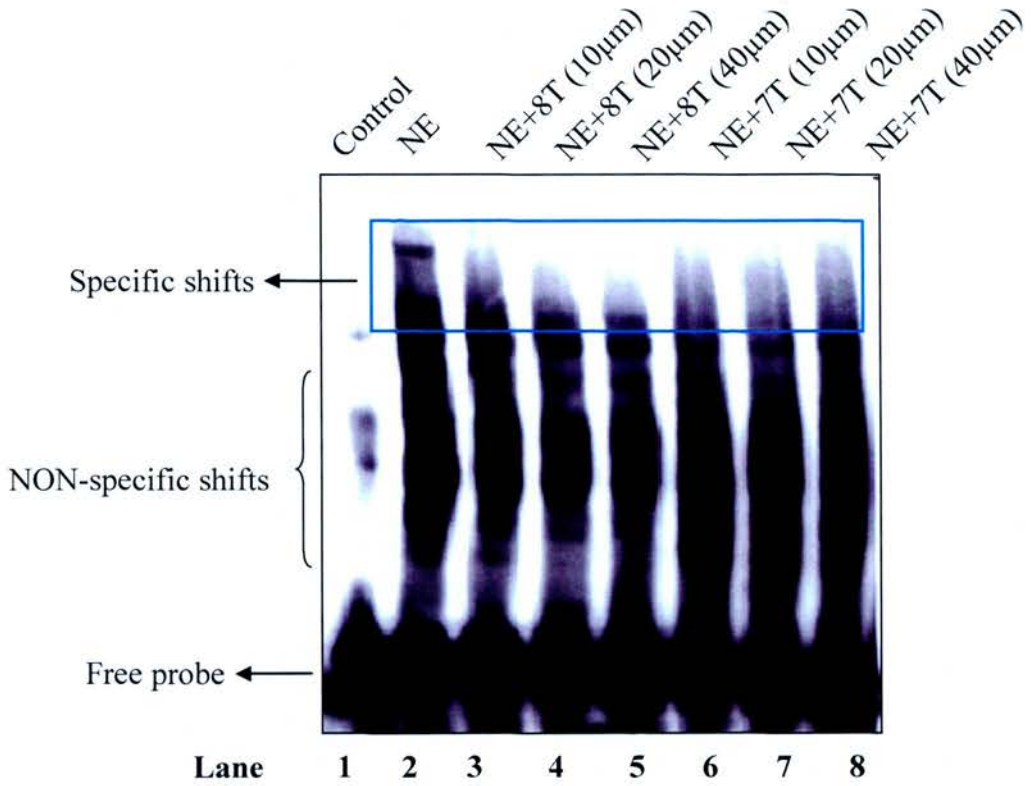
Figure 4.5 Competition EMSA showed -1663delT (7T) allele has higher binding affinity for proteins than -1663insT (8T) allele with biotin labelled 7T probe



EMSA was performed using 0.04μM biotin-labelled -1663delT (7T) probe (all lanes) and HOS TE85 cell nuclear extract (all lines). The bands presenting the specific binding of DNA-protein complex were attenuated more rapidly with increasing concentrations of the unlabeled competitor '7T' oligonucleotide (lane 5-7) compared with the unlabeled competitor '8T' oligonucleotide (lane 2-4). Assays were performed in triplicate. NE: nuclear extra; competitors: 8T (-1663insT) unlabeled probe; 7T (-1663delT) unlabeled probe.

Identical EMSA experiments were run, in triplicate, using biotin-labeled '8T' oligonucleotide probe instead of '7T' labelled probe (figure 4.6). Again the difference in binding affinity was found between the two competitors with the '7T' competitor binding with greater affinity than the '8T' competitor.

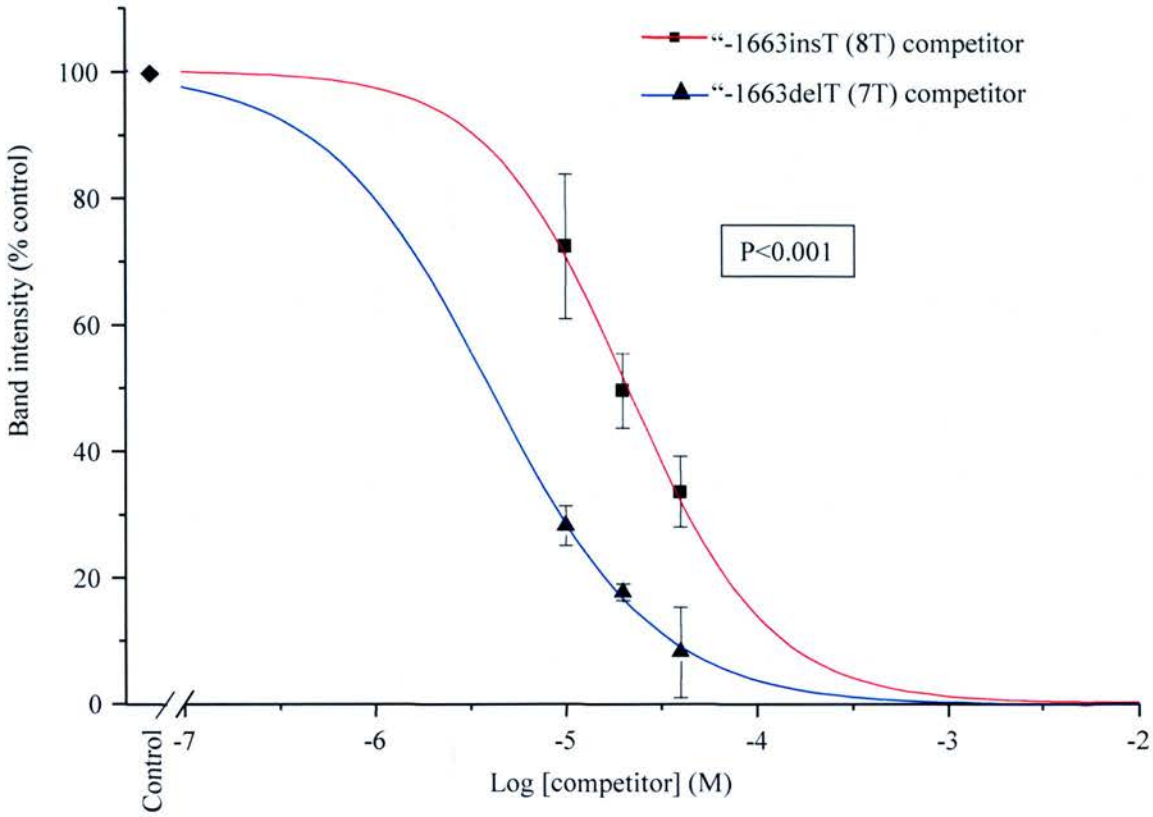
Figure 4.6 Competition EMSA showed -1663delT (7T) allele has higher binding affinity for proteins than -1663insT (8T) allele with biotin-labeled 8T probe



EMSA was peformed using 0.04μM biotin-labelled -1663insT (8T) probe (all lanes) and HOS TE85 cell nuclear extract (lane 2-8). The bands presenting the specific binding of DNA-protein complex were attenuated more rapidly with increasing concentrations of the unlabeled competitor '7T' oligonucleotide (lane 6-8) compared with the unlabeled competitor '8T' oligonucleotide (lane 3-5). Assay was performed in triplicate. NE: nuclear extra; competitors: 8T (-1663insT) unlabeled probe; 7T (-1663delT) unlabeled probe.

Band intensity results from three competitive EMSA experiments were quantified by Genetools software (SynGene). These results were averaged and analyzed by constructing a one-site competitive binding model ("Boltzman" model). As shown in Figure 4.7, band intensities (mean ± SEM) from lanes containing three different concentrations of competitors (lane 3-8, Figure 4.6) were expressed as a percentage of the band intensity in the absence of competitor (control, Figure 4.6, land 2). The band intensities of the two alleles were compared by GLM-ANOVA model (Minitab, version 12). Competitor "7T" had significantly greater binding affinity compared with the "8T" allele ($p<0.001$).

Figure 4.7 Competitive binding curves for -1663insT (8T) and -1663delT (7T) unlabeled probes when competing for biotin-labled "8T" – protein complex

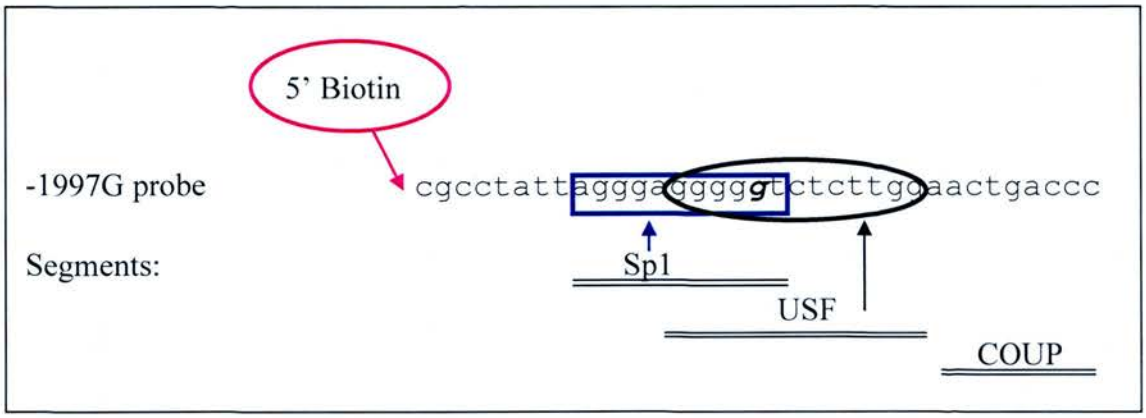


Band intensities of DNA-nuclear proteins complex from three EMSA experiments (mean ± SEM) were plotted against -1663insT (8T) and -1663delT (7T) unlabeled competitor concentration. Band intensity (mean ± SEM) at three concentrations was plotted as a percentage of band intensity in the absence of competitor for each competitor concentration (Figure 4.6, lane 2).For every competitor concentration "7T' competitor had greater binding affinity than "8T" competitor ($p<0.001$) and therefore a weaker band intensity was observed.

Sp1 Binds to the Region Surrounding the -1997T/G Polymorphism

Several proteins were predicted to bind the DNA surrounding -1997G/T polymorphism as shown in Figure 4.8.

Figure 4.8 Oligonucleotide sequence of -1997G probe used in EMSA and the potential binding motifs

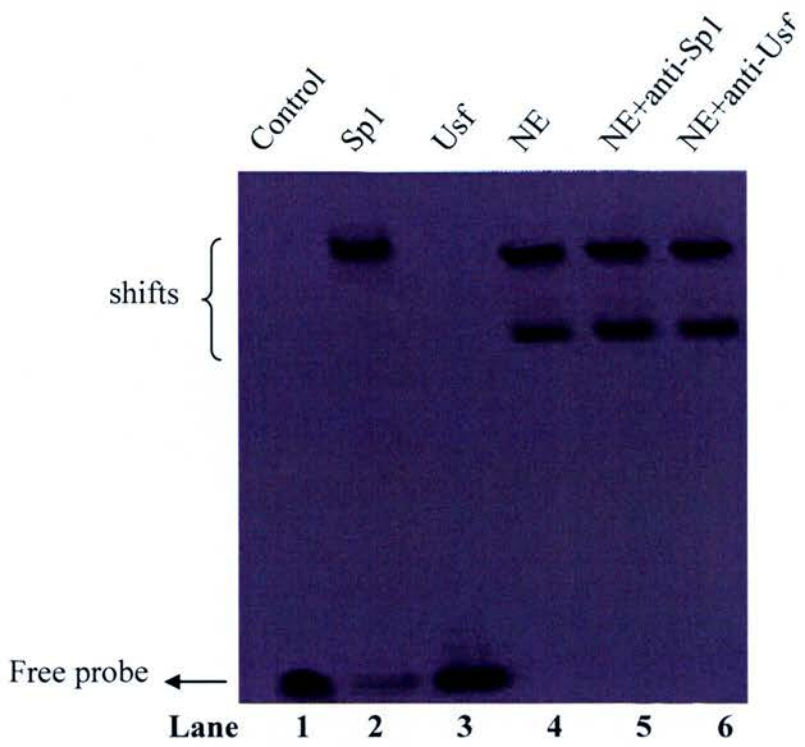


The oligonucleotide probe used in EMSA was 5'-end labeled with biotin. Three potential binding motifs were predicted by online program Alibaba 2.1. Two protein bindings investigated in this study (framed) were Sp1 and USF. The-1997G allele was **bold** *Italic*.
USF: Upstream Stimulatory Factor; COUP: Chicken Ovalbumin Upstream transcription factor

Since the DNA surrounding -1997G/T polymorphism contains a Sp1 site, I investigated the effects of the SNP on Sp1 protein binding. When Sp1 protein was used in the binding reaction which contained biotin-labeled -1997T oligonucleotide probe, one band was seen on the gel (figure 4.9, lane 2). When nuclear extracts from HOS TE85 cells were used as protein source in EMSA, two distinct bands were seen on the gel (lane 4), both different from the band in lane 2 representing the DNA-Sp1 complex. Addition of antibody to Sp1 did not change the position of DNA-nuclear protein complex (lane 5). To investigate the potential DNA-USF binding, Usf protein and its antibody were added into the binding reactions. As shown in Figure 4.9, addition of Usf protein did not result in any shift (lane 3), and the addition of USF

antibody did not further shift the DNA-nuclear protein complex, either (lane 6). Identical results were obtained when using biotin-labeled -1997G oligonucleotide probe (data not shown).

Figure 4.9 EMSA demonstrating binding of nuclear proteins to the biotin-labelled -1997T oligonucleotide probe

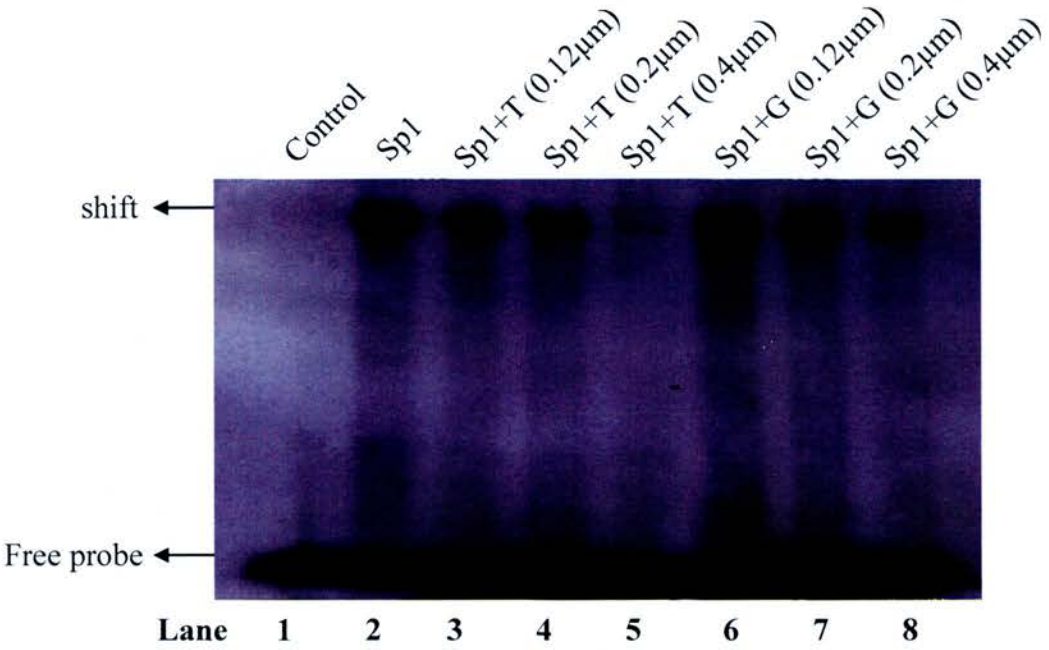


EMSA was performed with biotin-labeled -1997T oligonucleotide probe (all lanes). No protein was added to the control reaction (lane 1). DNA-Sp1 (lane 2) or DNA-nuclear protein complex (lane 4-6) were seen as shifts on gel. Usf protein showed no binding to the -1997T probe (lane 3). Addition of antibody of Sp1 and Usf did not change the binding of DNA-protein complex (lane 5-6).
NE, nuclear extracts from HOS TE85 cells.

-1997T Allele Had Greater Affinity for Sp1 than -1997G Allele

In the competition gel shift assays, biotin-labeled -1997T oligonucleotide probe was incubated with Sp1 protein and varying concentrations of either -1997T or -1997G unlabeled competitor oligonucleotides. The unlabeled -1997T competitor bound Sp1 protein with greater affinity than the unlabeled -1997G competitor resulting in weaker intensity of DNA-Sp1 band with -1997T competitor for each competitor concentration (Figure 4.10).

Figure 4.10 Competition EMSA showed -1997T allele had higher binding affinity for Sp1 protein than -1997G allele



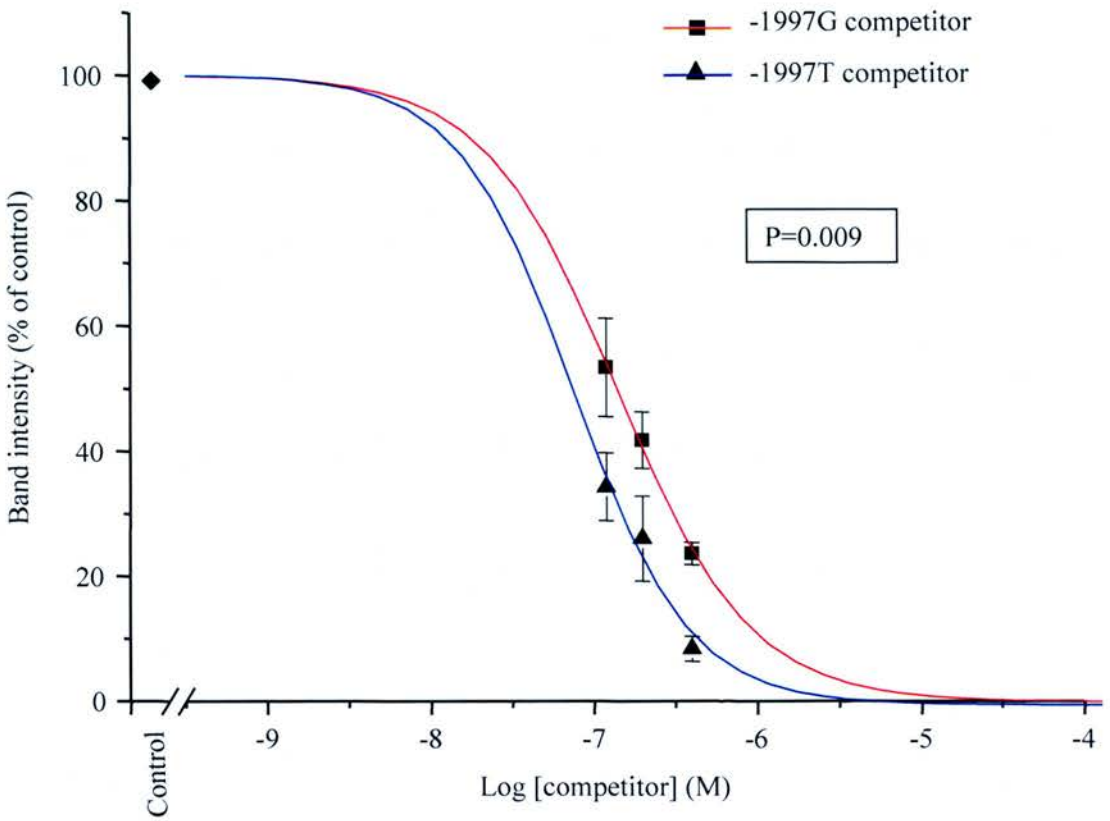
EMSA was performed with 0.04μM biotin-labelled -1997T oligonucleotide probe (all lanes) and Sp1 protein (lane 2-8). No addition of Sp1 in control reaction (lane 1). The bands presenting the specific binding of DNA-Sp1 complex were attenuated more rapidly with increasing concentrations of the unlabeled competitor 'T' oligonucleotide (lane 3-5) compared with the unlabeled competitor 'G' oligonucleotide (lane 6-8). Assay was performed in triplicate.

Identical EMSA experiments were run, in triplicate, using biotin-labeled -1997G oligonucleotide probe instead of -1997T labelled probe (data not shown). Again the difference in binding affinity for Sp1 protein was found between the two competitors

such that the -1997T competitor had greater binding affinity than the -1997G competitor.

To quantify the DNA-protein binding affinities between -1997G/T alleles, band intensity results from three competitive EMSA experiments were quantified by Genetools software (SynGene). These results were averaged and analyzed by constructing a one-site competitive binding model ("Boltzman" model). As shown in Figure 4.11, band intensities (mean ± SEM) from lanes containing three different concentrations of competitors (Figure 4.10, lane 3-8) were expressed as a percentage of the band intensity in the absence of competitor (control, Figure 4.10, land 2). The band intensities of two alleles were compared by GLM-ANOVA model (Minitab, version 12). Competitor T had significantly greater binding affinity compared with the G allele ($p=0.009$).

Figure 4.11 Competitive binding curves for -1997T and G unlabeled probes when competing for biotin-labeled T–Sp1 complex



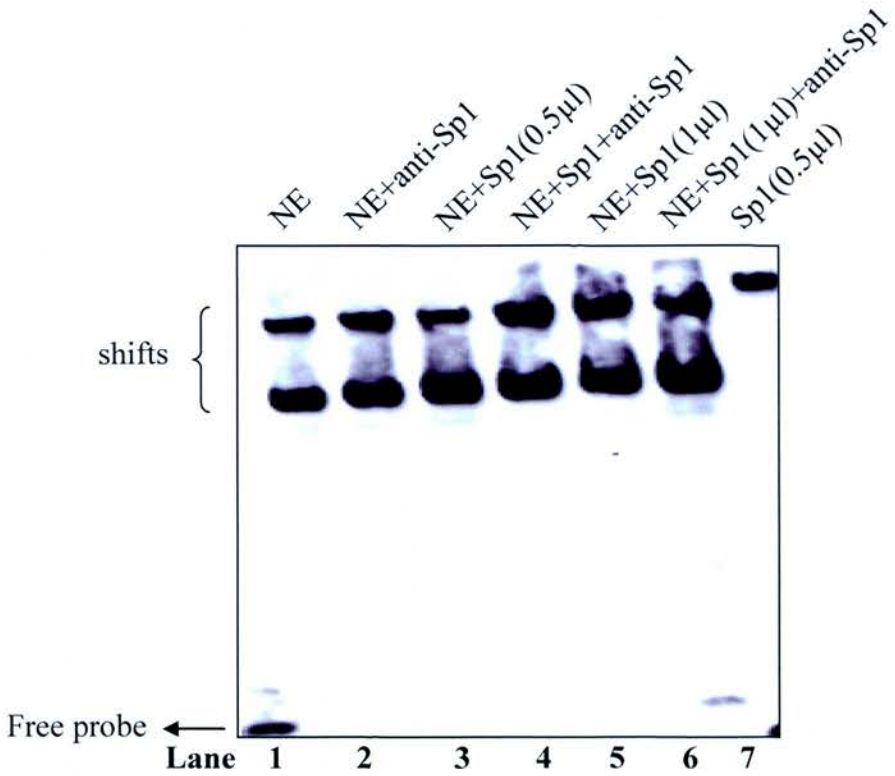
Band intensities of the DNA-Sp1 complex from Five EMSA experiments (mean ± SEM) were plotted against -1997T and G unlabeled competitor concentrations. Band intensity (mean ± SEM) at three concentrations was plotted as a percentage of band intensity in the absence of competitor for each competitor concentration (Figure 4.10, lane 2).For every competitor concentration T competitor had greater binding affinity than G competitor (p=0.009) and therefore a weaker band intensity was observed.

Unknown Nuclear Proteins Bind DNA Surrounding -1997G/T Polymorphism

When nuclear proteins from HOS TE85 cells were used as protein source in EMSA, two distinct bands were visualized (lane 1, Figure 4.12), which was different from the DNA-Sp1 band (lane 7). Addition of Sp1 protein or sp1 antibody did not change the binding pattern of DNA-nuclear protein complex (lane2-4). Increasing concentration of Sp1 protein had no impact on the gel shift either (lane 5-6).

Experiments were performed using nuclear extracts from MG63 cells and a biotin-labeled -1997G probe, with similar results (data not shown).

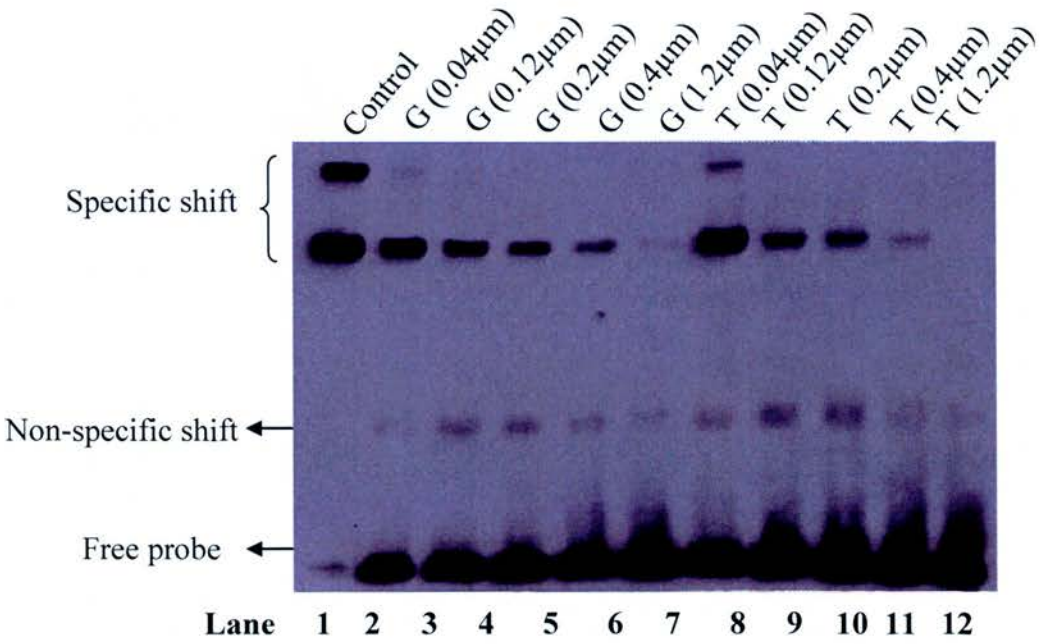
Figure 4.12 Unknown nuclear proteins bind the -1997T probe



When nuclear extracts from HOS TE 85 cells were used as protein source, two distinct bands were seen on the gel (lane 1). Addition of Sp1 antibody did not change the binding (lane 2). Addition of Sp1 with nuclear extracts produced the same shifts as lane 1 (lane 3). Sp1 antibody did not shift the DNA-nuclear protein binding (lane 4). Increased amounts of Sp1 protein (lane 5 & 6) did not cause any change in DNA-nuclear protein binding. Sp1 protein alone caused a DNA-Sp1 shift (lane 7).

Competition gel shift assays were performed to compare the difference in binding affinity of the polymorphic variants for nuclear proteins. Biotin-labeled -1997T oligonucleotide probe was incubated with nuclear extracts from HOS TE85 cells and varying concentrations of either T or G unlabeled competitor (Figure 4.13). Two bands were seen on the gel. The upper band was attenuated more rapidly with competitor -1997G compared with -1997T and disappeared completely when the amount of competitor was increased up to 0.12 μM (lane 2-4 & 8-9). The lower band disappeared completely with 1.2 μM of competitor T (lane12), but not with the same amount of competitor G (lane 7). Assays were repeated three times, and similar results were obtained. Identical experiments were performed using biotin-labeled -1997G probe, and again, similar results were obtained (data not shown).

Figure 4.13 The -1997G/T polymorphism shows different binding affinities



EMSA was run with 0.04μM biotin-labeled -1997T probe (all lanes) and nuclear extracts from HOS TE85 cells (lane 2-12). The upper band was attenuated more rapidly with increasing concentrations of the unlabeled G competitor (lane 2-3) compared with the unlabeled T competitor (lane 8). The lower band was attenuated more rapidly with T competitor than the G allele. The assays were performed in triplicate.

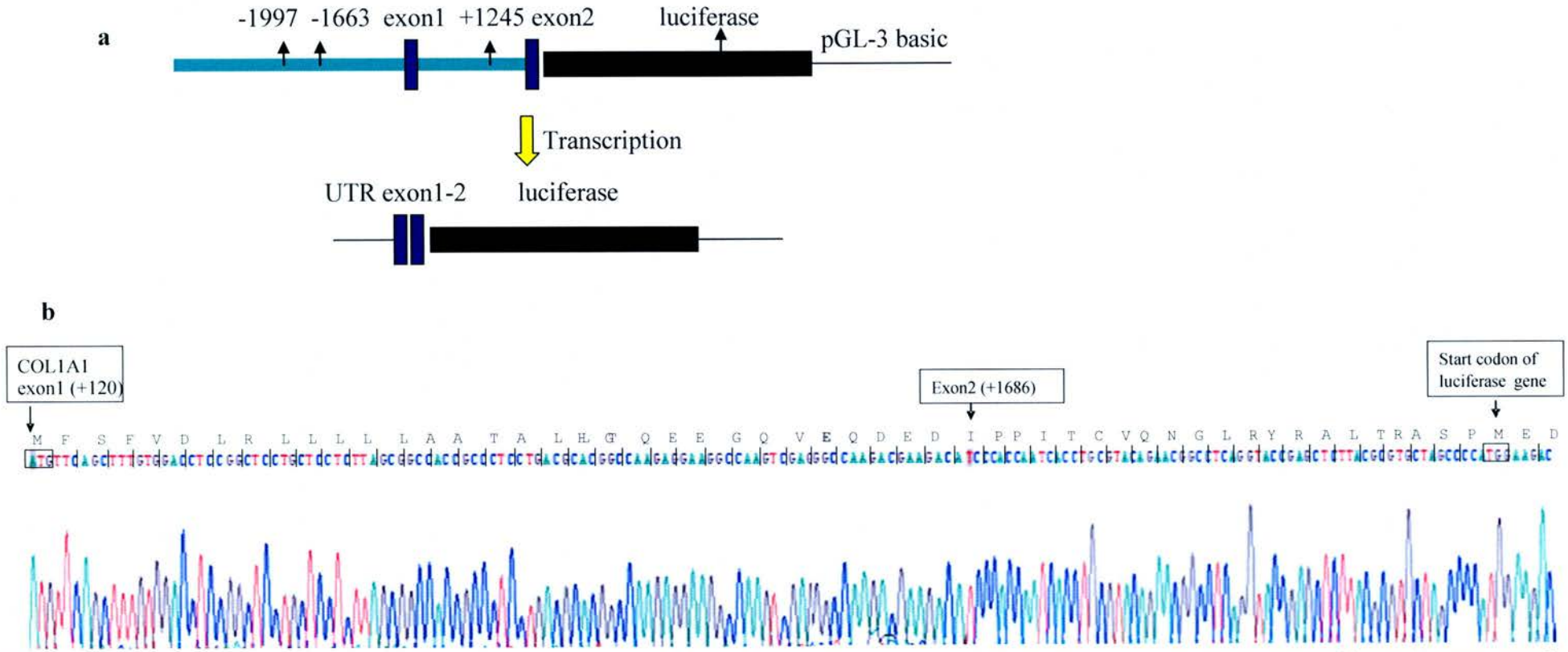
To quantify the DNA-protein binding affinities between -1997G/T alleles, band intensity results from three competitive EMSA experiments were analyzed by Genetools software (SynGene). Band intensities from lanes containing different concentrations of competitors were expressed as a percentage of the band intensity in the absence of competitor (control, lane 2, Figure 4.13). The band intensities of two alleles were compared by GLM-ANOVA model (Minitab, version 12). There was no statistically significant different binding affinity between two alleles ($p=0.3$ for the top band, $p=0.8$ for the bottom band).

Generation of *COL1A1*-luciferase Reporter Vectors

The pGL3-basic vector was modified to generate a fusion protein in which the open reading frame of the luciferase gene was inserted downstream of human *COL1A1* exon2. This resulted in production of a fusion protein containing the first 55 amino acids of *COL1A1* and the luciferase amino acid sequence (Figure 4.14). To confirm the presence of the *COL1A1*-luciferase fusion mRNA, total cell RNA was extracted twenty-four hours after the transfections and used as template to make cDNA by using reverse transcriptase. Then a 242bp fragment was amplified by a PCR reaction with a forward primer in *COL1A1* exon1 and a reverse primer in luciferase gene. The PCR product was then sequenced to confirm the presence of the fusion mRNA, as shown in Figure 4.14.

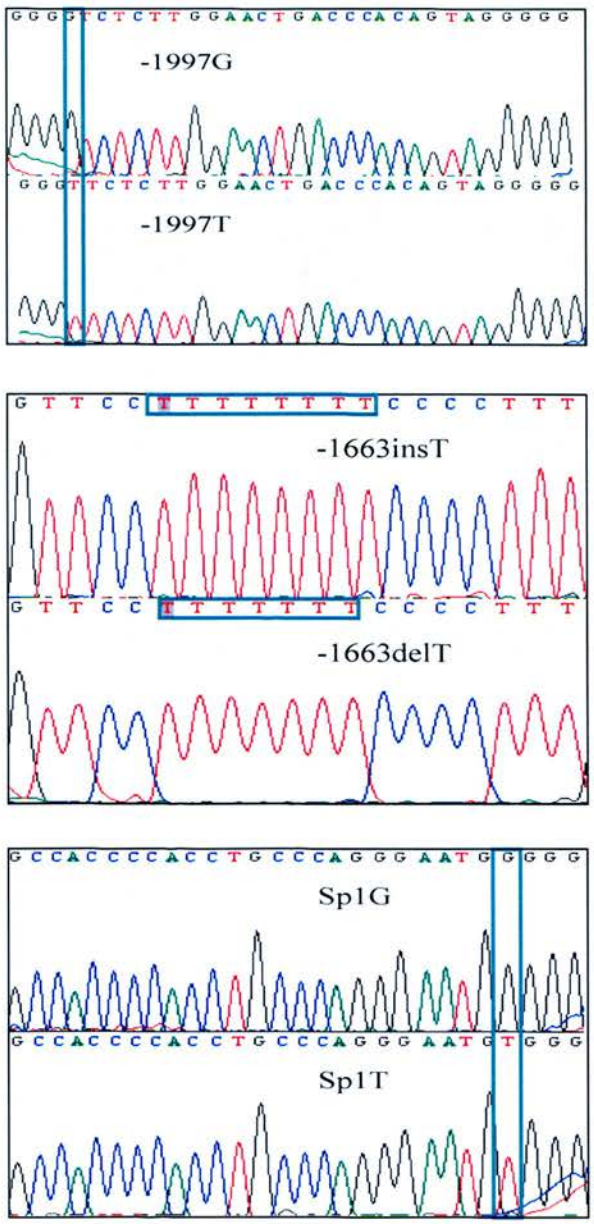
The vectors containing eight different haplotypes were generated by SDM reactions individually, and the sequence was confirmed by DNA sequencing analysis (Figure 4.15).

Figure 4.14 Sequence of cDNA and the amino acids of the fusion *COL1A1*-luciferase protein



The top panel (a) shows the structure of the reporter construct. The bright blue box represents the human *COL1A1* promoter and intron1 sequence; the dark blue boxes represent the human *COL1A1* exon1 and 2; the black box represents the luciferase cDNA.
The bottom panel (b) shows the sequence of the fusion *COL1A1*-luciferase cDNA. The translation codons were separated by vertical short lines. The amino acids were represented in upper case. The start codons of *COL1A1* and luciferase gene (ATG) were framed.

Figure 4.15 Sequencing reactions confirmed the Site-directed mutagenesis (SDM) at three polymorphic sites

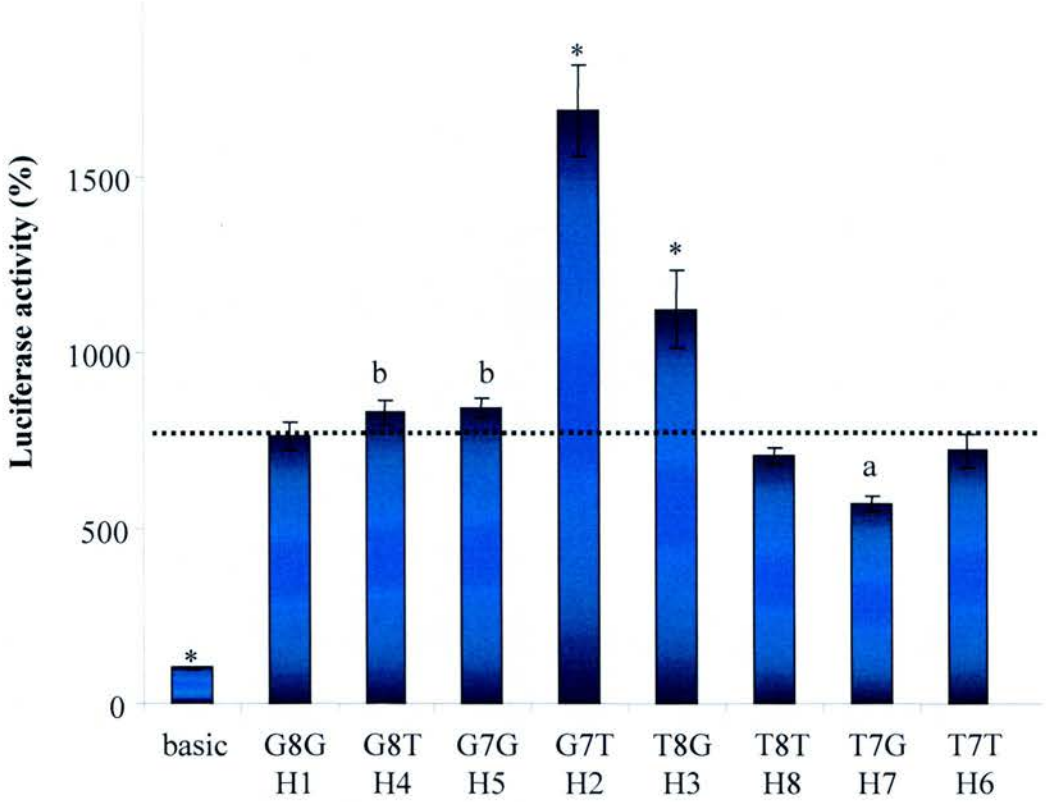


Three figures show base changes introduced by SDM at site -1997 (top panel), -1663 (middle panel) and +1245 Sp1 (bottom panel). The polymorphisms were represented in frame.

Luciferase Expression

The activities of *Firefly* luciferase (expressed from all pGL-3 reporter vectors) and *Renilla* luciferase (expressed from co-transfected pRL-TK vector) were measured sequentially from each sample by using the Dual-Luciferase® Reporter (DLR™) Assay System with the Plate-reader (Synergy HT, Bio-Tech). For each sample, the *Firefly* luciferase activity was normalized by the *Renilla* luciferase activity and then compared with the mean value of the pGL3-basic vector. As shown in Figure 4.16, the values of each vector represent the percentage of the relative activities to the pGL3-basic vector (where activity of the pGL-3 basic was set to 100). All the reporter vectors had significantly higher expression of luciferase than the pGL3-basic vector which lacks eukaryotic promoter and enhancer sequences ($p<0.001$). The vector containing haplotype 2 (G7T) had the highest luciferase activity, and this was significantly different from all the other vectors ($p<0.001$). The vector containing haplotype 3 (T8G) had the second highest luciferase activity, and this was significantly different to haplotype 4 G8T ($p=0.016$), to haplotype 5 G7G ($p=0.024$) and all the others ($p<0.001$). The vector containing haplotype 7 T7G had the lowest luciferase activity when compared with the vectors containing the other seven haplotypes. The activity of vector (haplotype7 T7G) differed significantly from those of haplotype 2(G7T) and 3 (T8G) ($p<0.001$), but also from the activity of G7G ($P=0.03$) and G8T ($P=0.054$). The rest of the vectors (G8G, G8T, G7G, T8T & T7T) had similar luciferase activities.

Figure 4.16 The luciferase activities expressed by pGL3-reporter vectors



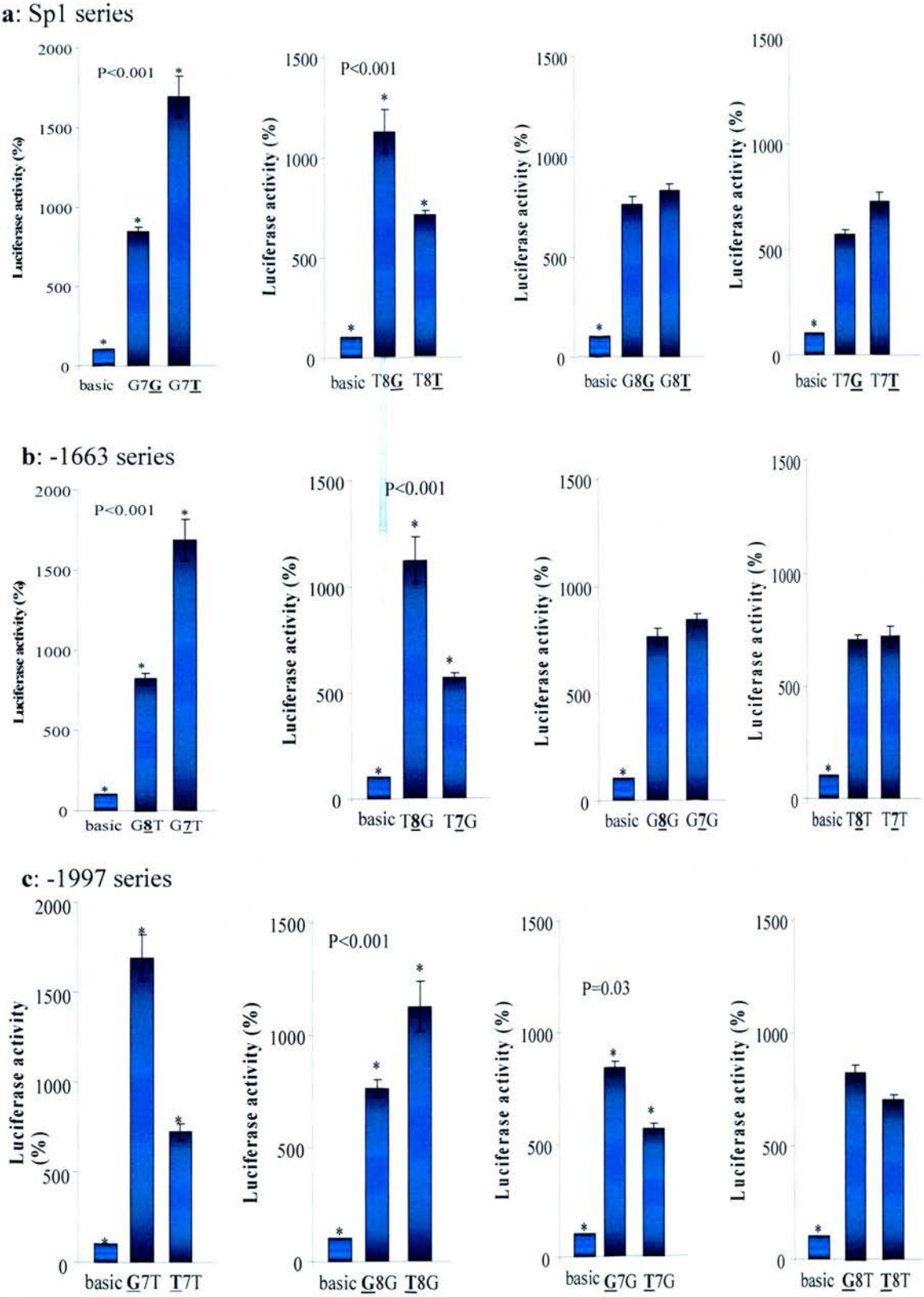
The values represent the percentage of the relative activities to the pGL3-basic vector (mean ± SEM). The transfection experiments were performed four times under the similar conditions. The results were analyzed by GLM-ANOVA.
a: Luciferase expression was significantly different from all the other vectors ($p<0.05$).
b: Luciferase expression was significantly different from a & c ($p\leq0.05$).
The dashed line represents the expression of the wild type vector (pGL-3 G8G).

The luciferase expressions were further analyzed by categorizing all the vectors according to three polymorphic sites (Figure 4.17). When vectors were grouped by Sp1 site, no consistent allele-specific effect on transcription was found between Sp1 T and G alleles. The Sp1T allele had higher luciferase expression than G allele when in context of '-1997G/-1663delT' (G7) ($p<0.001$), whereas T allele had lower luciferase when in context of '-1997T/-1663insT' (T8) ($p<0.001$). However, the Sp1 G and T allele had similar luciferase expression when in context of '-1997G/-1663insT' (G8) and '-1997T/-1663delT' (T7).

When vectors were grouped by -1663 site, no consistent allele-specific effect on transcription was observed, either. The -1663delT allele (7T) had higher luciferase expression than -1663insT allele (8T) only when in context of '-1997G/Sp1T' ($p<0.001$), whereas 8T allele had higher expression than 7T allele when in context of '-1997T/Sp1G' ($p<0.001$). There was similar luciferase expression between -1663 alleles when the other two sites had the same allele, such as '-1997G/Sp1G' or '-1997T/Sp1T'.

When vectors were grouped by -1997 site, no consistent allele-specific effect on transcription was observed. The -1997G allele showed higher luciferase expression than -1997T allele, when in context of '-1663delT', such as '-1663delT/Sp1G' (7G, $p=0.03$), '-1663delT/Sp1T' (7T, $p<0.001$). But -1997T allele showed higher luciferase expression than -1997G allele when in context of '-1663insT/Sp1G' (8G, $p<0.001$). There was no difference between -1997G/T alleles when in context of '-1663insT/Sp1T' (8T).

Figure 4.17 Comparisons of luciferase expressions by different categories



The luciferase expressions were analyzed by categorizing vectors into three series according to three polymorphic sites: Sp1 series (**a**); -1663 series (**b**), and -1997 series (**c**). Luciferase expression was represented as mean ± SEM.

Discussion

Previous studies showed that the +1245 Sp1T allele has significantly greater binding affinity for the transcription factor Sp1 than the G allele, in keeping with the fact that transcriptional efficiency of the T allele was increased compared with the G allele (Mann V *et al*, 2001). Another study showed that the -1663inT (8T) allele had higher binding affinity for nuclear proteins than the -1663delT (7T) allele and one of the binding proteins was identified as the architectural transcription factor Nmp4 (García-Giralt N *et al*, 2005).

The super-shift assays performed in this study confirmed that one of the nuclear proteins bound to the region surrounding -1663indelT polymorphism was Nmp4. I also showed that the transcription factor Osterix bound this region by using an Osterix antibody, which confirmed the potential Osterix binding site in this region. Since the super-shift with antibodies to Nmp4 and Osterix was the same, it is probable those two transcription factors bound to the same complex. However, it is not possible to identify which band represented which binding by super-shift assays as multiple shifts were seen on the gel and the super-shifts caused by two antibodies were identical. It is very likely that more than two nuclear proteins (Nmp4 & Osterix) bound this region and form DNA-multiple protein complexes. Although there are several potential binding sites in this region, as predicted by the online program Alibaba 2.1, I only tested a Sp1 binding site by using a Sp1 antibody. A single shift was seen when Sp1 protein was used as protein source in the gel shift and a super-shift was observed with addition of Sp1 antibody. The Sp1 antibody also resulted in a faint super-shift when nuclear extract was used as a protein source, whereas the NFATc1 antibody did not result in a super-shift. This suggested that Sp1 protein can bind the region surrounding -1663indelT polymorphism.

The -1663delT (7T) allele showed higher binding affinity for nuclear proteins than the -1663inT (8T) allele in competition gel shift assays with either labelled probe. By quantification of the band intensities from competitive EMSA, I found that the -1663delT (7T) allele had greater binding affinity than the "8T" allele. However, the overall effects of DNA-nuclear proteins binding are not clear. Osterix is a zinc finger-containing transcription factor encoded by the sp7 gene and expressed in osteoblasts of all endochondral and membranous bones. Osterix contains a DNA binding domain consisting three Cys2His2-type zinc fingers at its C terminus that share a high degree of identity with motifs in Sp1, Sp3 and Sp4 (Nakashima K *et al*, 2002). Osterix binds to Sp1 and EKLF consensus sequences and to other G/C–rich sequences.

Osterix is required for osteoblast differentiation and bone formation since in Osterix null mutant mice neither endochondral nor intramembranous bone formation occur (Nakashima K *et al*, 2002). Osterix regulates the expression of a number of important osteoblast genes such as osteocalcin, osteonectin, osteopontin, bone sialoprotein and type I collagen. In Osterix null mutant mice, *Col1a1* RNA levels in the mesenchyme were severely reduced, expressed at levels similar to those in skin fibroblasts (Nakashima K *et al*, 2002). Therefore, Osterix is a stimulator of *COL1A1* transcription.

On the other hand, nuclear matrix Nmp4 proteins are the only known Cys2His2 zinc finger proteins that bind to the AT-rich minor groove of DNA. These transcription factors regulate gene expression by virtue of their ability to bend or loop DNA, which thus facilitates communication between *cis*-elements and consequently modulates interactions between other *trans*-acting proteins (Wolffe AP, 1994). Some of the isoforms of Nmp4 recently have been identified as Cas-interacting zinc finger proteins (CIZ) that were characterized as nucleocytoplasmic shuttling proteins

localized at cell adhesion plaques. In CIZ-deficient mice, bone volume and the rate of bone formation were increased as well as the levels of mRNA expression of osteoblastic differentiation markers, such as alkaline phosphatase. Moreover, BMP2–induced bone formation on adult mouse calvariae *in vivo* was also enhanced by CIZ deficiency (Morinobu M *et al*, 2005). This is in agreement with the observations that Nmp4/CIZ over-expression suppressed the BMP2-enhanced expression of osteoblastic differentiation markers such as alkaline phosphatase and type I collagen (Shen ZJ *et al*, 2002). Therefore, Nmp4/CIZ is an inhibitor of the BMP signalling pathway and *COL1A1* transcription.

Two poly (dT) sequences in the rat *Col1a1* promoter region were identified as binding sites of Nmp4 (Alvarez M *et al*, 1998) and mutation of either site increased promoter activity in rat osteoblast-like cells in reporter assays (Thunyakitpisal P *et al*, 2001). One of the poly (dT) site (site B) is homologous to the human *COL1A1* -1663indelT site (García-Giralt N *et al*, 2002) and -1663inT (8T) showed increased binding affinity compared to the 7T allele (García-Giralt N *et al*, 2005). However our competition analysis showed that the 7T allele had greater binding affinity than the 8T allele. The difference between two experiments may be that double-stranded oligonucleotides were used as probes and competitors in our experiments while single-stranded oligonucleotides were used as probes and competitors in theirs. Moreover, our oligonucleotides are long enough to contain the whole potential Osterix binding site, which may have influenced the results.

Since the antibodies to Osterix and Nmp4 gave the same super-shift pattern and the two binding sites are next to each other, it is very likely that Osterix and Nmp4 interact with each other and form a DNA-multiple protein complex. That interaction may involve other nuclear proteins as the online program Alibaba 2.1 predicted

multiple binding sites overlapping each other at this region. Therefore, it is very difficult to predict how the different binding affinity for nuclear proteins between two alleles contributes to gene transcription simply based on the EMSA results. Further studies using reporter gene assays or CHIP assays would be of interest to explain these issues.

There was also an unexpected band appeared on the reaction with the actin antibody, which was very low molecular weight. Since there was no actin binding site on the oligonucleartide probe, one of the possible explain is that actin may interact with one of the transcription factors binding to the oligonucleartide probe. This interaction is needed to be investigated in future.

In the case of the -1997G/T polymorphism, the Sp1 protein bound to the region, because a single shift was seen on the gel when Sp1 protein was used as the protein source in gel shift and Sp1 antibody abolished this DNA-Sp1 shift. The -1997T allele showed increased binding affinity for Sp1 protein compared to the -1997G allele in competition assays. This is consistent with previous results that the +1245 Sp1T allele had greater binding affinity for Sp1 protein than the G allele (Mann V, *et al*, 2001). According to the prediction of online program Alibaba 2.1, Usf protein and antibody were also used in binding reactions. However, neither DNA-Usf shift nor super-shift was observed by using either of the probes, which suggested no DNA-Usf binding. Although I did not test the binding affinity of the Usf protein or Usf antibody to its known binding site, it is very unlikely both protein and the antibody do not work properly in this study.

Interestingly, two distinct shifts were observed when nuclear extracts from a osteosarcoma cell line (HOS TE 85 cells) were used as protein source. The migration

of either of these two shifts was different from the DNA-Sp1 shift. Moreover, addition of Sp1 antibody to nuclear extracts did not change the DNA-nuclear proteins binding pattern. It is possible that Sp1 protein is in competition with other nuclear proteins and loses the binding to DNA *in vivo*. Allelic differences in binding nuclear proteins were also detected by competition assays. However, the difference did not reach statistical significance.

In summary, three nuclear proteins bound to the region surrounding -1663indelT polymorphism were identified as Osterix, Nmp4 and Sp1 by super-shift assays. The -1663delT (7T) allele had greater binding affinity for nuclear proteins than the -1663inT (8T) allele. The -1997T allele had greater binding affinity for Sp1 protein than did the G allele *in vitro*. To clarify the contribution of allelic binding differences to gene transcription, more work is needed such as employment of reporter gene assays or Hap-CHIP assays.

This is the first study to investigate the transcriptional effects of all three polymorphisms on the *COL1A1* gene by reporter assays. Previous studies have shown that the Sp1T allele has increased binding affinity for Sp1 protein, and primary RNA transcripts derived from the Sp1T allele were approximately three times more abundant than Sp1G allele–derived transcripts in GT heterozygotes (Mann V *et al*, 2001). Collagen produced from osteoblasts cultured from Sp1 heterozygotes had an increased ratio of α1(I) protein relative to α2(I), and this was accompanied by an increased ratio of *COL1A1* mRNA relative to *COL1A2* (Mann V *et al*, 2001). A recent *in vitro* functional study showed that *COL1A1* promoter sequence containing haplotype '-1997G/-1663delT' (G7) drove luciferase activity more strongly than

haplotype '-1997T/-1663insT' (T8) (García-Giralt N *et al*, 2005). Moreover, this study also found that different transcription activity between two -1663insdelT alleles was dependent on the allele at -1997, and correlated with different binding capacities of the corresponding oligonucleotides to osteoblast nuclear proteins. However, the effect of the Sp1 polymorphism was not taken into account in this study.

Since the three polymorphisms are in high linkage disequilibrium and interact with each other to regulate BMD, it is also possible that all three 5' polymorphisms interact with each other to regulate gene transcription. In the current study, the luciferase reporter vectors containing 5' *COL1A1* sequence with all three polymorphisms were used to investigate the transcriptional activity of all eight possible haplotypes. The 5' *COL1A1* sequence was inserted into a modified pGL3-basic vector to produce a fusion protein containing the first 55 amino acids of *COL1A1* fused to the luciferase gene. The results showed that all the reporter vectors expressed significantly higher luciferase activity levels than pGL3-basic vector which lacks eukaryotic promoter and enhancer sequences ($p<0.001$). This means the *COL1A1*-luciferase fusion protein was functional and was expressed properly. This was also confirmed by sequencing the cDNA of the *COL1A1*-luciferase fusion protein.

The pGL3-G7T vector had the highest transcriptional activity compared with all the other haplotypes. This is in agreement with the results of García-Giralt N *et al* (2005) and Mann v *et al* (2001), as haplotype 'G7' and Sp1T allele showed higher transcriptional activity in their studies, respectively. Given the fact that haplotype 'G7T' was associated with decreased BMD at LS & FN (chapter 3), this also supports our hypothesis that the mechanism underlying the association of *COL1A1* haplotypes with osteoporosis is that increased transcript in the *COL1A1* gene results leads to an increased ratio of α1(I) protein relative to α2(I). However, other factors may also be

operative since the haplotype T8G which was associated with high BMD had the second highest transcription activities of the eight vectors studied.

The transcriptional activity of vector pGL3-T7G (haplotype7) was the lowest in eight reporter vectors, and significantly lower than those of pGL3-G7G (haplotype5, P=0.03) and pGL3-G8T (haplotype4, p=0.054), although it did not differ significantly from the wild type pGL3-G8G (haplotype1). But the haplotype7 (T7G) was not common in APOSS and was not associated with BMD. Five of the other reporter vectors containing haplotype1, 4, 5, 6 and 8 (G8G, G8T, G7G, T8T & T7T) had similar transcriptional activities.

The results also suggest that the three polymorphisms interact with each other to regulate transcription, which agrees with García-Giralt N *et al* (2005). When the eight reporter vectors were analyzed by categorizing into three series according to three polymorphic sites, no consistent allele-specific effect on transcription was observed at any polymorphic site. For example, when vectors were grouped by the Sp1 site, the Sp1T allele had higher luciferase expression than the G allele when in context of '-1997G/-1663delT' (G7) ($p<0.001$), whereas T allele had lower luciferase when in context of '-1997T/-1663insT' (T8) ($p<0.001$). However, the Sp1 G and T allele had similar luciferase expression when in context of '-1997G/-1663insT' (G8) and '-1997T/-1663delT' (T7). Therefore, the haplotype effect on reporter gene transcription could not be simply predicted by the individual effect of any polymorphism, either.

The gel shift assays showed that -1663delT allele (7T) had greater binding affinity for nuclear proteins, such as transcription factors Nmp4 and Osterix. The -1997T and Sp1T alleles had higher binding affinity for Sp1 protein than T allele. However, neither -1663delT (7T), -1997T, or Sp1T allele showed consistent effects on reporter gene transcription. This suggests that interactions among the three polymorphisms

have a more important effect on reporter gene transcription than individual polymorphisms.

However, the limitation of reporter gene assays is that chromatin structure is not taken into account. It is known that chromatin structure has an important effect on gene transcription. Therefore, to accurately investigate the effects of three polymorphisms on gene transcription, further study is required such as Hap-CHIP assay.

CHAPTER FIVE

Role of *COL1A1* Variants in the Mechanical Properties of Bone

Abstract

In chapter 3, I analyzed the relation between two *COL1A1* promoter polymorphisms and the Sp1 polymorphism and showed that they interacted to regulate BMD. In chapter 4, I showed that the regions surrounding both promoter polymorphisms were able to recognise nuclear proteins in gel shift assays and different alleles influenced DNA –protein binding affinities. The promoter polymorphisms interacted with the Sp1 polymorphism to affect reporter gene transcription *in vitro*. The Sp1 polymorphism of the *COL1A1* gene has been previously associated with reduced bone strength and abnormalities of bone mineralization *in vitro* and *in vivo*. In view of this, I sought to determine if haplotypes at the 5' flank of *COL1A1* were associated with the mechanical properties of bone.

Biomechanical tests were performed on trabecular bone samples from femoral heads obtained from 23 male and 71 female hip fracture patients. The stiffness and yield strength were determined from the stress-strain curve using standard techniques described previously. DNA was extracted from each bone sample and the genotypes at all three polymorphic sites were determined by PCR and DNA-sequencing analysis. Haplotypes were constructed using the 'PHASE' program. Statistical analysis (GLM-ANOVA) was used to compare bone volume, density, stiffness and yield strength according to genotypes and haplotypes. Bone cores derived from common homozygotes (-1997G/G, -1663delT/delT and Sp1G/G) had significantly increased bone volume, yield strength and stiffness compared with the cores derived from heterozygotes and rare homozygotes ($p<0.05$). The second most common haplotype 6 (-1997T/-1663delT/Sp1T, 12.24%) was associated with reduced bone volume, yield strength and stiffness ($p<0.05$). Interestingly, this haplotype was very rare in the general population and its frequency in the bone cores from the hip fracture patients was 401 times greater than in the APOSS population (0.03%) described in chapter 3. The associations remained significant after adjusting for gender, age, bone volume and the density of bone core. Therefore, the three common variants in 5' *COL1A1*

gene may predispose to osteoporosis by complex mechanisms involving a reduction in bone quality.

Introduction

Previous studies showed that the *COL1A1* Sp1 polymorphism acts as a predictor of the reduced BMD and increased risk of fragility fractures in several populations (Efstathiadou Z *et al*, 2001; Mann V *et al*, 2003). In a small study of 17 patients, mechanical tests also showed that the yield strength of bone derived from Sp1 G/T heterozygotes was reduced when compared with bone derived from Sp1 G/G homozygotes (Mann V *et al*, 2001). Subsequent studies showed that carriage of the *COLIA1* Sp1 T allele is associated with an impaired ability of osteoblast-like cells to form mineralized bone nodules *in vitro* and with abnormalities of bone mineralization *in vivo* (Stewart TL, *et al*, 2005). It is already known that specific abnormalities of collagen structure or quantity due to mutations in the *COL1A1* or *COL1A2* gene can cause osteogenesis imperfecta, a disease characterized by impaired material properties of bone and increased bone fragility. Some mutations that change the amino acid sequence can result in formation of branched fibers leading to brittle bone and abnormal mineralization (Baum J and Brodsky B, 1999). Therefore, the increased bone fragility in carriers of the Sp1T allele may result in part from defects in bone mineralization.

The effects of the promoter polymorphisms on biomechanical properties of bone have not yet been studied. In view of this, biomechanical properties of bone were studied according to genotypes and haplotypes defined by the 5' flank polymorphisms of the *COL1A1* gene on trabecular bone cores derived from femoral heads removed from hip fracture patients.

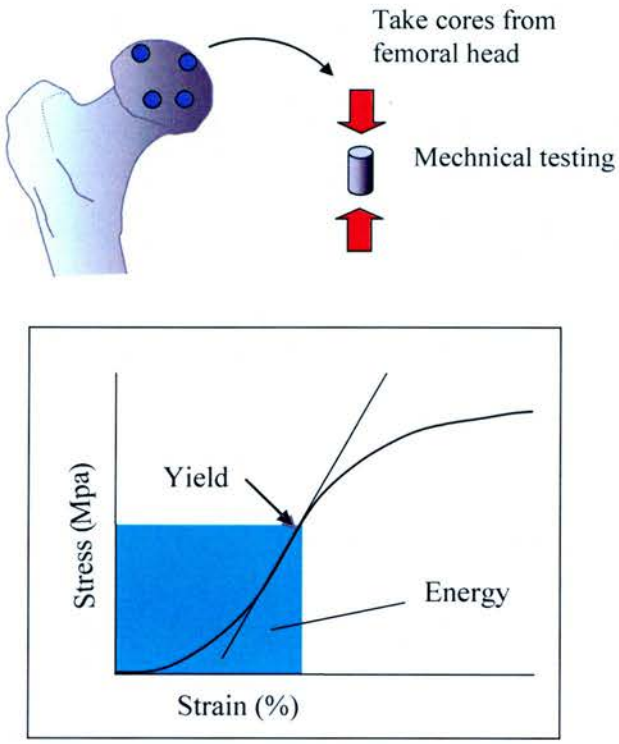
Methods

The biomechanical studies were performed on trabecular bone samples from femoral heads obtained from 23 male and 71 female patients undergoing hip replacement surgery as a result of osteoporotic hip fracture. The age at fracture ranges from 58 to 101 years with a average age of 81.6 ± 9.0 years.

Cylindrical cores of bone were removed from the superior in the femoral head and then subjected to an unconstrained compression test using an Instron 5564 materials testing machine (Instron, High Wycombe, UK), as shown in figure 5.1. The apparent density of washed and defatted bone samples was calculated by dividing the total mass by the sample volume. The stiffness and yield strength were determined from the characteristics of the stress-strain curve using standard techniques described previously (Li B and Aspden RM, 1997).

DNA was extracted from bone cores and the genotypes were determined by PCR followed by sequencing analysis as described in chapter 2. Haplotypes were constructed by the program 'PHASE'. Genotype and haplotype differences in bone volume, density of bone core, yield strength and stiffness were analyzed by comparing the common homozygotes with the heterozygotes and the rare homozygotes. GLM-ANOVA was used to adjust the effect of gender, age and density of bone core on bone volume, yield strength and stiffness by entering "gender", "age", "density of the bone core" and "genotype/ haplotype" into the "model". Multiple regression analysis and best subsets multiple regression analysis were used to evaluate associations between genotype, haplotype and bone volume, yield strength and stiffness.

Figure 5.1 The methods of biomechanical testing



Dean *et al* 1998

The top panel shows that trabecular bone cores were taken out from femoral head and subjected to mechanical testing. The bottom panel shows the stress-strain curve from which the "stiffness" and the "yield strength" were determined.
Mpa, megapascals.

Results

Characteristics of the Subjects

The average age at fracture of all subjects was 81.6±9.0 years. The average age at fracture was 82.5±9.3 years for 69 female, and 79.0±7.5 years for 23 males. There was no difference in genotype distribution according to age or gender ($p>0.05$, Table 5.1).

The genotype frequencies for all three polymorphisms were in Hardy-Weinberg equilibrium. Genotype frequencies of common homozygotes (-1997G/G, -1663In/In & Sp1G/G) were similar to those in younger individuals from the general population (APOSS) ($p>0.05$, Table 5.2). Eight possible haplotypes were predicted by the 'PHASE' program. Two common haplotypes accounted for 88% of alleles at the 5' flank of the *COL1A1* locus (Table 5.3). Haplotype1 (-1997G/ -1663InT/ +1245G, G8G) accounted for 76.53% of alleles, which is similar to its frequency in the general APOSS population (67.05%). But haplotype 2 and 3 were very rare in bone core subjects (1.53% and 1.02%, respectively), which were the second and third common haplotypes in the general population APOSS (17% and 12.21%, respectively). Conversely, haplotype 6 (-1997T/ -1663delT/ +1245G, T7T) had the second highest frequency in bone core subjects (12.24%), which was 400 times more common in this population than in the general APOSS population (0.03%). The other four haplotypes were equally rare in bone core subjects and the general population (<3%). The difference between haplotype distribution in bone core subjects and the APOSS population was highly significant ($p<0.001$, χ^2 test).

Table 5.1 Genotype frequencies in bone core samples in relation to gender and age

Genotype		Male (%) N=23	Female(%) N=69	Total(%) N=92	Age (years)*	p^a	p^b
-1997G/T	G/G	16 (69.57)	50 (72.46)	66 (71.74)	81.37±8.97	0.86	0.57
	G/T	7(30.43)	18 (26.09)	25 (27.17)	82.58±9.54		
	T/T	0	3 (4.35)	3 (3.26)	81.50±6.56		
-1663In/DelT	Ins/InsT	18 (78.26)	49 (71.01)	67 (72.83)	81.34±8.90	0.76	0.40
	Ins/DelT	5 (21.74)	17 (24.64)	22 (23.91)	82.14±9.99		
	Del/DelT	0	5 (7.25)	5 (5.43)	84.00±6.42		
sp1G/T	G/G	18 (78.26)	46 (66.67)	64 (69.57)	81.18±9.29	0.67	0.39
	G/T	5 (21.74)	22 (31.88)	27 (29.35)	83.04±8.60		
	T/T	0	3 (4.35)	3 (3.26)	81.00±6.68		

p^a, ANOVA test of genotype frequencies in gender; p^b, ANOVA test of genotype frequencies in age. Age*, age at fracture.

Table 5.2 Comparison of genotype frequencies in bone core samples and APOSS population

Genotype		Bone core (%) N=92 Age=81.6±9.0	APOSS (%) N=3011 Age=48.4±2.3	p
-1997G/T	G/G	66 (71.74)	2247 (74.63)	0.57
	G/T	25 (27.17)	699 (23.21)	
	T/T	3 (3.26)	65 (2.16)	
-1663In/DelT	Ins/InsT	67 (72.83)	1967 (65.85)	0.49
	Ins/DelT	22 (23.91)	869 (29.09)	
	Del/DelT	5 (5.43)	151 (5.06)	
sp1G/T	G/G	64 (69.57)	2145 (66.51)	0.94
	G/T	27 (29.35)	963 (29.86)	
	T/T	3 (3.26)	117 (3.63)	

p, ANOVA test of genotype frequencies in bone core subjects and APOSS.

Table 5.3 Comparison of haplotype frequencies in bone core samples and APOSS population

Haplotype	**Bone cores (%)** N=96*2	**APOSS (%)** N=3227*2
G8G (1)	150 (76.53)	4396 (67.05)
G7T (2)	3 (1.53)	1114 (17.00)
T8G (3)	2 (1.02)	800 (12.21)
G8T (4)	4 (2.04)	102 (1.56)
G7G (5)	5 (2.55)	135 (2.06)
T7T (6)	24 (12.24)	2 (0.03)
T7G (7)	3 (1.53)	3 (0.05)
T8T (8)	5 (2.55)	2 (0.03)

The haplotype frequencies in bone core samples are significantly different from those in APOSS population ($p<0.001$, χ^2-test). N: the number of chromosome, calculated by multiplying the total number of the subjects by 2.

Effects of Individual Polymorphisms on Biomechanical Properties of Bone

The relationship between *COL1A1* genotypes and biomechanical properties of bone is shown in Table 5.4. When three genotypes at each polymorphism were compared, the common homozygotes (-1997G/G, -1663in/inT, & Sp1G/G) were associated with increased yield strength and stiffness when compared with carriers of the less common allele (p^a <0.05). The Sp1G/G genotype was also associated with increased bone volume (p^a =0.017). There was no difference of genotype distribution in length of bone core or density of bone core (p^b >0.05), but all three polymorphisms were associated with bone volume, yield strength and stiffness in unadjusted and adjusted models. The -1997G/G genotype was associated with significantly increased bone volume, yield strength and stiffness (p^b=0.028, p^b<0.001, p^b<0.001, respectively). Similar associations were observed for the -1663in/delT and the Sp1G/T polymorphisms. The -1663in/inT and the Sp1G/G genotypes were associated with increased bone volume, yield strength and stiffness. After the adjustment for gender, age and "density of bone core", the associations remained significant (p^{adj}<0.05). These associations were similar when females (N=71) were analyzed separately, but no comparison was conducted in male group due to the small number (N=23).

Table 5.4 Biomechanical properties of bone according to *COLIA1* genotypes

Genotype	**-1997G/G N=68**	**-1997G/T N=26**	**-1997T/T N=4**	**p^a**	**p^b**	**p^{adj}**
length of core (mm)	9.29±1.23	9.19±1.45	8.56±1.80	0.55	0.78	
density of bone core (g/mm^3)	1.41±0.13	1.37±0.10	1.30±0.18	0.09	0.089	
bone volume (mm^3)	0.30±0.10	0.26±0.07	0.24±0.10	0.10	0.028	0.001
yield strength (MPa)	32.16±8.98§	25.35±7.26†	26.25±8.77	0.002	<0.001	0.001
Stiffness (MPa)	260.77±68.49§	217.7±53.20†	181.9±53.50	0.003	<0.001	0.001
Genotype	**1663Ins/InsT N=69**	**1663Ins/delT N=23**	**1663del/delT N=6**	**p^a**	**p^b**	**p^{adj}**
length of core (mm)	9.28±1.17	9.13±1.65	9.10±1.63	0.86	0.88	
density of bone core (g/mm^3)	1.40±0.12	1.40±0.13	1.35±0.17	0.65	0.86	
bone volume (mm)	0.30±0.09	0.26±0.08	0.25±0.08	0.07	0.017	0.017
yield strength (MPa)	32.14±8.86§	24.3±7.19†	29.00±8.17	0.001	<0.001	<0.001
Stiffness (MPa)	257.77±68.10§	222.2±61.2†	203.7±53.9	0.025	0.011	0.02
Genotype	**Sp1G/G N=66**	**Sp1G/T N=28**	**Sp1T/T N=4**	**p^a**	**p^b**	**p^{adj}**
length of core (mm)	9.34±1.23	9.00±1.51	9.15±1.07	0.53	0.46	
density of bone core (g/mm^3)	1.41±0.12	1.37±0.12	1.28±0.16	0.08	0.16	
bone volume (mm)	0.31±0.10§	0.25±0.07†	0.27±0.07	0.017	0.003	<0.001
yield strength (MPa)	32.85±8.85§	24.82±6.48†	22±6.22†	<0.001	<0.001	<0.001
Stiffness (MPa)	264.84±67.24§	208.66±52.86†	199.4±41.7	<0.001	<0.001	<0.001

p^a: GLM-ANOVA among three different genotype groups for each polymorphism; p^b: student t test between the common homozygotes and other genotype groups for each polymorphism; p^{adj}: adjusted for density of bone core & gender in GLM-ANOVA model.

§: significantly different from † group.

Regression analysis was used to investigate the relationship between the response (yield strength & stiffness) and the predictors by entering genotypes, gender, age at fracture, density of bone core, and bone volume into the model as predictors (Minitab). This regression model showed that three genotypes, density of bone core and bone volume predicted yield strength independently ($p<0.05$, Table5.5). The density of bone core and bone volume predicted stiffness independently ($p<0.001$ for both). The three polymorphisms did not predict stiffness significantly in the multiple regression model including gender, age at fracture, density of bone core, and bone volume.

Table 5.5 Multiple linear regression analysis of yield strength for the 5' flank *COL1A1* genotypes

Independent predictor	**R^2**	**p**
Bone density (g/mm^3)	13	<0.05
Bone volume (mm^3)	3.9	<0.05
-1997 genotypes	10.2	0.051
-1663 genotypes	7.5	0.038
Sp1 genotypes	19.9	0.001

R^2 value is the percentage of variance explained.

Effects of Haplotypes on Biomechanical Properties of Bone

Relevant characteristics of the bone cores in relation to the two common haplotypes are shown in Table 5.6. There was no significant association between any haplotype and age at fracture or v. Homozygotes for haplotype 1 had significantly increased bone volume (p=0.001), yield strength (p<0.001) and stiffness (p=0.002). Conversely, carriers of haplotype 6 had significantly reduced bone volume (p=0.001), yield strength (p<0.001) and stiffness (p=0.001). Because the homozygotes of haplotype 6 were very rare (n=2), they were combined with heterozygotes when analysis was carried out (p^b, Table 5.6).

Table 5.6 Biomechanical properties of bone according to two *COL1A1* haplotypes

haplotype 1 (G8G)	**0 copy** N=6	**1 copy** N=29	**2copies** N=56	p^a	p^{adj}
Age at op (yrs)	82.2±5.9	83.3±9.3	80.6±9.1	0.42	
Density of Bone core (g/mm^3)	1.387±0.139	1.390±0.128	1.414±0.122	0.649	
Bone volume (mm)	0.257±0.076	0.262±0.073 §	0.312±0.097†	0.028	0.001
Yield strength (MPa)	28±9.08	25.03±6.80 §	33.35±8.92†	<0.001	<0.001
Stiffness (MPa)	231.7±44.9	221.3±56.1 §	267.43±67.97 †	0.003	0.002
haplotype 6 (T7T)	**0 copy** N=75	**1 copy** N=17	**2copies** N=2	p^b	p^{adj}
Age at op (yrs)	81.70±9.1	81.88±9.15	77.50±4.95	0.92	
Density of Bone core (g/mm^3)	1.41±0.13	1.38±0.10	1.31±0.24	0.15	
Bone volume (mm)	0.30±0.09	0.25±0.09	0.30±0.11	0.034	0.001
Yield strength (MPa)	32.05±8.79	23.35±6.45	22.00±9.90	<0.001	<0.001
Stiffness (MPa)	259.60±65.1	203.60±0.10	227.10±0.24	0.001	0.001

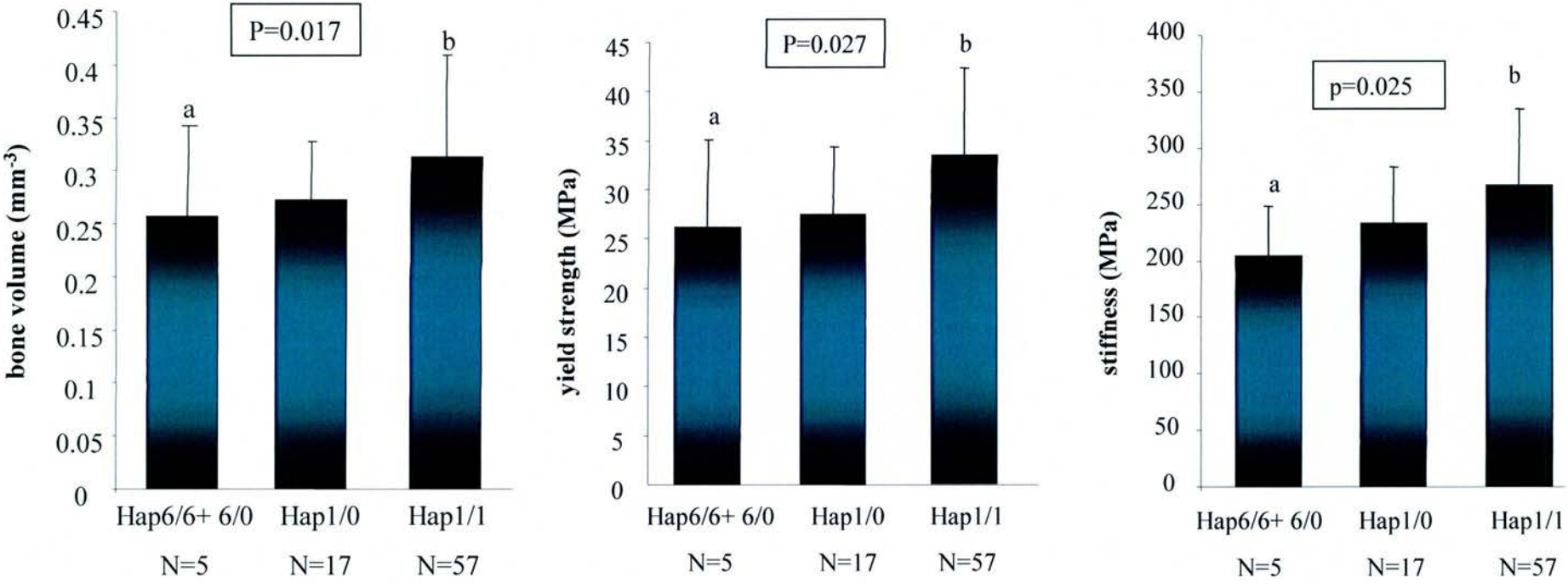
p^a: GLM-ANOVA model for three groups of haplotype1; p^b: student *t*-test for "0 copy" and "1 copy+2copies" of haplotype6; p^{adj}: adjusted for density of bone core & gender in GLM-ANOVA model by entering density of bone core, gender and the haplotypes into the model.
§: significantly different from † group.

Multiple regression analysis showed that haplotype 1 & 6 were also independent predictors of yield strength (p=0.023 & 0.007, respectively). In best subsets regression analysis, haplotype 1 accounted for 12.3% of the variance in yield strength, and haplotype 6 accounted for 13.6% of the variance in yield strength. However, neither of these haplotypes predicted stiffness significantly.

Since haplotype 1 and 6 had opposing effects on material properties of bone, all the samples were stratified according to combinations of these two haplotypes. This also showed significant differences in bone volume, yield strength and stiffness between the extreme haplotypes (Figure 5.1). Subjects who were homozygous for haplotype 1 had the highest bone volume, yield strength and stiffness, and carriers of haplotype 6 (homozygotes & heterozygotes) had the lowest bone volume, yield strength and stiffness. The difference between haplotypes was greatest for bone volume (p=0.017), but was also significant for yield strength (p=0.027) and stiffness (p=0.025). The difference in bone volume between the extreme haplotypes (homozygotes for haplotype 1, versus heterozgotes & homozygotes for haplotype 6) was 0.05 mm^{-3} (about 0. 6 SD). The difference between the extreme haplotypes was 7.15 MPa (about 0.8 SD) in yield strength, and 62.83 MPa (about 0.9 SD) in stiffness.

The relationships between individual haplotypes and unadjusted bone volume, yield strength and stiffness were also analyzed as illustrated in Figure 5.1. These also gave the similar results to those obtained with the adjusted values.

Figure 5.1. Relationship between *COL1A1* haplotypes and biomechanical properties of bone.



The values of bone volume (left), yield strength (middle) and stiffness (right) were adjusted for gender and density of bone core by ANOVA-GLM model in relation to carriage of haplotype 1 and 6 in the 5' flank of *COL1A1* gene. Columns are means and bars are SD.
6/6 represents homozygotes for haplotype 6; 6/0 represents heterozygotes for haplotype 6 who were not carriers of haplotype 1; 1/0 represents heterozygotes for haplotype 1 who were not carriers of haplotype 6; 1/1 represents homozygotes for haplotype 1.
a: significant different from group b.

Discussion

So far only one study was performed to study the relationship between the Sp1 polymorphism and bone quality (Mann V *et al,* 2001). In that study, the Sp1 "Ss" genotype (G/T heterozygote) was associated with reduced yield strength in a small study of 23 patients (p=0.031), but no association was found between genotype and stiffness (p=0.22) or density of the bone core (p=0.45). Bone cores derived from the Sp1 heterozygotes also had significantly reduced inorganic content (p=0.001) when compared with common homozygotes. A significant effect of density of bone core on yield strength was also reported in this study (p<0.0001).

This study is the first to examine the role of all three *COL1A1* polymorphisms and haplotypes in relation to bone mechanical properties including yield strength and stiffness. The Sp1 T allele ("s" allele) was associated with reduced yield strength (p<0.001), which is consistent with the results of a previous study (Mann V et al, 2001), but the association repeated here was stronger due to the larger sample size (n=94). No association between genotypes and density of bone core was observed in agreement with Mann's study. A possible explanation is that the density of the bone core in both studies is the density of the cancellous structure itself calculated by dividing the mass by the gross sample volume, which differs from the bone mineral density (BMD) that is used in most genetic association studies. In contrast to Mann's study, bone cores derived from the Sp1 G/G homozygotes ("SS") showed significantly increased bone volume and stiffness (p<0.001 for both) when compared with heterozygotes and rare

homozygotes. These differences could be due to the smaller sample size (17 females and 6 males) in the previous study.

For the two promoter polymorphisms, significant associations between genotypes and yield strength and stiffness were observed in present study. The -1997G/G homozygotes had increased yield strength and stiffness when compared with -1997G/T heterozygotes and -1997T/T homozygotes. There were also significant associations between the -1663insT/insT genotype and increased yield strength ($p<0.001$) and stiffness ($p=0.02$) compared with heterozygotes and rare homozygotes.

The associations between haplotypes defined by all three polymorphisms and bone quality were also very strong. Homozygotes of haplotype 1 containing three common alleles (-1997G/ -1663InT/ Sp1G) had significantly increased yield strength and stiffness when compared with carriers of haplotype 6 containing the three rare alleles (-1997T/ -1663DelT/ Sp1T). All three polymorphisms and haplotypes 1 & 6 were identified as independent predictors for yield strength.

Interestingly, the frequencies of some haplotypes were dramatically different from those in the general population (APOSS), although the genotype frequencies of all three polymorphisms did not show any difference between the two populations. For example, haplotype2 (-1997G/ -1663DelT/ Sp1 T, G7T) was the second common in APOSS (17%), but was very rare in bone cores samples (1.53%). The frequency of haplotype 3 (G8T) was 12.21% in APOSS but 1.02% in bone core samples. In contrast, haplotype 6 (T7T) was the second common in the bone core subjects (12.24%), which is 401 times more frequent than APOSS (0.03%). This is probably because the bone cores were

derived from a highly selected group (patients suffering from osteoporotic hip fractures that underwent hip replacement surgery) whereas the APOSS population comprised younger, normal healthy women of average age 48.4±2.3. Therefore the APOSS population is not the ideal control group for this fracture group, an age-matched control group is needed to confirm if the difference is due to age or fracture prevalence.

In summary, all three polymorphisms and haplotypes (1 & 6) were associated with yield strength and stiffness, and predicted yield strength significantly. This study therefore provides evidence to suggest that three *COL1A1* common variants that predispose to osteoporosis do so at least in part by determining bone quality and bone strength.

CHAPTER SIX

Association between *COL1A1* 5' Flank Polymorphisms and Fractures

Abstract

Previous studies have indicated that the 5' flank *COL1A1* polymorphisms are functional variants that influence BMD at the lumbar spine and the femoral neck, DNA protein binding, reporter gene expression and bone strength. The Sp1 polymorphism has been previously reported to be associated with osteoporotic fractures, particularly vertebral fracture, in several meta-analyse. However, there is no report about the association between the promoter polymorphisms, or haplotypes and fractures. Considering the fact that the three polymorphisms are in high linkage disequilibrium and interact with each other to regulate BMD, the aim of this chapter was to investigate the association of the three *COL1A1* 5'flank polymorphisms, haplotypes and osteoporotic fractures. The fracture groups comprised 303 female hip fracture patients and 151 female vertebral fracture patients, and 145 controls for hip fracture patients and 143 controls for vertebral fracture patients. There was no association between any of the individual SNP or haplotypes and vertebral fracture or hip fracture. Another interesting finding to emerge from this study was that allele frequencies and haplotype frequencies of *COL1A1* 5' flank polymorphisms altered with age and morbidity, which indicated that they are associated with mortality.

Introduction

The major clinical importance of osteoporosis is due to the fact that it is associated with an increased risk of fractures, most commonly affecting the hip, spine, and distal forearm. In the UK, the lifetime risk of symptomatic fracture for a 50 year old white woman has been estimated as 13% for the forearm, 11% for the vertebrae and 14% for the femoral neck. Fracture is a very complex phenotype which depends not only on fall-related factors but also on factors which influence bone-strength. Skeletal determinants of fracture risk include BMD, bone geometry, microarchitecture, bone turnover and the material properties of bone. Most of these factors are subject to strong genetic influences. For example, twin studies have suggested that 60-85% of the variance in BMD is genetically determined (Pocock NA *et al*, 1987 & Krall EA 1993). The genetic contribution to osteoporotic fracture risk has also been studied. Family history of hip fracture has consistently been shown to be a risk factor for fracture, independent of BMD in population-based studies (Seeman E *et al*, 1994; Cummings SR *et al*, 1995). Some family and twin studies showed the heritability of fracture was largely independent of BMD (Deng HW *et al*, 2000a; Andrew T *et al*, 2005). This indicated that predisposition may have been mediated through genetic influences on bone turnover, and bone geometry or other factors such as the risk of falling. Moreover, environmental factors become more important with age. For example, the study on a large cohort of 6021 Swedish twins has shown that the heritability of overall age-adjusted fracture was greater for first hip fractures before the age of 69 years (0.68; 95% CI, 0.41-0.78) and between 69 and 79 years (0.47; 95% CI, 0.04-0.62), but it dropped off rapidly with age to almost zero after 79 years old (0.03; 95% CI, 0.00-0.26) (Michaelsson K *et al*, 2005).

Among all candidate gene polymorphisms investigated in relation to osteoporosis so far, the *COL1A1* Sp1 polymorphism affecting a Sp1 binding site in the first intron is one of few validated functional polymorphisms which have been associated with BMD and fractures. In chapter 3, I found that two promoter polymorphisms were in strong linkage disequilibrium with the Sp1 polymorphism, and the *COL1A1* 5'flank haplotypes had stronger effects on BMD. The haplotypes were also found to influence reporter gene transcription (chapter 4) and bone strength in samples derived from hip fracture patients *ex vivo* (chapter 5). Therefore, the aim of this chapter is to study the associations of the three *COL1A1* 5'flank polymorphisms and haplotypes with fragility fracture.

Methods

There were two study groups in this chapter: one comprised 303 female hip fracture patients (average age 75.4±7.4 years) who were admitted to Aberdeen Royal Infirmary with osteoporotic hip fracture resulting from low-energy trauma and matched 145 female controls (average age 78.8 ± 8.6 years) recruited from patients admitted the same hospital who had fallen but not sustained a hip fracture. The other group comprised 151 female clinical vertebral fracture patients (average age 72.5 ± 6.5 years) and matched 143 controls (average age 71.1±5.6 years) also from the north east Scotland who had responded to an advertisement for volunteers to take part in a study of treatment for osteoporosis.

DNA was extracted and the genotypes for the three polymorphisms were determined by PCR and DNA sequencing methods. Haplotypes were constructed by the 'PHASE' program, as described in chapter 2. The probability threshold at which haplotype phases are called as "correct" is 90%.

This study had 80% power to detect a 10% difference in allele proportions (OR=1.5) in hip fracture (N=303) and control groups (N=145) based on the allele frequencies of the -1663delT and the Sp1T in normal Scottish women (20%, chapter 3). The power to detect the same effect in vertebral fracture (N=151) and control groups (N=143) was less than 60% (Minitab).

Results

Genotype Frequencies in Hip Fracture-control Groups

Genotypes for the -1997 G/T polymorphism and the Sp1 polymorphism were in Hardy-Weinberg equilibrium (HWE) in both hip fracture and control groups, but genotypes for the -1663in/delT polymorphism deviated significantly from HWE (p=0.03) due to under-representation of homozygotes for the -1663delT allele and over representation of heterozygotes in the hip fracture group. But HWE of -1663in/delT genotypes was observed in the combined samples (hip fracture and control groups). As shown in Table 6.1, there was no significant association of any genotype with hip fracture.

Table 6.1 Genotype frequencies in hip fracture-control groups

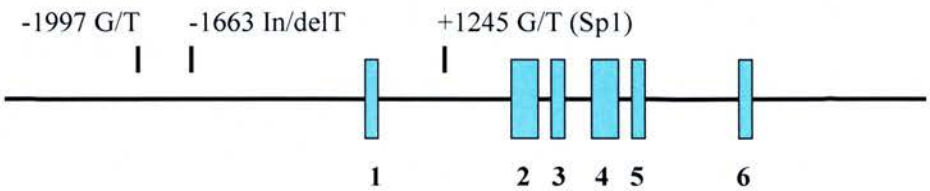
Genotype		**Hip fracture (%)** N=303	**Hip controls** (%) N=145	**p**
-1997G/T	G/G	242 (79.87)	123 (84.8)	0.18
	G/T	59 (19.47)	22 (15.2)	
	T/T	2 (0.66)	0	
-1663In/delT	In/inT	179 (59.08)	83 (57.2)	0.31
	In/delT	116 (38.28)	52 (35.9)	
	Del/delT	8 (2.64)	10 (6.9)	
Sp1G/T	G/G	192 (63.37)	85 (58.6)	0.67
	G/T	98 (32.34)	57 (39.3)	
	T/T	13 (4.29)	3 (2.1)	

P: χ^2 test of allele frequencies between hip fracture and control groups

Association of Haplotype Frequencies in Hip Fracture-Control Groups

Five haplotypes with minor-allele frequency (MAF) >1% were observed and these accounted for more than 90% of alleles in the hip fracture and control groups. As shown in Figure 6.1, the most common haplotype 1 (-1997*G*/-1663InT/sp1G, 'G8G') accounted for more than 45% in both groups. There was no significant difference in haplotype frequencies between hip fracture and control groups (χ^2=2.87, p=0.58).

Figure 6.1 Common haplotypes and frequencies in hip fracture-control group



		Hip fracture (%) (N=606)	Control (%) (N=290)
G - InsT ——— G	H1 (G8G)	299 (49.34)	136 (46.9)
G - DelT ——— T	H2 (G7T)	8 (1.32)	3 (1.03)
T - InsT ——— G	H3 (T8G)	59 (9.74)	22 (7.59)
G - InsT ——— T	H4 (G8T)	112 (18.48)	60 (20.69)
G - DelT ——— G	H5 (G7G)	124 (20.46)	69 (23.79)

Common haplotypes at the 5' flank of the *COLIA1* gene and their frequencies in hip fracture-control groups are shown in relation to the *COLIA1* gene structure. Shaded boxes represent exons. N represents the copy number of chromosomes, calculated by multiplying the total number of the population by 2.

Genotype Frequencies in Vertebral Fracture-Control Groups

The genotypes for all three polymorphisms were in Hardy-Weinberg Equilibrium ($p>0.05$). As shown in Table 6.2, there was no significant association between any genotype and vertebral fracture.

Table 6.2 Genotype frequencies in vertebral fracture-control groups

Genotype		Vertebral fracture (N=151, %)	Controls (N=143,%)	P*
-1997G/T	G/G	114 (75.50)	106 (74.13)	0.47
	G/T	32 (21.19)	35 (24.48)	
	T/T	5 (3.31)	2 (1.40)	
-1663indelT	In/inT	95 (62.91)	95 (66.43)	0.69
	In/delT	52 (34.44)	43 (30.07)	
	Del/delT	4 (2.65)	5 (3.50)	
Sp1G/T	G/G	101 (66.89)	95 (66.43)	0.99
	G/T	46 (30.46)	44 (30.77)	
	T/T	4 (2.65)	4 (2.80)	

P*: χ^2 test.

Haplotype Frequencies in Vertebral Fracture and Control Groups

Four haplotypes with minor-allele frequency (MAF) >1% were observed, and they accounted for more than 99% of alleles in the vertebral fracture and control groups. As shown in table 6.3, the most common haplotype 1 (-1997*G*/-1663*InT*/sp1*G*, 'G8G') accounted for more than 65% of both groups. There was no significant difference in haplotype frequencies between vertebral fracture and control groups (χ^2=2.23, p=0.53).

Table 6.3 Haplotype frequencies in vertebral fracture-control groups

	Vertebral fracture (%) (N=303)	Control (%) (N=286)
H1 (G8G)	198 (65.56)	194 (67.83)
H2 (G7T)	52 (17.22)	50 (17.48)
H3 (T8G)	42 (13.91)	37 (12.94)
H5 (G7G)	8 (2.65)	3 (1.05)

N, copy number of the chromosomes, calculated by multiplying the total number of the population by 2.

Allele and Haplotype Frequencies Change with Age

The frequency of the -1997G allele was increased from 86.23% in the APOSS population (average age =48.5 ± 2.4 years) to 92.4% in the hip fracture controls (average age=78.8 ± 8.6 years). The frequency of -1997T allele was correspondingly reduced in elderly hip fracture controls, compared with APOSS (p=0.003). The frequency of the -1663inT allele was decreased from 80.4% in the APOSS population to 75.2% in the hip fracture controls, and the -1663delT allele was correspondingly enriched in the hip fracture controls compared with APOSS (p=0.029). The frequency of the Sp1 genotype showed a similar change as the -1663in/delT genotype, but it did not reach statistical significance (p=0.176).

Table 6.4 Allele frequencies altered with age

Allele	**APOSS (%)** (48.5±2.4 yr)	**Vertebral fracture controls** (%) (71.1±5.6 yr)	**Hip fracture controls** (%) (78.8±8.6 yr)	**P**
-1997G	5193 (86.23)	247 (86.37)	268 (92.4)	0.003
-1997T	829 (13.77)	39 (13.64)	22 (7.6)	
-1663InT	4803 (80.40)	233 (81.47)	218 (75.2)	0.029
-1663DelT	1171 (19.60)	53 (18.53)	72 (24.8)	
Sp1G	5253 (81.44)	234 (81.82)	227 (78.3)	0.176
Sp1T	1197 (18.56)	52 (18.18)	63 (21.7)	

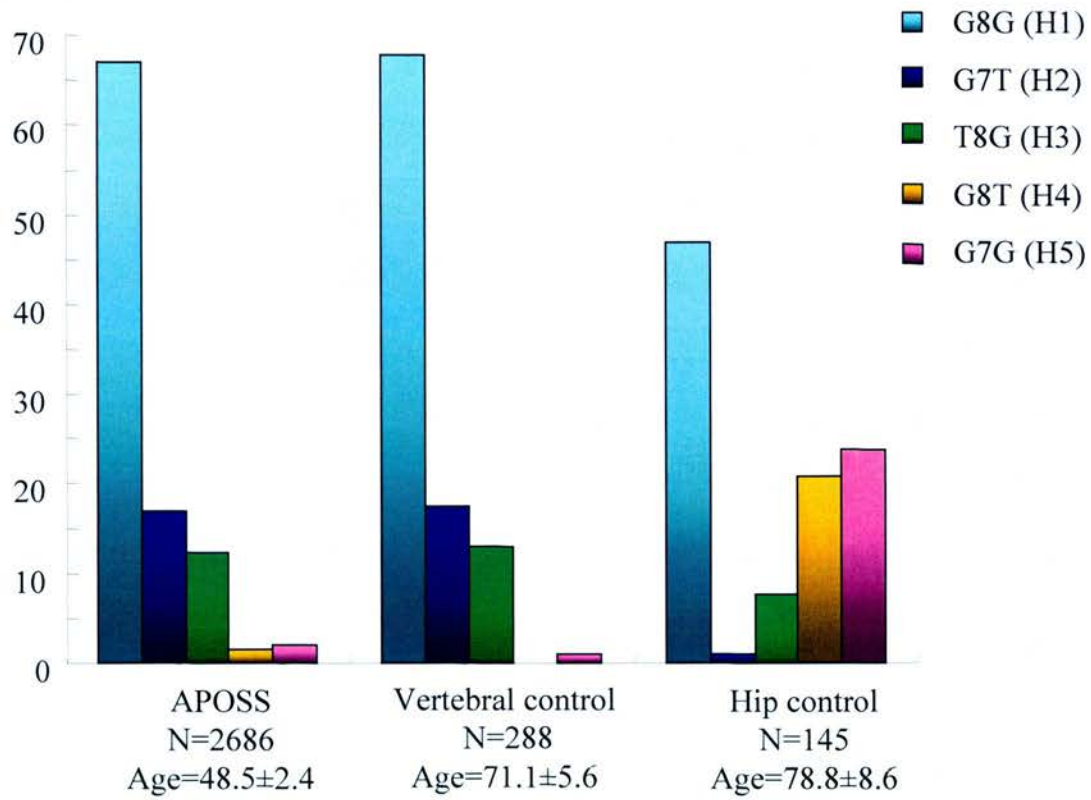
P, χ^2 –test between APOSS and hip control

When haplotype frequencies were compared among the hip fracture controls, the vertebral fracture controls and the general population (APOSS), I found that the frequencies for some haplotypes differed significantly in hip fracture controls and APOSS (Figure 6.2). Two haplotypes (4 & 5) had higher frequencies in the hip fracture controls than the APOSS population (P<0.001). Instead, the other three

haplotypes (1-3) were less common in hip fracture controls compared with APOSS. However, there was no significant difference in haplotype frequencies between vertebral fracture controls and APOSS (χ^2=1.53, p=0.68).

Figure 6.2 Haplotype frequencies differed in elderly controls and APOSS

	APOSS (%) (48.5±2.4 yr)	**Vertebral fracture control** (%) (71.1±5.6 yr)	**Hip fracture control (%)** (78.8±8.6 yr)
G8G(1)	4396 (67.05)	194 (67.83)	136 (46.9)
G7T(2)	1114 (17)	50 (17.48)	3 (1.03)
T8G(3)	800 (12.21)	37 (12.94)	22 (7.59)
G8T(4)	102 (1.56)	0	60 (20.69)
G7G(5)	135 (2.06)	3 (1.05)	69 (23.79)



The frequencies of haplotypes 1-3 (H1, H2 & H3) were decreased in elderly controls (hip control). In particular, the frequency of haplotype 2 dropped from 17% in APOSS to 1.03% in hip controls. Conversely, the frequencies of haplotypes 4-5 were increased from ~2% in APOSS to more than 20% in the hip fracture controls.

Discussion

The *COL1A1* gene is one of the most important functional candidate genes for osteoporosis due to the fact that type I collagen is the most abundant protein of bone. The Sp1 polymorphism of the *COL1A1* gene has been widely studied in relation to BMD and fractures. It has been considered a functional polymorphism because *in vitro* studies suggested that the rare 'T' allele is associated with more *COL1A1* transcript and protein product (Mann V *et al*, 2001). Two promoter polymorphisms (-1663In/delT & -1997G/T) were first identified in a Spanish population and have been associated with BMD in several other populations (Garcia-Giralt N *et al*, 2002; Liu PY *et al*, 2004; Yamada Y *et al*, 2005; Zhang YY *et al*, 2005 & Stewart TL *et al*, 2006). More recently, the Rotterdam study showed that the Sp1T allele was associated with a 2.3 times increased risk of fragility fracture (95%CI 1.4-3.9, P=0.001), and haplotype2 (-1997G/Sp1T) was associated with a 2.1 fold increased risk of fragility fracture in women (95%CI 1.2-3.7, P=0.001) (Yazdanpanah N *et al*, 2007). However, two promoter polymorphisms and the Sp1 polymorphism have not been studied together in the form of haplotypes in relation to osteoporotic fractures.

The present study of two promoter polymorphisms in relation to osteoporotic fractures is the first to be performed and also the first to report the results of haplotype analysis taking into account alleles at the Sp1 polymorphic site. In chapter 5, by performing biomechanical studies *ex vivo* in bone samples from hip fracture patients, I found that alleles -1997T, -1663delT and Sp1T were associated with reduced bone strength. Moreover, strong associations were found when subjects were categorised by haplotypes. The rare haplotype 6 '-1997T/-1663delT/Sp1T' was dramatically enriched in bone samples from hip fracture patients (12.24%), and was 401 more frequent than the general population (APOSS, 0.03%). This rare haplotype was

associated with reduced bone strength in hip fracture bone samples ($p<0.001$), but it was totally absent in the present study of fracture and control subjects. One of the poassible explanations is that the bone samples used in biomechanical test were removed from the patients who were admitted to Aberdeen Royal Infirmary and underwent hip replacement surgery as a result of osteoporotic hip fracture. This is a highly selected group with the age at fracture ranges from 58 to 101 years old with an average of 81.7 years, which is different from the hip fracture patients used in this association study.

In this case control study, the genotype frequency of the heterozygote -1663in/delT was significantly different between the hip fracture group and APOSS study ($p=0.017$). This was mainly due to a departure from Hardy-Weinberg Equilibrium caused by an over representation of the heterozygote -1663in/delT genotype in the hip fracture group. Possible reasons for the apparent HWE deviation include: genotyping error, random sampling error and because one genotype is predisposing to disease and therefore will be more frequent in subjects ascertained for that disease. To exclude the genotyping error, all the genotypes of three polymorphisms were repeated in duplicate. Since the genotype frequencies of the other two polymorphisms were not out of HWE and all the subjects were recruited from North East Scotland, random sampling error is not likely. Considering that only the hip fracture group showed a departure from HWE of the -1663ins/delT genotype, it may indicate an effect of the genotypic disease association or the functional nature of this polymorphism.

However, none of the genotypes or haplotypes was associated with increased risk of vertebral fracture or hip fracture in this study. It may be under-powered to detect the

modest effects that are expected from individual polymorphism on a complex disorder such as osteoporotic fracture in light of the fact that fracture is a very complex phenotype involving many risk factors and determined by multiple genes with minor effects. For example, the most recent meta-analysis involving 20,786 individuals from several European countries reported a non-significant trend for association of the Sp1 genotypes with vertebral fracture and a nominally significant association with incident vertebral fractures in females (OR=1.33 [95% CI, 1.00 to 1.77]; p=0.05). The study presented here only had about 50% power to detect the same effect of the Sp1T allele on hip fracture (N=303) and 30% on vertebral fracture (N=151) based on the frequency of Sp1T allele in the general population (20%, APOSS). Power was also low to detect similar effects of the -1997G/T and the -1663in/delT polymorphisms. Therefore, further studies in larger populations are required to detect the subtle effects of individual polymorphism on fracture.

Another interesting finding to emerge from this study was that the frequencies of genotypes and haplotypes differed markedly with age. For example, the frequencies of -1997G and -1663delT alleles were significantly increased in elderly hip fracture controls compared with those in the general population APOSS. The frequencies of haplotypes 4 & 5 were also significantly increased in elderly hip controls compared with those in APOSS. Conversely, the frequencies of haplotypes 1-3 were less common in elderly hip fracture controls. There was also a marked difference in haplotype frequencies in the bone core samples (chapter 5) compared with the general population.

Taken together, these observations suggest that *COL1A1* haplotypes might be associated with morbidity or mortality in the general population. There is evidence

that the Sp1T allele is associated with lower arterial compliance which may reflect the effects of the polymorphism on collagen type I composition of blood vessels (Brull DJ *et al*, 2001b).. Reduced arterial compliance precedes the development of hypertension which in turn is associated with the development and progression of atherosclerosis and increased cardiovascular risk (Noma T *et al*, 1999).

Another possible explanation is that the elderly controls for hip fracture were recruited from patients admitted to Aberdeen Royal Infirmary (ARI) with a fall who had not sustained a hip fracture, whereas the controls for vertebral fractures and APOSS population were randomly selected populations of normal healthy women.

In summary, there was no association of any genotype or haplotype with osteoporotic fracture in this study. But the frequencies of 5' flank *COL1A1* genotypes and haplotypes vary among populations, raising the possibility that *COL1A1* alleles are associated with morbidity or longevity, although further study will be needed to investigate this.

CHAPTER SEVEN

Towards Generation of An Animal Model for Osteoporosis Mediated by *COL1A1* Alleles

Abstract

The previous chapters have already revealed the possible mechanisms by which the three polymorphisms at 5' flank of *COL1A1* gene predispose to osteoporosis including regulating BMD and bone strength. I also found these polymorphisms have functional effects on gene transcription using promoter-reporter assays and affect DNA-nuclear protein binding in gel shift assays. To further investigate the function of these polymorphisms *in vivo*, the aim of the present study was to create an animal model of osteoporosis associated with the *COL1A1* haplotypes by gene targeting. Two targeting constructs containing 17.7 kb humanized mouse *Col1a1* sequence were generated by cloning and were transfected into mouse 129 embryo stem cells (ES) by electroporation. Recombination events were detected by long-PCR amplification of DNA extracted from ES cells which survived the G418 seletion. After three targeting attempts, 45 ES cell clones survived but none had the correct homologous recombination.

Introduction

In the previous chapters, I found that three polymorphisms at 5' flank of *COL1A1* gene interacted with each other to regulate BMD at lumbar spine and femoral neck in a peri-menopausal Scottish women population. The haplotypes defined by these polymorphisms also have strong effects on BMD: haplotype 2 (-1997G/-1663delT/Sp1T, "G7T") was associated with reduced BMD (p=0.007 at lumbar spine & p=0.008 at femoral neck) while haplotype 3 (-1997T/-1663inT/Sp1G, "T8G") was associated with increased BMD (p=0.007 at lumbar spine). The mechanisms of these associations are still not clear. The biomechanical tests in chapter 5 showed that the -1997 T, -1663 delT (7T) and Sp1 T alleles were associated with reduced bone strength ($p<0.001$), and haplotype 6 (T7T) was also associated with reduced bone strength *ex vivo* bone samples from hip fracture patients ($p<0.001$) with 401 times enrichment of its frequency in these bone samples. The -1663delT and -1997T alleles were found to have greater binding affinity for nuclear proteins of osteoblasts compared to their common alleles in gel shift assays (chapter 4). Furthermore, promoter-reporter assays in chapter 4 showed that haplotype 2 and 3 had higher transcriptional activities than other haplotypes for reporter gene ($p<0.001$).

Therefore, three 5' flank polymorphisms predispose to osteoporosis by the regulation of BMD and bone quality. Previous studies suggested that the Sp1 polymorphism predisposes to osteoporosis by increasing the ratio of collagen α1(I) to α2(I) mRNA in "Ss" heterozygotes (2.3 to 1), leading to abnormal protein production (homotrimer [$\alpha1(I)_3$]). This hypothesis is supported by the previous studies which have shown homotrimer [$\alpha1(I)_3$] is found in mouse fetal tissues and material derived from tumors and chronic fibrotic conditions and associated with impaired mechanical strength of

bone (Chipman SD *et al*, 1993 & McBride DJ *et al*, 1998). However, its presence has not been observed yet in human tissue due to the limited amounts of patient material available. The aim of this chapter was to investigate the molecular mechanisms by which the *COL1A1* haplotypes affect bone metabolism *in vivo* by generating a mouse model of osteoporosis associated with the polymorphisms by gene targeting.

Methods

Construction of Targeting Vectors

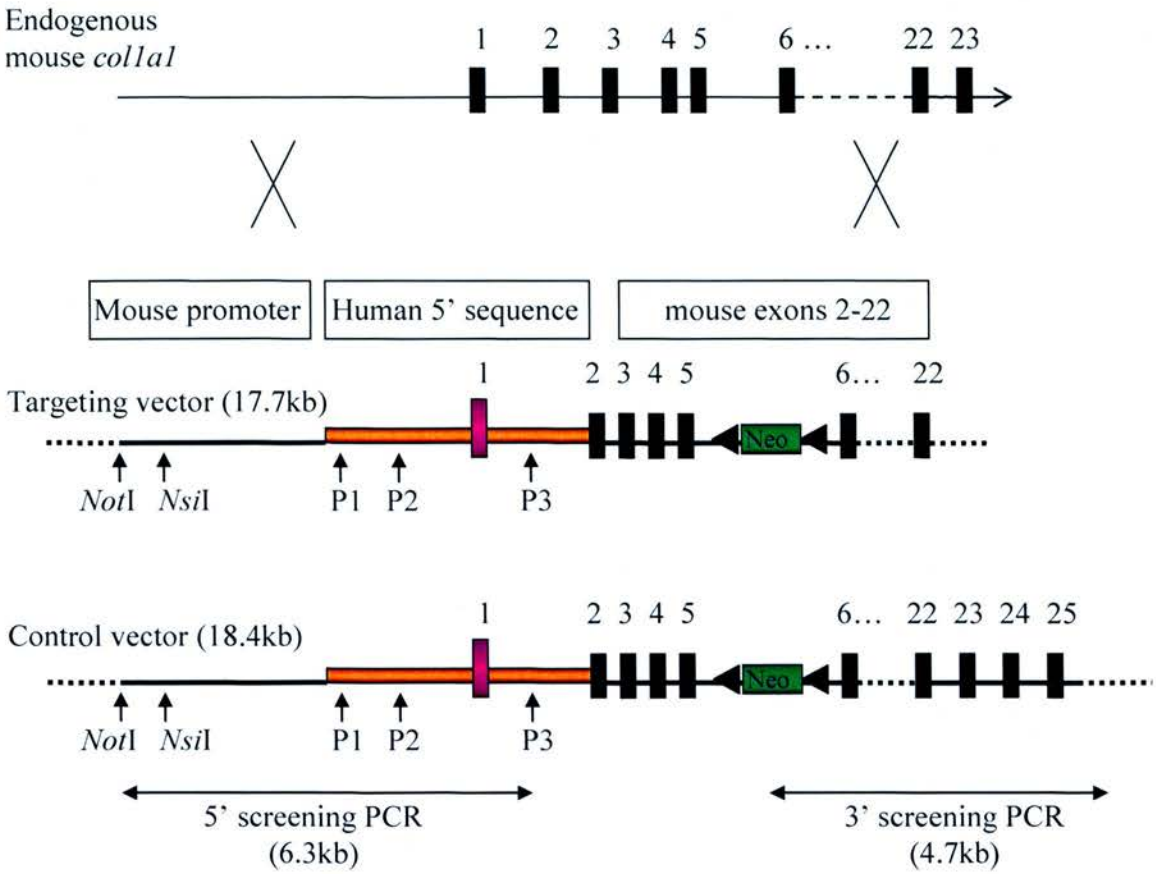
Targeting constructs containing 4.6kb of human *COL1A1* 5' sequence (-2104 to +1537; promoter, exon1 and intron1) were used to replace the proximal mouse *Col1a1* 5' sequence since the three polymorphisms of interest are absent from the mouse *Col1a1* gene. A neomycin-resistance gene (neo) was flanked by two loxp sites and the whole loxp-neo-loxp cassette was inserted into intron 5 of the mouse *Col1a1* gene (Figure 7.1).

Two targeting constructs were made corresponding to haplotypes 2 & 3. The base changes at each of the three polymorphic sites were introduced by a *Pfu* DNA polymerase-based mutagenesis with the Quick-Change site-directed mutagenesis kit (Strategene, UK). The same method was used to create restriction sites for cloning and to correct the PCR errors in the targeting constructs. The desired changes were confirmed by DNA sequence analysis.

To screen the recombination events at both 5' and 3' ends after targeting experiments, a control vector (18.4kb) was also constructed with the extended upstream sequence on mouse *Col1a1* 5' homologous sequence (-5320 to -1352) and extended downstream sequence on 3' homologous sequence (+1573 to +7590) (Figure 7.1).

All the details were described in chapter 2.

Figure 7.1 Schematic representation of endogenous mouse *Col1a1* gene, targeting vector and control vector.



This targeting event will replace the proximal promoter of the mouse *Col1a1* gene with the human *COL1A1* promoter, exon1 and intron1. Loxp-Neo-loxp cassette serves as a positive selection marker. A control vector with two extended homologous ends was generated to screen recombination events at both homologous ends using long-PCR reactions. The black boxes represent mouse *Col1a1* exons. The gold box represents human *COL1A1* promoter and intron1. The purple box represents human *COL1A1* exon1. P1, P2 & P3 represent three polymorphisms: -1997G/T, -1663ins/delT & sp1G/T, respectively. *Not*I and *Nsi*I site were used to linearize the targeting for the targeting experiments.

ES cell Targeting Method

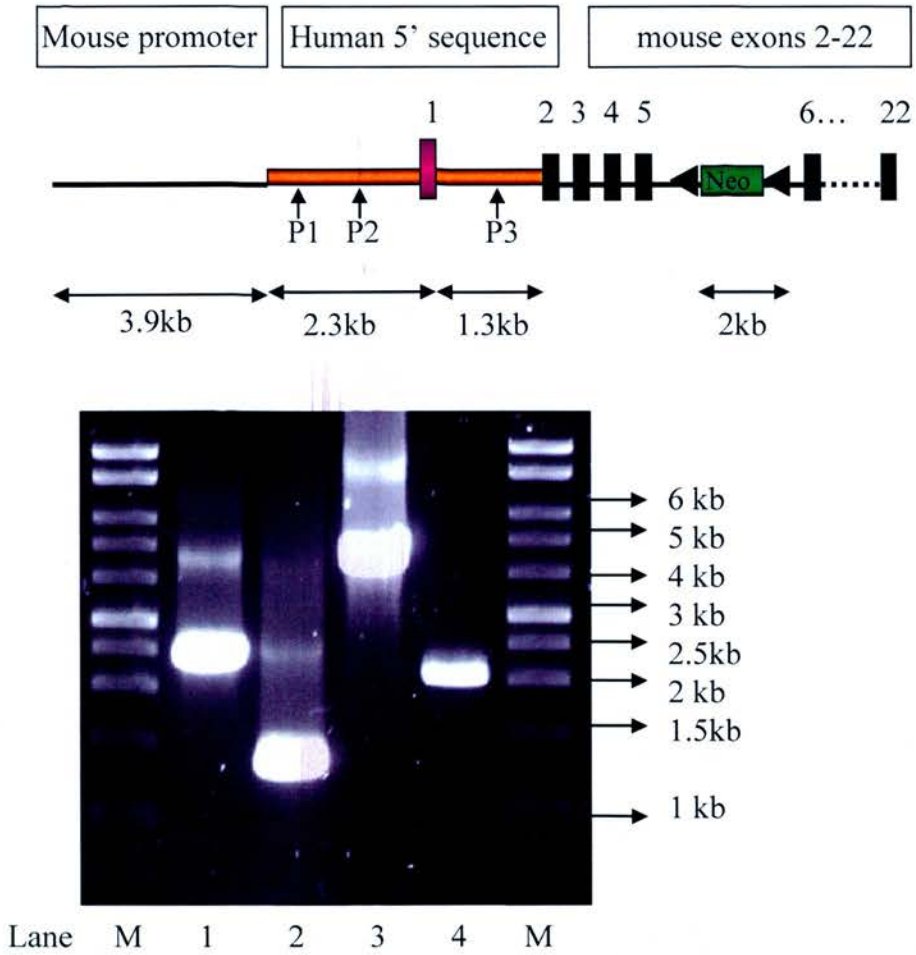
The two targeting vectors were purified and lineared using the restriction enzymes *Not*I and *Nsi*I. The linear DNA (16.6 kb) was then extracted using Elutrap Electro-seperation system (Schleicher & Schuell), and 100 μg purified linear DNA was used for electroporation. After the electroporation, ES cells were seeded at 10^6 per gelatinised petri plate (10 plates for targeted cells and 4 plates for control cells) and incubated at 37°C overnight before adding G418 (175μg/ml) for selection of recombinant colonies. The targeting was repeated three times under similar conditions by Fiona Kilanowski in Human Genetic Unit, MRC.. G418 resistant clones were picked, DNA was extracted and screened by myself using two long-PCR reactions to detect the recombinant integrations at both 5' and 3' end. The PCR conditions were optimized using the 18.4 kb control vector at different concentrations prior to the targeting. Details were described in chapter 2.

Results

Generation of Targeting Vectors by Long-PCR

Four PCR reactions were optimized to produce strong signals of 2.3 kb human *COL1A1* promoter and exon1, 1.3 kb human *COL1A1* intron1, 3.9 kb mouse *Col1a1* promoter and 2 kb loxp-neo-loxp cassette, respectively (Figure 7.2).

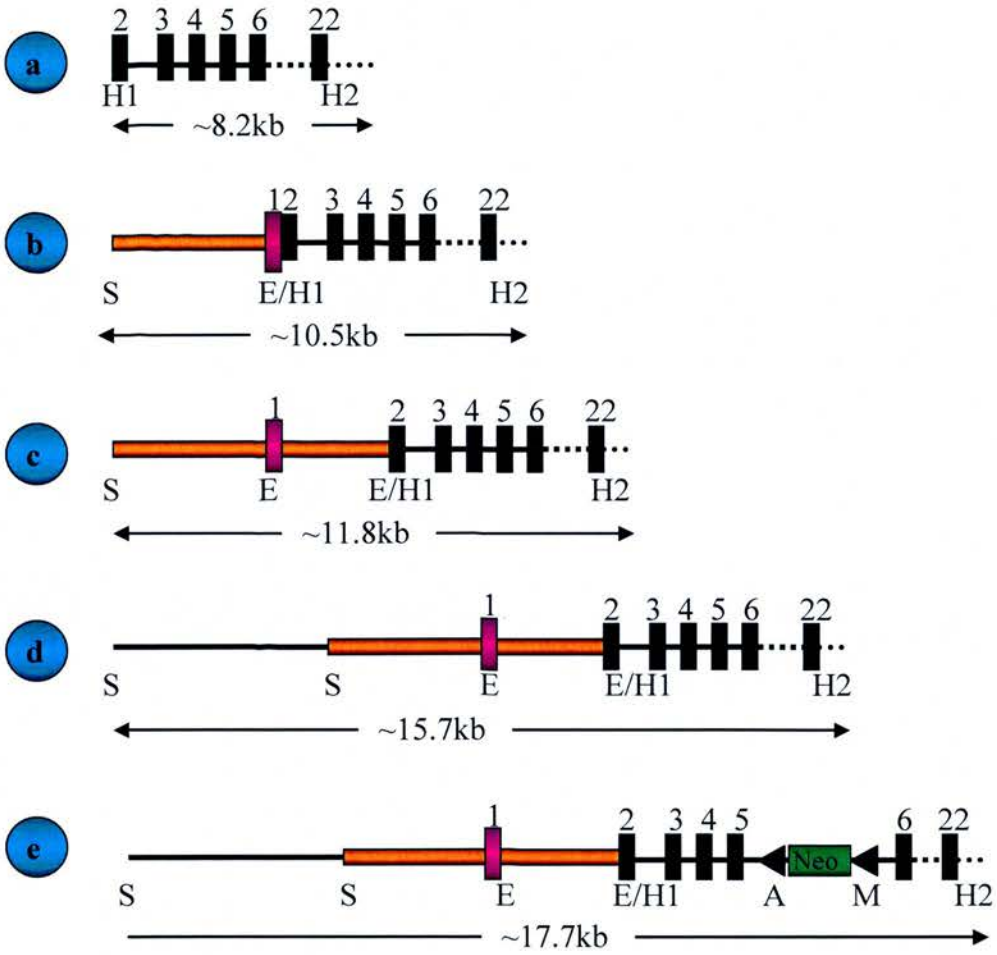
Figure 7.2 DNA fragments were made by long PCR



The top panel shows the targeting construct and the sizes of the fragments. The bottom panel shows the PCR products electrophoresed on a 1% agarose gel along with two 1kb DNA ladders on both sides.
Lane 1: 2.3 kb of human *COL1A1* promoter; lane 2: 1.3 kb of human intron1; lane 3: 3.9 kb of mouse promoter; lane 4: 2 kb of Loxp-Neo-Loxp cassette; M: 1 kb DNA ladder.

All these PCR fragments were subcloned into the pBLuescript SK (-) vector, and all ligations of the subclonings were checked by restriction digestions. After the sequence of each subclone was confirmed by sequencing reactions, DNA fragments were then ligated together to assemble the targeting vector as shown in Figure 7.3.

Figure 7.3 Schematic representation of assembling the targeting vector.

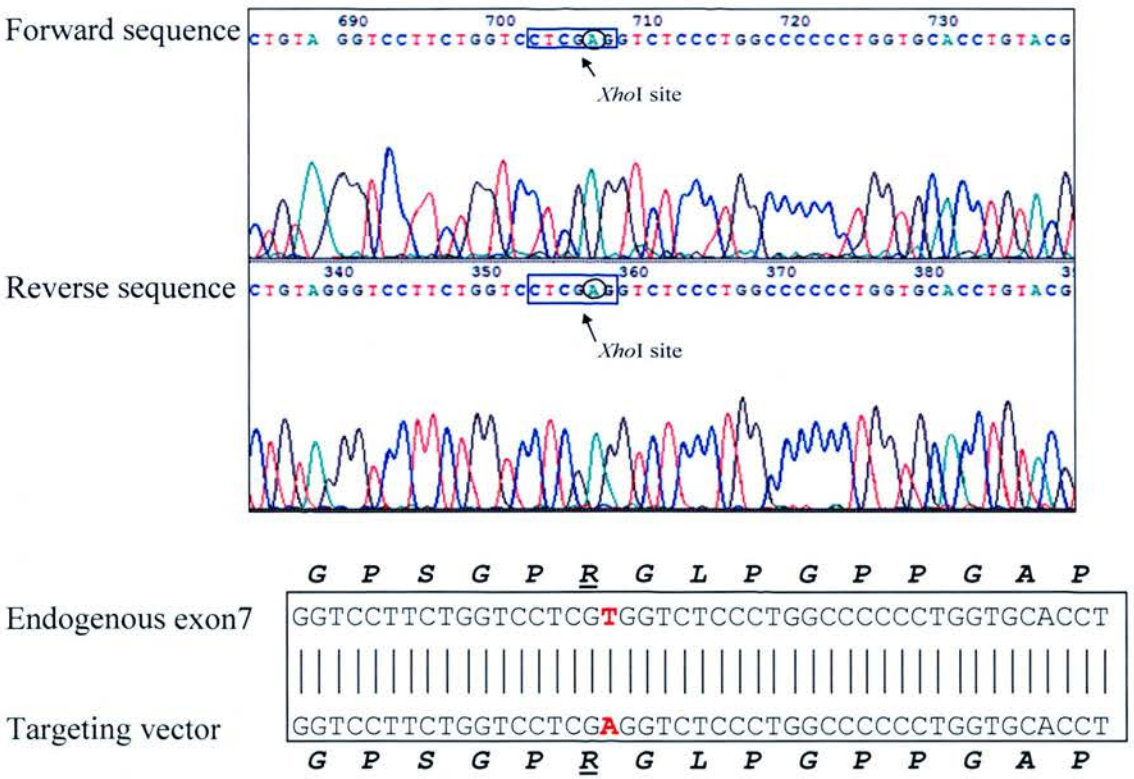


a: 8.2kb mouse *Col1a1* exon2-22 subclone; **b**: 10.5kb vector made by ligation of **a** & 5'human *COL1A1*; **c**: 11.8kb vector made by ligation of **b** & human *COL1A1* intron1; **d**: 15.7kb vector made by ligation of **c** & mouse *Col1a1* promoter; **e**: 17.7kb targeting vector made by ligation of **d** & Loxp-neo-loxp cassette.
Restriction enzymes used in each ligation: S, *Spe*I; E, *EcoR*I; A, *Afl*II; M, *Mfe*I; H1, *Hind*III; H2, *Hinc*II.

Sequencing Results

Two haplotypes created by three Site-directed mutagenesis (SDM) reactions at the human *COL1A1* -1663, -1997 and +2145 polymorphic sites were confirmed by DNA sequence analysis. The completed targeting vectors were sequenced successfully using primers covering the entire vector sequence at least twice. This revealed several base changes relative to the published sequence. A silent point mutation creating an *Xho*I site in mouse *Col1a1* exon7 was confirmed by both forward and reverse sequence reactions, which was introduced by Hormuzdi SG *et al* (1999) for quantification of allele-specific transcripts *in vivo* (Figure 7.4)

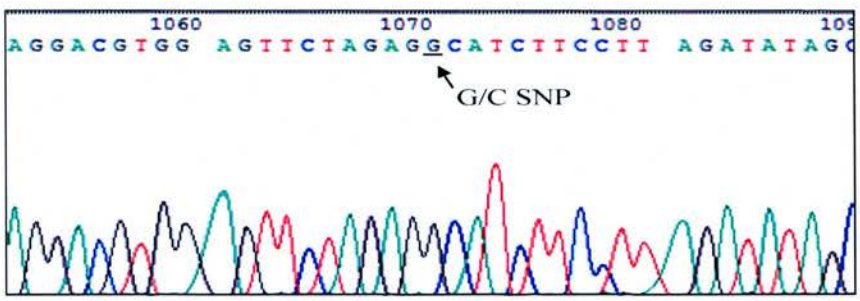
Figure 7.4 A *Xho*I site was introduced in mouse *Col1a1* exon7



This T→A substitution in exon7 resulted in a silent mutation, which would be used for quantification of allele-specific transcripts in vivo.
The *Xho*I site is represented in frame (top); T→A substitution is in red (bottom); amino acids are in **bold** *Italic* (bottom); the silent mutation is underlined.

The second base change was detected within the human *COL1A1* promoter. It was identified as a G/C SNP (rs2269336) with the G allele present in both targeting vectors (Figure 7.5). This is a validated SNP with an allele frequency of 0.838 for "C" and 0.164 for "G" in Central European population (CEU), according to the HapMap genotyping results.

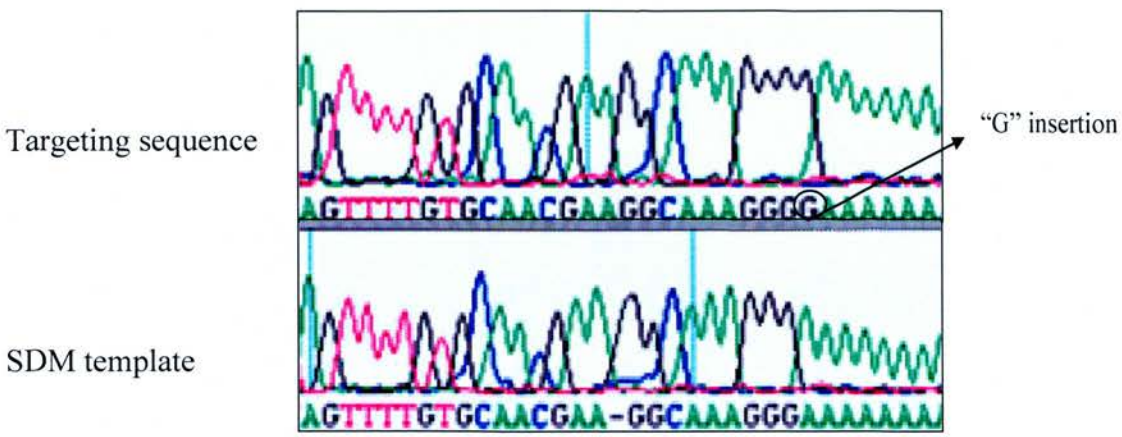
Figure 7.5 A G/C SNP was detected within human *COL1A1* promoter



This is a G/C SNP (rs2269336) with rare G allele present in both targeting vectors. The SNP is underlined.

The third base change (a "G" insertion) was detected at position -1545 of the mouse *Col1a1* promoter (Figure 7.6). It was confirmed as a SDM error after comparing with the SDM template sequence. To correct this error, another SDM was performed with new primers on plasmid DNA without human homologous sequence (chapter 2, *Material and methods*).

Figure 7.6 A "G" insertion introduced during an SDM reaction



Mouse (129/ola) *Col1a1* promoter GGTCTTTTTTTTTTTTCCCTTTGCCTTCGTTGC

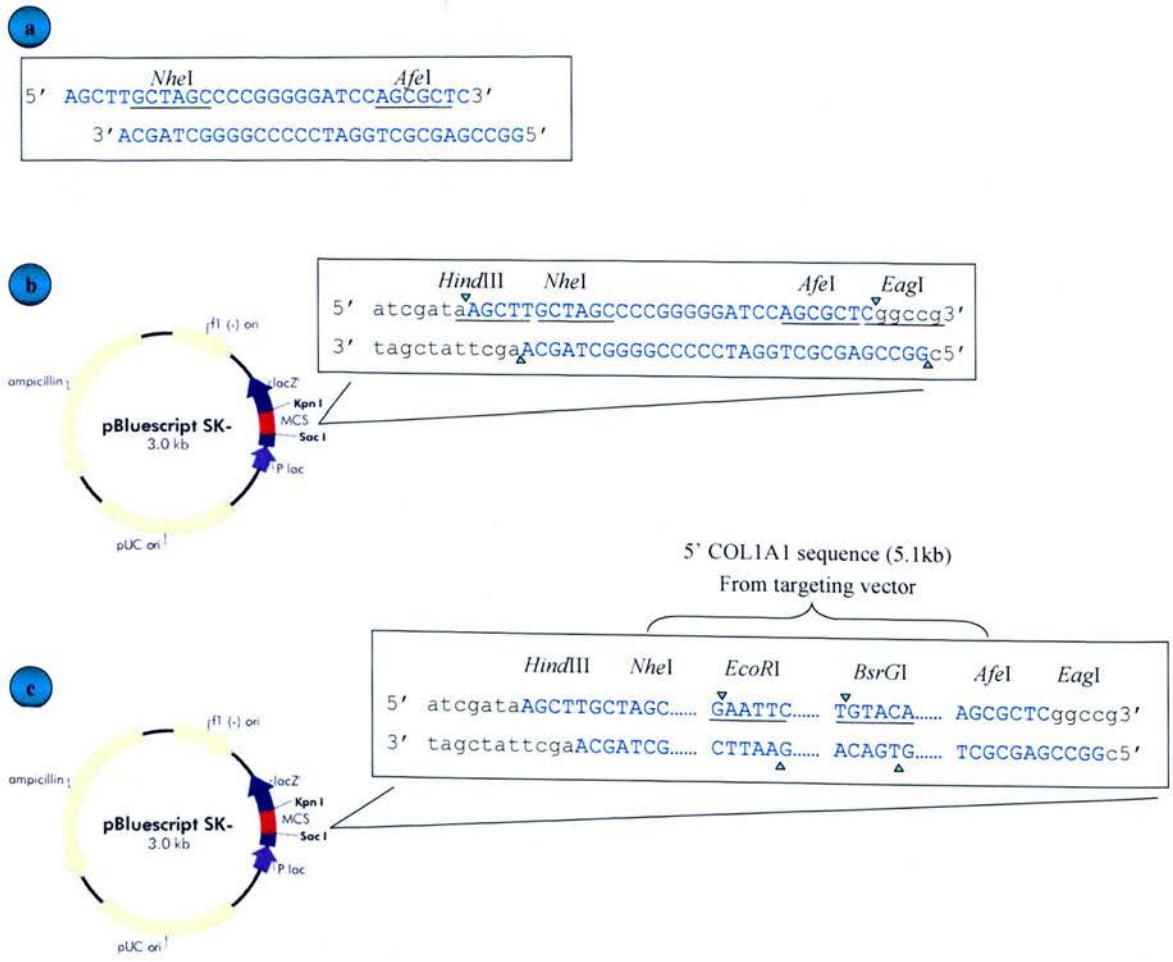
Homologous human sequence CTGTTCCTTTTTTTTCCCCTTTGCCTTCGTTGC

A "G" insertion at position -1545 of mouse *Col1a1* promoter (top) was found in the sequence reactions. This position is highly homologous to human *COL1A1* promoter (-1660 to -1670; bottom).

Finally, the sequencing reactions revealed that there were 21 extra bases of plasmid vector inserted into the human promoter which might be introduced by subcloning. To excise the extra 21bp sequence, the pBlueScript SK (-) vector was first modified to introduce *Nhe*I and *Afe*I restriction sites by inserting 20 base oligonucleotides containing both restriction sites into the multiple cloning site (Figure 7.7).

Secondly, the targeting vector was digested by enzymes *Nhe*I and *Afe*I to release two fragments: a 5.1kb fragment containing the 21 extra bases and a 12.6kb fragment. Both fragments were purified from gel and quantified. The 5.1kb fragment was then ligated with the modified pBlueScript SK (-) vector to make a new vector of about 8kb (Figure 7.7).

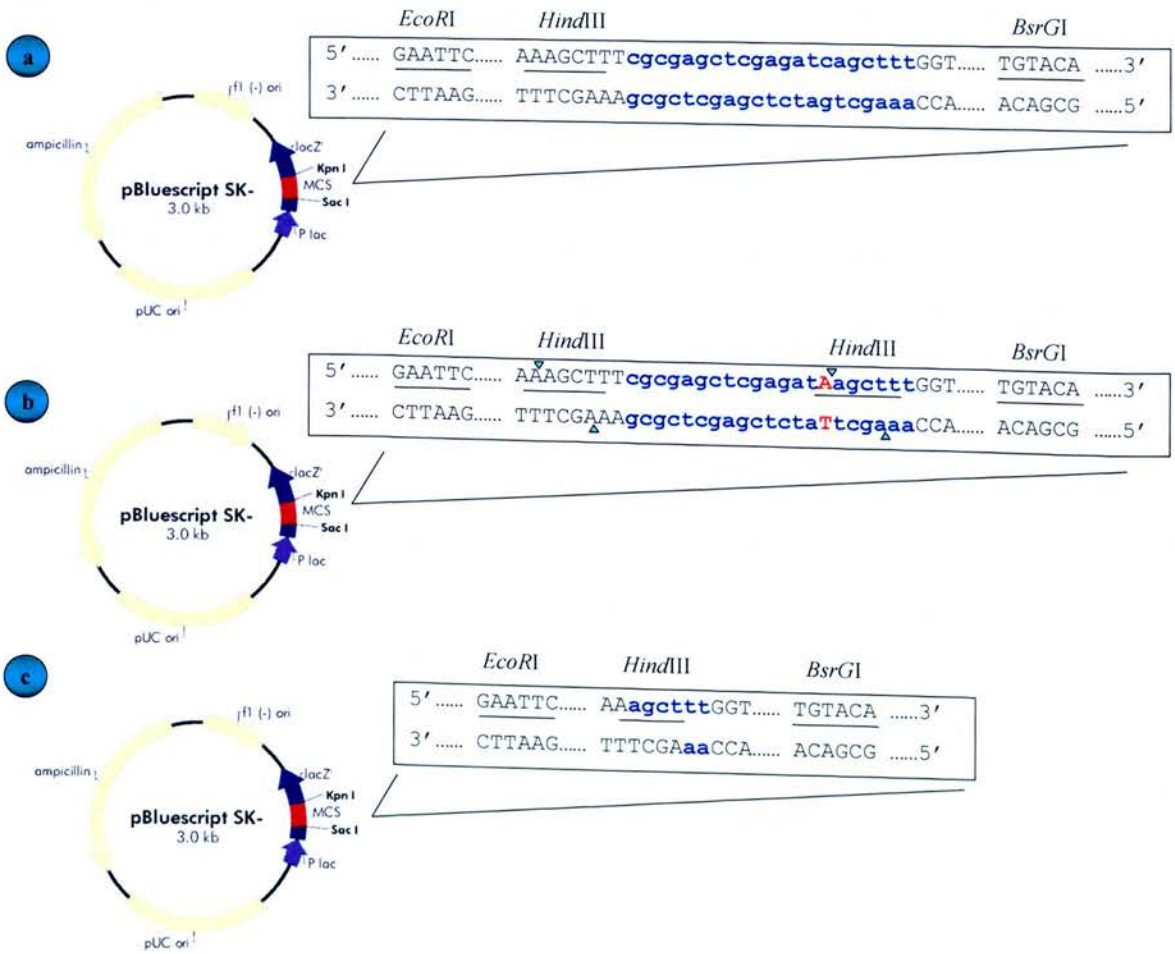
Figure 7.7 Modification of pBluescript SK(-) vector and generation of a new 8kb vector by *Nhe*I and *Afe*I digestion



a: double-stranded oligonucleotides containing *Nhe*I and *Afe*I sites. **b**: pBluescript SK(-) vector was modified by cloning the double-stranded oligos into the multiple cloning site (MCS) by *Hind*III and *Eag*I digestion. **c**: a new 8kb vector was made by NheI and AfeI digestion of targeting vector and modified pBluescript SK(-) vector, which contains 5.1kb 5' COL1A1 sequence including the extra 21 bases. Restriction sites are underlined.

Then a SDM reaction was carried out on a 'human promoter' subclone to create an extra *Hind*III restriction site at the first extra base (Figure 7.8). After SDM reaction, the extra 21 bases were removed by *Hind*III digestion, which was confirmed by sequencing reaction. After that a 1.5kb fragment containing the correct sequence was released from the 'human promoter' subclone by *BsrG*I and *EcoR*I digestion followed by ligation with the 8kb new plasmid digested by *BsrG*I and *EcoR*I. Finally the 5.1kb fragment with correct sequence was released from this 8kb new plasmid and cloned into the 12.6kb fragment digested by *Afe*I and *Nhe*I to make targeting vector.

Figure 7.8 Excision of extra 21 bases by *Hind*III digestion.

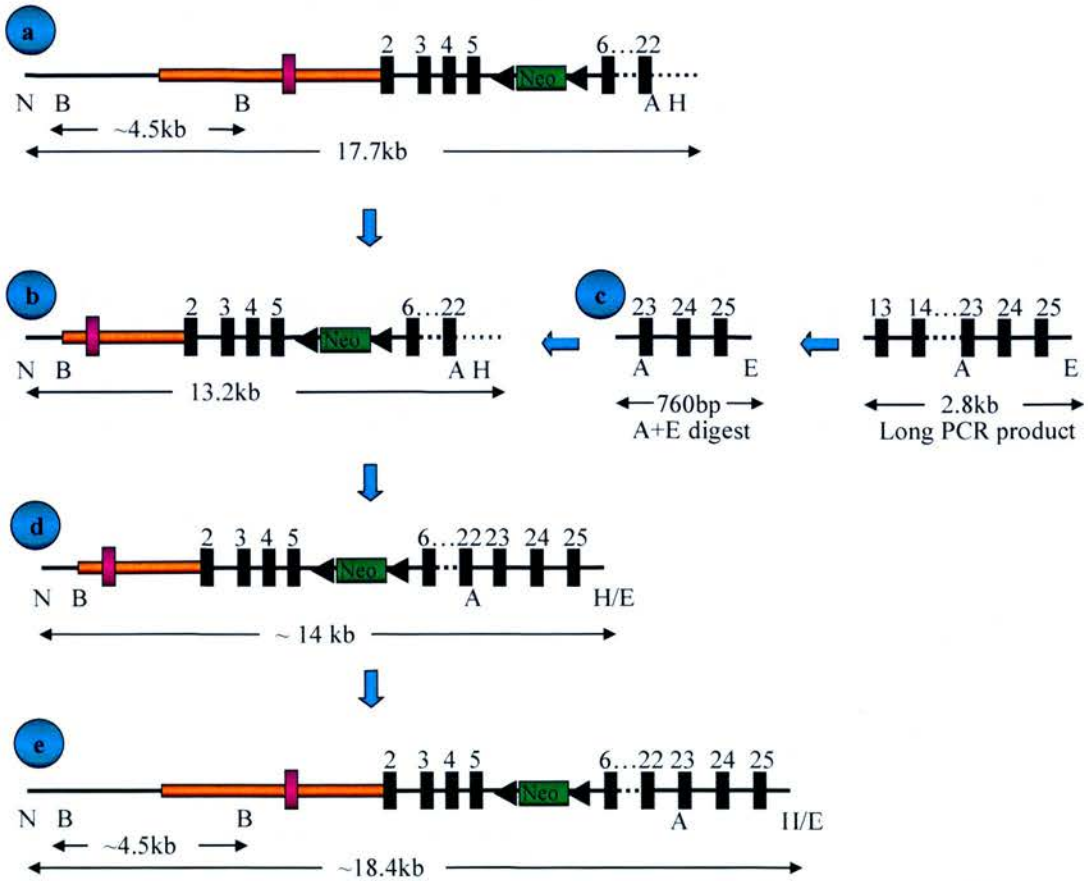


a: 'huamn promoter' subclone with extra 21 bases (in **bold blue low case**). **b**: another *Hind*III site was created by SDM (c→A substitution). **c**: 'huamn promoter' subclone with extra sequence removed by a *Hind*III digestion. Restriction site are underlined.

Generation of a control vector for PCR screening

A control construct with extended 3' homologous sequence was made for long PCR screening of homologous recombinant events after targeting. A 2838bp DNA fragment of mouse *Col1a1* intron 12-26 was amplified by long PCR and used to replace the mouse *Col1a1* intron 12-22 in the targeting vector. The extended 3' sequence was then inserted into a 13.2kb vector which was derived from the 17.7kb targeting vector by *BsrG*I digestion. Finally, this vector was ligated with the BsrGI digested fragment to generate the 18.4kb control vector, as shown in Figure 7.9.

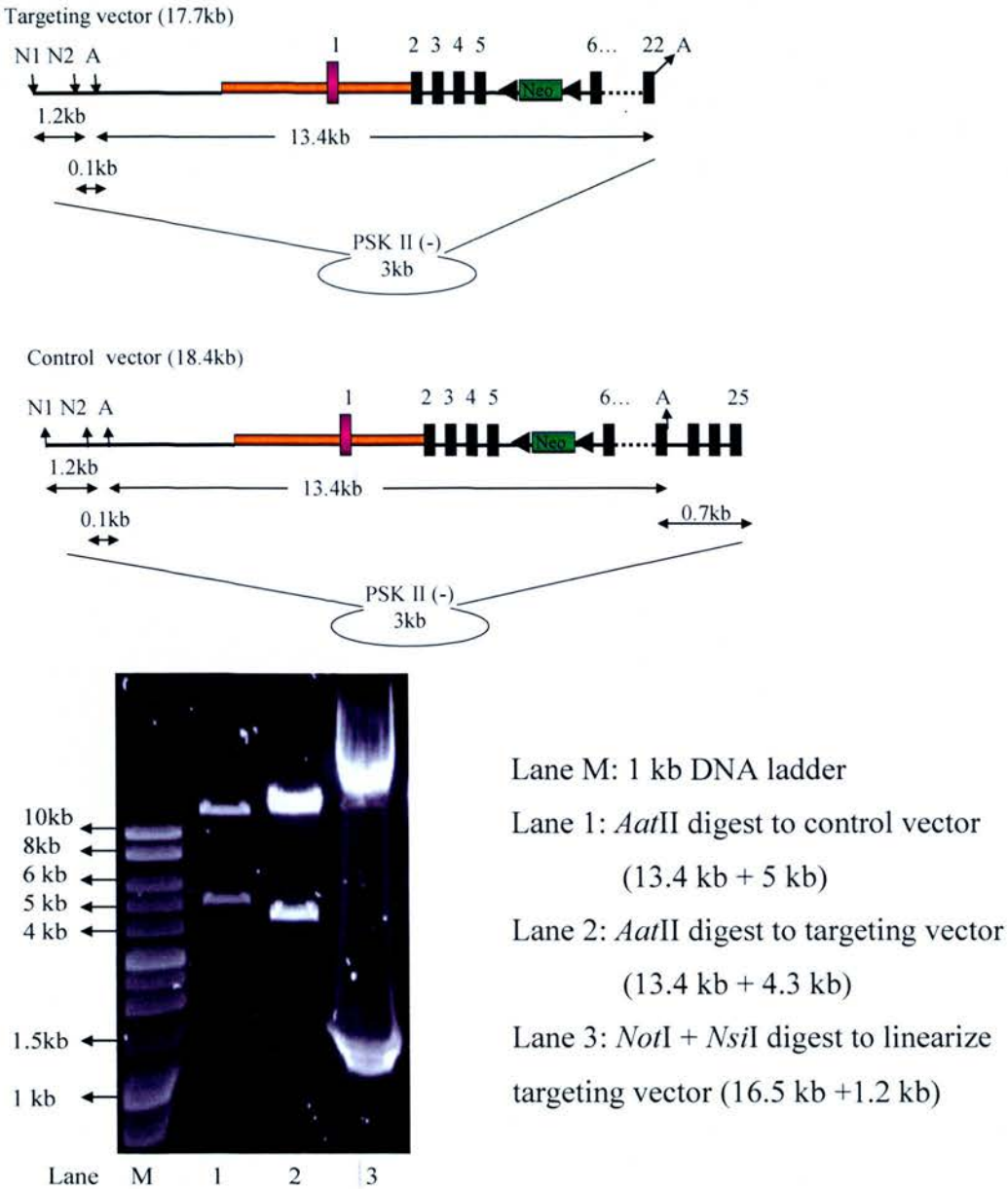
Figure 7.9 Schematic representation of the strategy to make control vector



a: 17.7kb targeting vector; **b**: 13.2kb vector made by *BsrG*I digestion; **c**: 760bp insert made by *Aat*II + *EcoR*V digestion of 2.8kb long PCR product; **d**: 14kb vector made by ligation of b & c; **e**: 18.4kb control vector made by ligation of 4.5kb BsrGI fragment & d.
Restriction enzymes: N, *Not*I; B, *BsrG*I; A, *Aat*II; H, *Hinc*II; E, *EcoR*V.

The completion of the targeting vectors and control vector were also confirmed by restriction digestions prior sequencing reactions. The digestion products were loaded on gel and shown in Figure 7.10.

Figure 7.10 Restriction digest products of targeting and control vectors

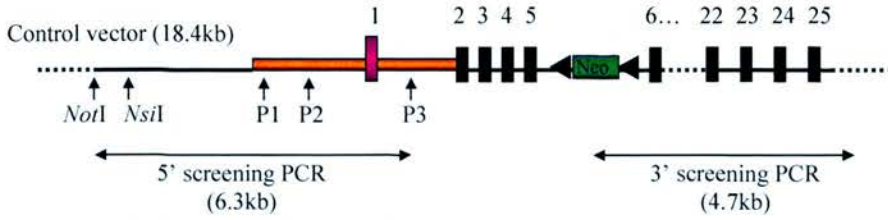


The top two panels show the restriction maps of the targeting vector and the control vector. The bottom panel shows the restriction products on gel along with 1kb DNA ladder (M). Restriction enzymes: N1, *Not*I; N2, *Nsi*I; A, *Aat*II. PSKII, pBluescripti SK (-) vector.

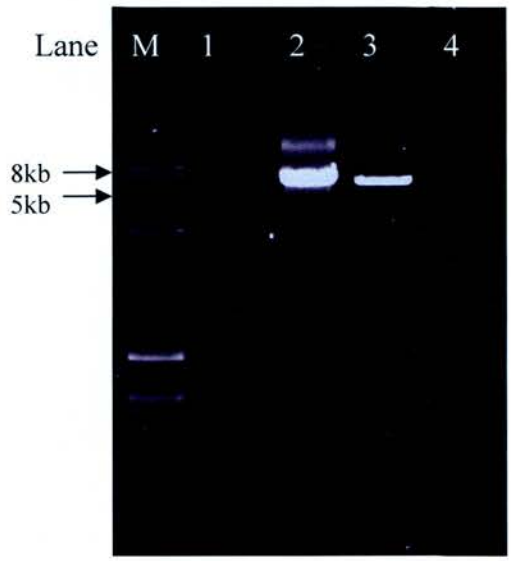
Long PCR reactions for screening the recombination events

Two PCR reactions for screening the recombinant integrations at 5' and 3' ends were also optimized with control vector at different concentrations. In this project, 5' screening PCR was optimized to give a 6330 bp band and 3' screening PCR still produced an unspecific amplification about 2000 bp apart from 4705 bp product after optimization (Figure 7.11).

Figure 7.11 Screening for recombination events by long PCR

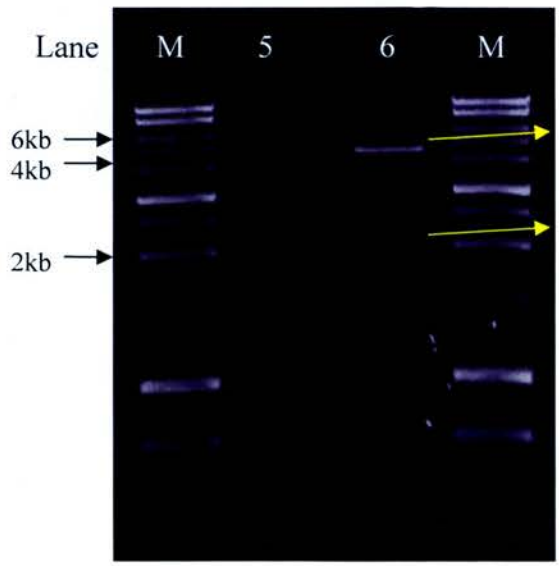


a. PCR screen for 5' integration



Lane 1: 80ng mouse 129/ola DNA
Lane 2: 129/ola +10copies control vector
Lane 3: 129/ola +1copy control vector
Lane 4: 129/ola +0.01copy control vector

b. PCR screening for 3' integration



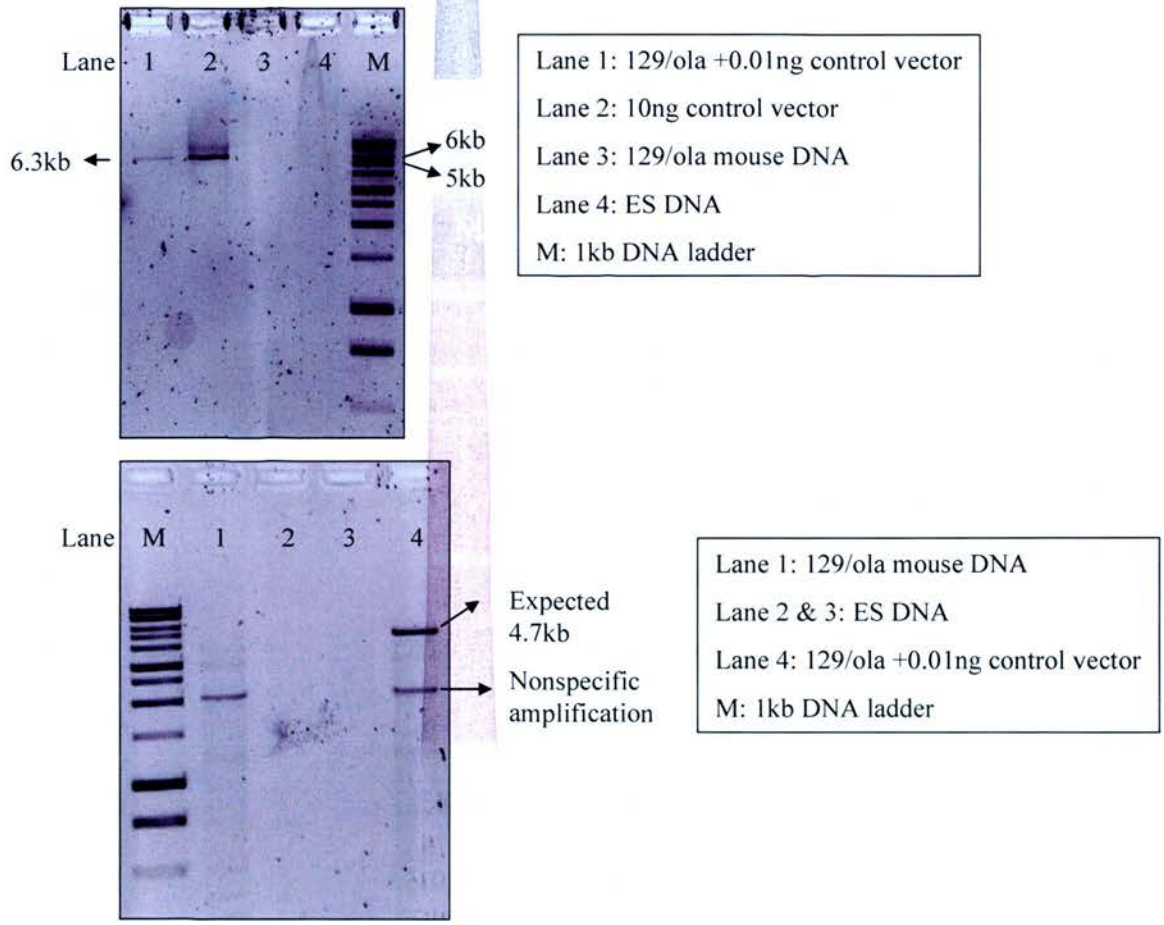
Lane 5: 80ng mouse 129/ola DNA
Lane 6: 129/ola +0.01copy control vector

The top panel shows that long PCR reactions were designed to detect the recombinant events at both 5' and 3' end. **a**: 5' screening PCR with forward primer locating extended homologous mouse promoter and reverse primer binding within human *COL1A1* intron1. The expected product is 6330bp. **b**: 3' screening PCR with forward primer binding within loxp-neo-loxp cassette and reverse primer located downstream of the targeting vector homology sequence (exon23). The expected product is 4705bp.

Gene Targeting Results

After the G418 selection, 45 resistant clones were picked of which 32 were from the first two targeting attempts and 13 from the third targeting. The quality and quantity of all 45 DNA samples extracted from the ES cells were confirmed by a PCR reaction which was designed to amplify a 300bp fragment of the mouse *Sqstm1* gene (not shown). Two screening PCR reactions were then carried out to screen the homologous recombination events at the 3' and the 5' arms on the ES cell DNA. However, only the control vector was amplified successfully, whereas none of the ES cell DNA samples gave the positive amplification signal (Figure 7.12).

Figure7.12 PCR screening for recombination events



The 5' and 3' screening PCR products (top and bottom, respectively). The ES cell DNA extracted from G418 resistant ES cell clones did not give any band at either ends (lane 4 on the top and lane 2&3 on the bottom).

Discussion

The development of gene-targeting technology combined with the sequencing of mouse genome has dramatically increased our ability to create mouse models in which the functional significance of genes, gene variants, and quantitative trait loci (QTL) can be determined. This can be achieved in a number of ways: microinjection of DNA into the pronucleus, embryonic stem cell transfer, retroviral transfer or the nuclear transfer. Embryonic stem cell (ES cells) transfer is based on homologous recombination at the both arms of interest. Embryonic stem cells are derived from the inner cell mass of a blastocyst, which is an early stage embryo. Embryonic stem cell transfer involves isolation of stem cells from a blastocyst, transfection of the transgene (by electroporation), selection of the stem cells in which the transgene has undergone homologous recombination at the desired locus, and injection of the selected stem cells into a host blastocyst-stage embryo. The resulting chimera will undergo further selective breeding for homozygous expression of the desired gene. This is also called gene-targeting.

To study the functions of the three 5' flank *COL1A1* polymorphisms *in vivo*, targeting vectors were designed to contain two haplotypes ("G7T" & "T8G"), which have been associated with BMD in the general population (chapter 3) and regulate reporter gene transcription (chapter 4). Since the three polymorphisms of interest are absent from the mouse *Col1a1* gene, the human *COL1A1* 5' sequence (-2104 to +1537) containing part of promoter, exon1 and intron1 were inserted in the targeting vector to replace the relevant mouse *Col1a1* gene sequence. A control vector was also created by cloning, which contained the same targeting gene but with extended 5' and 3' homologous mouse sequence used to screen the recombination events on both ends.

The targeting and control vectors were successfully completed and confirmed by restriction digestions and DNA sequence analysis. There were several base changes found by DNA sequencing. A "G" insertion at position -1545 of the mouse *Col1a1* promoter was introduced during SDM-PCR amplification, because the forward SDM primer bound human sequence as well as this highly homologous mouse sequence during the SDM-PCR amplification. This highly-conserved sequence may contain potential binding site of nuclear matrix protein 4 (Nmp4) because the binding site was already identified in homologous rat and human sequences (Alvarez M *et al*, 1998 & GARCIA-GIRALT N *et al*, 2005). Therefore, this error was corrected by performing SDM to the mouse sequence only. The DNA sequencing also revealed there were 21 extra bases in the human *COL1A1* promoter region, which derived from the multiple cloning sites of the pUC19 vector. These extra bases were excised successfully by SDM-ligation methods. Another G/C SNP (rs2269336) in human promoter sequence was also detected by sequencing reactions. The frequency is 0.836 for "C" allele and 0.164 for "G" allele in CEU population according to HapMap genotyping results. Although the "G" allele is rare, it was kept in both targeting vectors due to the same DNA template used in the initial PCR amplification. Finally there was a silent mutation in mouse *Col1a1* exon7, which was designed to create an *Xho*I site for quantification of allele-specific transcripts in *vivo*.

Two long-PCR reactions were optimized to detect the recombinant integrations at both 5' and 3' ends with 18.4 kb control vector which contains extended 5' upstream of homologous mouse promoter sequence and extended 3' homologous sequence downstream of the targeting vector (exon25). Because one primer was located in the human *COL1A1* intron1 and another was located in the loxp-neo-loxp cassette, PCR reactions should only give positive products when the homologous recombination has

occurred. Although the 3' screening PCR has a non-specific amplification about 2000 bp, both 5' and 3' screening PCR reactions were optimized to be able to amplify as little as 0.01 copy of control vector mixed with 80 ng mouse 129/ola genomic DNA. Therefore, this is a quick and sensitive method to screen for recombination, although southern blot would still be required to confirm homologous recombination after PCR screening. However, none of the 45 ES clones tested gave positive band in the 5' and 3' screening PCR reactions. The quality and quantity of these DNAs were checked by another PCR. It appears that no homologous recombination occurred in these 45 ES cell clones.

Technical errors in DNA purification, targeting process or ES cell culturing are excluded because the targeting was repeated three times by the same person with the same procedure. Some other targeting vectors which were purified and electroporated into ES cells in parallel were all successfully recombined into the mouse genome. Considering the very low survival of clones from G418 selection (average 13 clones out of 10 plates), it is possible that the loxp-neo-loxp selectable marker cassette did not function properly. This DNA cassette contains two loxp sites, the PGK promoter and neomycin resistance gene. In this project, DNA was amplified from a loxp-PGL-neo-loxp plasmid (gift from Omar Albagha) by PCR reaction with two primers containing *Mfe*I and *Afl*II restriction sites. However, the function of this PGK-neomycin resistant gene was not tested prior to targeting, although it worked properly in another project and the sequence was confirmed by sequencing reactions. To address this issue, the function of this selectable cassette needs be tested in ES cells in the future. Another possible explanation is that for some reason, the mouse genome does not favour the human *COL1A1* targeting construct. There is no relevant literature on this so far. Many foreign genes have been introduced into the mouse genome by

targeting including human Apolipoprotein E4 (cDNA) (Hamanaka H *et al*, 2000), human rhodopsin-GFP fusion gene (7.4kb DNA) (Chan F *et al*, 2004). Moreover, the part of human *COL1A1* sequence in the targeting vector is highly homologous to mouse *Col1a1* sequence. To make the human sequence more homologous to mouse sequence, some mouse sequences were inserted into the boundary of human promoter-exon1-intron1 to mimic the splicing procedure in the mouse genome (Appendix 2). Therefore, it is unlikely that the mouse genome rejects human *COL1A1* sequence during the targeting process.

In summary, two targeting vectors containing two haplotypes of 5' flanking human *COL1A1* gene and a control vector were successfully constructed. Three targeting attempts were performed to introduce targeting vectors into mouse 129/ola ES cells by electroporation. 45 neomycin resistant clones were picked. All extracted DNA was confirmed to have high quality and quantity by a PCR reaction, but none of them exhibited homologous recombination. Further study is in progress to try and repeat ES cell targeting to obtain homologous recombination.

CHAPTER EIGHT

Summary and Future Work

SUMMARY and Future Work

Osteoporosis is a multifactorial disease with a strong genetic component. Approaches such as linkage analysis, candidate gene association study and genetic analysis of mice have been successfully used in the identification of genes that are implicated in the pathogenesis of osteoporosis. Many candidate gene polymorphisms have been implicated in the pathogenesis of osteoporosis but one of the most widely studied is a polymorphism that affects a Sp1 binding site in the first intron of the *COLIA1* gene at position +1245. Type I collagen is the major protein of bone. Mutations within the exons of *COL1A1* and *COL1A2* genes, which code for procollagen α1(I) and α2(I) polypeptides, result in the abnormal bone quality and increased fragility which is associated with osteogenesis imperfecta. The Sp1 polymorphism has been associated with bone density, fragility fractures and other osteoporosis-related phenotypes such as postmenopausal bone loss, bone geometry, bone quality and bone mineralization. Functional analysis has shown that the osteoporosis-associated T allele of the Sp1 polymorphism is associated with increased DNA-protein binding, increased transcription, abnormally increased production of the collagen type I alpha 1 mRNA and bone mineralization (Stewart TL *et al*, 2005 & Mann V *et al*, 2003). Several retrospective meta-analyses of published studies concluded that carriage of the T allele is associated with reduced BMD at the lumbar spine and femoral neck and with vertebral fractures (Ralston SH *et al*, 2006b; Mann V *et al*, 2003; Efstathiadou Z *et al*, 2001 & Mann V *et al* 2003). However, two promoter polymorphisms of *COL1A1* (-1997G/T; rs1107946 and -1663indelT; rs2412298) were recently found to be in strong

LD with the Sp1 polymorphism. These polymorphisms were found to interact with each other and with the Sp1 polymorphism to regulate BMD in a small study of Spanish post-menopausal women (García-Giralt N *et al*, 2002). Also they were found to regulate reporter gene transcription in a promoter-reporter assay (García-Giralt N *et al*, 2005). The -1997G/T promoter polymorphism has been studied in relation to BMD in other populations and family based studied with mixed results although most of these studies have been of limited sample size (Yamada Y *et al*, 2005 & Liu PY *et al*, 2004).

In this study therefore, I examined the roles of all three polymorphisms of the 5' flank of *COL1A1* and their haplotypes in the determination of BMD and fracture, the roles of two promoter polymorphisms in DNA-protein binding and gene transcription and the roles of three polymorphisms and their haplotypes in mechanical properties of bone.

In chapter 3, I demonstrated that all three polymorphisms at 5' flank of *COL1A1* gene were associated with BMD at lumbar spine and femoral neck, and the haplotypes defined by them had stronger effects on BMD in a large cohort of Scottish perimenopausal women (N=3225). Homozygote carriers of haplotype 2 (-1997G /-1663delT / +1245T, "G7T") had reduced BMD at baseline (p=0.007 for LS-BMD; p=0.008 for FN-BMD), whereas homozygotes for haplotype 3 (-1997T / -1663insT / +1245G, "T8G") had increased BMD (p=0.007 for LS-BMD). I also found that the three polymorphisms were in strong linkage disequilibrium and the interactions between the -1997 and -1663 (or Sp1) polymorphisms with BMD, which is in

agreement with García-Giralt N *et al* (2002) and Bustamante M *et al* (2007). Therefore, these 5' flank polymorphisms interact to regulate BMD in this population. There was no association between any individual polymorphism or haplotype and fractures (hip / vertebral) probably due to lack of power of the study. Another interesting finding was that frequencies of some of *COL1A1* genotypes and haplotypes altered with age. This suggests that COL1A1 alleles may be associated with longevity, but this requires further study.

The possible mechanisms by which these polymorphisms regulate BMD have also been studied in this thesis. The gel shift and the super-shift assays in chapter 4 showed that the transcription factors Sp1, Nmp4 and Osterix bind to the DNA surrounding -1663ins/delT polymorphism. The "delT" (7T) allele had greater binding affinity for nuclear proteins of osteoblasts than "insT" (8T) allele. Since the antibodies to Osterix and Nmp4 gave the same super-shift pattern and the two binding sites are next to each other, it is very likely that Osterix and Nmp4 interact with each other and form a DNA-protein complex.

The Sp1 protein also bound to the -1997G/T probe, and the "T" allele showed greater binding affinity than the "G" allele. However, when nuclear extracts from osteoblasts were used as a protein source, there was no evidence of Sp1 binding. It is possible that other GC-box *trans*-acting proteins in nuclear extracts are in competition with Sp1 to bind the *cis*-acing element. For the -1997 G/T site, allelic differences in binding nuclear proteins were also detected by competition assays, but the difference did not reach statistical significance.

To further clarify the contribution of allelic binding differences to gene transcription, luciferase reporter assays were performed. Eight reporter vectors containing different haplotypes and a negative control vector (pGL-3 basic) were con-transfected with pRL-TK vector into osteoblast-like cells (HOS TE85). These reporter vectors were engineered to produce fused *COL1A1*-luciferase protein. All reporter vectors had significantly higher luciferase activities than pGL3-basic which lacks eukaryotic promoter and enhancer sequences ($p<0.001$). The pGL3-G7T (haplotype 2) vector had the highest gene transcription and differed from all the other vectors ($p<0.001$). The luciferase activity of pGL3-T8G (haplotype 3) was the second highest and also significantly differed from all the other vectors. The pGL3-T7G vector had the lowest luciferase activity. The luciferase activity of different haplotypes was not mediated by any individual polymorphism. Therefore, these results suggested that the COL1A1 5' flank polymorphisms interact with each other to regulate reporter gene transcription.

I also demonstrated that these polymorphisms were associated with reduced bone quality *ex vivo* by performing biomechanical tests on trabecular bone samples from femoral heads obtained from hip fracture patients. Not only the individual polymorphisms but also haplotype 6 (T7T) were associated with bone strength ($p<0.05$). Haplotype 6 was the second most common allele in these bone samples (12.24%), but very rare in the general population (0.03%). This is probably because the bone samples were derived from a highly selected group who had suffered from osteoporotic hip fractures and underwent hip replacement surgery, whereas the APOSS population is a randomly selected population of normal healthy women.

In summary, the studies suggest that haplotypes, rather than individual polymorphisms in the 5' flank of *COL1A1* predispose to osteoporosis, by affecting DNA protein interactions, gene expression and bone strength. Future work is required including repeating the gene targeting and the subsequent work on animals, further analyzing of *COL1A1* gene transcription by CHIP assays and studying the 5' flank polymorphisms in extend case and control populations.

REFERENCES

References

Abrahamsen B, Madsen JS, Tofteng CL, Stilgren L, Bladjberg EM. A common methylenetetrahydrofolate reductase (C677T) polymorphism is associated with low bone mineral density and increased fracture incidence after menopause: longitudinal data from the Danish Osteoporosis Prevention Study. *J Bone Miner Res* 2003; **18**: 723–9.

Ai M, Holmen SL, Van Hul W, Williams BO, Warman ML. Reduced affinity to and inhibition by DKK1 form a common mechanism by which high bone mass-associated missense mutations in LRP5 affect canonical Wnt signaling. *Mol Cell Biol.* 2005, **25(12):** 4946-55

Alliston, T. N. & Derynck, R. (2000) in Skeletal Growth Factors, ed. Canalis, E. (Lippincott Williams & Wilkins, Philadelphia), pp. 233–249.

Alvarez M, Thunyakitpisal P, Morrison P, Onyia J, Hock J, Bidwell JP. PTH-responsive osteoblast nuclear matrix architectural transcription factor binds to the rat type I collagen promoter. *J Cell Biochem* 1998; **69**: 336–52

Andrew T, Antioniades L, Scurrah KJ, MacGregor AJ, and Spector TD. Risk of wrist fracture in women is heritable and is influenced by genes that are largely independent of those influencing BMD. *J. Bone Miner. Res.* 2005, **20**: 67–74.

Annes JP, Munger JS, Rifkin DB: Making sense of latent TGF-β1 activation. *J Cell Sci* 2003, **116**: 217–224

Arai H, Miyamoto K, Taketani Y, Yamamoto H, Iemori Y, Morita K, Tonai T, Nishisho T, Mori S, Takeda E. A vitamin D receptor gene polymorphism in the translation initiation codon: effect on protein activity and relation to bone mineral density in Japanese women. *J Bone Miner Res.* 1997, **12(6)**:915-21

Arai H, Miyamoto KI, Yoshida M, Yamamoto H, Taketani Y, Morita K, Kubota M, Yoshida S, Ikeda M, Watabe F, Kanemasa Y, Takeda E. The polymorphism in the caudal-related homeodomain protein Cdx-2 binding element in the human vitamin D receptor gene. *J Bone Miner Res* 2001, **16**: 1256–1264.

Arden NK, Baker J, Hogg C, Baan K, and Spector TD. 1996. The heritability of bone mineral density, ultrasound of the calcaneus and hip axis length: A study of postmenopausal twins. *J Bone Miner Res.* 11: 530–534.

Arden NK & Spector TD. Genetic influences on muscle strength, lean body mass, and bone mineral density: a twin study. *J Bone Miner Res* 1997, **12**: 2076-2081

Asagiri M, Sato K, Usami T, Ochi S, Nishina H, Yoshida H, et al. Autoamplification of NFATc1 expression determines its essential role in bone homeostasis. *J Exp Med* 2005; **202**: 1261 - 9.

Ashby MF. The mechanical properties of cellular solids. *Metallurgical Transactions* 1983;**14-A**:1755-69.

Aubin JE, Heersche JNM. Cellular actions of parathyroid hormone on osteoblast and osteoclast differentiation. In: Bilezikian JP, Marcus R, Levine MA (eds.) The Parathyroids, Basic and Clinical Concepts, 2nd. Academic Press, San Diego, CA, USA, pp.199-211

Babij P, Zhao W, Small C, Kharode Y, Yaworsky PJ, Bouxsein ML, Reddy PS, Bodine PV, Robinson JA, Bhat B, Marzolf J, Moran RA, Bex F. High bone mass in mice expressing a mutant LRP5 gene. *J Bone Miner Res.* 2003, **18(6)**: 960-74

Balemans W, Devogelaer JP, Cleiren E, Van Hul W. A novel LRP5 mutation in a patient with increased bone mass results in reduced DKK1 inhibition. *Bone* 2006, **38**: S6–S7

Balemans W, Van Wesenbeeck L, and Van Hul W. A clinical and molecular overview of the human osteopetroses. *Calcif Tissue Int.* 2005, **77**: 263–274.

Baines M, Kredan MB, Usher J, Davison A, Higgins G, Taylor W, West C, Fraser WD, Ranganath LR. The association of homocysteine and its determinants MTHFR genotype, folate, vitamin B12 and vitamin B6 with bone mineral density in postmenopausal British women. *Bone.* 2007, **40(3)**: 730-6.

Bandrés E, Pombo I, González-Huarriz M, Rebollo A, López G, García-Foncillas J. Association between bone mineral density and polymorphisms of the VDR, ERalpha, COL1A1 and CTR genes in Spanish postmenopausal women. *J Endocrinol Invest.* 2005, **28(4)**: 312-21

Baron R. Anatomy and Ultrastructure of bone. *In* "Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism" (M J Favus, Ed), pp. 3-10, Lippincot Williams & Wilkins, Philadiphia.

Bathum L, von Bornemann-Hjelmorbg J, Christiansen L, Madsen JS, Skytthe A, Christensen K. Evidence for an association of methylene tetrahydrofolate reductase polymorphism C677T and an increased risk of fractures: results from a population-based Danish study. *Osteoporos Int* 2004, **15**: 659-664

Baum J and Brodsky B. Folding of peptide models of collagen and misfolding in disease. *Curr Opin Struct Biol* 1999, **9**:122-128

Bennett B, Beeson M, Gordon L, Carosone-Link P, Johnson TE. Genetic dissection of quantitative trait loci specifying sedative/hypnotic sensitivity to ethanol: mapping with interval-specific congenic recombinant lines. *Alcohol Clin Exp Res.* 2002, **26(11)**: 1615-24

Bilezikian JP, Morishima A, Bell J, and Grumbach MM. Increased bone mass as a result of estrogen therapy in a man with aromatase deficiency. *N Engl J Med.* 1998, **339**: 599–603.

Bogdanovic Z, Bedalov A, Krebsbach PH, Pavlin D, *et al.* Upstream regulatory elements necessary for expression of the rat COL1A1 promoter in transgenic mice. *J Bone Miner Res.* 1994; **9**; 285-292

Bogdanovic Z, Huang YF, Dodig M, Clark SH, Lichtler AC, Kream BE. Parathyroid hormone inhibits collagen synthesis and the activity of rat col1a1 transgenes mainly by a cAMP-mediated pathway in mouse calvariae. *J Cell Biochem* 2000; **77**: 149–58.

Bonewald LF, Mundy GR. Role of transforming growth factor-beta in bone remodeling. *Clin Orthop Relat Res* 1990, **250:** 261–276

Bonstein P, McKAy J. The first intron of the alpha 1(I) collagen gene constains several transcriptional regulatory elements. *J boil chem.* 1988, **263**, 1603-1606

Bornstein P, McKay J, Morishima JK, Devarayalu S, Gelinas RE. Regulatory elements in the first intron contribute to transcriptional control of the human alpha 1(I) collagen. *Proc Natl Acad Sci UAS* 1987; **84**; 8869-8873Bornstein P. Regulation of expression of α1 (I) collagen gene: a critical appraisal of the role of the first intron. *Matrix Biol* 1996; **15**:3–10

Bornstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet* 2003; **33**(suppl): 228-37

Boyce BF, Yoneda T, Lowe C, Soriano P, Mundy GR. Requirement of pp60c-src expression for osteoclasts to form ruffled borders and resorb bone in mice. *J Clin Invest* 1992, 90, 1622-1627.

Boyden LM, Insogna K, Lifton RP High-bone-mass disease and LRP5. *N Engl J Med* 2004, **350**:2096–2099

Boyden LM, Mao J, Belsky J, Mitzner L, Farhi A, Mitnick MA, Wu D, Insogna K, Lifton RP. High bone density due to a mutation in LDL-receptor-related protein 5. *N Engl J Med.* 2002, **346(20)**: 1513-21.

Brixen K, Beckers S, Peeters A, Nielsen TL, Wraae K, Piters E, Balemans W, Bathum L, Andersen M, Van Hul W, Abrahamsen B. Two polymorphisms in the gene encoding the low-density lipoprotein receptor-related protein-5 (LRP5) are associated with peak bone mass mainly in non-sedentary men of the Odense Androgen Study. *Calcif Tissue Int* 2006, **78**: S34–S35

Brull DJ, Montgomery HE, Sanders J, Dhamrait S, Luong L, Rumley A, Lowe GD, Humphries SE. Interleukin-6 gene -174g > c and -572g > c promoter polymorphisms are strong predictors of plasma interleukin-6 levels after coronary artery bypass surgery. *Arterioscler Thromb Vasc Biol.* 2001, **21,** 1458–1463.

Brull DJ, Murray LJ, Boreham CA, Ralston SH, Montgomery HE, Gallagher AM, McGuigan FE, Davey Smith G, Savage M, Humphries SE, Young IS. Effect of a

COL1A1 Sp1 binding site polymorphism on arterial pulse wave velocity: an index of compliance. *Hypertension.* 2001, **38**: 444-8

Bustamante M, Nogues X, Enjuanes A, Elosua R, Garcia-Giralt N, Perez-Edo L, Caceres E, Carreras R, Mellibovsky L, Balcells S, Diez-Perez A, Grinberg D. COL1A1, ESR1, VDR and TGFB1 polymorphisms and haplotypes in relation to BMD in Spanish postmenopausal women. *Osteoporos Int.* 2007, **18(2)**: 235-43

Byers PH 2001 Disorders of collagen biosynthesis and structure. In: Scriver CR, Beaudet AL, Sly WS, Valle D, Childs B, Vogelstein B (eds.) The Metabolic and Molecular Bases of Inherited Disease, 8th ed. McGraw-Hill, New York, NY, USA, pp. 5241-5285

Byers PH, Cole WG 2002 Osteogenesis imperfecta. In: Royce PM, Steinmann B (eds.) Connective Tissue and Its Heritable Disorders. Wiley-Liss Inc., New York, NY, USA, pp.385-430

Camichael CM & McGue M. A cross-sectional examination of height, weight, and body mass index in adult twins. *J Gerontol A Biol Sci Med Sci.* 1995 **50**: B237-44.

Canalis E, Centrella M, Burch W, McCarthy TL. Insulin-like growth factor I mediates selective anabolic effects of parathyroid hormone in bone cultures. *J Clin Invest.* 1989 **83(1)**: 60-5.

Carn G, Koller DL, Peacock M, Hui SL, Evans WE, Conneally PM, Johnston CC, Foroud T, Econs MJ. Sibling pair linkage and association studies between peak bone mineral density and the gene locus for the osteoclast-specific subunit (OC116) of the vacuolar proton pump on chromosome 11p12-13. *J Clin Endocrinol Metab.* 2002, **87(8)**: 3819-24

Casado-Díaz A, Cuenca-Acevedo R, Quesada JM, Dorado G. Individual single tube genotyping and DNA pooling by allele-specific PCR to uncover associations of polymorphisms with complex diseases. *Clin Chim Acta.* 2007, **376**: 155-62.

Cassella JP, Pereira R, Khillan JS, Prockop DJ, Garrington N, Ali SY. An ultrastructural, microanalytical, and spectroscopic study of bone from a transgenic mouse with a COL1A1 pro-alpha-1 mutation. *Bone* 1994; **15(6)**:611-9

Chan F, Bradley A, Wensel TG, Wilson JH. Knock-in human rhodopsin–GFP fusions as mouse models for human disease and targets for gene therapy. *PNAS.* 2004, **101**: 9109-9114

Chen SJ, Artlett CM, Jimenez, SA, Varga J. Modulation of human α1(I) procollagen gene activity by interaction with Sp1 and Sp3 transcription factors in vitro. *Gene* 1998; **215**: 101–110

Chipman SD, Sweet HO, McBride DJ Jr, Davisson MT, Marks SC Jr, Shuldiner AR, Wenstrup RJ, Rowe DW, Shapiro JR.: Defective pro alpha 2(I) collagen synthesis in a recessive mutation in mice: a model of human osteogenesis imperfecta. *Proc Natl Acad Sci* 1993; **90(5)**:1701-5.

Christian JC, Yu PL, Slemenda CW, Johnston CC. Heritability of bone mass: a longitudinal study in aging male twins. *Am J Hum Genet.* 1989, **44(3)**: 429-33.

Chu ML, de Wet W, Bernard M, Ding JF, Morabito M, Myers J, Williams C, Ramirez F. Human pro-α1 (I) collagen gene structure reveals evolutionary conservation of a pattern of introns and exons. *Nature* 1984, **310**:337–340.

Clover J, Gowen M. Are MG-63 and HOS TE85 human osteosarcoma cell lines representative models of the osteoblastic phenotype? *Bone.* 1994, **15(6):**585-91.

Colin EM, Weel AE, Uitterlinden AG, Buurman CJ, Birkenhäger JC, Pols HA, van Leeuwen JP. Consequences of vitamin D receptor gene polymorphisms for growth inhibition of cultured human peripheral blood mononuclear cells by 1, 25-dihydroxyvitamin D3. *Clin Endocrinol.* 2000, **52(2)**:211-6.

Cummings SR, Nevitt MC, Browner WS, Stone K, Fox KM, Ensrud KE, Cauley J, Black D, Vogt TM. Risk factors for hip fracture in white women. Study of Osteoporotic Fractures Research Group. *N Engl J Med.* 1995 **332(12)**: 767-73

Cummings SR, Melton LJ. Epidemiology and outcomes of osteoporotic fractures. *Lancet* 2002; **359**: 1761–67.

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nat Genet* 2001; **29**:229-232.

Darvasi A. Interval-specific congenic strains (ISCS): an experimental design for mapping a QTL into a 1-centimorgan interval. *Mamm Genome.* 1997, **8(3)**:163-7

de Wet W, Bernard M, Benson-Chanda V, Chu M-L, Dicson L, Well D, Ramirez F. Organization of human pro-α(I) collagen gene. *J Biol Chem* 1987, 262:16032–16036.

Deng HW, Chen WM, Recker S, Stegman MR, Li JL, Davies KM, Zhou Y, Deng H, Heaney R, Recker RR. Genetic determination of Colles' fracture and differential bone mass in women with and without Colles' fracture. *J Bone Miner Res.* 2000a, **15(7)**: 1243-52

Department of Health and Human Services. Bone health and osteoporosis: a report of the Surgeon-General. Rockville: US Department of Health and Human Services, Office of the Surgeon General, 2004.

Dick IM, Devine A, Li S, Dhaliwal SS, Prince RL. The T869C TGF beta polymorphism is associated with fracture, bone mineral density, and calcaneal quantitative ultrasound in elderly women. *Bone* 2003, **33**: 335–341.

Dodig M, Kronenberg MS, Bedalov A, Kream BE, *et al.* Identification of a TATA-containing motif required for high level expression of the COL1A1 promoter in differentiated osteoblasts of transgenic mice. *J Biol Chem* 1996; **271(27)**; 16422-16429

Dougall WC, Glaccum M, Charrier K, Rohrbach K, Brasel K, De Smedt T, et al. RANK is essential for osteoclast and lymph node development. *Genes Dev* 1999; **13**: 2412 - 24.

Ducy P, Desbois C, Boyce B, Pinero G, Story B, Dunstan C, Smith E, Bonadio J, Goldstein S, Gundberg C, Bradley A, Karsenty G. Increased bone formation in osteocalcin-deficient mice. *Nature*. 1996 **382**:448-52

Ducy P, Zhang R, Geoffroy V, Ridall AL, Karsenty G. Osf2/Cbfa1: a transcriptional activator of osteoblast differentiation. *Cell*. 1997 **89(5)**:747-54

Eccleshall TR, Garnero P, Gross C, Delmas PD, Feldman D. Lack of correlation between start codon polymorphism of the vitamin D receptor gene and bone mineral density in premenopausal French women: the OFELY study. *J Bone Miner Res*. 1998, **13(1)**: 31-5.

Efstathiadou Z, Tsatsoulis A, Ioannidis JP. Association of collagen Ialpha 1 Sp1 polymorphism with the risk of prevalent fractures: a meta-analysis. *J Bone Miner Res* 2001; **16(9):**1586-1592.

Ehrlich PJ, Lanyon LE. Mechanical strain and bone cell function: a review. *Osteoporos Int.* 2002, **13(9)**: 688-700

Ehringer MA, Thompson J, Conroy O, Yang F, Hink R, Bennett B, Johnson TE, Sikela JM. Fine mapping of polymorphic alcohol-related quantitative trait loci candidate genes using interval-specific congenic recombinant mice. *Alcohol Clin Exp Res.* 2002, **26(11)**: 1603-8

Falchetti A, Sferrazza C, Cepollaro C, Gozzini A, Del Monte F, Masi L, Napoli N, Di Fede G, Cannone V, Cusumano G, Pandolfo MC, Rini GB, Tanini A, Brandi ML. FokI polymorphism of the vitamin D receptor gene correlates with parameters of bone mass and turnover in a female population of the Italian island of Lampedusa. *Calcif Tissue Int.* 2007, **80(1)**:15-20

Fang Y, van Meurs JB, Bergink AP, Hofman A, van Duijn CM, van Leeuwen JP, Pols HA, Uitterlinden AG. Cdx-2 polymorphism in the promoter region of the human vitamin D receptor gene determines susceptibility to fracture in the elderly. *J Bone Miner Res*. 2003, **18(9)**:1632-41

Fang Y, van Meurs JB, d'Alesio A, Jhamai M, Zhao H, Rivadeneira F, Hofman A, van Leeuwen JP, Jehan F, Pols HA, et al. Promoter and 3' untranslated -region haplotypes in the vitamin D receptor gene predispose to osteoporotic fracture: The Rotterdam study. *Am J Hum Genet.* 2005, **77**: 807-823.

Fang Y, Rivadeneira F, van Meurs JB, Pols HA, Ioannidis JP, Uitterlinden AG. Vitamin D receptor gene BsmI and TaqI polymorphisms and fracture risk: A meta-analysis. *Bone* 2006; **39**; 938-945

Fearnhead NS, Wilding JL, Winney B, Tonks S, Bartlett S, Bicknell DC, Tomlinson IP, Mortensen NJ, Bodmer WF. Multiple rare variants in different genes account for multifactorial inherited susceptibility to colorectal adenomas. *Proc Natl Acad Sci* 2004, **101(45):** 15992-7

Felsenberg D, Boonen S. The bone quality framework: determinants of bone strength and their interrelationships, and implications for osteoporosis management. *Clin Ther*. 2005 **27(1)**: 1-11

Feng D, Ishibashi H, Yamamoto S, Hosoi T, Orimo H, Machida T, Koshihara Y. Association between bone loss and promoter polymorphism in the IL-6 gene in elderly Japanese women with hip fracture. *J Bone Miner Metab*. 2003, **21**, 225–228.

Ferrari SL, Ahn-Luong L, Garnero P, Humphries SE, Greenspan SL. Two promoter polymorphisms regulating interleukin-6 gene expression are associated with circulating levels of C reactive protein and markers of bone resorption in postmenopausal women. *J Clin Endocrinol Metab*. 2003, **88**, 255 - 259.

Ferrari SL, Deutsch S, Choudhury U, Chevalley T, Bonjour JP, Dermitzakis ET, Rizzoli R, Antonarakis SE. Polymorphisms in the low-density lipoprotein receptor-related protein 5 (LRP5) gene are associated with variation in vertebral bone mass, vertebral bone size, and stature in whites. *Am J Hum Genet.* 2004a, **74(5)**: 866-75.

Ferrari SL, Karasik D, Liu J, Karamohamed S, Herbert AG, Cupples LA., Kiel DP. Interactions of interleukin-6 promoter polymorphisms with dietary and lifestyle factors and their association with bone mass in men and women from the Framingham Osteoporosis Study. *J Bone Miner Res.* 2004b, **19**, 552–559.

Ferrari SL & Rizzoli R. Gene variants for osteoporosis and their pleiotropic effects in aging. *Mol Aspects Med.* 2005; **26(3):**145-67

Fishman D, Faulds G, Jeffery R, Mohamed-Ali V, Yudkin JS, Humphries S, Woo P. The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. *J Clin Invest*. 1998, **102**, 1369-1376.

Flicker L, Hopper JL, Rodgers L, Kaymakci B, Green RM, Wark JD. Bone density determinants in elderly women: a twin study. *J Bone Miner Res*. 1995 **10(11)**: 1607-13.

Fox SW, Fuller K, Bayley KE, Lean JM, Chambers TJ. *J Immunol*. 2000a **165**, 4957–4963.

Fox SW, Haque SJ, Lovibond AC, Chambers TJ. *J Immunol*. 2003, **170**, 3679–3687.

Frankel WN, Johnson EW, Lutz CM. Congenic strains reveal effects of the epilepsy quantitative trait locus, El2, separate from other El loci. *Mamm Genome* 1995, **6**, 839–843

Franzoso G, Carlson L, Xing L, Poljak L, Shores EW, Brown KD, et al. Requirement for NF-KB in osteoclast and B-cell development. *Genes Dev* 1997; **11**: 3482 - 96.

Frattini A, Orchard PJ, Sobacchi C, Giliani S, Abinun M, Mattsson JP, et al. Defects in TCIRG1 subunit of the vacuolar proton pump are responsible for a subset of human autosomal recessive osteopetrosis. *Nat Genet* 2000; **25**: 343 - 6.

Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, Perry JR, Elliott KS, Lango H, Rayner NW, Shields B, Harries LW, Barrett JC, Ellard S, Groves CJ, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*. 2007, **316(5826)**: 889-94

Fritz DT, Jiang S, Xu J, Rogers MB. A polymorphism in a conserved posttranscriptional regulatory motif alters bone morphogenetic protein 2 (BMP2) RNA:protein interactions. *Mol Endocrinol.* 2006, **20(7)**: 1574-86.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. *Science* 2002; **296**:2225-2229

García-Giralt N, Nogue's X, Enjuanesb A, Puig J. Two New Single-Nucleotide Polymorphisms in the COL1A1 Upstream Regulatory Region and Their Relationship to Bone Mineral Density. *J Bone Miner Res* 2002, **17(3);** 384-393

García-Giralt N, Enjuanesb A, Bustamantea M, Mellibovsky L, *et al.* In vitro functional assay of alleles and haplotypes of two COL1A1-promoter SNPs. *Bone* 2005; **36**; 902– 908

Garnero P, Arden NK, Griffiths G, Delmas PD, Spector TD. Genetic influence on bone turnover in postmenopausal twins. *Clin Endocrinol Metab* 1996, **81**: 140-146

Garnero P, Borel O, Sornay-Rendu E, Duboeuf F, Jeffery R, Woo P, Delmas PD. Association between a functional interleukin-6 gene polymorphism and peak bone mineral density and postmenopausal bone loss in women: the OFELY study. *Bone* 2002, **31,** 43–50.

Gennari L, Becherini L, Mansani R, Masi L, Falchetti A, Morelli A, Colli E, Gonnelli S, Cepollaro C, Brandi ML. FokI polymorphism at translation initiation site of the vitamin D receptor gene predicts bone mineral density and vertebral fractures in postmenopausal Italian women. *J Bone Miner Res.* 1999, **14(8)**:1379-86

Gjesdal CG, Vollset SE, Ueland PM, Refsum H, Drevon CA, Gjessing HK, Tell GS. Plasma total homocysteine level and bone mineral density: the Hordaland Homocysteine Study. *Arch Intern Med* 2006, **166**: 88-94

Glass DA 2nd, Karsenty G. In vivo analysis of Wnt signaling in bone. *Endocrinology*. 2007 Mar

Golbahar J, Hamidi A, Aminzadeh MA, Omrani GR. Association of plasma folate, plasma total homocysteine, but not methylenetetrahydrofolate reductase C677T polymorphism, with bone mineral density in postmenopausal Iranian women: a cross-sectional study. *Bone* 2004; **35**: 760–5.

Gong Y, Vikkula M, Boon L, Liu J, Beighton P, Ramesar R, Peltonen L, Somer H, Hirose T, Dallapiccola B, De Paepe A, Swoboda W, Zabel B, Superti-Furga A, Steinmann B, Brunner HG, Jans A, Boles RG, Adkins W, van den Boogaard MJ, Olsen BR, Warman ML. Osteoporosis-pseudoglioma syndrome, a disorder affecting skeletal strength and vision, is assigned to chromosome region 11q12-13. *Am J Hum Genet.* 1996, **59(1)**: 146-51.

Gowen M, Lazner F, Dodds R, Kapadia R, Feild J, Tavaria M, et al. Cathepsin K knockout mice develop osteopetrosis due to a deficit in matrix degradation but not demineralization. *J Bone Miner Res* 1999; **14**: 1654 - 63.

Ghosh AK. Factors Involved in the Regulation of Type I Collagen Gene Expression: Implication in Fibrosis. *Exp Biol Med* 2002; **227**: 301-14.

Gong Y, Slee RB, Fukai N, Rawadi G, Roman-Roman S, Reginato AM, Wang H, Cundy T, Glorieux FH, Lev D, Zacharin M, Oexle K, Marcelino J, Suwairi W, Heeger S, Sabatakos G, Apte S, Adkins WN, Allgrove J, et al. LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. *Cell.* 2001, **107(4)**: 513-23

Gross C, Eccleshall TR, Malloy PJ, Villa ML, Marcus R, Feldman D. The presence of a polymorphism at the translation initiation site of the vitamin D receptor gene is associated with low bone mineral density in postmenopausal Mexican-American women. *J Bone Miner Res.* 1996; **11(12)**: 1835-40.
Gross C, Krishnan AV, Malloy PJ, Eccleshall TR, Zhao XY, Feldman D. The vitamin D receptor gene start codon polymorphism: a functional analysis of FokI variants. *J Bone Miner Res*. 1998, **13(11)**: 1691-9.

Gueguen R, Jouanny P, Guillemin F, Kuntz C, Pourel J, and Siest G. Segregation analysis and variance components analysis of bone mineral density in healthy families. *J Bone Miner Res* 1995, **12**: 2017–2022.

Guo W, Fung WK. Combining the case-control methodology with the small size transmission/disequilibrium test for multiallelic markers. *Eur J Hum Genet.* 2005 **13(9):** 1007-12

Hamanaka H, Katoh-Fukui Y, Suzuki K, Kobayashi M, Suzuki R, Motegi Y, Nakahara Y, Takeshita A, Kawai M, Ishiguro K, Yokoyama M, Fujita SC. Altered cholesterol metabolism in human apolipoprotein E4 knock-in mice. *Hum Mol Genet.* 2000 **12**: 353-61

Hartung S, Jaenisch R, Breindl M. Retrovirus insertion inactivates mouse alpha 1(I) collagen by blocking initiation of transcription. *Nature* 1986; 320; 365-367

Herrington DM, Howard TD, Brosnihan KB, McDonnell DP, Li X, Hawkins GA, Reboussin DM, Xu J, Zheng SL, Meyers DA, Bleecker ER. Common estrogen receptor polymorphism augments effects of hormone replacement therapy on E-selectin but not C-reactive protein. *Circulation.* 2002, **105(16)**:1879-82

Hinke V, Seck T, Clanget C, Scheidt-Nave C, Ziegler R, Pfeilschifter J. Association of transforming growth factor-beta1 (TGFbeta1) T29 → C gene polymorphism with bone mineral density (BMD), changes in BMD, and serum concentrations of TGF-beta1 in a population-based sample of postmenopausal german women. *Calcif Tissue Int* 2001, **69:** 315–320.

Hitraya EG, Jimenez SA. Transcriptional activation of the α1(I) procollagen gene in systemic sclerosis dermal fibroblasts: role of intronic sequences. *Arthritis Rheum* 1996; **39**:1347–1354

Hitraya EG, Varga J, Artlett CM, Jimenez SA. Identification of elements in the promoter region of the α1(I) procollagen gene involved in its up-regulated expression in systemic sclerosis. *Arthritis Rheum* 1998; **41**:2048–2058

Hormuzdi SG, Penttinen R, Jaenisch R, Bornstein P. A Gene-Targeting Approach Identifies a Function for the First Intron in Expression of the α1(I) Collagen Gene. *Mol Cell Biol* 1998, **18** (6); 3368-3375

Hormuzdi SG, Strandjord TP, Madtes DK, Bornstein P. Mice with a targeted intronic deletion in the Col1a1 gene respond to bleomycin-induced pulmonary fibrosis with increased expression of the mutant allele. *Matrix Biol.* 1999, **18(3)**: 287-94

Horwood NJ, Elliott J, Martin TJ, Gillespie MT. Osteotropic agents regulate the expression of osteoclast differentiation factor and osteoprotegerin in osteoblastic stromal cells. *Endocrinology* 1998, **139(11)**: 4743-6

Huang JT, Welch JS, Ricote M, Binder CJ, Willson TM, Kelly C, Witztum JL, Funk CD, Conrad D, Glass CK. Interleukin-4-dependent production of PPAR-gamma ligands in macrophages by 12/15-lipoxygenase. Nature. 1999 Jul 22;400(6742):378-82

Hunter D, De Lange M, Snieder H, MacGregor AJ, Swaminathan R, Thakker RV, Spector TD. Genetic contribution to bone metabolism, calcium excretion, and vitamin D and parathyroid hormone regulation. *J Bone Miner Res* 2001, **16**: 371-378

Ichikawa S, Johnson ML, Koller DL, Lai D, Xuei X, Edenberg HJ, Hui SL, Foroud TM, Peacock M, Econs MJ. Polymorphisms in the bone morphogenetic protein 2 (BMP2) gene do not affect bone mineral density in white men or women. *Osteoporos Int.* 2006; **17(4)**:587-92.

Ichikawa S, Koller DL, Johnson ML, Lai D, Xuei X, Edenberg HJ, Klein RF, Orwoll ES, Hui SL, Foroud TM, Peacock M, and Econs MJ. Human ALOX12, but not ALOX15, is associated with BMD in white men and women. *J Bone Miner Res* 2006; **16**: 371-378

Ioannidis JP, Stavrou I, Trikalinos TA, Zois C, Brandi ML, Gennari L, Albagha O, Ralston SH, Tsatsoulis A. Association of polymorphisms of the estrogen receptor alpha gene with bone mineral density and fracture risk in women: a meta-analysis. *J Bone Miner Res* 2002; **17**: 2048-2060.

Ioannidis JP, Ralston SH, Bennett ST, Brandi ML, Grinberg D, *et al.* Differential genetic effects of ESR1 gene polymorphisms on osteoporosis outcomes. *JAMA.* 2004 **292**; 2105-14

Iotsova V, Caamano J, Loy J, Yang Y, Lewin A, Bravo R. Osteopetrosis in mice lacking NF- κ B1 and NF- κ B2. *Nat Med* 1997; **3**: 1285 - 9.

Janssens K, Gershoni-Baruch R, Guanabens N, Migone N, Ralston S, Bonduelle M, Lissens W, Van Maldergem L, Vanhoenacker F, Verbruggen L, Van Hul W. Mutations in the gene encoding the latency-associated peptide of TGF-beta 1 cause Camurati-Engelmann disease. *Nat Genet* 2000, **26:** 273–275.

Janssens K and Van Hul W. Molecular genetics of too much bone. *Hum Mol Genet.* 2002, **11**: 2385–2393.

Janssens K, ten Dijke P, Ralston SH, Bergmann C, Van Hul W. Transforming growth factor-beta 1 mutations in Camurati-Engelmann disease lead to increased signaling by altering either activation or secretion of the mutant protein. *J Biol Chem* 2003, **278**: 7718–7724.

Jilka RL, Weinstein RS, Bellido T, Roberson P, Parfitt AM, Manolagas SC. Increased bone formation by prevention of osteoblast apoptosis with parathyroid hormone. *J Clin Invest.* 1999, **104(4)**: 439-46

Johnson ML, Gong G, Kimberling W, Reckér SM, Kimmel DB, Recker RB. Linkage of a gene causing high bone mass to human chromosome 11 (11q12-13) *Am J Hum Genet.* 1997, **60(6)**: 1326-32.

Johnson ML, Harnish K, Nusse R, Van Hul W. LRP5 and Wnt signaling: a union made for bone. *J Bone Miner Res*. 2004, **19(11)**: 1749-57.

Johnson RS, Spiegelman BM, Papaioannou V. Pleiotropic effects of a null mutation in the c-fos proto-oncogene. *Cell* 1992; **71**: 577 - 86.

Jorgenssn HL, Madsen JS, Madsen B, Saleh MM, Abrahamsen B, Fenger M, et al. Association of a common allelic polymorphism (C677T) in the methylenetetrahydrofolate reductase gene with a reduced risk of osteoporotic fractures. A case control study in Danish postmenopausal women. *Calcif Tissue Int* 2002; **71**:386–92

Jurutka PW, Remus LS, Whitfield GK, Thompson PD, Hsieh JC, Zitzer H, Tavakkoli P, Galligan MA, Dang HT, Haussler CA, Haussler MR. The polymorphic N terminus in human vitamin D receptor isoforms influences transcriptional activity by modulating interaction with transcription factor IIB. *Mol Endocrinol.* 2000, **14(3)**: 401-20

Kadonaga JT, Carner KR, Masiarz FR, Tjian R. Isolation of cDNA encoding transcription factor sp1 and functional analysis of the DNA binding domain. *Cell* 1987; **51**; 1079-1090

Kammerer CM, Schneider JL, Cole SA, Hixson JE, Samollow PB, O'Connell JR, Perez R, Dyer TD, Almasy L, Blangero J, Bauer RL, Mitchell BD. Quantitative trait loci on chromosomes 2p, 4p, and 13q influence bone mineral density of the forearm and hip in Mexican Americans. *J Bone Miner Res* 2003; **18**:2245-2252.

Kannus P, Palvanen M, Kaprio J, Parkkari J, Koskenvuo M. Genetic factors and osteoporotic fractures in elderly people: prospective 25 year follow up of a nationwide cohort of elderly Finnish twins. *BMJ.* 1999 **319**,: 1334-7

Kanzler B, Foreman RK, Labosky PA, Mallo M.BMP signaling is essential for development of skeletogenic and neurogenic cranial neural crest. *Development* 2000, **127**: 1095–1104.

Kaprio, J., Rimpela, A., Winter, T., Viken, R.J., Rimpela, M., and Rose, R.J. Common genetic influences on BMI and age at menarche. *Hum. Biol.* 1995, **67**: 739–753.

Karaplis AC, Goltzman D. PTH and PTHrP effects on the skeleton. *Rev Endocr Metab Disord* 2000, **1**: 331-341

Karsenty G, de Crombrugghe B. Two different negative and one positive regulatory factors interact with a short promoter segment of alpha 1 (I) collagen gene. *J Biol Chem* 1990; **265**; 9934-9942

Karsenty G, de Crombrugghe B. Conservation of binding sites for regulatory factors in the coordinately expressed α1(I) and α2 (I) collagen promoters. *Biochem Biophys Res Commun* 1991, **177**: 538–544.

Karsenty G, Ravazzolo R, de Crombrugghe B. Purification and functional characterization of a DNA-binding protein that interacts with a negative element in the mouse α1(I) collagen promoter. J Biol Chem 266:24842–24848, 1991.

Kato M, Patel MS, Levasseur R, Lobov I, Chang BH, Glass DA, Hartmann C, Li L, Hwang TH, Brayton CF, Lang RA, Karsenty G, Chan L. Cbfa1-independent decrease in osteoblast proliferation, osteopenia, and persistent embryonic eye vascularization in mice deficient in Lrp5, a Wnt coreceptor. *J Cell Biol.* 2002, **15**; 157(2):303-14

Keen RW, Snieder H, Molloy H, Daniels J, Chiano M, Gibson F, Fairbairn L, Smith P, MacGregor AJ, Gewert D, Spector TD. Evidence of association and

linkage disequilibrium between a novel polymorphism in the transforming growth factor beta 1 gene and hip bone mineral density: A study of female twins. *Rheumatology (Oxford)* 2001, *40*: 48–54.

Kern B, Shen J, Starbuck M, Karsenty G. Cbfa1 contributes to the osteoblast-specific expression of type I collagen genes. *J Biol Chem* 2001; **276**:7101–7107

Khan E, Abu-Amer Y. Activation of peroxisome proliferator-activated receptor-gamma inhibits differentiation of preosteoblasts. J Lab Clin Med. 2003 Jul;142(1):29-34

Khillan JK, Olsen AS, Kontusaari S, Sokolov B, Prockop DJ. Transgenic mice that express a mini-gene version of the human gene for type I procollagen (COL1A1) develop a phenotype resembling a lethal form of osteogenesis imperfecta. *J Biol Chem* 1991; 266(12), 23373-23379

Kiel DP, Ferrari SL, Cupples LA, Karasik D, Manen D, Imamovic A, Herbert AG, Dupuis J. Genetic variation at the low-density lipoprotein receptor-related protein 5 (LRP5) locus modulates Wnt signaling and the relationship of physical activity with bone mineral density in men. *Bone*. 2007; **40(3)**: 587-96.

Kim CH, Takai E, Zhou H, von Stechow D, Muller R, Dempster DW, Guo XE. Trabecular bone response to mechanical and parathyroid hormone stimulation: the role of mechanical microenvironment. *J. Bone Miner. Res*. 2003, **18**: 2116–2125.

Kinoshita A, Saito T, Tomita H, Makita Y, Yoshida K, Ghadami M, Yamada K, Kondo S, Ikegawa S, Nishimura G, Fukushima Y, Nakagomi T, Saito H, Sugimoto T, Kamegaya M, Hisa K, Murray JC, Taniguchi N, Niikawa N, Yoshiura K. Domain-specific mutations in TGFB1 result in Camurati-Engelmann disease. *Nat Genet*. 2000, **26(1)**: 19-20

Kiviranta R, Morko J, Alatalo SL, NicAmhlaoibh R, Risteli J, Laitala-Leinonen T, et al. Impaired bone resorption in cathepsin K-deficient mice is partially compensated for by enhanced osteoclastogenesis and increased expression of other proteases via an increased RANKL/OPG ratio. *Bone* 2005; **36**: 159 - 72.

Klein RF, Mitchell SR, Phillips TJ, Belknap JK, Orwoll ES. Genetic analysis of bone mass in mice. J Bone Miner Res 1998; 13:1648-1656.

Klein RF, Allard J, Avnur Z, Nikolcheva T, Rotstein D, Carlos AS, Shea M, Waters RV, Belknap JK, Peltz G, Orwoll ES. Regulation of bone mass in mice by the lipoxygenase gene Alox15. Science 2004; 303:229-232.

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, Sangiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J. Complement factor H polymorphism in age-related macular degeneration. *Science* 2005; **308**:385-389.

Kobayashi T, Walsh PT, Walsh MC, Speirs KM, Chiffoleau E, King CG, et al. TRAF6 is a critical factor for dendritic cell maturation and development. *Immunity* 2003; **19**: 353 - 63.

Koga T, Inui M, Inoue K, Kim S, Suematsu A, Kobayashi E, et al. Costimulatory signals mediated by the ITAM motif cooperate with RANKL for bone homeostasis. *Nature* 2004; **428**: 758 - 63.

Koh JM, Jung MH, Hong JS, Park HJ, Chang JS, Shin HD, Kim SY, Kim GS. Association between bone mineral density and LDL receptor-related protein 5 gene polymorphisms in young Korean men. *J Korean Med Sci.* 2004, **19(3)**: 407-12

Koller DL, Rodriguez LA, Christian JC, Slemenda CW, Econs MJ, Hui SL, Morin P, Conneally PM, Joslyn G, Curran ME, Peacock M, Johnston CC, Foroud T. Linkage of a QTL contributing to normal variation in bone mineral density to chromosome 11q12-13. *J Bone Miner Res*. 1998, **13(12)**:1903-8.

Komori, T. Yagi H, Nomura S, Yamaguchi A, Sasaki K, Deguchi K, Shimizu Y, Bronson RT, Gao YH, Inada M, Sato M, Okamoto R, Kitamura Y, Yoshiki S, Kong YY, Yoshida H, Sarosi I, Tan HL, Timms E, Capparelli C, et al. OPGL is a key regulator of osteoclastogenesis, lymphocyte development and lymph-node organogenesis. *Nature* 1999; **397**: 315 - 23.

Kong YY, Feige U, Sarosi I, Bolon B, Tafuri A, Morony S, Capparelli C, Li J, Elliott R, McCabe S, Wong T, Campagnuolo G, Moran E, et al. Penninger JM. Activated T cells regulate bone loss and joint destruction in adjuvant arthritis through osteoprotegerin ligand. *Nature*. 1999, **402**: 304-9.

Kornak U, Kasper D, Bosl MR, Kaiser E, Schweizer M, Schulz A, et al. Loss of the ClC-7 chloride channel leads to osteopetrosis in mice and man. *Cell* 2001; **104**: 205 - 15.

Kishimoto T. . Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. *Cell* 1997, **89**, 755–764

Krall EA and Dawson-Hughes B. Heritable and life-style determinants of bone mineral density. *J Bone Miner Res* 1993, **8**: 1–9.

Kratochwil K, von der M, Kollar EJ, Jaenisch R, Mooslehner K, et al. Retrovirus-induced insertional mutation in Mov13 mice affects collagen I expression in a tissue-specific manner. *Cell* 1989; **57**; 807-816

Kream BE, LaFrancis D, Petersen DN, Woody C, Clark S, Rowe DW, Lichtler A. Parathyroid hormone represses alpha 1(I) collagen promoter activity in cultured calvariae from neonatal transgenic mice. *Mol Endocrinol* 1993; **7**: 399– 408

Kuhn R, Schwenk F, Aguet M, Rajewsky K. Inducible gene targeting in mice. *Science*, 1995; **269**; 1427-1429

Kuhn R, Torres RM. *Methods Mol. Biol* 2002; **180**, 175–204

Kuhn H, Walther M, Kuban RJ. Mammalian arachidonate 15-lipoxygenases structure, function, and biological implications. Prostaglandins Other Lipid Mediat. 2002 Aug;68-69:263-90

Kwee ML, Balemans W, Cleiren E, Gille JJ, Van Der Blij F, Sepers JM, Van HulW. An autosomal dominant high bone mass phenotype in association with craniosynostosis in an extended family is caused by an LRP5 missense mutation. *J Bone Miner Res* 2005, **20**: 1254–1260

Lakso M, Sauer B, Mosinger B, Lee EJ, Manning RW, Yu SH, Mulder KL, Westphal H. Targeted oncogene activation by site-specific recombination in transgenic mice. *Proc Natl Acad Sci U S A*. 1992, **89(14)**: 6232-6

Langdahl BL, Knudsen JY, Jensen HK, Gregersen N, Eriksen. EF. A sequence variation: 713-8delC in the transforming growth factor-beta 1 gene has higher prevalence in osteoporotic women than in normal women and is associated with very low bone mass in osteoporotic women and increased bone turnover in both osteoporotic and normal women. *Bone* 1997, **20**: 289–294.

Langdahl BL, Gravholt CH, Brixen K, Eriksen EF. Polymorphisms in the vitamin D receptor gene and bone mass, bone turnover and osteoporotic fractures. *Eur J Clin Invest*. 2000, **30(7)**:608-17

Langdahl BL, Carstens M, Stenkjaer L, Eriksen EF. Polymorphisms in the transforming growth factor beta 1 gene and osteoporosis. *Bone* 2003, **32**: 297–310.

Langdahl, B.L., Ralston, S.H., Grant, S.F.A., and Eriksen, E.F. 1998. An Sp1 binding site polymorphism in the COLIA1 gene predicts osteoporotic fractures in men and women. *J. Bone Miner. Res.* **13**:1384–1389.

Lau HH, Ng MY, Ho AY, Luk KD, Kung AW. Genetic and environmental determinants of bone mineral density in Chinese women. Bone. 2005, **36(4)**: 700-9

Lecka-Czernik B, Moerman EJ, Grant DF, Lehmann JM, Manolagas SC, Jilka RL. Divergent effects of selective peroxisome proliferator-activated receptor-gamma 2 ligands on adipocyte versus osteoblast differentiation. Endocrinology. 2002 Jun;143(6):2376-84

Li B and Aspden RM. Composition and mechanical properties of cancellous bone from the femoral head of patients with osteoporosis or osteoarthritis. *J Bone Miner Res.* 1997, **12**:641–651.

Li L, Artlett CM, Jimenez SA, Hall DJ, Varga J. Positive regulation of human alpha 1(I) collagen promoter activity by transcription factor Sp1. *Gene* 1995; **164**; 229-234

Li J, Sarosi I, Yan XQ, Morony S, Capparelli C, Tan HL, et al. RANK is the intrinsic hematopoietic cell surface receptor that controls osteoclastogenesis and regulation of bone mass and calcium metabolism. *Proc Natl Acad Sci* 2000; **97**: 1566 - 71.

Li YP, Chen W, Liang Y, Li E, Stashenko P. Atp6i-deficient mice exhibit severe osteopetrosis due to loss of osteoclast-mediated extracellular acidification. *Nat Genet* 1999; **23**: 447 - 51.

Liska DJ, Slack JL, Bornstein P. A highly conserved intronic sequence is involved in transcriptional regulation of the α1(I) collagen gene. *Cell Regul* 1990; **1**: 487–498

Little RD, Carulli JP, Del Mastro RG, Dupuis J, Osborne M, Folz C, Manning SP, Swain PM, Zhao SC, Eustace B, Lappe MM, Spitzer L, Zweier S, Braunschweiger K, Benchekroun Y, Hu X, et al. A mutation in the LDL receptor-related protein 5 gene results in the autosomal dominant high-bone-mass trait. *Am J Hum Genet.* 2002; **70(1)**: 11-9.

Liu PY, Lu Y, Long JR, Xu FH, Shen H, Recker RR, Deng HW Common variants at the PCOL2 and Sp1 binding sites of the COL1A1 gene and their interactive effect influence bone mineral density in Caucasians. *J Med Genet.* 2004, **41(10)**: 752-7

Livshits G, Trofimov S, Malkin I, Kobyliansky E. Transmission disequilibrium test for hand bone mineral density and 11q12-13 chromosomal segment. *Osteoporos Int.* 2002; **13(6)**:461-7

Lomaga MA, Yeh WC, Sarosi I, Duncan GS, Furlonger C, Ho A, et al. TRAF6 deficiency results in osteopetrosis and defective interleukin-1, CD40, and LPS signaling. *Genes Dev* 1999; **13**: 1015 - 24.

Lorentzon M, Lorentzon R, Nordstrom P. Interleukin-6 gene polymorphism is related to bone mineral density during and after puberty in healthy white males: a cross-sectional and longitudinal study. *J Bone Miner Res.* 2000, **15**, 1944–1949.

Lovibond AC, Haque SJ, Chambers TJ, Fox SW. *Biochem Biophys Res Commun.* 2003. **309,** 762–767.

Lowe C, Yoneda T, Boyce BF, Chen H, Mundy GR., Soriano P. Osteopetrosis in Src-deficient mice is due to an autonomous defect of osteoclasts. *Proc. Natl. Acad. Sci.* 1993, **90**, 4485-4490.

Mackay JP, CrossleyM. Zinc fingers are sticking together. *Trends Biochem. Sci.* 1998, **23**: 1–4

Mann V, Hobson EE, Li B, Stewart TL, Grant SF, Robins SP, Aspden RM, Ralston SH. A COL1A1 Sp1 binding site polymorphism predisposes to osteoporotic fracture by affecting bone density and quality. *J Clin Invest* 2001, **107**: 899–907

Mann V, Raslton SH. Meta-analysis of COL1A1 Sp1 polymorphism in relation to bone mineral density and osteoporotic fracture. *Bone* 2003; **32**: 711-717

Manolagas SC. The role of IL-6 type cytokines and their recrptors in bone. *Ann N Y Acad Sci.* 1998, **840**: 194-204

Marjamaa J, Tulamo R, Abo-Ramadan U, Hakovirta H, Frosen J, Rahkonen O, Niemela M, Bornstein P et al. Mice with a deletion in the first intron of the Col1a1 gene develop dissection and rupture of aorta in the absence of aneurysms: high-resolution magnetic resonance imaging, at 4.7 T, of the aorta and cerebral arteries. *Magn Reson Med.* 2006; **55(3)**: 592-7

Maruyama H, Toji H, Harrington CR, Sasaki K, Izumi Y, Ohnuma T, Arai H, Yasuda M, Tanaka C, Emson PC, Nakamura S, Kawakami H. Lack of an association of estrogen receptor alpha gene polymorphisms and transcriptional activity with Alzheimer disease. *Arch Neurol.* 2000, **57(2)**: 236-40

Medici M, van Meurs JB, Rivadeneira F, Zhao H, Arp PP, Hofman A, Pols HA, Uitterlinden AG. BMP-2 gene polymorphisms and osteoporosis: the Rotterdam Study. *J Bone Miner Res.* 2006, **21(6)**: 845-54.

McBride DJ, Jr Shapiro JR and Dunn MG. Bone geometry and strength measurements in aging mice with the oim mutation. *Calcif Tissue Int.* 1998, **62**: 172–176.

McGuigan FE, Armbrecht G, Smith R, Felsenberg D,Reid DM, Ralston SH. Prediction of osteoporotic fractures by bone densitometry and COLIA1 genotyping: a prospective, population-based study in men and women. *Osteoporos Int* 2001; **12**:91-96

McGuigan FE, Macdonald HM, Bassiti A, Farmer R, Bear S, Stewart A, Black A, Fraser WD, Welsh F, Reid DM, Ralston SH. Large-scale population-based study shows no association between common polymorphisms of the TGFB1 gene and BMD in women. *J Bone Miner Res.* 2007, **22(2)**: 195-202

McLean RR, Jacques PF, Selhub J, Tucker KL, Samelson EJ, Broe KE, et al. Homocysteine as a predictive factor for hip fractures in older persons. *N Engl J Med* 2004, **350**: 2042-2049

Michaelsson K, Melhus H, Ferm H, Ahlbom A, Pedersen NL. Genetic liability to fractures in the elderly. *Arch Intern Med.* 2005, **165(16)**: 1825-30

Misof K, Landis WJ, Klaushofer K, Fratzl P. Collagen from the osteogensis imperfecta mouse model (oim) shows reduced resistance against tensile strew. *J. Clin. Invest* 1997, **100**:40-45

Miyao M, Morita H, Hosai T, Kurihara H, Inoue S, et al. Association of methylenetetrehydrofolate reductase (MTHFR) polymorphism with bone mineral density in postmenopausal Japanese women. *Calcif Tissue Int* 2000; **66**:190–4.

Mocsai A, Humphrey MB, Van Ziffle JA, Hu Y, Burghardt A, Spusta SC, et al. The immunomodulatory adapter proteins DAP12 and Fc receptor γ-chain (FcR γ) regulate development of functional osteoclasts through the Syk tyrosine kinase. *Proc Natl Acad Sci* 2004; **101**: 6158 - 6163.

Moonga BS, Adebanjo OA, Wang HJ, Li S, Wu XB, Troen B, Inzerillo A, Abe E, Minkin C, Huang CL, Zaidi M. Differential effects of interleukin-6 receptor activation on intracellular signaling and bone resorption by isolated rat osteoclasts. *J Endocrinol.* 2002, **173(3)**:395-405.

Morel L, Yu Y, Blenman KR, Caldwell RA, Wakeland EK. Production of congenic mouse strains carrying genomic intervals containing SLE-susceptibility genes derived from the SLE-prone NZM2410 strain. *Mamm Genome* 1996, **7**, 335–339

Morinobu M, Nakamoto T, Hino K, Tsuji K, Shen ZJ, Nakashima K, Nifuji A, Yamamoto H, Hirai H, Noda M: The nucleocytoplasmic shuttling protein CIZ reduces adult bone mass by inhibiting bone morphogenetic protein-induced bone formation. *J Exp Med.* 2005 **201**:961-70

Morrison NA, Eisman JA. Nonhypercalcemic 1,25-(OH)2D3 analogs potently induce the human osteocalcin gene promoter stably transfected into rat osteosarcoma cells (ROSCO-2). *J Bone Miner Res.* 1991; **6(8)**: 893-9.

Morrison NA, Qi JC, Tokita A, Kelly PJ, Crofts L, Nguyen TV, Sambrook PN, Eisman JA. Prediction of bone density from vitamin D receptor alleles. *Nature* 1994, **367**:284–287

Murakami T, Yamamoto M, Yamamoto M, Ono K, Nishikawa M, Nagata N, Motoyoshi K, Akatsu T. *Biochem Biophys Res Commun.* 1998. **252**, 747–752.

Naito A, Azuma S, Tanaka S, Miyazaki T, Takaki S, Takatsu K, et al. Severe osteopetrosis, defective interleukin-1 signalling and lymph node organogenesis in TRAF6-deficient mice. *Genes Cells* 1999; **4**: 353 - 62.

Nakashima K, Zhou X, Kunkel G, Zhang Z, Deng JM, Behringer RR, de Crombrugghe B. The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. *Cell.* 2002, **108(1)**: 17-29

Nehls MC, Rippe RA, Veloz L, Brenner DA. Transcription factors nuclear factor I and Sp1 interact with them urine collagen alpha 1(I) promoter. *Mol cell boil* 1991; **11**; 4065-4073

Nehls MC, Grapilon ML, Brenner DA. NF1/Sp1 switch elements regulate collagen α1(I) gene expression. *DNA Cell Biol* 1992; **11**: 443–452

Noda M, Camilliere JJ. In vivo stimulation of bone formation by transforming growth factor-beta. *Endocrinology.* 1989, **124(6)**: 2991-4.

Noma T, Mizushige K, Yao L, Yu Y, Kiyomoto H, Hosomi N, Kimura S, Abe Y, Ohmori K, Matsuo H. Alteration in aortic wall stiffness and accumulation of collagen during the prediabetic stage of type II diabetes mellitus in rats. *Jpn Circ J.* 1999; **63**: 988–993

Nordstrom A, Gerdhem P, Brandstrom H, Stiger F, Lerner UH, Lorentzon M, Obrant K, Nordstrom P, Akesson K. Interleukin-6 promoter polymorphism is associated with bone quality assessed by calcaneus ultrasound and previous fractures in a cohort of 75-year-old women. *Osteoporos Int.* 2004, **15**, 820–826.

Orban PC, Chui D, Marth JD. Tissue- and site-specific DNA recombination in transgenic mice. *Proc Natl Acad Sci U S A.* 1992, **89(15)**: 6861-5.

Otto F, Thornell AP, Crompton T, Denzel A, Gilmour KC, Rosewell IR, Stamp GW, Beddington RS, Mundlos S, Olsen BR, Selby PB, Owen MJ. Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. *Cell.* 1997, **89(5)**:765-71

Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Sato H, Hori M, Nakamura Y, Tanaka T. Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. *Nat Genet* 2002; **32**:650-654

Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, Fisher SA, Roberts RG, Nimmo ER, Cummings FR, Soars D, Drummond H, Lees CW, Khawaja SA, Bagnall R, et al. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet.* 2007, **39(7)**: 830-2

Peltonen L, McMkusick VA. Genomic and medicine. Dissecting human disease in the postgenomic era. *Science* 2001; **291**: 1224-29

Pereira RF, Hume EL, Halford KW, Prockop DJ. Bone fragility in transgenic mice expressing a mutated gene for type I procollagen (COL1A1) parallels the age-dependent phenotype of human osteogenesis imperfecta. *J Bone Miner Res.* 1995; **10**:1837-43

Pocock NA, Eisman JA, Hopper JL, Yeates MG, Sambrook PN, and Eberl S. Genetic determinants of bone mass in adults: A twin study. *J Clin Inves.* 1987, 80: 706–710.

Poppleton HM, Raghow R. Transcriptional activation of the minimal human proalpha 1(I) collagen promoter: obligatory requirement for Sp1. *Biochem J* 1997; **323**; 225-231

Quinn JM, Itoh K, Udagawa N, Hausler K, Yasuda H, Shima N, Mizuno A, Higashio K, Takahashi N, Suda T, Martin TJ, Gillespie MT. *J Bone Miner Res.* 2001 **16**, 1787–1794.

Qureshi AM, Fogelman I, Grant SFA, Herd R, Blake G, Ralston SH. Polymorphisms of the COL1A1 gene predict BMD response to cyclical etidronate therapy. *Calcif Tiss Int*, 2002, **70**: 158–163

Qureshi AM, McGuigan FE, Seymour DG, Hutchison JD, Reid DM, Ralston SH. Association between COLIA1 Sp1 alleles and femoral neck geometry. *Calcif Tissue Int* 2001, **69**:67–72

Rahkonen O, Su M, Hakovirta H, Koskivirta I, Hormuzdi SG, Vuorio E, Bornstein P, Penttinen R. Mice with a deletion in the first intron of the Col1a1 gene develop age-dependent aortic dissection and rupture. *Circ Res* 2004; **94(1)**:83-90

Raisz LG. Physiology and pathophysiology of bone remodeling. *Clinical Chemistry*. 1999, **45**, 1353-1358

Ralston SH & de Crombrugghe B. Genetic regulation of bone mass and susceptibility to osteoporosis. *Genes Dev.* 2006, **20(18)**:2492-506

Ralston SH, Uitterlinden A, Brandi ML, Balcells S, Langdahl BL, Lips P, et al. Large-Scale Evidence for the Effect of the COLIA1 Sp1 Polymorphism on Osteoporosis Outcomes: The GENOMOS Study. *PLoS Med* 2006, **3(4):** e90.

Ramirez F, Bernard M, Chu ML, Dickson L, Sangiorgi F, Weil D, de Wet W, Junien C, Sobel M. Isolation and characterization of the human fibrillar collagen genes. *Ann N Y Acad Sci* 1985, **460**:117–129.

Rauch F, Glorieux FH. Osteogenesis imperfecta. *Lancet*. 2004; **363**: 1377-85

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES. Linkage disequilibrium in the human genome. *Nature* 2001; **411**:199-204.

Riancho JA, Valero C, Zarrabeitia MT. MTHFR polymorphism and bone mineral density: meta-analysis of published studies. *Calcif Tissue Int.* 2006, **79(5)**:289-93.

Rickels MR, Zhang X, Mumm S, Whyte MP. Oropharyngeal skeletal disease accompanying high bone mass and novel LRP5 mutation. *J Bone Miner Res* 2005, **20**: 878–885

Risch N, Teng J. The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. *Genome Res*. 1998, **8(12)**: 1273-88

Rossert J, Terraz C, Dupont S. Regulation of type I collagen genes expression. *Nephrol Dial Transplant* 2000; **15**; 66-68

Rowe DW. Osteoporosis Imperfecta. *In* "Bone and Mineral Research" (J. N. M. Heersche and J. A. Kanis, Eds.), pp.209-241, 1991, Elsevier, Amsterdam.

Ruf C, Carosone-Link P, Springett J, Bennett B. Confirmation and genetic dissection of a major quantitative trait locus for alcohol preference drinking. *Alcohol Clin Exp Res*. 2004, **28(11)**: 1613-21

Saftig P, Hunziker E, Wehmeyer O, Jones S, Boyde A, Rommerskirch W, et al. Impaired osteoclastic bone resorption leads to osteopetrosis in cathepsin-K-deficient mice. *Proc Natl Acad Sci* 1998; **95**: 13453 - 8.

Saitta B, Gaidarova S, Cicchillitti L, Jimenez SA. CCAAT binding transcription factor binds and regulates human COL1A1 promoter activity in human dermal fibroblasts: demonstration of increased binding in systemic sclerosis fibroblasts. *Arthritis Rheum* 2000, **43**: 2219–2229.

Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, König IR, Stevens SE, Szymczak S, et al. Genomewide Association Analysis of Coronary Artery Disease. *N Engl J Med*. 2007 Jul 18

Sauer B. Manipulation of Transgenes by Site-Specific Recombination: Use of Cre Recombinase, *Methods in Enzymology* 1993; **225**:890-900.

Seeman E, Tsalamandris C, Formica C, Hopper JL, McKay J. Reduced femoral neck bone density in the daughters of women with hip fractures: the role of low peak bone density in the pathogenesis of osteoporosis. *J Bone Miner Res*. 1994, **9(5)**: 739-43

Shastry BS. Transcription factor IIIA (TFIIIA) in the second decade. *J Cell Sci*. 1996, **109**, 535-9

Schnieke A, Harbers K, Jaenisch R. Embryonic lethal mutation in mice induced by retrovirus insertion into the α1(I) collagen gene. *Nature* 1983; **304:**315-320

Shen ZJ, Nakamoto T, Tsuji K, Nifuji A, Miyazono K, Komori T, et al. Negative regulation of bone morphogenetic protein/Smad signaling by Cas-interacting zinc finger protein in osteoblasts. *J Biol Chem* 2002; **277**: 29840– 6

Silbiger S, Lei J, Neugarten J. Estradiol suppresses type I collagen synthesis in mesangial cells via activation of activator protein-1. *Kidney Int* 1999, **55**:1268–1276

Silva MJ, Gibson LJ. Modeling the mechanical behavior of vertebral trabecular bone: Effects of age-related changes in microstructure. *Bone*. 1997; **21**: 191-199

Sims NA, Jenkins BJ, Quinn JM, Nakamura A, Glatt M, Gillespie MT, Ernst M, Martin TJ. Glycoprotein 130 regulates bone turnover and bone size by distinct downstream signaling pathways. *J Clin Invest*. 2004, **13(3)**:379-89.

Simonet WS, Lacey DL, Dunstan CR, Kelley M, Chang MS, Lüthy R, Nguyen HQ, et al. Osteoprotegerin: a novel secreted protein involved in the regulation of bone density. *Cell*. 1997, **89(2)**: 309-19

Slack JL, Liska DJ, Bonstein P. Regulation of expression of the type I collagen genes. *Am J Med Genet* 1993; **45**; 140-151

Slack JL, Parker MI, Bornstein P. Transcriptional repression of the α1(I) collagen gene by ras is mediated in part by an intronic AP1 site. *J Cell Biochem* 1995; **58**:380–392

Slemenda CW, Christian JC, Williams CJ, Norton JA, Johnston CC. Genetic determinants of bone mass in adult women: a reevaluation of the twin model and the potential importance of gene interaction on heritability estimates. *J Bone Miner Res*. 1991, **6(6)**: 561-7

Smith DM, Nance WE., Kang KW, Christian JC, and Johnston CC. Genetic factors in determining bone mass. *J Clin Invest*. 1973, **52**: 2800–2808.

Snieder H, MacGregor AJ, and Spector TD. Genes control the cessation of a women's reproductive life: a twin study of hysterectomy and age at menopause. *J Clinc Emdocrio Metabo* 1998, **83**: 1875-1880

Soriano P, Montgomery C, Geske R, Bradley A. Targeted disruption of the c-src proto-oncogen leads to osteopetrosis in mice. *Cell* 1991, **64**, 693-702.

Spinella-Jaegle S, Roman-Roman S, Faucheu C, Dunn FW, Kawai S, Gallea S, Stiot V, Blanchet AM, Courtois B, Baron R, Rawadi G. Opposite effects of bone morphogenetic protein-2 and transforming growth factor-β1 on osteoblast differentiation. *Bone* 2001, **29**:323–330.

Stewart TL, Roschger P, Misof BM, Mann V, Fratzl P, Klaushofer K, Aspden RM, Ralston SH. Association of COL1A1 Sp1 alleles with defective bone nodule formation in vitro and abnormal bone mineralization in vivo. *Calcif Tiss Int* 2005, **77**, 113-118

Stewart TL, Jin H, McMuigan FE, ALbagha OM, Garcia-Giralt N, Bassiti A et al. Haplotypes defined by promoter and intron1 polymorphisms of COL1A1 gene regulate bone mineral density in women. *J Clin Endocrinol Metab* 2006

Stewart TL, Mann V. Extraction of nucleic acids from bone. *Methods Mol Med.* 2003; **80**: 425-32

Stephens M, Smith N, and Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J of Hum Genet* 2001, **68**, 978--989.

Stephens M and Donnelly P. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J of Hum Genet*, 2003, **73(5)**:1162-9

Styrkarsdottir U, Cazier JB, Kong A, Rolfsson O, Larsen H, Bjarnadottir E, Johannsdottir VD, Sigurdardottir MS, Bagger Y, Christiansen C, Reynisdottir I, Grant SF, Jonasson K, Frigge ML, Gulcher JR, Sigurdsson G, Stefansson K. Linkage of osteoporosis to chromosome 20p12 and association to BMP2. *PLoS*

Biol 2003; **1**:1-9.

Suda T, Takahashi N, Udagawa N, Jimi E, Gillespie MT, Martin TJ. Modulation of osteoclast differentiation and function by the new members of the tumor necrosis factor receptor and ligand families. *Endocr Rev* 1999, **20(3)**: 345-57.

Teitelbaum SL. Bone resorption by osteoclast. *Science* 2000, **289**: 1504-1508

Terry CF, Loukaci V, Green FR. Cooperative influence of genetic polymorphisms on interleukin-6 transcriptional regulation. *J Biol Chem.* 2000, **275**, 18138 - 18144.

Thakkinstian A, DÉste C, Eisman J, Nguyen T, Attia J. Meta-analysis of molecular association studies: vitamin D receptor gene polymorphisms and BMD as a case study. *J Bone Miner Res* 2004, **19**, 419-428

Thunyakitpisal P, Alvarez M, Tokunaga K, Onyia JE, Hock J, Ohashi N, Feister H, Rhodes SJ, Bidwell JP. Cloning and functional analysis of a family of nuclear matrix transcription factors (NP/NMP4) that regulate type I collagen expression in osteoblasts. *J Bone Miner Res.* 2001, **16(1)**:10-23.

Thies RS, Bauduy M, Ashton BA, Kurtzberg L, Wozney JM, Rosen V. Recombinant human bone morphogenetic protein-2 induces osteoblastic differentiation in W-20-17 stromal cells. *Endocrinology* 1992, **130**:1318–1324.

Thirunavukkarasu K, Miles RR, Halladay DL, Yang X, Galvin RJ, Chandrasekhar S, Martin TJ, Onyia JE. *J Biol Chem* 2001. **276**, 36241–36250.

Tintut Y, Parhami F, Le V, Karsenty G, Demer LL. Inhibition of osteoblast-specific transcription factor Cbfa1 by the cAMP pathway in osteoblastic cells. Ubiquitin/ proteasome-dependent regulation. *J Biol Chem.* 1999, **274(41)**: 28875-9

Tondravi MM, McKercher SR, Anderson K, Erdmann JM, Quiroz M, Maki R and Teitelbaum SL. Osteopetrosis in mice lacking haematopoietic transcription factor PU.1. *Nature* 1997, **385**: 81-84

Torgerson DJ, Thomas RE, Campbell MK, Reid DM. Randomized trial of osteoporosis screening. Use of hormone replacement therapy and quality-of-life results. *Arch Intern Med.* 1997, **157(18)**: 2121-5

Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, Lowe CE, Szeszko JS, Hafler JP, Zeitels L, Yang JH, Vella A, Nutland S, et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet.* 2007, **39(7)**: 857-64

Turgeman G, Zilberman Y, Zhou S, Kelly P, Moutsatsos IK, Kharode YP, Borella LE, Bex FJ, Komm BS, Bodine PV, Gazit D. Systemically administered rhBMP-2

promotes MSC activity and reverses bone and cartilage loss in osteopenic mice. *J Cell Biochem* 2002, **86:** 461–474

Turner CH, Burr DB. Basic biomechanical measurements of bone: A tutorial. *Bone*. 1993; **14**: 595 608.

Turner CH Biomechanics of bone: Determinants of skeletal fragility and bone quality *Osteoporos lnt*. 2002; **13**: 97-104

Turner CH. Bone strength: current concepts. *Ann N Y Acad Sci*. 2006, **1068**: 429-446

Uitterlinden AG, Burger H, Huang Q, Yue F, McGuigan FE, Grant SF, Hofman A, van Leeuwen JP, Pols HA, Ralston SH. Relation of alleles of the collagen type I α 1 gene to bone density and risk of osteoporotic fractures in postmenopausal women. *N. Engl. J. Med.* **338**:1016–1022.

Uitterlinden AG, Ralston SH, Brandi ML, Carey AH, Grinberg D, *et al.* GENOMOS study. The association between common vitamin D receptor gene variations and osteoporosis: a participant-level meta-analysis. *Ann Intern Med.* 2006; **15**; 255-64.

Uitterlinden AG, Weel AE, Burger H, Fang Y, van Duijn CM, Hofman A, van Leeuwen JP, Pols HA. Interaction between the vitamin D receptor gene and collagen type I α1 gene in susceptibility for fracture. *J Bone Miner Res* 2001, **16**:379–385

Urano T, Shiraki M, Ezura Y, Fujita M, Sekine E, Hoshino S, Hosoi T, Orimo H, Emi M, Ouchi Y, Inoue S. Association of a single-nucleotide polymorphism in low-density lipoprotein receptor-related protein 5 gene with bone mineral density. *J Bone Miner Metab*. 2004; **22(4)**:341-5.

Utomo ARH, Nikitin AU, Lee WH. Temporal, spatial, and cell type−specific control of Cre-mediated DNA recombination in transgenic mice. *Nature Biotechnology* 1999; **17,** 1091 - 1096

Van der Mark K. Structure, biosynthesis, and gene regulation of collagens in cartilage and bone. In: Dynamics of bone and cartilagen metabolism. Seibel MJ, Robins SP & Bilezikian JP (eds.) Academic Press 1999, pp3-29.

Van Hul E, Gram J, Bollerslev J, Van Wesenbeeck L, Mathysen D, Andersen PE, Vanhoenacker F, Van Hul W. Localization of the gene causing autosomal dominant osteopetrosis type I to chromosome 11q12–13. *J Bone Miner Res* 2002, **17**: 1111–1117

van Meurs JB, Rivadeneira F, Jhamai M, Hugens W, Hofman A, van Leeuwen JP, Pols HA, Uitterlinden AG. Common genetic variation of the low-density lipoprotein receptor-related protein 5 and 6 genes determines fracture risk in elderly white men. *J Bone Miner Res*. 2006, **21(1)**:141-50

van Wesenbeeck L, Cleiren E, Gram J, Beals RK, Bénichou O, Scopelliti D, Key L, Renton T, Bartels C, Gong Y, Warman ML, De Vernejoul MC, Bollerslev J, Van Hul W. Six novel missense mutations in the LDL receptor-related protein 5 (LRP5) gene in different conditions with an increased bone density. *Am J Hum Genet.* 2003, **72(3)**: 763-71

Vergee WP, Sogo JM, Pretorius PJ, de Vries WN. Interaction of AP1, AP2, and Sp1 with the regulatory regions of the human pro-alpha 1(I) collagen gene. *Arch Biolchem Bioph* 2000, **377**; 69-79

Viguet-Carrin S, Garnero P, Delmas PD. The role of collagen in bone strength. *Osteoporosis Int* 2006; **17**: 319-336

Villadsen MM, Bu¨ nger MH, Carstens M, Stenkjaer L, Langdahl BL. Methylenetetrahydrofolate reductase (MTHFR) C677T polymorphism is associated with osteoporotic vertebral fractures, but is a weak predictor of BMD. *Osteoporos Int* 2005, **16**: 411-416

Von der Mark. Structure, biosynthesis, and gene regulation of collagens in cartilage and bone. In "*Dynamics of bone and cartilagen metabolism*" pp3-29. Editted by Seibel MJ, Robins SP & Bilezikian JP. Academic Press.

Walsh MC, Choi Y. Osteoclast Biology of the TRANCE axis. *Cytokine Growth Factor Rev* 2003, **14**: 251-263

Wang ZQ, Ovitt C, Grigoriadis AE, Mohle-Steinlein U, Ruther U, Wagner EF. Bone and haematopoietic defects in mice lacking c-fos. *Nature* 1992; **360**: 741 - 5.

Weichetová M, Stepán JJ, Michalská D, Haas T, Pols HA, Uitterlinden AG. COLIA1 polymorphism contributes to bone mineral density to assess prevalent wrist fractures. *Bone*. **26**:287–290.

Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 2007, **447(7145)**: 661-78.

Whyte MP. Hypophosphatasia and the role of alkaline phosphatase in skeletal mineralization. Endocr Rev 1994, **15**: 439-461

Whyte MP, Reinus WH, Mumm S High-bone-mass disease and LRP5. *N Engl J Med* 2004, **350**: 2096–2099

Williams FMK and Spector TD. Recent advances in the genetics of osteoporosis. *J Muscu Neuro Int* 2006; **6(1)**:27-35

Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational inferences about departures from Hardy-Weinberg equilibrium. *Am J Hum Genet.* 2005; **76(6):**967-86.

Wolffe AP: Architectural transcription factors. *Science* 1994, **264**:1100-1101

Yamada Y, Ando F, Niino N, Shimokata H. Association of a -1997G-->T polymorphism of the collagen Ialpha1 gene with bone mineral density in postmenopausal Japanese women. *Hum Biol.* 2005, **77(1)**: 27-36

Yamada Y, Miyauchi A, Goto J, Takagi Y, Okuizumi H, Kanematsu M, Hase M, Takai H, Harada A, Ikeda K. Association of a polymorphism of the transforming growth factorbeta1 gene with genetic susceptibility to osteoporosis in postmenopausal Japanese women. *J Bone Miner Res* 1998, **13**: 1569–1576

Yamada Y, Hosoi T, Makimoto F, Tanaka H, Seino Y, Ikeda K. Transforming growth factor beta-1 gene polymorphism and bone mineral density in Japanese adolescents. *Am J Med* 1999, **106**: 477–479.

Yasuda H, Shima N, Nakagawa N, Yamaguchi K, Kinosaki M, Mochizuki S, Tomoyasu A, Yano K, Goto M, Murakami A, Tsuda E, Morinaga T, Higashio K, Udagawa N, Takahashi N, Suda T. Osteoclast differentiation factor is a ligand for osteoprotegerin/osteoclastogenesis-inhibitory factor and is identical to TRANCE/RANKL. *Proc Natl Acad Sci U S A.* 1998, **95(7)**:3597-602.

Yazdanpanah N, Rivadeneira F, van Meurs JB, Zillikens MC, Arp P, Hofman A, van Duijn CM, Pols HA, Uitterlinden AG. The -1997 G/T and Sp1 Polymorphisms in the Collagen Type I alpha1 (COLIA1) Gene in Relation to Changes in Femoral Neck Bone Mineral Density and the Risk of Fracture in the Elderly: The Rotterdam Study. *Calcif Tissue Int.* 2007

Yui MA, Muralidharan K, Moren-Altamirano B, Perrin G, Chestnut K, Wakeland EK. Production of congenic mouse strains carrying NOD-derived diabetogenic genetic intervals: an approach for the genetic dissection of complex traits. Mamm Genome 1996, **7**, 331–334

Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science.* 2007, **316(5829)**: 1336-41

Zhang DE, Hetherington CJ, Chen HM, Tenen DG. The macrophage transcription factor PU.1 directs tissue-specific expression of themacrophage colony-stimulating factor receptor. *Mol Cell Biol* 1994; **14**: 373 - 81.

Zhang YY, Lei SF, Mo XY, Wang YB, Li MX, Deng HW. The -1997 G/T polymorphism in the COLIA1 upstream regulatory region is associated with hip bone mineral density (BMD) in Chinese nuclear families. *Calcif Tissue Int.* 2005; **76(2):** 107-12.

Zmuda JM, Sheu YT, Moffett SP. The search for human osteoporosis genes. *J Musculoskelet Neuronal Interact* 2006; **6(1)**:3-15

APPENDICES

Appendix 1 Primers Used for Sequencing of Targeting Vectors

1.1 Primers Located in Mouse col1a1 Promoter Region

Name	Sequence (5'→3')	Position in targeting vector
Reverse primer		
MouProSeq1F	GACTGCAGGGGCTATAAT	5917
MouProSeq2F	GCCTGAGGTTCTGATGGA	5382
MouProSeq3F	GAATGTAGGGCTTTGTCA	4958
MouProSeq4F	GGCATTTAGGTCTGGAGT	4495
MouProSeq5F	CCGCAGGAACATGGAGAC	3920
MouProSeq6F	CCTAAGGGACTCGCTCAT	3453
MouProSeq7F	GTATAGGGCTGGTGAGAT	2869
MouProSeq8F	GCCCCGGCTTCCTCTACT	2458
Forward primer		
MouProSeq9R	TCAGGCCCAAAGATGTAT	2679
MouProSeq10R	ACTTGGGAGCTGCGATGC	3221
MouProSeq11R	TCCAATCACTCTCTCCAT	3687
MouProSeq12R	TAGTACTGAAGGTGAGCT	4306
MouProSeq13R	AACTCAGAGCTCAGCACT	4802
MouProSeq14R	CAATCACCAGTTCATACC	5394
MouProSeq15R	TGGGATAATTGATGAAGT	5822

1.2 Primers Located in 5' Flank of Human COL1A1 Gene

Name	Sequence (5'→3')	Position in targeting vector
Reverse primer		
HumProSeq1F	GTGCCCCAGCCAATCAGA	8227
HumProSeq2F	GCAGGCAAGGGAGTTTTA	7674
HumProSeq3F	TGGGCTGCAGAGGATGAT	7122
HumProSeq4F	CCCGACCATGTGGCAGCA	6466
HumIntSeq1F	CTCATCCCGCCCCCATTC	9552
HumIntSeq2F	GCCAAAGTTTCTCATCCT	9010
HumIntSeq3F	CCCTCATCATCTCCCTTC	8645
Forward primer		
HumProSeq5R	ACTCCCCGCTTCCACGTT	6747
HumProSeq6R	ATGATCACAGGCCTCCCA	7359
HumProSeq7R	CTTAGCTCTTGCCAGGAT	7855
HumProSeq8R	CTCCGAGGGGCAGGGTTC	8246
HumIntSeq4R	CGCGGCGTGAGTTGGTGC	8736
HumIntSeq5R	TGCGGACTTTTTGGTTCT	9317
HumIntSeq6R	TGGAAAGAGTTACAGCCT	9770

1.3 Primers Located in Mouse Exons (2-22)

Name	Sequence (5'→3')	Position in targeting vector
Reverse primer		
MouExSeq1F	TGTGACTTACGGGTTCTC	16479
MouExSeq2F	CGGAGGAAAGGACAAATC	15865
MouExSeq3F	GCAGTGCCAGGAGGTCCA	15288
MouExSeq4F	ATGGTGACTTGGGGCTTT	14675
MouExSeq5F	CCTGCTGGGAAGATTTCA	14102
MouExSeq6F	GAGGGGCTTTTGATATCT	13537
MouExSeq8F	CCTGCACCAAGAGACGGA	10515
MouExSeq9F	AGGATGTGTGCGGGTCAG	9898
Forward primer		
MouExSeq10R	GGTTCTGGTCCTAGGGTT	10119
MouExSeq11R	TGTGACTTCCCGTTTTGT	10645
MouExSeq12R	CCTGGCTTTCAATTTACC	13186
MouExSeq13R	CTGGTGAACCTGGCGAGC	13785
MouExSeq14R	TGTAGGGGGAAGCTGGCA	14396
MouExSeq15R	TGAATGACCTGTGCCCAA	15034
MouExSeq16R	CATCCACAGGGTCCCACC	15575
MouExSeq17R	CCTTGGGCTAAAATGTCT	16164
MouExSeq18R	AGGAGCCCGTGGTGAGCC	16704

1.4 Primers Located in pBlueScript II (-) Plasmid

Name	Sequence (5'→3')	Position in targeting vector
M13F	GTAAAACGACGGCCAGT	17000
M13R	GGAAACAGCTATGACCAT	2136

1.5 Primers Located in Loxp-neo-loxp Cassette

Name	Sequence (5'→3')	Position in targeting vector
Loxpseq2	GATGCTCTTCGTCCAGATCA	12042
Loxpseq4	CTCCTCTTCCTCATCTCCGG	11505
Loxpseqrev	CGTAAACTCCTCTTCAGACC	10988
Loxpseqfor	ACTCTAGAGCTTGCGGAACC	12829

Appendix 2 The Full Sequence of Targeting Vector

```
   1 GTGGCACTTT TCGGGGAAAT GTGCGCGGAA CCCCTATTTG TTTATTTTTC TAAATACATT
  61 CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAAT GCTTCAATAA TATTGAAAAA
 121 GGAAGAGTAT GAGTATTCAA CATTTCCGTG TCGCCCTTAT TCCCTTTTTT GCGGCATTTT
 181 GCCTTCCTGT TTTTGCTCAC CCAGAAACGC TGGTGAAAGT AAAAGATGCT GAAGATCAGT
 241 TGGGTGCACG AGTGGGTTAC ATCGAACTGG ATCTCAACAG CGGTAAGATC CTTGAGAGTT
 301 TTCGCCCCGA AGAACGTTTT CCAATGATGA GCACTTTTAA AGTTCTGCTA TGTGGCGCGG
 361 TATTATCCCG TATTGACGCC GGGCAAGAGC AACTCGGTCG CCGCATACAC TATTCTCAGA
 421 ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT TACGGATGGC ATGACAGTAA
 481 GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC TGCGGCCAAC TTACTTCTGA
 541 CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTTGCA CAACATGGGG GATCATGTAA
 601 CTCGCCTTGA TCGTTGGGAA CCGGAGCTGA ATGAAGCCAT ACCAAACGAC GAGCGTGACA
 661 CCACGATGCC TGTAGCAATG GCAACAACGT TGCGCAAACT ATTAACTGGC GAACTACTTA
 721 CTCTAGCTTC CCGGCAACAA TTAATAGACT GGATGGAGGC GGATAAAGTT GCAGGACCAC
 781 TTCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA TAAATCTGGA GCCGGTGAGC
 841 GTGGGTCTCG CGGTATCATT GCAGCACTGG GGCCAGATGG TAAGCCCTCC CGTATCGTAG
 901 TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA
 961 TAGGTGCCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA AGTTTACTCA TATATACTTT
1021 AGATTGATTT AAAACTTCAT TTTTAATTTA AAAGGATCTA GGTGAAGATC CTTTTTGATA
1081 ATCTCATGAC CAAAATCCCT TAACGTGAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG
1141 AAAAGATCAA AGGATCTTCT TGAGATCCTT TTTTTCTGCG CGTAATCTGC TGCTTGCAAA
1201 CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTTGCCGGA TCAAGAGCTA CCAACTCTTT
1261 TTCCGAAGGT AACTGGCTTC AGCAGAGCGC AGATACCAAA TACTGTCCTT CTAGTGTAGC
1321 CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA
1381 TCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA
1441 GACGATAGTT ACCGGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGGTTCG TGCACACAGC
1501 CCAGCTTGGA GCGAACGACC TACACCGAAC TGAGATACCT ACAGCGTGAG CTATGAGAAA
1561 GCGCCACGCT TCCCGAAGGG AGAAAGGCGG ACAGGTATCC GGTAAGCGGC AGGGTCGGAA
1621 CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCTG GTATCTTTAT AGTCCTGTCG
1681 GGTTTCGCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATG CTCGTCAGGG GGGCGGAGCC
1741 TATGGAAAAA CGCCAGCAAC GCGGCCTTTT TACGGTTCCT GGCCTTTTGC TGGCCTTTTG
1801 CTCACATGTT CTTTCCTGCG TTATCCCCTG ATTCTGTGGA TAACCGTATT ACCGCCTTTG
1861 AGTGAGCTGA TACCGCTCGC CGCAGCCGAA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG
1921 AAGCGGAAGA GCGCCCAATA CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCATTAAT
1981 GCAGCTGGCA CGACAGGTTT CCCGACTGGA AAGCGGGCAG TGAGCGCAAC GCAATTAATG
2041 TGAGTTAGCT CACTCATTAG GCACCCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT
2101 TGTGTGGAAT TGTGAGCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC CATGATTACG
2161 CCAAGCGCGC AATTAACCCT CACTAAAGGG AACAAAAGCT GGAGCTCCAC CGCGGTGGCG
2221 GCCGCTCTAG AACTAGtGGA TCCCAAGATT AGGGCAGACC CTCATTCACA ACCCGGTCTG
2281 TTGCAGTCTT TGGTTATGAG GTTTATTTCC GTCTCTTTAT CAGATGTGGG AATCCACAGG
2341 GTCTCTGCAG CTCGCTGAAT CCCAAGAGGA CCAAGAGGAT GAACCCTTGA TGTTTAAGCA
2401 CTGTGGGGTG ACTACAACAG AAGTGAACCC TAACTTCTCA GTAGAGGAAG CCGGGGCAGG
```

2461 GGGGGGGCAA AAACCCTGTG GGGAAGAACA TGGGACCATC ATATCCTCCT TTATACTGGA
2521 TTCTTTTTTT TTTTAAGATT TATTTATTTA TTTTATGTAT ATGAGTACAC TGGAGCTGTA
2581 CAGATGGCTG TGAGCCTTTA TGTGGTCATT GGGAATTGAA TTTTTAGGAC CTCTGGTTGC
2641 TCTGGTCAAC CCTGCTCCCT CCAGTCAGCC CTGCTCGCTC AGGCCCAAAG ATGTATTTAA
2701 TATTATACGT AAATACACTA TAGCTGACTT CAGACCCACC AGAAGAGGGC ATAGGATCTC
2761 ATTACAGGTG GTTGTGAGCC ACCATGTGGT TGCTGGGATT TGAACTCAGG ACCTTCAGAA
2821 GAACAATCAG CGTCTCTTAC CCGCTGAGCC ATCTCACCAG CCCTATACTG GATTTTTTCT
2881 GCTGCAAGCC ACATTCTGTT CAGCCAGACC CCTAAGGAGG GATCAAAGGA GGCTGGTACA
2941 TAAGTGTAAG ATTGAAGGCA TCTGTGGCTG AGGCTTACCT GGAGAATGGA GACAATATCT
3001 GGATAGAGCC CCTGACTCAA ACACCATCTC CTCACCCAGC TCTACATAGA CACACTCTAA
3061 AACTTGAAAA TCAAGCTTCA TTGGATTCTC TGCATTTATA TCCTTGTAGT TCAAGCTGGA
3121 ATCTGGTGGT GATACAGGAC CCTTAGGCCT CTTCCCCTCC CGAAGGCAAC TGGACATCCT
3181 TCTCTTGCTG AGAAAGGATG CAGATACTGG AGAACTGGAG ACTTGGGAGC TGCGATGCCA
3241 AGCAAGGATG TTAGAAGAGA AGACAAAGGG GAAAACACTG CATGGGTGGA GGGGTGGGGC
3301 AATGGACAGA GGGAAAGAGA GGACACTGTG AAAAATGCAT GAAAAGGAGA CTTCGAGGGA
3361 GAGGAGCAGA AGGGCAGGGG GAAGAGAAGT GCAAAGCAAA GGGGGGGGGG CGGAGAAAGG
3421 TGAAGCTAGG GAGGAATGAG CGAGTCCCTT AGGGACTAAA TGAAGGTCCA GGCACTTGAT
3481 GCCTCCAGAA TTCAGATAAC AGGAAGGAGG GGTCTGACGT CTCCTCTAAA GAGGAGTAAA
3541 GAAGAGCAGG AGTTTTCCTT CAGAACCTCC CCTCCGTCCT TTTGACATAC CTGTATCCCC
3601 ACTCACTTTA TGGCTGAAGG CTTTCTTTGC CCGCCATGTT TTACTGAGCT CTGCAGTTCT
3661 CTTTGGAGCA CCCACCTCTT CCTATTTCCA ATCACTCTCT CCATCCCTCT CTTCCTTCCC
3721 TTCTCCCTCA ATAGGCCATT CTTGTCCCAT CTCCTGTGTG TGGTCCTCTT CTCCCAAATC
3781 CTTCACTCTG AGGTGGCTCG AGACTTGGCT CTAGCCCTCA GCTCCCCCGC CCCCATGCTT
3841 TCTCTGGGGA CAACTACACA CACACACACA CACACACACA CACACACACA CACACAGCTT
3901 CAGTCTCCAT GTTCCTGCGG TCCCCCCCCC CATGTTTCTT CTCTAAAATC TAGAATGCAG
3961 AAGCCAAGGA TTTCCTTTTC TCTCTTCTCT TTCTTTTTTC CTTTTTCTCT TTCCTCCCCC
4021 CCCCCCAAGA TGGAGTTTCT CTGTGTAGCC CTGGCTGTCC TGGAACTCAC TCTATAGACC
4081 AGGTTAGCCT CAAACTTAGA GATCTGCCTG CCTCTGGCTC TTGAGTGCTT AGAATTGAAG
4141 GTGTAAACCA CTGCTACCTG ACAGAGGATT TCTTACACAT GTCTGACTGG AGCGCTAAGA
4201 GGCATCAAAC TCAGCATAAG CAGACCTGAA CTTGTTGGCT TTCCCTTTAA ATCAGCCAGG
4261 AAAGGAGGGA TGGGGAGAGG AAGGAAATGC TCAGAAGGAA AGGAGTAGTA CTGAAGGTGA
4321 GCTGGAGGAA CCCAGATCGG GCCATACAGA AGAAAAAGGA GAAGCAACGC ACCCACACAC
4381 CTAGGCTCCA CCCACAGATG TTGCTCTGGG TACCTCTGTC TGAAGACTGT AGGGCTTTGT
4441 GATGGAGGGT GGGGTATGCT TCATGCCCAT GGCCTTTACT CCAGACCTAA ATGCCCACCC
4501 CCACATACAG CTGCTCTTTT TCCCTGCCCT TCTCCCGAGA GACCAGCTCT CCATCCCTGG
4561 TCTGCAGCCA AGGCTGGGGG CAGAAGAACT TTCTGGAGGA TCTGAGTGAG AAAAGCAAGA
4621 GAGCCTCGAA TAGGGACTAG AACCACTGGG GCGGCAAGTG CAGAGGAGAC CGGGTGTGTG
4681 CCCTACCTGC TACATTTATA CCTAGGCTGC CTCTGCTTCT GTTCCACACC GTAGAGTCCA
4741 GACGGGTCTT CTCCCTCCCT CGTGGGGGGG GAAACATTTG GGAAAATTTG GAGAGAGACA
4801 GAACTCAGAG CTCAGCACTT TCCTCTTTCT GTTTTTCTTC TTGAGGAATT TTTTTCCCCC
4861 AACGGCTGAT GACTTTACCA TTCTTGGGGG TGGGGGGGGG TGGAGATTCT GGCTTTTGCT
4921 CTCCCTACAC TCCAAGTGCC TGACAAAGCC CTACATTCCT CAAAAAGCCA GAGCTTTAGA
4981 GTTTCCTAAA GATGAGGTGG CGTGGCGTGG CGTGGCGTTC TCCTCCCTTC CCAGCTCCAA
5041 CTCCCCCTCC CCCAGTCTCC AGCCCTCAGC CCAGCCAGGG AGGCCCCGCC AGGCTGGGAG

5101 GAGACCCCAA GCACATTCTT CCTCTCTCTG TCATGCTGCA GAAATTAAAG ACACATCTCT
5161 GAGCTGGGCA CCCGCCAATC GTTTCAAGTA GAGAGTGGCA GAGGAGGTCC TGAGCTGCAG
5221 CTCCATGCCA CGTGTAAAGG AAGCTTGGAA CCCATTGCCT TCAACTCCTG GGGCAAAGGC
5281 CTGGAAATCA GACACAGGGT GACGGCTGTC ACTGAATGCA CACACACCAA TCTCCCCCAC
5341 ACCCATTCTA GTGGTCTCCA AGTTTCCATC AGAACCTCAG GCCCTTAATG TAGCAATCAC
5401 CAGTTCATAC CTGTTAGGCT CTTGTCTCTT CTTCCAGGTG AGCCTTGGGG ATGCTAGATT
5461 GGGAGCAGGG AAAGGGAGCT TTGGATAAGT GCTTTTCTGA TACAAAACCC AACAAAGGGT
5521 GAGCAGATCA CACCGCGCGC GCGCGTGTGT GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT
5581 GTGCGCACGC TTCGTGGCAT TCTACCCTGG ACACAGCCAC CTCAAACACC TGTTCCCTCT
5641 GAGGTACACC ATCCTTGTCC ACTGTTGGGG AGTGGGGCAT CCTTTGGAAC TGACCACAGT
5701 GGAAGGCAGG ACTTTGGTGA GTTCCGGAAC TACCATCTCA GCATCTCAGC CCCAGCCTTG
5761 CCCTACAGGC ACTGGCATTA GGCGGGGGCA GATCCTGGGC CACAAGTTAC TGCCACATGG
5821 TTGGGATAAT TGATGAAGTC CTGTCCTTCC ATTGCTGTCT CCAGCTCTGC TTCTCTGGAA
5881 ACTCTATATT TTCCCTTTAA TTATAGCCCC TGCAGTCTCC CTCCGCCACC CCACCCGCAC
5941 TGCTCGGCCT AGCTGCCCAC GGCCAGCGAC GTGGCTCCCT CCCCTTCTGT TCCCTTGGTC
6001 TTTTTTTTTT TTCCCTTTGC CTTCGTTGCA CAAAACTAGC TCAGGGAGGG CGTGAAGGGG
6061 GGGGGGGGAG CAATGGAATC TTGGATGGTT TGGAGGAGGC GGGACTCCCT GCTTCCACGT
6121 TTACAGCTCT AAAGAAGCCG GTGGGGGAAG TGATACAGGA TGTCTATGGG CCCTGCAAGG
6181 AGACCCCAGT GCCTTCCCAC ACCACCacta gTCACTAACC CTCATACTAC CAAGCCCTGC
6241 GGCACCCTGC CCTAGACCAC CACTCTAAAT GTCTGTTCCC TCCAAAAACA GGACCCCTGT
6301 CGCCTATTAG GGAGGGGTTC TCTTGGAACT GACCCACAGT AGGGGGCAGG ACTTTGGTGG
6361 GTTCAAGAAC TGCCATCTCA GCACCTCAGC CCCCTAGTCC TGCCCTGCAG TCGCTGGCAC
6421 TAGGCGGGGG CAGACCCTGG GCCACAAGTT GCTGCCACAT GGTCGGGATA ATTGATGAAG
6481 GTCCATCCCT CCATTGCTGT CTCCAGCCCT GCCTCTCTGG AAACTCTATA TTTTCCCTTT
6541 AATTATAGCC CCTGCAGTCT CCCTCTGCTG CCCCACCCGC ACCGCTCATC CTGGCTGCCC
6601 ACGGCCAGCC GGCCAGCCGA CGTGGCTCCC TCCCCTTCTG TTCCTTTTTT TTCCCCTTTG
6661 CCTTCGTTGC ACAAAACCAG CTGGGGGAGG GCGTGGAGAG GGGCGGGGGG AGGCAATGGA
6721 ATCTTGGATG GTTTGGGGGA GGCGGGACTC CCCGCTTCCA CGTTTGCAGC TCTGGAGCAC
6781 CCGGGGTGGG GAGCTGCACA GGAGGGAGAG AAATGAACAG GGCACTGCAA GGAGACCCCC
6841 AGGCCTTCTC TCAGCCCTAC AGAGTTTCTC AGGACGAGGT AGATTGGGGT TGAGGCAGAG
6901 CCTTGTTGGG GGAATGGGAC ATGGAGGAAG AAAGGACGTG GAGTTCTAGA GCCATCTTCC
6961 TTAGATATAG CCTGCTGTCC TTCGGGTCCC CAGACCCTTT CAGAGTGTAC AGATGATTCT
7021 CTCTGGTTCC TAAGGCATAG AGCAATGACC GGGATTTTCA AGAAAGAGAT GAGGCAGTGG
7081 GAAGTAGCCC CTAAAACAAA GTCAATCATC CTCTGCAGCC CATCCCACAC CCCCAAAGGA
7141 AAGTTTCACC CAGACACCCA AAATATCCCA TACATCCCCA ACACTGAGTC CAGGTACAAC
7201 TGGAGAAGGG GCTTTATGCA GCTCCCAGAA AGACACCCCT TTAGCTAAGT GCCCTCCCTC
7261 CACCCAGGTT CTCTCTGGTT TGACTGTGCT GGGAAGGAGG GTCTCTAAGC AGCCCCTGGC
7321 CACAGCCATG GCAAACAAAA CTCTTCTCTA AGTCACCAAT GATCACAGGC CTCCCACTAA
7381 AAATACTTCC CAACTCTGGG GTGGAAGAGT TTGGGGGATG AATTTTTAGG GGATTGCAAG
7441 CCCCAATCCC CACCTCTGTG TCCCTAGAAT CCCCCACCCC TACCTTGGCT GCTCCATCAC
7501 CCAACCACCA AAGCTTTCTT CTGCAGAGGC CACCTAGTCA TGTTTCTCAC CCTGCACCTC
7561 AGCCTCCCCA CTCCATCTCT CAATCATGCC TAGGGTTTGG AGGAAGGCAT TTGATTCTGT
7621 TCTGGAGCAC AGCAGAAGAA TTGACATCCT CAAAATTAAA ACTCCCTTGC CTGCACCCCT
7681 CCCTCAGATA TCTGATTCTT AATGTCTAGA AAGGAATCTG TAAATTGTTC CCCAAATATT

7741 CCTAAGCTCC ATCCCCTAGC CACACCAGAA GACACCCCCA AACAGGCACA TCTTTTTAAT
7801 TCCCAGCTTC CTCTGTTTTG GAGAGGTCCT CAGCATGCCT CTTTATGCCC CTCCCTTAGC
7861 TCTTGCCAGG ATATCAGAGG GTGACTGGGG CACAGCCAGG AGGACCCCCT CCCCAACACC
7921 CCCAACCCTT CCACCTTTGG AAGTCTCCCC ACCCAGCTCC CCAGTTCCCC AGTTCCACTT
7981 CTTCTAGATT GGAGGTCCCA GGAAGAGAGC AGAGGGGCAC CCCTACCCAC TGGTTAGCCC
8041 ACGCCATTCT GAGGACCCAG CTGCACCCCT ACCACAGCAC CTCTGGCCCA GGCTGGGCTG
8101 GGGGGCTGGG GAGGCAGAGC TGCGAAGAGG GGAGATGTGG GGTGGACTCC CTTCCCTCCT
8161 CCTCCCCCTC TCCATTCCAA CTCCCAAATT GGGGGCCGGG CCAGGCAGCT CTGATTGGCT
8221 GGGGCACGGG CGGCCGGCTC CCCCTCTCCG AGGGGCAGGG TTCCTCCCTG CTCTCCATCA
8281 GGACAGTATA AAAGGGGCCC GGGCCAGTCG TCGGAGCAGA CGGGAGTTTC TCCTCGGGGT
8341 CGGAGCAGGA GGCACGCGGA GTGTGAGGCC ACGCATGAGC GGACGCTAAC CCCCTCCCCA
8401 GCCACAAAGA GTCTACATGT CTAGGGTCTA GACATGTTCA GCTTTGTGGA CCTCCGGCTC
8461 CTGCTCCTCT TAGCGGCCAC CGCCCTCCTG ACGCACGGCC AAGAGGAcag taagttccaa
8521 actctgggGA ATTCGGGAGT GCAAGGATAC TCTATATCGC GCCTTGCGCT TGGTCCCGGG
8581 GGCCGCGGCT TAAAACGAGA CGTGGATGAT CCGGAGACTC GGGAATGGAA GGGAGATGAT
8641 GAGGGCTCTT CCTCGGCGCC CTGAGACAGG AGGGAGCTCA CCCTGGGGCG AGGTTGGGGT
8701 TGAACGCGCC CCGGGAGCGG GAGGTGAGGG TGGAGCGCGG CGTGAGTTGG TGCAAGAGAG
8761 AATCCCGAGC GCGCAACCGG GGAAGTGGGG ATCTGGGTGC AGAGTGAGGA AAGCACGTCG
8821 AAGATGGGAT GGGGGCGCCG AGCGGGGCAT TTGAAGCCCA AGATGTAGAA GCAATCAGGA
8881 AGGCCGTGGG ATGATTCATA AGGAAAGATT GCCCTCTCTG CGGGCTAGAG TGTTGCTGGG
8941 GCCGTGGGGG TGCTGGGCAG CCGCGGAGGG GGTGCGGAGC GTGGGCGGGT GGAGGATGAG
9001 AAACTTTGGC GCGGACTCGG CGGGGCGGGG TCCTTGCGCC CCCTGCTGAC CGATGCTGAG
9061 CACTGCGTCT CCCGGTCCAA CGCTTACTGG GGCAGGAGCC GGAGCGGGAA GACCCGGGTT
9121 ATTGCTGGGT GCGGACCCCC ACCTCTAGAT CTGGAAAGTA AAGCCAGGGA TGGGGCAGCC
9181 CAAGCCTCTT AAAGAGGTAG TCGGGCCGGT GAGGTCGGCC CCGCCCCGGC CCCATTGCTT
9241 AGCGTTGCCC GACACCTAGT GGCCGTCTGG GGAGCCGCTA GCGCGGTGGG AGTGGTTAGC
9301 TAACTTCTGG ACTATTTGCG GACTTTTTGG TTCTTTGGCT AAAAGTGACC TGGAGGCATT
9361 GGCTGGCTTT GGGGGACTGG GGATGGCCCC GAGAGCGGGC TTTTAAGATG TCTAGGTGCT
9421 GGAGGTTAGG GTGTCTCCTA ATTTTGAGGT ACATTTCAAG TCTTGGGGGG GCCTCCCTTC
9481 CAATCAGCCG CTCCCATTCT CCTAGCCCCG CCCCCGCCAC CCCACCTGCC CAGGGAATGG
9541 GGGCGGGATG AGGGCTGGAC CTCCCTTCTC TCCTCCCTCG CCCTCCTCCT GTCTCTACCA
9601 CGCAAGCCAC TCCCCACGAG CCTGCCCTCC CGATGGGGCC CCTCCTATTC TCCCCCCGCC
9661 CTCCCCCTCT CACCCTGTGG TTTTTATTTC ACTTGGCTTC AGCGCCAATG GGCTGAGGTT
9721 GGAGTTGGAA GCCACCGCGG ACTAAAGCTT TGTTTAAATT CCTGAGAACT GGAAAGAGTT
9781 ACAGCCTCCC TGGCCAGGCG CCTCGGCGCT GTCACCCGCG CTGATGAGGA GCGAaTTCGA
9841 TATCAAGCTT ATCGATCCCT GGTGCTCCGG GGTAAGGTCT CTGACCCGCA CACATCCTGT
9901 GTCTCTTCCC ATTTGCTCAG TCCCTGAAGT CAGCTGCATA CACAATGGCC TAAGGGTCCC
9961 CAATGGTGAG ACGTGGAAAC CCGAGGTATG CTTGATCTGT ATCTGCCACA ATGGCACGGC
10021 TGTGTGCGAT GACGTGCAAT GCAATGAAGA ACTGGACTGT CCCAACCCCC AAAGACGGGA
10081 GGGCGAGTGC TGTGCTTTCT GCCCGGAAGA ATACGGTACG GTTCTGGTCC TAGGGTTAAG
10141 GGAGGTTGGA GTCCCTCCCT CGCCTATGTC TTGAAAGCCT ATAACACCCC CAGGTACCTT
10201 AGTAAATTTA CTTCTAACTG AACGTTTTTC TCCCCAGTAT CACCAAACTC AGAAGATGTA
10261 GGAGTCGAGG TAATAATCTG CTCCCTACTT TCCTCCCACC CGTCCACGAC CCCCACTGTC
10321 CTCCAATAAA CTGCAGTTTT CTTTGCTTCT CTTAACGGCC TTTTGTTTCT TCATCCTCAA

10381 CCCCACAGGG ACCCAAGGGA GACCCTGGCC CCCAAGGCCC AAGGGTAAGC ATCGTATTCT
10441 AAATGAGCTG GGAAACTGGA GGGGGCGTGG CTCTGATTTG TGCTCATAGC AGCCCCCTCC
10501 GTCTCTTGGT GCAGGGACCC GTTGGCCCCC CTGGACGAGA TGGCATCCCT GGACAGCCTG
10561 GACTTCCTGG TCCTCCTGGT CCCCCTGGGC CCCCCGGACC CCCTGGCCTT GGAGGAGTAA
10621 GTAGATCTTG GAGACCCCAG AGGGGTGTGA CTTCCCGTTT TGTTTGTTTG TTTGTCTGTC
10681 TGTTTGTTTG TTTTTCTTTG CTTTGGTTTT GTCAGACATA GTTTTCTATT TTAAAATAAT
10741 TATAAACTTA CAGAAAATAG AATAGAATAG AACTTCTATA CACTCTTCAA CACAACCGCT
10801 TTGCCACTTT CTGGGCACCC AGTTCTCTAT GTCGCTGTAT GTCCTGGACC ATTTGAGACA
10861 TAAGTTGCCA AGAACGTGCC CGTGACTTAC GATCTTAAgG ATCAATTCCG ATCATATTCA
10921 ATAACCCTTA ATATAACTTC GTATAATGTA TGCTATACGA AGTTATTAGG TCTGAAGAGG
10981 AGTTTACGTC CAGCCAAGCT TGATATCGAG GGCCCCTGCA GGTCAATTCT ACCGGGTAGG
11041 GGAGGCGCTT TTCCCAAGGC AGTCTGGAGC ATGCGCTTTA GCAGCCCCGC TGGGCACTTG
11101 GCGCTACACA AGTGGCCTCT GGCCTCGCAC ACATTCCACA TCCACCGGTA GGCGCCAACC
11161 GGCTCCGTTC TTTGGTGGCC CCTTCGCGCC ACCTTCTACT CCTCCCCTAG TCAGGAAGTT
11221 CCCCCCCGCC CCGCAGCTCG CGTCGTGCAG GACGTGACAA ATGGAAGTAG CACGTCTCAC
11281 TAGTCTCGTG CAGATGGACA GCACCGCTGA GCAATGGAAG CGGGTAGGCC TTTGGGGCAG
11341 CGGCCAATAG CAGCTTTGCT CCTTCGCTTT CTGGGCTCAG AGGCTGGGAA GGGGTGGGTC
11401 CGGGGGCGGG CTCAGGGGCG GGCTCAGGGG CGGGGCGGGC GCCCGAAGGT CCTCCGGAGG
11461 CCCGGCATTC TGCACGCTTC AAAAGCGCAC GTCTGCCGCG CTGTTCTCCT CTTCCTCATC
11521 TCCGGGCCTT TCGACCTGCA GCCAATATGG GATCGGCCAT TGAACAAGAT GGATTGCACG
11581 CAGGTTCTCC GGCCGCTTGG GTGGAGAGGC TATTCGGCTA TGACTGGGCA CAACAGACAA
11641 TCGGCTGCTC TGATGCCGCC GTGTTCCGGC TGTCAGCGCA GGGGCGCCCG GTTCTTTTTG
11701 TCAAGACCGA CCTGTCCGGT GCCCTGAATG AACTGCAGGA CGAGGCAGCG CGGCTATCGT
11761 GGCTGGCCAC GACGGGCGTT CCTTGCGCAG CTGTGCTCGA CGTTGTCACT GAAGCgGGAA
11821 GGGACTGGCT GCTATTGGGC GAAGTGCCGG GGCAGGATCT CCTGTCATCT CACCTTGCTC
11881 CTGCCGAGAA AGTATCCATC ATGGCTGATG CAATGCGGCG GCTGCATACG CTTGATCCGG
11941 CTACCTGCCC ATTCGACCAC CAAGCGAAAC ATCGCATCGA GCGAGCACGT ACTCGGATGG
12001 AAGCCGGTCT TGTCGATCAG GATGATCTGG ACGAAGAGCA TCAGGGGCTC GCGCCAGCCG
12061 AACTGTTCGC CAGGCTCAAG GCGCGCATGC CCGACGGCGA TGATCTCGTC GTGACCCATG
12121 GCGATGCCTG CTTGCCGAAT ATCATGGTGG AAAATGGCCG CTTTTCTGGA TTCATCGACT
12181 GTGGCCGGCT GGGTGTGGCG GACCGCTATC AGGACATAGC GTTGGCTACC CGTGATATTG
12241 CTGAAGAGCT TGGCGGCGAA TGGGCTGACC GCTTCCTCGT GCTTTACGGT ATCGCCGCTC
12301 CCGATTCGCA GCGCATCGCC TTCTATCGCC TTCTTGACGA GTTCTTCTGA GGGGATCGAT
12361 CCGCTGTAAG TCTGCAGAAA TTGATGATCT ATTAAACAAT AAAGATGTCC ACTAAAATGG
12421 AAGTTTTTCC TGTCATACTT TGTTAAGAAG GGTGAGAACA GAGTACCTAC ATTTTGAATG
12481 GAAGGATTGG AGCTACGGGG GTGGGGGTGG GGTGGGATTA GATAAATGCC TGCTCTTTAC
12541 TGAAGGCTCT TTACTATTGC TTTATGATAA TGTTTCATAG TTGGATATCA TAATTTAAAC
12601 AAGCAAAACC AAATTAAGGG CCAGCTCATT CCTCCCACTC ATGATCTATA GATCTATAGA
12661 TCTCTCGTGG GATCATTGTT TTTCTCTTGA TTCCCACTTT GTGGTTCTAA GTACTGTGGT
12721 TTCCAAATGT GTCAGTTTCA TAGCCTGAAG AACGAGATCA GCAGCCTCTG TTCCACATAC
12781 ACTTCATTCT CAGTATTGTT TTGCCAAGTT CTAATTCCAT CAGAAGCTGA CTCTAGAGCT
12841 TGCGGAACCC TTAATATAAC TTCGTATAAT GTATGCTATA CGAAGTTATT AGGTCCCTCG
12901 AAGCTTATCG AcaattGCAT TCCATTGTCT CATTCCTTTG GTTCCCACAA ATGTAGACTG
12961 TTTCTGAGAC TCTATGTCAT CCTTGACTTG GGAAGGTTTT GAAGAGTCTG GCTGTTGGGC

13021 AGAGGGTGTT ATTTTATAGG TTCTTCTGTA GTTTGGGATT GTCTGGTTAG ATTAAATTCA
13081 GTCGAGGTGT CTGGTACAGG AATATCTTAC AGGCTATATT GCTGTCTCCC TGAGTTAATT
13141 ACTTCCGGAA GCCTACATCC GTTTGTCCCA TTACTGCTAA TGCCTCCTGG CTTTCAATTT
13201 ACCATCCAGA GTCCAATTTC TCACACATTC TCTCTTGTTT TTCTAGAACT TTGCTTCCCA
13261 GATGTCCTAT GGCTATGATG AAAAATCAGC TGGAGTTTCC GTGCCTGGCC CCATGGTGAG
13321 CCACCAGGGG GAGCAGTGAC CACACAGAGG ATAGGCATCC CGAAGAGAAA GGCCAATGAT
13381 GTCCTGCAAG GGGGTGGTAC AGATAGCTTG GGAGGTAAAG GCCACCTTAG GGACTTGAGA
13441 TCTACAGTAG ATGCAAGTCA GACAGTGATC CCCACAGGAA GAGGTAGAAA CAGATTTCTG
13501 CCATACTGCC TTTATTCACA GATATCAAAA GCCCCTCTTT TTCCTTCTGT AGGGTCCTTC
13561 TGGTCCTCGT GGTCTCCCTG GCCCCCCTGG TGCACCTGTA CGTATCTAAG GGTCTTCATG
13621 GATCATTTTG GGCATGTACT GTTGCGGGGG TGGGGGAGAT GGGGATGAGG GCAGAAGGCT
13681 GGTAGACATC TCTTGTAAGA GGGAGACAGC CCGGGAGGAA CCTGCCTTTC TGAGCCTATC
13741 CTTCTCTCGT GTTCTTCAGG GTCCACAAGG TTTCCAAGGC CCCCCTGGTG AACCTGGCGA
13801 GCCTGGCGGT TCAGTAAGTT AGTCCTCCAT GCCGACACAC CTCCATCTCC ACAAATTCTC
13861 CTATGGAAAG TCACCTGTAC TGAGAGAGGG GTAGCTTGTC ACCTCCTCGC TCCAGTCTTC
13921 CTCCTTATTG TCTCTCATCT AAGAGTTCTA AATGCTTCTC CTGTCCCCTC TAGGGTCCAA
13981 TGGGTCCCCG AGGTCCCCCT GGCCCTCCTG GCAAGAATGG AGATGATGTA AGTACCCCTG
14041 GGGAAAGACA CTATCAAATC CTGTGGCCTT CTCTATGTGG GTCCTGAAAT CTTCCCAGCA
14101 GGGTCTCAGC ACCAATAACC AGGAGAAGTC AAGATGACCC TTGGCAGTCA TCTTTCCACT
14161 GAATAGAATG TTTTAGTGAG GCAGCCTCCA TCCCCACTTG CATCCTCACC CCCATCCATC
14221 TCCATCCCTT GGAGTAATCC AGGTTTAAGA GTTCTGGGTG AGCAGATGTG GAAATGAATA
14281 CACCACTGTA GATATAAACT GAGTTTTAGG TTAGAAGTTA GATTTGGTCC CTTAAACTGA
14341 ATCCACCCTT CAATGGGGGA TAGAGGAGCA TTGCCATAAT GGGCTTTTCT CTCTCTGTAG
14401 GGGGAAGCTG GCAAGCCCGG CCGTCCTGGT GAGCGTGGAC CTCCTGGACC TCAGGTGAGC
14461 AGGATGTGAG GGTGGGCCCA GGGCTTACTG CACTGGGGAT ACATTTAGGA GCAATAGTCA
14521 CAGAGCCCTT CCTTCTGACA CCCTTCCAAG TCCTCATTTG CCTTTCTCTC CAGGGTGCTC
14581 GTGGATTGCC TGGAACAGCT GGCCTCCCTG GAATGAAGGG ACACCGAGTA AGTCTCTTTG
14641 GAGTCACTGA AGCTCCCAAA GCCCCAAGTC ACCATTCTCT CCCCTTCCCT AAGAGATCCT
14701 GTTGTGAATC ATGTTAGATG GTAGCTAGAG ATATCTCAAA GGTCATAGGT TAGACATCTG
14761 AAAGGACTTC CTTTTGTGGG GGAAGAAGTC CCAATAAAGA GACTGGCTGG ATGATAGGTG
14821 AATTGGGATC TACTTTAGCT CAGAGACAGA GAGATGGCTG GAGATCTTCC AAAAATTATT
14881 GTCAAATCCA TGTGGGTGTC AAAGTGGTAT ACTCCCCCAA CAGGGGTCTT ACTTCATCTC
14941 TTTTCCCCCC CATTAAGGGC TTCAGTGGTT TGGATGGTGC CAAAGGAGAT GCTGGTCCTG
15001 CTGGTCCTAA GGTAAGATAC TGATGAGAGG TCCTGAATGA CCTGTGCCCA ACAGTCCTAC
15061 TATGAATGCT TCACTCTGTT CTCTGTTCCT AGGGAGAGCC CGGCAGTCCT GGTGAAAACG
15121 GAGCTCCTGG CCAGATGGTA AGTAGGCCAG CTTGAGTGAC AGGGGGCGCT CCAAGAGGGA
15181 GGCTGCAGAA GAGGGTGTCT CCATCTCATT CGACCTCCCT TGGTTTCGTT TTTAGGGTCC
15241 CCGAGGTCTG CCCGGTGAGA GAGGTCGCCC TGGACCTCCT GGCACTGCTG TAAGTACTCC
15301 TGCCCCCCTT GGATGCCCCG GGCTCTGGTT GCAGTTCCCT GGGGACTAAG GGTACTGGCC
15361 ACTGCTCAGG GGACATAATT GGGCTTCCCC TCTCTCTTAC AGGGTGCTCG CGGTAACGAT
15421 GGTGCTGTTG GTGCTGCTGG ACCCCCTGTA AGTATGGCCT TGACTTTGAG TATGGGAGTA
15481 GGGTGAGACC TCAGGGTCAG CTCTTGGGGC TGACATCAAC TTTTGAAACT TAGAAGCTCT
15541 TAGTCTCCTC CACCTTGTAG CTAAACTTCC ATTTCATCCA CAGGGTCCCA CCGGCCCCAC
15601 TGGCCCTCCT GGCTTCCCTG GTGCAGTTGG TGCTAAGGTG AGGCACGTTC TTTCTTTGAA

15661 CATGATTGTC ATAGGCAATG GGCTAGGCAT CACCCAGATC AGAGAAGCCC AGGGTGGCAA
15721 GAATGAAGCA CACTCTACAA TCCTGGACTC AATCCTGGTG TCCCCTTCCT AGCTCTCAAA
15781 GTCCTCGTCC TTGCTCCCCC GCCCTCCAAC ACCTGACATA CAGAAGTAAT GCTGGCTGCC
15841 CCAATCTGAT TTGTCCTTTC CTCCGATCTA GGGTGAAGCT GGTCCCCAAG GAGCTAGAGG
15901 CTCTGAAGGT CCCCAGGGTG TGCGTGGTGA GCCCGGACCC CCTGGCCCTG CTGGTGCTGC
15961 CGGCCCTGCT GTGAGTGTCC TTCCTGCTGC TGACCCCTGT GCTACCCGCC AACTTCACAG
16021 TTCTAGGCTT TCACTCACCC TCCTTTCTCT ATGTGACAGG GAAACCCTGG TGCTGATGGA
16081 CAACCTGGCG CTAAAGGTGC CAATGTAAGT CTCTAGCAGC CCTCCTCCCA TGAGCCACCT
16141 GCCACCTGCC CCATGCCCTC CACCCTTGGG CTAAAATGTC TGTCCTGACT CCTGCAGGGT
16201 GCTCCTGGTA TTGCTGGTGC TCCTGGCTTC CCTGGTGCCC GAGGCCCCTC TGGACCCCAG
16261 GGCCCCAGCG GCCCTCCAGG TCCCAAGGGT AACAGTGTAA GTATAAGACT CTCCCTTCTG
16321 CCTCGCTTGT TCATGAGCTC CAGGGTGCCC CCTTCTGGAG AAACATACAC TTGCGAGCCA
16381 ACAGAGCAAC TTACCCTCTT TCTCTCCCTA CACAGGGTGA ACCTGGTGCT CCTGGCAACA
16441 AAGGAGACAC TGGTGCCAAA GGAGAACCCG TAAGTCACAC ATCACCCCTT CTCCCCCTCT
16501 GCCCCCAGGG TCAGGAAGCC TCAACGGGGA GCTCCAGGGA GGAAGGGAGC CCCCTCCTCA
16561 TACCCAGAAG TCCTATTACT AGAGGCCAGC ATCATAGGCT CCTGGGCTTG TACTCCCTCA
16621 CCAAGGCTAC TCCCTCTCTG CCCACTCCAG GGTGCTACTG GAGTTCAAGG TCCCCCAGGC
16681 CCTGCCGGAG AAGAAGGAAA ACGAGGAGCC CGTGGTGAGC CTGGACCTTC CGGACTGCCT
16741 GGACCTCCTG GCGAGCGTGT AAGTGTCCCC TGTCCCCTGT GTCTCTCCTC ATTTGCTCTT
16801 AGTGCTTGCC CAGTGCTGTG ATGGTAATAG GCTCACTGGC CTTTTTTCCT TCTCCCTGCA
16861 GGGTGGACCT GGTAGCCGTG GTTTCCCTGG TGCTGATGGT GTTGCTGGCC CCAAGGTAAT
16921 CTCTCTTCCT GGCAGAGAAG CTGACTCCCC CAGCTGCCCG ACGTCACTTC CCTGACAGTA
16981 CGGTTGACCT CGAGGGGGGG CCCGGTACCC AATTCGCCCT ATAGTGAGTC GTATTACGCG
17041 CGCTCACTGG CCGTCGTTTT ACAACGTCGT GACTGGGAAA ACCCTGGCGT TACCCAACTT
17101 AATCGCCTTG CAGCACATCC CCCTTTCGCC AGCTGGCGTA ATAGCGAAGA GGCCCGCACC
17161 GATCGCCCTT CCCAACAGTT GCGCAGCCTG AATGGCGAAT GGGACGCGCC CTGTAGCGGC
17221 GCATTAAGCG CGGCGGGTGT GGTGGTTACG CGCAGCGTGA CCGCTACACT TGCCAGCGCC
17281 CTAGCGCCCG CTCCTTTCGC TTTCTTCCCT TCCTTTCTCG CCACGTTCGC CGGCTTTCCC
17341 CGTCAAGCTC TAAATCGGGG GCTCCCTTTA GGGTTCCGAT TTAGTGCTTT ACGGCACCTC
17401 GACCCCAAAA AACTTGATTA GGGTGATGGT TCACGTAGTG GGCCATCGCC CTGATAGACG
17461 GTTTTTCGCC CTTTGACGTT GGAGTCCACG TTCTTTAATA GTGGACTCTT GTTCCAAACT
17521 GGAACAACAC TCAACCCTAT CTCGGTCTAT TCTTTTGATT TATAAGGGAT TTTGCCGATT
17581 TCGGCCTATT GGTTAAAAAA TGAGCTGATT TAACAAAAAT TTAACGCGAA TTTTAACAAA
17641 ATATTAACGC TTACAATTTA G