

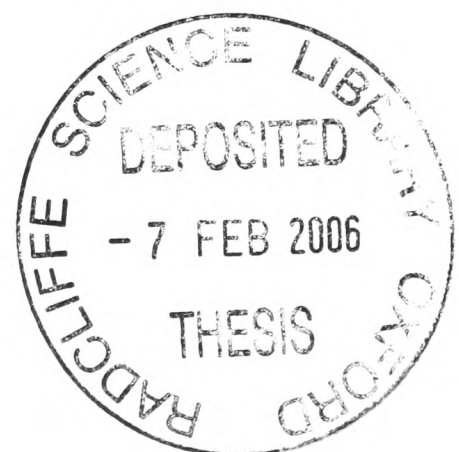
Gene Hunting in Primary Osteoarthritis

Tracy Forster

**Exeter College,
University of Oxford**

A thesis submitted in partial fulfilment of the requirements for the
degree of Doctor of Philosophy.

Nuffield Department of Clinical Laboratory Sciences
Michaelmas Term, 2004



Gene Hunting In Primary Osteoarthritis
Tracy Forster, Exeter College, University of Oxford
D.Phil. Michaelmas 2004

Osteoarthritis (OA) is a common, disabling disease of the elderly. It is uncommon in individuals under 45 years of age, and prevalence is higher in females than males. OA is characterised by focal cartilage loss accompanied by osteophytes, sclerosis and often some inflammatory response,

A significant genetic component to OA has been demonstrated by twin-pair, sibling risk and segregation studies

Our group previously carried out a two-stage genome-wide linkage scan and identified six chromosomal regions as potentially harbouring OA susceptibility genes.

The aim of this project was to examine three of these linkage regions, one on chromosome 4 and two on chromosome 16, in more detail in order to further localise and determine the identity of the OA susceptibility gene(s) within these regions.

Initially the three regions were linkage mapped using a staged approach. Linkage mapping of the region on chromosome 4q identified a region showing evidence of linkage to female hip OA with a multipoint LOD (MLS) score of 3.1. For chromosome 16 linkage mapping identified two regions with MLS of 1.7 and 1.9 respectively. The first region was restricted to families with female siblings concordant for hip OA while the second is accounted for by families with affected female siblings.

The ADAMTS3 gene was identified as a strong candidate gene within the linkage region on chromosome 4 as a result of its role in procollagen processing. SNPs within the gene were identified and typed in cohorts of cases and controls. There was no evidence for an association with any of these SNPs and OA disease.

Within the first linkage region on chromosome 16 the IL4R gene was identified as a strong candidate gene as a result of its proposed role in the normal response of chondrocytes to mechanotransduction. Again SNPs within the gene were identified and typed. Four SNPs show evidence of association with OA and for three of these there is evidence for a relationship between genetic variation and evidence for linkage. It appears that there is a dosage effect with at least two copies of different variant alleles being necessary to increase OA disease susceptibility.

Contents

Table of contents		i
List of abbreviations		viii
Chapter 1 – Introduction		1
1.1	Definition of Osteoarthritis	1
1.2	Normal articular joint structure	3
1.2.1	Cartilage	3
1.2.2	Subchondral bone	5
1.3	Normal age associated changes in articular joint structure	6
1.4	The OA disease process	7
1.4.1	Cartilage	8
1.4.2	Bone	12
1.5	The Genetic Component of OA	13
1.5.1	Epidemiological studies	13
1.5.1.1	Twin studies	14
1.5.1.2	Sibling risk studies	15
1.5.1.3	Segregation studies	16
1.6	Non-genetic risk factors for OA	19
1.6.1	Age	19
1.6.2	Sex	20
1.6.3	Body mass index	20
1.6.4	Bone density	21
1.6.5	Local biomechanical factors/ joint injury	23
1.6.6	Additional proposed factors	23
1.7	Summary of disease associated factors	24
1.8	Complex trait analysis	24
1.8.1	Linkage analysis	27
1.8.2	Association studies	30
1.8.2.1	Linkage disequilibrium	31
1.8.2.2	Patterns of linkage disequilibrium	31
1.8.2.3	Haplotype blocks and recombination hotspots	32
1.8.3	Functionality testing	36
1.9	Conditioning analysis	37
1.9.1	The example of type II diabetes	37

1.10	Gene finding in primary OA	38
1.10.1	Linkage analysis of candidate genes	38
1.10.2	Genome-wide scans	39
1.10.3	Association studies of candidate genes	41
1.11	Background and Aims	42
1.11.1	Background	42
1.11.2	Aims	44
Chapter 2 - Materials and Methods		47
2.1	Families	47
2.2	Amplification of DNA by the polymerase chain reaction (PCR)	48
2.3	Linkage analysis	49
2.3.1	Microsatellite Markers	49
2.3.2	Separation of PCR products	50
2.3.3	Microsatellite genotyping and statistical analysis	50
2.3.3.1	One-LOD support interval	51
2.3.4	Stratification	51
2.4	Conditioning analysis	52
2.4.1	Determination of evidence for a statistical interaction between regions	52
2.4.2	Conditioning analysis	53
2.4.2.1	Modelling for a positive interaction	53
2.4.2.2	Modelling for a negative interaction	54
2.5	Candidate gene identification	54
2.6	DNA sequencing	55
2.6.1	PCR amplification	55
2.6.2	Precipitation	56
2.6.3	Separation of products	56
2.6.4	Sequence analysis	57
2.6.5	Association studies	57
2.6.6	Variant typing	57
2.6.7	Analysis of results	58
2.6.8	Linkage disequilibrium studies	59
2.7.3.1	Measures of LD	59
2.7.4	Investigation of the relationship between genetic variation and the evidence for linkage	60
2.8	Expression studies	61
2.8.1	RNA extraction from cartilage	61

2.8.2	cDNA synthesis	61
2.8.3	Amplification of cDNA for IL4R	62
Chapter 3 - Linkage mapping of chromosomes 4 & 16		63
3.1	Introduction	63
3.2	Initial (MD10) linkage mapping of chromosome 4	65
3.3	Finer (HD5) linkage mapping of chromosome 4	66
3.3.1	Female and female-THR families	66
3.3.2	TKR families	67
3.4	Additional finer linkage mapping of chromosome 4	68
3.5	Initial (MD10) linkage mapping of chromosome 16	69
3.6	Finer (HD5) linkage mapping of chromosome 16	70
3.7	Additional finer linkage mapping of chromosome 16	71
3.7.1	Region 1 (D16S500 - D16S415)	72
3.7.2	Region 2 (D16S3140 - D16S3091)	73
3.8	Discussion	74
Chapter 4 – Conditioning analysis		78
4.1	Introduction	78
4.2	Conditioning analysis	82
4.2.1	Correlation calculations	82
4.2.2	Conditioning analysis	84
4.3	Discussion	86
Chapter 5 - ADAMTS3 candidate gene investigation		89
5.1	Introduction: Candidate gene identification on chromosome 4q	89
5.1.1	Cataloguing of genes within the linkage region	90
5.1.2	Identification of candidate genes	90
5.2	Introduction: The ADAMTS3 (a disintegrin and	91

	metalloproteinase with thrombospondin motifs 3) gene	
5.2.1	The potential role of ADAMTS3 in normal cartilage function and the OA disease process	91
5.2.2	ADAMTS3 gene structure	95
5.3	Sequencing of the ADAMTS3 gene	96
5.4	Association studies - Stage 1	97
5.4.1	SNP selection	97
5.4.2	SNP genotyping	98
5.4.3	Association analysis	98
5.5	Haplotype analysis	100
5.6	Linkage disequilibrium studies	100
5.7	Association studies - Stage 2	101
5.7.1	Haplotype analysis	103
5.8	Relationship between genetic variation and the evidence for linkage	104
5.9	Discussion	105
Chapter 6 - IL4R candidate gene investigation		111
6.1	Introduction: Candidate gene identification on chromosome 16	111
6.1.1	Cataloguing of genes within the linkage region	111
6.1.2.	Identification of candidate genes	111
6.2	Introduction: The Interleukin - 4 receptor alpha chain gene	112
6.2.1	The role of IL4R in OA	112
6.2.2	IL4R gene structure	115
6.2.3	IL4R chain protein structure and function	117
6.2.4	IL4R SNPs and haplotypes	119
6.3	Sequencing of the IL4R gene	121
6.4.1	Association studies - Stage 1	122
6.4.2	SNP selection	122
6.4.3	SNP genotyping	123
6.4.4	Association analysis	123
6.4.5	SNP genotyping and association studies - Stage 2	125
6.5	Linkage disequilibrium studies	127
6.6	Haplotype analysis	128

6.7	Relationship between genetic variation and the evidence for linkage	129
6.7.1	N142N	130
6.7.2	S411L	131
6.7.3	S727A	132
6.8	Examination of dosage effects	132
6.8.1	Association analysis of SNP genotypes	133
6.8.2	Association analysis for SNP genotype combinations	133
6.9	Splicing analysis	135
6.10	Discussion	136
 Chapter 7 - Summary and conclusions		 147
7.1	Introduction	147
7.2	Chromosome 4	148
7.2.1	Linkage mapping	148
7.2.2	Conditioning and sibling TDT	148
7.2.3	ADAMTS3 candidate gene identification	148
7.2.4	Future studies	150
7.2.4.1	Further ADAMTS3 studies	150
7.2.4.2	Investigation of other candidate genes	150
7.2.4.3	Systematic association studies	150
7.3	Chromosome 16	151
7.3.1	Linkage mapping	151
7.3.2	Conditioning and sibling TDT	151
7.3.3	IL4R candidate gene identification	152
7.3.4	Future studies	153
7.3.4.1	Further IL4R studies	153
7.3.4.2	Investigation of other candidate genes	154
7.3.4.3	Systematic association studies	156
7.4	Concluding remarks	157
 Presentations, publications and prizes		 159
 Bibliography		 161
 Appendix 1 Primer nomenclature and PCR conditions for microsatellite markers		

Appendix 2 Family NPL scores for the regions showing maximum evidence for linkage on chromosomes 4q, 16p and 16q in the 146 female-THR families

Appendix 3 Family NPL scores for the regions showing maximum evidence for linkage on chromosomes 16p and 16q in the 218 female families

Appendix 4 Primer nomenclature and PCR conditions for ADAMTS3 and IL4R DNA sequencing

Appendix 5 Primer nomenclature and PCR conditions for ADAMTS3 and IL4R restriction digests

Appendix 6 Enzymes and conditions used for ADAMTS3 and IL4R restriction digests

Appendix 7 Summary of ADAMTS3 sequencing. Sequence covered, positions of primers and SNP locations are shown in relation to the genomic sequence.

Appendix 8 Summary of IL4R sequencing. Sequence covered, positions of primers and SNP locations are shown in relation to the genomic sequence.

List of Abbreviations

ABI	Applied Biosystems
ADAM	a disintegrin and metalloproteinase
ADAMTS	a disintegrin and metalloproteinase with thrombospondin motifs
APS	ammonium persulphate
ASP	affected sibling pairs
BHD	Beukes hip dysplasia
BMD	bone mineral density
BMI	body mass index
BMP	bone morphogenetic protein
bp	basepair
cAMP	cyclic adenosine monophosphate
CAPN10	Calpain 10
CARD15	CED4/Apaf-1 family member with a N-terminal caspase recruitment domain
CD1, CD5 etc	CD stands for cluster of differentiation, which refers to groups of monoclonal antibodies on the surface of leukocytes that show similar reactivity with certain subpopulations of antigens.
cDNA	complementary deoxyribonucleic acid
CFTR	cystic fibrosis transmembrane conductance regulator
CHR	cytokine binding homology domain
CI	confidence interval
CLECSF1	cartilage-derived, c type lectin

cM	centimorgan
COL2A1	collagen, type II, alpha 1
COX-2	cyclooxygenase 2
CRR	cysteine rich region
CRTL	cartilage link protein 1
MATN1	cartilage matrix protein (Matrilin 1)
CS	chondroitin sulfate
CTLA4	cytotoxic T lymphocyte antigen 4 gene
dbSNP	database single nucleotide polymorphism
DEPC	diethylpyrocarbonate
dH ₂ O	distilled water
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleotides
ECM	extracellular matrix
EDS	Ehlers-Danlos syndrome
EM	expectation maximisation
ESE	exonic splice enhancer
EtBr	ethidium bromide
EXO1	exonuclease I
g,mg,µg,ng	grams, milligrams, micrograms, nanograms
GAG	glycosaminoglycan
GAS	Genetic Analysis System
GOLD	graphical overview of linkage disequilibrium
HD5	high density 5

Hz	hertz
IBD	identity by descent
IgE	immunoglobulin E
IGFBP-3	insulin-like growth factor binding protein 3
IL	interleukin
IL1R	interleukin 1 receptor
IL1RN	interleukin 1 receptor antagonist
IL4-BP	interleukin 4 binding protein
IL4R	interleukin 4 receptor, alpha subunit
iNOS	inducible nitric oxide synthase
IRS	insulin receptor substrate
ITIM	immuno-receptor tyrosine-based inhibitory motif
JAK	janus kinase
Kb	kilobasepair
kDa	kilodalton
l,ml, µl	litre, millilitre, microlitre
LD	linkage disequilibrium
LIF	leukaemia inhibitory factor
LMS	linkage mapping set
LOD	logarithm of the odds
M, mM, µM	Molar, millimolar, micromolar
mA	milliampere
Mb	megabases
MD10	medium density 10 ABI set
MgCl ₂	magnesium chloride

MHC	major histocompatibility region
MIM	Mendelian inheritance in man
MLS	multipoint logarithm of the odds score
MMP	matrix metalloproteinase
mRNA	messenger ribonucleic acid
MT-MMP	membrane-type metalloproteinase
NCBI	National Center for Biotechnology Information
NIDDM	non-insulin dependant diabetes mellitus
NMD	nonsense-mediated decay
NPL	non parametric linkage
OA	osteoarthritis
OMIM	online Mendelian inheritance in man
p-value	probability value
PCR	polymerase chain reaction
PGE ₂	prostaglandin E ₂
PTC	premature termination codon
RNA	ribonucleic acid
rpm	revolutions per minute
RT-PCR	reverse transcriptase polymerase chain reaction
SAP	shrimp alkaline phosphatase
SIL4R	soluble interleukin 4 receptor
SH2	src homology 2
SHP	src homology 2-containing phosphatases
SK	small potassium channel
SMN1	survival motor neurone gene

SNP	single nucleotide polymorphism
SOCS	suppressers of cytokine signalling
SR	serine/arginine protein
STAT	signal transducer and activator of transcription
TBE	tris borate EDTA buffer
TDT	transmission/disequilibrium test
TEMED	N, N, N', N', tetraethyl-ethylenediamine
TGFB1	transforming growth factor, beta 1
THR	total hip joint replacement
TKR	total knee joint replacement
TIMP	tissue inhibitor of metalloproteinases
TNFA	tumour necrosis factor alpha
TYK	tyrosine kinase
UCSC	University of California, Santa Cruz
UTR	untranslated region
V	volts
VDR	vitamin D receptor
XYLTI	xylosyltransferase I

Chapter 1 - Introduction

1.1 Definition of Osteoarthritis

Primary osteoarthritis (OA) is the most common form of arthritis affecting approximately 8 million people in the UK, and is a major cause of morbidity and disability especially amongst the elderly. It is recognised histologically by focal cartilage loss accompanied by osteophytes, sclerosis and often some inflammatory response (Jones et al. 1995). No standard definition was available until 1986 when definitions were proposed both by the Subcommittee on OA of the American College of Rheumatology Diagnostic and Therapeutic Criteria Committee (Altman et al. 1986) and at the Workshop on the Etiopathogenesis of OA (Mankin et al. 1986). Both definitions recognise OA as a group of overlapping, distinct diseases and not a single disorder.

In 1995, at a workshop in the USA, a further definition was developed recognising OA as involving the whole synovial joint: “Osteoarthritic diseases are a result of both mechanical and biological events that destabilise the normal coupling of degradation and synthesis of articular cartilage chondrocytes and extracellular matrix, and subchondral bone. Although they may be initiated by multiple factors, including genetic, developmental, metabolic, and traumatic, osteoarthritic diseases involve all of the tissues of the diarthrodial joint. Ultimately, osteoarthritic diseases are manifested by morphologic, biochemical, molecular and biomechanical changes of both cells and matrix which lead to a softening, fibrillation, ulceration, loss of articular cartilage, sclerosis and eburnation of subchondral bone, osteophytes and subchondral cysts. When clinically evident, osteoarthritic diseases are characterised by joint pain, tenderness, limitation of

movement, crepitus, occasional effusion, and variable degrees of inflammation without systemic effects” (Kuettner and Goldberg 1995).

In summary, it is proposed that OA is the consequence of events, either mechanical or biological, which result in uncoupling of the normal balance between synthesis and degradation of the components of the extracellular matrix of cartilage and bone (Reginato and Olsen 2002). Triggers are likely to be diverse and include genetic, metabolic and traumatic factors.

Definition of the OA phenotype in clinical practice is often difficult with most epidemiological studies using radiographic measures based on the scales proposed by Kellgren and Lawrence in 1957 and on updated grading schemes. Such radiographic observations however, often correlate poorly with clinical symptoms and thus algorithms based on clinical symptoms may also be used in disease definition (e.g. Altman et al. 1986). Some studies incorporate data from combinations of symptoms, physical findings, laboratory data and radiographic evidence (e.g. Flores and Hochberg 1998).

Although the OA disease process ultimately involves the whole synovial joint much work has concentrated on the role of cartilage, and to a lesser extent bone, as these are the tissues commonly implicated as being important in the initiation and progression of OA.

1.2 Normal articular joint structure

Figure 1.2a illustrates the structure of a normal articular knee joint. Both knee and hip joints belong to the diarthrodial family of joints. In summary, at the diarthrodial joint the surface of the bones are covered with a thin layer of hyaline articular cartilage. The synovial membrane, of which the capsule is the tough outer layer, surrounds the joint and produces the synovial fluid. The capsule in conjunction with the ligaments, which run within or just outside the capsule, act to prevent excessive movement of the bones. Some diarthrodial joints are completely or incompletely divided by an articular disk or meniscus.

1.2.1 Cartilage

Articular cartilage forms a thin layer lining the articulating ends of diarthrodial joints and functions to provide a smooth, frictionless surface, to evenly distribute load applied across the joint and to minimise the contact stress on the underlying bone (Ateshian et al. 1994). Cartilage is a composite solid matrix whose major structural components are collagen fibrils and proteoglycans. Between 65-85% of total tissue weight is water (Mankin and Thrasher 1975). Mature cartilage is avascular and aneural and the only cells are chondrocytes, which account for less than 5% total volume (Poole et al. 2002). These cells are responsible for both the generation and the maintenance of the extracellular matrix (Goldring 2000b). The cartilage matrix is divided into four zones classified as superficial, middle, deep and calcified (Aigner and McKenna 2002). The arrangement of molecules within the cartilage is zone dependent, with interzonal differences in

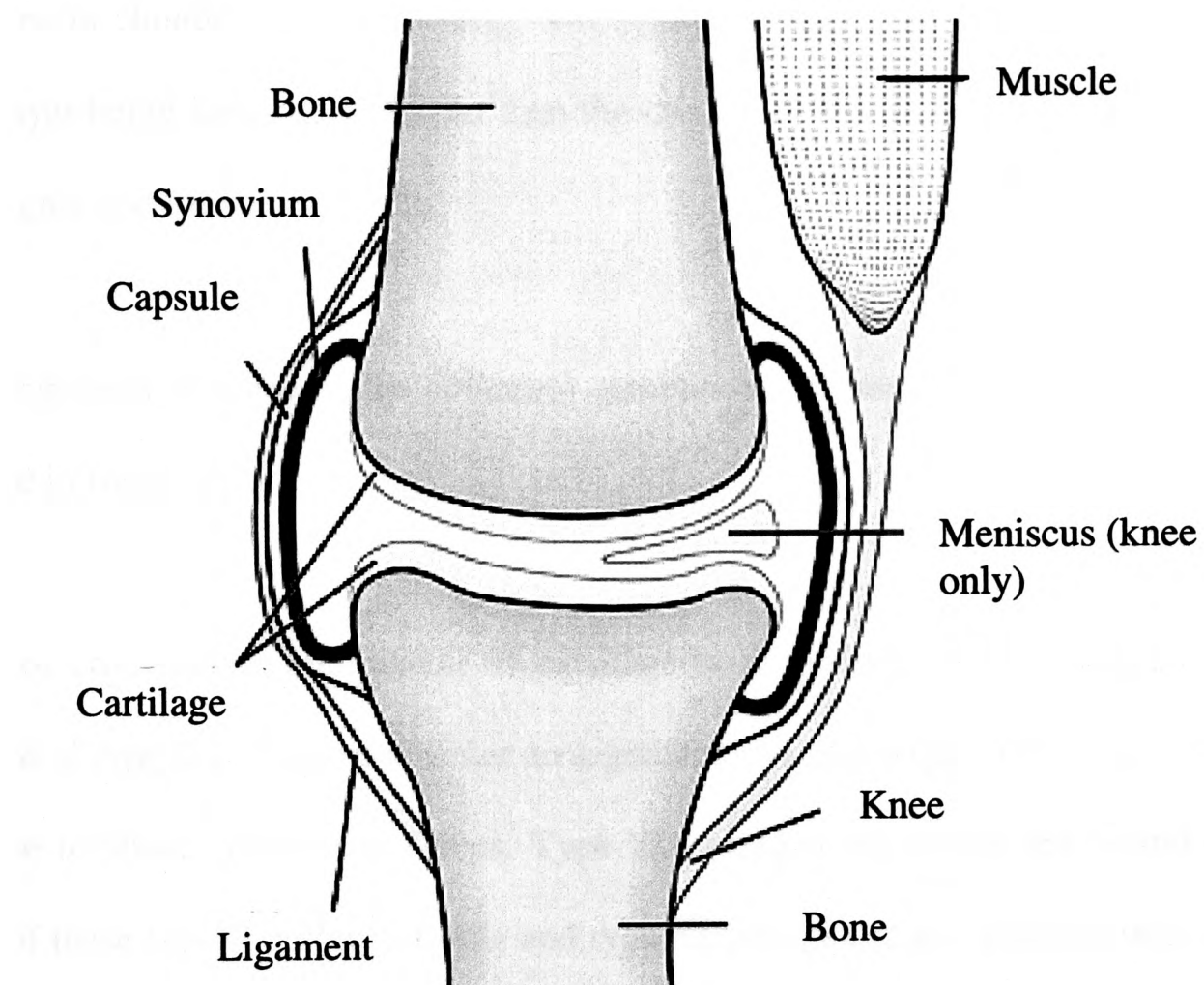


Figure 1.2a: Diagrammatic representation of a normal articular knee joint (from www.arc.org.uk/about_arth/booklets/6025/6025.htm).

proteoglycan and collagen concentration, orientation and width of collagen fibrils and water content (Heinegard and Lorenxo 1998; Eyre 2002). There are also interzonal differences in chondrocyte morphology and behaviour, with, for example, cells in the middle layer being larger and rounder than the small flat chondrocytes of the superficial layer (Aigner and McKenna 2002).

The arrangement of some of the structural components of normal articular cartilage is illustrated in figure 1.2b.

The major collagenous component of cartilage is accounted for by collagen fibrils composed of type II collagen molecules arranged in a classical triple helix. These provide resistance to shear and tensile forces. Type IX collagen molecules are bound to the surface of these type II collagen fibrils and type XI collagen is also present, both on the surface and within the fibril (Poole et al. 2002). Once laid down during development there is apparently little scope for recapitulation of the collagen architecture by the chondrocytes if the network is disrupted by injury or degeneration (Eyre 2002).

The major proteoglycan in articular cartilage is aggrecan, composed of a protein core with a number of distinctive domains. Chondroitin sulphate and other glycosaminoglycans (GAGs) bind along the surface at defined regions and the aggrecan molecules bind to hyaluronan (this interaction is stabilised via link protein) to form highly charged macromolecules which interact with the collagen fibrils (Carney and Muir 1988). The large negative charge provided by these molecules is important in binding

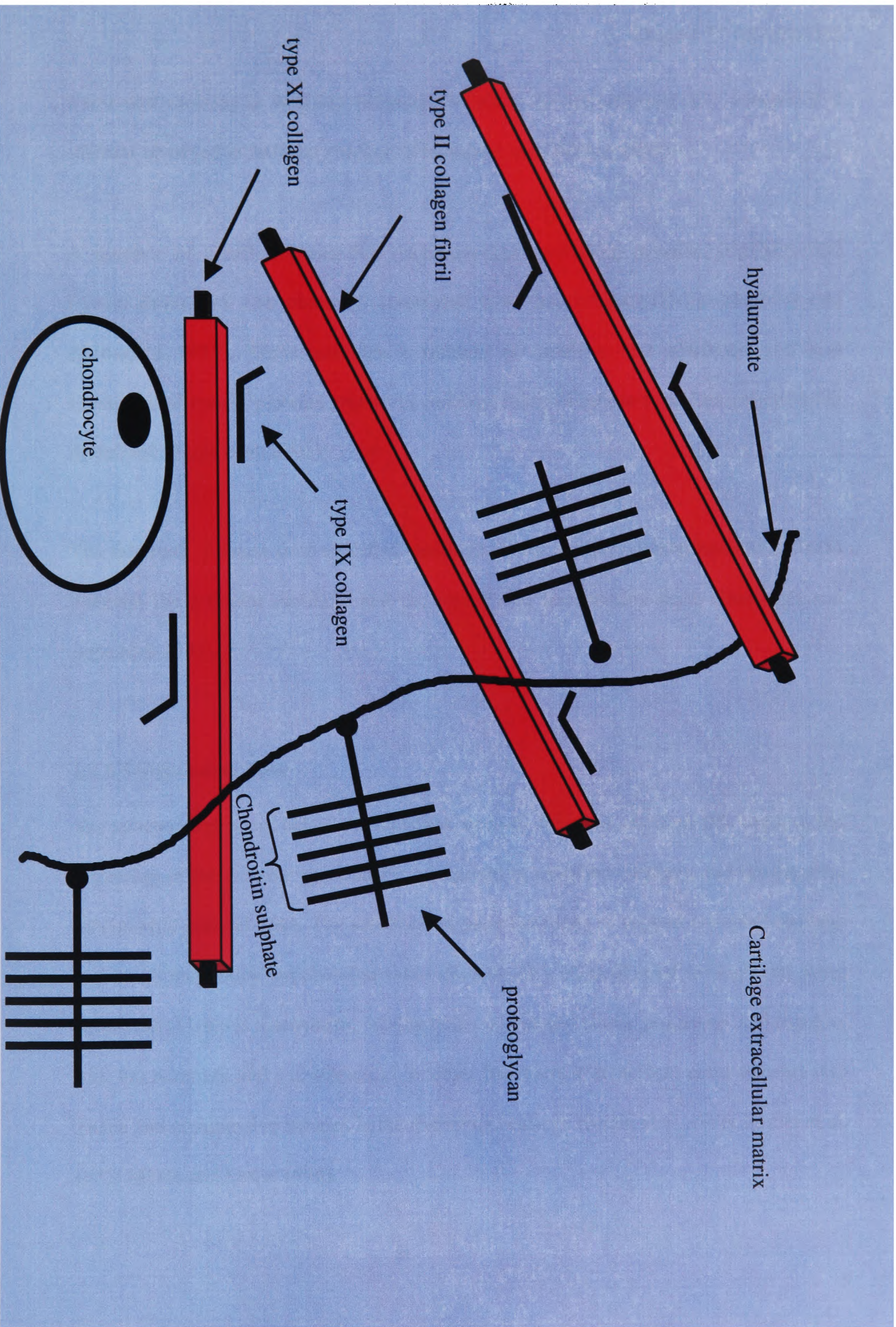


Figure 1.2b: Diagrammatic representation of some of the components of the articular cartilage matrix. Figure courtesy of Dr John Loughlin.

water and creating a swelling pressure, resulting in the characteristic compressive stiffness observed in articular cartilage (Heinegard and Oldberg 1989).

A number of small, leucine-rich proteoglycans, including decorin, biglycan and fibromodulin, are also observed associated with the collagen fibrils (Hedbom and Heinegard 1993). There are also a number of smaller, less abundant and less characterised matrix proteins present in cartilage many of whose functions are currently ill defined (Poole et al. 2002).

The tidemark, formed from calcified cartilage, forms a layer of intermediate stiffness between the articular cartilage and the underlying subchondral bone (Bullough and Jagannath 1983).

1.2.2 Subchondral bone

The subchondral plate, below the cartilage, extends from the calcified tide mark to the beginning of the marrow space and is composed of calcified cartilage and the lamellar subchondral cortical bone. The subchondral plate functions to support articular cartilage and in normal joints subchondral bone absorbs 30 - 50% of load through the joint. Subchondral bone is viscoelastic, having both viscous and elastic properties, and together with the tidemark and osteochondral junction functions to transform sheer stresses into tensile and compressive stresses in the overlying cartilage (Radin et al. 1970; Redler et al. 1975; Mente and Lewis 1994).

Bone consists of a number of cell types, including osteoblasts and osteoclasts, embedded in a matrix composed of both organic and inorganic molecules (Boyle et al. 2003; Harada and Rodan 2003). The organic matrix, which accounts for approximately 35% of the dry weight of bone, is composed of collagen (primarily type I collagen which accounts for 90% of the organic matrix), proteoglycans and non-collagenous matrix proteins including osteocalcin and osteonectin. The inorganic matrix consists of both calcium hydroxyapatite, which gives bone its compressive strength, and osteocalcium phosphate (Revell 1998). In contrast to articular cartilage, subchondral bone is highly vascular, and innervated by unmyelinated nerve fibres (Clark and Huber 1990).

The structure of bone is such as to allow for maximum strength with minimum mass. Its microstructure is continually moulded, shaped and repaired throughout life, a process that involves both resorption and synthesis. In normal adult bone the rate of resorption by osteoclasts is exactly matched by the rate of synthesis by osteoblastic cells in a process tightly controlled by both systemic and local factors (Revell 1998; Sambrook et al. 1999).

1.3 Normal age associated changes in articular joint structure

After the age of about 30 years the tensile properties of articular cartilage show a progressive decline. This decline is more prevalent in some joints, i.e. hip and knee, than others, i.e. ankle (Poole et al. 2002). Collagen fibrils undergo modification and an increase in cross-linking resulting in increased stiffness. There is also an age-related turnover of type II collagen at all cartilage sites (in younger cartilage turnover is limited

to pericellular sites). Collagen damage progresses to deeper sites with age resulting in a zone of damage, distinct from the zone of healthy cartilage below (Poole et al. 2002). Aggrecan molecules also show increasing heterogeneity with age as they are degraded at varying rates (Hardingham and Bayliss 1990). In addition, duplication and migration of the tidemark occurs leaving a thinner layer of non-calcified cartilage (Green et al. 1970).

In bone, both the organic and the inorganic components decrease with age as osteoblasts become less active than osteoclasts. After the age of 40 women lose about 8%, and men 3% of their total bone mass per decade. This osteopenia (loss of bone) results in bones which are more brittle, less elastic and more prone to fracture (Moore and Dalley 1999; Saladin and Fox 2002).

1.4 The OA disease process

Figure 1.4 illustrates a generalised end-stage OA articular knee joint. Despite extensive research much of the detail of the OA disease process remains unresolved and enigmatic. Research has been confounded by the lack of availability of early-stage OA tissue with most research being conducted on tissue from joint replacement, representing the end-stage of the disease. Such difficulties limit both the elucidation of events involved early in the disease process and determination of the order of events resulting in end-stage phenotype.

1.4.1 Cartilage

The cartilaginous damage, resulting in destruction and failure of the extracellular matrix (ECM), observed in OA appears to be the result of either the action of normal loading on abnormal cartilage or to the action of abnormal loading or stress on normal cartilage (Aigner and McKenna 2002). The cellular reaction patterns observed in OA appear at first very heterogeneous and differ both between the different zones of cartilage and depending on stage of the disease process (Sandell and Aigner 2001). This heterogeneity is reflected in a lack of consensus between independent studies aimed at characterising the OA disease process.

The first overt sign of OA is the onset of fibrillation at the articular surface, followed by the development of focal lesions at points of direct mechanical contact between articular surfaces (Neuhold et al. 2001). Collagen fibrils are damaged and proteoglycan lost resulting in increased cartilage water content, swelling and softening of the articular cartilage and loss of tensile strength (Aigner and McKenna 2002; Poole et al. 2002). Damage appears to be due to both mechanical wear and enzymatic degradation. There is also observed a calcification of the articular cartilage matrix and advancement of the tidemark. This results in a thinning of the articular cartilage with a reciprocal thickening of the underlying subchondral bone (Karvonen et al. 1994).

The release of fragments from the degraded cartilage matrix components serves to further induce the expression and secretion, by chondrocytes, of a number of molecules involved

in cartilage degradation including matrix metalloproteinases (MMPs) and cytokines (Homandberg et al. 1992; Liu et al. 2000).

Chondrocyte proliferation is observed early on in the disease process (chondrocytes in normal tissue are postmitotic) but is limited by the physical bulk of collagen fibrils (Hulth et al. 1972). In the early stages the chondrocytes show increased synthesis of type II collagen, aggrecan, matrix degrading enzymes and catabolic cytokines. However, normal cartilage matrix architecture cannot be regenerated by adult articular chondrocytes (Aigner and McKenna 2002). As the disease progresses there is a suppression of expression of aggrecan and type II collagen mRNA (Goldring 2000a).

In later stages of the disease process a significant proportion of chondrocytes take on a distinct phenotype, comparable to that of the chondroprogenitor cells seen in foetal skeletal development (Sandell and Aigner 2001). Many cells show an increased expression of interleukin 1 receptor (IL1R) type 1, insulin-like growth factor binding protein 3 (IGFBP-3), inducible nitric oxide synthase (iNOS), cyclo-oxygenase 2 (COX-2) and an increased synthesis of collagens VI, III, X and IIA (Goldring 2000a; Sandell and Aigner 2001). These differentiated, hypertrophic chondrocytes also express increased levels of matrix metalloproteinases (MMPs) and are apoptotic compared with chondrocytes from normal cartilage (Aigner and Kim 2002). Alterations in matrix structure, in turn, serve to alter the mechanical forces acting on the chondrocyte resulting in further degenerative pathology (Poole et al. 2002).

Several cartilage-degrading proteases have been implicated as having a role in the pathology of OA. The matrix metalloproteinases (MMP) family are thought to have a major role in the structural changes seen in OA cartilage. At least four, MMP -1, -8, -13, and -14, have been implicated in type II collagen degradation with MMP-13 thought to have the major role (Dahlberg et al. 2000; Wu et al. 2002b). Although MMPs (especially MMP-3, -8 and membrane-type metalloproteinase [MT-MMP]) have also been proposed to have a role in aggrecan cleavage the more recently identified aggrecanases ADAMTS4 and ADAMTS5 (members of the a disintegrin and metalloproteinase with thrombospondin motifs [ADAMTS] family of extracellular proteases) are likely to be the major players in this process (Abbaszade et al. 1999; Hurskainen et al. 1999; Tortorella et al. 1999). MMP action is regulated by the action of tissue inhibitors of metalloproteinases (TIMPs) which are able to form tight complexes with the MMPs. MMP -2 and -9, Cathepsins B, L and D and plasminogen activator have been implicated in the degradation of other matrix components (Poole 1997).

It is predicted that the balance between catabolic and anabolic events, and hence integrity of cartilage, both normally and in disease, is regulated through the action of cytokines on the chondrocytes. Cytokines may originate from activated synoviocytes, mononuclear cells or from the chondrocytes themselves, although evidence suggests that the majority of the cytokines involved in cartilage metabolism originate from the chondrocytes (Moos et al. 2000). An excess of pro-inflammatory cytokines is thought to be responsible for many of the clinical manifestations of OA and they have been proposed to play an important role in disease initiation and progression (Fernandes et al. 2002).

Two of the key degradative cytokines implicated in OA are interleukin 1 (IL1) and tumour necrosis factor α (TNFA) which are expressed by chondrocytes in the majority of OA patients. TNFA is thought to be more important in the onset of OA whilst IL1 is important in both early and late stages (Fernandes et al. 2002). Studies by Moo et al (2001) have suggested that chondrocytes in OA cartilage may display one of two mutually exclusive phenotypes characterised by elevations in either TNFA (TNFA^{high} phenotype) or IL1B (TNFA^{low} phenotype). Other catabolic, pro-inflammatory cytokines have also been implicated as having a role in OA. These include IL17, IL18, IL6, IL8, leukaemia inhibitory factor (LIF) and oncostatin M (Goldring 2000a; Fernandes et al. 2002).

The other group of cytokines important in the OA disease process are the inhibitory or anti-inflammatory cytokines. This group, which includes IL4, IL10 and IL13, act mainly by decreasing the synthesis or activity of the pro-inflammatory cytokines (Goldring 2000b). Studies on synovial fibroblasts have demonstrated that IL4, IL10 and IL13 act through different mechanisms to decrease TNFA-induced release of prostaglandin E₂ (PGE₂). IL4 and IL10 have both been demonstrated to be chondroprotective *in vivo* and all three have been implicated in decreasing production of IL1B, TNFA and MMPs and in upregulating both interleukin 1 receptor antagonist [IL1RN] (which in turn blocks PGE₂ synthesis, degradation of the cartilage matrix and production of collagenases and nitric oxide) and TIMP-1 synthesis (Alaaeddine et al. 1999; Fernandes et al. 2002).

1.4.2 Bone

There is an acceleration in the turnover of subchondral bone in OA with increases in MMP-2 and alkaline phosphatase (key enzymes in bone metabolism) expression by osteoblasts compared with normal or osteoporotic bone (Mansell et al. 1997; Mansell and Bailey 1998). This increased metabolism appears to favour synthesis over degradation and results in an hypomineralised, thickened, stiffened bone (Bailey et al. 2001). This increased and deregulated turnover often results in the formation of osteophytes and subchondral cysts, possibly in an attempt to stabilise the otherwise unstable joint (Potenger et al. 1990).

There has been much recent discussion as to whether the observed changes in bone precede or are subsequent to those observed in cartilage. For some time it was presumed that changes in cartilage were the initiating factor in the disease which then triggered the observed changes in bone turnover. An increasing number of studies however, have challenged this assumption. Some animal models appear to suggest that the increased subchondral bone density observed in many OA patients may precede the cartilaginous changes observed in OA. For example, in guinea pig models of OA lowered rates of bone metabolism appear to be correlated with decreased susceptibility to OA (Watson et al. 1996; Huebner et al. 2002). In humans there also appears to be a positive correlation between bone mineral density and OA (Dequeker 1997; Stewart and Black 2000). In addition, studies suggest that artificial lowering of bone turnover rates with the use of bisphosphonates is chondroprotective (Christgau et al. 2001) and that bone cells from OA patients can induce alterations in cartilage metabolism (Westacott et al. 1997). Bailey et

al (2001) provide an overview of the numerous animal and human studies that suggest subchondral bone changes precede the initiation of cartilage fibrillation. Results of many studies however, have been contradictory or inconclusive and the majority of investigators still appear to presume that the cartilaginous changes observed in OA precede any changes in bone metabolism (e.g. Lane and Nevitt 2002) with many more articles and reviews on the role of cartilage rather than bone in OA initiation still being published. There is, however, an increasing interest in the involvement of bone especially with regards disease progression and much investigation is being undertaken with the aim of increasing understanding of the role of changes in bone metabolism on OA initiation and progression (Bailey et al. 2001). There is also an increasing agreement, that an increase in subchondral bone density is associated with and indeed may even be necessary for disease progression (Burr 1998).

In the absence of further evidence it seems necessary at present to conclude that OA is a disease of unknown origin which results in degenerative changes in both bone and cartilage in such a way that suggests a positive feedback loop between them (Burr 1998).

1.5 The Genetic Component of OA

1.5.1 Epidemiological studies

A role for genetic factors in determining susceptibility to primary OA was recognised as early as 1941 when Stecher et al reported on familial clustering of nodal disease (OA characterised by Heberden's nodes at the distal interphalangeal joints and Bouchard's

nodes at the proximal interphalangeal joints of the hand). Subsequent work by Kellgren (1963), who sub-divided generalised OA (OA involving multiple joints, typically defined by combined presence of hand and large joint OA) according to the involvement of interphalangeal joints, proposed a strong genetic component in patients with generalised OA that involved the interphalangeal joints. Kellgren (1963) also suggested a less well-defined and possibly polygenic component in patients with generalised OA which didn't involve the hands.

In 1977 Lawrence published results of twin studies showing approximately double the concordancy for nodal OA in monozygotic compared with dizygotic twins further confirming the role of genetic factors in OA.

More recently, a number of epidemiological studies, including twin pair and family risk studies, have been undertaken to formally investigate the genetic component of OA (for a comprehensive review see Loughlin 2001).

1.5.1.1 Twin studies

Five, large scale, twin studies have been undertaken in recent years to estimate the heritability of OA (Kaprio et al. 1996; MacGregor et al. 1998; Sambrook et al. 1999; Spector et al. 1999; Page et al. 2003). In summary these studies estimate heritability for OA in females to be between 39 and 79% depending on joint considered, with estimates for heritability of OA at the knee and hand being between 39 and 65% and at the hip 58% (table 1.5a). In the study by Kaprio et al (1996), who investigated OA at any joint,

Twin cohort ¹	Sex	Affected joint	Diagnostic criteria	Country	Heritability	Reference
130 MZ pairs and 120 DZ pairs	Female	Hand or knee	Radiographic	United Kingdom	39% - 65%	Spector et al. 1996
135 MZ pairs and 277 DZ pairs	Female	Hip	Radiographic	United Kingdom	58%	MacGregor et al. 2000
577 MZ pairs and 1180 DZ pairs	Male	Any joint	Physician diagnosed	Finland	No heritability for men	Kaprio et al. 1996
836 MZ pairs and 1502 DZ pairs	Female				44% for women	
172 MZ pairs and 154 DZ pairs	Male and Female	Spine	Magnetic resonance imaging	United Kingdom and Australia	63 - 79% (no sex stratification)	Sambrook et al. 1999
1224 MZ pairs and 1137 DZ pairs	Male	Hip	Radiographic	United States	53 - 61%	Page et al. 2003

¹ MZ = monozygotic twins, DZ = dizygotic twins

Table 1.5a: Summary of recent twin-pair studies for primary osteoarthritis

stratification by sex revealed a striking difference with evidence for a genetic component of 44% in females but no evidence for such a component in males. The study by Page et al (2003) is the only study to have focused exclusively on male twins. They estimate that an additive genetic component accounts for 53 - 61% of the variance in liability of radiographically determined OA of the hip.

1.5.1.2 Sibling risk studies

There have been four, large scale, sibling risk studies using severely affected probands and their siblings (Chitnavis et al. 1997; Ingvarsson et al. 2000; Lanyon et al. 2000; Jonsson et al. 2003). All four studies identified a higher risk of OA for siblings of probands with severe OA than for a control population. Estimates of this increased risk range from 2.3 to 14.4 times the control risk (table 1.5b). Chitnavis et al (1997) investigated probands in an UK cohort who had undergone total hip or knee joint replacement for OA and their siblings. They found a 2.3 times increased risk for joint replacement in siblings versus controls. A second study, also in the UK, by Lanyon et al (2000) identified a 9.8 times increased risk of hip OA in siblings of hip OA probands compared with a control population. Interestingly, this study found a higher risk for male siblings of cases (relative risk of 14.4) than females (relative risk of 7.7) apparently contradicting the results of the Kaprio et al (1996) twin study.

The remaining two studies investigated the sibling risk of hip (Ingvarsson et al. 2000) and hand (Jonsson et al. 2003) OA in Icelandic cohorts. Siblings of probands with hip OA were found to have a three times increased risk of hip replacement for OA compared with

Cohort	Sex	Affected joint	Diagnostic criteria	Country	Sibling risk	Reference
402 probands, 1171 siblings, 376 spouses	Male and female	Hip or Knee	Joint replacement	United Kingdom	2.3	Chitnavis et al. 1997
392 probands, 604 siblings, 1718 controls	Male and female	Hip	Joint replacement	United Kingdom	Females 7.7 Males 14.4	Lanyon et al. 2000
2713 patients ¹	Male and female	Hip	Joint replacement	Iceland	3.1	Ingvarsson et al. 2000b
2919 patients (2449 female and 470 male) ¹	Male and female	Hand	Clinical and radiographic	Iceland	2.0 - 6.9 depending on sex and classification	Jonsson et al. 2003

¹ No control group used as investigators had access to population-based list of patients with hip joint replacement for OA/hand OA and the total numbers of individuals in the population born in the period of interest.

Table 1.5b: Summary of recent sibling risk studies for primary osteoarthritis

the general population. For hand OA the relative risk ranged from 2.0 to 6.9 for female siblings of women with hand OA depending on the joints involved. Jonsson et al (2003) reported that males and females show similar relative risks for hand OA but that prevalence is lower in males. The genetic component of hand OA also appears to increase with severity (Jonsson et al. 2003). Both Ingvarsson et al (2000) and Jonsson et al (2003) discovered that Icelandic patients with OA of the hip or the hand descended from a smaller number of founders than did the control population indicating that these patients are significantly more related to each other than to matched controls.

1.5.1.3 Segregation studies

Other family based studies have also been used to measure the degree of heritability of OA. Felson et al (1998), investigating the inheritance of generalised OA in 337 nuclear families participating in the Framingham OA longitudinal study in the United States, found clear correlation of OA between parent-offspring and sibling-sibling pairs and observed that these correlations were higher for mother-child pairs than father-child pairs. Again, as for Kaprio et al (1996), this suggests a greater heritability of OA in females than in males. Felson et al (1998) also propose that their results imply the involvement of a major recessive gene with a multifactorial component accounting for the genetic susceptibility to OA. A segregation study in 295 Chuvashian families with hand OA by Kalichman et al (2003) also found evidence for a significant familial aggregation, which could best be explained by a mixed Mendelian model. They suggest that a putative major gene accounts for between 49 and 52% of the OA phenotype after adjustment for age and sex.

Analysis of data from the Baltimore longitudinal study on ageing for 167 families with hand OA and 157 families with knee OA (Hirsch et al. 1998) revealed a role for a genetic factor in determining severity of disease for hand and for hand plus knee OA. This study, in contrast to many others, however, found no evidence for a genetic component in determining susceptibility to knee OA alone.

In the Rotterdam study of the elderly, Bijkerk et al (1999), found evidence for the involvement of a genetic component in hand and spine OA but no such component for OA of the knee or hip. This may reflect too small a sample size to detect heritability for OA at these two joint sites (30 probands & 17 siblings for hip OA and 70 probands & 47 siblings with knee OA compared with 97 probands & 192 siblings and 95 probands & 213 siblings for hand and spinal OA respectively) since these results appear to be in contrast with the majority of other epidemiological studies. Alternatively, such conflicting results may reflect the complex heterogeneity of OA disease so that this cohort consists of individuals whose OA phenotype is not accounted for to a significant extent by genetic influences.

A more recent epidemiological study by MacGregor et al (2002) suggests that different genes may account for disease susceptibility at the hip compared to those for the knee.

Although there are a small number of conflicting results, probably reflecting the complex nature of disease susceptibility, taken together these numerous epidemiological studies provide strong evidence that OA susceptibility has a significant genetic component, with an average heritability of ~ 50%, but also that other, environmental factors play a role.

They also infer that the size and nature of the genetic component of susceptibility may vary depending on both sex and the joint(s) affected.

There are a number of studies which have reported on rare pedigrees in which primary OA (not secondary to any other factor such as dysplasias/injury) appears to be transmitted in a Mendelian fashion (Palotie et al. 1989; Meulenbelt et al. 1997; Ingvarsson et al. 2001). These pedigrees are usually characterised by a severe, early onset form of OA. The extent to which the genetic component in these families can be extrapolated to the more common, complex form of OA is as yet unknown.

There are also a number of rare, inherited diseases (osteoochondrodysplasias) caused by mutations in genes that encode for structural molecules of cartilage or bone which display OA as a secondary phenotype (for a review see Jimenez and Williams 1999). One such dysplasia, multiple epiphyseal dysplasia, characterised by an early onset of OA of large joints has been attributed to mutations in the matrilin-3 gene (MATN3) [Chapman et al. 2001]. MATN3 is a non-collagenous cartilage extracellular matrix protein which may be involved in the inhibition of normal development and maintenance of cartilage and bone. A recent study by Stefansson et al (2003), in Icelandic patients concordant for hand OA affecting the first metacarpal joint, found evidence for linkage of OA to the short arm of chromosome 2 where the MATN3 gene resides. Further investigation found evidence for co-segregation of a missense mutation of MATN3 with hand OA in a number of families. This potentially represents a gene involved in an osteochondrodysplasia which also has a role in primary OA susceptibility. Investigations of genes involved in rare dysplasias

however, have often produced conflicting results as to their involvement in encoding susceptibility to common, primary OA.

1.6 Non-genetic risk factors for OA

Several epidemiological studies have also identified a number of non-genetic risk factors for OA. These have been summarised in reviews by Felson and Zhang (1998) and by Felson et al (2000). Although studies have often produced conflicting results for a number of the factors proposed to predispose to disease, a number of factors have been generally accepted as conferring increased risk for OA. Even for these generally accepted factors, however, as is the case for the genetic studies, there is not complete concordance between studies or final agreement as to the degree of increased risk accountable to them. Such apparent conflicts may be the result of a number of variables including inter study design variations, differences in disease phenotype definition, or insufficient sample sizes. It is likely that a proportion of this observed discordancy also reflects the complex nature and heterogeneity of the OA disease process.

1.6.1 Age

OA is generally uncommon in individuals under the age of 45 and increases in prevalence up to the age of 65 with incidence appearing to level off or decline at about 80 years of age (Felson et al. 1995; Oliveria et al. 1995; Creamer and Hochberg 1997). It has been proposed that a decline in the ability of the chondrocyte to maintain the integrity of the articular cartilage alongside destabilisation of the joint, due to failure of ligaments,

decreased joint proprioception and a decline in neuromuscular function, may contribute to this age-related effect (Jones and Doherty 1995; Haq et al. 2003; Martin and Buckwalter 2003).

1.6.2 Sex

Both incidence and prevalence of OA are higher in males before the age of 50 (especially in the knee, proposed to be secondary to injury) but then subsequently become increasingly higher in women (Van Saase et al. 1989, March and Bagga 2004). This female preponderance after menopausal age alongside the recognition in some females of rapidly progressing ‘menopausal OA’ of the hand has led to the proposal of an hormonal involvement in OA. This is further supported by some studies showing that hormone replacement therapy (HRT) offers some protection against development of OA (Hannan et al. 1990; Dennison et al. 1998). As mentioned previously, epidemiological studies also appear to show that the genetic component of OA is higher in females than in males (Kaprio et al. 1996; Felson et al. 1998).

1.6.3 Body mass index

Body mass index (BMI) is generally a well-recognised factor associated with OA of the knee, with obesity (BMI >25) accounting for more knee OA in females than any other currently recognised factor. An elevated BMI has been shown to be a risk factor for both OA initiation and progression at the knee (Felson 1990; Hart et al. 1999; Manek et al. 2003).

For hip OA, results of different studies into the effects of BMI have been conflicting. A review of published reports finds only moderate evidence for a positive association between obesity and hip OA (Lieveense et al. 2002), whilst a prospective study carried out in Scandinavia found that women in the highest BMI quartile had an approximately three times increased risk of total hip replacement due to OA compared with those in the lowest quartile (Flugsrud et al. 2002). More recently a prospective study of 120,000 women in Massachusetts, USA found evidence that an increased risk of hip OA requiring joint replacement surgery was associated with elevated BMI and older age (Karlson et al. 2003).

It has been proposed that obesity may increase the risk of OA either by increasing the load across susceptible joints or through the action of an obesity-associated circulating factor accelerating articular cartilage breakdown (Hartz et al. 1986; Van Saase et al. 1988; Manek et al. 2003). The association of obesity with OA is higher in women than in men.

1.6.4 Bone density

The integrity of the articular cartilage is dependent at least in part on the mechanical properties of the underlying subchondral bone. It is suggested that stiffening of the bone due to increases in bone mineral density (BMD) may cause increased cartilage degradation on repeated loading (Radin and Rose 1986). This is supported both by the generally accepted inverse correlation observed in a number of studies between osteoporosis and osteoarthritis (see reviews by Dequeker 1997; Brandi et al. 2001) and by

the increased bone density observed in OA patients versus controls in several studies (Dequeker 1997; Stewart and Black 2000). However, more recent studies have implicated a complex relationship between BMD and OA, with a potential stage-dependent effect. Some studies suggest that high BMD may be predictive of progression of the disease (Bruyere et al. 2003) whilst others find correlation between high BMD and slower disease progression (Zhang et al. 2000). Lane and Nevitt (2002) summarise the results of a number of studies on the relationship between BMD and OA and conclude that the relationship between them is complex and requires further investigation.

It has been proposed that osteophyte formation, often seen in the OA disease process, may be linked to high bone mass thus proposing a role for some form of circulating bone growth factor (Zhang et al. 2000; Hart et al. 2002). Hart et al (2002) demonstrated that there is an association between increased BMD of the spine and hip and incident knee OA characterised by osteophytes. Osteophytes are formed by endochondral ossification in contrast with the subchondral bone osteogenesis seen in OA which is linked with bone resorption (Brandt 2003). It is thought that osteophytes are likely to represent an attempt to repair or stabilise the joint (Neuman et al. 2003).

1.6.5 Local biomechanical factors/ joint injury

Articular cartilage may be damaged by two mechanisms:

- i) Sudden major trauma, such as ligament tears. This is a common cause of secondary OA. In the Framingham study (Felson 1990) males with a history of major knee injury were found to be at 5 - 6 times increased risk of knee OA compared to those with no such history. OA can also be induced in animal models by damaging the meniscus and/or the surrounding structures (Elliott et al. 1999).
- ii) Stereotyped repetitive joint use, such that the joints are overworked and the protecting muscle fatigued increasing force transmitted to cartilage. Certain occupations carry an increased risk of OA as a result of this i.e. cotton mill workers had an increased prevalence of Heberden's nodes whilst farmers have a high risk of hip OA (Lawrence 1961; Croft et al. 1992; Thelin et al. 2004).

1.6.6 Additional proposed factors

Several other factors have been implicated in conferring risk to OA. For example, ethnicity, diet, levels of recreational activity and muscle weakness. Evidence for these, however has been less consistent and compelling with much interstudy variability (Felson and Zhang 1998; Felson et al. 2000).

1.7 Summary of disease associated factors

In summary, epidemiological, physiological and genetic studies appear to suggest that, whilst OA can (and often does) result directly from severe joint injury, the disease appears to frequently be the result of the interaction of a number of systemic (i.e. obesity, sex or genetic) and local (i.e. joint instability or dysplasia) factors (Jones and Doherty 1995). It is envisaged that systemic factors serve both to increase vulnerability of cartilage to injury and reduce its capacity for repair. Local factors acting on this vulnerable cartilage may then serve to initiate the joint breakdown characteristic of OA (Felson and Zhang 1998).

The involvement of multiple factors in determining both the onset and progression of the disease has led to the description of OA as an oligogenic, multifactorial disease.

1.8 Complex trait analysis

Although historically diseases were classified as monogenic (where mutations at a single locus are predictive of phenotype) or multifactorial (involving multiple disease associated genetic loci and/or environmental factors), there is increasing evidence for the existence of a continuum between the two extremes. It appears that few, if any, traits are truly monogenic with the emergence of increasing instances of both locus heterogeneity and the effects of disease modifiers (Badano and Katsanis 2002). In cystic fibrosis, the prototypic monogenic disease, for example, disease phenotype has now been shown to be affected by at least 6 modifier genes with some individuals with milder forms of the

disease identified as completely lacking mutations in the CFTR (cystic fibrosis transmembrane conductance regulator) gene (Groman et al. 2002). In addition, heterozygous mutations of the CFTR gene have been identified as being associated with rhinosinusitis; an identified complex trait (Wang et al. 2000).

Complex diseases, such as OA, are characterised by multiple genetic, developmental and environmental factors. They are non-Mendelian; but show familial aggregation and are generally more common than single-gene disorders. It is hoped that identifying the genes and mutations involved in complex disease susceptibility will lead to a better understanding of the pathogenesis of these disorders and result ultimately in improved treatments and diagnostic ability and perhaps eventually prevention.

The identification of genes in complex traits however, is complicated by a number of factors including the presence of locus and allelic heterogeneity, epistasis, low penetrance and variable expressivity.

The past two decades have seen a sharp increase in attempts to map the genes involved in susceptibility to complex diseases. Such attempts have used both linkage and association-based techniques. Linkage studies, traditionally genome-wide have the benefit of being both systematic and allowing for the study of the whole genome using only a few hundred markers. However, they have limited power for detecting disease causing alleles with only modest effects (Weeks and Lathrop 1995). Association studies, by contrast, concentrate on a particular gene or region of interest (genome-wide association studies are currently becoming feasible due to the drop in costs of marker genotyping [e.g.

Sawcer et al. 2004]). Such studies have much higher power than linkage studies for detecting disease-causing variants with modest effects - risk alleles with a frequency of 1% can readily be detected in association but not in linkage studies (Weeks and Lathrop 1995; Botstein and Risch 2003). In practise, the two techniques are often used as fully complementary approaches, with linkage studies used to identify a targeted genomic region that may then be investigated by association analysis (See reviews by Glazier et al. 2002; Vink and Boomsma 2002; Botstein and Risch 2003).

There is much debate as to the best methods by which to localise and identify complex trait susceptibility genes accentuated by the dearth of success stories to date (Risch 2000). Although there is some disagreement between different studies, many suggest the following chronological steps:

- 1) The use of genome-wide linkage analysis to identify a chromosomal region(s) showing evidence for linkage to the disease phenotype. Such scans commonly utilise highly polymorphic microsatellite markers which reduce the effects of non-informativity often seen for less polymorphic markers (Weeks and Lathrop 1995).
- 2) Finer mapping of the identified region(s) using linkage or association analysis. Such analysis aims to narrow the region facilitating identification of candidate genes.
- 3) Sequence analysis of the candidate region(s) or genes within the region(s) identified by linkage to identify genetic variants, which are then prioritised and tested for association with the disease phenotype in either family-based or case-control cohorts. The markers of choice for such association analyses are usually single nucleotide polymorphisms (SNPs). These biallelic markers are common, accounting for about 90% of sequence variation in

humans, and are amenable to high throughput, low cost, automated genotyping (Akey et al. 2003).

4) Functional testing of candidate genes and variants showing association is undertaken in model systems or organisms to prove functionality. Additional circumstantial evidence for the role of the candidate gene may be provided in the forms of patterns of gene expression, similarities to phenotypes associated with naturally occurring or engineered mutations in animal models and strong mechanistic evidence.

1.8.1 Linkage analysis

Linkage studies in complex traits often focus on the use of affected sibling pairs (ASP). The use of ASPs allows linkage analysis to be undertaken in the absence of traditional Mendelian pedigrees and also enables the investigator to use non-parametric analysis (Vink and Boomsma 2002). In the absence of linkage at a given marker it would be expected that siblings would share 0, 1 and 2 alleles identical by descent (IBD) [i.e. both alleles descended from the same ancestral chromosome] with frequencies of 0.25, 0.50 and 0.25 respectively. ASP linkage studies are based on the idea that linkage at a given marker locus is supported if the siblings share alleles at that locus more often than would be expected by chance alone. This method is most powerful when IBD status can be determined. This ideally involves genotyping of parents which is often not possible for late onset diseases such as OA. In such situations genotyping of unaffected siblings can be used to assist in IBD determination. Non parametric methods, requiring no prior model definition, are normally used since information on disease gene frequency, penetrance or mode of inheritance are normally unknown. The ability of such methods to

detect linkage depends on the contribution to heritability of the locus being considered (Weeks and Lathrop 1995; Vink and Boosma 2002).

Published linkage studies differ widely in design reflecting the lack of consensus as to the best method(s) to maximise detection of susceptibility loci. There have been numerous published reports on the use of genome-wide linkage scans to identify regions harbouring susceptibility genes for complex disease (e.g. Gretarsdottir et al. 2002; Hakonarson et al. 2002; Olson et al. 2002). A review by Altmuller et al (2001) identified 101 such studies with varying design and outcome. They found that the only predictors for a successful outcome in terms of significant or suggestive linkage results were increased numbers of individuals and sampling from an ethnically homogeneous group. Beyond this the authors conclude that there is no gold standard applicable to complex diseases.

There is much debate as to what constitutes a significant outcome in linkage analysis for complex traits. Much emphasis is placed on the recommendations by Lander and Kruglyak (1995) who suggest that significant evidence of linkage requires a LOD score of ≥ 3.6 and suggestive evidence of linkage a LOD score of ≥ 2.2 , in a dense, complete genome scan. However, less than a third of reports to date have identified regions with evidence of suggestive linkage using these criteria (Wiltshire et al. 2002), which are based on the assumption of both dense-marker genotyping (rarely the case in genome-wide scans) and near complete extraction of inheritance information (not possible if parental genotypes are unknown). It has been suggested that such criteria are too strict, and Lander and Kruglyak (1995) themselves acknowledge that the use of these criteria may result in the rejection, as false positives, of results which indicate the presence of

susceptibility genes with modest effect. As a result many investigators have recommended the use of empirical calculations of significance based on the individual study design. Sawcer et al (1997) carried out such simulations on their data for a genome-wide scan for multiple sclerosis (MS) disease loci and identified values for significance almost 25% lower than that recommended by Lander and Kruglyak. These results underline the poorly understood discrepancy between theoretical simulations and simulations based on real data. Using simulations Wiltshire et al (2002) determined that under the conditions encountered during a typical genome-wide scan, with a 10 cM marker map and approximately 15% missing genotypes, you would expect a locus with a LOD score of between 1.5 and 1.6 to occur once per genome (compared to a locus with a LOD of 2.2 predicted by Lander and Kruglyak [1995]).

Wiltshire et al (2002) also propose locus counting as a complementary approach to the use of significance levels. This involves comparing the number of independent loci identified in a genome-wide screen with the empirically-determined number expected by chance. They conclude that for a typical genome-wide scan of a complex trait using a 10 cM marker density you would expect 2 - 3 independent loci with a LOD score ≥ 1.2 by chance. The exact number being dependent on the details of the study design. Obviously, however, this method does not allow determination as to which regions are true and which are false positives.

In practice, there appears to be large variation in published literature as to the LOD scores reported as indicating 'significant' evidence for linkage. Under such circumstances, the

most compelling evidence of the involvement of a region in encoding for disease susceptibility remains the independent replication of the original linkage results.

Following the identification of a region showing evidence for linkage, work is often undertaken to narrow the interval. This may involve the use of additional linkage mapping with a higher density of markers or the use of association studies across the region.

1.8.2 Association studies

Once a region harbouring a susceptibility allele(s) has been identified and narrowed then association analysis may be undertaken to identify the disease causing variant(s).

Association studies are comparable to classic epidemiological case-control studies and traditionally involve comparison of allele frequency distributions between cohorts of cases and population controls (Tabor et al. 2002).

Association analyses may be systematic, covering the whole of the region under investigation (Patil et al. 2001; Gabriel et al. 2002), or candidate gene based, identifying variation in genes prioritised according to their potential role in disease biology (Johnson and Todd 2000; Risch 2000). Both methods have their advantages. The systematic association approach necessitates the use of a greater number of markers than those employed in the candidate gene based approach thereby increasing cost and time of the study. However, this approach has the advantage over the candidate gene based approach

of requiring no prior assumption as to the gene(s) likely to be involved in a disease process which is often poorly understood. Systematic approaches also ensure coverage of the regulatory regions of genes which are often ill defined and may be distal to the gene itself (Vink and Boomsma 2002; Botstein and Risch 2003).

1.8.2.1 Linkage disequilibrium

When a disease-causing variant occurs it does so on one specific background of other genetic variation (a haplotype). Initially, this variant is in complete linkage disequilibrium (LD) with all these other variants. Over time however, LD is eroded by recombination. The rate at which the LD is eroded is dependent on the distance of loci from the disease variant so that association is lost more quickly for distal than proximal markers. Association studies are based on the assumption that marker alleles that are associated with a particular phenotype are therefore closely linked to the disease mutation. This is probably an over simplistic picture however as patterns of LD are also complicated by for example new or recurrent mutations (of the disease or marker loci) and by the non-random distribution of recombination (Kauppi et al. 2003).

1.8.2.2 Patterns of linkage disequilibrium

Patterns of LD across genomic regions differ according to a number of factors including age of mutation, drift, selection, demographic history and local recombination rates (Shifman et al. 2003). LD can be destroyed by recurrent mutations, differing allele frequencies and recombination/gene conversion events creating complex patterns of LD such that markers can be in stronger LD with distal markers than with markers in close

proximity (Cardon and Bell 2001). Patterns of LD differ between ethnic groups too, with older populations i.e. sub-Saharan Africans showing less extensive LD than that observed in more recent populations i.e. Europeans (Shifman et al. 2003).

.

LD extends over much greater distances (of up to 14 Mb) between microsatellites than between SNPs. This is probably due to microsatellites representing multiallelic markers with a higher mutation rate; so that these markers carry more recent alleles compared with the biallelic, often ancient SNP markers (Kauppi et al. 2003). SNPs are fast becoming the markers of choice for association studies as a result of their mutational stability, relative abundance and amenability to high throughput genotyping. It has been suggested however that microsatellites could be useful if the disease under consideration is accounted for by rare variants since microsatellite markers allow for the identification of rare as well as common haplotypes (Varilo et al. 2003).

1.8.2.3 Haplotype blocks and recombination hotspots

Many studies of the patterns of LD across the genome suggest that recombination is not uniformly distributed and that blocks of little or no recombination (cold spots) are separated by regions showing high recombination rates (hotspots) [Petes 2001; Phillips et al. 2003; Stumpf and Goldstein 2003]. Such block-like patterns have been experimentally confirmed in at least one region examined, the major histocompatibility (MHC) class II region. By looking at recombination in sperm and LD patterns, Kauppi et al (2003) were able to define a 60 kb LD block in three populations (Saami, UK North Europeans and Zimbabweans) which ended at experimentally verified recombination hotspots. Some

recombination within this LD block was observed (the rate of which varied between populations) but at a much lower rate than was observed in the hotspots. Kauppi et al (2003) suggest therefore, that in this region at least, recombination hotspots have played a large role in shaping patterns of LD. The observation of a less clear block structure in the more ancient Zimbabwean population, differences in LD across weaker hotspots, the decline in LD across the block and the presence of at least four SNPs showing evidence for an obligatory recombination event, however, suggests that other, as yet undetermined factors, are also important in determining patterns of LD in this region. In contrast to observations of other regions of the genome the MHC region appears to contain very few cosmopolitan haplotypes (haplotypes shared between populations) suggesting that this region displays a high haplotype turnover and therefore, that observed haplotypes are not the result of ancient recombination events. This could be driven by the known high selection pressures on the MHC region. As a result of this, the extent to which observations in the MHC region can be extrapolated to other regions of the genome is as yet unclear.

The wider identification and confirmation of haplotype blocks in other regions has been complicated by the somewhat arbitrary and subjective definitions of haplotype block boundaries used in many studies. In some instances, as for the MHC class II region, haplotype blocks have been defined as regions between experimentally established recombination hotspots. Such definitions often result in relatively large haplotype blocks consisting of fifteen or more common haplotypes (Kauppi et al. 2003). By contrast, more stringent definitions such as that used by Patil et al (2001) have required a certain percentage of diversity within a block to be accounted for by common haplotypes. The

percentage of diversity and the number of haplotypes required to define a block being left to the individual investigator. Other studies have investigated pairs of SNPs, looking for pairs where LD falls below a certain (subjective) level to delineate block boundaries. Gabriel et al (2002) using this method identified a number of blocks which were defined by between three and five common haplotypes accounting for >90% of all chromosomes examined. Block definition also depends on the density and allele frequencies of SNPs used, with dense maps identifying shorter blocks and the use of SNPs with a higher minor allele frequency defining longer blocks. All of these different methods result in differing block boundaries and therefore differences in block sizes. In addition, there is some argument as to whether such blocks exist at all (Couzin 2002).

Much further investigation is needed to determine the presence, frequency and usefulness of such blocks (Stumpf 2002; Cardon and Abecasis 2003). Even if haplotype blocks can be defined, modelling has shown that the block-like structure observed experimentally could be due to the stochastic distribution of recombination across the genome and is not necessarily the result of recombination hotspots (Wang et al. 2002). Recombination hotspots have been demonstrated to exist in other organisms such as the yeast *Saccharomyces cerevisiae* (Gerton et al. 2000). In humans Jeffreys et al (2001) and Kauppi et al (2003) have demonstrated the presence of recombination hot spots affecting the pattern of linkage disequilibrium in the MHC region. The extent to which these observations may be extrapolated to the genome as a whole however is unclear.

If haplotype blocks are shown to exist they will be useful in defining the areas identified by association studies within which candidate complex disease susceptibility genes are

likely to reside. Understanding of the patterns of LD in a region may also be utilised to select a subset of SNPs for genotyping that will retain most of the genetic variation within a region whilst reducing the number of markers needing to be genotyped in a larger sample. A number of methods to select such a subset of SNPs have been proposed, some based on the haplotype structure of a region (e.g. <http://www-gene.cimr.cam.ac.uk/clayton/software/stata/htSNP/htsnp.pdf> [Clayton 2001]) and others based on decomposition of the pairwise LD between markers (e.g. Meng et al. 2003).

The power of an association study to detect disease susceptibility-encoding variants can be influenced by a number of factors including sample size, density of marker map, levels of LD within the investigated region and the age and frequency of both the disease mutation and marker alleles. In view of this many investigators advocate the use of large sample sizes, in order to maximise power to detect association (especially relevant if the genetic effect of the variant is small), and the use of dense maps, to enable location of ancient, disease susceptibility encoding alleles where recombination will have broken down all LD except at the closest linked markers. Optimal study design also includes careful control population selection in order to minimise the likelihood of false positive results due to population stratification. Population stratification, where cases and controls come from different subpopulations with differing allele frequencies, has often been reported as the reason for failure to replicate initial positive results in subsequent association studies. A recent paper by Marchini et al (2004) found that even modest population substructure could have serious consequences especially with the large sample sizes required for genome wide association studies. Methods to correct for these

differences, such as Genomic Control (Devlin and Roeder 1999) are only effective if enough loci are considered.

In addition to the presence of population substructure, failure to replicate initial positive results may also be due to other factors such as over interpretation of initial results, inadequate statistical power and biological and phenotypical complexity of the trait investigated (Cardon and Bell 2001).

Once a variant(s) showing association with disease phenotype has been identified there remains the long and often arduous task of identifying the disease causing variant(s). This may be the variant(s) showing association or alternatively other variant(s) in LD with them. Determination of LD patterns and allele frequencies in distinct ethnic groups may help to further narrow the region containing the susceptibility allele by allowing for determination of ancestral breakpoints (Johnson and Todd 2000).

1.8.3 Functionality testing

For complex traits functionality tests are necessary to prove the disease causing nature of the variant(s). This may involve knock-in technology, gene targeting or transgenic complementation (Rocha et al. 1998; Nebert et al. 2000; Kamath et al. 2003). Additional evidence as to the disease causing nature of the variants may be gained through tissue expression patterns, mechanistic support and observations of similar phenotypes in naturally occurring or engineered mutations in animal models (Picciotto et al. 2000; Crabbe 2002).

Many variations on this protocol have been used to varying success. There are a limited number of confirmed success stories for isolating genes causing susceptibility to complex traits in humans. A selection of these successes are summarised in table 1.8. The small number of examples to date precludes any generalisations both as to the best method(s) to identify genes in complex diseases and as to the likely nature of disease causing variants in complex traits.

1.9 Conditioning analysis

Conditioning analysis allows for the identification of potential interactions (such as epistasis) between regions showing evidence of linkage in genome scans. It is achieved by weighting the families used for linkage analysis in one region according to their evidence for linkage, as determined by their non-parametric linkage (NPL) scores, at another, independent region. In this way the analysis involves looking for correlation between family NPL scores at the two independent loci (Cox et al. 1999). Such analysis is envisioned as being useful in allowing the regions identified in genome scans to be narrowed, increasing confidence in the validity of loci and possibly allowing replication of results in independent samples where susceptibility genes may be contributing to susceptibility to different levels.

1.9.1 The example of type II diabetes

One disease where such analysis has been successful is type II diabetes. An original scan in Mexican American families identified a major susceptibility gene, designated

Disease	Chromosome region	Gene	Identification of disease causing variants	Additional evidence	Select references
Type 1 diabetes	6p	HLA-DQ β	Multiple linkage analysis and association studies implicate DQ region in disease susceptibility. All disease-associated DQ β chains have Ala/Val/Ser at amino acid 57. Those chains with low frequency in Caucasian cases have Asp.	NOD (non obese diabetic) mouse A β (mouse equivalent of DQ β) has Ser at amino acid 57, different from non diabetic mice who have Asp. Potential functional effect of change in humans predicted from modelling of molecule.	Todd et al. 1988; Davies et al. 1994; Concannon et al. 1998
Type 2 diabetes	2q	CAPN10	Genome-wide linkage scan and conditioning analysis in Mexican Americans. Followed by association studies. Detected haplotypes associated with disease. Same haplotypes associated in separate cohort of South Indians.	Some variants shown to have functional effects. Association of the genetic variation with the evidence for linkage. Increased risk for compound heterozygotes.	Horikawa et al. 2000; Cox 2001; Cassell et al. 2002
Alzheimers disease	19q	APOE	Genome-wide linkage scan in families with late onset disease. Identification of association of APOE4 allele and disease. Confirmation in association studies.	Gene dosage effect. Demonstrations of a role for APOE in facilitation of amyloid plaque formation in a isoform specific and quantitative manner	Pericak-Vance et al. 1991; Corder et al. 1993; Bales et al 1999
Crohns disease	16		followed by positional cloning and candidate gene analysis. Identification of at least 3 common variants and many rare variants associated with disease. Subsequent confirmation of linkage. Confirmation of some associations. Conditioning analysis reveals interaction with locus on chromosome 5q.	Potential genotype/phenotype correlations. Potential functional effects of variants Gene is strong biological candidate.	rugot et al. 2001; Ogura et al. 2001; Lesage et al. 2002; Murillo et al. 2002; Mirza et al. 2003

Table 1.8: A selection of genes identified as conferring susceptibility to complex traits following genome-wide linkage scans.

NIDDM1, located on chromosome 2. Conditioning analysis was then undertaken to look for interaction between this locus and a number of other loci where nominal evidence for linkage had been observed (Cox et al. 1999). Significant correlation was observed between the NPL scores for NIDDM1 and the NPL scores for CYP19 (a region on chromosome 15 showing nominal evidence for linkage to type II diabetes) with an increase in LOD for the CYP19 region from 1.3 to 4.1 after conditioning analysis. Similar increases were observed in the LOD score for NIDDM1 when reanalysed after weighting for CYP19. Simulations to estimate significance assigned an empirical p-value to these results of 0.0001 (uncorrected). Evidence for the involvement of this second locus (CYP19) in susceptibility to type II diabetes was further supported by genome-wide analysis in a second, independent sample in which the only two replicated regions of linkage were CYP19 and NIDDM1 (Cox et al. 1999).

1.10 Gene finding in primary OA

Attempts to localise the genes involved in OA disease susceptibility may be roughly divided into three categories; linkage analysis of candidate genes, genome-wide linkage scans and association analysis of candidate genes.

1.10.1 Linkage analysis of candidate genes

Linkage analysis has been used in at least two studies to investigate candidate genes for OA. The first study, by Loughlin et al (1994), found no evidence of linkage to OA for three candidate genes, α 1 chain type II collagen (COL2A1), cartilage link protein-1

(CRTL1) and matrillin-1 (MATN1). In the second study, Mustafa et al (2000) looked at eleven potential candidate genes. They found no evidence of linkage for nine of the genes with suggestive evidence for linkage of COL9A1 and HLA/COL11A2 with p-values of between 0.002 and 0.05 in the female families and in those female families concordant for hip OA. Subsequent investigations however, have found no evidence of association for the COL9A1 gene and OA although there is some evidence for haploinsufficiency of COL9A1 in OA patients (J. Loughlin, personal communication).

1.10.2 Genome-wide scans

There have been a number of both genome-wide and more localised linkage scans undertaken for OA. Many studies have used affected sibling pair (ASP) methods, which can be used without assumptions of mode of inheritance of the disease and in the absence of information from parents who are rarely available in late-onset diseases such as OA. A minority of studies have focussed on large pedigrees where a primary OA phenotype appears to be segregating as a Mendelian trait. The studies show much variability in design, with differences in such variables as the ethnic group studied, density and type of markers used and disease phenotype definitions. Some of the most noticeable differences in study design are those of patient ascertainment criteria and the form of OA investigated. Some studies have used joint replacement, and therefore severe, end-stage disease, as a requirement for patient selection whilst others have selected patients on the basis of radiographic or clinical diagnosis. These differences reflect the lack of an absolute clinical, radiological or pathological standard against which the epidemiology of OA can be tested (Spector et al. 1993). Most of the studies have focussed on OA at a

particular joint group, i.e. hip, knee or hand, or on generalised OA (involving at least three joint groups). Stratification of disease according to joint group affected is considered sensible due to the extensive evidence from epidemiological and genetic studies for the nature of OA disease susceptibility differing between different joint sites. Similar differences in disease susceptibility are also thought to exist between the sexes and therefore a number of investigators have also chosen to stratify patient groups according to sex. Table 1.10a summarises the results of the linkage studies undertaken to date that have found some evidence for linkage to OA. Additional, finer linkage mapping, of the loci identified by the Oxford group (Chapman et al. 1999; Loughlin et al. 1999) and the Icelandic group investigating hand OA (Stefansson et al. 2003) has also been undertaken using higher densities of markers and/or additional families. Results of these finer mapping studies are shown in table 1.10b.

In addition to the studies listed in table 1.10a which have found evidence for linkage to OA there have been a number of other studies which have failed to replicate evidence for linkage. A subsequent study by the group from Nottingham (Gillaspy et al. 2002), with 410 families from the Nottingham area with generalised OA, failed to replicate the linkage to chromosome 2q observed by Wright et al (1996). This led Gillaspy et al (2002) to conclude that the original study may represent a false positive result due to the small sample size of this study and the possibility of bias from outlying families. Stankovich et al (2003) also failed to find any evidence for linkage on chromosome 2q in 68 Tasmanian families with familial hand OA. Kalichman et al (2003) investigated the region on chromosome 11q, identified by Demissie et al (2002) as showing evidence for linkage to hand OA, in 295 families from Chuvastia concordant for hand OA. They found no

Study Type ¹	Centre	Sample size	Phenotype/ Sex	Chromosome	Evidence ⁴	Reference
ASP, results for chromosome 2 only reported	Nottingham, UK	44 families	Generalised nodal OA	2q23-q35	LOD > 0.6	Wright et al.1996 ⁵
ASP, genome-wide	Oxford, UK	481 families	Hip OA	2q12-q31	MLS = 2.2	Loughlin et al. 1999
			Female Hip OA	4q12-q21.2	MLS = 3.9	Loughlin et al. 1999
			Hip OA	6p23-q22.1	MLS = 2.9	Loughlin et al. 1999
			Female Hip OA	11q12-q13	MLS = 3.0	Chapman et al. 1999
			Female Hip OA	16p13.1-q12.1	MLS = 2.1	Loughlin et al. 1999
			Female OA	16q21-q23	MLS = 2.0	Loughlin et al. 1999
ASP, genome-wide	Finland	27 families	DIP ² joint OA	2q12-q21	P = 0.0001	Leppavuori et al. 1999
				4q26-q27	P = 0.0001	
				7p15-p21	P = 0.001	
				Xcen	Not available	
ASP, genome-wide	Iceland	329 families	CMC1 ³ joint OA	2p23.3	LOD = 2.2	Stefansson et al. 2003
			DIP ² joint OA	3p13	LOD = 2.2	
			Hand OA	4q32.2	LOD = 2.6	
			DIP ² joint OA	4q32.1	LOD = 2.6	
Pedigree, genome-wide	Netherlands	Four generation Dutch pedigree	Knees, hands and spine OA	2q24.3-q32.3	LOD > 2.5	Meulenbelt et al. 1999
Pedigree, genome-wide	Iceland	Affected female and 17 children	Severe hip OA	16p13.13-p12.1	LOD 2.6	Ingvarsson et al. 2001
Parent-offspring, genome-wide	Framingham Study, USA	296 two generational families	Hand OA	1p32.1-p22.1	MLS = 3.0	Demissie et al. 2002
				2p24.2-p21	MLS = 2.2	
				7p14.1-p12.3	MLS = 2.3	
				9q21.13-21.33	MLS = 2.3	
				11q13.2-q14.2	MLS = 1.6	
				12q24.33	MLS = 1.7	
				13q14.11-q14.3	MLS = 1.6	
				19q12-q13.33	MLS = 1.8	

¹ ASP = affected sibling pair

² DIP = distal interphalangeal joint

³ CMC1 = first carpometacarpal joint

⁴ LOD = logarithm of the odds score, MLS = multipoint LOD score, p-values calculated using the B statistic

⁵ See text

Table 1.10a: Summary of linkage scans undertaken for OA. Column one indicates study type, with most using either affected sibling pairs (ASP) or pedigrees in which OA appeared to segregate as a Mendelian trait. The Framingham study (Demissie et al. 2002) used a cohort of parents and offspring followed longitudinally. Column 2 shows the country in which the study was co-ordinated. Column 3 details the number of families used in the study. Column 4 indicates the phenotype and if appropriate sex of individuals showing evidence for linkage. Column 5 designates the chromosome, and where available the region of the chromosome, showing linkage to OA. Column 6 indicates the strength and type of evidence for linkage. Primary references are shown in column 7.

Chromosome	Evidence ¹	Phenotype/ Sex	Reference
Oxford			
2q24.3-q31.1	MLS = 1.6	Hip	Loughlin et al. 2002b
6p12.3-q13	MLS = 4.0	Female hip	Loughlin et al. 2002a
11q13.4-q14.3	MLS = 2.4 and 1.8 ²	Female hip	Chapman et al. 2002
Iceland			
2p24.1	LOD = 5	CMC1 ³	Stefansson et al. 2003
4q32.1	LOD = 3.3	DIP ⁴	
3p13	LOD = 1.8	DIP ⁴	

¹ LOD = Logarithm of the Odds Score, MLS -= multipoint LOD score

²This chromosome appears to contain 2 closely linked susceptibility loci

³First carpometacarpal joint

⁴Distal interphalangeal joint

Table 1.10b: Summary of the finer mapped loci from the Oxford and Icelandic studies.

evidence for a quantitative trait locus in this region using model-based linkage analysis and only a possible linkage between microsatellite markers on 11q12-q13 and hand OA using the transmission/disequilibrium (TDT) test.

Figure 1.10 illustrates the chromosomal regions identified as showing evidence for linkage with OA and demonstrates the regions of overlap. Although there are a small number of regions, on chromosomes 2p, 2q, 4q, 11q and 16p that show overlap between the studies, many regions have been identified by only a single study. Although such failures to replicate positive results may reflect the original result being a false positive, there are several other possible explanations. Genetic heterogeneity, epistasis, the presence of multiple disease alleles at a given locus and the interaction of environmental and genetic factors may all act to reduce concordance between studies (Duerr et al. 2000). This may be even more apparent when different ethnic populations are considered. In addition, initial positive reports are likely to represent an overestimation of the true effect of the locus as a result of random fluctuations while subsequent studies will often regress towards the true value and therefore be less significant (Lander and Kruglyak 1995).

1.10.3 Association studies of candidate genes

A large number of conventional case-control association studies have been undertaken to investigate the role of potential candidate genes, selected on the basis of a biologically plausible role in the OA disease process rather than on any linkage evidence suggesting that the gene resides within an OA locus (for a summary see Loughlin 2001). Such studies have mainly focussed on genes that are either mutated in the

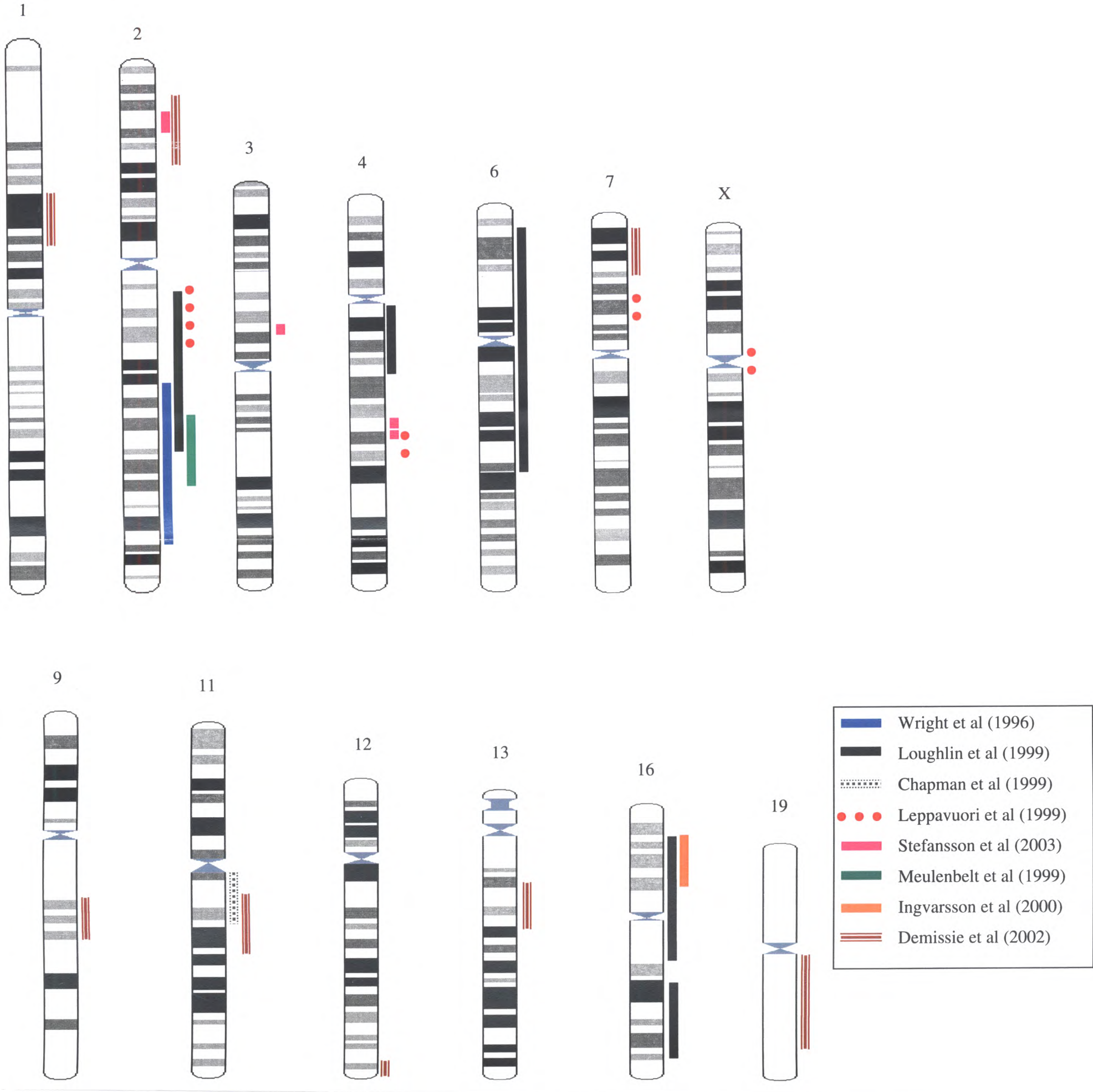


Figure 1.10: Chromosomal regions showing evidence for linkage to OA in linkage scans. Coloured lines depict the regions identified by individual studies.

osteochondrodysplasias (rare diseases of the musculoskeletal system that often have secondary OA as one of their phenotypic components) or on genes which are involved in the regulation of bone density. This approach has been largely unsuccessful with many studies yielding negative results or, failing to replicate the positive findings of earlier studies (Newman and Wallis 2002; Loughlin 2003).

1.11 Background and Aims

1.11.1 Background

A two stage genome-wide linkage scan was previously carried out by our group in a cohort of 481 OA families containing at least one pair of siblings who had undergone total hip (THR), total knee (TKR) or both total hip and total knee joint replacement for primary OA (Chapman et al. 1999). In stage one linkage analysis using a sparse map of microsatellite markers (average intermarker distance of approximately 15 cM) was carried out in 297 of the families. Any marker showing evidence for linkage in this stage was then genotyped in the 184 stage two families in order to look for further evidence of linkage.

Analysis of all data, from stages one and two, revealed two loci, one on chromosome 2q31-q32.1 and one on chromosome 11q12-q13 as showing evidence for linkage to OA in all 481 families. As many epidemiological and genetic studies have suggested that the genetic contribution to OA may differ both between the sexes and between different joint sites, the results were then stratified according to sex and joint replaced. Stratification of

the linkage data for chromosome 11 revealed that the locus on chromosome 11 was accounted for largely by that subset of families with affected female sibling pairs concordant for hip OA with a multipoint LOD score (MLS) of 3.0 (Chapman et al. 1999). Stratification of the chromosome 2 results showed that families with sibling pairs concordant for THR were the major contributor to linkage on this chromosome with a MLS of 2.2 (Loughlin et al. 2000a). Stratification of the stage one data (297 families) revealed a further four loci. In the stratum of families containing sibling pairs concordant for hip OA there was one locus, on chromosome 6p23-q22.1 with a MLS of 2.9. In the stratum of families containing affected female siblings concordant for hip OA there were two loci, on chromosomes 4q12-q21.2 and 16p13.1-q12.1 with MLS of 3.9 and 2.1 respectively. A second chromosome 16 locus at 16q21-q23 was also observed in the stratum of families containing female sibling pairs with a MLS of 2.0 (Chapman et al. 1999; Loughlin et al. 1999; Loughlin et al. 2000a).

Figure 1.11 shows the results of the multipoint linkage analysis performed by our group for chromosomes 4 and 16.

Other groups have also reported suggestive linkage of OA both to chromosome 4 (Leppavouri et al. 1999; Roby et al. 1999), where regions identified are telomeric to our region, and to chromosome 16 (in a large Icelandic pedigree with severe primary hip OA defined by joint replacement) in a region superimposed on our chromosome 16p region (Ingvarsson et al. 2001).

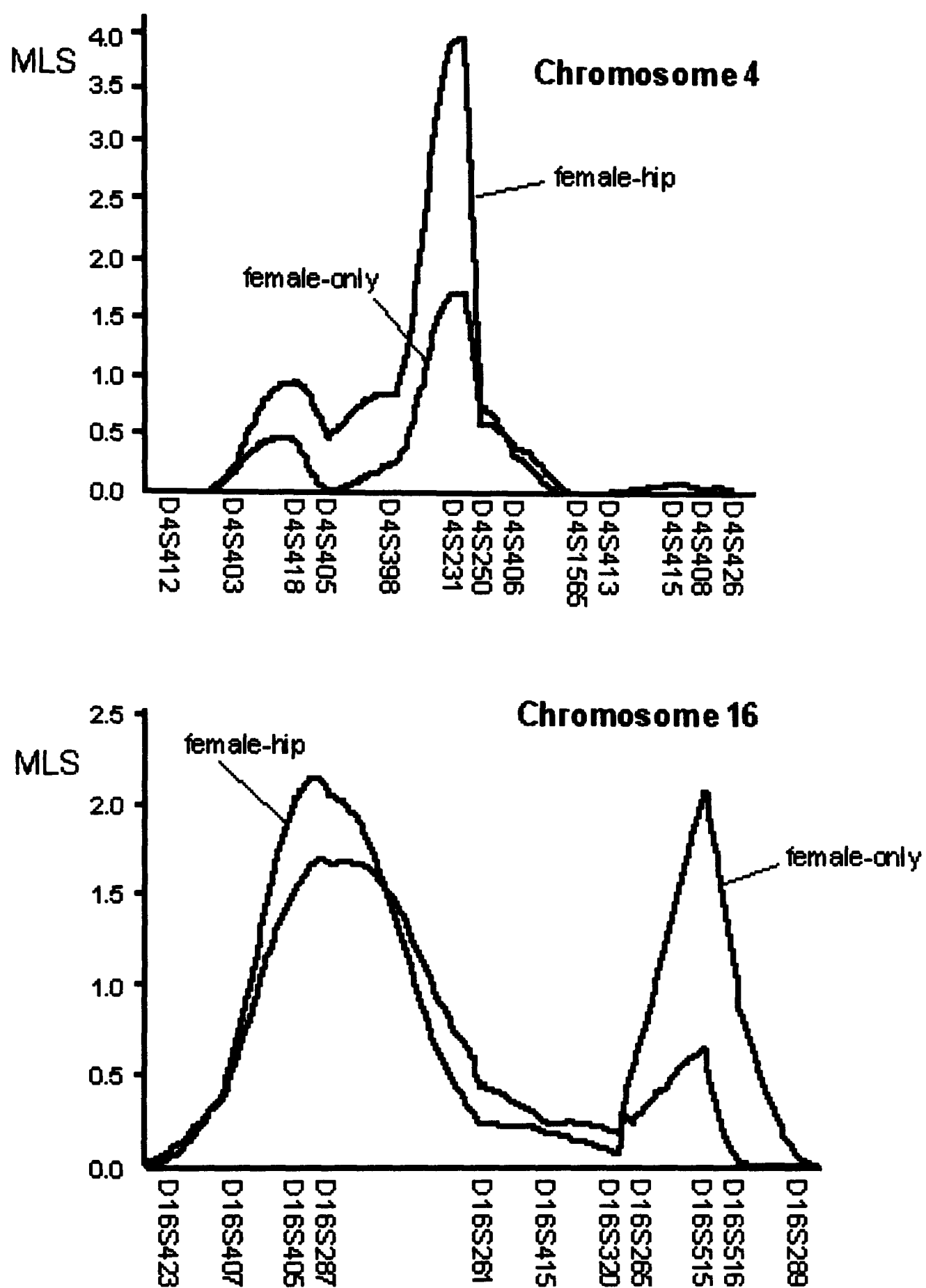


Figure 1.11: Multipoint linkage analysis of chromosomes 4 and 16

1.11.2 Aims

Analysis of chromosomes 4 and 16 in the genome-wide linkage scan used markers at a relatively low density of approximately one marker every 15 cM. In addition stage one of the genome scan used only 297 families of the 571 currently available to this group. Furthermore, minor errors in the maps used in the original scan have become apparent following the completion of the sequencing of the human genome, with a number of the markers genotyped now considered to have unreliable map positions. All of these factors may have limited both the power of the original study and the potential to detect and refine linkage of the initial scan.

The initial aim of this project therefore, was to linkage scan the whole of chromosomes 4 and 16 using all 571 available families with a more dense panel of microsatellite markers ordered as per the most recent, reliable genetic maps. The reasoning behind this is three-fold. Firstly, increasing the number of families will increase the power of the study to detect any regions of linkage with OA and may allow detection of other regions of linkage on these chromosomes. This will allow us to determine if our cohort shows evidence of linkage to the regions identified by Roby et al (1999) and Leppavuori et al (1999) on chromosome 4. Secondly, the increase in families and markers will allow us to ascertain if the loci identified are restricted to the strata in which they were originally detected. Finally, using a higher density of markers should allow us to narrow the regions identified.

Following linkage analysis and narrowing of the regions work will be undertaken to identify candidate genes within the regions. Single nucleotide polymorphisms (SNPs)

from within these genes will then be identified, genotyped in case-control cohorts and tested for association with OA.

Chapter 2 - Materials and Methods

2.1 Families

A total of 571 families, containing at least one pair of siblings whom had undergone total hip (THR) or total knee (TKR) or total knee and total hip joint replacement for primary OA, were recruited for linkage analysis. All families recruited were of UK Caucasian origin with collection being co-ordinated through the Nuffield Orthopaedic Centre in Oxford. All patients were reviewed and examined by trained clinical research fellows and clinical research nurses and primary OA status confirmed. Clinical, radiographic, operative and histological findings were used to support the primary diagnosis. Subjects who had undergone joint replacement secondary to any other disorder, such as fracture or rheumatoid arthritis were excluded. Since OA is a late-onset disorder only three families had parents available for participation. Therefore, in order to determine identity by descent (IBD) transmission additional siblings who had not undergone knee or hip joint replacement were also recruited. The 571 families comprise a total of 1239 affected individuals (729 female and 510 males) and 320 additional siblings (165 female and 155 male). The average age of patients at time of first joint replacement was 66 years (age range 40 – 86 years). Table 2.1a gives a summary of the families. Of the 939 individuals who had undergone only THR, 32 % had undergone right THR only, 23 % left THR only and 45 % bilateral THR. Of the 227 individuals who had undergone TKR 30 % had undergone right TKR only, 23 % left TKR only and 47 % bilateral TKR.

For finer mapping, only the subset of 218 families with affected female sibling pairs were used. Table 2.1b lists the family structure of this cohort of 218 female families. The 218 families were composed of 461 affected female subjects and 156 additional siblings. The mean age of the subjects at the time of joint replacement was 66 ± 8.5

Total no. of families			571
Stratified	With affected	Sibpairs	498
		Trios	58
		Quartets	11
		Other*	4
	Total number of female	Families	218
		Pairs	197
		Trios	21
	Total number of male	Families	117
		Pairs	108
		Trios	8
		Other*	1
	Total number of THR	Families	378
		Pairs	328
		Trios	38
		Quartets	6
		Other*	6
	Total number of TKR	Families	60
		Pairs	57
		Trios	3
	Total number of female THR	Families	146
		Pairs	136
		Trios	10
	Total number of female TKR	Families	26
		Pairs	25
		Trios	1
	Total number of male THR	Families	83
		Pairs	77
		Trios	5
		Other*	1
	Total number of male TKR	Families	11
		Pairs	11
Total number of affected individuals	Females		729
	Males		510
Stratified	Number of THR		939
	Number of TKR		227
	Number of THR and TKR		73
	Number of female THR		535
	Number of female TKR		146
	Number of female THR & TKR		48
	Number of male THR		404
	Number of male TKR		81
	Number of male THR & TKR		25
Number of additional siblings	Females		165
	Males		155
Total Individuals			1559

*Pairs of relatives such as cousins, uncles and aunts

Table 2.1a: Family Structures for all 571 families. THR = total hip joint replacement, TKR = total knee joint replacement

Family structure	Total
Total no. of female families	218
With affected sibling pairs	195
With affected sibling trios	20
Other*	3
Total no. of female affected individuals	461
No. of additional siblings	156
Total no. of female-THR families	146
With affected sibling pairs	136
With affected sibling trios	10
Total no. of female-THR affected individuals	302
No. of additional siblings	126
Total no. of female-TKR families	25
With affected sibling pairs	24
With affected sibling trios	1
Total no. of female-TKR affected individuals	51
No. of additional siblings	12
Total no. of female-THR and -TKR families	47
With affected sibling pairs	44
With affected sibling trios	3
Total no. of female-THR and -TKR affected individuals	97
No. of additional siblings	20

* An affected sibling pair plus an affected aunt, an affected sibling trio plus an affected cousin, and an affected sibling pair plus two affected nieces.

Table 2.1b. Family structures of the female siblings affected with osteoarthritis.

years. Three hundred and forty three of the subjects had undergone THR (uni- or bi-lateral) only, 93 had undergone TKR (uni- or bi-lateral) only and 25 both TKR and THR.

For association studies collection of a case-control cohort consisting of female-THR patients and controls was co-ordinated through the Nuffield Orthopaedic Centre in Oxford. All cases had undergone THR because of primary OA. Patients who had undergone THR secondary to any other disorder, such as fracture or rheumatoid arthritis were excluded. A total of 390 cases were collected for the ADAMTS3 candidate gene association study with an additional 66 being collected before the IL4R candidate gene association study was undertaken. A total of 399 age-matched female controls consisting of individuals showing no overt clinical signs of OA were also collected. All cases and controls were UK Caucasians aged 55 years or over.

For each subject, 25 ml of venous blood was collected and DNA was extracted by conventional techniques. Ethical approval for the study was obtained from the Central Oxford Research Ethics Committee and informed consent was obtained from all subjects.

2.2 Amplification of DNA by the polymerase chain reaction (PCR)

PCR conditions were optimised for each reaction by varying magnesium chloride (MgCl_2) concentration and annealing temperature, and DNA was then amplified in 15 μl multiplex or singleplex PCR reactions. PCR reaction mix contained 1.5 μl GeneAmp® 10X reaction PCR buffer II (Applied Biosystems), 0.375 μl dNTPs (at 8mM) [Sigma], 0.15 - 0.60 μl of each primer pair (at 5 μM) [Applied Biosystems],

0.08 µl of Amplitaq Gold DNA polymerase (at 5U/µl) [Applied Biosystems], MgCl₂ (Perkin Elmer) to a final concentration of 1-3 mM depending on the primers used, and 5 µl DNA (at approximately 10ng/µl). 5% dimethyl sulfoxide (DMSO) was added to facilitate DNA strand separation in GC rich regions when amplification of DNA using standard conditions was inefficient.

All PCR cycling conditions consisted of denaturing at 95 °C for 15 minutes followed by 35 cycles of 94 °C for 45 seconds, 60 seconds at the appropriate annealing temperature and 45 seconds at 72 °C. Full details of reaction conditions for each primer pair are given in Appendices 1, 4 and 5.

2.3 Linkage Analysis

2.3.1 Microsatellite Markers

Linkage analysis of chromosomes 4 and 16 was carried out using fluorescently labelled polymorphic microsatellite markers. Oligonucleotide primers for each marker were purchased from Applied Biosystems (ABI) Prism Linkage Mapping Sets (LMS) MD10 and HD5 (Version 2/Version 2.5, PE Applied Biosystems) or were custom made using primer sequences from the Genome Database (<http://www.gdb.org>) with the forward primer fluorescently labelled. Details of fluorescent label and primer sequence, where available, are given in Appendix 1.

Markers were then amplified in 15 µl multiplex or singleplex PCR reactions as detailed in section 2.2. Full details of reaction conditions for each marker are given in Appendix 1.

For both chromosomes 4 and 16, initial mapping was carried out with a coarse panel of markers from the LMS MD10 sets to crudely define the region of linkage. Additional markers, using both the LMS HD5 sets and custom made primers, were then added in the regions of interest in order to further refine the regions.

2.3.2 Separation of PCR products

All markers were arranged in panels of up to eleven markers. For each panel the PCR products were pooled and 2.5 µl of the pool added to a 3.5 µl mix containing 0.4 µl Genescan®-500[Rox]TM size standard (Applied Biosystems), 0.4 µl loading dye (Applied Biosystems), 1.7 µl deionised formamide (Sigma) and 1 µl sterile water. The samples were denatured at 96 °C for between 6 and 8 minutes and then snap cooled on ice. The denatured samples were loaded on a 6% polyacrylamide gel (25 ml Long Ranger 5%, 6M Urea premix gel [Biowhittaker Molecular Applications], 16.5 µl TEMED [N N N', N'-Tetramethylethylenediamine] [Sigma] and 125 µl 10% ammonium persulphate [Sigma]). The PCR products were then separated by electrophoresis using an ABI Prism 377 DNA sequencer at 3000 V (60.0 mA, 200 Watts) for between 120 and 150 minutes.

2.3.3 Microsatellite genotyping and linkage analysis

The alleles of the microsatellite markers were sized using Genescan and Genotyper software (Applied Biosystems) as described previously (Chapman et al. 1999). Genotyping and Mendelian inconsistencies were monitored using the RECODE v1.4 and PEDCHECK (O'Connell and Weeks 1998) programs. Microsatellite allele frequencies were calculated from the input data using the Genetic Analysis System

(GAS) program (<http://users.ox.ac.uk/~ayoung/gas.html>) and two-point linkage analysis was performed using ANALYZE (<http://hpcio.cit.nih.gov>). Markers were ordered according to the DeCODE genetic map (Kong et al. 2002) and by reference to the Marshfield Map (http://research.marshfieldclinic.org/genetics/Map_Markers/maps/IndexMapFrames.html). Non-parametric multipoint linkage analysis was carried out using ASPEX (<http://hpcio.cit.nih.gov/Lserver.htm>). Any siblings additional to the affected sibling pairs were classified as of “unknown” disease status in the linkage analysis.

2.3.3.1 One-LOD support interval

The 1-LOD support interval was defined as the region corresponding to the peak LOD score minus one and is the region expected to contain the responsible gene in 80 - 90% of cases (Kruglyak and Lander 1995).

2.3.4 Stratification

Initial analyses were carried out with the 571 families as a whole and following stratification for sex, joint replaced and sex plus joint replaced.

For families which contain more than two affected siblings of different sexes the sibling(s) of opposite sex to the sibling pair was given the status “unknown” in the stratified linkage analysis.

For joint replacement a TKR family consists of a family with sibling pairs concordant for TKR, whilst a female-THR family contains a pair of female siblings concordant

for THR. Any other affected individuals within the family of a different sex or with a different joint replaced were designated as having an “unknown” status in the linkage analysis. We did not stratify the TKR families according to sex because of the small size of this stratum (only 60 families).

Subsequent analyses were carried out in the subset of 218 families with affected female sibling pairs (female families) and the 146 female-THR families.

2.4 Conditioning analysis

2.4.1 Determination of evidence for a statistical interaction between regions

The family non-parametric LOD (NPL) scores at the regions showing maximum evidence for linkage on chromosomes 4q, 16p and 16q were determined using GENEHUNTER PLUS (Kruglyak et al. 1996; Kong and Cox 1997). Correlation between the NPL scores was determined using the MedCalc[®] version 5.00.019 (copyright© 1993-1999 Frank Schoonians) program to calculate the Pearson product-moment correlation coefficient, p-values and 95% confidence intervals. The correlation coefficient, r , for two variables X and Y is calculated using the following formula:

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N}) (\sum Y^2 - \frac{(\sum Y)^2}{N})}}$$

Where N = the total number of observations

r is an appropriate statistic to use in this situation since NPL scores have a normal distribution with a mean of 0 and a variance of 1.

r can take any value between -1 and +1 with values of -1 and +1 indicating that the two variables X and Y lie on a straight line. A r value of 0 indicates that there is no linear relationship between the two variables. r is positive if higher variables of X are associated with higher variables of Y and negative if higher values of one variable are associated with lower values of the other variable.

2.4.2 Conditioning analysis

Conditioning analysis was performed only for regions which showed significant evidence for correlation. By performing the less labour intensive correlation calculations first we were able to determine which combinations of regions to focus on for conditioning analysis. Conditioning analysis has the advantage over correlation analysis of allowing for observation of any increase in LOD score and narrowing of the linkage region subsequent to the analysis.

2.4.2.1 Modelling for a positive interaction

Families were assigned a weight according to their evidence for linkage at the first “conditioning” locus. Families with a NPL score > 0 were assigned a weight of 1 whilst families with a NPL score ≤ 0 were assigned a weight of 0. Linkage analysis was then performed using Allegro (Gudbjartsson et al. 2000) for the second “conditioned” locus using these weights.

One thousand permutations were undertaken to estimate the empirical significance of any increase in LOD score.

2.4.2.2 Modelling for a negative interaction

Families were again assigned a weight according to their evidence for linkage at the first “conditioning” locus. Families with a NPL score > 1 were assigned a weight of 0 whilst families with a NPL score ≤ 0 were assigned a weight of 1. Linkage analysis was then performed using Allegro (Gudbjartsson et al. 2000) for the second “conditioned” locus using these weights.

2.5 Candidate gene identification

Database searching, using National Center for Biotechnology Information (NCBI) [<http://www.ncbi.nlm.nih.gov/genome/guide/human>], Ensembl (<http://www.ensembl.org/>), and University of California, Santa Cruz (UCSC) [<http://genome.ucsc.edu/>] was undertaken to catalogue all reported genes within the identified regions. Where possible functions and expression patterns were identified for each gene using the above databases plus information from Online Mendelian Inheritance in Man (OMIM) [<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>], PubMed (<http://ncbi.nlm.nih.gov/entrez/query.fcgi>) and LocusLink (<http://www.ncbi.nlm.nih.gov/LocusLink/>). Potential candidates were selected on the basis of a reported or suggested role in the normal synthesis, development, maintenance or degradation of any tissue within the joint organ (especially cartilage or bone) and of known expression in joint tissues.

2.6 DNA Sequencing

2.6.1 PCR amplification

All reactions were prepared using the ABI Prism® BigDye™ Terminator Cycle Sequencing Ready Reaction kit (version 3) according to the manufacturers instructions (Applied Biosystems). In summary, the DNA to be sequenced was amplified using sequence specific unlabelled primers, with M13 tags (sequences - 21M13 (FOR): TGT AAA ACG ACG GCC AGT, M13 REV (REV): CAG GAA ACA GCT ATG ACC) at the 5' end, in 15 µl PCR reactions as detailed in section 2.2. Full details of primer sequences and reaction conditions are given in Appendix 4.

Prior to sequencing 8 µl PCR product was treated with 1 µl each of Shrimp Alkaline Phosphatase (SAP) [Amersham] and Exonuclease I (EXO1) [Amersham] and incubated at 37 °C for 15 minutes. Exonuclease I degrades any remaining single-stranded template and primers whilst SAP degrades any unused dNTPs. The enzymes were then deactivated by heating to 80 °C for 15 minutes.

A 5 µl reaction containing 2.7 µl of SAP/EXO1 treated PCR product, 2 µl of Ready Dye mix (Applied Biosystems) and 0.3 µl of either the M13 forward, M13 reverse or a specific internal primer (all at 10 µM) was used for sequencing reactions. Reaction conditions were 25 cycles of 96 °C for 10 seconds, 50 °C for 5 seconds and 60 °C for 4 minutes. Products were then kept at 4 °C until the precipitation stage.

2.6.2 Precipitation

Sequencing reactions were precipitated using an ethanol method based on the manufacturers protocol (Applied Biosystems). Briefly, the 5 µl of sequencing reaction was added to 15 µl sterile distilled water. Products were precipitated using 80 µl of ethanol solution containing 3 µl 3M sodium acetate (pH 4.6) [Sigma], 62.5 µl 95% ethanol (Merck Eurolab), 14.4 µl sterile water and 0.4 µl of Pellet Paint® NF Co-Precipitant (Merck Eurolab). After being left for a minimum of ten hours and a maximum of eighteen hours the DNA was pelleted by centrifuging at 3750 rpm (1600 xg) for 45 minutes in a Centra CL3 bench top centrifuge (Thermo Electron Corporation). The pellet was then washed using 70 % ethanol and centrifuged for a further 10 minutes at 3750 rpm (1600 xg). The DNA pellet was dried by inverted spinning at 500 rpm (30 xg) for 5 seconds and then sealed and stored at -20 °C.

2.6.3 Separation of products

Prior to loading on the gel, pellets were resuspended in 2 µl of a dye/formamide mix (containing loading dye [Applied Biosystems] and deionised formamide [Sigma] at a 5:1 ratio), denatured at 95 °C for 2 minutes and snap cooled on ice.

The samples were loaded on to a 4.25% polyacrylamide gel (30 ml 4.25%, 6M sequencing mix [Severn Biotech], 19 µl TEMED [N N N', N'-Tetramethylethylenediamine] [Sigma] and 150 µl 10% ammonium persulphate [Sigma]). The products were then separated by electrophoresis using an ABI Prism 377 DNA sequencer at 3000V (60.0 mA, 200 Watts) for 225 minutes.

2.6.4 Sequence analysis

PE Applied Biosystems Sequence Analysis 3.3 and Sequence Navigator programs were used to perform sequence analysis.

2.7 Association studies

2.7.1 Variant typing

Single nucleotide polymorphisms (SNPs) were typed by PCR-restriction enzyme analysis using either natural or forced restriction sites that were created or abolished by the SNP of interest. For each SNP 5 μ l DNA (approximately 10ng/ μ l) was amplified in a 15 μ l PCR reaction as detailed in section 2.2. Details of primer sequences and PCR conditions can be found in Appendix 5.

Five μ l digestion mix consisting of 1 μ l of a 10X stock of the manufacturer supplied buffer (New England Biolabs) and between 2 and 10 units of the appropriate restriction enzyme (New England Biolabs) were added to 5 μ l PCR product and incubated for 3 hours at the manufacturer recommended digestion temperature. Where applicable the enzyme was subsequently denatured at the manufacturer recommended temperature for 20 minutes. Details of restriction enzymes, buffer, digestion and inactivation temperatures are listed in Appendix 6.

Three μ l loading dye (0.5mg/ml bromophenol blue [Sigma]) was added to the digestion mix and digested products separated by electrophoresis through a 3 % agarose gel made from 3 g of agarose (Eurogentec) and 100 ml of 1X TBE (10X stock of TBE [Tris Borate EDTA Buffer] 1.0 mM Tris, 0.9 M Boric Acid, 0.01 M EDTA,

pH @ 25°C = 8.4 ±0.10] [Gibco]) stained with 1 µl ethidium bromide (10mg/ml) [Sigma]. Products were electrophoresed for 1 hour at 110 V and then visualised over an UV imager and a photographic image of the gel taken.

Genotyping of individuals was done by eye with digestion being repeated for any individuals where genotype was unclear. Genotypes were confirmed by sequencing in a minimum of 40 individuals.

2.7.2 Analysis of results

Distribution of alleles between cohorts of cases and controls was compared by χ^2 using standard contingency table analysis with the MedCalc® version 5.00.019 (<http://www.medcalc.be>, copyright© 1993-1999 Frank Schoonians) program. All calculations include a Yates' continuity correction. Odds ratios and their 95% confidence intervals were also calculated using the MedCalc® program. All markers were tested for compliance with Hardy Weinberg equilibrium in the controls.

The frequencies of SNP genotypes and the combination of alleles at different loci were also compared between cases and controls by χ^2 using standard contingency table analysis. Again, all calculations include a Yates' continuity correction.

The estimate haplotype program EH-PLUS (<http://web1.iop.kcl.ac.uk/IoP/Departments/PsychMed/GepiBst/software.shtml>) [Zhao et al. 2000] was used to estimate the haplotype frequencies for combinations of SNPs. The differences in haplotype frequencies between cases and controls was then compared by χ^2 using

standard contingency table analysis. ODDs ratios and 95% confidence intervals were also calculated using the MedCalc® program.

2.7.3 Linkage disequilibrium studies

The pair-wise linkage disequilibrium (LD) coefficients D' and r^2 were calculated using the GOLD (graphical overview of linkage disequilibrium) program (Abecasis and Cookson 2000) for all SNP pairs. The GOLD program was also used to display the pattern of pair-wise LD graphically on a physical map.

2.7.3.1 Measures of LD

Many different statistics have been proposed to measure LD, each having different strengths depending on the context considered. Two of the most commonly reported measures are D' and r^2 . For both measures it is first necessary to calculate the value D which is defined as, $D = P_{AB} - P_A \times P_B$, where P_{AB} = observed frequency of the haplotype consisting of alleles A and B, P_A = frequency of allele A at the first locus and P_B = frequency of allele B at the second locus. D' is obtained by dividing D by its maximum possible value at the observed allele frequencies. A D' of 1 indicates that 2 SNPs have not been separated by recombination (or recurrent mutation or gene conversion) during the history of the sample. Values of $D' < 1$ indicate complete ancestral LD has been disrupted but the relative magnitude of values has no clear interpretation. In addition, estimates of D' are strongly inflated in small samples. Therefore, statistically significant values of D' at or near 1 are a useful indicator of minimal historical recombination but smaller values have limited use as a measure of the extent of LD. Because of these limitations, r^2 has emerged as the measure of

choice in a number of studies. It is calculated by dividing D^2 by the product of the four allele frequencies at the two markers under consideration. Unlike for D' , r^2 is equal to 1 if and only if the two markers have the same allele frequency and have not been separated by recombination. Intermediate values of r^2 are easy to interpret as its value represents the statistical correlation between the two loci. It has the additional advantage of there being an inverse relationship between the sample size required to detect association and the value of r^2 (Pritchard and Przeworski 2001; Ardlie et al. 2002; Nordberg and Tavaré 2002).

2.7.4 Investigation of the relationship between genetic variation and the evidence for linkage.

The relationship between genetic variation and the evidence for linkage was determined using the method recommended by Cox (2002). In brief, the 146 female-THR families used in the finer linkage mapping analysis were divided in to two groups. The first group contained families where the proband had at least one copy of the associated allele for the SNP under consideration and the second group those families where the proband was homozygous for the non-associated allele. Multipoint non-parametric linkage analysis was then undertaken for the two groups separately using the ASPEX program (<http://hpcio.cit.nih.gov/Lserver.htm>). The linkage results were examined for a disproportionate evidence for linkage in those families where the proband has at least one copy of the associated allele.

2.8 Expression studies

2.8.1 RNA extraction from cartilage

The OA genetics group has extracted RNA from the articular cartilage of knee and hip joints collected following joint replacement for OA. In summary, a scalpel blade was used to remove the cartilage from the joints within two hours of the joint replacement operation. Cartilage was snap-frozen in liquid nitrogen and ground to a powder with a Retsch mixermill 200 (Retsch Limited, Leeds, UK). RNA extraction was then carried out as described by McKenna et al (2000), ending up with approximately 1ml of RNA in ethanol per column. DNA was also extracted from the same cartilage samples.

2.8.2 cDNA synthesis

cDNA was synthesised from cartilage total RNA, using the Invitrogen Superscript™ First-Strand Synthesis System for RT-PCR kit according to the manufacturer's instructions. Briefly, for first strand cDNA synthesis a mix containing 1 µl random hexamers [50 ng/µl] (Invitrogen), 1 µl dNTP mix [10 µM] (Invitrogen) and 3 µl diethylpyrocarbonate (DEPC)-treated water was added to 5 µl of each RNA solution. A DEPC-treated water blank and the kit control were also included. The random hexamers were then annealed at 65 °C for 5 minutes. A mix containing 2 µl 10X reaction buffer (200 mM Tris HCl, 500 mM KCl, pH 8.4), 4 µl of MgCl₂ (25mM), 2 µl of 0.1M DTT (dithiothreitol) and 1 µl RNase OUT inhibitor (all from the Invitrogen Superscript™ First-Strand Synthesis System for RT-PCR) was added to the RNA mix, mixed gently and incubated at 25 °C for 2 minutes. One µl of superscript II reverse transcriptase (Invitrogen) was added and the solution mixed. The samples were then incubated at 25 °C for 10 minutes, 42 °C for 50 minutes and 70 °C for 15

minutes to synthesise the cDNA. Following synthesis 1 µl RNase H (Invitrogen) was added to degrade the RNA template from the RNA:DNA hybrid, and the samples incubated at 37 °C for 20 minutes. The samples were then stored at 4 °C.

2.8.3 Amplification of cDNA for IL4R

Primers were designed to anneal in exons 6 and 8 of IL4R, these exons flank exon 7 which contains N142N, the SNP of interest. Primer sequences were: Forward 5' TGT GGA AGG GCT CCT TCA AG 3' and Reverse 5' AGG GTT CTA GGT AGG TCA CG 3' (MWG-Biotech). Two µl cDNA was amplified in a 28 µl PCR reaction. Reaction mix contained 2.4 µl MgCl₂ (Applied Biosystems) to a final concentration of 2mM, 3.0 µl GeneAmp® 10X PCR reaction buffer (Applied Biosystems), 0.75 µl dNTPs (at 8 mM) [Sigma], 0.60 µl of each of the forward and reverse primers (at 10 µM) [MWG-Biotech], 0.16 µl of Amplitaq Gold DNA polymerase (at 5 U/µl) [Applied Biosystems]. PCR cycling conditions consisted of denaturing at 95 °C for 15 minutes followed by 35 cycles of 94 °C for 45 seconds, 57 °C for 60 seconds and 45 seconds at 72 °C.

Five µl loading dye (0.5mg/ml bromophenol blue [Sigma]) was added to the PCR products and products separated by electrophoresis through a 2 % agarose stretch gel made from 2 g agarose (Eurogentec) and 100 ml 1X TBE (10X stock of TBE [Tris Borate EDTA Buffer] 1.0 mM Tris, 0.9 M Boric Acid, 0.01 M EDTA, pH @ 25°C = 8.4 ±0.10) [Gibco]) stained with 1 µl ethidium bromide (10mg/ml) [Sigma]. A φ 174 /Hae III DNA ladder (Promega) was also run to allow sizing of the PCR products. Products were electrophoresed for 2 hours at 110 V and then visualised over an UV imager and a photographic image of the gel taken.

Chapter 3 - Linkage mapping of chromosomes 4 & 16

3.1 Introduction

Complex diseases are characterised by multiple genetic, developmental, and environmental risk factors. They are non-Mendelian; but show familial aggregation and are generally more common than single-gene disorders.

Epidemiological, physiological and genetic studies reveal that primary OA appears to be the result of the interaction of a number of systemic (i.e. obesity, sex, genetic) and local factors (i.e. joint instability, dysplasia) which determine onset and progression of disease (Jones and Doherty 1995). A significant genetic component to OA has been demonstrated by twin-pair, sibling risk and segregation studies (reviewed by Loughlin 2001). The involvement of multiple factors in determining onset and progression of the disease has led to the description of OA as a complex, multifactorial disease. Differences have been observed, in the heritability and the factors associated with disease, both between the sexes and between different joint sites (Chitnavis et al. 1997; MacGregor and Spector 1999; Lanyon et al. 2000).

The past two decades have seen a sharp increase in attempts to map genes involved in susceptibility to complex diseases. Many groups have reported the use of genome-wide linkage scans, followed by mapping with a higher density of markers in those regions showing evidence for linkage, in order to identify narrowed regions of the genome harbouring susceptibility to complex human diseases. Such work has been undertaken for a number of diseases including type I and II diabetes, familial Alzheimer's disease, and Crohn's disease (Pericak-Vance et al. 1991; Hanis et al. 1996; Nistico et al. 1996; Hugot et al. 2001).

Our group carried out a genome-wide linkage scan on a cohort of OA families containing at least one pair of siblings who had undergone total hip (THR), total knee (TKR) or both total hip and total knee joint replacement for primary OA (Chapman et al. 1999; Loughlin et al. 1999; Loughlin et al. 2000a). Stratification of the linkage data according to sex and joint replaced revealed six loci. In the stratum of families containing sibling pairs concordant for hip OA (THR-only families) there were two loci, on chromosomes 2q and 6 with multipoint LOD scores (MLS) of 2.2 and 2.9 respectively. In the stratum of families containing affected female siblings concordant for hip OA (female-THR families) there were three loci, on chromosomes 4 (MLS = 3.9), 11 (MLS = 2.1) and 16 (MLS = 2.1). A second chromosome 16 locus, with a MLS of 2.0, was also observed in the stratum of families containing female sibling pairs with hip, knee or both hip and knee OA (female families). This genome-wide linkage scan used markers at a relatively low density of approximately one marker every 15 cM, resulting in relatively broad linkage intervals. In the original genome scan Chapman et al (1999) genotyped fourteen markers on chromosome 4 and eleven markers on chromosome 16. Following stratification by sex and joint replaced a locus was detected centred on 4q12-4q21.2 for the subset of families containing female sibling pairs concordant for hip OA (85 families). On chromosome 16 linkage analysis in this same stratum identified a locus on 16p centred on 16p13.1-16q12.1. When the subset of families with affected female sibling pairs (132 families) were analysed a second locus on chromosome 16 was identified. This locus on 16q was centred on 16q21-16q23.

Since the completion of the genome scan the collection of affected sibling pair families has increased and now stands at 571 families. In order to confirm and narrow the regions identified on chromosomes 4 & 16, linkage analysis was carried out on all 571 families with OA affected sibling pairs using a panel of microsatellite markers at an average density

of one marker every 10 cM. Regions identified as showing evidence for linkage were then further defined and narrowed using higher densities of microsatellite markers.

For all PCR reactions in this chapter, the reaction conditions, temperatures and, where available, full primer sequences are given in Appendix 1. All markers for multipoint analysis were ordered as per the deCODE map (Kong et al. 2002). Markers genotyped that are not on this map were included at positions representative of their position on the Marshfield map (http://research.marshfieldclinic.org/genetics/map_markers/maps).

3.2 Initial (MD10) linkage mapping of chromosome 4

Initial linkage mapping was undertaken using twenty microsatellite markers from the ABI Prism Linkage Mapping Set v2 (LMSv2) MD10. These markers are an average of 10 cM apart and cover the whole of chromosome 4. Two-point (table 3.2) and multipoint analysis (figures 3.2a and b) was undertaken both for all data and following stratification according to sex, joint replaced and sex plus joint replaced (for details of strata see section 2.1).

Two-point p-values at or below a nominal significance of 0.05 are highlighted in red (table 3.2). The region between D4S391 and D4S1534 contains a number of markers with two-point p-values at or approaching significance in several of the strata, most noticeably that of females and female-THRs. This is supported by a peak with a maximum LOD score of 1.2 in this region in the female families and 1.3 in the female-THR families in the multipoint analysis (figure 3.2a). The location of this peak, 84 cM from the p-telomere, corresponds with that seen in the original screen. The 1-LOD support interval for this region is large, spanning over 60 cM.

Marker	Distance from p-telomere in cM	All Families		Females		Female-THRs		Males		Male-THRs		THRs		TKRs	
		LOD	P-value ¹	LOD	P-value ¹	LOD	P-value ¹	LOD	P-value ¹	LOD	P-value ¹	LOD	P-value ¹	LOD	P-value ¹
D4S412	4.42	0.0	ns	0.0	ns	0.0	ns	0.0	ns	0.0	ns	0.0	ns	0.0	ns
D4S2935	11.85	0.1	ns	0.4	0.094	0.5	0.072	0.0	ns	0.0	ns	0.4	0.096	0.0	ns
D4S403	26.71	0.0	ns	0.0	ns	0.2	ns	0.0	ns	0.0	ns	0.0	ns	0.0	ns
D4S419	34.97	0.0	ns	0.0	ns	0.0	ns	0.0	ns	0.0	ns	0.0	ns	0.0	ns
D4S391	47.33	0.5	0.059	1.0	0.018	0.8	0.026	0.0	ns	0.0	ns	0.2	ns	0.0	ns
D4S405	60.23	0.0	ns	0.0	ns	0.1	ns	0.0	ns	0.0	ns	0.0	ns	0.0	ns
D4S1592	72.71	0.2	ns	0.2	ns	0.5	0.060	0.0	ns	0.2	ns	0.2	ns	0.0	ns
D4S392	80.38	0.0	ns	1.0	0.018	1.6	0.004	0.0	ns	0.1	ns	0.2	ns	0.0	ns
D4S2964	88.45	0.7	0.034	0.6	0.054	0.7	0.036	0.1	ns	0.0	ns	0.6	0.052	0.2	ns
D4S1534	93.19	0.1	ns	0.3	ns	0.4	0.092	0.4	0.098	0.6	0.048	0.3	ns	0.0	ns
D4S1572	107.52	0.1	ns	0.0	ns	0.0	ns	0.3	ns	0.0	ns	0.0	ns	0.7	0.041
D4S406	114.74	0.1	ns	0.0	ns	0.1	ns	0.1	ns	0.0	ns	0.0	ns	0.2	ns
D4S1575	131.59	0.4	0.085	0.1	ns	0.0	ns	0.0	ns	0.0	ns	0.0	ns	0.8	0.025
D4S424	138.87	0.3	ns	0.1	ns	0.0	ns	0.0	ns	0.0	ns	0.0	ns	1.3	0.008
D4S413	150.31	0.6	0.048	0.0	ns	0.0	ns	0.4	0.087	0.5	0.062	0.4	0.091	0.8	0.028
D4S1597	163.65	0.1	ns	0.6	0.050	0.3	ns	0.1	ns	0.0	ns	0.0	ns	0.3	ns
D4S1539	170.01	0.2	ns	0.3	ns	0.2	ns	0.3	ns	0.3	ns	0.0	ns	0.7	0.040
D4S415	173.24	0.5	0.066	0.3	ns	0.0	ns	0.3	ns	0.1	ns	0.0	ns	0.4	0.083
D4S1535	189.38	0.1	ns	0.0	ns	0.1	ns	0.5	0.066	0.4	0.091	0.0	ns	0.4	0.097
D4S426	202.69	0.0	ns	0.0	ns	0.0	ns	0.5	0.061	0.2	ns	0.0	ns	0.4	0.100

¹ ns = not significant

Table 3.2: Two-point linkage analysis of chromosome 4 for the MD10 markers. Results for all families and following stratification by sex, joint replaced and sex plus joint replaced. Only p-values ≤0.1 are shown. P-values ≤0.05 are highlighted in red. The genetic distance of markers from the p-telomere is also shown. Markers are from the ABI Prism Linkage Mapping Set MD10. THR = total hip joint replacement, TKR = total knee joint replacement

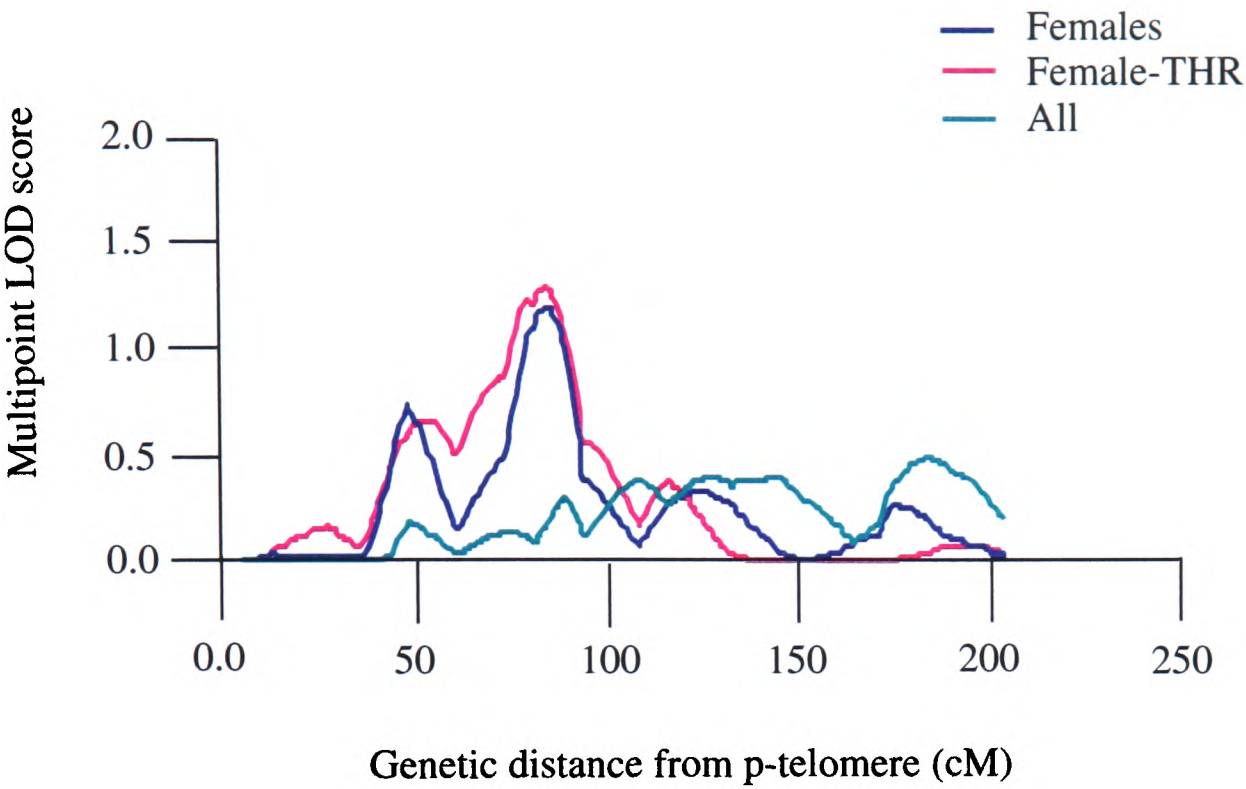


Figure 3.2a: Multipoint linkage analysis of chromosome 4 with markers at an average density of one marker every 10 cM. Results for all data, female and female-THR families. THR= total hip joint replacement, TKR= total knee joint replacement.

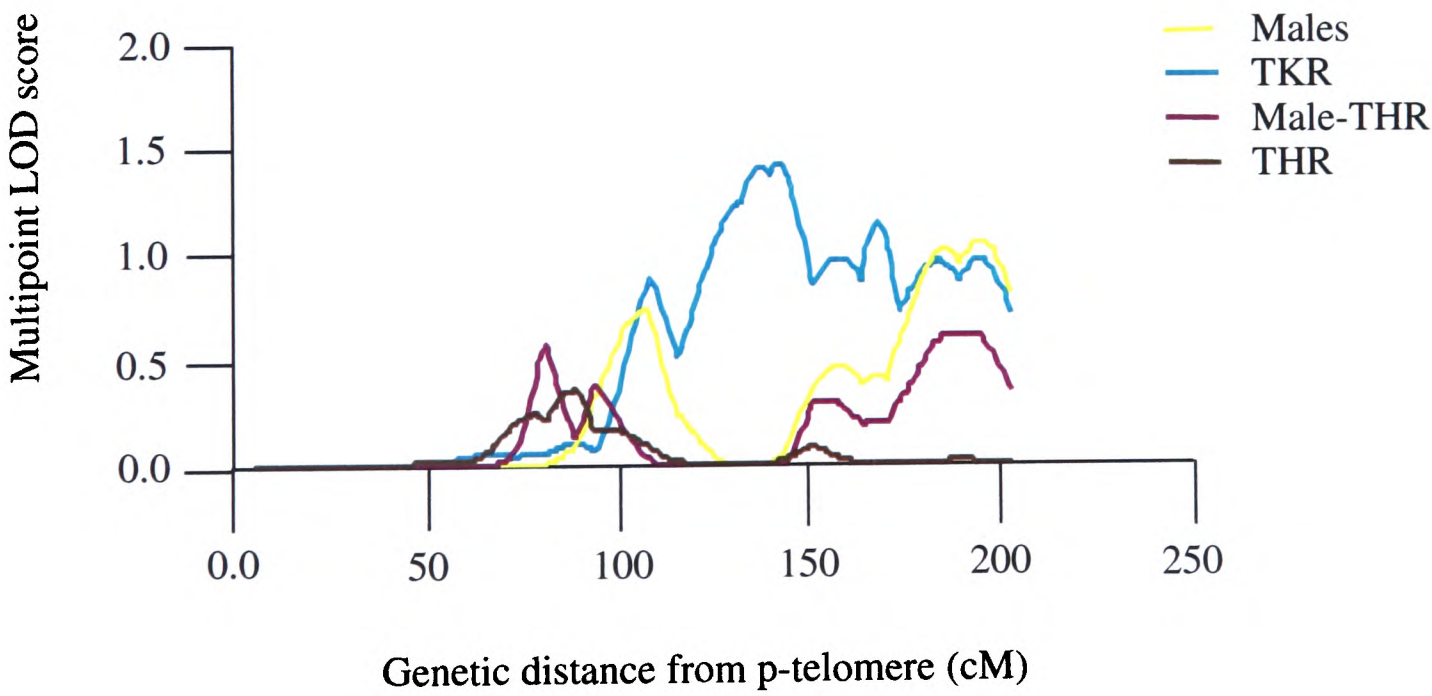


Figure 3.2b: Multipoint linkage analysis of chromosome 4 with markers at an average density of one marker every 10 cM. Results for THR, TKR, male and male-THR families. THR= total hip joint replacement, TKR= total knee joint replacement.

A second region of interest, not seen in the first screen, is also detected. Within the knee stratum five markers between D4S1572 and D4S1539 have significant p-values (table 3.2). Again this is reflected in the multipoint analysis with a corresponding peak reaching a maximum LOD of 1.4 centred at 140 cM from the p-telomere (figure 3.2b). This region is very large, with a 1-LOD support interval of over 100 cM.

No other strata tested show evidence for linkage to OA (figures 3.2a and b).

Since the regions showing evidence for linkage identified in this initial mapping are large, it was decided to undertake finer linkage mapping of these regions. In the 218 families with affected female sibling pairs, finer linkage mapping was performed between the markers D4S403 and D4S406, the markers encompassing the region showing evidence of linkage. For the 60 families with siblings concordant for TKR, finer linkage mapping was performed between markers D4S2964 and the q-telomere, thus encompassing the region that had shown evidence for linkage in this stratum. It was hoped such analysis would narrow the regions showing evidence for linkage prior to association analysis.

3.3 Finer (HD5) linkage mapping of chromosome 4

3.3.1 Female and female-THR families

Eleven additional microsatellite markers, located between D4S403 and D4S406, were genotyped in the 218 female families. These markers are from the ABI Prism LMSv2.5 HD5 set. In combination with the ten MD10 markers already genotyped within this region, the average density of markers was now one marker every 4 cM. Table 3.3a lists the two-point LOD scores and p-values for the eleven HD5 markers in addition to the previously genotyped MD10 markers. Again, p-values with a nominal significance ≤ 0.05 are

Marker	Distance from p-telomere in cM	Females		Female-THRs	
		LOD	P-value ¹	LOD	P-value ¹
D4S403	26.71	0.0	ns	0.2	ns
D4S419	34.97	0.0	ns	0.0	ns
D4S2994	38.61	0.1	ns	0.4	0.093
D4S3022	42.43	1.1	0.014	1.4	0.005
D4S391	47.33	1.0	0.018	0.8	0.026
D4S2912	50.77	0.0	ns	0.1	ns
D4S1587	54.21	0.0	ns	0.6	0.055
D4S405	60.23	0.0	ns	0.1	ns
D4S2971	66.78	1.0	0.015	1.3	0.008
D4S428	69.01	0.1	ns	1.2	0.009
D4S1592	72.71	0.2	ns	0.5	0.060
D4S398	75.67	0.4	0.092	0.6	0.049
D4S3004	77.86	0.7	0.038	0.9	0.021
D4S392	80.38	1.0	0.018	1.6	0.004
D4S3042	85.42	0.5	0.065	0.5	0.058
D4S2964	88.45	0.6	0.054	0.7	0.036
D4S1534	93.19	0.3	ns	0.4	0.092
D4S2460	96.96	0.4	0.100	0.8	0.029
D4S2986	104.31	0.2	ns	0.2	ns
D4S1572	107.52	0.0	ns	0.0	ns
D4S406	114.74	0.0	ns	0.1	ns

¹ ns = not significant

Table 3.3a: Two-point linkage analysis of chromosome 4 for the MD10 (black) and HD5 (blue) markers in the female and female -THR families. Only p-values ≤ 0.1 are shown. P-values ≤0.05 are highlighted in red. The genetic distance of markers from the p-telomere is also shown. All markers are from the ABI Prism Linkage Mapping Sets version 2/2.5. THR = total hip joint replacement.

highlighted in red. In the female-THR families nine markers between D4S3022 and D4S2460 have significant p-values. Figures 3.3a and b show the results of multipoint linkage analysis for all 218 female families and for the 146 female-THR families for all twenty-one markers, overlaid on the MD10 results. The addition of extra markers has resulted in an increase in the MLS both for all 218 female families and for the 146 female-THR families. For the female families the MLS is now 2.4, compared to 1.2 for the MD10 markers, whilst for the female-THR families the MLS is 2.7, compared to 1.3 for the MD10 markers. The position (80 cM from the p-telomere) and width of the region are very similar to those observed for the MD10 markers, although there has been some narrowing of the region towards the q-telomere. The 1-LOD support interval has been narrowed though and is now approximately 15 cM for all 218 female families and also for the 146 female-THR families.

It was clear from the above that the linkage to chromosome 4 was principally accounted for by the 146 female-THR families from our cohort of 218 female families: the MLS for the 146 female-THR families in both stages of the screening was higher than that for all 218 families. It was decided, therefore, to focus on the female-THR families and to genotype additional markers in the centre of the region showing evidence of linkage in this stratum only in order to further narrow and define the linked region.

3.3.2 TKR families

Thirteen additional microsatellite markers, located between D4S2964 and the q-telomere, were genotyped in the 60 TKR families. These markers are from the ABI Prism LMSv2.5 HD5 set. In combination with the twelve MD10 markers already genotyped within this region, the average density of markers was now one marker every 5 cM. Table 3.3b lists the

Female Families

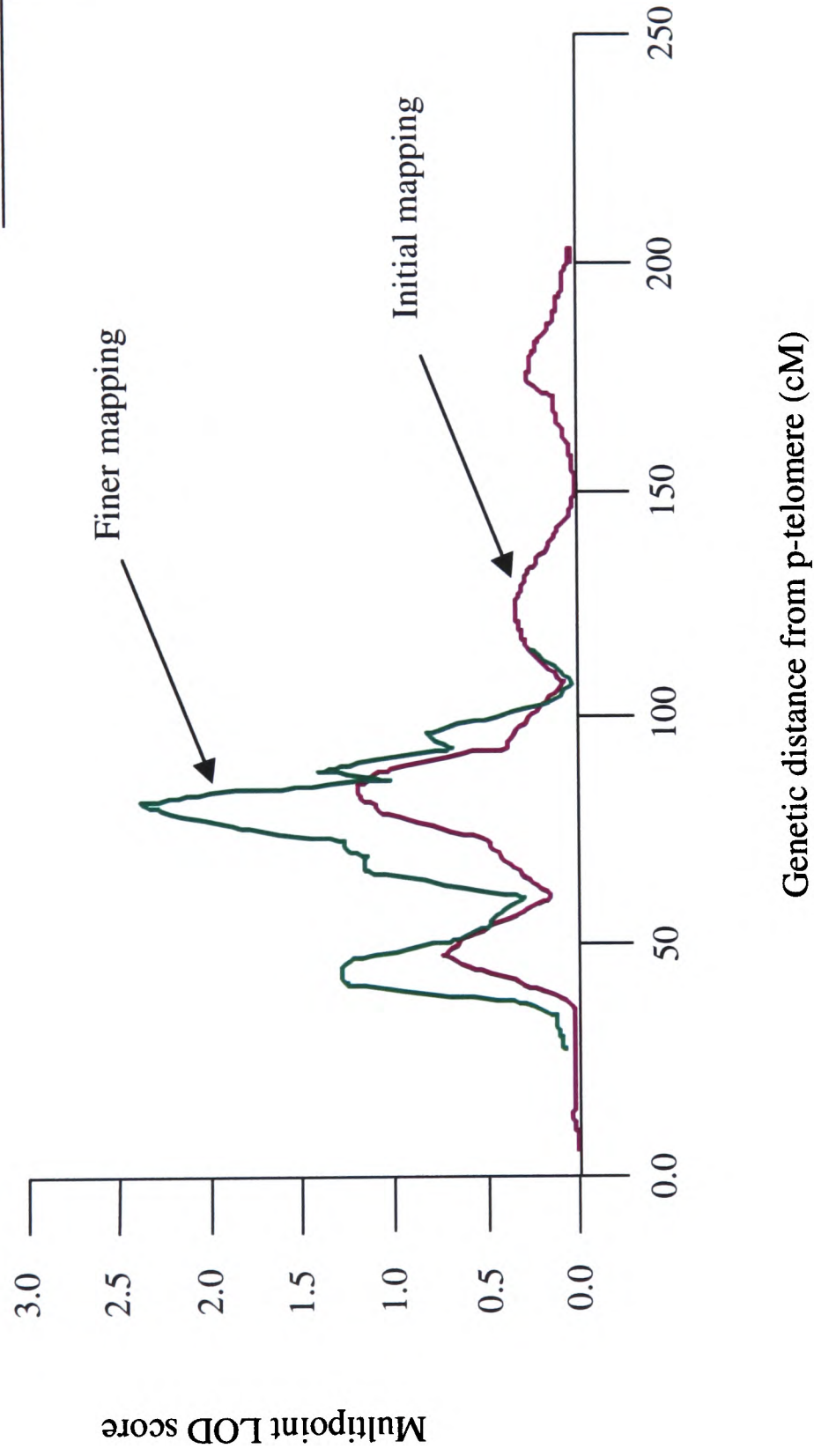


Figure 3.3a: Multipoint linkage analysis of chromosome 4 for the 218 female families, with markers at an average of one marker every 4 cM, overlaid on the results of the initial mapping with one marker every 10 cM.

Female-THR Families

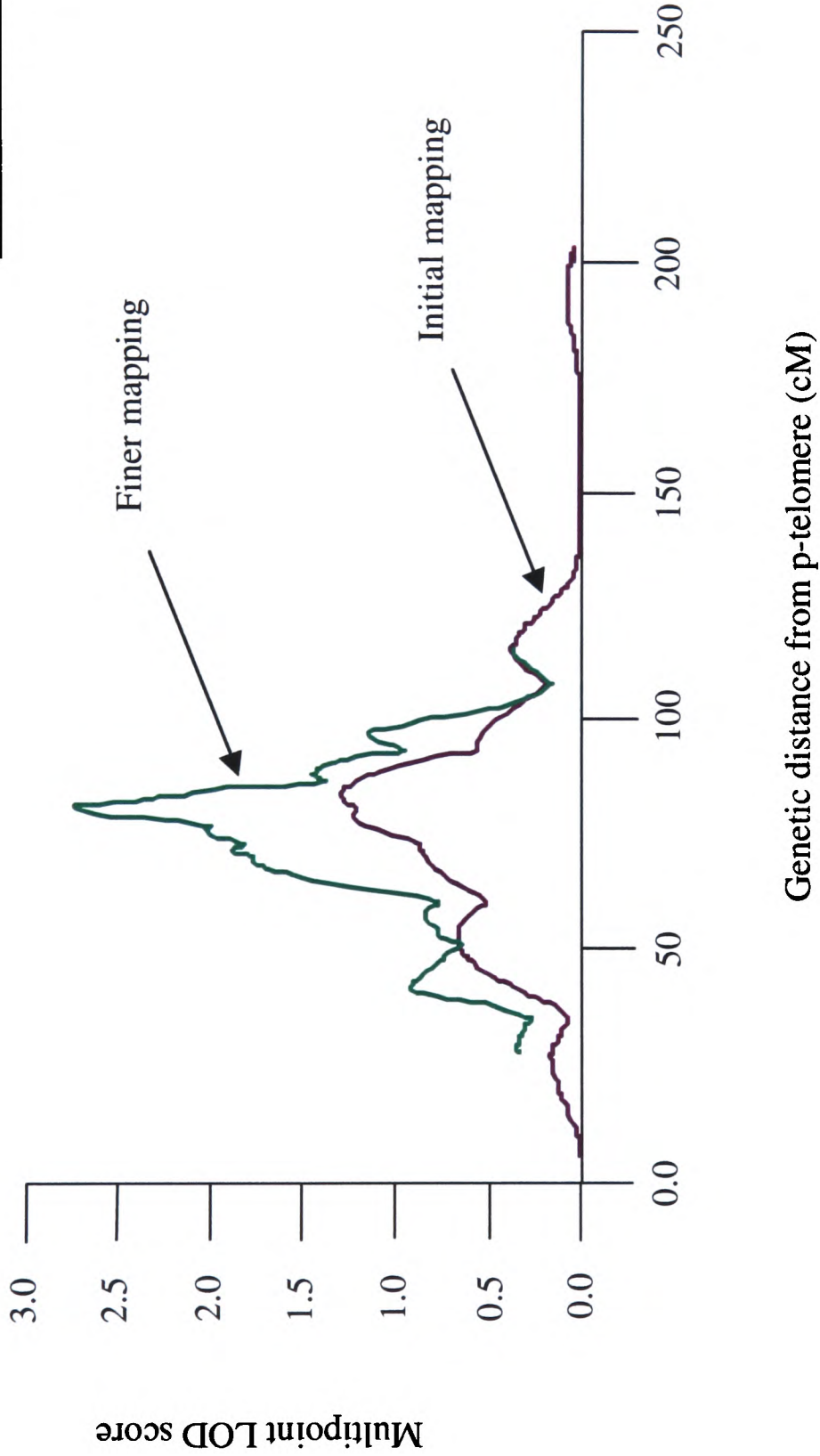


Figure 3.3b: Multipoint linkage analysis of chromosome 4 for the 146 female-THR families, with markers at an average of one marker every 4 cM, overlaid on the results of the initial mapping with one marker every 10 cM.

Marker	Distance from p-telomere in cM	TKRs	
		LOD	P- value ¹
D4S2964	88.45	0.2	ns
D4S1534	93.19	0.0	ns
D4S2460	96.96	0.1	ns
D4S2986	104.31	0.5	0.063
D4S1572	107.52	0.7	0.041
D4S406	114.74	0.2	ns
D4S1615	126.83	1.0	0.017
D4S1575	131.59	0.8	0.025
D4S1579	136.82	0.1	ns
D4S424	138.87	1.3	0.008
D4S1586	141.21	0.1	ns
D4S2962	144.74	0.2	ns
D4S413	150.31	0.8	0.028
D4S3046	155.87	0.4	0.081
D4S2952	160.72	0.1	ns
D4S1597	163.65	0.3	ns
D4S1595	168.07	0.0	ns
D4S1539	170.01	0.7	0.040
D4S415	173.24	0.4	0.083
D4S2920	186	0.1	ns
D4S1535	189.38	0.4	0.097
D4S2924	193.15	0.2	ns
D4S3051	200.02	0.2	ns
D4S426	202.69	0.4	0.100
D4S2930	204.47	0.5	0.064

¹ns = not significant

Table 3.3b: Two-point linkage analysis of chromosome 4 for the MD10 (black) and HD5 (blue) markers in the TKR families. Only p-values ≤ 0.1 are shown. P-values ≤0.05 are highlighted in red. The genetic distance of markers from the p-telomere is also shown. All markers are from the ABI Prism Linkage Mapping Sets version 2/2.5. TKR = total knee joint replacement.

two-point LOD scores and p-values for both the thirteen HD5 markers and the previously genotyped MD10 markers in this region. Again, p-values with a nominal significance ≤ 0.05 are highlighted in red. Only one of the additional markers, D4S1615, has a significant p-value. Figure 3.3c shows the results of multipoint linkage analysis with all twenty-five markers overlaid on the MD10 results. The addition of extra markers has resulted in a decrease in the MLS from 1.4 to 1.0 and has failed to narrow the region. This result would appear to suggest that this is not a genuine susceptibility region. No further analysis was undertaken.

3.4 Additional finer linkage mapping of chromosome 4

Five additional microsatellite markers in the centre of the region showing evidence of linkage (D4S403 - D4S406) were genotyped in the 146 female-THR families in order to further define the linked region. In combination with the markers already genotyped within this region, the average density of markers is now one marker every 3 cM. Table 3.4 lists the two-point LOD scores and p-values for the five additional markers in addition to the previously genotyped markers in the region. P-values with a nominal significance ≤ 0.05 are highlighted in red. All five additional markers have p-values at or below 0.02, with two, D4S1600 and D4S3253, having p-values of 0.001, representing the most significant two-point p-values for this chromosome. Figure 3.4 shows the results of multipoint linkage analysis for all markers genotyped in the 146 female-THR families, overlaid on the previous MD10 plus HD5 results. The addition of these five extra markers has resulted in both an increase in MLS, from 2.7 to 3.1, and a narrowing of the 1-LOD support interval, from 15 to 4 cM.

TKR Families

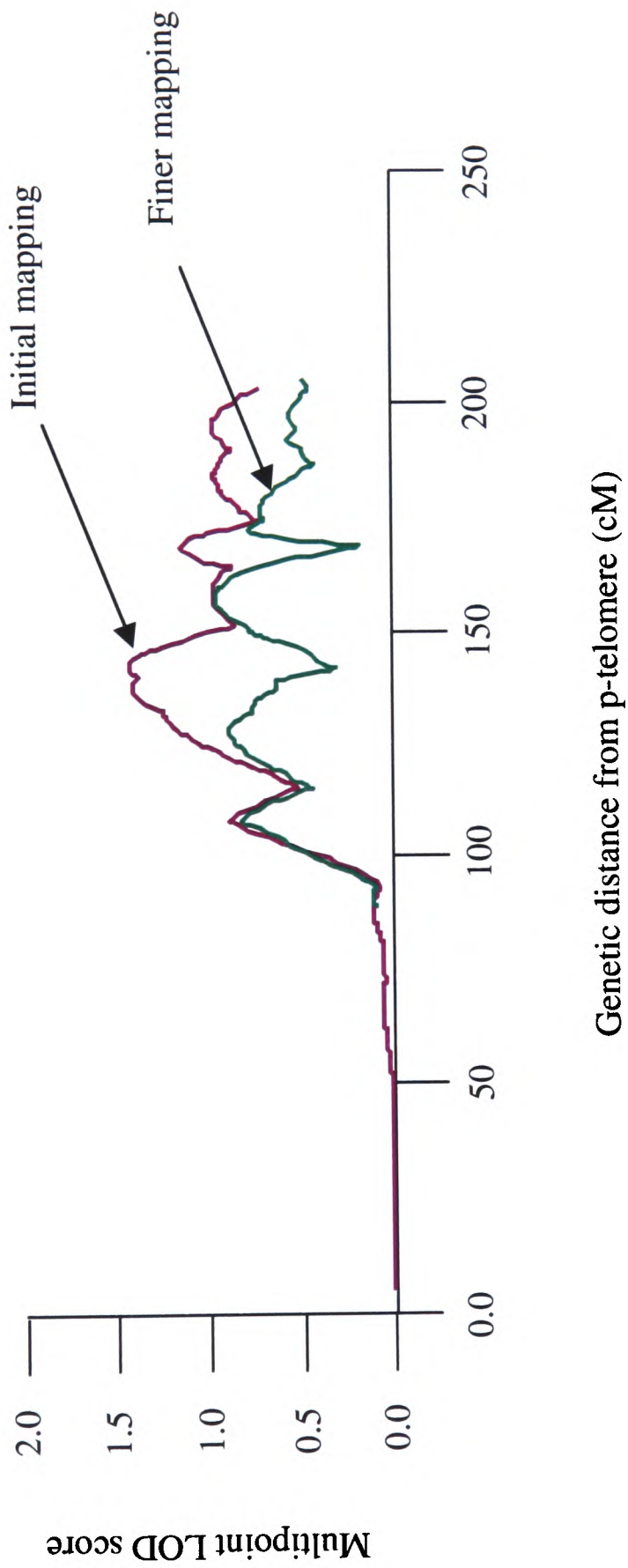


Figure 3.3c: Multipoint linkage analysis of chromosome 4 for the 60 TKR families, with markers at an average of one marker every 5 cM, overlaid on the results of the initial mapping with one marker every 10 cM.

Marker	Distance from p-telomere in cM	Female-THR	
		LOD	P-value ¹
D4S403	26.71	0.2	ns
D4S419	34.97	0.0	ns
D4S2994	38.61	0.4	0.093
D4S3022	42.43	1.4	0.005
D4S391	47.33	0.8	0.026
D4S2912	50.77	0.1	ns
D4S1587	54.21	0.6	0.055
D4S405	60.23	0.1	ns
D4S2971	66.78	1.3	0.008
D4S428	69.01	1.2	0.009
D4S1583	71.58	1.0	0.018
D4S1592	72.71	0.5	0.060
D4S398	75.67	0.6	0.049
D4S1600	76.18	2.1	0.001
D4S3004	77.86	0.9	0.021
D4S2987	78.03	1.8	0.002
D4S3253	78.68	2.0	0.001
D4S1519	79.3	0.9	0.022
D4S392	80.38	1.6	0.004
D4S3042	85.42	0.5	0.058
D4S2964	88.45	0.7	0.036
D4S1534	93.19	0.4	0.092
D4S2460	96.96	0.8	0.029
D4S2986	104.31	0.2	ns
D4S1572	107.52	0.0	ns
D4S406	114.74	0.1	ns

¹ ns = not significant

Table 3.4: Two-point linkage analysis of chromosome 4 for the MD10 (black), HD5 (blue) and custom made (green) markers. Only p-values ≤ 0.1 are shown. P-values ≤0.05 are highlighted in red. The genetic distance of markers from the p-telomere is also shown. THR = total hip joint replacement.

Female-THR Families

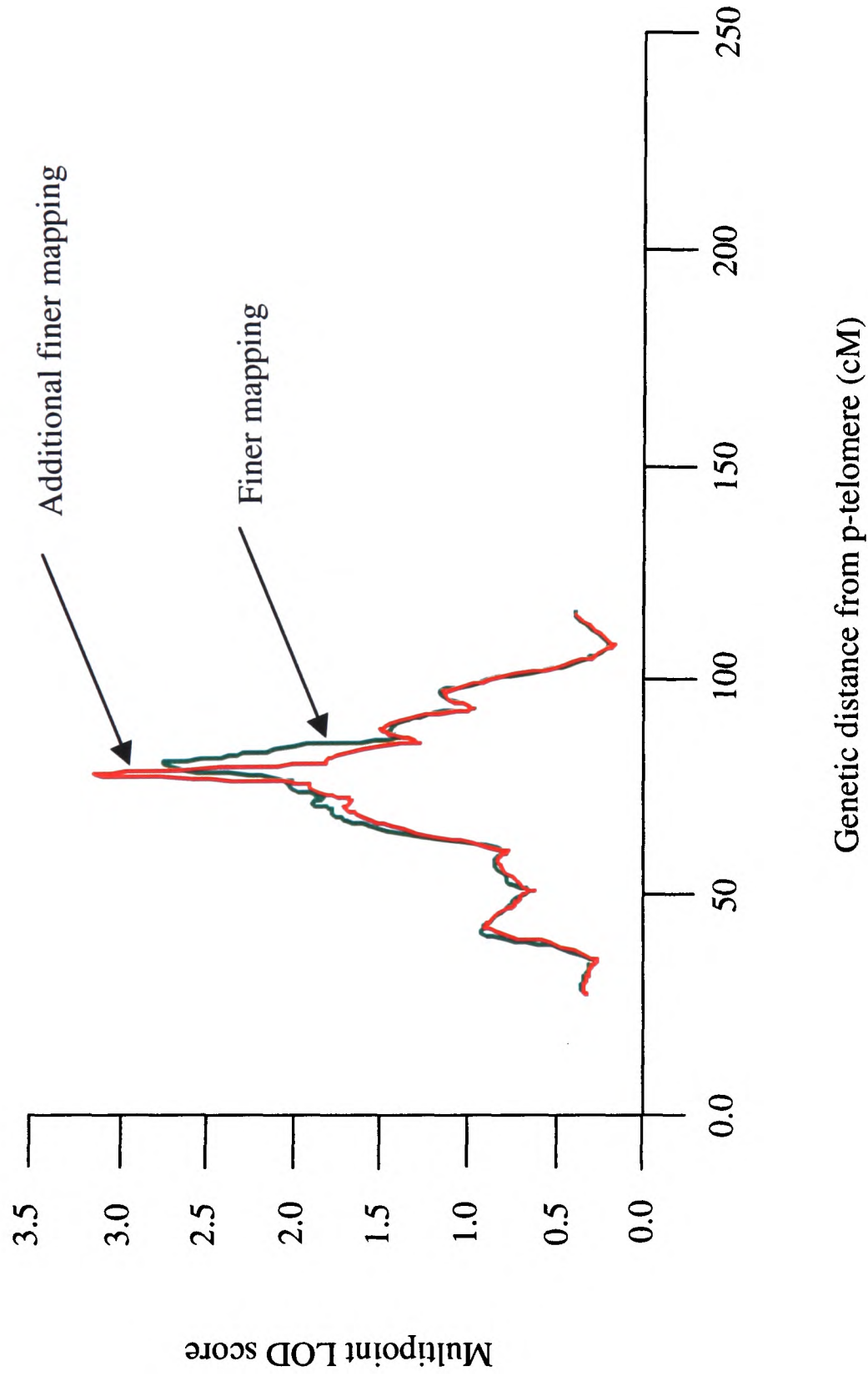


Figure 3.4: Multipoint linkage analysis of chromosome 4 for the 146 female -THR families, showing the results of an additional five markers in the centre of the region of interest overlaid on the results of finer mapping with markers at an average of one marker every 4 cM.

3.5 Initial (MD10) linkage mapping of chromosome 16

Twelve microsatellite markers from the ABI Prism LMSv2 MD10 were used for the initial mapping of chromosome 16. These markers are an average of 9 cM apart and cover the whole of chromosome 16. Two-point (table 3.5) and multipoint analysis (figures 3.5a and b) was undertaken both for all data and following stratification according to sex, joint replaced and sex plus joint replaced.

Two-point p-values at or below a nominal significance value of 0.05 are highlighted in red (table 3.5). The only two strata with significant two-point p-values are the female families, for markers D16S515 and D16S516, and the female-THR families, for markers D16S3103 and D16S3046.

The two regions identified by multipoint analysis as showing evidence for linkage, both with maximum LOD scores of 1.4 (figure 3.5a), are in similar positions and of similar widths as those in the original scan. The first region is centred at 50 cM from the p-telomere and has a 1-LOD support interval of 19 cM. The second region is centred at 93 cM from the p-telomere and has a 1-LOD support interval of 23 cM. In the original scan the first region was restricted to female families and the second to female-THR families. In this analysis however both regions are accounted for by female families. No other strata show evidence for linkage to OA (figures 3.5a and b). As for chromosome 4, the regions showing evidence for linkage are large. Therefore, it was decided to undertake finer linkage mapping between markers D16S3075 and D16S3091, the markers encompassing the two regions showing evidence for linkage, in the 218 families with affected female sibling pairs.

Marker	Distance from p-telomere in cM	All		Females		Female-THRs		Males		Male-THRs		THRs		TKRs	
		LOD	P-value ¹	LOD	P-value ¹	LOD	P-value ¹	LOD	P-value ¹	LOD	P-value ¹	LOD	P-value ¹	LOD	P-value ¹
D16S423	14.05	0.0	ns	0.0	ns	0.0	ns	0.0	ns	0.0	ns	0.0	ns	0.2	ns
D16S404	24.89	0.0	ns	0.0	ns	0.0	ns	0.0	ns	0.0	ns	0.0	ns	0.5	0.061
D16S3075	28.63	0.0	ns	0.0	ns	0.0	ns	0.1	ns	0.0	ns	0.0	ns	0.0	ns
D16S3103	37.97	0.2	ns	0.1	ns	0.8	0.027	0.3	ns	0.1	ns	0.5	0.063	0.0	ns
D16S3046	42.99	0.0	ns	0.4	0.085	0.9	0.019	0.2	ns	0.2	ns	0.4	0.091	0.0	ns
D16S3068	49.5	0.6	0.052	0.3	ns	0.2	ns	0.4	0.086	0.1	ns	0.2	ns	0.0	ns
D16S3136	60.35	0.0	ns	0.0	ns	0.0	ns	0.0	ns	0.0	ns	0.0	ns	0.0	ns
D16S415	66.13	0.0	ns	0.0	ns	0.1	ns	0.0	ns	0.0	ns	0.0	ns	0.1	ns
D16S503	83.09	0.0	ns	0.4	0.077	0.0	ns	0.0	ns	0.0	ns	0.0	ns	0.0	ns
D16S515	92.06	0.1	ns	0.6	0.048	0.2	ns	0.0	ns	0.0	ns	0.4	0.098	0.0	ns
D16S516	97.99	0.0	ns	1.3	0.008	0.2	ns	0.2	ns	0.1	ns	0.0	ns	0.0	ns
D16S3091	109.59	0.0	ns	0.0	ns	0.0	ns	0.3	ns	0.4	0.087	0.0	ns	0.2	ns

¹ ns = not significant

Table 3.5: Two-point linkage analysis of chromosome 16 for the MD10 markers. Results for all families and following stratification by sex, joint replaced and sex plus joint replaced. Only p-values ≤ 0.1 are shown. P-values ≤0.05 are highlighted in red. The genetic distance of markers from the p-telomere is also shown. Markers are from the ABI Prism Linkage Mapping Set MD10. THR = total hip joint replacement, TKR = total knee joint replacement.

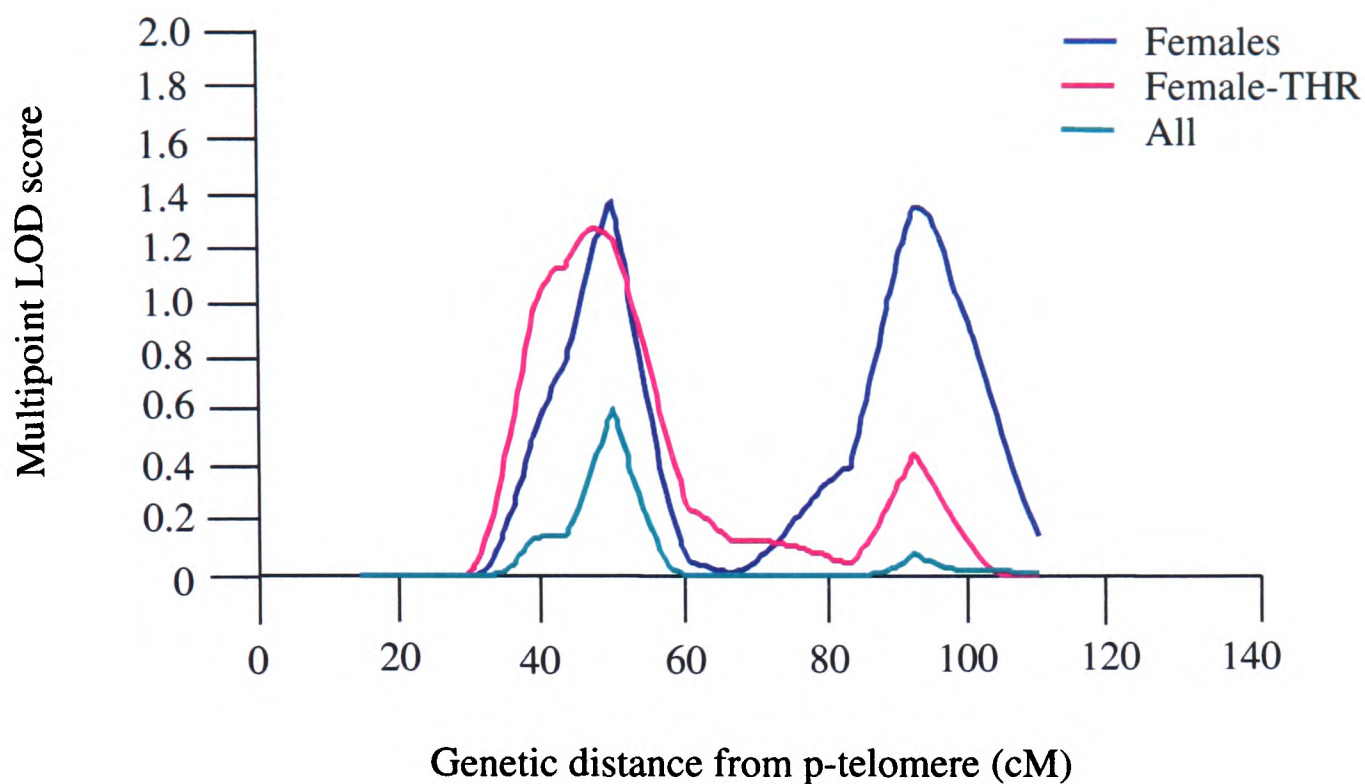


Figure 3.5a: Multipoint linkage analysis of chromosome 16 with markers at an average density of one marker every 9 cM. Results for all data, female and female-THR families. THR= total hip joint replacement, TKR= total knee joint replacement.

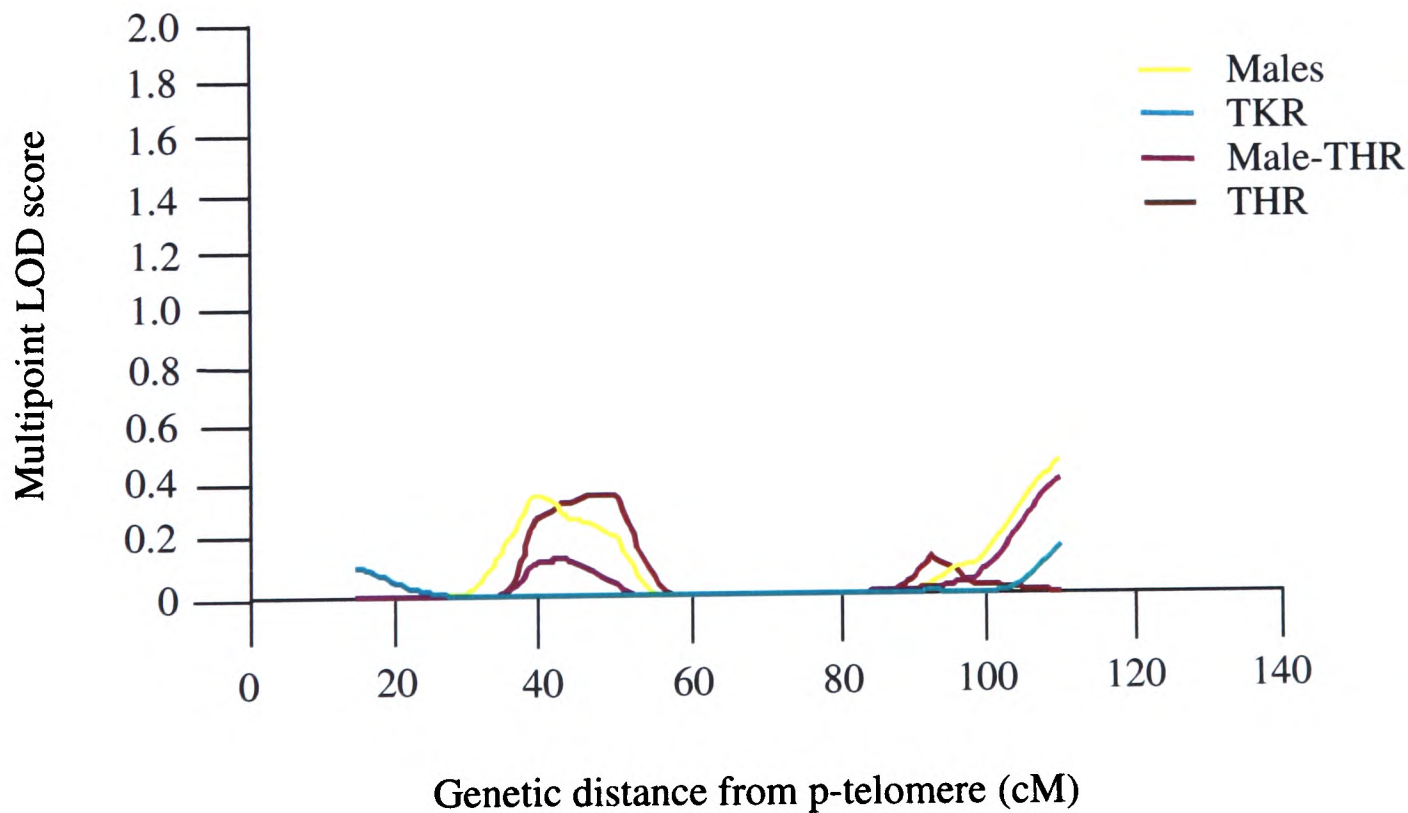


Figure 3.5b: Multipoint linkage analysis of chromosome 16 with markers at an average density of one marker every 9 cM. Results for THR, TKR, male and male-THR families. THR= total hip joint replacement, TKR= total knee joint replacement.

3.6 Finer (HD5) linkage mapping of chromosome 16

Eleven additional microsatellite markers, located between D16S3075 and D16S3091, were genotyped in the 218 female families. These markers are from the ABI Prism LMSv2 HD5 set. In combination with the ten MD10 markers already genotyped within this region, the average density of markers was now one marker every 4 cM. Table 3.6 lists the two-point LOD scores and p-values for the HD5 markers in addition to the previously typed MD10 markers. P-values with a nominal significance ≤ 0.05 are highlighted in red.

In the 146 female-THR families all four markers between D16S3103 and D16S403 have significant p-values with the HD5 marker D16S3041 having the lowest ($P = 0.003$). The HD5 marker D16S3066 also has a significant p-value in this stratum. In the female families, seven markers between D16S514 and D16S505 have significant p-values with the HD5 marker D16S3066 having the lowest ($P = 0.002$) [table 3.6].

Figures 3.6a and b show the results of multipoint linkage analysis for all 218 female families and for the 146 female-THR families respectively, for all twenty-one markers overlaid on the MD10 results. Multipoint analysis in the female families (figure 3.6a) shows two regions with evidence for linkage in a very similar position to those observed in the MD10 screen. The first region shows a slight decrease in MLS from 1.4 to 1.3 whilst the second shows an increase from 1.4 to 1.7.

Multipoint analysis in the female-THR families (figure 3.6b) identifies the same two regions. The first region has an increased MLS compared with the MD10 results (1.6 versus 1.3). The second region also has an increased MLS (1.1 versus 0.5). When comparing the female families to the female-THR families, it is apparent that the 146 female-THR

Marker	Distance from p-telomere in cM	Females		Female-THR	
		LOD	P-value ¹	LOD	P-value ¹
D16S3075	28.63	0.0	ns	0.0	ns
D16S3102	29.63	0.0	ns	0.0	ns
D16S500	33.8	0.0	ns	0.0	ns
D16S3103	37.97	0.1	ns	0.8	0.027
D16S3041	41.80	0.4	0.096	1.7	0.003
D16S3046	42.99	0.4	0.085	0.9	0.019
D16S403	45.58	0.2	ns	0.7	0.041
D16S3068	49.5	0.3	ns	0.2	ns
D16S3100	51.64	0.3	ns	0.3	ns
D16S3136	60.35	0.0	ns	0.0	ns
D16S415	66.13	0.0	ns	0.1	ns
D16S3140	71.50	0.2	ns	0.0	ns
D16S514	79.95	0.7	0.032	0.2	ns
D16S503	83.09	0.4	0.077	0.0	ns
D16S3066	88.61	1.9	0.002	1.4	0.006
D16S515	92.06	0.6	0.048	0.2	ns
D16S3049	96.17	0.9	0.019	0.2	ns
D16S516	97.99	1.3	0.008	0.2	ns
D16S3040	100.51	0.8	0.029	0.1	ns
D16S505	105.09	0.6	0.044	0.4	0.095
D16S3091	109.59	0.0	ns	0.0	ns

¹ ns = not significant

Table 3.6: Two-point linkage analysis of chromosome 16 for the MD10 (black) and HD5 (blue) markers. Only p-values ≤ 0.1 are shown. P-values ≤ 0.05 are highlighted in red. The genetic distance of markers from the p-telomere is also shown. All markers are from the ABI Prism Linkage Mapping Sets version 2. THR = total hip joint replacement.

Female Families

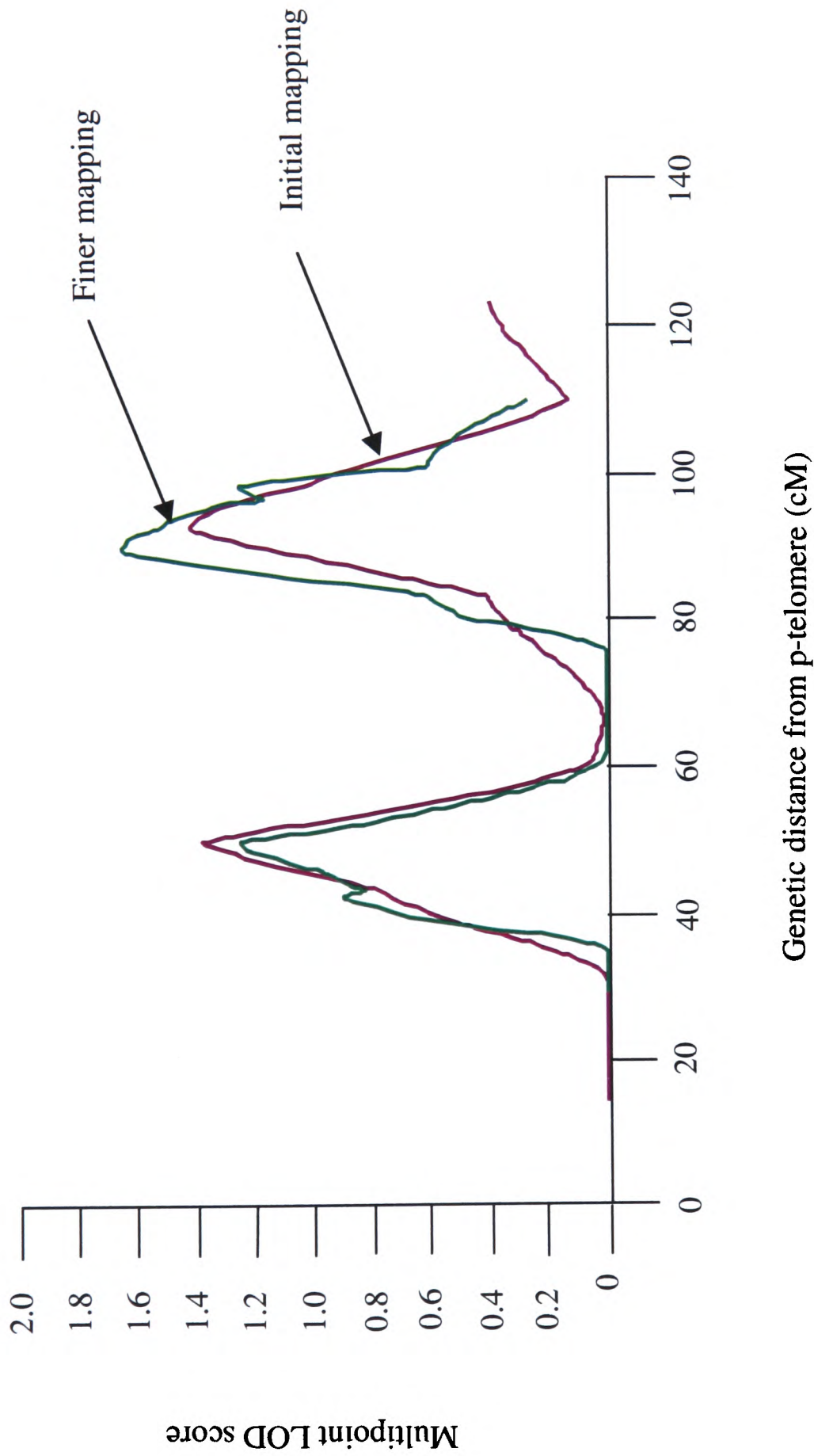


Figure 3.6a: Multipoint linkage analysis of chromosome 16 for the 218 female families, with markers at an average of one marker every 4 cM, overlaid on the results of the initial mapping with one marker every 9 cM.

Female-THR Families

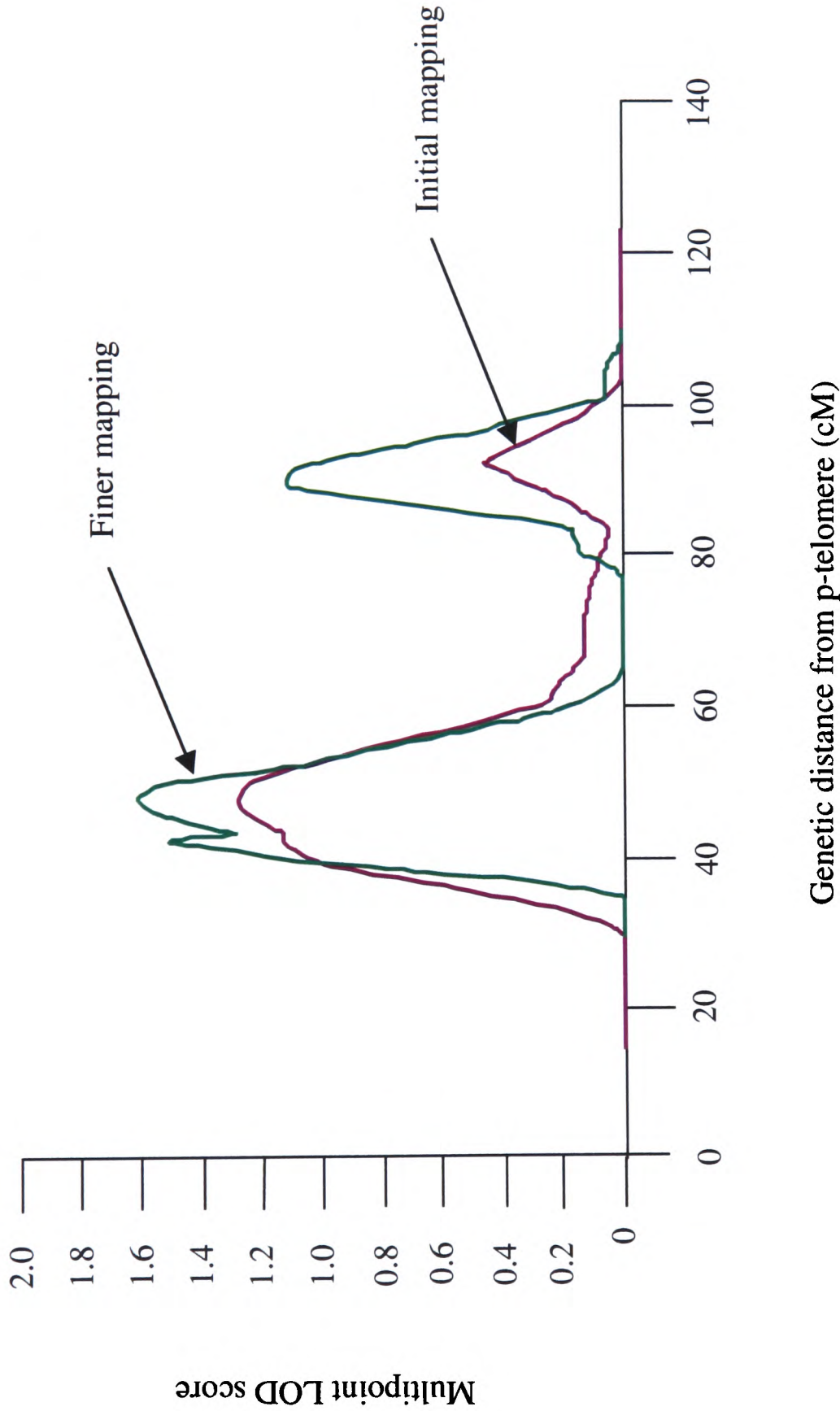


Figure 3.6b: Multipoint linkage analysis of chromosome 16 for the 146 female-THR families, with markers at an average of one marker every 4 cM, overlaid on the results of the initial mapping with one marker every 9 cM.

families account for the first linkage region, whereas the second region is accounted for by all 218 female families.

The addition of extra markers has had no effect on the 1-LOD support interval for the first region, in either strata, which remains at 19 cM. There has been some narrowing of the second region with a 1-LOD support interval of 17 cM in the female families following this finer mapping compared with 23 cM in the initial mapping. This limited narrowing may reflect the contribution of only a small number of families to the evidence for linkage at this region, limiting the number of informative meioses and therefore restricting the ability to narrow the region. Alternatively, each region identified may represent more than one region, which cannot be resolved by finer linkage mapping. A final possibility is that the markers used are still not of a sufficiently high density and that the addition of further markers would allow for narrowing of these regions.

Since the identified regions are still too large to allow for systematic association analysis it was decided to genotype additional markers within the two regions showing evidence for linkage (D16S500 - D16S415 and D16S3140 - D16S3091) in the 218 female families in an attempt to further narrow and define them.

3.7 Additional finer linkage mapping of chromosome 16

Eleven additional microsatellite markers in each of the two regions (a total of twenty-two additional markers) showing evidence for linkage (D16S500 - D16S415 and D16S3140 - D16S3091) were identified from the Marshfield map (http://research.marshfieldclinic.org/genetics/map_markers/maps). At the time of marker selection this was reported to be the most reliable and widely used available map. Subsequent to the publication of the deCODE

genetic map (Kong et al. 2002) markers were ordered according to their position in the deCODE map. The deCODE map was constructed using 5,136 short tandem repeat markers in a total of 146 families and represents a five-fold increased resolution compared to the best previously available maps. The increase in precision of marker location achieved by this map serves to improve the statistical power of multipoint linkage analysis. Any markers genotyped in this study that are not positioned in the deCODE map were included at positions which are representative of their positions, relative to other markers, in the Marshfield map. Two of the markers genotyped, D16S403 and D16S412, have the same genetic position in the deCODE map. Although D16S403 is an HD5 marker and was used in earlier analysis, for additional finer mapping analysis we used D16S412.

Table 3.7 lists the two-point LOD scores and p-values for the twenty-two additional markers in addition to the previously genotyped markers in the two regions. The additional twenty-two markers in combination with the markers already typed within the regions of interest represent an average density of markers of one marker every 1.8 cM in the first region (D16S500 - D16S415) and one marker every 1.9 cM in the second (D16S3140 - D16S3091). P-values with a nominal significance ≤ 0.05 are highlighted in red.

3.7.1 Region 1 (D16S500 - D16S415)

For the female families, two of the additional markers, D16S412 and D16S3113 have significant p-values. In the female-THR families, five markers between D16S3041 and D16S3113 (including the additional markers D16S3036, D16S412 and D16S3113) have p-values at or approaching significance (table 3.7).

Marker	Distance from p-telomere in cM	Females		Female-THR _s	
		LOD	P- value ¹	LOD	P-value ¹
D16S500	33.8	0.0	ns	0.0	ns
D16S3103	37.97	0.1	ns	0.8	0.027
D16S499	39.08	0.0	ns	0.0	ns
D16S501	39.62	0.0	ns	0.0	ns
D16S3056	40.23	0.0	ns	0.0	ns
D16S3041	41.80	0.4	0.096	1.7	0.003
D16S3036	42.07	0.1	ns	0.4	0.086
D16S3046	42.99	0.4	0.085	0.9	0.019
D16S412	45.58	0.7	0.030	1.6	0.003
D16S420	47.57	0.1	ns	0.2	ns
D16S3113	48.13	1.0	0.014	0.5	0.065
D16S3133	48.5	0.4	0.088	0.3	ns
D16S3068	49.5	0.3	ns	0.2	ns
D16S769	50.54	0.0	ns	0.0	ns
D16S3100	51.64	0.3	ns	0.3	ns
D16S409	58.41	0.0	ns	0.1	ns
D16S411	59.49	0.1	ns	0.1	ns
D16S3136	60.35	0.0	ns	0.0	ns
D16S415	66.13	0.0	ns	0.1	ns
D16S3140	71.50	0.2	ns	0.0	ns
D16S3071	72.19	0.0	ns	0.0	ns
D16S3057	73.98	0.0	ns	0.0	ns
D16S3132	78.64	0.0	ns	0.0	ns
D16S514	79.95	0.7	0.032	0.2	ns
D16S3129	81.25	0.2	ns	0.1	ns
D16S503	83.09	0.4	0.077	0.0	ns
D16S3031	84.88	0.4	0.082	0.1	ns
D16S3095	86.16	0.8	0.030	0.3	ns
D16S3066	88.61	1.9	0.002	1.4	0.006
D16S3018	90.85	1.1	0.012	0.5	0.068
D16S515	92.06	0.6	0.048	0.2	ns
D16S3138	93.44	1.5	0.005	1.0	0.017
D16S518	95.45	1.1	0.011	0.3	ns
D16S3049	96.17	0.9	0.019	0.2	ns
D16S516	97.99	1.3	0.008	0.2	ns
D16S504	98.99	0.3	ns	0.1	ns
D16S3040	100.51	0.8	0.029	0.1	ns
D16S3073	101.51	0.1	ns	0.0	ns
D16S505	105.09	0.6	0.044	0.4	0.095
D16S3091	109.59	0.0	ns	0.0	ns

¹ ns = not significant

Table 3.7: Two-point linkage analysis of chromosome 16 for the MD10 (black), HD5 (blue) and custom made (green) markers. Only p-values ≤ 0.1 are shown. P-values ≤ 0.05 are highlighted in red. The genetic distance of markers from the p-telomere is also shown. THR = total hip joint replacement.

Figures 3.7a and b show the results of multipoint linkage analysis for all markers genotyped, for the 218 female families and for the 146 female-THR families, overlaid on the previous (MD10 plus HD5) results. The addition of the extra eleven markers in region 1 has caused a slight increase in MLS in the female families from 1.3 to 1.5 (figure 3.7a) and in the female-THR families from 1.6 to 1.7 (figure 3.7b). Again, as for the finer mapping, this suggests that the evidence for linkage in this region is largely accounted for by families with affected female-THR sibling pairs. The addition of extra markers in this region has also narrowed the 1-LOD support interval from 19 cM to 12 cM in the female-THR families.

3.7.2 Region 2 (D16S3140 - D16S3091)

Eleven markers in the interval between D16S514 and D16S505 have a significant p-value for the female families with a further two approaching significance, providing strong support for evidence of linkage. In the female-THR families two markers in this region have a significant p-value (table 3.7).

Figures 3.7c and d show the results of multipoint linkage analysis for all markers genotyped, for the 218 female families and for the 146 female-THR families, overlaid on the previous (MD10 plus HD5) results. The addition of the extra eleven markers has caused an increase in MLS in the female families from 1.7 to 1.9 (figure 3.7c) and in the female-THR families from 1.1 to 1.3 (figure 3.7d). This again suggests, as for all previous analyses, that the evidence for linkage in this region is accounted for by affected female sibling pairs and is not restricted to those pairs concordant for THR as is the case for the first region. The addition of extra markers has also narrowed the 1-LOD support interval, from 17 cM to 10 cM in the female families.

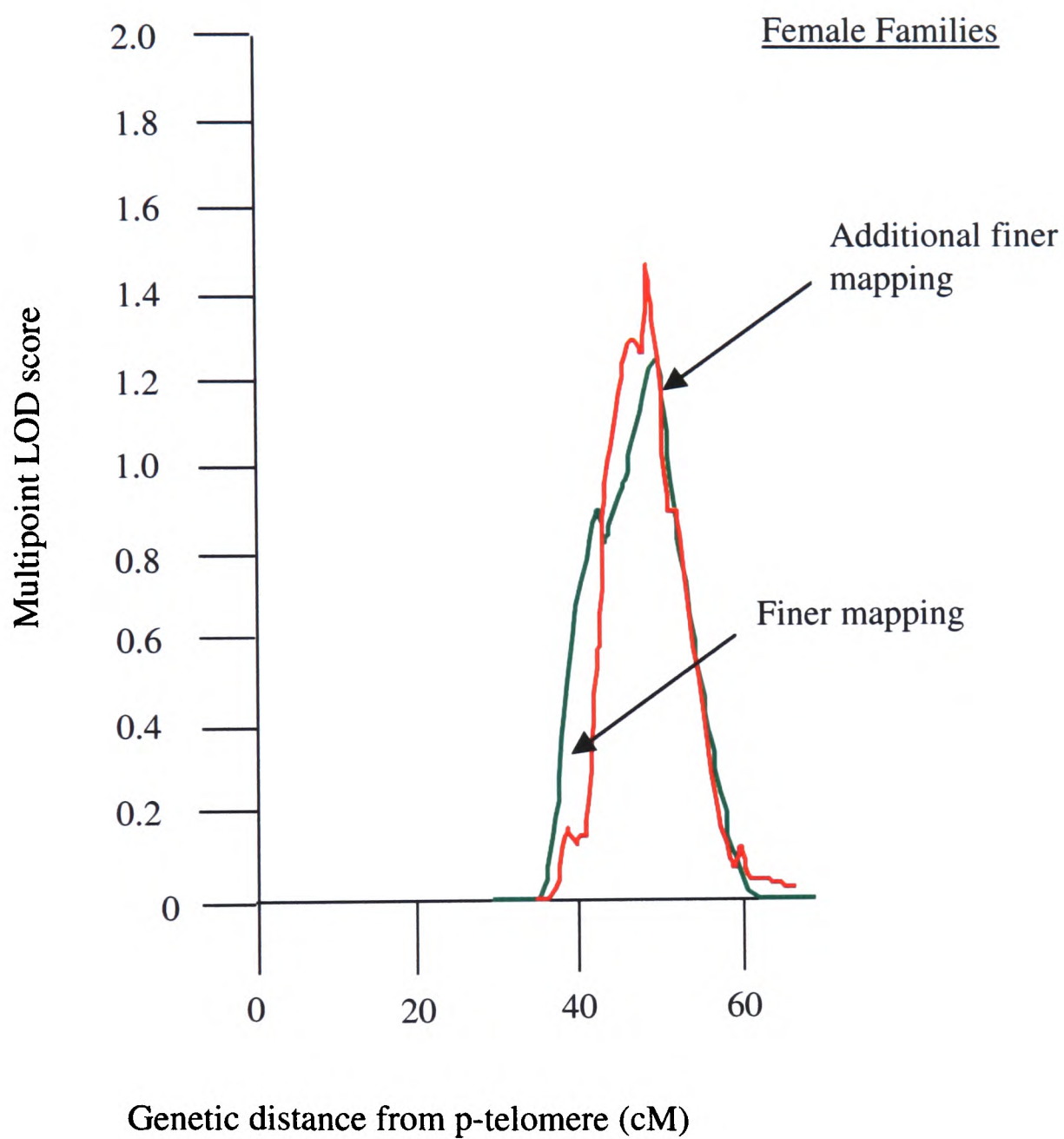


Figure 3.7a: Multipoint linkage analysis of chromosome 16, region 1 for the 218 female families, with markers at an average of one marker every 1.8 cM, overlaid on the results of the finer mapping with one marker every 4 cM.

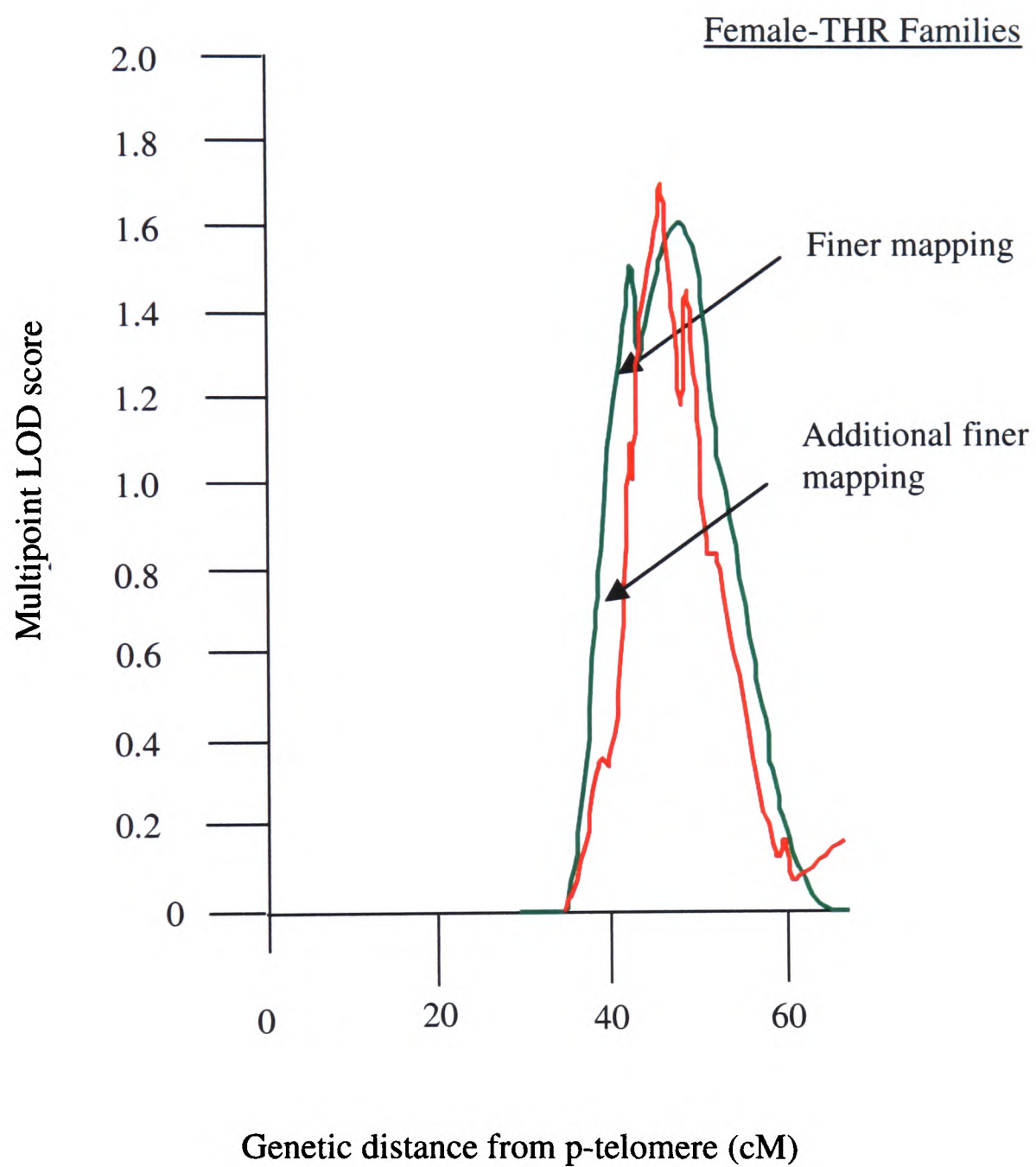


Figure 3.7b: Multipoint linkage analysis of chromosome 16, region 1 for the 146 female-THR families, with markers at an average of one marker every 1.8 cM, overlaid on the results of the finer mapping with one marker every 4 cM.

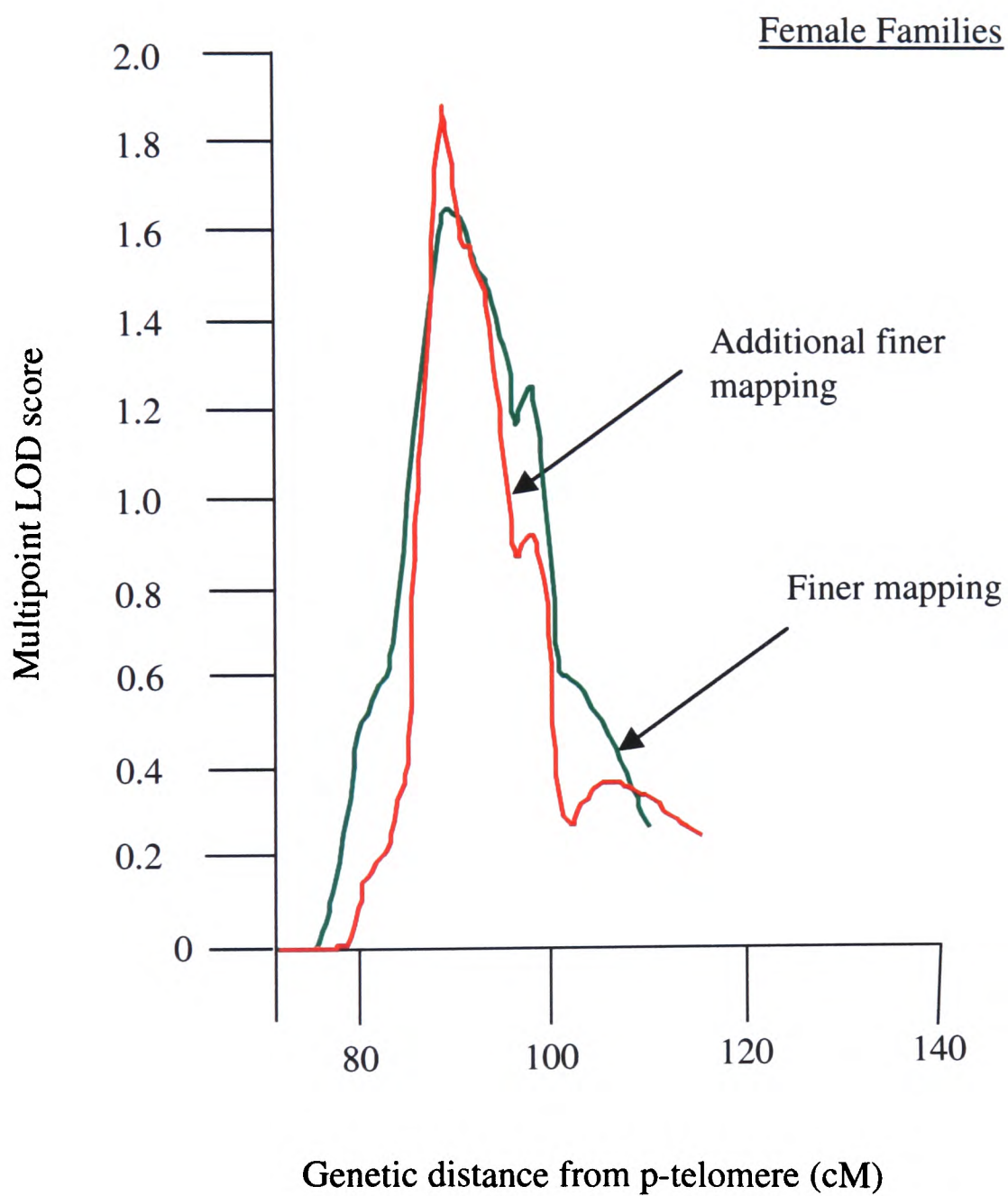


Figure 3.7c: Multipoint linkage analysis of chromosome 16, region 2 for the 218 female families, with markers at an average of one marker every 1.9 cM, overlaid on the results of the finer mapping with one marker every 4 cM.

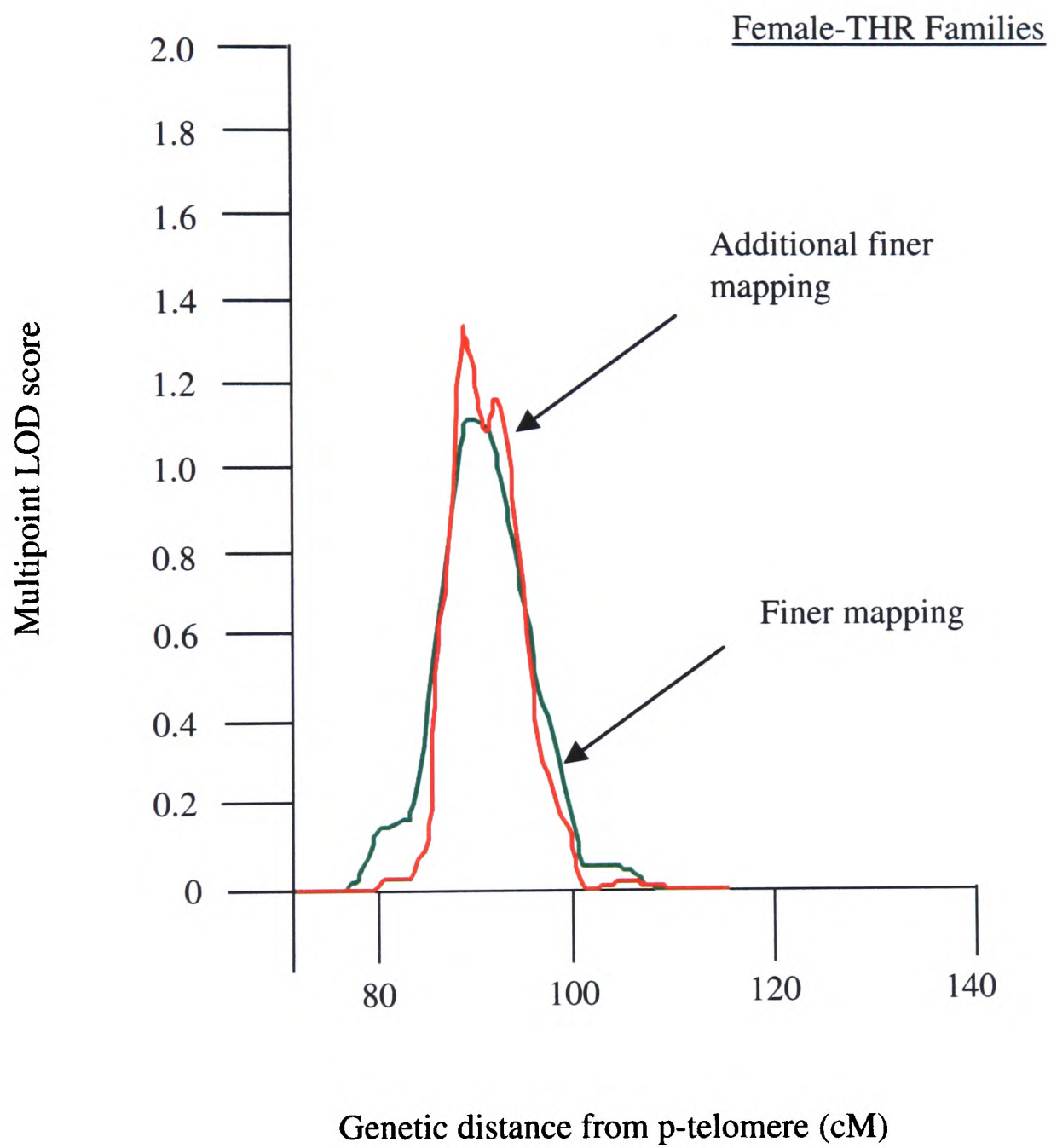


Figure 3.7d: Multipoint linkage analysis of chromosome 16, region 2 for the 146 female-THR families, with markers at an average of one marker every 1.9 cM, overlaid on the results of the finer mapping with one marker every 4 cM.

3.8 Discussion

The region showing evidence for linkage on chromosome 4 has now been mapped to an average density of one marker every 3 cM within the region of interest. Initial mapping, with markers at an average density of one marker every 10 cM, confirmed the presence of the region showing evidence of linkage to OA identified in the original scan by Loughlin et al (1999), with a multipoint LOD score (MLS) of 1.2 and 1.3 for the 218 female families and 146 female-THR families respectively. Finer mapping, with markers at an average distance of one marker every 4 cM resulted in increases in MLS to 2.4 for the female families and 2.7 for the female-THR families, with some narrowing of the region. The addition of extra markers in the centre of the region of interest for the female-THR families has allowed the region to be further narrowed to a 1-LOD support interval of only 4 cM with an increase in MLS to 3.1. This region is located at 4q13.1-4q13.2, with a genetic position between 76 cM and 80 cM from the chromosome 4 p-telomere. This interval represents approximately 6 Mb of DNA.

Two other groups have also published reports of linkage of an OA phenotype to chromosome 4. The first group, Roby et al (1999) investigated a large South African kindred of Dutch origin with familial bilateral dysplasia of the hip and an accompanying secondary osteoarthritis. Termed Beukes hip dysplasia (BHD), the disease displays an autosomal dominant mode of inheritance with evidence for incomplete penetrance. A genome-wide linkage scan revealed an 11 cM region on chromosome 4q35 as being the likely location for the BHD gene with a maximum two-point LOD score of 5.7.

The second group, Leppavuori et al (1999) performed a two stage genome-wide scan in twenty-seven Finnish sibships with severe, radiologically defined distal interphalangeal

joint (DIP) OA. In addition to a number of other regions they identified a region showing evidence for linkage on chromosome 4 at 4q26-q27 with a Z value of 1.9.

Initial mapping of the region, first identified by Loughlin et al (1999), on chromosome 4 with the MD10 markers allowed us to exclude any overlap of our region of linkage with those identified by these two groups. See figure 3.8a for positions of these regions relative to the region identified by this scan. The differences in the positions of the regions showing evidence for linkage in these studies probably represent the heterogeneity of disease with different genes contributing to susceptibility in different OA phenotypes (characterised by the involvement of different joints, differences in age of onset etc) as has been suggested by epidemiological studies (Spector et al. 1996; Hirsch et al. 1998).

The two regions, identified by Loughlin et al (1999), on chromosome 16 have also been finer mapped to an average density of one marker every 2 cM. Initial mapping with markers at an average density of one marker every 9 cM confirmed the presence of two regions showing evidence of linkage in the female families, both with MLS of 1.4. Subsequent analysis with markers at an average density of one marker every 4 cM resulted in an increase in MLS to 1.6 for the first region for the 146 female-THR families and to 1.7 for the second region for all 218 female families. There was also some narrowing of the second region. Finally, analysis with a higher density of markers, at an average of one marker every 1.8 cM in the first region and of one marker every 1.9 cM in the second region, resulted in an increase in MLS to 1.7 for the first region in the female-THR families and to 1.9 for the second region in all 218 female families. The 1-LOD support intervals of the two regions are now 12 cM and 10 cM respectively. The first region is located at 16p12.3-16p12.1, with a genetic position between 40 cM and 52 cM from the chromosome 16 p-telomere. The second is located at 16q22.1-16q23.1, with a genetic position between 85 cM

Chromosome 4

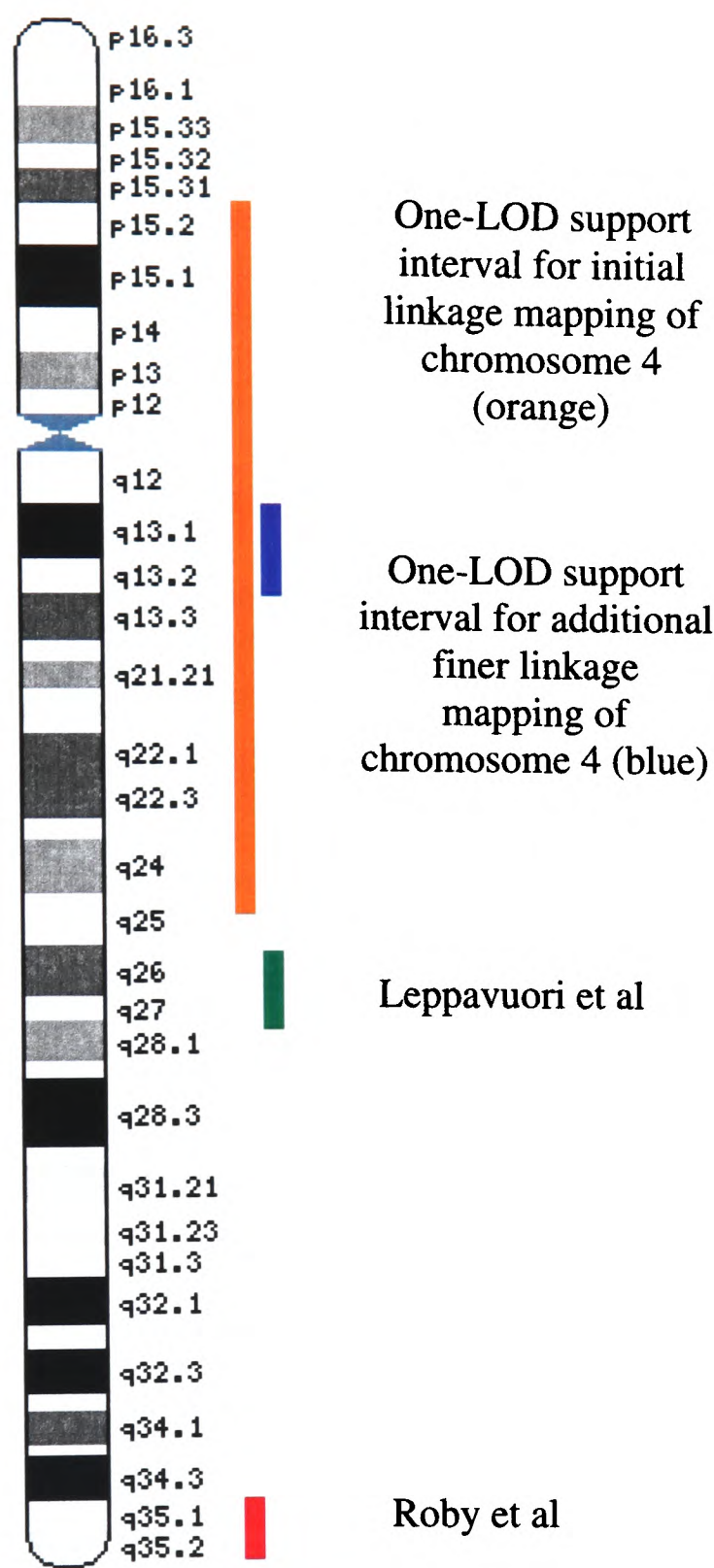


Figure 3.8a: The approximate positions of the loci identified by Roby et al (1999) and Leppavuori (1999) et al in relation to the region identified on chromosome 4 in the linkage scan reported here.

and 95 cM from the chromosome 16 p-telomere. These two regions on chromosome 16 represent approximately 8 Mb and 9 Mb of DNA sequence respectively.

A second group have also reported linkage of an OA phenotype with a region on chromosome 16p. Ingvarsson et al (2001) report the results of a genome-wide linkage scan in a large Icelandic pedigree where at least four generations have been affected with severe hip OA. The OA in this pedigree is characterised by an early onset (before 35 years of age) and there is reportedly no evidence for an underlying dysplasia or other joint disease, with symptoms being indistinguishable from those of primary hip OA. Their linkage scan identified a region showing evidence for linkage on chromosome 16p with a MLS of 2.6. The 1-LOD support interval for this region is 19 cM residing between markers D16S500 and D16S3068. The large overlap of this region (figure 3.8b) with the telomeric end of the region identified in the analysis detailed in this chapter on chromosome 16p may represent independent replication of results. Although overlap of these two regions is not complete it is possible that they represent the same susceptibility region. A simulation study by Roberts et al (1999) found that location estimates for the same gene might vary considerably between studies especially where the expected LOD score is modest. This replication increases confidence in the existence of an OA susceptibility gene within this region.

The addition of extra families has caused a decrease in the maximum LOD score of the regions identified both on chromosome 4 and 16 compared with the results of the initial genome-wide scan by Loughlin et al (1999). This is not totally unexpected and similar results have been reported by other groups working on complex traits. For example, Tavigian et al (2001) working on pedigrees with prostate cancer found evidence of linkage to chromosome 17 decreased with the addition of extra families since not all affected individuals within the additional pedigrees shared the same haplotype. Further studies by

Chromosome 16

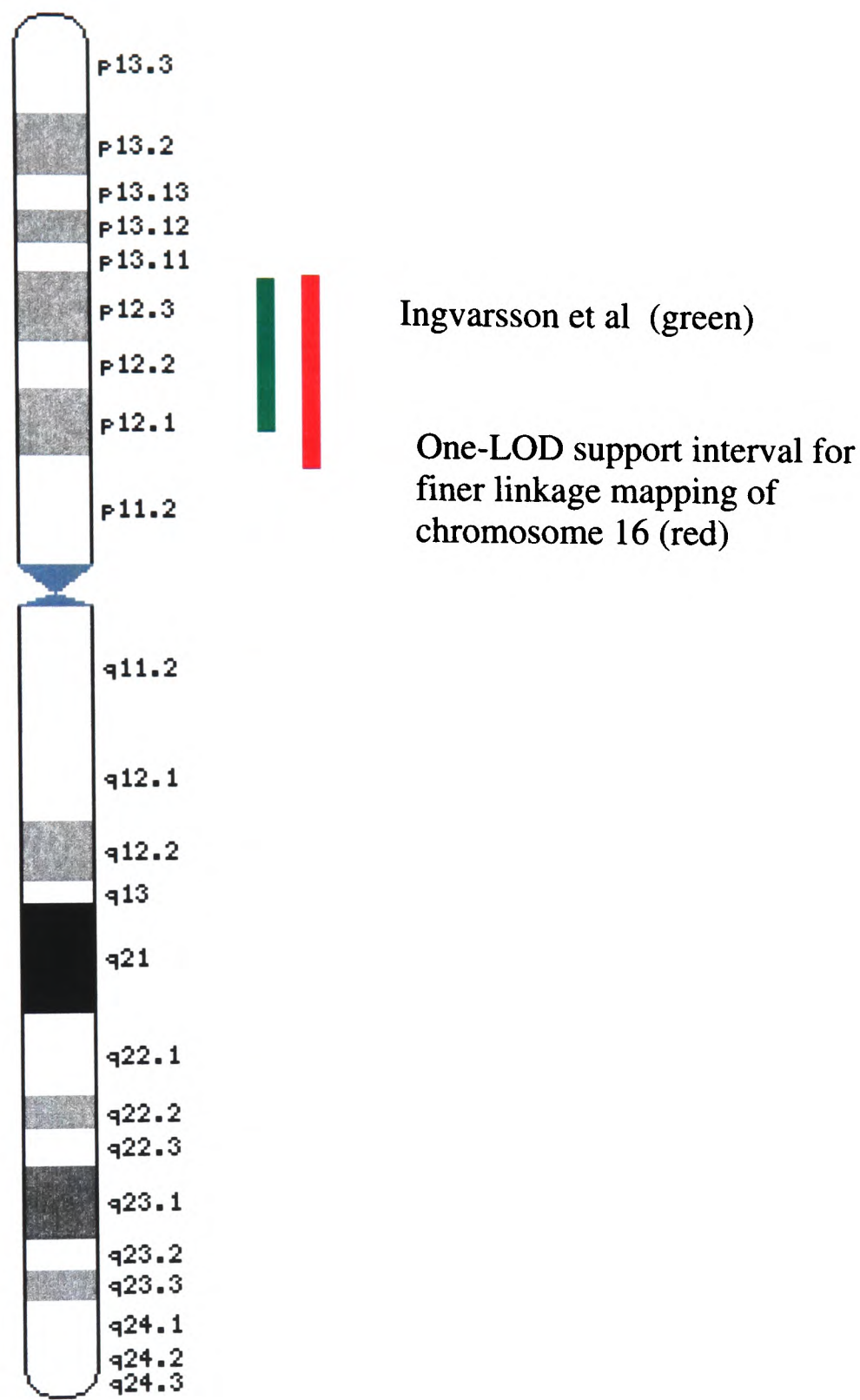


Figure 3.8b: The approximate positions of the loci identified by Ingvarsson et al (2001) in relation to the region identified on chromosome 16 in the linkage scan reported here.

this group, though identified disease associated mutations of a gene within the linked region in at least a subset of pedigrees increasing confidence that this was a genuine susceptibility locus.

In the case of our study on OA it is possible in the work reported here that in the initial scan by Loughlin et al (1999) there was an over-representation of families contributing to evidence of linkage at chromosome 4 or 16, which allowed for identification of the regions in the first place. In this case, the addition of extra families may now have brought the LOD scores towards their genuine levels.

In addition, the LOD score of 3.9 for chromosome 4 in the original study appears to have been accounted for largely by a single marker, D4S231. This marker was not used in subsequent work due to unreliability of its genetic position in more recent maps. However, in the additional finer mapping reported here, four separate markers within the region of linkage on chromosome 4 show two-point LOD scores of >1.5 .

In conclusion, the regions on chromosome 4 and 16 have now been sufficiently narrowed to allow for association analysis.

Chapter 4 - Conditioning Analysis

4.1 Introduction

Complex traits are characterised by the contribution of multiple genes and environmental factors to disease susceptibility. In addition, there often exists a series of epistatic and/or synergistic interactions both between different genes and between genes and the environment (Smith and Lusi 2002). Complex trait analysis can also be further confounded by the existence of reduced penetrance, variable expressivity and pleiotropy (Glazier et al. 2002). The power of linkage studies to detect the genes encoding susceptibility to complex traits depends on a number of factors including the increased risk associated with the gene of interest, the existence and nature of interactions and the significance threshold applied (Weeks and Lathrop 1995).

Genome-wide scans for a number of complex traits have often produced conflicting results with concordance between regions of linkage identified by different studies generally rare. Although some of these conflicts may reflect the lack of statistical power of some studies due to small sample sizes, or the existence of false positive results, many likely reflect the etiological complexity and heterogeneity of complex traits (Wiltshire et al. 2001). For example, the existence of locus heterogeneity may mean that evidence for linkage detected in one subset of families will not be replicated in an independent set of families where a different gene(s) contributes to disease susceptibility.

The limited power of linkage studies as a result of such complexities may be overcome to some extent by the use of large samples, careful selection of participants to minimise ethnic and clinical diversity and the use of specifically developed statistical approaches (Weeks and Lathrop 1995; Wiltshire et al. 2001). In addition,

statistical exploratory methods such as conditioning analysis (Cox et al. 1999) may be used to provide increased power.

Studies in animal models have provided clear evidence for regions which showed no significant evidence for linkage in primary scans but which have been shown to contribute substantially to a trait when analysed allowing for gene-gene or gene-environment interactions (Kuida et al. 2000; Leiter et al. 1998)

Conditioning analysis is one method which allows for the re-evaluation of evidence for linkage in one region taking in to account the evidence for linkage at a second region, unlinked to the first. It was designed in order to increase the power of linkage studies by taking account of the involvement of multiple genes in encoding susceptibility for complex traits (Cox et al. 1999).

As a preliminary stage to test for a statistical interaction between two regions, the family non-parametric LOD (NPL) scores, calculated on the basis of allele sharing, at the two unlinked regions, are tested for correlation. Correlation studies test the hypothesis that allele sharing at the two regions is independent within families. Correlation is expected only if both regions considered contribute to disease susceptibility and significant correlation therefore serves to increase confidence in the involvement of the regions in disease susceptibility. Positive correlation indicates a significant statistical interaction between the two regions and may be indicative of an interaction between genes in the two regions e.g. epistasis or an additive effect. Negative correlation can be an indicator of genetic heterogeneity.

Any regions showing evidence of correlation may then be assessed using conditioning analysis. Linkage analysis is undertaken for the first (“conditioned”) locus with each family assigned a weight according to their NPL score at the second (“conditioning”) locus (Cox et al. 1999). To model for positive interactions families with an NPL score of 0 or less are assigned a weight of 0 and families with a positive NPL score a weight of 1. To model negative interactions families with a positive NPL score are given a weight of 0 and those with an NPL score of 0 or less a weight of 1.

Where an increase in LOD score is observed following this weighted analysis the significance must be assessed empirically (Cox et al. 1999). A significant increase in LOD score serves to increase confidence in the genuineness of the regions as encoding susceptibility to disease. In addition, such conditioning studies may also serve to narrow regions showing evidence for linkage (Horikawa et al. 2000).

A number of studies have reported on the use of conditioning analysis in complex trait investigation (e.g. Cox et al. 1999; Wiltshire et al. 2001; Xu et al. 2001). Both Cox et al (1999) and Wiltshire et al (2001) have used such analyses to further investigate regions identified in genome-wide scans as showing evidence for linkage to type II diabetes. The study by Cox et al (1999), who investigated two regions, NIDDM1 and CYP19, is summarised in chapter 1.9. Briefly, they found increased evidence for linkage, with an increase in LOD score from 4.0 to 7.3, at region NIDDM1 when they weighted Mexican American families proportionately according to their evidence for linkage at region CYP19. This result has an uncorrected empirical p-value of 0.0001 providing highly significant evidence for a statistical interaction between the two loci. A similar increase was also seen in the LOD score for region CYP19 when

conditioned on NIDDM1. These two regions represent the only two type II diabetes regions to be replicated in an independent study of Mexican Americans (Hanis et al. 1996) increasing confidence in the genuineness of both regions as type II diabetes susceptibility regions. This conditioning analysis also served to decrease the 1-LOD support interval of NIDDM1 from 12 to 7 cM (Horikawa et al. 2000). A gene within this narrowed region, CAPN10, has since been implicated in susceptibility to type II diabetes in Mexican Americans highlighting the potential usefulness of conditioning analysis in complex trait susceptibility gene detection (Horikawa et al. 2000).

The study by Wiltshire et al (2001) investigated the correlation between NPL scores for regions showing evidence for linkage in a genome-wide scan for type II diabetes. They identified a positive correlation between regions on chromosomes 1q and 10q and a negative correlation between the region on 1q and a region on chromosome 5q. Subsequent conditioning analysis using weights of 0 and 1 revealed an increase in LOD, from 2.0 to 3.1, for the region on chromosome 1q weighted on chromosome 10q and, from 2.0 to 2.8, for chromosome 10q weighted on chromosome 1q, both under a model of a positive interaction. These results have p-values of 0.017 and 0.032 respectively. Under a model of negative correlation, there was an increase in LOD from 2.0 to 3.0 ($P = 0.023$) for the region on chromosome 1q when conditioned on chromosome 5q (Wiltshire et al. 2001). Again, as for Cox et al (1999), these results serve to increase confidence in the genuineness of these regions as containing susceptibility encoding genes.

4.2 Conditioning analysis

Correlation analysis was undertaken as the first step in testing for an interaction between the regions identified as showing evidence for linkage to OA on chromosomes 4q, 16p and 16q. Family NPL scores from the regions showing maximum evidence for linkage, as defined by Allegro (Gudbjartsson et al. 2000) analysis, were tested for correlation by calculation of the correlation coefficient r . This is a relatively quick test compared with conditioning analysis and allows a number of regions to be tested easily. Any regions showing evidence for a statistical interaction ($P \leq 0.05$) in these preliminary correlation studies were then analysed using conditioning analysis in an attempt to increase LOD score and narrow the linkage region.

4.2.1 Correlation calculations

Family NPL scores at the regions showing maximal evidence for linkage on chromosomes 4q, 16p and 16q were calculated using Allegro for the 146 female-THR families used in the linkage analysis (NPL scores are listed in appendix 2). Although the region on 16q showed a higher LOD score in all 218 female families, and was not restricted to the 146 female-THR families, NPL calculations were initially restricted to the 146 female-THR families to enable comparison between this region and the regions on chromosome 4q and 16p. Figures 4.2a-c show the scatter plots of the NPL scores for the region on chromosome 4q plotted against the NPL scores for both the region on chromosome 16p and the region on chromosome 16q and the NPL scores for the region on chromosome 16p plotted against the NPL scores for the region on 16q. From observation of the scatter plots it appears that there is little correlation

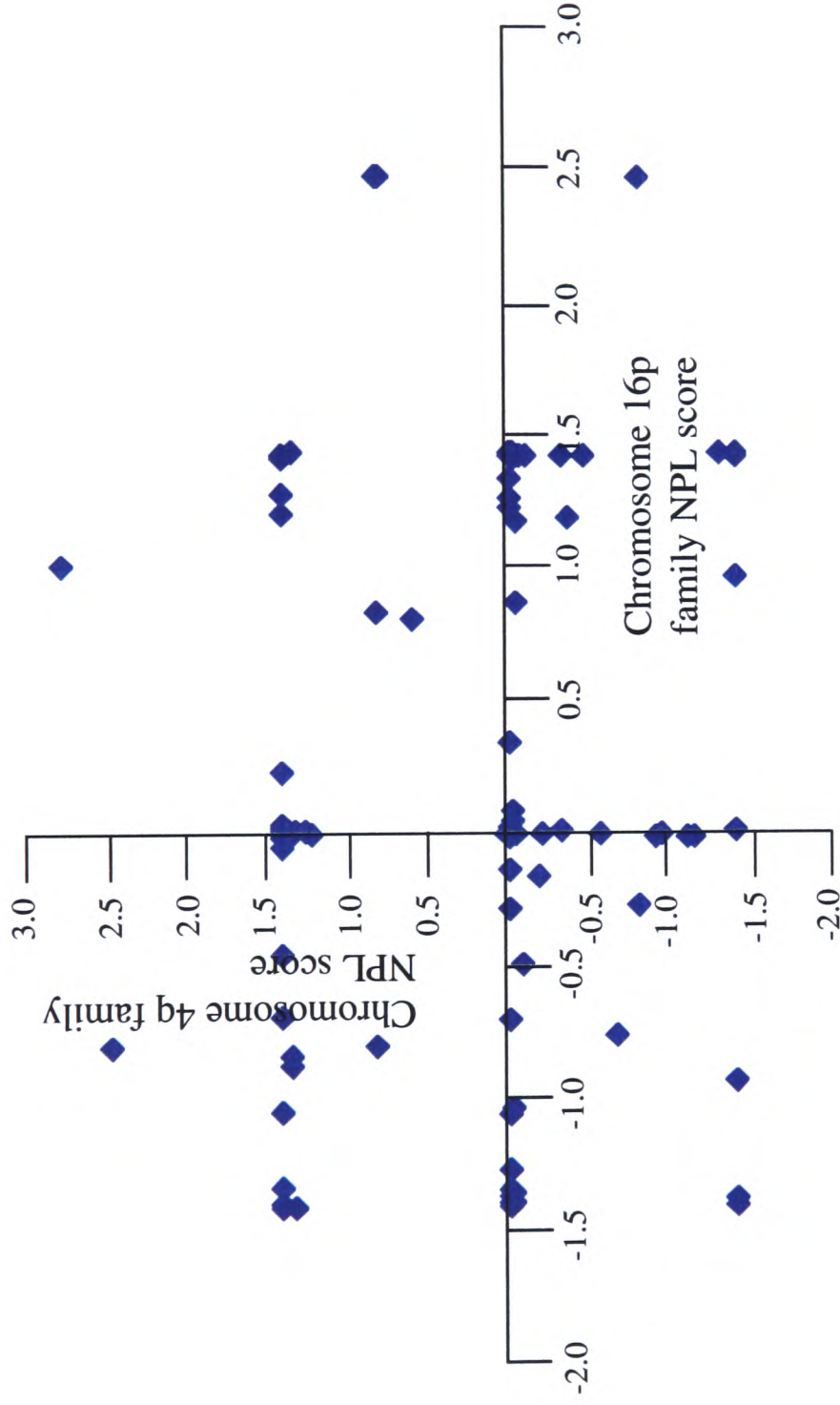


Figure 4.2a: Scatter plot showing relation between family NPL scores of the 146 female-THR families at the regions showing maximum evidence for linkage on chromosomes 4q and 16p. Each diamond represents one family. Diamonds in top right sector represent families with positive NPL scores for chromosome 4q and 16p, diamonds in bottom right sector represent families with a positive NPL score for chromosome 16p and a negative NPL score for chromosome 4q, diamonds along the horizontal axis represent families with a NPL score of zero for chromosome 4q etc.

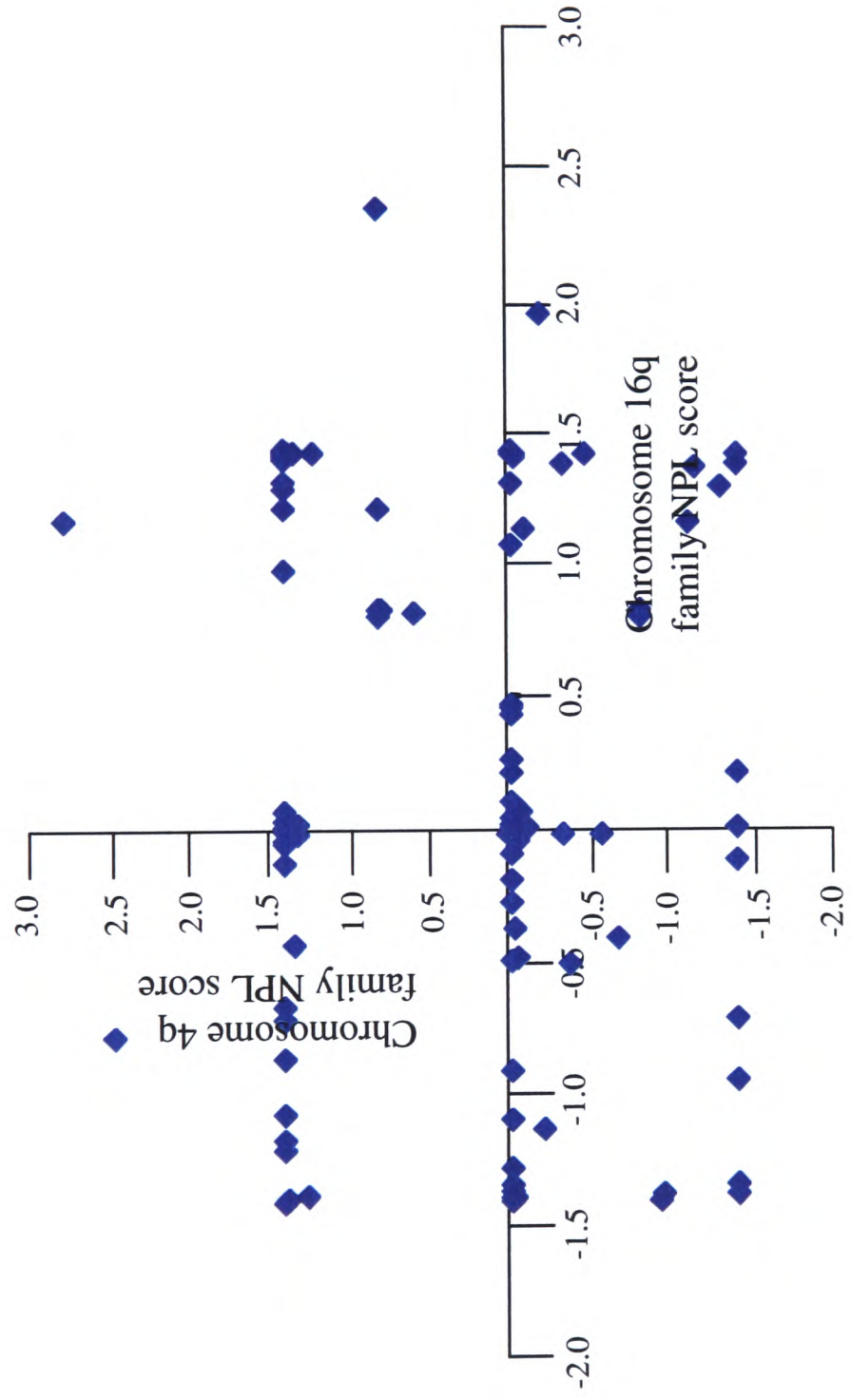


Figure 4.2b: Scatter plot showing relation between family NPL scores for the 146 female-THR families at the regions showing maximum evidence for linkage on chromosomes 4q and 16q. Each diamond represents one family. Diamonds in top right sector represent families with positive NPL scores for chromosome 4q and 16q, diamonds in bottom right sector represent families with a positive NPL score for chromosome 16q and a negative NPL score for chromosome 4q, diamonds along the horizontal axis represent families with a NPL score of zero for chromosome 4q etc.

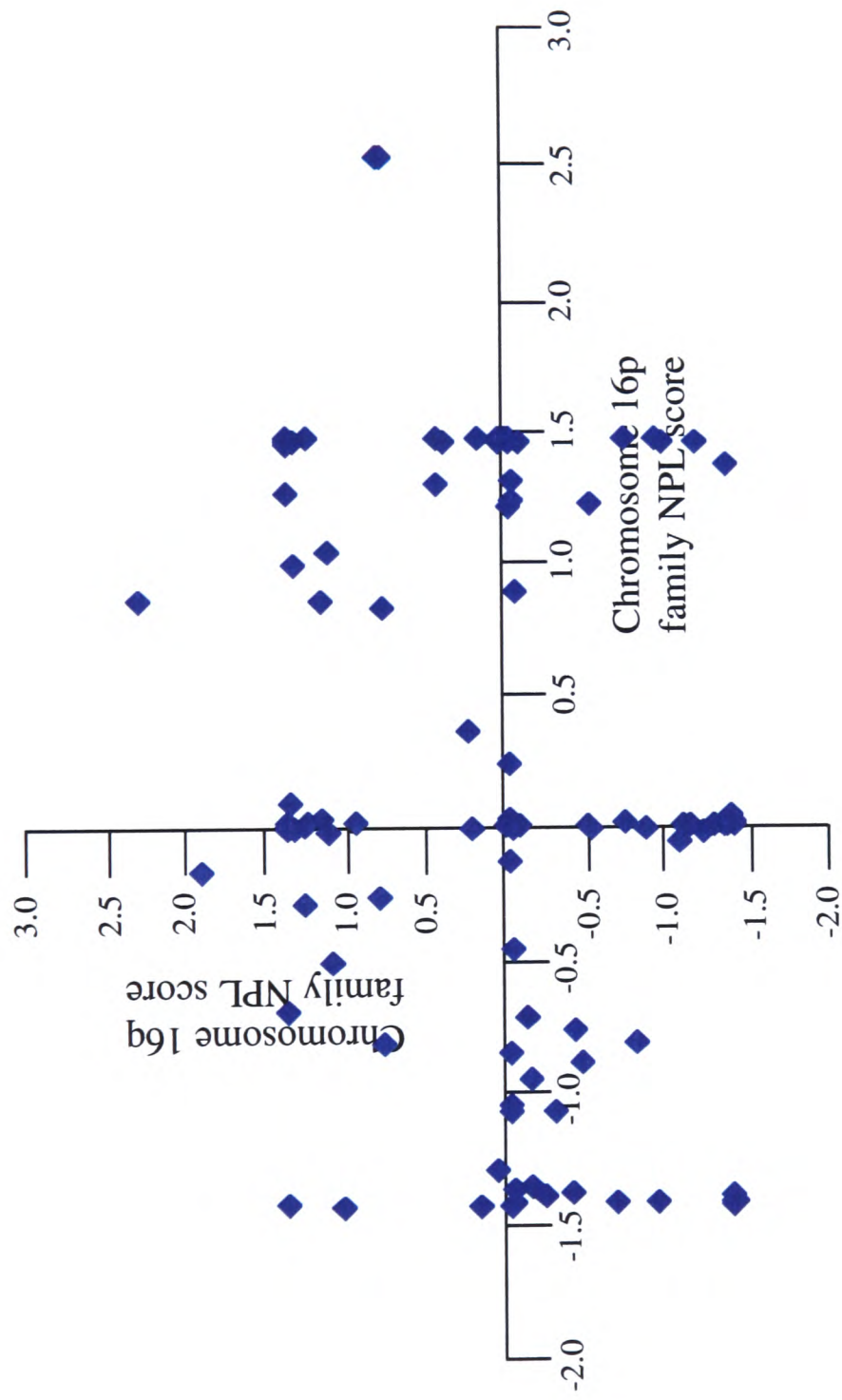


Figure 4.2c: Scatter plot showing relation between family NPL scores for the 146 female-THR families at the regions showing maximum evidence for linkage on chromosomes 16p and 16q. Each diamond represents one family. Diamonds in top right sector represent families with positive NPL scores for chromosome 16q and 16p, diamonds in bottom right sector represent families with a positive NPL score for chromosome 16p and a negative NPL score for chromosome 16q, diamonds along the horizontal axis represent families with a NPL score of zero for chromosome 16q etc.

between the NPL scores for either the regions on chromosomes 4q and 16p or the regions on chromosomes 4q and 16q with points in all four sectors and along the axes (figure 4.2a,b). There appears to be a higher degree of correlation between the NPL scores for the two chromosome 16 regions with more points tending towards the positive diagonal (figure 4.2c). To formally test the statistical evidence for correlation the correlation coefficient r , p -values and 95% confidence intervals were calculated for all three comparisons. Neither of the comparisons between the region on 4q and the regions on chromosomes 16p or 16q shows significant evidence for correlation. Comparison between the NPL scores for the region on chromosome 4q and the region on chromosome 16p gives a correlation coefficient of -0.003 ($P = 0.98$, 95% confidence interval -0.6 to 0.16) whilst comparison of NPL scores for the region on chromosome 4q and the region on chromosome 16q results in a correlation coefficient of 0.04 ($P = 0.63$, 95% confidence interval -0.12 to 0.20). Comparison between the region on chromosome 16p and the region on chromosome 16q provides significant evidence for a correlation between these two regions with a correlation coefficient of 0.31 ($P = 0.0001$, 95% confidence interval 0.16 to 0.45). Even after correction for multiple testing (multiplying the p -value by 3, the number of tests) and using the strict criteria suggested by Cox et al (1999) of a p -value lower than 0.01 this result remains highly significant at $P = 0.0003$. The confidence interval however, is fairly wide indicating the sample size is not sufficient to accurately predict the degree of correlation.

Since the locus on chromosome 16q had a higher LOD score in all 218 female families and was not restricted to the 146 female-THR families, family NPL scores at the regions showing maximal evidence for linkage were calculated, using Allegro, for

the regions on chromosomes 16p and 16q in all 218 female families (it was not possible to calculate the NPL scores for all 218 female families for the region on chromosome 4 since only the 146 female-THR families were used in the additional finer mapping of this chromosome). If a correlation between NPL scores exists, this would allow for conditioning analysis of the region on chromosome 16q, weighted on the NPL scores at the region on chromosome 16p, for all 218 female families. Family NPL scores for all 218 female families at the regions of maximum linkage on chromosomes 16p and 16q are listed in appendix 3. Figure 4.2d shows the scatter plot of the NPL scores for the region on chromosome 16p plotted against the NPL scores for the region on chromosome 16q. Again, as for the plot of 16p versus 16q for the 146 female-THR families there is a clustering of points towards the positive diagonal. Formal comparison of the NPL scores gives a correlation coefficient of 0.28 ($P < 0.0001$, 95% confidence interval 0.16 to 0.40). Again, this provides highly significant evidence for a statistical interaction between these two regions, this time in the 218 female families. The confidence interval however, is again wide indicating that the sample size is still too small to accurately predict effect size.

4.2.2 Conditioning analysis

Conditioning analysis was undertaken for the two regions on chromosomes 16p and 16q. Firstly, conditioning analysis with chromosome 16p as the “conditioning” locus and 16q as the “conditioned” locus was undertaken in all 218 female families (as these were the families showing maximal evidence for linkage for the region on 16q [chapter 3.7.2]). Secondly, conditioning analysis was undertaken with chromosome 16q as the “conditioning” locus and 16p as the “conditioned” locus in the 146 female-THR families (as these were the families showing maximal evidence for linkage for

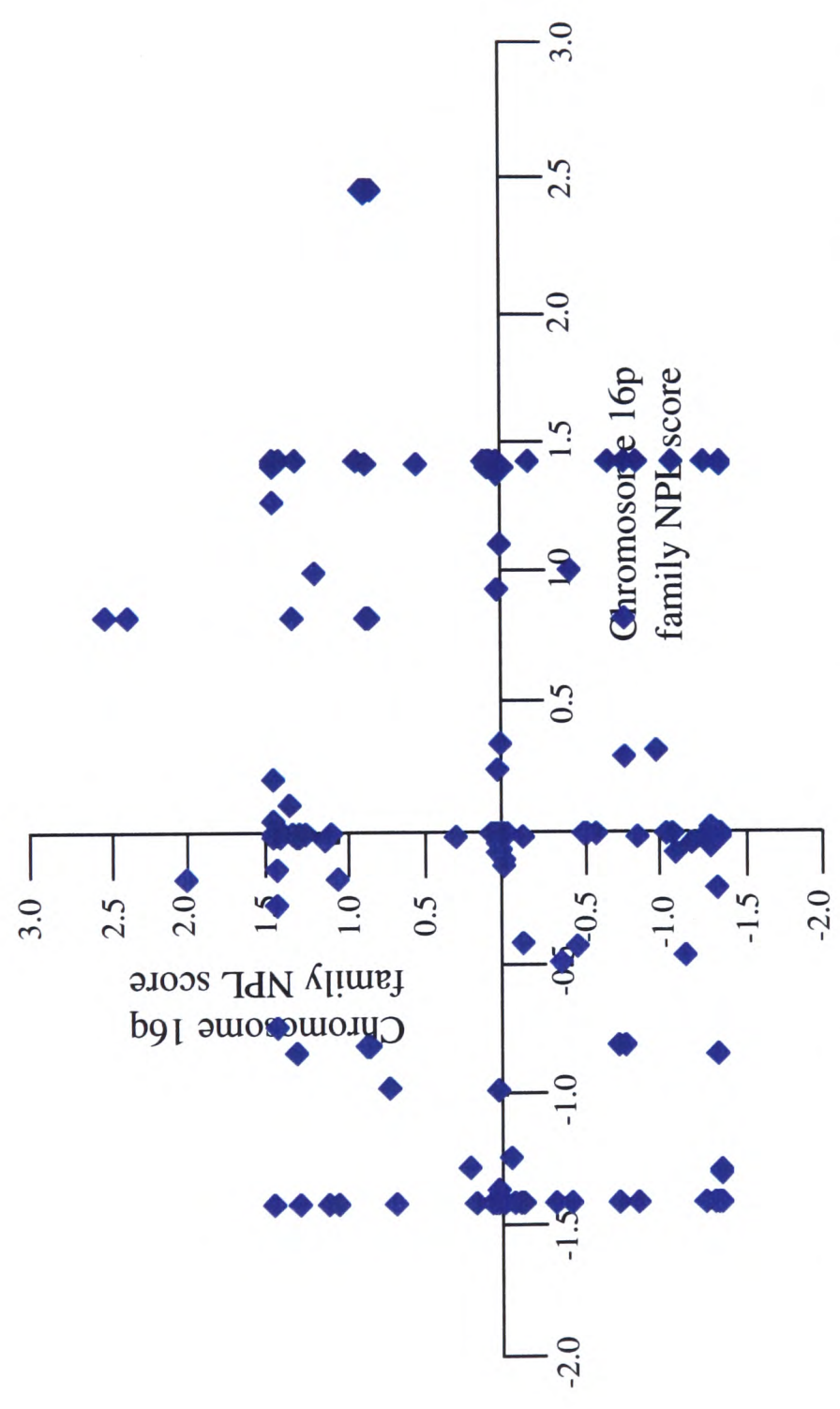


Figure 4.2d: Scatter plot showing relation between family NPL scores for the 218 female families at the regions showing maximum evidence for linkage on chromosomes 16p and 16q. Each diamond represents one family. Diamonds in top right sector represent families with positive NPL scores for chromosome 16q and 16p, diamonds in bottom right sector represent families with a positive NPL score for chromosome 16p and a negative NPL score for chromosome 16q, diamonds along the horizontal axis represent families with a NPL score of zero for chromosome 16q etc.

the region on chromosome 16p [chapter 3.7.1]). For each comparison analysis was undertaken to model for both positive and negative interactions.

The results of conditioning analysis modelling for positive interactions (families with NPL score ≤ 0 assigned weight of 0) are shown in figures 4.2e and f. For the region on chromosome 16p conditioning analysis, weighted according to the NPL scores for the region on chromosome 16q, in the 146 female-THR families, has resulted in an increase in multipoint LOD score (MLS) from 1.1 (unweighted) to 4.4 (weighted) [the difference between LOD scores reported here and those reported in chapter 3 are due to the use of different statistical packages - ASPEX in chapter 3 and Allegro here]. There has also been a narrowing of the 1-LOD support interval from 20 cM to 7 cM. For the region on chromosome 16q conditioning analysis, weighted according to the NPL scores for the region on chromosome 16p, in the 218 female families, has resulted in an increase in MLS from 1.7 (unweighted) to 5.7 (weighted). Again, there has been a narrowing of the 1-LOD support interval from 10 cM to 5 cM. In order to calculate the significance of these increases in LOD score 10,000 permutations (involving the permutation of family-specific weights) were carried out. This gives an empirical p-value of <0.001 for the region on chromosome 16p and <0.001 for the region on chromosome 16q.

Next, conditioning analysis modelling for a negative interaction was undertaken (families with NPL score ≥ 0 assigned weight of 0). For the region on chromosome 16p conditioning analysis, weighted according to the NPL scores for the region on chromosome 16q, in the 146 female-THR families, resulted in a decrease in MLS from 1.1 (unweighted) to 0 (weighted). For the region on chromosome 16q

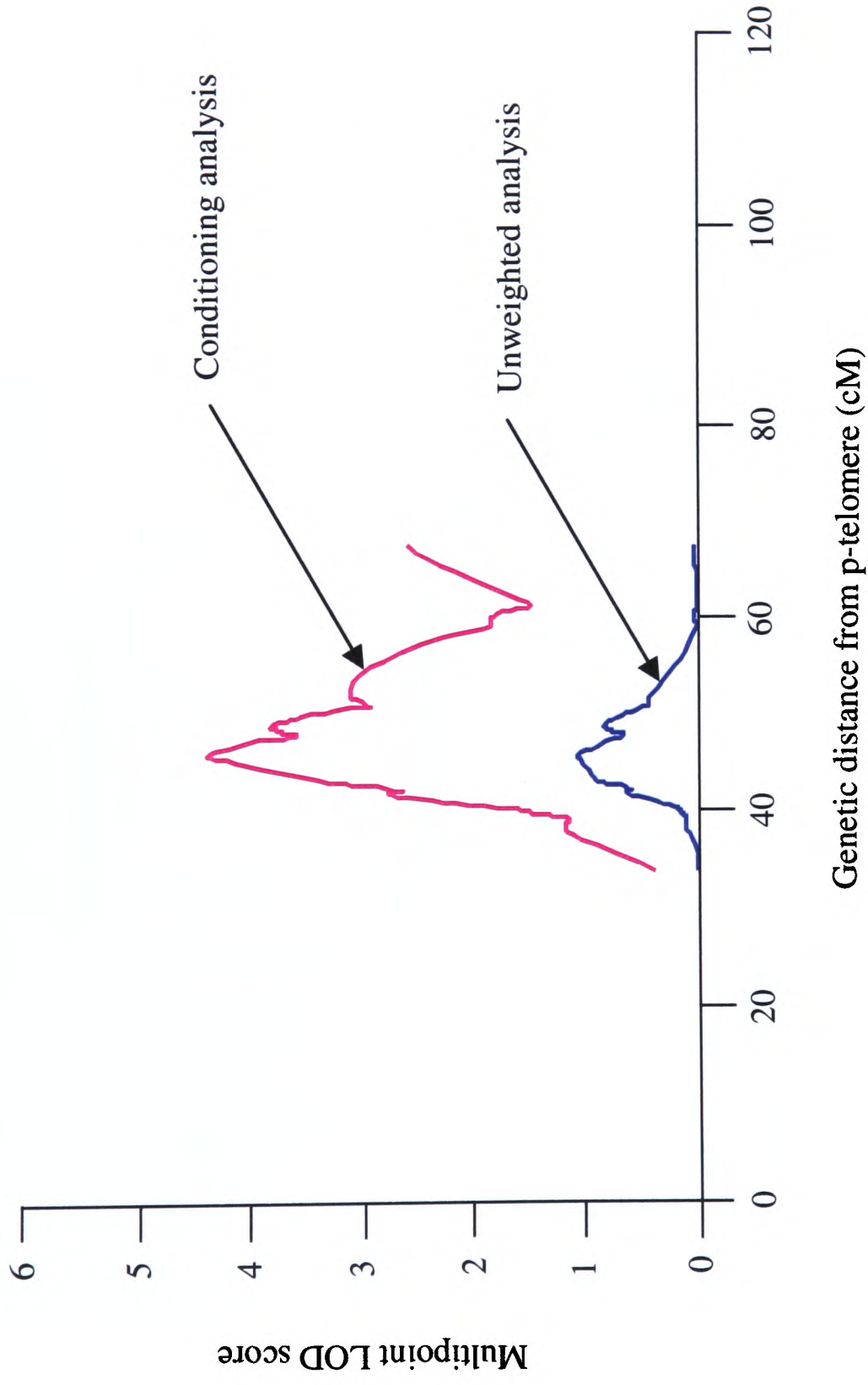


Figure 4.2e: Results of conditioning analysis, under a positive interaction model, for the region on chromosome 16p conditioned on the NPL scores for the peak on 16q in the 146 female-THR families.

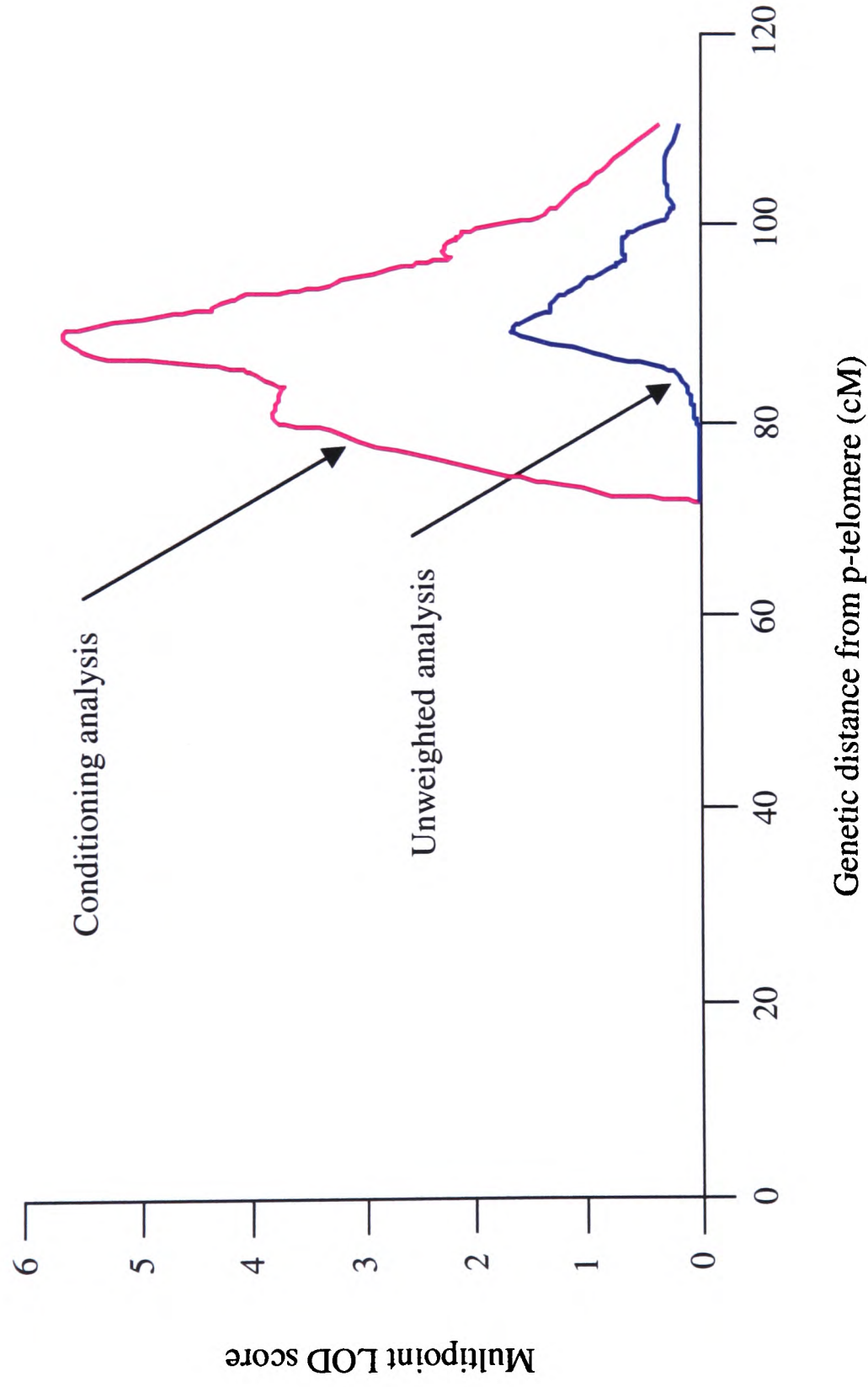


Figure 4.2f: Results of conditioning analysis, under a positive interaction model, for the region on chromosome 16q conditioned on the NPL scores for the peak on 16p in the 218 female families.

conditioning analysis, weighted according to the NPL scores for the region on chromosome 16p, in the 218 female families, resulted in a decrease in MLS from 1.7 (unweighted) to 0.01 (weighted).

As there is no evidence of correlation between the NPL scores for the regions on chromosome 4q and chromosome 16p or between the regions on chromosome 4q and chromosome 16q, conditioning analysis was not undertaken for these regions.

4.3 Discussion

Calculation of the correlation coefficient r between family specific NPL scores for the regions on chromosomes 16p and 16q revealed evidence for a highly significant positive correlation between the two regions in both the 146 female-THR families and the 218 female families.

Conditioning analysis in the 146 female-THR families resulted in an increased MLS from 1.1 to 4.4, for the region on chromosome 16p when conditioned on the NPL scores for the region on chromosome 16q using a positive interaction model. This increase is highly significant with an empirical p -value of <0.001 . This result shows that the same families are contributing to the evidence for linkage at the region on chromosome 16p and the region on chromosome 16q. This may be indicative of some form of epistasis or interaction between susceptibility genes in these two regions or be the result of genes in these regions having an additive effect.

Conditioning analysis in the 218 female families resulted in an increased MLS, from 1.7 to 5.7, for the region on chromosome 16q when conditioned on the NPL scores for

the region on chromosome 16p using a positive interaction model. Again, this is a highly significant result with a p-value of <0.001 . As for the conditioning analysis of the region on chromosome 16p, this result indicates that the same families are contributing to evidence for linkage at the regions on chromosomes 16p and 16q.

In addition to the increases in the MLS, the conditioning analysis has also served to narrow the two regions. The 1-LOD support interval for chromosome 16p has been narrowed from 20 cM to 7 cM, whilst the 1-LOD support interval for chromosome 16q has been narrowed from 10 cM to 5 cM.

Conditioning analysis modelling for a negative interaction between the two regions on chromosome 16 showed no evidence for a negative interaction between the two regions, with a decrease in MLS for both the 218 female families and the 146 female-THR families.

Whilst the increases in MLS observed under the positive interaction model serve to increase the confidence of the regions on chromosomes 16p and 16q as containing susceptibility genes, since both regions are on the same chromosome, the results must be interpreted with caution. In order for conditioning analysis to be effective it is necessary for the two regions under consideration to be unlinked. Although the centre of the two regions on chromosome 16p and 16q are approximately 44 cM apart (and therefore there would be expected to be close to random segregation for these two regions) there is only approximately 20 cM between the centromeric edges of the two regions. This makes it unlikely that the two regions are totally unlinked. This could

account for some or all of the increase in LOD score observed with conditioning analysis.

Calculation of the correlation coefficient for family NPL scores for both chromosome 4q versus chromosome 16p and for chromosome 4q versus chromosome 16q revealed no evidence for a significant statistical correlation between the regions. This result does not exclude the possibility of an interaction between the regions however, as some biological interactions would not be expected to result in detectable correlations (Cox et al. 1999).

Conditioning analysis between the regions on chromosomes 4, 16p and 16q and the other regions identified by our group, on chromosomes 2, 6 and 11, was undertaken by another member of the group in the 146 female-THR families (Dr Kay Chapman, personal communication). No evidence for any interaction was observed under either positive or negative interaction models.

Chapter 5 - ADAMTS3 candidate gene identification

5.1 Introduction: Candidate gene identification on chromosome 4q

Association analysis is emerging as the method of choice for detecting the genetic variation encoding susceptibility to complex disease. It has increased power compared with linkage analysis for detecting alleles with small effects. Risk alleles with a frequency of 1% can readily be detected in association studies but not in linkage studies (Botstein and Risch 2003). However, due to the prohibitively large number of markers that would be required to carry out genome-wide association studies (and the concurrent cost) researchers have tended to focus on the use of association studies to investigate candidate regions identified by linkage studies (Kaplan and Morris 2001a, b). Genome-wide association studies have been recently reported, however (Sawcer et al. 2004).

Candidate region association analysis may be systematic, covering the whole of the region under investigation, or candidate gene based, identifying variation in genes prioritised according to their potential role in disease biology. Both methods have their advantages. The greater number of markers necessary for the systematic method compared with the candidate gene based approach increases both the cost and time of the study. However, the systematic approach has the advantage of requiring no prior assumption as to the gene(s) likely to be involved in a disease process which may often be poorly understood. Systematic approaches also enable regulatory elements distal to the coding sequence to be covered (Vink and Boomsma 2002; Botstein and Risch 2003).

The region on chromosome 4 identified by finer linkage mapping (chapter 3.4) is still too large to allow for an economically viable systematic association analysis, with the 1-LOD support interval spanning approximately 6 Mb of DNA and at least eighteen known genes

(in addition to fifty hypothetical or predicted genes). Candidate gene-based association studies were therefore undertaken for this region.

5.1.1 Cataloguing of genes within the linkage region

Genes within the linkage region identified on chromosome 4 (chapter 3.4) were identified using the three major databases of Ensembl, NCBI and UCSC. Initial investigation concentrated on the 1-LOD support interval (4q12-4q13, 63 Mb-69 Mb). There were however, at the time of candidate gene selection, no obvious candidates within this interval. Investigation of genes outside this region was therefore carried out in a stepwise process. Table 5.1 details an integrated list from all three databases of all confirmed genes within the 1-LOD support region (marked by double lines) in addition to genes approximately 5 cM either side of this region, correct as of July 2003. Databases are frequently updated and therefore tables of genes within a region and their positions relative to genetic markers constantly refined. A number of the genes detailed in the databases are hypothetical and are not included.

5.1.2 Identification of candidate genes

Many of the genes within the interval have not yet been fully characterised and, as such knowledge about their function(s) is incomplete. Even for those genes that have been highly investigated novel functions are still discovered, often resulting from alternative splicing (Grabowski 1998; Smith and Valcarcel 2000). Knowledge of expression patterns for many of the genes is also often limited with the tissues of the joint poorly represented in many expression studies. In an effort to overcome such limitations, investigations into gene function are being carried out, databases are being constantly updated and

Gene symbol	Name	LocusID ¹	Function
KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	3815	Type 3 transmembrane receptor for mast cell growth factor. Role in cell growth and signal transduction
KDR	Kinase insert domain receptor (a type III receptor tyrosine kinase	3791	Regulating angiogenesis-related functions. Vascular endothelial growth factor receptor activity. Role in signalling
TPARL ²	TPA regulated locus	55858	Function unknown
CLOCK	Clock homolog (mouse)	9575	Role in circadian rhythm regulation. Transcription factor activity. Role in signal transduction
NMU	Neuromedin U	19874	Neuropeptide signalling. Regulation of smooth muscle contraction. Digestion
SEC3 ²	Sec3-like	55763	Role in targeting exocytic vesicles to docking sites on plasma membrane
LOC132948 ²	2-aminoadipic 6-semialdehyde dehydrogenase	132948	Function unknown
PPAT	Phosphoribosyl pyrophosphate amidotransferase	5471	Catalyses first step of de novo purine nucleotide biosynthesis
PAICS	Phosphoribosylaminoimidazole carboxylase, phosphoribosylaminoimidazole succinocarboxamide synthetase	10606	Catalyses steps six and seven in the purine nucleotide biosynthesis pathway
SRP72	Signal recognition particle 72kDa	6731	Signal recognition and protein export
HOP ²	Homeodomain-only protein	84525	Associated with lung cancer, function unknown
SPINK2	Serine protease inhibitor, Kazal type, 2 (acrosin-trypsin inhibitor)	6691	Serine protease inhibitor activity, function unknown
REST	RE1-silencing transcription factor	5978	Represses neuronal genes in non-neuronal tissues
POLR2B	Polymerase (RNA) II (DNA directed) polypeptide B, 140kDa	5431	Part of RNA polymerase II complex responsible for synthesising mRNA in eukaryotes

IGFBP7	Insulin-like growth factor binding protein 7	3490	Negative regulation of cell proliferation
LPHN3	Latrophilin 3	23284	Synaptic calcium dependent alpha-latrotoxin (LTX) receptor.
EPHA5	EphA5	2044	Receptor tyrosine kinase
CENPC1	Centromere protein C1	1060	Centromere autoantigen and component of inner kinetochore plate
BRDG1	BRCA downstream signalling 1	26228	Docking protein in B cell antigen receptor signalling. Role in humoral immune response
GNRHR	Gonadotrophin-releasing hormone receptor	2798	Stimulates phosphoinositide turnover and membrane depolarisation. Role in development
HAT ²	Airway trypsin-like protease	9407	Possible role in host defence system on mucous membrane.
GRINL1B	Glutamate receptor, ionotropic, N-methyl D-aspartate-like 1B	84534	Function unknown
YT521 ²	Splicing factor YT521-B	91746	Splicing factor
DESC1 ²	DESC1 protein	28983	Proposed function of regulating or binding carbohydrate side chains. Enzymatic activity
UGT2B17	UDP glycosyltransferase 2 family, polypeptide B17	7367	Transfer of glucuronic acid to a variety of substrates including steroids. Role in steroid and lipid metabolism
UGT2B15	UDP glycosyltransferase 2 family, polypeptide B15	7366	Transfer of glucuronic acid to a variety of substrates including steroids. Role in steroid and lipid metabolism
UGT2B10	UDP glycosyltransferase 2 family, polypeptide B10	7365	Transfer of glucuronic acid to a variety of substrates including steroids. Role in steroid and lipid metabolism
UGT2B7	UDP glycosyltransferase 2 family, polypeptide B7	7364	Transfer of glucuronic acid to a variety of substrates including steroids. Role in steroid and lipid metabolism
UGT2B11	UDP glycosyltransferase 2 family, polypeptide B11	10720	Transfer of glucuronic acid to a variety of substrates including steroids. Role in steroid and lipid metabolism
UGT2B28	UDP glycosyltransferase 2 family, polypeptide B28	54490	Transfer of glucuronic acid to a variety of substrates including steroids. Role in steroid and lipid metabolism

UGT2B4	UDP glycosyltransferase 2 family, polypeptide B4	7363	Transfer of glucuronic acid to a variety of substrates including steroids. Role in steroid and lipid metabolism
UGT2A1	UDP glycosyltransferase 2 family, polypeptide A1	10941	Transfer of glucuronic acid to a variety of substrates including steroids. Role in steroid and lipid metabolism
SULT1B1	Sulfotransferase family, cytosolic, 1B, member 1	27284	Catalyses the sulphate conjugation of many hormones, neurotransmitters, drugs, and xenobiotic compounds.
STE	Sulfotransferase, oestrogen-preferring	6783	Catalyses sulphate conjugation of oestrogen
CSN1S1	Casein alpha s1	1446	Role in digestion
CSN2	Casein beta	1447	Milk protein
STATH	Statherin	6779	Roles in ossification and bone mineralisation regulation in teeth. Structural component of tooth enamel. Maintenance of oral health.
HTN3	Histatin 3	3347	Salivary protein. Non immunological, anti-microbial function in the oral cavity. Role in ossification
HTN1	Histatin 1	3346	Salivary protein. Non immunological, anti-microbial function in the oral cavity. Role in ossification
CSN1S2A	Casein alpha s2-like A	286828	Function unknown
C4orf7	Chromosome 4 open reading frame 7	260436	Expressed only in certain activated immune cells, function unknown
CSN3	Casein kappa	1448	Mammalian milk protein, function unknown
NYD-SP26 ²	Testis development protein NYD-SP26	85438	Function unknown
PROL5	Proline-rich 5 (salivary)	26952	Function unknown
PROL3	Proline rich 3	10879	Function unknown
PROL1	Proline rich 1	58503	Function unknown

MUC7	Mucin 7, salivary	4589	Involved in structure and biosynthesis of human salivary mucins
ENAM	Enamelin	10117	Structural component of tooth enamel. Role in odontogenesis
SAS10 ²	Disrupter of silencing 10	57050	Function unknown
RIPX ²	RAP2 interacting protein X	22902	Potential role in Ras-GTPase signalling pathways
GRSF1	G-rich RNA sequence binding factor 1	2926	Selectively recruits cellular and viral mRNA with 5'UTR GRSF-1 binding sites to polyribosomes
DCK	Deoxycytidine kinase	1633	Phosphorylation of several deoxyribonucleosides. Deficiency associated with drug resistance
SLC4A4	Solute carrier family 4, sodium bicarbonate cotransporter, member 4	8671	Small molecule transport
GC	Group-specific component (vitamin D binding protein	2638	Role in regulation of fasting insulin levels. ?Precursor to macrophage-activating factor. ?Role in thyroid autoimmunity. Part of plasma scavenging system.
GPR74	G protein-coupled receptor 74	10886	G protein receptor signalling. Role in cell growth and/or maintenance. Perception of abiotic stimulus
ADAMTS3	A disintegrin-like and metalloproteinase (reprolysin type) with thrombospondin type 1 motif, 3	9508	Metalloproteinase activity. Role in collagen fibril organisation and collagen catabolism
GTAR ²	Gene trap ankyrin repeat	26057	Potential role in liver development
ALB	Albumin	213	Carrier protein in the serum. Role in stabilisation of extracellular fluid volume
AFP	Alpha-fetoprotein	174	Fetal counterpart of serum albumin
AFM	Afamin	173	Serum transport protein
RASSF6	Ras association (RA1GDS/AF-6) domain family 6	166824	Putative RAS binding protein, function unknown

IL8	Interleukin 8	3576	Major mediator of immune response. Acts as chemoattractant and potential role in angiogenesis. Proinflammatory action
CXCL6	Chemokine (C-X-C motif) ligand 6 (granulocyte chemotatic protein 2)	6372	Chemotactic factor for granulocytes. Leads to MMP9 release from granulocytes. Role in inflammatory response, signal transduction and cell-cell signalling
PF4V1	Platelet factor 4 variant 1	5197	Chemokine. Role in heparin binding and immune response
CXCL1	Chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)	2919	Multiple functions including regulation of cell proliferation, chemotaxis, inflammatory response and growth factor activity
PF4	Platelet factor 4 (chemokine (C-X-C motif) ligand 4)	5196	Roles in chemotaxis, immune response and cell proliferation. Induces IL-8 release from natural killer cells
PPBP	Pro-platelet basic protein (chemokine (C-X-C motif) ligand 7)	5473	Promotes neutrophil adhesion to vascular endothelium. Role in glucose transport, growth factor activity, immune response, and cell proliferation. Structural component of extracellular matrix
CXCL5	Chemokine (C-X-C motif) ligand 5	6374	Multiple roles including cell proliferation, chemotaxis, signal transduction, inflammatory response
CXCL3	Chemokine (C-X-C motif) ligand 3	2921	Multiple roles including chemotaxis, immune and inflammatory response and signalling
SPBPBP ²	DNA-binding protein amplifying expression of surfactant protein B	10895	Transcription factor activity
CXCL2	Chemokine (C-X-C motif) ligand 2	2920	Multiple roles including chemotaxis, immune and inflammatory response and signalling
EREG	Epiregulin	2069	Member of epidermal growth factor family. Multiple roles in cell signalling, proliferation, angiogenesis etc
AREG	Amphiregulin (schwannoma-derived growth factor)	374	Member of epidermal growth factor family. Growth factor for epithelial cells
BTC	Betacellulin	685	Member of epidermal growth factor family. Growth factor activity
ZNF363	Zinc finger protein 363	25898	Promotes p53 degradation. Interacts with androgen receptor

CDKL2	Cyclin-dependent kinase-like 2 (CDC2-related kinase)	8999	Multiple functions including signal transduction, role in sex differentiation and amino acid phosphorylation
G3BP2 ²	Ras-GTPase activating protein SH3 domain-binding protein 2	9908	Role in RAS protein signal transduction
VDP ²	Vesicle docking protein p115	8615	Role in intracellular protein transport
PPEF2	Protein phosphatase, EF hand calcium-binding domain 2	5470	Role in visual system.
CXCL9	Chemokine (C-X-C motif) ligand 9	4283	Specific function unknown. Role in signalling, inflammatory response, chemotaxis etc
CXCL10	Chemokine (C-X-C motif) ligand 10	3627	Multiple roles including stimulation of monocytes, natural killer and T cell migration. Role in signalling, chemotaxis, inflammatory response etc
CXCL11	Chemokine (C-X-C motif) ligand 11	6373	Chemotaxis in activated T cells. Multiple roles in signal transduction, immune response, chemotaxis etc
ART3	ADP-ribosyltransferase 3	419	ADP-ribosylation of proteins

¹ www.ncbi.nlm.nih.gov/LocusLink

² Denotes interim gene name and symbol. All other gene names and symbols as per HGNC official gene name and symbol.

Table 5.1: Integrated list of genes within the region identified by linkage analysis on chromosome 4. The double lines mark the approximate boundaries of the 1-LOD support interval.

knowledge of gene function is increasing at an exponential rate. In addition, the work of the Gene Ontology Consortium (www.geneontology.org) of classifying genes according to their molecular function, biological process and cellular component has made the identification of genes involved in specific functions much more efficient.

Selection of candidate genes concentrated on looking for those genes with a biologically plausible role in normal joint function or in the OA disease process. A strong candidate gene was identified. The ADAMTS3 (a disintegrin and metalloproteinase with thrombospondin motifs 3) gene which resides at 4q13.3. Its position relative to the linkage region identified by finer linkage mapping on chromosome 4 (chapter 3.4) is illustrated in figure 5.1. A number of studies have reported that estimates of the location of a disease locus from linkage analyses may vary moderately from the location of the true disease locus (Darvasi et al. 1993; Hovatta et al. 1998; Roberts et al. 1999). Also, in some cases linkage may be the result of variation within regulatory regions of the gene, sometimes many hundreds of kilobases distal to the gene itself. Therefore, the location of the ADAMTS3 gene slightly outside the centre of the region of linkage, was not felt to exclude it as a strong candidate for encoding susceptibility to OA.

5.2 Introduction: The ADAMTS3 (a disintegrin and metalloproteinase with thrombospondin motifs 3) gene

5.2.1 The potential role of ADAMTS3 in normal cartilage function and the OA disease process

The ADAMTS (a disintegrin and metalloproteinase with thrombospondin motifs) proteins are a relatively newly identified family of extracellular proteases. In similarity with the

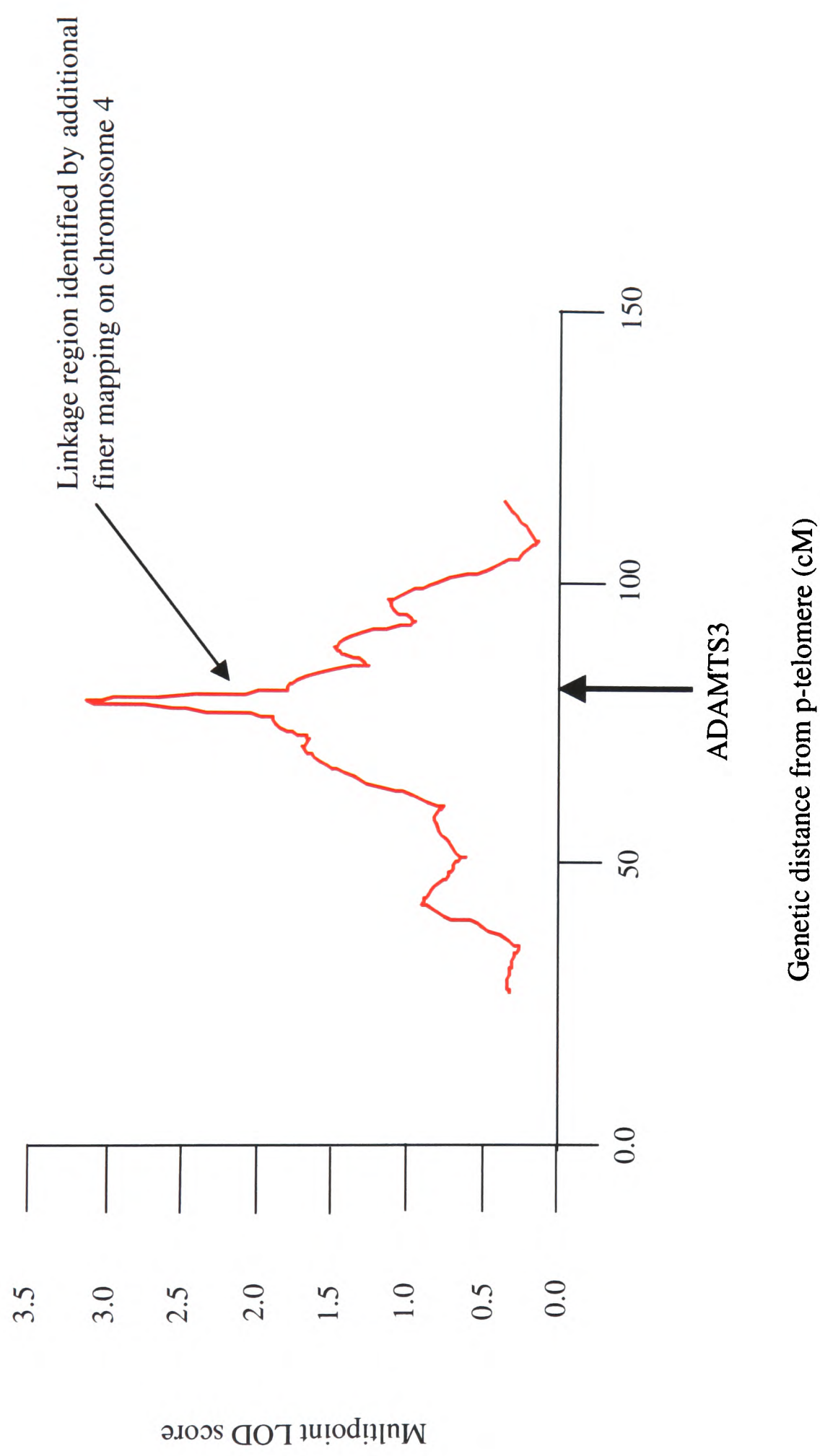


Figure 5.1: Approximate position of the ADAMTS3 gene relative to the chromosome 4 linkage region.

ADAM (a disintegrin and metalloproteinase) family of proteases they are characterised by multiple domains, including protease, disintegrin-like and cysteine-rich domains, but may be distinguished from the ADAM family by the presence of multiple copies of thrombospondin 1-like repeats (Tang 2001). They also have similarities with the matrix metalloproteinase (MMP) family of proteins. The ADAMTS proteins are predicted to be extracellular matrix proteins with diverse functions assigned to individual members including procollagen N-proteinase activity (ADAMTS2) and aggrecanase activity (ADAMTS4). They are predicted to be developmentally regulated, playing an important role in morphogenesis. This theory is supported by their low expression levels in adult tissues (Tang 2001). Members of the mammalian ADAMTS family display between 20 and 40% similarity to each other.

Type I and type II collagen are synthesised as soluble, procollagen molecules, consisting of a large collagenous region flanked at either end by propeptide molecules. Correct assembly of collagen fibrils requires the removal of these propeptides. The procollagen C-proteinase for both types of collagen molecule has been identified as the bone morphogenetic protein-1 (BMP-1) (Li et al. 1996). Recently, ADAMTS2 was identified by Colige et al (1999) as being a specific procollagen N-proteinase for type I and type II collagen.

Mutations in the ADAMTS2 gene are known to cause Ehlers-Danlos syndrome (EDS) type VIIC characterised by severe skin fragility as a result of defective N-terminal propeptide processing of type I procollagen (Colige et al. 1999). It has been observed however, that patients with this syndrome display normal processing of type II collagen in cartilage and show no evidence for chondrodysplasias or osteoarthritis, suggesting the presence of an additional procollagen N-proteinase in this tissue. In 1997, Nagase et al

identified a partial cDNA for ADAMTS3, lacking the 5' coding sequence, in human brain. Then, in 1999 Fernandes et al identified the ADAMTS3 gene as coding for a protein which belongs to the ADAMTS family, showing similarity in both amino acid sequence and domain structure to ADAMTS2. They investigated the role for both ADAMTS2 and ADAMTS3 in processing procollagen II in RCS-LS cells, a Swarm rat chondrosarcoma-derived cell line which normally fails to process the type II procollagen peptide. Similar effects were observed for both enzymes with addition of purified forms of the enzymes resulting in partial rescue of type II collagen processing. Investigations into expression levels of the enzymes revealed ADAMTS3 mRNA to be expressed at levels five-fold higher than ADAMTS2 mRNA in human cartilage. These observations led Fernandes et al (1999) to propose that ADAMTS3 and not ADAMTS2 is the physiologically relevant procollagen N-proteinase in cartilage tissue.

There are two types of procollagen II created by the alternative splicing of exon 2 of the type II collagen gene (COL2A1). This exon contains a cysteine rich repeat (CRR), which constitutes almost the entire globular domain of the N-propeptide of collagen II. Type IIA procollagen, which contains this CRR region, has been shown to be expressed in prechondrocytes, mesenchymal and epithelial cells and a number of other embryonic tissues (Bornstein 2002). Type IIB collagen lacks exon 2 and is expressed in differentiating and adult cartilage and other adult tissues (Bornstein 2002). This suggests a role for the CRR of the N-propeptide in the development of a number of tissues including cartilage in addition to its role in normal collagen fibril assembly (Bornstein 2002). During skeletal development and chondrogenesis type IIA procollagen retains the N-propeptide further supporting a role for the CRR in development (Fukui et al. 2002). The CRR of the N-propeptide of type IIA procollagen is able to bind to the morphogenetic factors TGFB1 and BMP2 (Zhu et al. 1999) and full-length procollagen

type IIA functions to inactivate BMP (Larrain et al. 2000). Cleavage of the N-propeptide of type IIA collagen in the extracellular matrix may therefore be envisaged to have a significant effect on the binding or presentation of the BMPs and in regulation of these growth factors in skeletal development. A role for the type IIA N-propeptide and its cleavage in growth plate remodelling has also been proposed (Nah and Upholt 1991). Studies have suggested that MMPs in addition to ADAMTS3 may be physiologically relevant in cleaving the N-propeptide at a number of different sites and that some level of proteinase redundancy may exist (Fukui et al. 2002).

Abnormalities in the ADAMTS3 gene that result in either total failure of, or reduction in, the ability of the protein to correctly process the type II collagen molecule could therefore affect both normal collagen fibril formation and skeletal development. Such abnormalities may then act as factors predisposing to the development of OA later in life as a result of an altered joint structure or functionally impaired cartilage.

In OA cartilage, but not normal aged cartilage, type IIA procollagen is synthesised probably reflecting reversion of the chondrocytes to a chondroprogenitor cell phenotype (Aigner et al. 1999). In unaffected adults chondrocytes express instead type IIB mRNA (Aigner et al. 1999).

In summary, the position of ADAMTS3 at 4q13, within the linkage region identified on chromosome 4 coupled with its potential role in type II collagen processing make it a strong candidate for encoding susceptibility to OA.

5.2.2 ADAMTS3 gene structure

The ADAMTS3 gene consists of twenty-two exons, five of which may be alternatively spliced (www.ncbi.nih.gov/IEB/Research/Acembly/). It is a very large gene spanning approximately 290 Kb from the start of exon 1 to the end of the 3'UTR. The mRNA molecule is predicted to be between 5822 and 5826 bp with a 5'UTR of approximately 40 bp and a long 3'UTR of approximately 2200 bp. It is proposed that alternative splicing may result in the generation of at least four transcripts (www.ncbi.nih.gov/IEB/Research/Acembly/). This is in keeping with other members of the ADAMTS family, where multiple transcripts of the genes are produced in a tissue-specific manner (Hurskainen et al. 1999; Fernandes et al. 2001). All three major genome databases (Ensembl, NCBI and UCSC) agree as to the size of the exons and the 3'UTR. There are however some slight discrepancies in intron sizes between the databases and between different builds of the same databases. All future references to intron sizes or nucleotide positions within the genomic DNA are as per the Ensembl database (April 2003 release). A summary of the size of the introns and the exons of the ADAMTS3 gene is found in table 5.2.

The gene encodes for a 135.6 kDa protein of 1205 amino acids of which the first twenty comprise a signalling peptide which is cleaved to generate the mature protein.

To study the potential role of the ADAMTS3 gene in encoding susceptibility to OA variants within the gene were identified by direct DNA sequencing. Common (minor allele frequency ≥ 0.05) variants were then genotyped in cohorts of probands, cases and controls and the pattern of linkage disequilibrium across the region established. Potential associations between alleles and/or haplotypes and OA were tested by χ^2 analysis using standard contingency tables. ODDs ratios were also calculated. The relationship between

Exon N°	Intron N°	Size (bp)
1		69
	1	1292
2		28
	2	18489
3		407
	3	133506
4		157
	4	75121
5		200
	5	16396
6		84
	6	2143
7		157
	7	750
8		106
	8	382
9		144
	9	627
10		133
	10	2600
11		114
	11	2035
12		146
	12	1210
13		176
	13	1109
14		134
	14	1527
15		124
	15	3329
16		81
	16	1906
17		164
	17	5474
18		166
	18	2490
19		133
	19	4591
20		208
	20	1986
21		118
	21	5046
22		568
	3'UTR	2170

Table 5.2: Exon and Intron sizes for the ADAMTS3 gene as per the Ensembl database, April 2003 release (www.ensembl.org).

genetic variation and the evidence for linkage was also examined. All investigations focussed on affected female patients who have undergone total hip replacement for OA (female-THR patients) as it was those families with female sibling pairs concordant for THR in which evidence for linkage to chromosome 4 was found (chapter 3.4).

For all PCR reactions, sequencing and restriction enzyme digests in this chapter, the reaction conditions, temperatures and full primer sequences are given in Appendices 4-6.

5.3 Sequencing of the ADAMTS3 gene.

All twenty-two exons, 5' and 3' flanking intronic sequences and the 3'UTR of the ADAMTS3 gene were sequenced in 48 female-THR OA cases in order to identify variants within the gene in our population. Details of primers and PCR conditions may be found in Appendix 4.

Results of the number of individuals successfully sequenced, exon size and the length of 5' and 3' intronic flanking sequence covered are summarised in table 5.3a. Sequences were aligned with each other and with a reference sequence taken from the Ensembl database and any sequence changes noted. Table 5.3b details the nature and position of polymorphisms discovered, minor allele frequencies and database SNP (dbSNP) reference (www.ncbi.nlm.nih.gov/SNP/) where available. Full details of sequence and position of SNPs may be found in Appendix 7. In summary, there are ten intronic SNPs, five of which have a minor allele frequency $\geq 5\%$ and may therefore be classed as common. The 3'UTR contains two SNPs, one of which has a minor allele frequency $\geq 5\%$. There are four exonic SNPs, all of which are non-synonymous with one having a minor allele frequency $\geq 5\%$. All four non-synonymous SNPs are conservative resulting

Exon	Exon size	N° of sequenced intronic bases 5' of exon	N° of sequenced intronic bases 3' of exon	Average N° of individuals sequenced
1	106	28	13	46
2	28	57	40	46
3	407	63	281	45
4	157	74	69	44
5	200	75	59	47
6	84	71	78	48
7	157	71	72	41
8	106	66	72	43
9	144	64	73	48
10	133	71	54	46
11	114	62	11	47
12	146	85	52	45
13	176	79	64	46
14	134	125	79	44
15	124	75	67	47
16	81	90	70	45
17	164	74	71	44
18	166	65	72	45
19	133	85	77	46
20	208	72	66	48
21	118	94	56	46
22 ¹	2736	95	59	45

¹ Sequence includes 569 bases exon 22 and the 3'UTR

Table 5.3a: Summary of sequencing for ADAMTS3 gene

Exon/Intron number	SNP	Nature	Distance from Exon	Amino acid substitution	Nucleotide number ⁵	Amino acid position	Minor allele frequency	rs reference ⁶
Intron 1	Indel ¹	Intronic	-7 ²		1391		0.47	
Intron 2	T/A	Intronic	-244 ²		19670		0.34	
Intron 2	A/G	Intronic	-128 ²		19786		0.33	
Exon 3	T/C	Non-synonymous		Arginine/Lysine	20230	138	0.34	788908
Intron 5	C/A	Intronic	-5 ²		245697		0.01	
Intron 5	G/A	Intronic	-4 ²		245698		0.01	
Intron 7	A/C	Intronic	+27 ³		248112		0.01	
Intron 10	G/A	Intronic	+34 ³		250261		0.01	
Intron 11	G/T	Intronic	+11 ³		252952		0.03	
Intron 12	G/T	Intronic	-19 ²		256314		0.05	
Exon 13	T/C	Non-synonymous		Asparagine/Serine	256341	585	0.01	Not in dbSNP
Intron 15	G/A	Intronic	+66 ³		259468		0.20	
Exon 20	T/C	Non-synonymous		Asparagine/Serine	277838	942	0.02	Not in dbSNP
Exon 22	C/T	Non-synonymous		Serine/Asparagine	285410	1122	0.01	Not in dbSNP
3'UTR	A/C	3'UTR	+922 ⁴		286585		0.04	
3'UTR	C/T	3'UTR	+937 ⁴		286600		0.05	

¹ Insertion/deletion of single A nucleotide

² Position relative to preceding exon

³ Position relative to preceding exon

⁴ Position relative to last nucleotide of termination codon

⁵ First nucleotide of exon 1 = nucleotide 0.

⁶ rs = reference SNP from www.ncbi.nlm.nih.gov/SNP/. Blank cell indicates not possible to determine if SNP is in dbSNP.

Table 5.3b: Summary of SNPs in ADAMTS3 detected by sequencing.

in the substitution of an amino acid with the same polarity and hydrophobicity. Only the most common of the exonic SNPs (in exon 3) has been previously reported and is in dbSNP. It is currently difficult to locate particular intronic SNPs in dbSNP and it was therefore not possible to determine how many of the identified intronic SNPs are novel.

The coding region of the ADAMTS3 gene is highly conserved with only one common (minor allele frequency $\geq 5\%$) polymorphism. Sequence diversity across the region sequenced, of one SNP approximately every 600 nucleotides, is comparable with current estimates of a genome-wide average density of one polymorphism approximately every three hundred to a thousand bases pairs (Kwok et al. 1996; Nickerson et al. 2000; Krugylak and Nickerson 2001).

5.4 Association studies - Stage 1

5.4.1 SNP selection

Linkage analyses have power only to detect common variants contributing to disease phenotype (Smith and Lusk 2002). Therefore, the evidence for linkage on chromosome 4 would appear to indicate the presence of common variant(s) in this region contributing to disease susceptibility. In addition, currently available evidence appears to indicate that many common diseases are due at least in part to genes with a small number of frequent, disease-associated alleles (Lohmueller et al. 2003). For these reasons those SNPs with a minor allele frequency greater than 5% (generally considered to be common) were selected for genotyping. The two SNPs in intron 2 were in complete LD and so only the first of the two, T/A at nucleotide 19670, was selected for genotyping.

Table 5.4a summarises the six SNPs selected for association studies. These represent four intronic SNPs, (three of which are single nucleotide changes and one of which represents an indel of a single A nucleotide) one exonic SNP (non-synonymous) and one SNP in the 3'UTR.

5.4.2. SNP genotyping

All six SNPs were genotyped using either forced or natural restriction enzyme digest analysis in the 146 female probands from the 146 female-THR families used in the linkage analysis studies (chapter 2.1) and in 399 age-matched female controls. Initial studies focused on the probands since these represent the 146 female-THR families showing evidence for linkage and therefore should provide maximum potential for detecting association. Genotyping accuracy was confirmed by sequencing in a subset of at least 42 individuals. Details of primers, enzymes and conditions may be found in Appendices 5 and 6.

5.4.3 Association analysis

Table 5.4b shows the numbers and frequencies of plus (+) and minus (-) alleles for all six SNPs in the probands and in the controls. Plus alleles correspond to the presence of a cut site for the specific restriction enzyme used whilst minus alleles correspond to no cut site. Also shown are the p-values derived from comparison of the (+) and (-) allele numbers in probands versus controls by χ^2 analysis using standard contingency tables. All SNPs conform to Hardy-Weinberg equilibrium in the controls. Two of the SNPs (Intron1indel and Intron12A/C) show a significant association ($P \leq 0.05$) with OA with p-values of 0.04 and 0.01 respectively. For SNP Intron1indel the minor allele is under-represented in the

SNP	Nucleotide position ¹	Distance from SNP above	Base change (non coding strand)	Base change (coding strand)	Amino Acid change	Nomenclature
1	1391		Indel	Indel		Intron1indel
2	19670	18279	T/A	A/T		Intron2A/T
3	20230	560	T/C	A/G	Arginine/Lysine	R138K
4	256314	236084	T/G	A/C		Intron12A/C
5	259468	3154	G/A	C/T		Intron15C/T
6	286600	27132	C/T	G/A		3'UTRG/A

¹ First nucleotide of exon 1 = nucleotide 0.

Table 5.4a: Summary of the ADAMTS3 SNPs selected for association analysis.

SNP	Controls		146 Female-THR Probands		
	(+)	(-)	(+)	(-)	P-value
	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	
Intron1indel	417(0.54)	353(0.46)	172(0.61)	108(0.39)	0.04
Intron2A/T	543(0.69)	241(0.31)	190(0.67)	94(0.33)	0.51
R138K	300(0.38)	484(0.62)	97(0.33)	195(0.67)	0.16
Intron12A/C	759(0.97)	23(0.03)	269(0.93)	19(0.07)	0.01
Intron15C/T	634(0.81)	148(0.19)	239(0.82)	51(0.18)	0.68
3'UTRG/A	724(0.91)	68(0.09)	259(0.90)	29(0.10)	0.53

Table 5.4b: Numbers and frequencies of (+) and (-) alleles for each of the ADAMTS3 SNPs in a cohort of female probands from the 146 female -THR families and a cohort of 399 age-matched female controls. The final column shows the p-value for the χ^2 test using standard contingency table analysis.

probands (0.39 versus 0.46) whilst for Intron12A/C the minor allele is over-represented (0.07 versus 0.03). No other SNPs show a significant difference ($P \leq 0.05$) in allele frequencies between probands and controls. SNP R138K shows a tendency toward significant association with a p-value of 0.16. For this SNP, the minor allele is under-represented in the probands compared with the controls (0.33 versus 0.38).

Genotype frequencies between the 146 female-THR probands and the 399 controls were also compared by χ^2 analysis using standard 2 x 3 contingency tables (table 5.4c). The only SNP showing a significant association is SNP Intron12A/C with a p-value of 0.01. This is the same as the p-value for comparison of allele frequencies (table 5.4b), indicating that there is no increased risk associated with a particular genotype.

ODDs ratios and 95% confidence intervals were calculated with respect to the minor allele for all six SNPs (table 5.4d). Only two SNPs have confidence intervals that do not encompass 1, the value reflecting 'no effect'. For SNP Intron1indel the ODDs ratio is 0.74 (95% confidence interval 0.86 to 0.98). This SNP also shows a significant association with χ^2 analysis ($P = 0.04$). The narrowness of the confidence interval increases confidence that the study has reasonable power to detect the true effect. However, the confidence interval is very close to 1 which indicates that any effect of this SNP is very small. SNP Intron12A/C has an ODDs ratio of 2.33 (95% confidence interval 1.25 to 4.35). This SNP also had a significant p-value ($P = 0.01$) with χ^2 analysis. The confidence interval for this SNP however, is wide indicating that the sample size may be too small to precisely estimate the effect of this SNP.

SNP	Controls			146 female-THR probands			P-value
	(++)	(+-)	(--)	(++)	(+-)	(--)	
	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	
Intron1indel	112(0.29)	193(0.50)	80(0.21)	54(0.39)	64(0.46)	22(0.16)	0.10
Intron2A/T	180(0.46)	183(0.47)	29(0.07)	65(0.46)	60(0.42)	17(0.12)	0.23
R138K	52(0.13)	196(0.50)	144(0.37)	18(0.12)	61(0.42)	67(0.46)	0.15
Intron12A/C	368(0.94)	22(0.06)	0(0)	126(0.88)	17(0.12)	1(<0.01)	0.01 ¹
Intron15C/T	255(0.65)	124(0.32)	12(0.03)	96(0.66)	47(0.32)	2(0.01)	0.55
3'UTRG/A	331(0.84)	62(0.16)	3(<0.01)	116(0.81)	27(0.19)	1(<0.01)	0.69

¹ Analysed with 2 x 2 contingency table, clumping (++) and (--) cells together

Table 5.4c: Genotype numbers and frequencies for the six ADAMTS3 gene SNPs for the 146 female-THR probands and the 399 age-matched female controls. The final column shows the p-value for comparison of probands and controls by χ^2 analysis using standard contingency table analysis.

SNP	ODDs ratio	95% confidence interval
Intron1indel	0.74	0.86 to 0.98
Intron2A/T	1.11	0.83 to 1.49
R138K	0.80	0.60 to 1.07
Intron12A/C	2.33	1.25 to 4.35
Intron15C/T	0.91	0.64 to 1.30
3'UTRG/A	1.19	0.75 to 1.88

Table 5.4d: ODDs ratios and 95% confidence intervals with respect to the minor allele of the ADAMTS3 SNPs following restriction digest in a cohort of female probands from the 146 female -THR families and a cohort of 399 age-matched female controls.

5.5 Haplotype analysis

The frequencies of haplotypes (complex alleles) composed of between two and six SNPs were estimated using the EH-PLUS program (Zhao et al. 2000) and compared between probands and controls. Haplotypes showing some evidence for association with a global $p\text{-value} \leq 0.1$ are shown in table 5.5. No haplotype has a global $p\text{-value}$ less than that for its constituent SNPs. This indicates that there is no increased risk associated with any particular haplotype. All of the haplotypes with a $p\text{-value} \leq 0.1$ contain the Intron12A/C SNP and it is the presence of this SNP which is generating the $p\text{-values} \leq 0.1$.

5.6 Linkage disequilibrium studies

Pairwise linkage disequilibrium (LD) patterns for the six SNPs were estimated from the dataset of all 545 individuals genotyped (146 female-THR probands and 399 controls) using the GOLD (graphical overview of linkage, Abecasis and Cookson 2000) program (table 5.6). Figures 5.6a and b display the graphical overview of the pair-wise linkage disequilibrium for D' and r^2 respectively on the physical map.

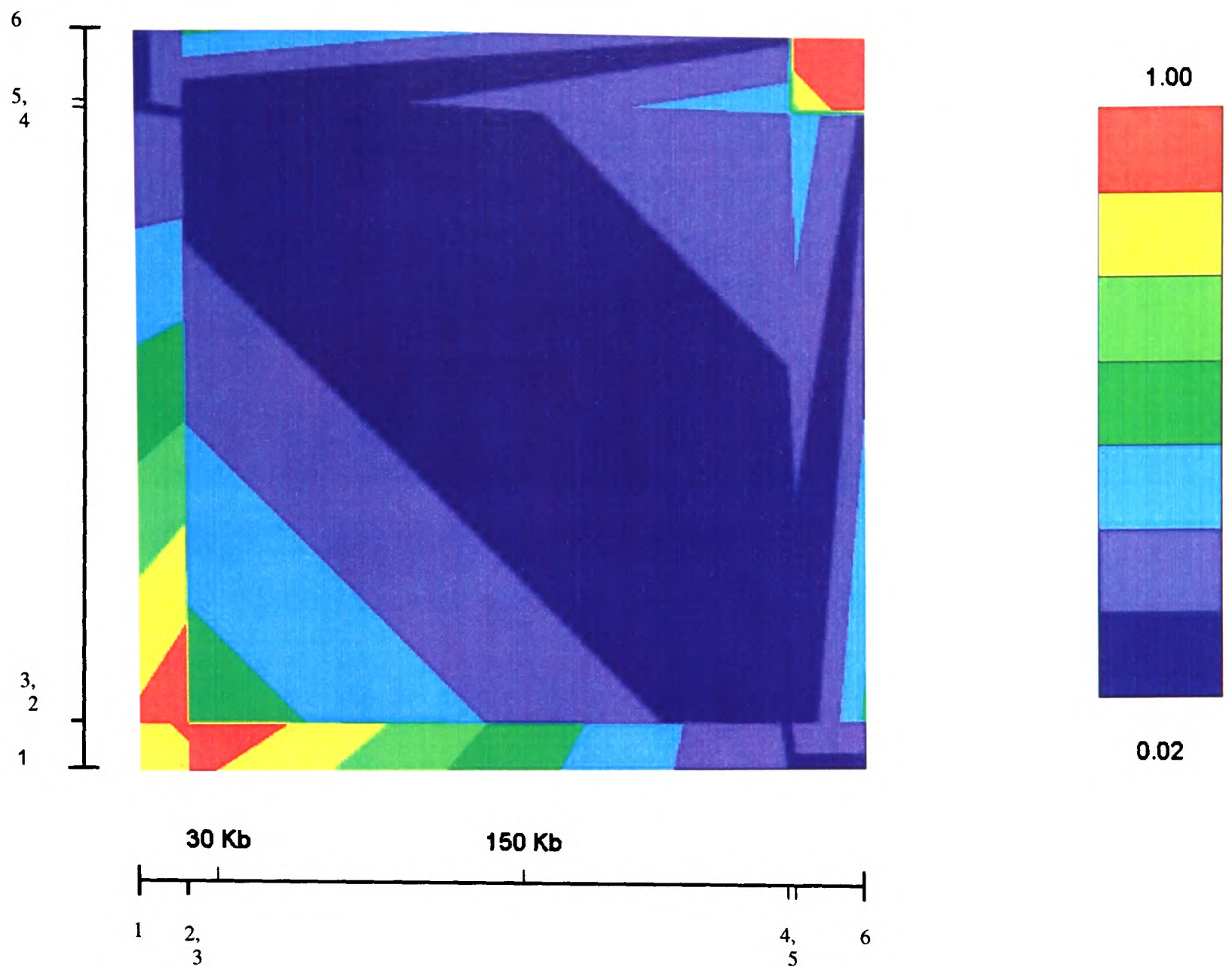
Consideration of D' indicates a high degree of LD between SNPs Intron1indel, Intron2A/T and R138K and between Intron15C/T and 3'UTRG/A with LD breaking down substantially between them. Pairwise comparisons with r^2 reveal a more modest pattern of LD with only SNP pairs Intron1indel & Intron2A/T, Intron1indel & R138K and Intron2A/T & R138K showing any evidence for disequilibrium (r^2 value greater than 0.1) with values of 0.233, 0.600 and 0.271 respectively. This indicates that for the majority of SNP pairs it is not possible to predict which allele is present at one SNP on the basis of the allele at the other SNP. Such a low level of LD will considerably compromise the power of EH-PLUS to reliably estimate haplotype frequencies and therefore the

SNPs defining haplotype	P-value	Lowest p-value of constituent SNPs
Intron1indel, Intron12A/C	0.01	0.01
Intron2A/T, Intron12A/C	0.084	0.01
R138K, Intron12A/C	0.036	0.01
Intron12A/C, Intron15C/T	0.087	0.01
Intron12A/C, 3'UTRG/A	0.065	0.01
Intron1indel, Intron2A/T, Intron12A/C	0.036	0.01
Intron1indel, R138K, Intron12A/C	0.064	0.01
Intron1indel, Intron12A/C, Intron15C/T	0.051	0.01
Intron1indel, Intron12A/C, 3'UTRG/A	0.046	0.01
R138K, Intron12A/C, Intron15C/T	0.076	0.01

Table 5.5: Results of haplotype analysis using EH-PLUS to estimate differences in haplotype frequencies between the female probands from the 146 female-THR families and the 399 age-matched female controls. Only haplotypes showing a p-value ≤ 0.1 are shown. No haplotype combination shows a p-value less than that for its constituent SNPs.

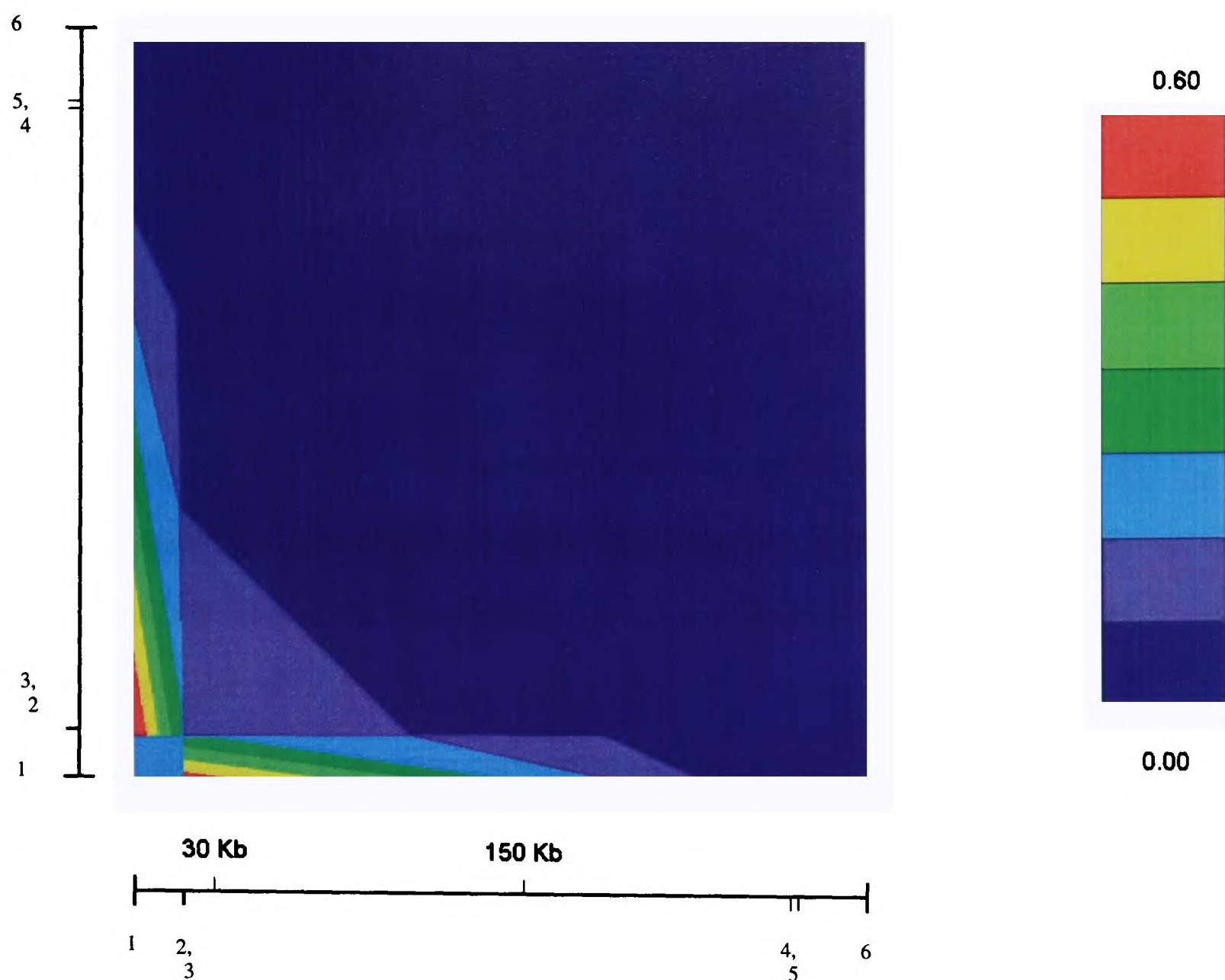
SNP 1	SNP 2	r ²	D'
Intron1indel	Intron2A/T	0.233	0.798
Intron1indel	R138K	0.600	0.894
Intron1indel	Intron12A/C	0.001	0.157
Intron1indel	Intron15C/T	0.002	0.113
Intron1indel	3'UTRG/A	0.001	0.126
Intron2A/T	R138K	0.271	1.000
Intron2A/T	Intron12A/C	0.002	0.149
Intron2A/T	Intron15C/T	0.007	0.258
Intron2A/T	3'UTRG/A	0.012	0.231
R138K	Intron12A/C	0.000	0.058
R138K	Intron15C/T	0.000	0.019
R138K	3'UTRG/A	0.013	0.472
Intron12A/C	Intron15C/T	0.001	0.400
Intron12A/C	3'UTRG/A	0.000	0.136
Intron15C/T	3'UTRG/A	0.023	1.000

Table 5.6: Pairwise LD values for the six ADAMTS3 gene SNPs.



SNP 1 = Intron1indel
 SNP 2 = Intron2A/T
 SNP 3 = R138K
 SNP 4 = Intron12A/C
 SNP 5 = Intron15C/T
 SNP 6 = 3'UTRG/A

Figure 5.6a: Graphical representation of pairwise LD, measured by D' , across the ADAMTS3 gene. Horizontal and vertical axes are scaled according to physical distance between SNP markers. The axis of the gene runs along the diagonal from the bottom left to the top right. The D' values are colour coded with red representing SNPs in linkage disequilibrium.



SNP 1 = Intron1indel
 SNP 2 = Intron2A/T
 SNP 3 = R138K
 SNP 4 = Intron12A/C
 SNP 5 = Intron15C/T
 SNP 6 = 3'UTRG/A

Figure 5.6b: Graphical representation of LD, measured by r^2 , across the ADAMTS3 gene. Horizontal and vertical axes are scaled according to physical distance between SNP markers. The axis of the gene runs along the diagonal from the bottom left to the top right. The r^2 values are colour coded with red representing SNPs tending towards linkage disequilibrium.

significance of results of the haplotype analysis must be treated with caution. Indeed, because of the low level of LD between most of the SNPs genotyped the use of EH-PLUS to predict haplotypes is likely to be of limited value for this gene.

5.7 Association studies - Stage 2

In order to try and determine if the associations observed in the female-THR probands represent genuine positive results, the two SNPs showing evidence for association (Intron1indel and Intron12A/C) were genotyped in a separate cohort of 244 unrelated, female-THR cases. SNP R138K was also genotyped in this additional cohort both because the p-value for association approached significance in the proband cohort and because it is the only non-synonymous SNP (and therefore a potentially functional SNP) with a minor allele frequency ≥ 0.05 .

The allele numbers for each of the three SNPs in the 244 female cases were compared to those in the 399 female controls by χ^2 analysis using standard contingency tables. Allele numbers, frequencies and p-values for the 244 female-THR cases, the 146 female-THR probands and for the 399 female controls are shown in table 5.7a. Also shown are allele numbers, frequencies and p-values for all probands and cases combined. None of the SNPs tested show significant evidence for association ($P \leq 0.05$) in the additional 244 females compared with the 399 controls. Observation of the allele frequencies for SNPs Intron1indel and R138K show that the allele that has a raised frequency in the 244 cases compared with the controls is the alternate allele to that that was raised for the female-THR probands i.e. for SNP Intron1indel the (-) allele is raised in the 244 female-THR case cohort when compared with the controls, whilst for the 146 female-THR probands cohort it is the (+) allele that is relatively raised. This is reflected by a p-value

approaching 1 for both SNPs when all cases are combined (final column of table 5.7a). This result would appear to show that the evidence for association observed in the female-THR proband cohort for these two SNPs is a false positive result.

For SNP Intron12A/C there is some evidence for association in the female-THR case cohort, with a p-value of 0.1, although this is not significant. This may be due to the association observed in the female-THR proband cohort being a false positive or may reflect the heterogeneous nature of the disease with a higher proportion of individuals in the female-THR proband cohort compared with the female-THR case cohort having a susceptibility allele at this locus.

Genotype frequencies in the 244 female cases and in the 146 female probands and 244 female cases combined ($n = 390$) were compared with the genotype frequencies in the 399 female controls by χ^2 analysis using standard contingency table analysis (table 5.7b). The only SNP showing a significant association in the combined cases and probands is Intron12A/C with a p-value of 0.02. This value is the same as the p-value for comparison of allele frequencies for this SNP (table 5.7a). This indicates that there is no increased risk associated with any particular Intron12A/C genotype.

ODDs ratios were calculated with regards to the minor allele for the three SNPs in the 244 additional female cases and for the 390 female cases and female probands combined (table 5.7c). The only SNP having a confidence interval that does not span 1, the value reflecting 'no effect', is SNP Intron12A/C in the probands and cases combined. The ODDs ratio for this SNP is 1.95 (95% confidence interval 1.16 to 3.27). This SNP also shows a significant association with χ^2 analysis ($P = 0.02$). The confidence interval for this SNP is

SNP	Controls				244 Female-THR cases				P-value	Probands and cases combined (n=390)				P-value
	(++)	(+-)	(--)		(++)	(+-)	(--)			(++)	(+-)	(--)		
	N° (Freq)	N° (Freq)	N° (Freq)		N° (Freq)	N° (Freq)	N° (Freq)			N° (Freq)	N° (Freq)	N° (Freq)		
Intron1indel	112(0.29)	193(0.50)	80(0.21)		63(0.27)	106(0.46)	61(0.27)		0.26	117(0.32)	170(0.46)	83(0.22)		0.52
R138K	52(0.13)	196(0.50)	144(0.37)		45(0.19)	111(0.47)	79(0.34)		0.14	63(0.17)	172(0.45)	146(0.38)		0.29
Intron12A/C	368(0.94)	22(0.06)	0(0)		212(0.91)	19(0.08)	2(0.01)		0.16 ¹	338(0.90)	36(0.10)	3(<0.01)		0.02 ¹

¹ Analysed with 2 x 2 contingency table, clumping (++) and (--) cells together

Table 5.7b: Genotype numbers and frequencies for three ADAMTS3 gene SNPs for the 390 female-THR probands and cases combined and the 399 age-matched female controls. The final column shows the p-value for comparison of probands and cases versus controls by χ^2 analysis using standard contingency table analysis.

SNP	244 Female-THR cases		Probands and cases combined (n = 390)	
	ODDs ratio	95% confidence interval	ODDs ratio	95% confidence interval
Intron1indel	1.1609	0.9212 to 1.4631	0.9825	0.8023 to 1.2031
R138K	1.2055	0.9553 to 1.5213	1.0361	0.8443 to 1.2716
Intron12A/C	1.7133	0.9500 to 3.0900	1.9466	1.1589 to 3.2697

Table 5.7c: ODDs ratios and 95% confidence intervals with respect to the minor allele (defined as per the 146 female-THR probands) of three ADAMTS3 SNPs following restriction digest in the 244 female-THR cases and the 390 female-THR probands and cases combined in comparison with the 399 age-matched female controls.

wide indicating that the sample size is too small to allow a precise estimation of the effect of this SNP.

5.7.1 Haplotype analysis

The frequencies of two and three SNP haplotypes comprised of SNPs Intron1indel, R138K and Intron12A/C were estimated using the EH-PLUS program (Zhao et al. 2000) and compared between all 390 probands and cases combined and the 399 controls (table 5.7d). Haplotypes where the global p-values are less than those for any of the constituent SNPs are highlighted in red.

The only haplotype with a global p-value less than that for the constituent SNPs is that comprised of Intron1indel and Intron12A/C with a p-value of 0.01 compared with 0.91 and 0.02 for SNPs Intron1indel and Intron12A/C respectively.

In order to investigate this further the four possible haplotypes for the Intron1indel and Intron12A/C SNPs were estimated in the combined probands and cases and then individually compared to the estimated frequencies in the 399 controls (table 5.7e). One haplotype - that containing the common allele at SNP Intron1indel and the variant allele at SNP Intron12A/C is significantly over represented in the combined probands and cases compared with controls (OR 3.26; 95% CI 1.65 to 6.44; $p < 0.0005$). The wide confidence interval however, indicates that this result should be treated with caution as the sample size may be too small to accurately estimate the effect of this SNP. Also, and as was noted previously (table 5.6), the pairwise r^2 and D' values for SNPs Intron1indel and Intron12A/C are very low: 0.01 and 0.157 respectively. Such low values impair the ability of the EH-PLUS program to accurately estimate haplotype frequencies.

SNPs defining haplotype	Probands and cases combined (n=390)	Lowest p-value for constituent SNPs
Intron1indel, Intron12A/C	0.010	0.02
R138K, Intron12A/C	0.020	0.02
Intron1indel, R138K, Intron12A/C	0.069	0.02

Table 5.7d: Results of haplotype analysis using EH Plus to estimate differences in haplotype frequencies between all 390 female-THR probands and cases combined; and the 399 age-matched female controls. Only haplotypes showing a p-value ≤ 0.1 are shown. P-values less than those for any of the constituent SNPs are highlighted in red.

Haplotype combination		Frequency		ODDs ratio	P-value
Intron1indel	Intron12A/C	Probands and cases combined (n = 390)	Controls (n = 399)		
1	1	0.50	0.53	0.90(0.74 - 1.11)	0.34
1	2	0.05	0.02	3.26 (1.65 - 6.44)	<0.0005
2	1	0.44	0.44	0.99 (0.81 - 1.22)	1.00
2	2	0.01	0.01	0.69 (0.27 - 1.79)	0.55

Table 5.7e : Haplotype frequencies estimated from the genotypes of the 390 probands and cases combined and from the 399 controls. Intron1indel, allele 1, insertion, allele 2, deletion; Intron12A/C, allele 1, A, allele 2, C. The ODDs ratio and 95% confidence interval of each haplotype relative to the other haplotypes as a group are shown. The final column shows the p-value for comparison of cases and controls by χ^2 analysis using standard contingency table analysis. P-values ≤ 0.05 are highlighted in red.

5.8 Relationship between genetic variation and the evidence for linkage

In order to investigate the association of genetic variation at SNP Intron12A/C, the only SNP showing evidence for association in the probands and cases combined, with the evidence for linkage to chromosome 4, each of the 146 female-THR families used in the linkage analysis were classified according to the genotype of the proband at SNP Intron12A/C. Two groups were defined; those eighteen families where the proband has at least one copy of the associated (C) allele and those 126 families where the proband was homozygous for the non-associated (A) allele (two probands failed in the digestion analysis). Linkage analysis was then undertaken for the two groups separately. Under the null hypothesis of no association between SNP Intron12A/C and the disease, it would be expected that evidence for linkage would be randomly distributed amongst the two groups and would be proportional to the size of the group. If, however, the SNP is associated with the disease then it would be expected that the group with at least one copy of the associated allele would show a disproportionately high level of evidence for linkage (Cox 2002).

Figure 5.8 shows the results of linkage analysis for the two groups overlaid on the results for chromosome 4 for all 146 families. Removal of the eighteen families in which the proband has at least one copy of the associated allele results in a slight decrease in LOD score from 3.1 to 3.0. Linkage analysis on these eighteen families results in a LOD score of just over 0.1. Although the evidence for linkage is not strictly proportional to the size of the two groups (which would result in LOD scores of 0.4 and 2.7) it is clear that the eighteen families are not disproportionately contributing to linkage. This implies that the (C) allele of SNP Intron12A/C is not associated with the evidence for linkage observed at this locus.

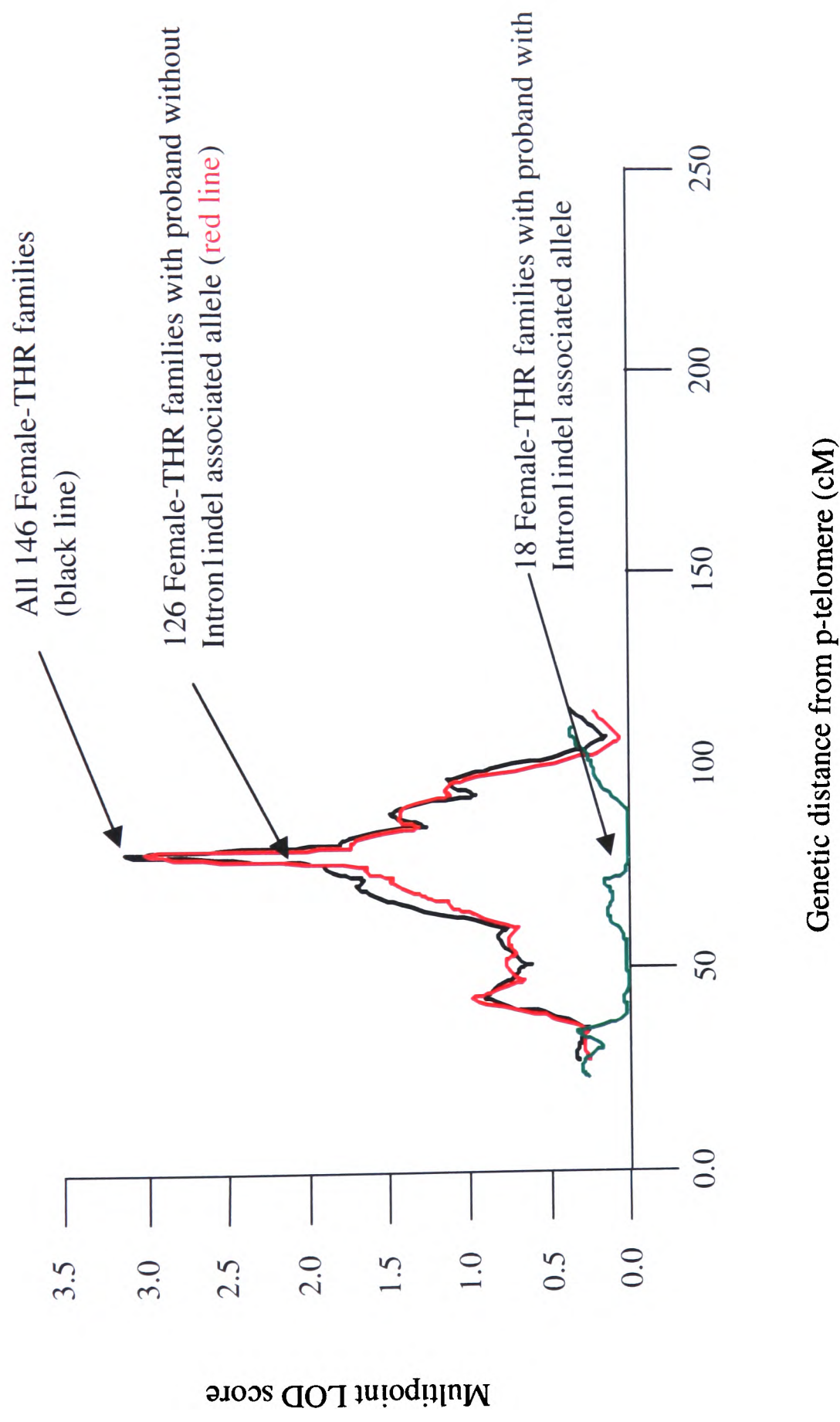


Figure 5.8: Multipoint linkage analysis of chromosome 4 for the 146 female-THR families, overlaid with results of linkage analysis for both the eighteen families whose proband has a copy of the associated (C) allele at SNP Intron12A/C and for the 126 families where the proband is homozygous for the non-associated (A) allele of this SNP.

This result, in addition to the lack of replication in the 244 female-THR cases of the initial positive result that had been detected in the 146 female-THR probands, would appear to suggest that this SNP does not encode for susceptibility to OA.

5.9 Discussion

The identification of genes encoding susceptibility for complex traits is often a long and arduous task complicated by a number of factors including the presence of locus and allelic heterogeneity, epistasis, low penetrance and variable expressivity. In addition, a number of susceptibility alleles will be present in any population with each allele only contributing to a subset of the total number of cases. Despite these potential obstacles association analysis has been used successfully in the identification of variants encoding for susceptibility to complex disease. For example, the identification of polymorphisms within the CARD15 gene encoding susceptibility to Crohns disease (Hugot et al. 2001; Ogura et al. 2001).

The ADAMTS3 gene was identified as a strong candidate for OA susceptibility based both on its position within the region showing evidence for linkage to OA on chromosome 4 and the biological plausity of a role for the protein product in normal and diseased cartilage homeostasis.

All genetic variation in and immediately around the exonic sequence for ADAMTS3 has been determined by the sequencing of the gene in at least 41 females who have undergone total hip joint replacement for OA (female-THR cases). The gene shows a moderate level of nucleotide diversity with sixteen SNPs in the total of 9018 nucleotides sequenced. This represents an average frequency of one SNP every 600 nucleotides which is comparable

to the current estimates of a genome-wide average density of one polymorphism every three hundred to a thousand base pairs (Kwok et al. 1996; Nickerson et al. 2000; Krugylak and Nickerson 2001). Of these sixteen SNPs, seven may be classified as common (minor allele frequency ≥ 0.05) consisting of one exonic (non-synonymous, conservative) and six non-coding (five intronic, one 3'UTR) SNPs.

Comparison of allele frequencies of six of these common SNPs (two of the SNPs are in complete LD) in a cohort of female probands from the 146 female-THR families used in the additional finer linkage mapping analysis for chromosome 4 (chapter 3.4) and a cohort of 399 age-matched female controls provided significant evidence for an association between two of the SNPs (Intron1indel and Intron12A/C) and OA ($P = 0.04$ and 0.01 respectively). A third SNP (R138K), the only common, exonic SNP within the gene, approached significance with a p-value of 0.16 .

In order to try and determine if any of these results represent genuine positives, these three SNPs were genotyped in a separate cohort of 244 female-THR cases. None of the SNPs showed significant ($P \leq 0.05$) evidence of association with disease in this second cohort, although SNP Intron12A/C approached significance ($P = 0.1$). When the probands and additional cases were combined only SNP Intron12A/C showed any evidence for association ($P = 0.02$). This is slightly less significant than the p-value of 0.01 observed for the probands alone. These results appear to suggest that the moderate evidence for association for SNPs Intron1indel, Intron12A/C and R138K observed in the female-THR probands represent false positive results.

There is a very low level of LD between the SNPs genotyped, with only three out of the fifteen pairwise combinations having r^2 values greater than 0.1 . Such low levels of LD are

not entirely unexpected as a result of the large distance separating many of the SNPs. Indeed the SNPs showing highest levels of LD are some of those in closest proximity. The examination of haplotypes (complex alleles), using a predictive method, provided little additional information at least partly as a result of this low level of linkage disequilibrium.

Investigation of the association of genetic variation with the evidence for linkage for SNP Intron12A/C reveals no evidence for partitioning of the evidence of linkage in the way that would be expected if the minor allele of this SNP were associated with OA disease. This suggests that neither this SNP, nor one in strong LD with it, is responsible for the evidence for linkage seen on chromosome 4. This observation, in conjunction with the failure to replicate the significant female proband result with the 244 female-THR cases, appears to exclude a role for this SNP in encoding for susceptibility to OA.

In an association study of a complex disease it is difficult to conclusively exclude a gene as having a role in encoding for disease susceptibility. Although it appears that the polymorphisms identified within and close to the exons in the ADAMTS3 gene are not encoding for susceptibility to OA, there remains the possibility of the involvement of this gene in the OA disease process.

It is possible that variation(s) within the introns or in distal regulatory elements of the ADAMTS3 gene may be involved in determining OA susceptibility. Such variations could affect the splicing and/or expression of the ADAMTS3 gene.

Novel splicing enhancers and silencers are constantly being discovered in both coding and non-coding regions of DNA and it appears that splicing is regulated in a complex and

tightly co-ordinated way (Faustino and Cooper 2003). It is possible therefore, that variation within non coding regions of the ADAMTS3 gene may act to subtly alter the splicing of the ADAMTS3 transcript. In a study of autoimmune disease patients Ueda et al (2003) found evidence for a non-coding 6.1 kb region 3' of the cytotoxic T lymphocyte antigen 4 gene (CTLA4) being associated with the efficiency of splicing of the mRNA molecule and the relative levels of the soluble form of the protein.

Several studies have suggested that variation in expression levels may also be important in a number of genes that encode susceptibility to complex traits (Drysdales et al. 2000; Cowles et al. 2002). In some skeletal dysplasias, mutations in the regulatory regions of genes have been associated with milder phenotypes than those severe phenotypes often associated with either loss of function or dominant negative mutations (Dreyer et al. 1998).

Since the ADAMTS3 gene is large, spanning approximately 290 Kb, and as regulatory regions for both splicing and expression may be found throughout the gene or distal to it, it is likely that at least some of the variation that affects splicing or expression will be in almost total linkage equilibrium with the SNPs genotyped in this study. This is supported by the low level of linkage disequilibrium between the SNPs investigated. If SNPs influencing expression or splicing of the ADAMTS3 gene are responsible for the OA phenotype therefore it may be unlikely that any association would be detected with the SNPs used in this study.

For these reasons it is not possible, at this stage, to totally exclude ADAMTS3 as the chromosome 4 OA susceptibility gene. Further studies investigating the expression or

splicing patterns of this gene in OA patients versus normal controls may yield further insights into the potential role of this gene in OA.

If this gene is not responsible for the linkage to OA detected in this region then the evidence for linkage must be due to variation within another gene or genes. The genome databases are being constantly updated and the functions of genes which were previously unknown are being investigated. In addition novel functions for already characterised genes are frequently discovered. Although ADAMTS3 was, at the time of candidate gene selection, and still is, the most obvious candidate within the linkage region on chromosome 4, a number of other genes could potentially be envisaged as having a role in the OA disease process.

There is a cluster of genes encoding CXC chemokines (a small family of pro-inflammatory cytokines characterised by the presence of one amino acid residue separating the first two conserved cysteine residues), including IL-8, within the linkage region on chromosome 4. The chemokines are pro-inflammatory and have functionally diverse roles within the immune and nervous systems in addition to a role in angiogenesis (Frederick and Chapman 2001). Although OA is not defined as an inflammatory disease, a variable degree of inflammation is observed and recent evidence suggests that cartilage integrity is maintained by the balance between pro- and anti- inflammatory cytokines (Fernandes et al. 2002). Any one or a number of these chemokines could therefore be envisaged as having a role in the cartilage degradation seen in OA. IL-8 in particular has been demonstrated to be expressed by articular chondrocytes (Lotz et al. 1992).

A number of studies have implicated a potential role for vitamin D in OA either directly or as a consequence of its role in normal bone metabolism (Dieppe et al. 1993;

McAlindon et al. 1996). The vitamin D binding protein gene, GC, could therefore potentially have a role in the OA disease process by modifying vitamin D function.

There are also a number of genes within the 4q linkage interval that encode for growth factors and proteins involved in cell and nuclear signalling. If any of these genes are found to be expressed in joint tissue then they become reasonable candidates for detailed investigation.

In conclusion, no substantial evidence has been found for an association between OA and the common ADAMTS3 DNA variants genotyped. As a consequence of the large size of this gene, and the low level of linkage disequilibrium across it however, further work, for example investigating expression or splicing patterns, would be needed before it can be confidently excluded as a susceptibility gene for OA. There are a number of additional genes within the region which may be envisaged as having a role in the OA disease process and which may also be investigated in future studies.

Chapter 6 -

IL4R

**candidate gene
identification**

6.1 Introduction: Candidate gene identification on chromosome 16

As for chromosome 4 (chapter 5), candidate gene investigation was undertaken for the two regions identified on chromosome 16. The regions on chromosome 16p and 16q identified by linkage mapping (chapter 3.7) are still too large to allow for economically viable systematic association analysis, with the 1-LOD support intervals spanning approximately 8 Mb of DNA on chromosome 16p and 9 Mb of DNA on chromosome 16q. A search for candidate genes within these regions was therefore undertaken.

6.1.1 Cataloguing of genes within the linkage region

Genes within the 1-LOD support intervals of the two regions identified on chromosome 16, defined as 16p12.3-16p12.1 (19 Mb - 27 Mb) and 16q22.1-16q23.1 (69 Mb - 78 Mb) respectively, were identified using the three major databases of Ensembl, NCBI and UCSC as for Chromosome 4 (section 5.1.1). Tables 6.1a and b detail integrated lists from all three databases of all confirmed genes within the regions (correct as of July 2003). A number of the genes in these regions are hypothetical and are not included in the tables.

6.1.2. Identification of candidate genes

Following identification and classification of the genes within the regions, one gene was selected as being a notable candidate. The interleukin 4 receptor α chain (IL4R) gene resides to the edge of the 1-LOD support region on 16p at 16p12.1-p11.2. Its position relative to the linkage region is illustrated in figure 6.1.

Gene symbol	Name	LocusID ¹	Function
RPS15A	Ribosomal protein S15a	6210	Cytosolic small structural protein of the ribosome. Involved in protein biosynthesis
ARL6IP	ADP-ribosylation factor-like 6 interacting protein	23204	Integral membrane protein. Function unknown
SMG1 ²	PI-3-kinase-related kinase SMG-1	23049	Role in nonsense-mediated mRNA decay (NMD)
TMC7 ²	TMC7 protein	79905	Belongs to a novel family of transmembrane channel-like proteins. Function unknown
COQ7	Coenzyme Q7 Homolog, ubiquinone (yeast)	10229	Role in ubiquinone biosynthesis and gluconeogenesis
LOC51760 ²	B/K protein	51760	Transporter activity
TMC5 ²	TMC5 protein	79838	Belongs to a novel family of transmembrane channel-like proteins. Function unknown
MIR16 ²	Membrane interacting protein of RGS16	51573	Integral membrane glycoprotein. Function unknown
CP110 ²	CP110 protein	9738	Function unknown
MGC16824 ²	Esophageal cancer associated protein	57020	Function unknown
GPRC5B	G protein-coupled receptor, family C, group 5, member 5	51704	Receptor activity. Specific function unknown
GP2	Glycoprotein 2 (zymogen granule membrane)	2813	Associated with zymogen granule membrane. Function unknown
UMOD	Uromodulin (uromucoid, Tamm-Horsfall glycoprotein)	7369	Role in normal renal function

BUCS1	Butyryl Coenzyme A synthetase 1	1166285	Role in lipid metabolism	
SAH	SA hypertension-associated homolog (rat)	6296	Function unknown. RFLP associated with hypertension	
LOC81691 ²	Exonuclease NEF-sp	81691	Function unknown	
LOC57156 ²	Promethin	57146	Function unknown	
ZP2	Zona pellucida glycoprotein 2 (sperm receptor)	7783	Structural protein of zona pellucida. Role in sperm penetration	
CRYM	Crystallin, mu	1428	Binds thyroid hormone and may have a role in regulation or development	
LAT1-3TM ²	LAT1-3TM protein	81893	Function unknown	
DREV1 ²	DORA reverse strand protein	51108	Function unknown	
IGSF6	Immunoglobulin superfamily, member 6	10261	Transmembrane receptor with a role in immune response	
OTOA	Ontancorin	146183	Expressed in inner ear only. Candidate gene for isolated deafness	
UQCRC2	Ubiquinol-cytochrome c reductase core protein II	7385	Role in electron transport chain in mitochondria	
EEF2K ²	Elongation factor-2 kinase	29904	Translation factor activity. Role in cell division	
RPC5 ²	RNA polymerase III 80 kDa subunit RPC5	55718	Subunit of RNA polymerase III	
CDR2	Cerebellar degeneration-related protein 2, 62kDa	1039	Function unknown	

HS3ST2	Heparan sulfate (glucosamine) 3-O-sulfotransferase 2	9956	Role in generating heparan sulfate. Expressed mainly in brain
SCNN1G	Sodium channel, nonvoltage-gated 1, gamma	6340	Controls reabsorption of sodium in the kidney, colon, lung and sweat glands
COG7	Component of oligomeric golgi complex 7	91949	Role in protein transport
GGA2	Golgi associated, gamma adaptin ear containing, ARF binding protein 2	23062	Role in protein trafficking between trans-golgi network and the lysosome
UBPH ²	Similar to ubiquitin binding protein	56061	Function unknown
NDUFAB1	NADH dehydrogenase (ubiquinone) 1, alpha/beta subcomplex, 1, 8 kDa	4706	Role in lipid metabolism
MGC3248 ²	Dynactin 4	84516	Function unknown
PLK	Polo-like kinase (Drosophila)	5347	Role in regulation of cell cycle via interaction with cyclin B1
ERN2	ER to nucleus signalling 2	10595	Role in ER stress response
LOC63928 ²	Hepatocellular carcinoma antigen gene 520	63928	Potential tumour antigen
PRKCB1	Protein kinase C, beta 1	5579	Calcium activated kinase. Phosphorylates a range of cellular proteins. Role in B cell activation
CACNG3	Calcium channel, voltage-dependent, gamma subunit 3	10368	Component of L-type calcium channel. Role in neurones
TNRC6	Trinucleotide repeat containing 6	27327	Putative transcription factor in brain. May have a role in regulating neurodevelopment or neuroplasticity
RBBP6	Retinoblastoma binding protein 6	5930	Role in regulation of cell cycle and proliferation

TNCR6	Trinucleotide repeat containing 6	27327	Putative transcription factor expressed in brain. ? Role in regulating neuro development and neuro plasticity
KST1 ²	Sodium/myo-inositol cotransporter 2	115584	Putative sodium-coupled transporter
RICH1 ²	Homolog of rat nadrin	55114	GTPase activating protein for Cdc42 and Rac1
LCMT1	Leucine carboxyl methyltransferase 1	51451	Methylation of protein phosphatase 2A
AQP8	Aquaporin 8	343	Water channel protein in pancreas and colon
NSE1 ²	Non-SMC (structural maintenance of chromosomes) element 1 protein	197370	Role in cell proliferation and DNA repair
IL4R	Interleukin 4 receptor	3566	Role in cell signalling. Involved in immune response
IL21R	Interleukin 21 receptor	50615	Role in proliferation of T, B and NK cells
GTF3C1p	General transcription factor IIIC, polypeptide 1, alpha 220 kDa	2975	Component of RNA polymerase III. Role in 5sRNA, tRNA and VAI RNA transcription

¹ www.ncbi.nlm.nih.gov/LocusLink

² Denotes interim gene name and symbol. All other gene names and symbols as per HGNC official gene name and symbol.

Table 6.1a: Integrated list of genes within the 1-LOD support region identified by linkage analysis on chromosome 16p.

Gene symbol	Name	LocusID ¹	Function
PDPR ²	Pyruvate dehydrogenase phosphatase regulatory subunit	55066	Potential role in glycine catabolism
MTR3 ²	Homolog of yeast mRNA transport regulator 3	118460	Subunit of the exosome involved in mRNA degradation
WWP2 ²	Nedd-4-like ubiquitin - protein ligase	11060	Role in ubiquitin-protein ligase activity. Specific function unknown
AARS	Alanyl-tRNA synthetase	16	Role in tRNA synthesis
DDX19	DEAD (Asp-Glu-Ala-As) box polypeptide 19	11269	Role in export of mRNA from the nucleus
SIAT4B	Sialyltransferase 4B (beta-galactosidase alpha-2,3-sialyltransferase	6483	Transfer of sialic acid in the Golgi apparatus
FUK ²	L-fucose kinase	197258	Involved in salvaging of fructose from degradation of oligosaccharides
COG4	Component of oligomeric Golgi complex 4	25839	Role in Golgi apparatus structure and capacity for both intracellular transport and glycoprotein modification
SF3B3	Splicing factor 3b, subunit 3, 130kDa	23450	Involved in assembly of the spliceosome
ZNF23	Zinc finger 23 (KOX 16)	7571	Transcription factor. Potential role in embryonic development
ZNF19	Zinc finger protein 19 (KOX 12)	7567	Putative transcription factor. Function unknown
CHST4	Carbohydrate (N-acetylglucosamine 6-O) sulphotransferase 4	10164	Potential involvement in synthesis of corneal keratan sulphate

AP1G1	Adaptor-related protein complex 1, gamma 1 subunit	164	Role in membrane trafficking between the trans-Golgi network and endosomes
DHODH	Dihydroorotate dehydrogenase	1723	Involved in de novo pyrimidine biosynthesis. Located on outer surface of inner mitochondrial membrane
CALB2	Calbindin 2, 29 kDa (calretin)	794	Enhancer of gene expression in neuronal cells. Role at network level in cerebellar physiology
TAT	Tyrosine aminotransferase	6898	Catalyses conversion of L-tyrosine to p-hydroxyphenylpyruvate in liver
HP	Haptoglobin	3240	Role in haemoglobin metabolism and iron elimination. Also immunomodulatory effects on Th1/Th2 balance
HPR	Haptoglobin-related protein	3250	Function unclear. May have arisen by gene conversion from haptoglobin
DHX38	DEAD (Asp-Glu-Ala-As) box polypeptide 38	9785	ATPase with a role in pre-mRNA splicing
PMFBP1	Polyamine modulated factor 1 binding protein 1	83449	Function unknown
ATBF1	AT-binding transcription factor 1	463	Transcription factor regulating alpha-fetoprotein expression (found in hepatoma cell lines)
PSMD7	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 7 (Mov34 homolog)	5713	Subunit of the 19S regulator of the 26S proteasome
FAXDC1	Fatty acid hydroxylase domain containing 1	79152	Function unknown
NIN283 ²	Nerve injury gene 283	84937	Potential role in ubiquitin-mediated protein modification in the central nervous system
GLG1	Golgi apparatus protein 1	2734	Biogenesis and function of Golgi apparatus

LDHD	Lactate dehydrogenase D	197257	Role in metabolism.
ZFP1 ²	Zinc finger protein 1	162239	Potential transcription factor. Function unknown
CTRB1	Chymotrypsinogen B1	1504	Serine protease in the gastrointestinal tract
BCAR1	Breast cancer anti-oestrogen resistance 1	9564	Adaptor molecule involved in cytoskeletal organisation and embryonic development. Also involved in oncogenic transformation
CFDP1	Craniofacial development protein 1	10428	Putative role in craniofacial development
CHST6	Carbohydrate (N-acetylglucosamine 6-O) sulphotransferase 6	4166	Sulphate modification of glyco conjugates. Potential involvement in synthesis of corneal keratan sulphate
CHST5	Carbohydrate (N-acetylglucosamine 6-O) sulphotransferase 5	23563	Sulphate modification of glyco conjugates. Expressed in small intestine and colon
ADAT1	Adenosine deaminase, tRNA-specific 1	23536	Responsible for deamination of adenosine 37 to inosine in tRNA
GABARAPL2	GABA(A) receptor-associated protein-like 2	11345	Soluble transport factor. May modulate intra Golgi transport
KARS	Lysyl-tRNA synthetase	3735	Involved in charging tRNAs with their cognate amino acids
TERF2IP	Telomeric repeat binding factor 2, interacting protein	54386	Located at telomeres and affects telomere length
CASPR4 ²	Cell recognition protein CASPR4	85445	Function in vertebrate nervous system as cell adhesion molecule/receptor

ADAMTS18	A disintegrin-like and metalloproteinase (reprolysin type) with a thrombospondin type I motif 18	170692	Mainly expressed in lung, liver and kidney foetal tissues. Also expressed in some adult tissues. Specific function unknown
CLECSF1	C-type (calcium dependent, carbohydrate-recognition domain) lectin, superfamily member 1 (cartilage-derived)	10143	Potential role in cartilage matrix organisation or remodelling
WWOX	WW domain containing oxidoreductase	51741	Potential roles in steroid metabolism and apoptosis

¹ www.ncbi.nlm.nih.gov/LocusLink

² Denotes interim gene name and symbol. All other gene names and symbols as per HGNC official gene name and symbol.

Table 6.1b: Integrated list of genes within the 1-LOD support region identified by linkage analysis on chromosome 16q.

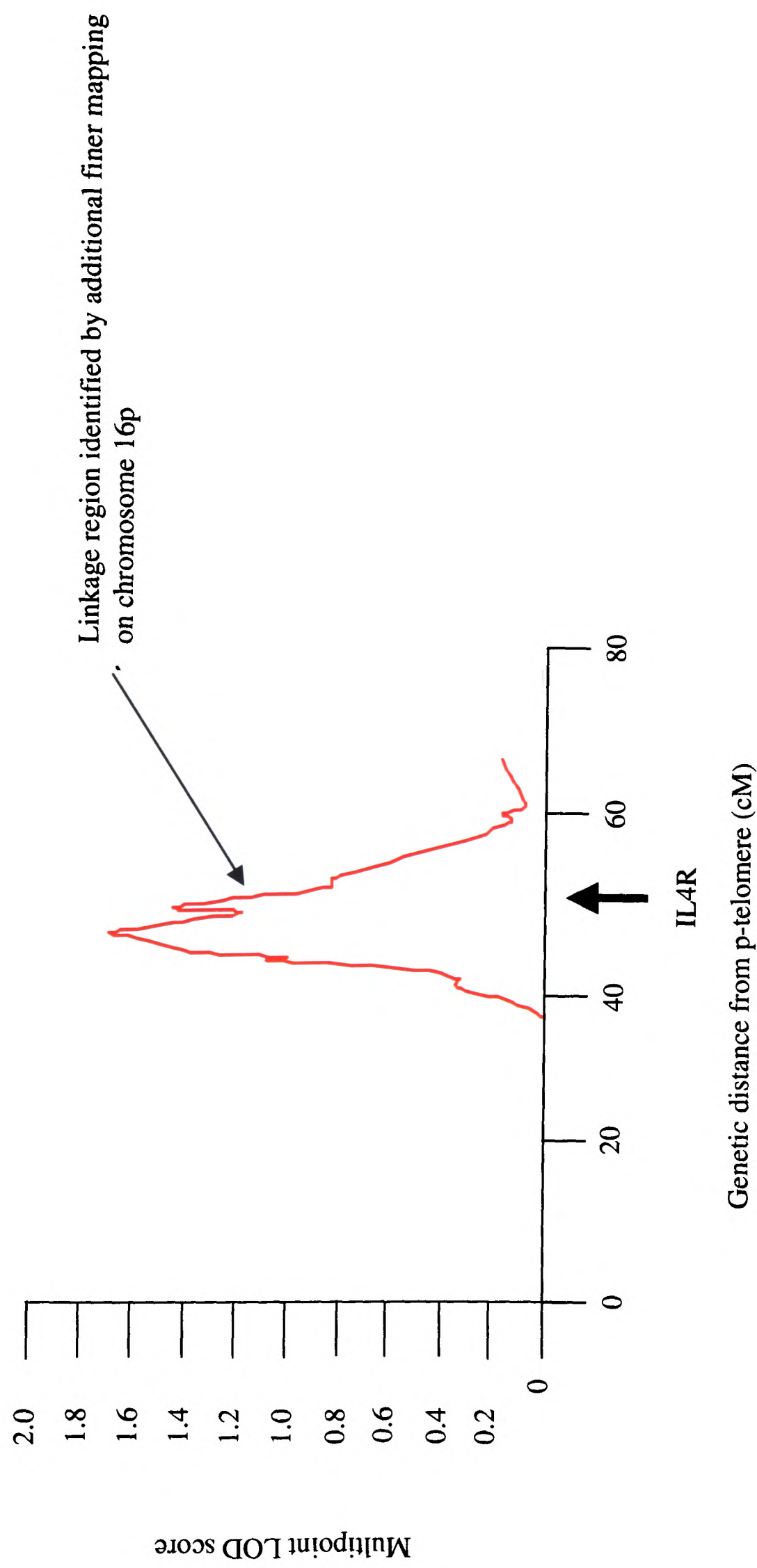


Figure 6.1: Approximate position of the IL4R candidate gene relative to the chromosome 16p linkage region.

The strong biological evidence for a role of the IL4R gene in the OA disease process (see below), coupled with its position within both the linkage region identified by this study and within the region on chromosome 16 identified by Ingvarsson et al (2001) made this gene a strong candidate. Therefore, association studies were focussed on the IL4R gene.

6.2 Introduction: The interleukin - 4 receptor alpha chain gene

6.2.1 The role of IL4R in OA

Recent evidence suggests that chondrocyte function, and thus cartilage integrity, is partly regulated by mechanical stimulation. Optimal stimulation, such as that associated with physiologic levels of exercise, is thought to maintain cartilage integrity whilst abnormal over or under loading has been shown to result in cartilage matrix degradation and tissue remodelling and can lead to the development and progression of OA (Brandt et al. 1991; Jortikka et al. 1997; Blain et al. 2001; Long et al. 2001).

Mechanical stimulation *in vitro* of chondrocytes from normal tissue has been demonstrated to result in (amongst other anabolic responses), increased production of aggrecan, increases in proteoglycan synthesis, decreased matrix metalloproteinase 3 (MMP3) mRNA synthesis, changes in membrane potential and increases in intracellular calcium and cAMP levels (Buschmann et al. 1995; Millward-Sadler et al. 2000). These changes are dependent on $\alpha_5\beta_1$ integrins and SK (calcium sensitive small potassium) channels. The cytokine interleukin 4 (IL4) has been demonstrated in *in vitro* studies to be the active autocrine or paracrine signalling molecule in this

mechanotransduction pathway (Millward-Sadler et al. 1999; Salter et al. 2002). Figure 6.2a details the proposed pathway of action of mechanotransduction in normal chondrocytes.

Although IL4 signalling pathways in the immune system are highly characterised there is currently little information on IL4 mediated signalling pathways in non-immune cells. In order to initiate signalling IL4 binds to the IL4R chain which then heterodimerizes with either the common gamma (γ) chain (also a component of receptors for IL2, 7, 9, 15 and 21), to form the Type I receptor, or with the IL13RA1 subunit, to form the Type II receptor. Formation of the different receptor types results in the activation of alternative downstream signalling pathways. Type I receptor signalling acts via Janus Kinase 1 (JAK1) and JAK3 whilst Type II receptor signalling appears to act via JAK1 and Tyrosine kinase 2 (TYK2) (Ryan et al. 1996; Murata and Puri 1997). IL4 mediated signalling results in the upregulation of a number of transcription factors including signal transducer and activator of transcription 6 (STAT6). This transcription factor acts to enhance IL4R gene expression in addition to upregulating the expression of many other IL4 responsive genes. There is likely to also be a role for a number of additional molecules, both inhibitory and activating, including SOCS (suppressors of cytokine signalling) and SHPs (SH2-containing phosphatases) with many of these molecules acting in a cell-type specific manner (Haque et al. 1998). In cells of a non-haematopoietic lineage IL4 appears to signal through the Type II receptor (Pan and Rothman 1999).

In vitro studies of chondrocytes from OA cartilage have revealed a number of differences in response to mechanotransduction compared with chondrocytes from

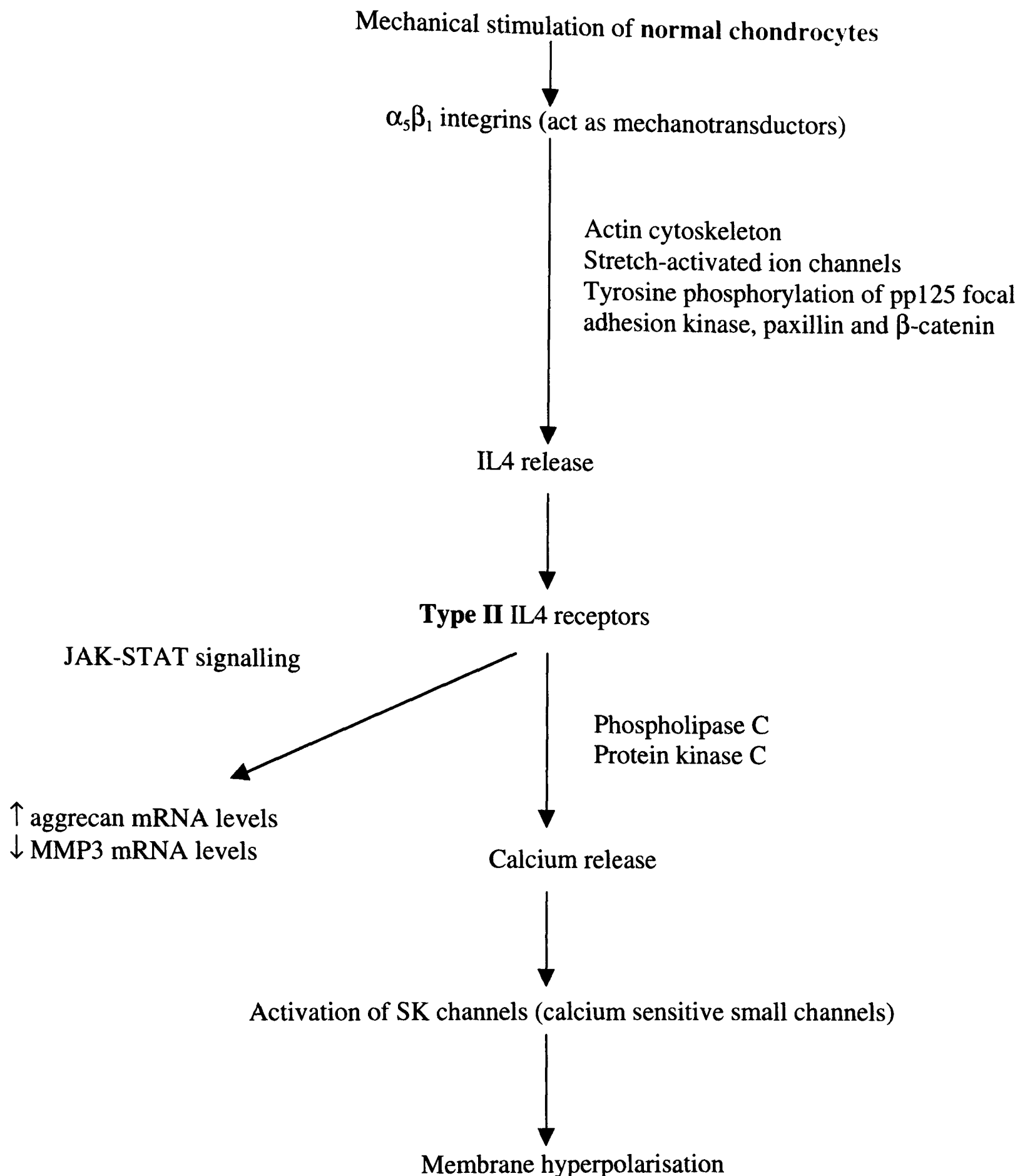


Figure 6.2a: Proposed mechanism of action for the observed effects of mechanical stimulation on normal chondrocytes. Many of the details of the pathway, including downstream signalling of the IL4R are yet to be determined. Based on work by Salter et al (2002).

normal tissue. In OA chondrocytes a membrane depolarisation is seen in response to cyclic stimulation, with pressure pulses at 50 kPa above atmospheric pressure at 0.33 Hz, in contrast to the hyperpolarisation observed in normal chondrocytes under these conditions. In addition the anabolic response, seen in normal chondrocytes, of increased aggrecan and decreased MMP3 production is not observed in OA chondrocytes (Salter et al. 2002). *In vitro* studies have demonstrated that while OA chondrocytes show a similar mechanotransduction pathway up to the release of IL4 (with the possible exception of the response of OA chondrocytes to mechanical stimulation not involving the cytoskeleton or protein kinase C), the downstream signalling pathways appear to be very different. In OA chondrocytes IL4 signalling appears to be via the Type I IL4R and results in the activation of tetrodotoxin-sensitive sodium channels with a subsequent membrane depolarisation. This is as opposed to the signalling via Type II receptors, activation of apamin-sensitive SK channels and membrane hyperpolarisation observed in normal chondrocytes (Salter et al. 2002). Activation of the different IL4 receptor types is likely to result in signalling via different JAK/STAT pathways possibly accounting for the differences in gene expression observed between normal and OA chondrocytes in response to mechanical stimulation. In OA chondrocytes membrane depolarisation may be blocked by antibodies either against IL4 or IL1B, suggesting an additional role for IL1B in OA chondrocytes, not observed in normal chondrocytes (Salter et al. 2002). Figure 6.2b details the proposed pathway for the response of OA chondrocytes to mechanotransduction. These studies suggest a potentially catabolic response in OA chondrocytes to mechanotransduction compared with the anabolic response seen in normal chondrocytes. These observed differences between normal and OA

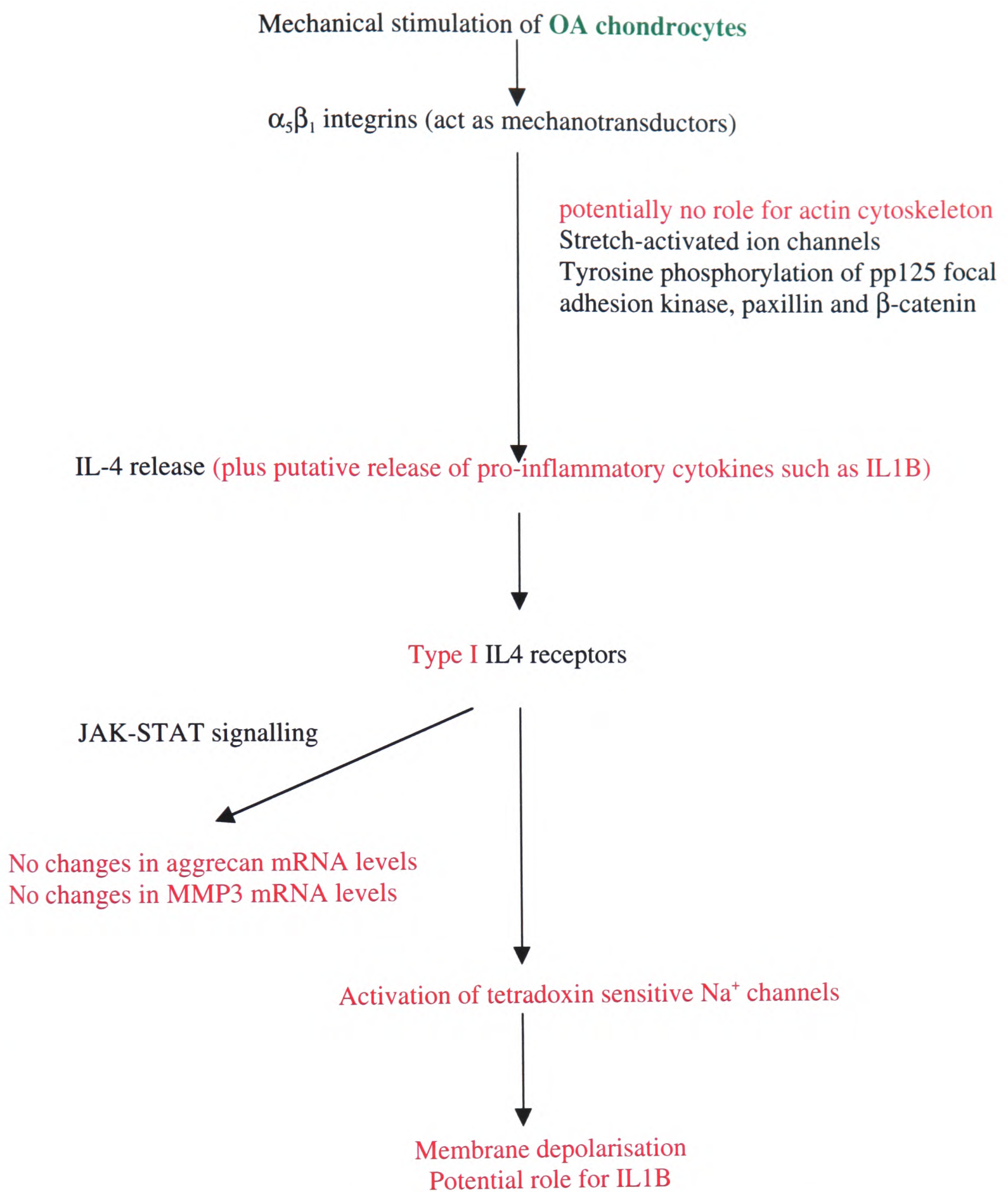


Figure 6.2b: Proposed mechanism of action for the observed effects of mechanical stimulation on OA chondrocytes. Many of the details of the pathway, including downstream signalling of the IL4R are yet to be determined. Differences from responses in normal chondrocytes are highlighted in red. Based on work by Salter et al (2002).

chondrocytes are difficult to explain at present but Salter et al (2002) suggest that they may be related to cytokine receptor expression or function.

There has recently been a resurgence of interest in the possibility of the involvement of cytokines generally in OA. It is thought that the balance between catabolic and anabolic events in normal cartilage is maintained through the action of cytokines, especially IL1 and TNFA, on the chondrocyte (Goldring 2000a, b). IL4 has also been demonstrated in *in vivo* and *in vitro* studies to be directly chondroprotective, regulating the turnover of tissue inhibitors of metalloproteinases (TIMPs) and MMPs (Cawston et al. 1996; Nemoto et al. 1997; Lubberts et al. 1999). IL4 has also been shown to act indirectly, both by suppressing the actions of IL1 and IL17 (which act to mediate MMP3 production and inhibit proteoglycan synthesis) and by inhibiting TNFA-induced PGE₂ production (Alaaeddine et al. 1999; Lubberts et al. 1999, 2000).

The IL4R gene then clearly is expressed in tissues of the arthritic joint and has a biologically plausible role in the OA disease process. Its position at 16p12.1-p11.2, places it well within the region showing evidence for linkage on 16p identified by this study. It also resides within the region of linkage detected by Ingvarsson et al (2001) in the Icelandic hip OA pedigree (section 3.8). For these reasons it is considered to be a strong candidate OA susceptibility gene.

6.2.2 IL4R gene structure

The IL4R gene encodes for a 90 KDa protein of 825 amino acids. The first twenty-five amino acids comprise a signalling peptide which is cleaved to generate the mature protein. Although all three major databases (Ensembl, NCBI and UCSC) are

concordant on the size and number of the coding exons there is some discrepancy as to the number and size of non-coding exons 5' of the coding region. This discrepancy is also observed between publications and between different builds of the genomic databases. Ensembl predicts that the transcript is 3545 bp long and consists of twelve exons of which the first three are non-coding. The start codon resides within the fourth exon. The current build of both the UCSC and NCBI databases, which share much detail, agree with the first exon and with exons 5 through 12 but have an extended exon 2, which is 81 bp longer than that predicted by Ensembl, and no exon 3 compared with the Ensembl database. In addition they have one extra base in exon 4. For the ease of communication, all references to exon numbers in this and subsequent chapters refer to numbering as per the Ensembl database, April 2003 release. It should be remembered however, that the exact number and sequence of the non-coding exons remains unclear. It is possible that multiple non-coding exons occur which may be alternatively spliced to yield different transcripts as has been suggested to be the case for numerous genes (Graveley 2001). The absence of a TATA box preceding the proposed mRNA start site has led to the postulation of multiple transcription initiation sites of the IL4R gene in humans (Kruse et al. 1999b). A summary of the size of the introns and exons of the human IL4R gene is found in table 6.2

In addition to the constitutive exons there are reports of an alternative 114 bp exon between exons 8 and 9 which, when present in the mature transcript, results in the formation of a truncated, soluble form of the IL4R protein (sIL4R/splice) This truncation is the result of inclusion of a premature stop codon, in the alternative exon, which creates a protein lacking the transmembrane and intracellular regions. A distinct soluble form of the receptor can also be produced by proteolytic shedding

Exon N°	Intron N°	Size (bp)
1		24
	1	16045
2		52
	2	9670
3		28
	3	369
4		89
	4	1847
5		139
	5	2609
6		152
	6	1446
7		152
	7	5921
8		157
	8	3111
9		100
	9	3008
10		79
	10	1771
11		50
	11	1436
12		1576
	3'UTR	947

Table 6.2: Exon and intron sizes for the IL4R gene, as per Ensembl database April 2003 release (www.ensembl.org).

(Kruse et al. 1999a, b). This soluble form of the IL4R protein is designated sIL4R/prot is proposed to play a physiological role in either inhibiting or co-stimulating IL4 signalling (Jung et al. 1999a, b).

6.2.3 IL4R chain protein structure and function

Two systems have been used for numbering amino acids for the IL4R chain protein. This chapter follows the example of Deichmann et al (1997) of numbering amino acids according to their position in the mature protein.

The extracellular domain of the protein, encoded for by amino acids 1 to 207 of the mature peptide (Hage et al. 1998), is termed the IL4 binding protein (IL4-BP) and is responsible for almost all of the binding affinity of the receptor complex (Hage et al. 1998; Zhang et al. 2002). It represents the minimal version of a cytokine binding homology domain (CHR) and has two fibronectin type III domains (Mueller et al. 2002). It is likely that this epitope consists of two independent binding clusters, each of which contains a central polar or charged side chain (at amino acids 72 and 183) surrounded by hydrophobic groups (Hage et al. 1999; Zhang et al. 2002).

The transmembrane domain of the protein consists of about 23 or 24 amino acids between amino acids 208 and 230/231 of the mature protein (predicted).

The intracellular domain, from amino acids 232 to 800 of the mature peptide (predicted), contains five conserved tyrosine residues at amino acids 472, 550, 578, 606 and 688. These are termed Y1-5 respectively (Kashiwada et al. 2001). All five are sites of potential receptor phosphorylation, creating binding sites for a number of

downstream signalling molecules. There are four defined functional regions in the intracellular domain, three of which are centred around these residues (Pan and Rothman 1999). The region immediately proximal to the membrane is involved in interactions with Jak-1. The second region, between amino acids 412 and 532, is termed the insulin/IL4 receptor (I4R) motif. This region encompasses Y1 and is important in the process of recruitment and activation of a number of signalling molecules involved in IL4 induced proliferation including IRS (insulin receptor substrate) (Keegan et al. 1994; Pernis et al. 1995). Distal to amino acid 532 is the gene-activation domain which contains the functionally redundant tyrosine residues Y2-Y4. This region is important in the induction of signalling pathways leading to gene activation and is likely to be the docking site for STAT6 (Ryan et al. 1996; Wang et al. 1996). The final region is the ITIM (immuno-receptor tyrosine-based inhibitory motif). Containing Y5 this carboxy terminal domain is postulated to have a role in the negative regulation of IL4 induced proliferation and signalling pathways, possibly via the dephosphorylation of activated STAT6 (Kashiwada et al. 2001; Hanson et al. 2003). Cells expressing mutant forms of this region of the IL4R chain show a hyperproliferative response to IL4 stimulation (Kashiwada et al. 2001). It may also have a role in the observed anti apoptotic effect of IL4 signalling (Pan and Rothman 1999).

The three-dimensional structure for the free IL4R chain has not been determined so far.

6.2.4 IL4R SNPs and haplotypes

The IL4R gene shows a high level of sequence variation, much higher than that observed for other cytokine receptor genes (Hackstein et al. 1999). This has led to speculation that natural selection, possibly as a result of the activity of pathogens, has influenced the evolution of this gene (Wu et al. 1999). Because of its reported potential association with asthma and atopy there are numerous published studies on polymorphisms in the IL4R gene and on their putative functional effects.

Most, if not all work, investigating the functional effects of SNPs have focussed on cells of the haemopoietic/immune system where the IL4R chain forms a heterodimer with the common γ chain to form a type I receptor. It is possible therefore that such results can not be extrapolated exactly to IL4 mediated signalling via the type II receptor as is seen in normal chondrocytes.

It is generally agreed that the I50V SNP, which results in the substitution of a valine for an isoleucine at amino acid 50 in the extracellular region of the receptor, upregulates the response of the receptor to IL4 and augments STAT6 transcription factor activation (Mitsuyasu et al. 1999). The mechanism by which this upregulation is achieved however is unclear as the variant does not influence the binding affinity of the receptor. The S478P polymorphism, resulting in the substitution of a proline for a serine at amino acid 478 has been associated with alterations in STAT6 binding. This substitution, in the intracellular portion of the receptor, alters the polarity of the protein and as such may induce a conformational change in the receptor (Kruse et al. 1999a). The Q551R polymorphism results in the substitution of a glutamine for an arginine residue at amino acid 551 adjacent to Y2. The variant form of this SNP

changes the sequence of the Src homology 2 (SH2) domain and has been reported to result in a decreased binding of the negative regulator SHP-1 (src homology 2-containing phosphatase-1: a phosphotyrosine phosphatase involved in signal termination) (Kruse et al. 1999). Hershey et al (1997) demonstrated that peripheral blood B-lymphocytes from individuals with the R551 variant had increased CD23 (cluster of differentiation 23) induction in response to IL4 stimulation. Failure to replicate these results using transfected cell lines however, suggests that the action of this SNP may be context dependent, having functional effects only in the presence of other polymorphisms, or be due to linkage disequilibrium with another SNP (Mitsuyasu et al. 1999; Wang et al. 1999). Hackstein et al (1999) found evidence for six non-synonymous SNPs (I50V, E375A, C406R, S411L, S478P and Q551R) being associated with lowered IgE levels. In addition to the action of these coding SNPs, the presence of the promoter SNP C-3223T, occurring 3223 bases upstream of the start of exon 1, has been found to be significantly associated with lowered sIL4R levels with the observation of a possible gene dosage effect (Hackstein et al. 2001).

In addition to these effects of individual SNPs, there have been several reports of potential interactions or cumulative effects. It has been reported that the presence of both the S478P and Q551R polymorphisms in the same receptor may affect the IRS pathway, which is not affected by either SNP in isolation. This results in a magnification of effects on receptor signalling (Kruse et al. 1999a). A recent study by Risma et al (2002) found no evidence for effect of either the V50 (in contradiction to other reports) or R551 alleles alone on CD23 induction but found that the presence of both substitutions resulted in expression of an IL4R with enhanced sensitivity to IL4. These results were supported by association analysis, with a significant association to

atopic asthma being observed only in the presence of the V50R551 combination (Risma et al. 2002).

To study the potential role of the IL4R chain in encoding susceptibility to OA, variants within the gene were identified using both published reports and direct sequencing and were genotyped in cohorts of probands, cases and controls. The pattern of LD across the region was established and associations between alleles, genotypes and haplotypes and disease were examined by χ^2 analysis using standard contingency tables. ODDs ratios were also calculated. The relationship between genetic variation and the evidence for linkage was also examined.

Association work was concentrated on female cases who had undergone total hip replacement (THR) for OA as it was the female-THR families in whom the highest evidence for linkage to chromosome 16p was detected (chapter 3.7.1). For all PCR reactions and restriction digests in this chapter, the reaction conditions and full primer sequences are given in Appendices 4-6.

6.3 Sequencing of the IL4R gene

The IL4R gene was sequenced both to investigate the frequencies of previously identified polymorphisms within our population and to determine any novel polymorphisms in individuals with OA that could encode for susceptibility to disease.

All coding exons, 5' and 3' flanking intronic sequences and the 3'UTR of the IL4R were sequenced in at least 39 OA cases. Details of primers and PCR conditions may

be found in Appendix 4. Exons 2 and 3, which are non-coding, were also sequenced. Exon 1 failed, after repeated attempts, to amplify, and so was not included.

Results of both number of individuals successfully sequenced and length of sequence flanking the exons covered are summarised in table 6.3a. Table 6.3b details SNPs discovered, position, nature of the polymorphism, minor allele frequency and database SNP reference (www.ncbi.nlm.nih.gov/SNP/) where available. Full details of sequence and position of SNPs may be found in Appendix 8. In summary, there are seven intronic SNPs, four of which have a minor allele frequency $\geq 5\%$ and may therefore be classified as common. There are sixteen exonic SNPs, six of which are synonymous with four of these having a minor allele frequency $\geq 5\%$. Of the ten non-synonymous SNPs, eight have a minor allele frequency $\geq 5\%$. Most of the SNPs identified have been previously reported.

6.4 Association studies - Stage 1

6.4.1 SNP selection

A subset of the SNPs identified by sequencing were selected for use in the association study based both on their frequency and reported or potential functional effects. Based on epidemiological association studies, it has been suggested that the variants used in association studies should, where possible be those most likely to have functional effects and therefore be disease causing (Tabor et al. 2002). SNPs selected initially, were therefore those most likely to have a functional effect that in addition had a minor allele frequency ≥ 0.05 .

Exon	Exon size	N° of bases 5' of exon sequenced	N° of bases 3' of exon sequenced	Average N° of individuals sequenced ¹
2	52	135	104	48
3	28	11	71	39
4	89	70	53	48
5	139	89	93	46
6	152	18	16	45
7	152	110	66	45
8	157	87	99	47
9	100	182	76	39
10	79	147	78	46
11	50	109	109	44
12 ²	2523	96	17 ³	43

¹ For some exons more than one sequencing reaction was necessary to cover the whole exon

² Sequence includes 3'UTR

³ 17 bp 3' of 3'UTR sequenced

Table 6.3a: Summary of sequencing of the IL4R gene

Exon/Intron number	SNP	Nature	Distance from Exon	Amino acid substitution	Nucleotide number ³	Amino acid position ⁴	Minor allele frequency	rs reference ⁵
Intron 4	C/T	Intronic	-85 ¹		11971		0.17	3024544
Intron 4	G/T	Intronic	-27 ¹		12029		0.01	not in database
Exon 5	C/T	Synonymous			12093	11	0.02	not in database
Exon 5	C/T	Non-synonymous		Arginine/Cysteine	12160	34	0.01	not in database
Exon 6	A/G	Non-synonymous		Isoleucine/Valine	14817	50	0.40	1805010
Intron 6	A/G	Intronic	-4 ¹		16398		0.09	3024570
Exon 7	C/T	Synonymous			16541	142	0.10	3024571
Intron 8	T/C	Intronic	-147 ¹		25596		0.08	3024636
Intron 8	G/A	Intronic	-59 ¹		25684		0.04	not in database
Intron 11	C/T	Intronic	-96 ¹		32091		0.01	not in database
Intron 11	A/C	Intronic	-15 ¹		32172		0.24	3024676
Exon 12	T/C	Synonymous			32226	288	0.05	2234897
Exon 12	A/C	Non-synonymous		Glutamate/Alanine	32486	375	0.16	1805011
Exon 12	G/T	Synonymous			32529	389	0.16	2234898
Exon 12	T/C	Non-synonymous		Cysteine/Arginine	32578	406	0.18	1805012
Exon 12	C/T	Synonymous			32580	406	0.01	2234899
Exon 12	T/C	Synonymous			32586	408	0.19	2234900
Exon 12	C/T	Non-synonymous		Serine/Leucine	32594	411	0.06	1805013
Exon 12	T/C	Non-synonymous		Serine/Proline	32794	478	0.24	1805015
Exon 12	A/G	Non-synonymous		Glutamine/Arginine	33014	551	0.33	1801275
Exon 12	C/T	Non-synonymous		Proline/Serine	33310	650	0.02	3024678
Exon 12	T/G	Non-synonymous		Serine/Alanine	33541	727	0.06	1805016
Exon 12	T/C	Non-synonymous		Serine/Proline	33643	761	0.01	1805014
3'UTR	T/C	3'UTR	+6 ²		33771		0.10	not in database
3'UTR	T/C	3'UTR	+8 ²		33773		0.01	3024680
3'UTR	C/G	3'UTR	+88 ²		33853		0.05	2234924
3'UTR	G/T	3'UTR	+280 ²		34045		0.01	not in database
3'UTR	G/A	3'UTR	+392 ²	34156			0.38	1049631
3'UTR	A/G	3'UTR	+636 ²	34401			0.36	8832
3'UTR	C/T	3'UTR	+738 ²	34503			0.04	8674
3'UTR	G/C	3'UTR	+833 ²	34598			0.01	not in database

¹Position relative to preceding exon

²Position relative to termination codon

³First nucleotide of exon 2 (Ensembl April 2003 release) = nucleotide 1

⁴Numbering as per mature protein

⁵rs = reference SNP from www.ncbi.nlm.nih.gov/SNP/.

Table 6.3b: Summary of SNPs in IL4R detected by sequencing.

Table 6.4a summarises the ten SNPs used for association studies. Selected were the seven non-synonymous SNPs identified by sequencing that have a minor allele frequency ≥ 0.05 in addition to one synonymous SNP, N142N, included to provide more complete coverage of the gene. In addition to these eight SNPs identified by sequencing, two common promoter SNPs were also genotyped (Hackstein et al. 2001).

6.4.2 SNP genotyping

All ten SNPs were genotyped using either forced or natural restriction digest analysis in the female probands from the 146 female-THR families used in the finer mapping linkage analysis studies for chromosome 16p (chapter 3.7.1) and in 399 age-matched female controls. Details of primer sequence, amplification conditions and restriction enzymes can be found in Appendices 5 and 6. C406R demonstrated unreliable amplification as a result of multiple additional SNPs within the region and was therefore excluded from further studies. It is clear that this SNP is in very strong LD with E375A, and as such little novel information was lost by its exclusion. The accuracy of SNP genotyping was confirmed by sequencing in a minimum of 43 individuals.

6.4.3 Association analysis

The allele frequencies for each of the SNPs in probands and controls were compared by χ^2 analysis using standard 2 x 2 contingency tables. Table 6.4b shows the numbers and frequencies of plus (+) and minus (-) alleles for all nine SNPs in the probands and the controls as well as the p-values derived from comparison of the allele numbers. A

SNP	Position on chromosome 16 ¹	Distance from SNP above	Exon	Base change	Amino Acid change	Nature of change	Nomenclature	rs reference ²
1	27249353		NA	C/T	NA	promoter	C-3223T	2057768
2	27250662	1309	NA	T/C	NA	promoter	T-1914C	2107356
3	27283461	32799	6	A/G	Isoleucine/Valine	non- synonymous	I50V	1805010
4	27285185	1724	7	C/T	NA	synonymous	N142N	3024571
5	27301130	15945	12	A/C	Glutamate/Alanine	non- synonymous	E375A	1805011
6	27301222	92	12	T/C	Cysteine/Arginine	non- synonymous	C406R	1805012
7	27301238	16	12	C/T	Serine/Leucine	non- synonymous	S411L	1805013
8	27301438	200	12	T/C	Serine/Proline	non- synonymous	S478P	1805015
9	27301658	220	12	A/G	Glutamine/Arginine	non- synonymous	Q551R	1801275
10	27302185	527	12	T/G	Serine/Alanine	non- synonymous	S727A	1805016

¹ As per dbSNP (June 2003) (www.ncbi.nlm.nih.gov/SNP/)

² rs = reference SNP from www.ncbi.nlm.nih.gov/SNP/.

Table 6.4a: Summary of the IL4R SNPs used for association analysis. SNP nomenclature as per amino acid position in mature peptide for coding SNPs. Promoter SNPs numbered as per position upstream of gene. Position 125165 of human Chromosome 16 BAC clone CIT987SK-582J2 is designated as position -1 (Hackstein 2001). NA = Not Applicable.

SNP	Controls		146 Female-TIIR Probands		
	(+)	(-)	(+)	(-)	P-value
	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	
C-3223T	547(0.70)	231(0.30)	202(0.72)	80(0.28)	0.74
T-1914C	441(0.57)	339(0.43)	160(0.56)	128(0.44)	0.83
I50V	327(0.43)	441(0.57)	126(0.45)	156(0.55)	0.59
N142N	732(0.93)	54(0.07)	259(0.90)	29(0.10)	0.11
E375A	87(0.12)	669(0.88)	28(0.10)	246(0.90)	0.64
S411L	733(0.97)	21(0.03)	271(0.94)	17(0.06)	0.03
S478P	663(0.85)	115(0.15)	242(0.83)	50(0.17)	0.40
Q551R	142(0.18)	642(0.82)	58(0.20)	230(0.80)	0.51
S727A	726(0.97)	24(0.03)	269(0.93)	19(0.07)	0.02

Table 6.4b: Numbers and frequencies of (+) and (-) alleles for each of the IL4R SNPs in a cohort of female probands from the 146 female-THR families used in linkage analysis and a cohort of 399 age-matched female controls. The final column shows the p-value for the χ^2 test using standard contingency table analysis.

(+) allele corresponds to a cut site for the specific restriction enzyme used being present whilst a (-) allele corresponds to no cut site. All SNPs conform to Hardy-Weinberg equilibrium in the controls. Two of the SNPs, S411L and S727A, show a significant association ($P \leq 0.05$) with OA with p-values of 0.03 and 0.02 respectively. For both SNPs the minor allele is over represented in the cases compared with the controls. None of the other SNPs show a significant difference ($P \leq 0.05$) in allele frequencies between cases and controls. N142N however shows a suggestion of association with a p-value of 0.11. Comparison of allele frequencies for this SNP shows that, as for S411L and S727A, the minor allele is elevated in the cases compared with the controls.

Genotype frequencies for the 146-THR probands and the 399 controls were also compared by χ^2 analysis using standard 2 x 3 contingency tables (table 6.4c). Three SNPs show significant associations. For SNP S727A the p-value (0.02) is the same as the p-value for the comparison of allele frequencies ($P = 0.02$) [table 6.4b] indicating that there is no increased risk associated with a particular genotype. For SNP S411L, the p-value for comparison of genotypes ($P = 0.02$) is slightly less than that for comparison of allele frequencies ($P = 0.03$, [table 6.4b]). Finally, for SNP N142N, the p-value for comparison of genotype frequencies ($P = 0.05$) is more significant than that for the comparison of allele frequencies ($P = 0.11$ [table 6.4b]). Observation of the genotype frequencies shows that the (+/-) genotype is elevated in the cases compared with the controls.

ODDs ratios and 95% confidence intervals were calculated with respect to the minor allele for all nine SNPs (table 6.4d). Only two SNPs have confidence intervals that do

SNP	Controls			146 Female-THR probands			P-value
	(++)	(+-)	(--)	(++)	(+-)	(--)	
	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	
C-3223T	192 (0.49)	163 (0.42)	34 (0.09)	77(0.55)	48(0.34)	16(0.11)	0.23
T-1914C	128 (0.33)	185 (0.47)	77 (0.20)	45(0.31)	70(0.49)	29(0.20)	0.94
I50V	70 (0.18)	187 (0.49)	127 (0.33)	26(0.18)	74(0.52)	41(0.29)	0.67
N142N	343 (0.87)	46 (0.12)	4 (0.01)	115(0.80)	29(0.20)	0 (0)	0.05 ¹
E375A	5 (0.01)	77 (0.20)	296 (0.78)	1(<0.01)	26(0.19)	110(0.80)	0.80
S411L	356 (0.94)	21 (0.06)	0 (0)	127(0.88)	17(0.12)	0 (0)	0.02 ²
S478P	285 (0.73)	93 (0.24)	11 (0.03)	98(0.67)	46(0.32)	2(0.01)	0.15
Q551R	14 (0.04)	114 (0.29)	264 (0.67)	5(0.03)	48(0.33)	91(0.63)	0.64
S727A	351 (0.94)	24 (0.06)	0 (0)	125(0.87)	19(0.13)	0 (0)	0.02 ²

¹ Analysed with 2 x 2 contingency table, combining (+-) and (--) cells

² Analysed with 2 x 2 contingency table, ignoring empty cells

Table 6.4c: Genotype numbers and frequencies for the nine IL4R gene SNPs for the 146 female-THR probands and the 399 female controls. The final column shows the p-value for comparison of probands and controls by χ^2 analysis.

SNP	ODDs ratio	95% confidence interval
C-3223T	0.94	0.69 to 1.27
T-1914C	1.04	0.79 to 1.37
I50V	1.09	0.83 to 1.43
N142N	1.52	0.95 to 2.44
E375A	0.88	0.56 to 1.37
S411L	2.19	1.14 to 4.21
S478P	1.19	0.83 to 1.71
Q551R	1.14	0.81 to 1.60
S727A	2.14	1.15 to 3.96

Table 6.4d: ODDs ratios and 95% confidence intervals with respect to the minor allele of the IL4R SNPs in a cohort of female probands from the 146 female -THR families and a cohort of 399 age-matched female controls.

not encompass 1, the value reflecting 'no effect'. For SNP S411L the ODDs ratio is 2.19 (95% confidence interval 1.14 to 4.21). This SNP also shows a significant association with χ^2 analysis ($P = 0.03$). The wideness of the confidence interval indicates however that the study size is too small to accurately predict the true effect of this SNP. SNP S727A has an ODDs ratio of 2.14 (95% confidence interval 1.15 to 3.96). This SNP also had a significant p-value ($P = 0.02$) with χ^2 analysis. The confidence interval for this SNP is also wide indicating that the sample size may be too small to precisely estimate the effect of this SNP.

6.4.5 SNP genotyping and association studies - Stage 2

The two SNPs, S411L and S727A, showing evidence for association, plus SNP N142N with a p-value approaching significance and a potential significant association of the (+/-) genotype, were then genotyped in a separate cohort of 310 unrelated, female-THR cases. Again the allele numbers for each of the SNPs in cases and controls were compared by χ^2 analysis using standard 2 x 2 contingency tables. Allele numbers, frequencies and p-values for the 310 female-THR cases in addition to the previously typed 146 female-THR probands and controls, are shown in table 6.4e. Evidence for association ($P \leq 0.05$) is observed for all three SNPs in the new cohort of 310 female-THR cases: N142N ($P = 0.04$), S411L ($P = 0.04$) and S727A ($P = 0.007$).

Since all of these SNPs show evidence for association in this case cohort, and as evidence for association for N142N is higher in this cohort than for the 146 female-THR probands, it was decided to genotype all other SNPs in this case cohort too. This allowed for investigation of haplotype frequencies and the structure of linkage disequilibrium across the gene in all affected individuals as well as the controls.

SNP	Controls		146 Female-THR Probands				310 Female-THR Cases				Probands and cases combined (n=456)			
			(+) (-)		(+) (-)		(+) (-)		(+) (-)		(+) (-)		(+) (-)	
	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	P-value	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	P-value
N142N	732(0.93)	54(0.07)	259(0.90)	29(0.10)	29(0.10)	0.11	547(0.90)	61(0.10)	61(0.10)	0.04	806(0.90)	90(0.10)	90(0.10)	0.03
S411L	733(0.97)	21(0.03)	271(0.94)	17(0.06)	17(0.06)	0.03	577(0.95)	31(0.05)	31(0.05)	0.04	848(0.95)	48(0.05)	48(0.05)	0.01
S727A	726(0.97)	24(0.03)	269(0.93)	19(0.07)	19(0.07)	0.02	529(0.93)	37(0.07)	37(0.07)	0.007	798(0.93)	56(0.07)	56(0.07)	0.004

Table 6.4e: Numbers and frequencies of (+) and (-) alleles for three of the IL4R SNPs in a cohort of female probands from the 146 female-THR families, a cohort of 310 female-THR cases and for all 456 female-THR probands and cases combined, and for a cohort of 399 age-matched female controls. The final column of each block shows the p-value for the χ^2 test using standard contingency table analysis.

Table 6.4f details results of association analysis for all nine SNPs in the 310 female-THR cases and in the 146 female-THR probands. There are a total of four SNPs which show evidence for association in this second cohort. In addition to N142N, S411L and S727A, there is evidence for association for Q551R with p-value of 0.04. Additionally, SNP S478P shows a p-value of 0.07 which approaches significance. Observation of allele frequencies shows all differences between cases and controls are due to an elevation of the minor allele in the cases (the (+) allele for N142N, S411L, S478P and S727A and the (-) allele for Q551R). The association of SNPs S411L and S727A in both the probands and the cases serves to increase confidence of their involvement in OA susceptibility. SNP N142N also showed some suggestion of association in the female-THR proband cohort ($P = 0.11$). SNPs S478P and Q551R showed no evidence of association in the female-THR proband cohort. This may reflect lack of power to detect association in the female-THR proband cohort due to its relatively small size.

In order to maximise power and to minimise the number of tests, all subsequent analysis was carried out for all 465 probands and cases combined.

ODDs ratios and 95% confidence intervals were calculated with respect to the minor allele for all nine SNPs for all probands and cases combined compared with controls (table 6.4g). Only three SNPs have confidence intervals that do not encompass 1. For SNP N142N the ODDs ratio is 1.51 (95% confidence interval 1.06 to 2.51). This SNP also shows a significant association with χ^2 analysis ($P = 0.03$). This confidence interval is wide indicating that the study has limited power to accurately measure the true effect. The confidence interval is also close to 1 which indicates that the effect of

SNP	Controls		146 Female-THR Probands				310 Female-THR Cases				Probands and cases combined (n=456)			
	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	P-value		N° (Freq)	N° (Freq)	P-value		N° (Freq)	N° (Freq)	P-value	
C-3223T	547(0.70)	231(0.30)	202(0.72)	80(0.28)	0.74		451(0.73)	167(0.27)	0.31		654(0.73)	247(0.27)	0.34	
T-1914C	441(0.57)	339(0.43)	160(0.56)	128(0.44)	0.83		343(0.59)	271(0.44)	0.85		503(0.56)	399(0.44)	0.79	
I50V	327(0.43)	441(0.57)	126(0.45)	156(0.55)	0.59		257(0.42)	351(0.58)	0.95		383(0.43)	507(0.57)	0.89	
N142N	732(0.93)	54(0.07)	259(0.90)	29(0.10)	0.11		547(0.90)	61(0.10)	0.04		806(0.90)	90(0.10)	0.03	
E375A	87(0.12)	669(0.88)	28(0.10)	246(0.90)	0.64		76(0.13)	528(0.87)	0.60		104(0.12)	774(0.88)	0.90	
S411L	733(0.97)	21(0.03)	271(0.94)	17(0.06)	0.03		577(0.95)	31(0.05)	0.04		848(0.95)	48(0.05)	0.01	
S478P	663(0.85)	115(0.15)	242(0.83)	50(0.17)	0.39		500(0.81)	114(0.19)	0.07		742(0.82)	164(0.18)	0.08	
Q551R	142(0.18)	642(0.82)	58(0.20)	230(0.80)	0.51		140(0.23)	472(0.77)	0.04		198(0.22)	702(0.78)	0.06	
S727A	726(0.97)	24(0.03)	269(0.93)	19(0.07)	0.02		529(0.93)	37(0.07)	0.007		798(0.93)	56(0.07)	0.004	

Table 6.4f: Numbers and frequencies of (+) and (-) alleles for each of the nine IL4R SNPs in a cohort of female probands from the 146 female-THR families, a cohort of 310 female-THR cases, for all 456 probands and cases combined, and for a cohort of 399 age-matched female controls. The final column of each block shows the p-value for the χ^2 test using standard contingency table analysis.

SNP	ODDs ratio	95% confidence interval
C-3223T	0.89	0.72 to 1.11
T-1914C	1.03	0.85 to 1.25
I50V	1.02	0.84 to 1.24
N142N	1.51	1.06 to 2.15
E375A	1.03	0.76 to 1.40
S411L	1.98	1.17 to 3.33
S478P	1.27	0.98 to 1.65
Q551R	1.28	1.00 to 1.62
S727A	2.12	1.30 to 3.46

Table 6.4g: ODDs ratios and 95% confidence intervals with respect to the minor allele for each of the nine IL4R SNPs in all 456 female-THR probands and cases and for a cohort of 399 age-matched female controls.

this SNP may be small. SNP S411L has an ODDs ratio of 1.98 (95% confidence interval 1.17 to 3.33). This SNP also has a significant p-value ($P = 0.01$) with χ^2 analysis. SNP S727A has an ODDs ratio of 2.12 (95% confidence interval 1.30 to 3.46). This SNP also has a significant p-value ($P = 0.004$) with χ^2 analysis. The confidence intervals for these two SNPs are also wide indicating that the sample size may be too small to precisely estimate the effect of these SNPs.

6.5 Linkage disequilibrium studies

Pairwise LD values for all nine SNPs were estimated for the dataset of all 855 individuals genotyped (all probands, cases and controls) using the GOLD (graphical overview of linkage, Abecasis and Cookson 2000) program (table 6.5). Figures 6.5a and b display the graphical overview of the pair-wise linkage disequilibrium for D' and r^2 on the physical map.

Consideration of D' values indicates a high degree of LD across the gene. Pairwise comparisons with r^2 however, reveal a much more varied pattern of LD as a result of the dependence of this statistic on allele frequencies at the two loci.

SNPs S478P and Q551R show a high level of LD with a r^2 value of 0.79. It may be therefore that the trend towards association observed for S478P is due to its strong LD with Q551R and not due to an independent association. SNPs S411L and S727A also show strong evidence of LD with a r^2 value of 0.80 and a D' value of 0.97. SNP N142N shows some evidence of LD with both SNPs S411L and S727A (r^2 values for comparisons are 0.33 and 0.36 respectively). Observation of the investigated

SNP 1	SNP 2	r ²	D'
C-3223T	T-1914C	0.305	0.993
C-3223T	I50V	0.461	0.929
C-3223T	N142N	0.032	0.926
C-3223T	E375A	0.009	0.426
C-3223T	S411L	0.012	0.836
C-3223T	S478P	0.023	0.541
C-3223T	Q551R	0.019	0.436
C-3223T	S727A	0.014	0.841
T-1914C	I50V	0.543	0.966
T-1914C	N142N	0.063	0.926
T-1914C	E375A	0.068	0.804
T-1914C	S411L	0.031	0.940
T-1914C	S478P	0.108	0.832
T-1914C	Q551R	0.124	0.791
T-1914C	S727A	0.039	0.839
I50V	N142N	0.128	1.000
I50V	E375A	0.007	0.195
I50V	S411L	0.035	0.774
I50V	S478P	0.044	0.410
I50V	Q551R	0.025	0.272
I50V	S727A	0.038	0.739
N142N	E375A	0.002	0.439
N142N	S411L	0.332	0.834
N142N	S478P	0.084	0.420
N142N	Q551R	0.057	0.393
N142N	S727A	0.359	0.811
E375A	S411L	0.001	0.478
E375A	S478P	0.662	0.993
E375A	Q551R	0.522	0.993
E375A	S727A	0.002	0.562
S411L	S478P	0.211	1.000
S411L	Q551R	0.168	1.000
S411L	S727A	0.802	0.969
S478P	Q551R	0.792	1.000
S478P	S727A	0.247	0.983
Q551R	S727A	0.194	0.981

Table 6.5: Pairwise LD values for the nine IL4R gene SNPs.

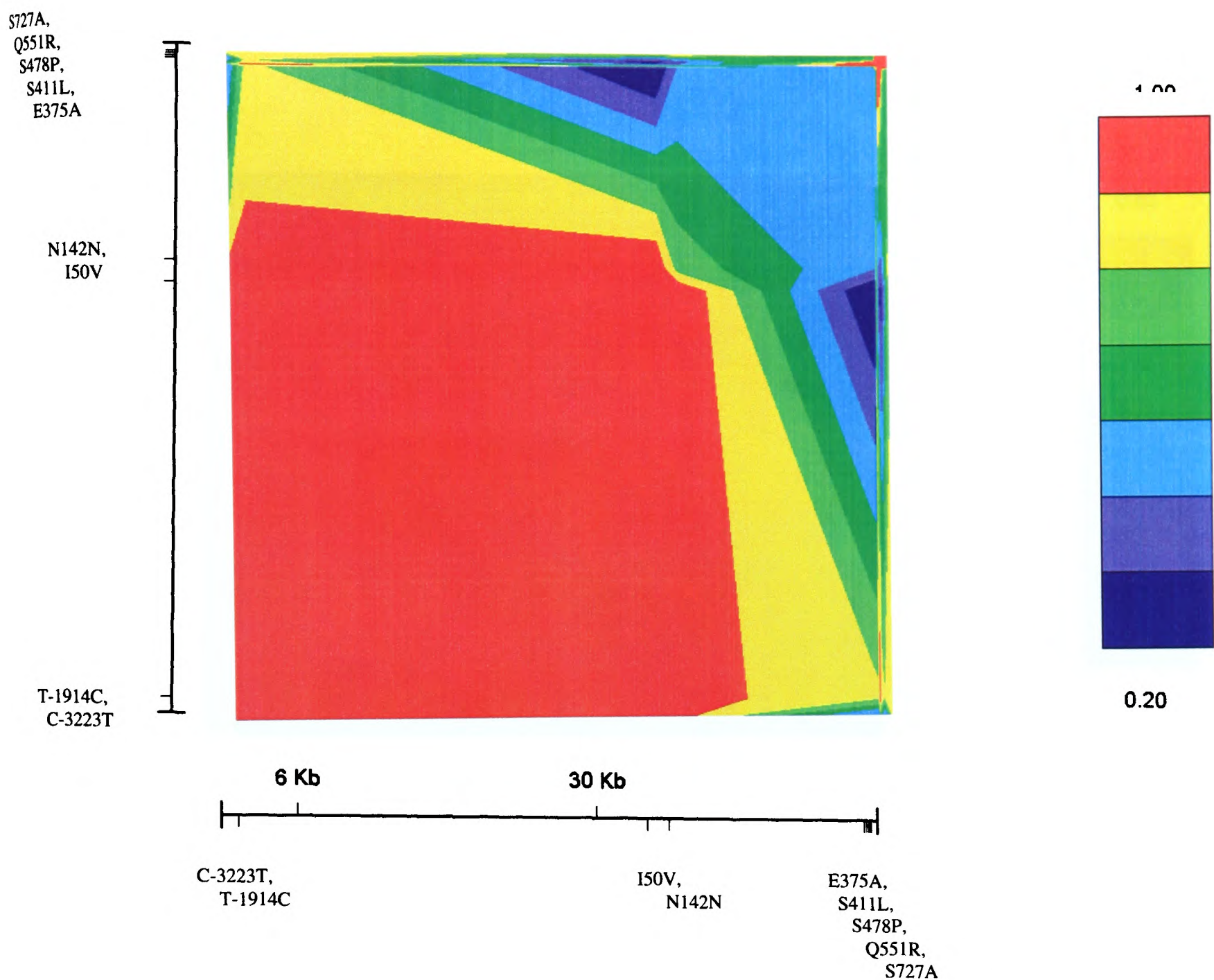


Figure 6.5a: Graphical representation of pairwise LD, measured by D' , across the IL4R gene. Horizontal and vertical axes are scaled according to physical distance between SNP markers. The axis of the gene runs along the diagonal from the bottom left to the top right. The D' values are colour coded with red representing SNPs in linkage disequilibrium.

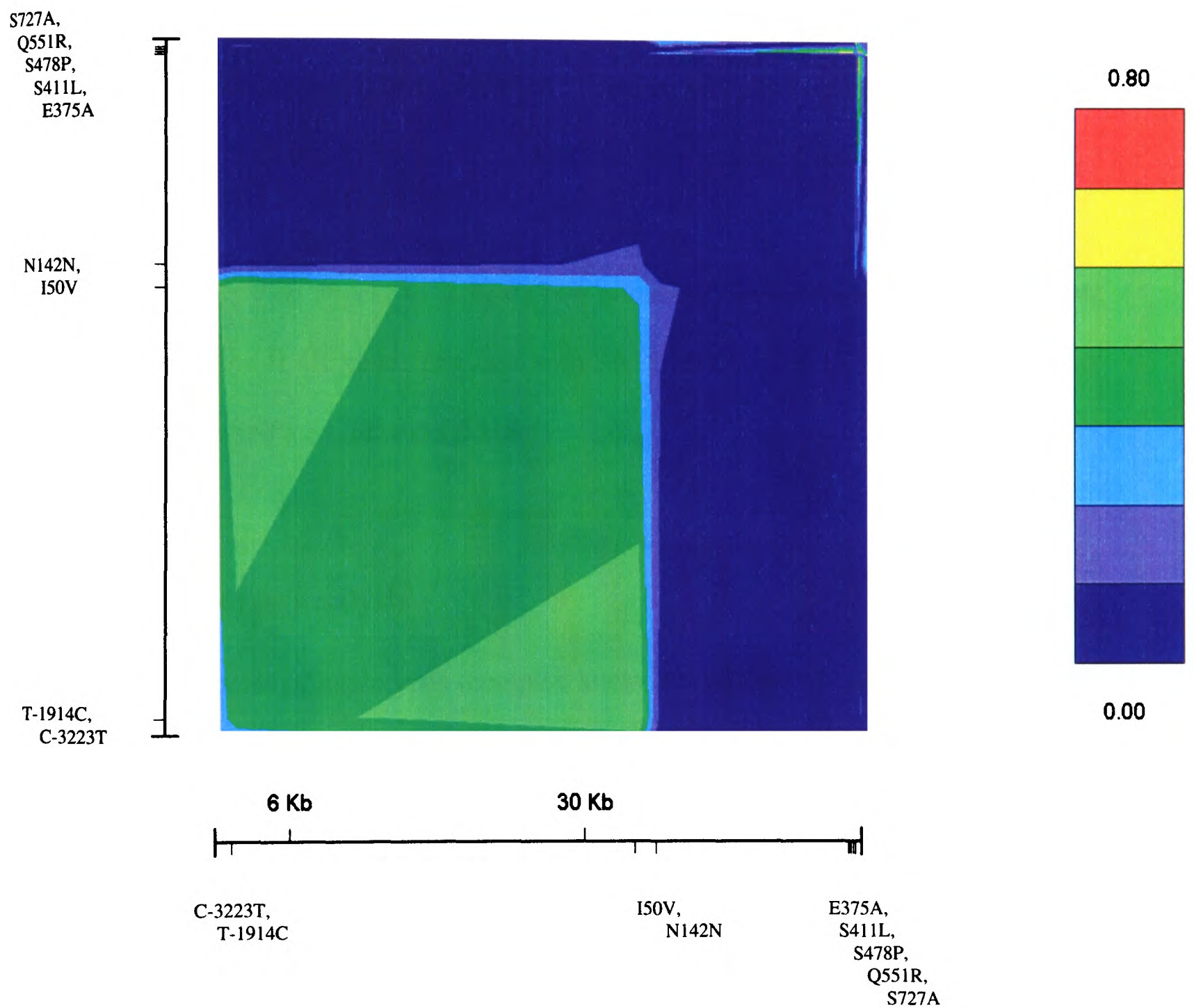


Figure 6.5b: Graphical representation of LD, measured by r^2 , across the IL4R gene. Horizontal and vertical axes are scaled according to physical distance between SNP markers. The axis of the gene runs along the diagonal from the bottom left to the top right. The r^2 values are colour coded with red representing SNPs approaching linkage disequilibrium.

individuals genotypes for these three SNPs reveals that a large fraction of individuals with the minor allele at S411L or S727A also have a minor allele at N142N, although the reverse cannot be said to be true because of the higher frequency of the N142N SNP. There is very little LD between SNPs N142N and S478P or Q551R with r^2 values of 0.08 and 0.06 respectively. This suggests that the observed association for N142N is independent of that at Q551R. There is only weak evidence of LD between S478P or Q551R and SNPs S411L or S727A with r^2 values of between 0.17 and 0.25.

These results then appear to indicate two potentially independent associations with OA within the IL4R gene. The first with SNPs N142N, S411L and S727A and the second with SNPs S478P and Q551R.

6.6 Haplotype analysis

The frequencies of haplotypes (complex alleles) composed of between two and nine SNPs were estimated using the EH-PLUS program (Zhao et al. 2000) and compared between the 456 probands and cases combined and the controls. Haplotypes showing some evidence for association with global p-values ≤ 0.1 are shown in table 6.6a. Haplotypes where the global p-value is less than that for any of the constituent SNPs are highlighted in red. Although there are a number of such haplotypes, for many the decrease in p-value compared to that for the individual SNPs is marginal. The haplotype showing the largest deviation between the global p-value for the haplotype and that of the constituent SNPs is that for N142N plus S478P with a p-value of 0.005 for the haplotype compared with a p-value of 0.03 for N142N. This represents an almost ten-fold decrease in p-value. It should be noted however, that the r^2 value for these two SNPs is very low, at 0.08, meaning that the power of EH-PLUS to reliably

SNPs defining haplotype	Probands and cases combined (n = 456)	Lowest p- value for constituent SNPs
C-3223T, N142N	0.053	
C-3223T, S411L	0.034	
C-3223T, S727A	0.023	
T-1914C, N142N	0.085	
T-1914C, S411L	0.045	
T-1914C, S727A	0.043	
I50V, S411L	0.082	
I50V, S478P	0.016	0.08
I50V, Q551R	0.028	0.06
I50V, S727A	0.054	
N142N, S411L	0.052	
N142N, S478P	0.005	0.03
N142N, Q551R	0.008	0.03
N142N, S727A	0.038	
E375A, S411L	0.069	
E375A, S478P	0.014	0.08
E375A, S727A	0.008	
S411L, S478P	0.013	
S411L, Q551R	0.031	
S411L, S727A	0.021	
S478P, S727A	0.048	
Q551R, S727A	0.046	
C-3223T,I50V, S478P	0.069	0.08
C-3223T, I50V, Q551R	0.088	
C-3223T, N142N, S478P	0.008	0.03
C-3223T, N142N, Q551R	0.013	0.03
C-3223T, E375A, S478P	0.031	0.08
C-3223T, E375A, S727A	0.082	
C-3223T, S478P, S727A	0.080	
C-3223T, Q551R, S727A	0.084	
T-1914C, I50V, Q551R	0.058	
T-1914C, I50V, S727A	0.033	
T-1914C, N142N, S478P	0.028	
T-1914C, N142N, Q551R	0.048	
T-1914C, N142N, S727A	0.042	
T-1914C, S411L, S727A	0.100	
I50V, N142N, S478P	0.035	
I50V, N142N, Q551R	0.044	
I50V, E375A, S411L	0.066	
I50V, E375A, S478P	0.033	0.08
I50V, E375A, Q551R	0.066	
I50V, E375A, S727A	0.004	
I50V, S411L, S478P	0.028	
I50V, S411L, Q551R	0.077	
I50V, Q551R, S727A	0.062	
I50V, S478P, S727A	0.030	
I50V, Q551R, S727A	0.067	
N142N, E375A, S478P	0.096	
N142N, E375A, S727A	0.035	
N142N, S411L, S478P	0.091	
N142N, S411L, S727A	0.071	
N142N, S478P, S727A	0.063	
N142N, Q551R, S727A	0.049	
T-1914C, I50V, E375A, S727A	0.064	

Table 6.6a: Results of haplotype analysis using EH-PLUS to estimate differences in haplotype frequencies between the combined female-THR cohorts and the 399 age-matched female controls. Only haplotypes showing a p-value ≤ 0.1 are shown. P-values less than those for any of the constituent SNPs are highlighted in red and the corresponding lowest p-value for any constituent SNP marked in the final column.

estimate haplotypes is limited. The significance of this result must therefore be treated with caution.

In order to investigate this further, EH-PLUS was used to estimate individual haplotype frequencies in the probands and cases combined and in the controls (table 6.6b). ODDs ratios and p-values were calculated to test for an association between haplotypes and OA. Haplotypes showing significant p-values are highlighted in red. Several haplotypes show evidence for over or under representation in the probands and cases versus the controls. Many of the haplotypes with the lowest p-values consist of the associated alleles at N142N, S478P and Q551R (allele 2 for N142N and S478P and allele 1 for Q551R). Although this result may potentially be interesting it likely reflects at least in part the small number of individuals with the minor alleles (as indicated by the large 95% confidence interval for the ODDs ratio). Generally, the confidence intervals for all of the haplotypes showing significant evidence for association are wide. This indicates that the results should be interpreted with much caution, as the sample size is likely to be too small to accurately estimate the effect of these haplotypes.

6.7 Relationship between genetic variation and the evidence for linkage.

In order to investigate the relationship between genetic variation and the evidence for linkage, each of the 146 female-THR families used in the original additional finer linkage mapping (chapter 3.7.1) were classified according to the genotype of their proband. For each SNP, two groups were defined; those where the proband had at

SNPs forming haplotype						Frequency		ODDs ratio ^{1,2}	P-value ²
C-3223T	I50V	N142N	E375A	S478P	Q551R	Cases	Controls		
	1			1		0.30	0.34	0.81 (0.66 to 0.99)	0.05
	1			2		0.13	0.08	1.73 (1.25 to 2.39)	0.001
	2			1		0.52	0.51	0.97 (0.80 to 1.18)	0.66
	2			2		0.05	0.07	0.71 (0.47 to 1.09)	0.17
	1				1	0.14	0.09	1.68 (1.23 to 2.30)	0.001
	1				2	0.29	0.34	3.99 (3.00 to 5.31)	0.05
	2				1	0.08	0.09	0.83 (0.58 to 1.17)	0.31
	2				2	0.49	0.48	1.04 (0.85 to 1.26)	0.75
		1		1		0.78	0.81	0.83 (0.65 to 1.06)	0.15
		1		2		0.12	0.12	0.98 (0.73 to 1.32)	0.97
		2		1		0.04	0.04	0.88 (0.54 to 1.43)	0.71
		2		2		0.06	0.03	2.61 (1.54 to 4.43)	<0.0005
		1			1	0.16	0.16	1.04 (0.80 to 1.36)	0.79
		1			2	0.74	0.78	0.82 (0.65 to 1.03)	0.09
		2			1	0.06	0.03	2.47 (1.47 to 4.17)	0.0008
		2			2	0.04	0.04	0.90 (0.55 to 1.45)	0.76
			1	1		0.00	0.001	0.91 (0.06 to 13.95)	0.56
			1	2		0.12	0.12	1.00 (0.74 to 1.36)	0.94
			2	1		0.85	0.82	0.78 (0.60 to 1.02)	0.09
			2	2		0.03	0.06	2.11 (1.28 to 3.46)	0.07
1	1			1		0.07	0.07	0.89 (0.61 to 1.30)	0.64
1	1			2		0.10	0.07	1.56 (1.08 to 2.24)	0.02
1	2			1		0.51	0.49	1.09 (0.90 to 1.33)	0.41
1	2			2		0.04	0.07	0.62 (0.40 to 0.96)	0.04
2	1			1		0.23	0.28	0.80 (0.64 to 1.00)	0.05
2	1			2		0.03	0.01	3.12 (1.40 to 6.95)	0.007
2	2			1		0.009	0.01	0.92 (0.34 to 2.46)	0.95
2	2			2		0.002	0.002	0.79 (0.08 to 7.63)	0.89
1		1		1		0.54	0.53	1.05 (0.86 to 1.27)	0.66
1		1		2		0.09	0.11	0.77 (0.56 to 1.07)	0.15
1		2		1		0.04	0.04	0.96 (0.59 to 1.58)	0.99
1		2		2		0.06	0.02	2.61 (1.52 to 4.48)	<0.0005
2		1		1		0.24	0.28	0.82 (0.66 to 1.03)	0.10
2		1		2		0.03	0.01	2.48 (1.17 to 5.26)	0.02
2		2		1		<0.001	0.003	0.33 (0.03 to 3.26)	0.52
2		2		2		<0.001	0.00		
1		1			1	0.12	0.14	0.83 (0.62 to 1.11)	0.24
1		1			2	0.51	0.50	1.04 (0.86 to 1.26)	0.73
1		2			1	0.06	0.03	2.45 (1.44 to 4.17)	0.0009
1		2			2	0.04	0.04	0.98 (0.60 to 1.61)	0.95
2		1			1	0.04	0.02	2.08 (1.14 to 3.79)	0.02
2		1			2	0.23	0.27	0.80 (0.64 to 1.00)	0.06
2		2			1	<0.001	0.00		
2		2			2	<0.001	0.004	0.31 (0.03 to 3.04)	0.53
1			1	1		0.001	0.00	0.85 (0.05 to 13.62)	0.56
1			1	2		0.09	0.105	0.85 (0.61 to 1.18)	0.37
1			2	1		0.59	0.57	1.05 (0.86 to 1.28)	0.66
1			2	2		0.06	0.03	1.81 (1.09 to 3.01)	0.03
2			1	1		0.00	0.00		
2			1	2		0.02	0.01	1.96 (0.89 to 4.30)	0.12
2			2	1		0.24	0.28	0.80 (0.64 to 1.01)	0.07
2			2	2		0.007	<0.001	5.35 (0.65 to 44.24)	0.19
	1		1	1		0.00	0.00		
	1		1	2		0.07	0.05	1.44 (0.95 to 2.19)	0.11
	1		2	1		0.30	0.35	0.80 (0.64 to 0.99)	0.04
	1		2	2		0.06	0.03	2.29 (1.33 to 3.95)	0.004
	2		1	1		0.001	0.00	0.92 (0.06 to 13.94)	0.56
	2		1	2		0.04	0.06	0.67 (0.43 to 1.06)	0.12
	2		2	1		0.52	0.51	1.06 (0.87 to 1.29)	0.63
	2		2	2		0.007	0.006	1.05 (0.30 to 3.68)	0.96

¹ Where the frequency in cases or controls was 0, odds ratios were calculated under the assumption that a single individual carried the haplotype.
²Where frequency in both cases and controls ≤ 0.01 then ODDs ratio and χ^2 calculations were not undertaken.

Table 6.6b : Haplotype frequencies estimated from the genotypes of 456 probands and cases combined and 399 controls. C-3223T, allele 1, C, allele 2, T; I50V, allele 1, G, allele 2, A; N142N, allele 1, C, allele 2, T; E375A, allele 1, A, allele 2, C; S478P, allele 1, T, allele 2, C; Q551R, allele 1, G, allele 2, A. The ODDs ratio and 95% confidence interval of each haplotype relative to the other haplotypes as a group are shown. Differences in haplotype frequencies between probands and cases combined and controls were calculated using a χ^2 test using standard contingency table analysis and P-values are shown in the final column. P-values ≤ 0.05 are highlighted in red.

least one copy of the disease associated allele and those where the proband was homozygous for the non-associated allele. Linkage analysis was then undertaken for the two groups separately. Under the null hypothesis of no association between the SNP in consideration (or one in LD with it) and the disease, it is expected that the evidence for linkage will be randomly distributed amongst the two groups and will therefore be proportional to the size of the group. If, however, the SNP is associated with the disease then it is expected that the group with at least one copy of the disease associated allele will show a disproportionately high level of evidence for linkage (Cox 2001). Figures 6.7a - c show the results of the linkage analysis of the two groups for each of the three SNPs showing some evidence for association in the 146 female-THR probands (N142N, S411L and S727A), overlaid on the original additional finer linkage mapping results.

6.7.1 N142N

Linkage analysis for chromosome 16p was repeated for the twenty-nine families in which the proband had at least one copy of the associated allele at N142N and also for the 115 families where the proband is homozygous for the non-associated allele (figure 6.7a). Two families were not included as the proband failed to amplify in the SNP genotyping. The results show strong evidence for non-random partitioning of linkage between the two groups. The group of families where the proband has at least one copy of the associated allele represents only approximately 20% of the families and yet has a LOD score of 1.2 compared to the LOD score of 0.9 for the other 115 families. This observation suggests that this allele, or one in tight LD with it, is associated with the evidence for linkage at chromosome 16 therefore increasing confidence in a role for the IL4R gene in OA.

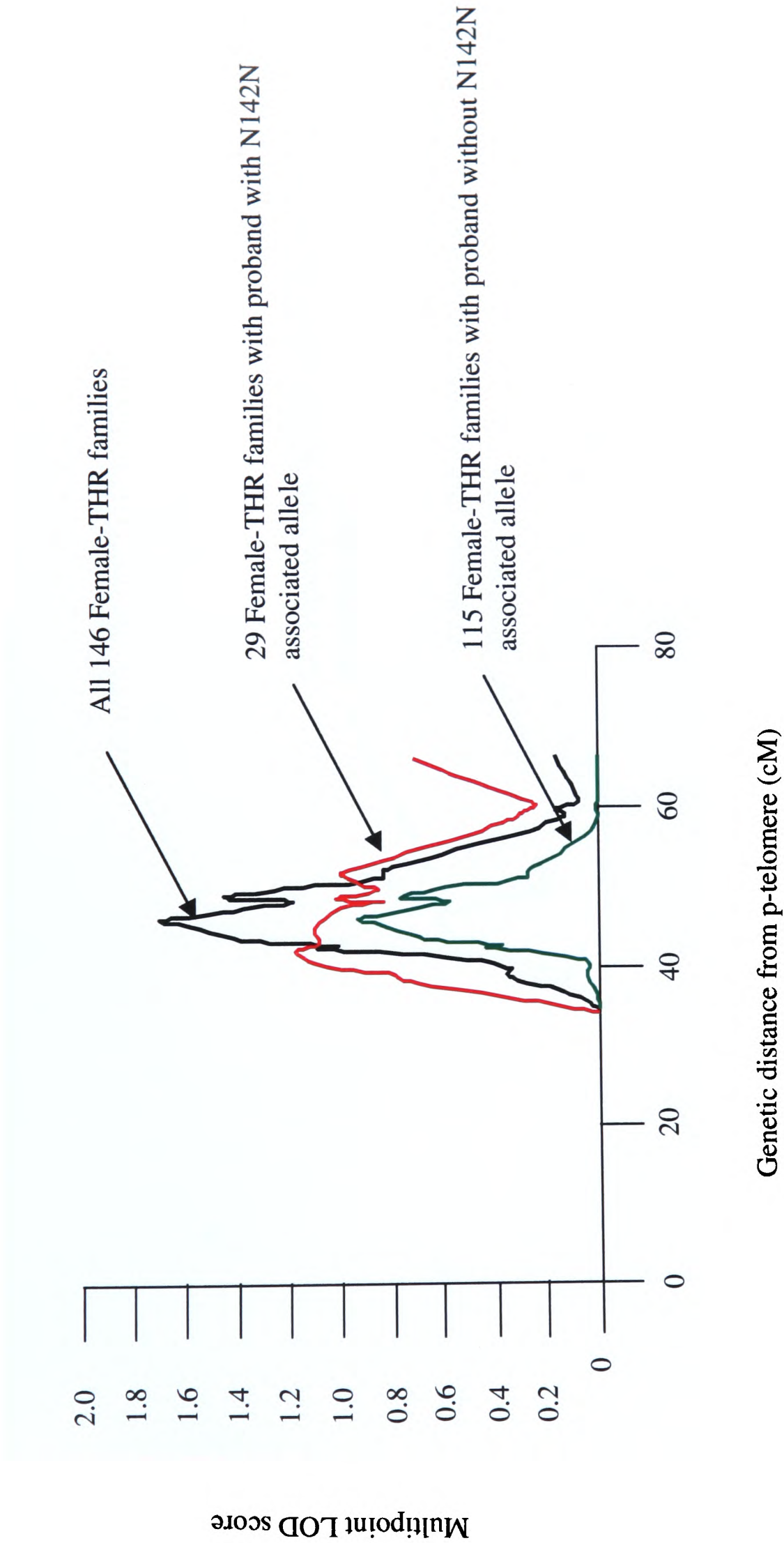


Figure 6.7a: Multipoint linkage analysis of chromosome 16p, for the 146 female-THR families, overlaid with results of linkage analysis for both the 29 families whose proband has a copy of the associated allele of SNP N142N and for the 115 families where the proband is homozygous for the non-associated allele of this SNP.

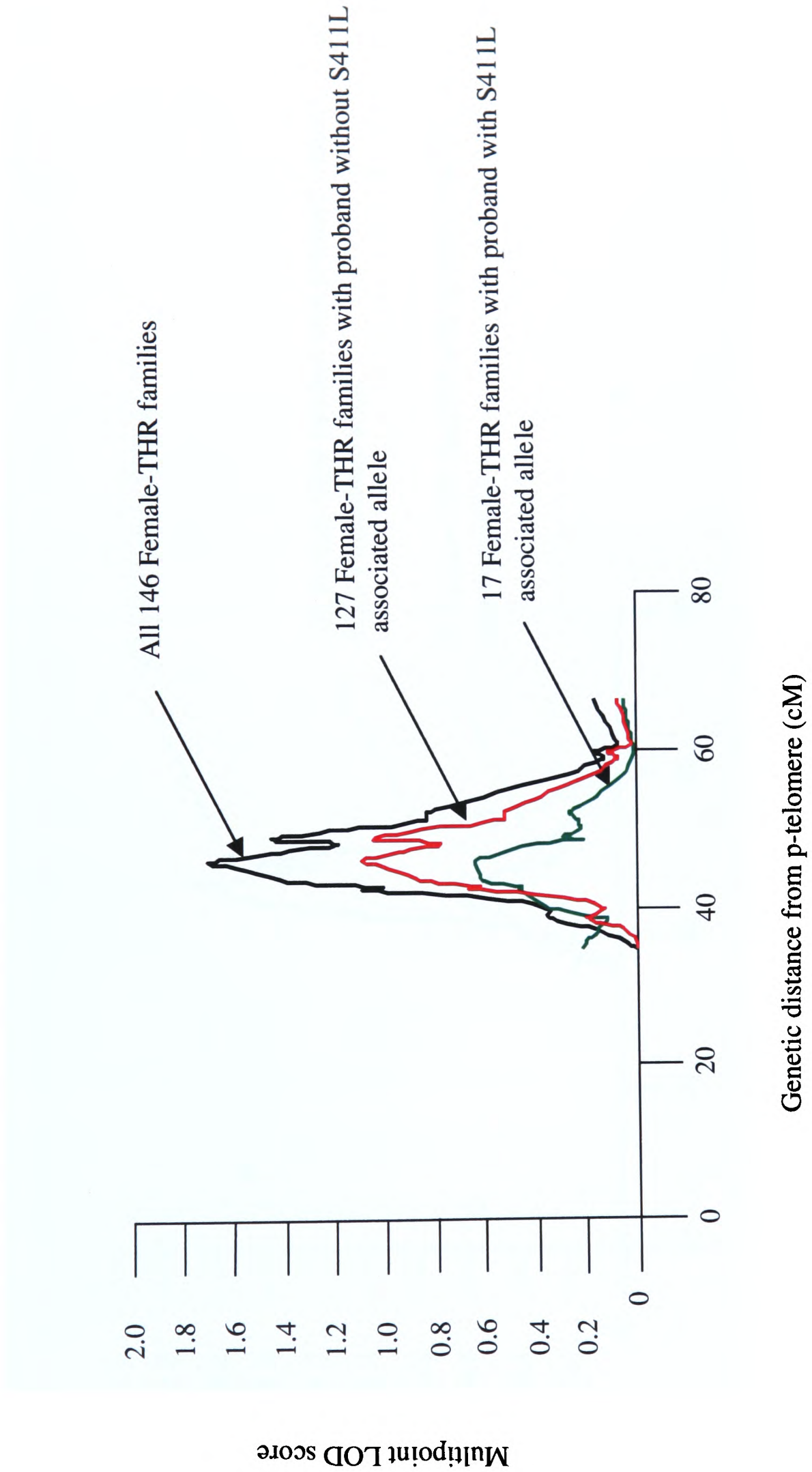


Figure 6.7b: Multipoint linkage analysis of chromosome 16p, for the 146 female-THR families, overlaid with results of linkage analysis for both the 17 families whose proband has a copy of the associated allele of SNP S411L and for the 127 families where the proband is homozygous for the non-associated allele of this SNP.

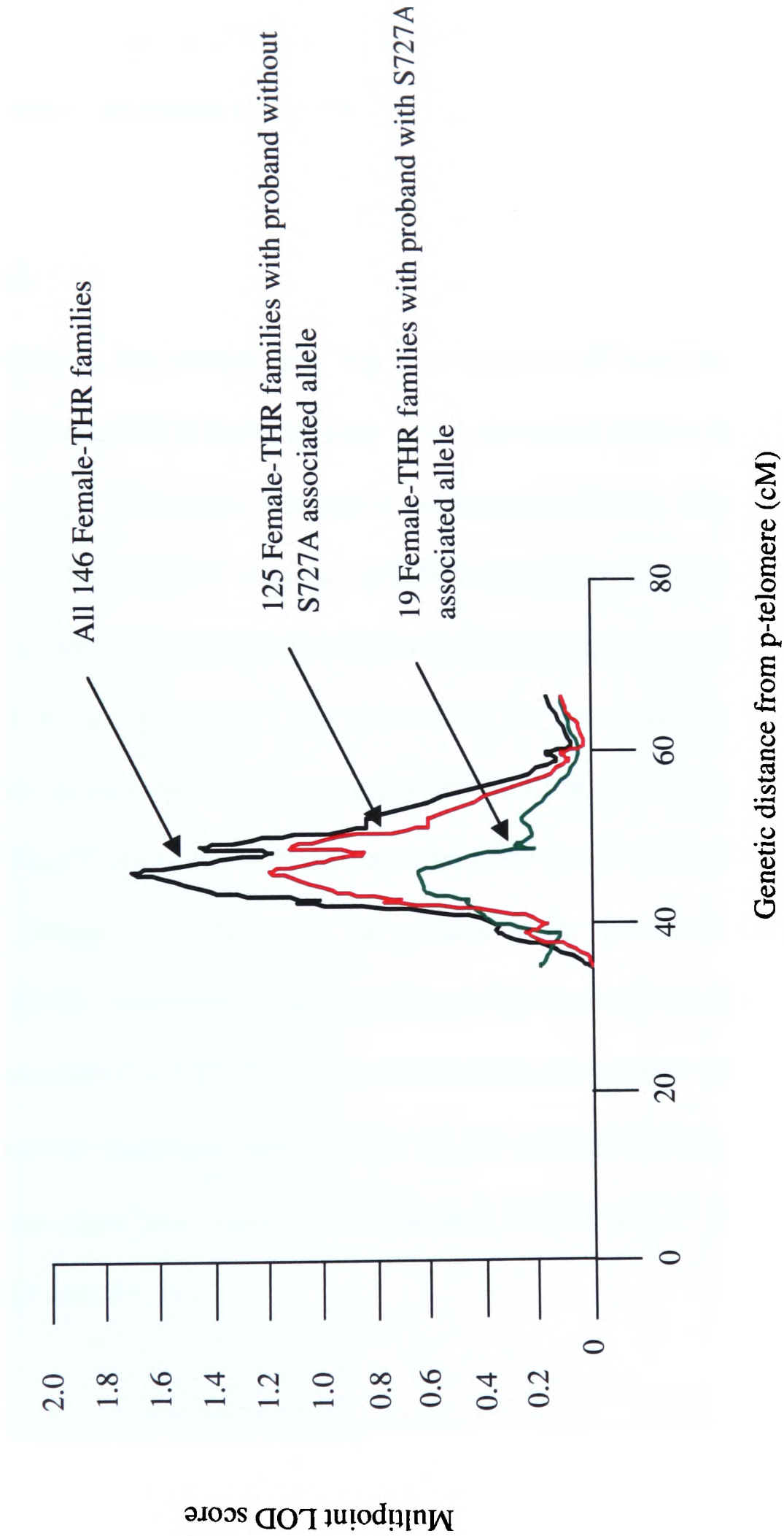


Figure 6.7c: Multipoint linkage analysis of chromosome 16p for the 146 female-THR families, overlaid with results of linkage analysis for both the 19 families whose proband has a copy of the associated allele of SNP S727A and for the 125 families where the proband is homozygous for the non-associated allele of this SNP.

SNP N142N however, does not appear to account for all the linkage to chromosome 16p since the remnant of families without the N142N associated allele still show some evidence for linkage. This may indicate the presence of another variant(s) within the region also contributing to susceptibility.

6.7.2 S411L

Linkage analysis for chromosome 16p was repeated for the seventeen families in which the proband had at least one copy of the associated allele at S411L and also for the 127 families where the proband is homozygous for the non-associated allele (figure 6.7b). Two families were not included as the proband failed to amplify in the SNP genotyping. As for N142N the effects on linkage analysis are disproportionate to the size of the groups with a LOD score of 0.7 for the seventeen families with the minor allele in the probands compared with a LOD score of 1.1 for the other 127 families. This is compared with the expected LOD scores of approximately 0.2 and 1.5 if the evidence for linkage was proportional to the size of the two groups. This result for S411L is not totally unexpected due to the relatively strong LD between this SNP and N142N ($r^2 = 0.33$, $D' = 0.83$). Indeed there are fourteen families in which the proband has the associated alleles at both N142N and S411L. This result implies that part, but not all, of the observed association at N142N may be accounted for by it being in LD with S411L.

6.7.3 S727A

Linkage analysis for chromosome 16p was repeated for the nineteen families in which the proband had at least one copy of the associated allele at S727A and also for the 127 families where the proband is homozygous for the non-associated allele (figure 6.7c). Two families were not included as the proband failed to amplify in the SNP genotyping. The results for this analysis are very similar to those of S411L with LOD scores of 0.7 and 1.2 for the groups with and without a proband with the associated allele respectively. This is compared with the expected LOD scores of approximately 0.2 and 1.5 expected if the evidence for linkage were proportional to the size of the two groups. This disproportionate partitioning of linkage is not unexpected due to the strong LD between this SNP and S411L ($r^2 = 0.80$, $D' = 0.97$). In fact all seventeen of the families with the associated allele at S411L also have an associated allele at S727A.

These results appear to further suggest that the association at both S411L and S727A may be dependent on their association with N142N or reciprocally, that part of the association observed for SNP N142N may be dependent on its LD with these two SNPs. It appears however, that the evidence for association at N142N cannot be entirely accounted for by LD with either S411L or S727A.

6.8 Examination of dosage effects

Many investigators have reported an increased risk of complex disease associated with two variant alleles at a susceptibility gene (e.g. Corder et al. 1993; Hugot et al. 2001). This may correspond to either two copies of a single variant allele (i.e.

homozygosity), or to a combination of variant alleles at different sites (i.e. compound heterozygosity). Since there are also reports of potential interaction between alleles within the IL4R gene (Kruse et al. 1999; Mitsuyasu et al. 1999) the effects of dosage of the associated IL4R SNPs were examined.

6.8.1 Association analysis of SNP genotypes

Genotype frequencies between all 456 female-THR probands and cases and the 399 controls were compared by χ^2 analysis using standard 2 x 3 contingency tables for all nine SNPs (table 6.8a). The p-values for all SNPs are either equal to or less significant than those for comparison of the allele frequencies. This indicates that there is no increased risk associated with any particular genotype compared to having an associated allele for any of the individual SNPs.

6.8.2 Association analysis for SNP genotype combinations

In order to examine if combinations of variant alleles at different sites are important in determining susceptibility to OA the numbers and frequencies of variant alleles at pairs of SNP loci were examined. Since at each SNP loci an individual can have 0, 1 or 2 variant alleles, when two loci are considered an individual can have between 0 and 4 variant alleles. For each possible combination of SNPs N142N, S411L, S478P, Q551R and S727A the number and frequency of individuals with 0, 1, 2, 3 or 4 variant alleles was determined both for all 456 female-THR probands and cases and for the 399 female controls. Results are shown in table 6.8b. In order to make a preliminary comparison of the distribution of the number of variant alleles between cases and controls, the cohorts were divided in to two groups. The first group

SNP	Female Controls			Probands and cases combined (n = 456)			P-value
	(++)	(+-)	(--)	(++)	(+-)	(--)	
	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	
C-3223T	192 (0.49)	163 (0.42)	34 (0.09)	236 (0.52)	181 (0.40)	33 (0.07)	0.59
T-1914C	128 (0.33)	185 (0.47)	77 (0.20)	143 (0.32)	217 (0.48)	91 (0.20)	0.94
I50V	70 (0.18)	187 (0.49)	127 (0.33)	82 (0.18)	219 (0.49)	144 (0.32)	0.98
N142N	343 (0.87)	46 (0.12)	4 (0.01)	363 (0.81)	80 (0.18)	5 (0.01)	0.04
E375A	5 (0.01)	77 (0.20)	296 (0.78)	3 (0.01)	98 (0.22)	338 (0.77)	0.53
S411L	356 (0.94)	21 (0.06)	0 (0)	400 (0.89)	48 (0.11)	0 (0)	0.01 ¹
S478P	285 (0.73)	93 (0.24)	11 (0.03)	304 (0.67)	134 (0.30)	15 (0.03)	0.15
Q551R	14 (0.04)	114 (0.29)	264 (0.67)	24 (0.05)	150 (0.33)	276 (0.61)	0.15
S727A	351 (0.94)	24 (0.06)	0 (0)	373 (0.87)	52 (0.12)	2 (0.01)	0.005 ²

¹ Analysed with 2 x 2 contingency table, ignoring empty cells

² Analysed with 2 x 2 contingency table, clumping (++) and (--) cells together

Table 6.8a: Genotype numbers and frequencies for the nine IL4R gene SNPs for the cohort of all 456 female-THR probands and cases and the cohort of 399 controls. The final column shows the p-value for comparison of cases and controls by χ^2 analysis.

SNP Combination	Probands and cases combined (n = 456)					Female Controls				
	0	1	2	3	4	0	1	2	3	4
	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)	N° (Freq)
N142N & S411L	351(0.80)	49(0.11)	36(0.08)	5(0.01)	0(0)	325(0.87)	29(0.08)	18(0.05)	2(<0.01)	0(0)
N142N & S478P	270(0.61)	114(0.26)	47(0.11)	12(0.03)	2(<0.01)	257(0.66)	101(0.26)	21(0.05)	8(0.02)	0(0)
N142N & Q551R	243(0.55)	132(0.30)	50(0.11)	15(0.03)	2(<0.01)	237(0.61)	119(0.31)	26(0.07)	8(0.02)	0(0)
N142N & S727A	332(0.79)	45(0.11)	40(0.09)	3(0.01)	2(<0.01)	318(0.86)	32(0.09)	19(0.05)	2(<0.01)	0(0)
S411L & S478P	297(0.67)	97(0.22)	40(0.09)	11(0.02)	0(0)	271(0.73)	75(0.20)	21(0.06)	5(0.01)	0(0)
S411L & Q551R	269(0.61)	115(0.26)	45(0.10)	13(0.03)	0(0)	248(0.66)	95(0.25)	25(0.07)	5(0.01)	0(0)
S411L & S727A	369(0.87)	10(0.02)	44(0.10)	2(<0.01)	0(0)	336(0.94)	3(0.01)	20(0.06)	0(0)	0(0)
S478P & Q551R	276(0.62)	22(0.05)	126(0.28)	8(0.02)	15(0.03)	263(0.68)	21(0.05)	93(0.24)	1(<0.01)	11(0.03)
S478P & S727A	283(0.67)	87(0.21)	42(0.10)	10(0.02)	2(<0.01)	267(0.72)	74(0.20)	22(0.06)	6(0.02)	0(0)
Q551R & S727A	255(0.61)	105(0.25)	48(0.11)	11(0.03)	2(<0.01)	244(0.66)	94(0.25)	26(0.07)	6(0.02)	0(0)

Table 6.8b: Numbers and frequencies of individuals with zero, one, two, three and four variant alleles for different SNP pair combinations in the cohort of all 456 female-THR probands and cases and the cohort of 399 controls.

contained those individuals who have a total of 0 or 1 variant alleles and the second group contained those individuals who have a total of 2 or more variant alleles. The numbers in each group were compared between the cases and probands combined and the control cohorts using χ^2 analysis with standard 2 x 2 contingency tables. Results are shown in table 6.8c. Two SNP combinations have a p-value which is more significant than that observed for either of the constituent SNPs. The combination of N142N & S478P has a p-value of 0.006, compared with individual SNP p-values of 0.03 (N142N) and 0.08 (S478P) (table 6.4f) while that of N142N & Q551R has a p-value of 0.007, compared with individual SNP p-values of 0.03 (N142N) and 0.06 (Q551R) [table 6.4f]. Observation of the frequencies shows that for both SNP combinations there is an elevation in the cases and probands combined compared with the controls in the frequency of individuals with two or more variant alleles.

In order to investigate further these two SNP combinations, the cases and probands combined and control cohorts were further subdivided according to both the total number of variant alleles present and the composition of these variant alleles (tables 6.8d and e). For both SNP combinations, there is clearly an elevated number of individuals in the cases and probands combined versus the controls who have at least one variant allele at each of the SNPs (those groups are indicated with a ¹ in tables 6.8d and e). In order to formally compare the cases and probands combined and the controls, χ^2 analysis using standard 2 x 2 contingency tables (with the number of individuals with at least one variant allele at both SNPs versus the number of individuals without at least one variant allele at each SNP) was undertaken. P-values are given in tables 6.8d and e. The p-value for at least one variant allele at both SNP N142N and SNP S478P has a value of 0.0008 compared to the p-value of 0.006

SNP Combination	Probands and cases combined (n = 456)				Female Controls				P-value
	<2 variant alleles		≥2 variant alleles		<2 variant alleles		≥2 variant alleles		
	N°	Freq	N°	Freq	N°	Freq	N°	Freq	
N142N & S411L	400	0.91	41	0.09	354	0.95	20	0.05	0.047
N142N & S478P	384	0.86	61	0.14	358	0.93	29	0.07	0.006
N142N & Q551R	375	0.85	67	0.15	356	0.91	34	0.09	0.007
N142N & S727A	377	0.89	45	0.11	350	0.94	21	0.06	0.017
S411L & S478P	394	0.89	51	0.11	346	0.93	26	0.07	0.042
S411L & Q551R	384	0.87	58	0.13	343	0.92	30	0.08	0.028
S411L & S727A	379	0.89	46	0.11	339	0.94	20	0.06	0.013
S478P & Q551R	298	0.67	149	0.33	284	0.73	105	0.27	0.058
S478P & S727A	370	0.87	54	0.13	341	0.92	28	0.08	0.024
Q551R & S727A	360	0.86	61	0.14	338	0.91	32	0.09	0.017

Table 6.8c: Number and frequency of individuals with less than two variant alleles and two or more variant alleles in the cohort of all 456 female-THR probands and cases and the cohort of 399 controls. The final column shows the p-value for χ^2 analysis using standard 2 x 2 contingency tables.

Number of variant alleles	Composition	Probands and cases combined (n = 456)	Controls
0		270	257
1	N142N	28	26
1	S478P	86	75
2	2xN142N	0	2
2	2xS478P	4	5
2	N142N, S478P ¹	43	14
3	2xN142N, S478P ¹	3	2
3	2xS478P, N142N ¹	9	6
4	2xN142N, 2xS478P ¹	2	0
≥ 1 copy of each variant allele		57	22
< 1 copy of each variant allele		388	365
P-value		0.0008	

¹ Rows added together to calculate numbers of individuals with at least one copy of each associated allele.

Table 6.8d: Numbers of individuals with different combinations of variant alleles of SNPs N142N and S478P. The p-value for comparison of individuals with ≥ 1 copy of each variant allele and individuals with < 1 copy of each variant allele in probands and cases combined versus controls is shown in the bottom row.

Number of variant alleles	Composition	Probands and cases combined (n = 456)	Controls
0		243	237
1	N142N	27	25
1	Q551R	105	94
2	2xN142N	0	2
2	2xQ551R	10	8
2	N142N, Q551R ¹	40	16
3	2xN142N, Q551R ¹	3	2
3	2xQ551R, N142N ¹	12	6
4	2xN142N, 2xQ551R ¹	2	0
≥ 1 copy of each variant allele		57	24
< 1 copy of each variant allele		385	366
P-value		0.002	

¹ Rows added together to calculate numbers of individuals with at least one copy of both variant alleles.

Table 6.8e: Numbers of individuals with different combinations of variant alleles of SNPs N142N and Q551R. The p-value for comparison of individuals with ≥ 1 copy of each variant allele and individuals with < 1 copy of each variant allele in probands and cases combined versus controls is shown in the bottom row.

observed when any combination of two or more variant alleles was considered. This almost ten-fold decrease in p-value when individuals with at least one variant allele at each SNP is considered would appear to indicate that possessing at least one copy of each variant allele (compound heterozygosity) increases OA susceptibility risk. For the combination of SNPs N142N and Q551R the p-value, when individuals with at least one variant allele at both SNPs are considered, of 0.002 also represents a decrease when compared with that for any combination of two or more variant alleles (p-value = 0.007). It is likely that the result for N142N and Q551R is not independent of that for N142N and S478P due to the high level of LD between SNPs S478P and Q551R ($r^2 = 0.79$).

6.9 Splicing analysis

At least 60% of human genes express two or more transcripts generated by alternative splicing (Modrek and Lee 2002) resulting in functionally diverse protein isoforms which are often tissue and developmental stage specific (Grabowski 1998; Smith and Valcarcel 2000). A new group of auxiliary, cis-acting elements which influence splicing have recently been identified. These elements aid in the recognition of exons and are especially relevant for regulating alternative splicing (Smith and Valcarcel 2000). A subgroup of these cis-acting elements, the exonic splice enhancers (ESEs) are predicted to be in the majority of exons (Schaal and Maniatis 1999) and are likely to function by acting as binding sites for specific serine/arginine (SR) proteins which promote exon definition (Robberson et al. 1990). It is predicted that a number of exonic point mutations previously classified as silent may disrupt the sequence of an ESE resulting in aberrant splicing with a subsequent effect on protein function.

Since the N142N SNP is an apparently silent polymorphism, it was postulated that its observed association with OA may be due to an effect of this variation on splicing of the IL4R gene. In order to test this, cDNA was synthesised from total cartilage RNA collected from the removed joint in individuals who had undergone total hip or total knee joint replacement for primary OA. This cDNA was amplified, using primers in exons 6 and 8 of the IL4R gene (the exons flanking exon 7 which contains the N142N SNP) in eleven individuals heterozygous for the N142N SNP (there were no individuals homozygous for the associated allele) and in two individuals homozygous for the non-associated allele to act as controls (figure 6.9). All individuals show a single band at 219 bp, the size expected for correct splicing of exon 7. There is no difference between the individuals homozygous or heterozygous for the N142N SNP. If SNP N142N did have an effect on IL4R splicing in chondrocytes then it would be expected that heterozygous individuals would have an additional band of a different size. This additional band could be larger, if some of the intronic flanking sequence was spliced in, or smaller, if some or all of the exon was excluded. No such band was seen.

6.10 Discussion

The results reported in this chapter provide substantial evidence for the involvement of the IL4R gene in encoding susceptibility for OA. The IL4R gene was identified as a strong candidate for involvement in the OA disease process, as a result of its position both within the linkage region on chromosome 16p identified by work reported here (chapter 3.7) and the region identified by Ingvarsson et al (2001), and as a result of its observed chondroprotective effect (Cawston et al. 1996; Nemoto et al. 1997;

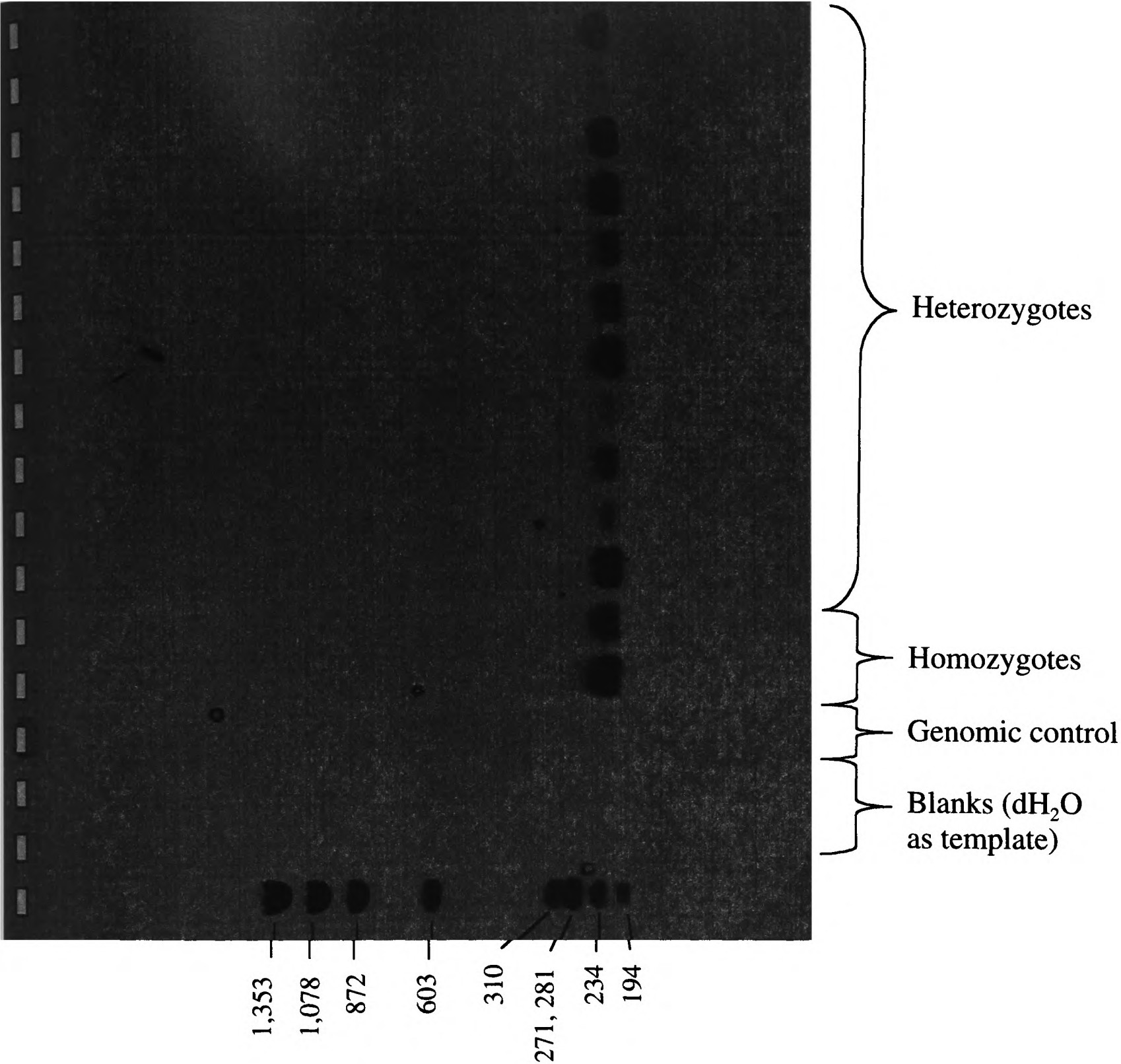


Figure 6.9: Results of amplification across exon 7 of the IL4R gene for OA joint cartilage cDNA from 11 individuals heterozygous and 2 individuals homozygous for the non-associated allele of the N142N SNP.

Alaaeddine et al. 1999; Lubberts et al. 1999) and potential involvement in normal chondrocyte mechanotransduction (Millward-Sadler et al. 1999; Salter et al. 2001).

Sequencing of the IL4R gene both confirmed many previous published reports as to the high level of variation within the gene (Hackstein et al. 1999; Wu et al. 1999), with a total of twelve common, exonic SNPs, and allowed for determination of allele frequencies in our population. All common, non-synonymous variants detected had been previously reported.

Preliminary investigations of nine SNPs within the IL4R gene, in the probands of the 146 female-THR families used in the additional finer linkage analysis of chromosome 16p and in 399 age-matched female controls, found significant evidence for association between SNPs S411L ($P = 0.03$) and S727A ($P = 0.02$) and OA disease with a further SNP, N142N, approaching significance ($P = 0.11$). Subsequent analysis with a separate cohort of 310 female-THR cases provided further evidence for association for SNPs N142N, S411L and S727A (with p-values of 0.04, 0.04 and 0.07 respectively). A fourth SNP, Q551R also showed significant association in this cohort ($P = 0.04$) with a fifth, S478P, having a p-value approaching significance ($P = 0.07$). Combination of information from all 456 female-THR patients provided significant and increased evidence for association at SNPs N142N, S411L and S727A compared to either cohort individually (with p-values of 0.03, 0.01 and 0.004 respectively). SNPs S478P ($P = 0.08$) and Q551R ($P = 0.08$) show a trend towards association in the combined dataset. For all five SNPs the minor allele is elevated in the patients compared with the controls.

Linkage disequilibrium studies revealed a relatively high level of LD between SNPs N142N, S411L and S727A and between SNPs S478P and Q551R with minimal LD between these two groups. This suggests the possibility of two independent associations of the IL4R gene with disease.

Haplotype analysis added little further information to that gained for individual SNPs. This may be partly attributed to the limitation of haplotype estimation programmes to accurately estimate haplotype frequencies when SNPs are in near linkage equilibrium. No haplotypes comprised of alleles in LD with a r^2 value ≥ 0.2 showed an increase in association when compared with the evidence for association of their constituent SNPs.

When the relationship between genetic variation and the evidence for linkage was examined, evidence for non-random partitioning of the evidence for linkage with the associated allele was seen for SNPs N142N, S411L and S727A, with a much higher level of linkage in those families whose proband has a copy of the associated allele than would be expected for a group of that size. This increases confidence for the involvement of one or more of these SNPs in encoding for OA susceptibility. SNPs S411L and S727A are in tight LD and both show a reasonable level of LD with N142N. The results of this partitioning analysis appear to indicate that the association at both S411L and S727A may be dependent on their LD with N142N, or conversely, that part of the association observed for SNP N142N may be dependent on its LD with these two SNPs. The evidence for association at N142N does not however, appear to be entirely accounted for by LD with either S411L, S727A or both. In addition, the residual evidence for linkage in those families without the associated

alleles at any of these three SNPs appears to suggest that a further variant(s) within this region, either within or without of the IL4R gene, is contributing to OA susceptibility.

Some studies have demonstrated an increased risk or severity of complex diseases associated with two variant alleles. This suggests a role for a gene dosage effect on disease susceptibility (e.g. Corder et al. 1993; Horikawa et al. 2000; Hugot et al. 2001). This dosage effect may be due either to two copies of a single variant (i.e. homozygosity) or be the result of a combination of variants at different sites (i.e. compound heterozygosity). There are also reports that interactions between variants within the IL4R gene can affect protein function (Kruse et al. 1999a; Mitsuyasu et al. 1999). Therefore, the effects of dosage of the SNPs in the IL4R gene which show evidence of association with OA were examined. No increased evidence of association (compared with the evidence for association of individual alleles) was found when frequencies of individual SNP genotypes were compared between cases and probands combined and controls. This indicates that two copies of the same variant (homozygosity) are not associated with an increased risk of disease. When combinations of two SNPs were compared, highly significant evidence of association ($P = 0.0008$) was found for the group of individuals, who carry at least one copy of the variant allele at both of SNPs N142N and S478P (compound heterozygosity). A significant association ($P = 0.002$) was also observed for the group of cases with at least one copy of the variant allele at both of SNP N142N and Q551R. It is probable however, that this result is not independent of that for N142N and S478P due to the high level of LD between SNPs Q551R and S478P. These results suggest that compound heterozygosity at SNPs N142N and S478P is important in increasing

susceptibility to OA. The requirement of at least one copy of a variant allele at two SNP loci (compound heterozygosity) may account for the non-significant association results for N142N, S478P and Q551R in the smaller, 146 female-THR proband cohort. It is not possible from the present study to determine whether it is important for both variant alleles to reside on the same protein molecule, or if the effect is greatest when both copies of the gene carry at least one variant allele.

Neither SNP S411L nor SNP S727A showed increased association with disease in combination with either SNP S478P or SNP Q551R. This result, in combination with the studies of the relationship between variation and evidence for linkage, appear to indicate that while LD of N142N with S411L and S727A may account for some of the observed association at N142N it does not account for all of it.

The observation that several variants within the IL4R gene are implicated in OA susceptibility appears to be in line with other complex traits where multiple variants within a gene, some rare and some common, have been identified as encoding for susceptibility to disease. For example, in Crohns disease at least three common variants and several rare variants in the CARD15 gene have been implicated as encoding for susceptibility (Lesage et al. 2002).

SNPs S411L, S478P, Q551R and S727A are all non-synonymous and may therefore be envisaged as having a direct role on protein function. Investigations in to the functional effects for S478P and Q551R have been reported. The S478P polymorphism, which results in an amino acid substitution in the intracellular portion of the receptor, alters the polarity of the IL4R protein and may consequently affect its

conformation. This polymorphism has been associated with alterations in the binding of the STAT6 transcription factor (Kruse et al. 1999a). The Q551R polymorphism results in a glutamine for an arginine substitution within the Src homology 2 (SH2) domain and has been associated with decreased binding of the negative regulator SHP-1 (a phosphotyrosine phosphatase involved in signal termination) (Kruse et al. 1999a). It is not currently clear if this variant affects protein function with two investigations (Mitsuyashu et al. 1999; Wang et al. 1999) failing to replicate the results of an earlier study by Hershey et al (1997) who demonstrated an increased CD23 induction in response to IL4 stimulation in individuals with the variant form of the IL4R protein. It has been suggested, that the action of SNP Q551R may be context dependent, having functional effects only in the presence of other polymorphisms, or that the effects observed by Hershey et al (1997) may be the result of linkage disequilibrium of the Q551R polymorphism with another SNP (Mitsuyashu et al. 1999; Wang et al. 1999). It may be possible that the observed results for Q551R in the study by Hershey et al (1997) is due to the action of SNP S478P with which Q551R is in strong LD. It is also possible, in light of the gene dosage studies reported above, that the effect of Q551R on protein function may be dependent on the co-presence of a variant allele at N142N, or another SNP in tight LD with it.

Functional studies of the effects of S411L and S727A have not so far been reported. S411L resides on the boundary between the region involved in interaction with JAK-1 and the I4R motif. The variant allele results in the substitution of an aliphatic, hydrophobic residue for a polar, hydrophilic one, which could potentially have a profound effect on protein function. The S727A polymorphism, which also results in the substitution of an aliphatic, hydrophobic residue for a polar, hydrophilic one,

resides within the terminal cytoplasmic domain of the protein, in the region involved in negative regulation of IL4 signalling (Kashiwada et al. 2001; Hanson et al. 2003). Again, such a substitution can be envisaged as having a substantial effect on the protein.

Since N142N does not result in a change in amino acid in the IL4R protein, it is difficult to immediately envisage its role in encoding susceptibility to OA. Although it is possible that some of the evidence for association seen for SNP N142N is accounted for by its LD with the potentially functional, non-synonymous SNPs S411L and S727A it is clear from the study of the relationship between genetic variation and the evidence for linkage, and from the study of dosage effects that these variants cannot account for all the observed association at N142N. No other common SNPs that could be classified as functional were found in LD with this SNP and thus it remains that either this SNP, or another as yet unidentified SNP(s) in LD with N142N, are likely responsible for at least a degree of the association observed for this gene.

Traditionally, DNA sequence variation has been classified as functional only if it results in either the substitution of an amino acid in the protein or obviously disrupts a splice junction or recognised promoter sequence. Many, more recent, studies however, challenge this classification as too narrow, with increasing evidence for mutation outside these regions resulting in disease phenotype (Mendell and Dietz 2001; Knight 2003). Increasing understanding of the mechanisms involved in both gene expression and mRNA processing reveal the importance of an increasing number of DNA regions, some potentially distal to the gene of interest itself, in controlling correct protein structure and function.

Control of splicing is increasingly emerging as a highly complex process. In addition to the involvement of consensus sequences at the splice junctions and branch sites, various other auxiliary cis-acting exonic and intronic sequences have been identified as having an influence on splicing. One set of these, termed exonic splice enhancers (ESEs), act as binding sites for serine/arginine rich (SR) proteins and aid in the recognition of exons (Schaal and Maniatis 1999). It is predicted that such elements may be in the majority of exons and that a considerable number of mutations within exons may disrupt these elements resulting in aberrant splicing patterns (Cartegni et al. 2002). A synonymous mutation in exon 7 of the SMN1 gene has been identified as being causal for spinal muscular atrophy as a result of exon skipping (Lorson et al. 1999). Since consensus sequences for such elements are currently poorly understood it is not possible to accurately predict which sequence changes will affect splicing. It was therefore postulated that the N142N SNP could interrupt an exonic splice enhancer or silencer, or similar sequence, affecting IL4R protein structure/function at the level of splicing.

Investigation of cDNA synthesised from the cartilage RNA of joints replaced for OA, found no evidence for an alternative splice product in individuals heterozygous for the N142N SNP compared with individuals homozygous for the non-associated allele. However, since numerous variants that affect splicing result in the introduction of an in-frame premature termination codon (PTC) leading to the degradation of the mRNA by nonsense-mediated decay (NMD), it is possible that only the normal copy of the mRNA would be detected in cartilage samples. Further work in to the relative levels

of the two mRNA transcripts in heterozygous individuals may yield further insight in to the potential effect of N142N on splicing.

If it emerges that this SNP does not itself have a functional effect it is possible that it is acting as a marker for the genuine disease encoding SNP(s). The pattern of LD across the genome is emerging as complex, dependent on a number of factors including recombination rates, population history and allele frequencies (Phillips et al. 2003). In this way a SNP in LD with N142N, for which N142N is acting as a marker, may be some distance from it. In addition, the SNPs showing evidence for association in this study do not appear to account for all the linkage observed on chromosome 16p and it is possible that additional variant(s) within the IL4R, not in LD with the investigated SNPs, are also involved. Since all common, non-synonymous variants within the IL4R have been genotyped in this study, any additional variants contributing to OA susceptibility within this gene must either represent rare non-synonymous variants (a number of such variants were identified by sequencing but were not investigated further due to their low frequency) or be due to synonymous or non-coding variation(s) which affects splicing or mRNA expression. Some investigators have proposed that such variants, which do not result in an amino acid substitution, may be expected to be involved in susceptibility to complex traits, as their effects will often be less severe than those seen for amino acid changes, giving rise to a more subtle, environmentally and genetically modifiable effect on phenotype (Mendell and Dietz 2001).

In addition to the DNA sequences proposed to be involved in the control of splicing detailed above, numerous other DNA regions are implicated as having roles in

controlling gene expression, mRNA processing and message degradation. It appears that gene expression is dependent both on the promoter region, often many tens of Kb upstream of the translation start site, and on enhancer and silencer sequences which may be a considerable distance 5' or 3' of the transcription start site. Consensus sequences for such regions are again poorly characterised making identification of such regions based on sequence information alone difficult (Werner 2002). Some predictions can be made however by alignment and comparison of non-coding sequence between species (Videl et al. 2003). The promoter region of the IL4R gene is approximately 30 Kb upstream of the translation start site and covers at least 4 Kb of DNA (Hackstein et al. 2001). This region was not sequenced in this study and it is possible that variation within the promoter or within as yet unidentified enhancer or silencer elements may also affect gene expression thereby encoding susceptibility to OA.

Protein synthesis is also regulated at the level of mRNA processing and degradation. Following transcription the mRNA molecule undergoes a number of modifications including splicing, capping and polyadenylation. Control of splicing has been detailed above.

The 5'UTR of the mRNA molecule plays an important role in protein translation, being responsible for the recruitment and scanning of the 40s ribosomal subunit. Mutations within this region can result in defects in both translation initiation and regulation and in transcript stability (Mendell and Dietz 2001). The 5'UTR of the IL4R gene was not covered in this study and it is possible that variants, which encode susceptibility to OA by affecting translation, may be present in this region. At the

other end of the gene, the 3'UTR is important for interaction of the mRNA with specific proteins and exosomes which can result in the rapid degradation of the molecule (Bevilacqua et al. 2003). Mutations in this region can therefore also have important effects on protein synthesis by affecting message stability (Bevilacqua et al. 2003). Numerous mutations within the 3'UTR of the IL4R gene were detected and it is possible that any one or a number of these may be involved in altering mRNA stability and therefore contributing to disease.

The regions described above, as potentially harbouring additional variants conferring susceptibility to disease, are very large and disparate, potentially covering hundreds of Kb. In addition, understanding as to the consensus sequence of the elements involved in regulation of protein expression and function and also the potential consequence of variations within these elements is limited. It would therefore seem that a sensible approach to dissect further the role of this gene in the OA disease process would be to undertake *in vivo* functional studies in to gene expression and transcript processing in addition to examining the effects of identified non-synonymous mutations on protein function.

Although it is very likely that the IL4R gene has a role in OA disease susceptibility it cannot be excluded that other gene(s) within the region of linkage may additionally encode susceptibility for OA. There are, at the present time, however, no other obvious, strong candidates within the 1-LOD support region of linkage on chromosome 16p.

In summary then, there is strong evidence for the involvement of the IL4R gene in the OA disease process, with at least three SNPs showing consistent evidence for association with both disease and evidence for linkage. In addition gene dosage studies appear to show that combinations of variant alleles are important in increasing risk to OA. The exact nature and effects of disease encoding variants may now be investigated by functional studies.

Chapter 7 - Summary and conclusions

7.1 Introduction

Osteoarthritis (OA) is a common, disabling disease of the elderly. It is uncommon in individuals under 45 years of age, and prevalence is higher in females than males. OA is characterised by focal cartilage loss accompanied by osteophytes, sclerosis and often some inflammatory response (Felson et al. 2000).

Primary OA has a large genetic component with heritability for hip and knee OA estimated from twin studies to be between 39 and 65% (MacGregor and Spencer 1999). Other systemic, local and environmental factors have also been shown to be involved in the initiation and progression of the disease leading to the description of OA as an oligogenic, multifactorial disease.

Identification of OA susceptibility genes will enhance our understanding of the biological basis of this disease leading to the development of new treatments.

Our group previously carried out a two-stage genome-wide linkage scan with a total of 481 families containing at least one pair of affected siblings (Chapman et al. 1999). Following stratification for sex, joint replaced and sex & joint replaced 6 chromosome regions - on chromosomes 2, 4, 6, 11, 16p and 16q - were identified as potentially harbouring OA susceptibility genes.

The aim of this study was to refine the regions on chromosomes 4 and 16 and then to test for association candidate genes identified within these regions.

7.2 Chromosome 4

7.2.1 Linkage mapping

Linkage mapping of chromosome 4 has identified a region showing evidence of linkage to OA that confirms and refines the region first identified by Loughlin et al in 1999. This region is restricted to the families with female siblings concordant for hip OA and has a multipoint LOD score (MLS) of 3.1. The 1-LOD support interval for this region is only 4 cM and resides at 4q13.1-4q13.2.

7.2.2 Conditioning analysis

There was no evidence for correlation between the family NPL scores at the region on chromosome 4 and the NPL scores for either the region on chromosome 16p or that on chromosome 16q (or with any of the other regions identified by this group).

7.2.3 ADAMTS3 candidate gene identification

At the time of the study there were no obvious candidate genes within the 1-LOD support interval on chromosome 4q and therefore the search for candidates was expanded outside this region. The ADAMTS3 gene, residing at 4q13.3 (approximately 4.5 Mb from the 1-LOD support interval) was identified as a strong candidate with a potentially important role in the processing of type II collagen in cartilage.

The exonic sequence, the 5' and 3' flanking intronic sequences and the 3'UTR of the ADAMTS3 gene were sequenced and SNPs identified. The ADAMTS3 is a highly conserved gene with only one common exonic SNP identified. Comparison of allele

frequencies between probands from the 146 female-THR families and controls for six of the SNPs identified by sequencing revealed two SNPs as showing significant evidence for an association with OA disease. A third SNP had a p-value which approached significance. None of these SNPs showed significant evidence for association to OA in an additional case cohort. When the results for this additional case cohort were combined with the results for the probands only one SNP, Intron12A/C, showed any evidence for association. This SNP showed no evidence for association of genetic variation with the evidence for linkage and therefore is unlikely to represent a genuine disease causing/associated polymorphism.

There is very little linkage disequilibrium between the six SNPs used in this study and therefore haplotype analysis provided very little useful information.

Although this study indicates that variants within the coding regions of the ADAMTS3 gene are not responsible for susceptibility to OA it remains possible that variants that affect splicing or expression, which would not likely have been detected by this study, may predispose to OA disease. If the ADAMTS3 gene is not the susceptibility gene on chromosome 4 there are a number of other genes which could be considered. These include the CXC chemokines and the vitamin D binding protein, GC.

7.2.4 Future studies

7.2.4.1 Further ADAMTS3 studies

Since ADAMTS3 is a good candidate gene it may warrant further investigation. Splicing studies, like those undertaken for IL4R, may yield insights as to the splicing patterns of the ADAMTS3 gene in diseased cartilage. In addition, quantitative allelic expression of the ADAMTS3 gene may be investigated. Expression studies have been undertaken by our group using methods based on those proposed by Cowles et al (2002). If either of these studies provide interesting results then functional studies involving model systems, cell culture, mutagenesis etc may yield further insights.

7.2.4.2 Investigation of other candidate genes

In addition to the ADAMTS3 gene there are other potential candidate genes within the linkage region on chromosome 4. Initial studies of the type used for the candidate genes reported here could be used to identify genes showing association with OA disease. Any genes showing evidence for association could then be investigated further either by using splicing or expression studies or by looking at the functional effects of variations.

7.2.4.3 Systematic association studies

Systematic association studies across the region identified by linkage mapping may be of use in narrowing the region and identifying disease-associated variants. This technique has been used successfully by Mira et al (2004) in leprosy. Initially they used a sparse map of SNPs within a 6.4 Mb region of DNA identified by linkage studies to define a region showing evidence for association. Then, using a denser

panel of markers, they were able to define an 80 Kb block of the genome showing evidence for association. Multivariate analysis was then used to identify two SNPs as responsible for the observed association. These reside in the upstream regulatory regions of two genes. This result has been confirmed in an independent sample. Such studies could conceivably be used across the 1-LOD support interval on chromosome 4 which spans only 7 Mb of DNA. The success of such studies would depend on many factors including the pattern of LD across the region and the choice of SNPs.

7.3 Chromosome 16

7.3.1 Linkage mapping

Linkage mapping of chromosome 16 identified two regions showing evidence of linkage to OA which confirm and refine the regions first identified by Loughlin et al in 1999. The first region, on chromosome 16p, is restricted to the families with female siblings concordant for hip OA and has a MLS of 1.7. The 1-LOD support interval for this region is 12 cM and resides at 16p12.3-16p12.1. The second region, on chromosome 16q, is accounted for by the families with female siblings and has a MLS of 1.9. The 1-LOD support interval for this region is 10 cM and resides at 16q22.1-16q23.1.

7.3.2 Conditioning analysis

Conditioning analysis, for the region on 16p conditioned on the region on 16q and for the region on 16q conditioned on the region on 16p, resulted in increased MLS and a narrowing of the regions. The region on chromosome 16p was narrowed to a 1-LOD

support interval of 7 cM with a MLS of 4.4. The region on chromosome 16q was narrowed to a 1-LOD support interval of 5 cM with a MLS of 5.7. These results need to be treated with caution however due to the two regions residing on the same chromosome.

7.3.3 IL4R candidate gene identification

The IL4R gene, residing at 16p12.1, towards the edge of the 1-LOD support interval on 16p was identified as a strong candidate gene. The IL4R has been proposed to have a role in the normal response of healthy chondrocytes to mechanical stimulation. Different types of the receptor for IL4 and downstream signalling pathways have been shown to be important in chondrocytes from normal versus osteoarthritic cartilage at least in model systems (Millward-Sadler et al. 1999; Millward-Sadler et al. 2000; Salter et al. 2001; Salter et al. 2002).

The IL4R is a highly polymorphic gene and the most polymorphic of all the cytokine receptor genes. There are many published reports on both the common variations within the gene and the potential role of these variants on protein function.

The coding exonic sequence, the 5' and 3' flanking intronic sequences and the 3'UTR were sequenced and variants identified. In addition two non-coding exons were also sequenced. Seven common exonic SNPs were selected for genotyping in addition to two SNPs in the promoter region of the gene. Comparison of allele frequencies between probands from the 146 female-THR families and controls for these SNPs identified two SNPs (S411L, S727A) as showing significant evidence for association with OA disease. A third SNP (N142N) had a p-value approaching significance. All

three of these SNPs and an additional SNP (Q551R) showed significant evidence for association to OA in an additional case cohort.

Linkage disequilibrium studies reveal a complex pattern of LD across the IL4R gene and suggest that there are potentially two independent associations between the IL4R gene and OA disease. The first association consists of SNPs N142N, S411L and S727A whilst the second consists of SNPs S478P and Q551R.

The three SNPs N142N, S411L and S727A all show evidence for a relationship between genetic variation and the evidence for linkage. This increases confidence for the involvement of these variants in encoding for OA susceptibility.

When gene dosage effects were investigated, it appears there is a requirement for at least one copy of the variant allele at each of SNPs N142N and S478P or SNPs N142N and Q551R in order to increase susceptibility to OA.

Splicing analysis of cDNA from cartilage RNA did not detect any aberrant splicing in individuals heterozygous for the N142N polymorphism.

7.3.4 Future studies

7.3.4.1 Further IL4R studies

It is hard to envisage how N142N could have an effect on IL4R protein function. It is suggested that this SNP may either directly affect expression of the gene or cause degradation of the transcript via nonsense mediated decay. Expression studies, such as

those referred to for the ADAMTS3 gene, could therefore be undertaken to investigate if this is the case. If such studies provide interesting results then functional studies involving, for example, protein constructs, cell cultures and/or model systems may yield further insights.

There are several reports as to the effects of many of the common IL4R SNPs on protein function but these are generally limited to cells of the immune system and little, if any, work has been carried out on IL4R and cartilage chondrocytes. In addition, no functional studies as to the effect(s) of SNPs S411L and S727A have been undertaken. Again, functional studies on chondrocytes, potentially using the human articular primary chondrocyte monolayer and mechanotransduction system first employed by Millward-Sadler et al (1999) may yield useful insights in to the effects of polymorphisms in the IL4R in cartilage. Since the polymorphisms under consideration are common it should be relatively easy to attain cartilage samples from individuals with the polymorphisms negating the need for site directed mutagenesis/transfection. Such systems could also be used to investigate the effect of combinations of SNPs on protein function.

It would also be reassuring to confirm the positive associations between the IL4R SNPs and OA in additional OA case cohorts.

7.3.4.2 Investigation of other candidate genes

Since the time of original candidate gene selection, when there were no obvious candidate genes within the 1-LOD support region on chromosome 16q, two potential

candidate genes have been positioned within this region. The first on 16q22.1 encodes for haptoglobin, a protein involved in binding haemoglobin facilitating its clearance by the liver. The protein therefore, has a role in the control of haemoglobin turnover and iron availability. Iron availability has been demonstrated to have a fundamental role in bone metabolism with depletion of iron being associated with demineralisation of bone (Angus et al. 1988). The capacity of haptoglobin to bind haemoglobin is dependent on a polymorphism within the haptoglobin gene and Pescarmona et al (2001) found an association between a particular haptoglobin genotype and increased risk of post menopausal osteoporosis. Given the inverse correlation observed between bone mineral density and OA (Dequeker 1997; Stewart et al. 2000) it is possible that polymorphisms within this gene could result indirectly in defects of bone metabolism by influencing iron availability, and therefore increase risk of OA disease.

The second potential candidate gene on 16q is the cartilage-derived, c type lectin (CLESCF1) gene on 16q23. Expression of this protein has been found to be restricted to cartilage and is especially prevalent in the upper hypertrophic and proliferating zones of developing rib. It is a trimeric protein with 3 globular domains probably involved in ligand binding and is closely related to tetranectin. Although its function is currently unknown it is postulated to have a role in either organisation of the cartilage matrix or as a protease precursor with a role in regulation of cartilage remodelling (Neame et al. 1999). This gene could therefore clearly have a role in both normal cartilage function and the pathogenesis of OA.

For both the regions on chromosome 4q and chromosome 16q there are a large number of hypothetical genes and genes with an unknown or poorly defined function. Any of these could in subsequent studies be identified as having potential roles within the OA disease process. In addition, novel functions are frequently discovered for well characterised genes, so that any number of genes could, in the future, be implicated as potential OA susceptibility genes. In addition, a number of genes within the regions e.g. Butyryl Coenzyme A synthetase 1 (BUCS1) and Ubiquinol-cytochrome c reductase core protein II (UQCRC2) encode for proteins that are involved in mitochondrial pathways that are active in a number of tissues. Unless studies in to tissue expression patterns have been carried out it is not possible to totally exclude such genes as being involved in cartilage/bone function and therefore being candidates for OA susceptibility. Limitations in knowledge such as those listed above are being constantly overcome as a result of the availability of increasing quantities of curated data, on numerous genes from multiple sources, within the human genome database.

7.3.4.3 Systematic association studies

As for the region on chromosome 4q it is possible that a systematic association study could be used across the 1-LOD support interval on 16q which spans 10 Mb of DNA. Again, the success of such a study would depend on the pattern of LD across this region and the choice of SNPs.

7.4 Concluding remarks

In conclusion, work reported here has identified regions on chromosome 4q, 16p and 16q as encoding for susceptibility to OA. There is strong evidence for association of polymorphisms in the IL4R gene with OA disease.

The research conducted during my DPhil has been carried on by others within the group.

The IL4R data reported in this thesis is now being tested from a functional viewpoint in a collaboration with Professor Donald Salter of Edinburgh University. As previously mentioned, Professor Salter has developed techniques for investigating the response of chondrocytes to mechanical stimulation. In the collaboration chondrocytes with different genotypes at the IL4R SNPs are being investigated to determine whether genotype influences response.

The finer mapping of chromosome 4 has been extended with a high density microsatellite association analysis of the linkage interval. This revealed a number of microsatellites demonstrating evidence for association. The genes positioned near to the positive microsatellites are currently being assessed. Intriguingly, one of the positive microsatellites is positioned immediately 5' to the ADAMTS3 gene. This has prompted the group to investigate whether polymorphism in regulatory elements of ADAMTS3 exist and, if they do, whether they are associated with OA. A project grant application has been awarded by the Arthritis Research Campaign to facilitate the identification of the chromosome 4 OA susceptibility gene.

Future investigation of all three regions and the IL4R gene will yield further insights in to the poorly understood OA disease process paving the way for creation of novel and more effective treatments for this debilitating disease.

Presentations, publications and prizes

Presentations

- 2001. Poster 'Gene Hunting in Primary Osteoarthritis' at Research into Ageing grant holders symposium, University College, London, UK
- 2002. Poster 'Unravelling The Genetics Of Primary Osteoarthritis - A Common Disabling Disease Of The Elderly' at EuroConference on Normal Ageing, Longevity and Age-Related Diseases, Spetses, Greece
- 2002. Poster 'Linkage analysis and candidate gene investigation in osteoarthritis' at American Society of Human Genetics annual conference, Baltimore, USA
- 2003. Poster and oral presentation as invited speaker 'Identification and Analysis of Genes Encoding Susceptibility for Osteoarthritis' at Research into Ageing grant holders symposium, University College, London, UK

Publications

- Forster T, Chapman K, Loughlin J (2004a) Common variants within the interleukin 4 receptor alpha gene (IL4R) are associated with susceptibility to osteoarthritis. Hum Genet 114:391-5

- Forster T, Chapman K, Marcelline L, Mustafa Z, Southam L, Loughlin J (2004b) Finer linkage mapping of primary osteoarthritis susceptibility loci on chromosomes 4 and 16 in families with affected women. *Arthritis Rheum* 50:98-102

- Kawahara C, Forster T, Chapman K, Carr A, Loughlin J (2004) Genetic association analysis of the IGFFBP7, ADAMTS3 and IL8 genes as the potential osteoarthritis susceptibility that maps to chromosome 4q. *Annals of the Rheumatic Diseases* (in press)

Prizes

- 2003. Prize for best poster at Research into Ageing grant holders symposium, University College, London, UK

- Abbaszade I, Liu RQ, Yang F, Rosenfeld SA, Ross OH, Link JR, Ellis DM, Tortorella MD, Pratta MA, Hollis JM, Wynn R, Duke JL, George HJ, Hillman MC, Jr., Murphy K, Wiswall BH, Copeland RA, Decicco CP, Bruckner R, Nagase H, Itoh Y, Newton RC, Magolda RL, Trzaskos JM, Burn TC (1999) Cloning and characterization of ADAMTS11, an aggrecanase from the ADAMTS family. *J Biol Chem* 274:23443-50
- Abecasis GR, Cookson WO (2000) GOLD--graphical overview of linkage disequilibrium. *Bioinformatics* 16:182-3
- Aigner T, Zhu Y, Chansky HH, Matsen FA, 3rd, Maloney WJ, Sandell LJ (1999) Reexpression of type IIA procollagen by adult articular chondrocytes in osteoarthritic cartilage. *Arthritis Rheum* 42:1443-50
- Aigner T, Kim HA (2002) Apoptosis and cellular vitality: issues in osteoarthritic cartilage degeneration. *Arthritis Rheum* 46:1986-96
- Akey JM, Zhang K, Xiong M, Jin L (2003) The effect of single nucleotide polymorphism identification strategies on estimates of linkage disequilibrium. *Mol Biol Evol* 20:232-42
- Alaaeddine N, Di Battista JA, Pelletier JP, Kiansa K, Cloutier JM, Martel-Pelletier J (1999) Inhibition of tumor necrosis factor alpha-induced prostaglandin E2 production by the antiinflammatory cytokines interleukin-4, interleukin-10, and interleukin-13 in osteoarthritic synovial fibroblasts: distinct targeting in the signaling pathways. *Arthritis Rheum* 42:710-8
- Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, Christy W, Cooke TD, Greenwald R, Hochberg M, et al. (1986) Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis Rheum* 29:1039-49
- Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M (2001) Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet* 69:936-50
- Angus RM, Sambrook PN, Pocock NA, Eisman JA (1988) Dietary intake and bone mineral density. *Bone Miner* 4:265-77
- Ardlie KG, Kruglyak L, Seielstad M (2002) Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 3:299-309
- Ateshian GA, Lai WM, Zhu WB, Mow VC (1994) An asymptotic solution for the contact of two biphasic cartilage layers. *J Biomech* 27:1347-60
- Badano JL, Katsanis N (2002) Beyond Mendel: an evolving view of human genetic disease transmission. *Nat Rev Genet* 3:779-89

- Bevilacqua A, Ceriani MC, Capaccioli S, Nicolin A (2003) Post-transcriptional regulation of gene expression by degradation of messenger RNAs. *J Cell Physiol* 195:356-72
- Bijkerk C, Houwing-Duistermaat JJ, Valkenburg HA, Meulenbelt I, Hofman A, Breedveld FC, Pols HA, van Duijn CM, Slagboom PE (1999) Heritabilities of radiologic osteoarthritis in peripheral joints and of disc degeneration of the spine. *Arthritis Rheum* 42:1729-35
- Blain EJ, Gilbert SJ, Wardale RJ, Capper SJ, Mason DJ, Duance VC (2001) Up-regulation of matrix metalloproteinase expression and activation following cyclical compressive loading of articular cartilage in vitro. *Arch Biochem Biophys* 396:49-55
- Bornstein P (2002) The NH(2)-terminal propeptides of fibrillar collagens: highly conserved domains with poorly understood functions. *Matrix Biol* 21:217-26
- Botstein D, Risch N (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet* 33 Suppl:228-37
- Brandi ML, Gennari L, Cerinic MM, Becherini L, Falchetti A, Masi L, Gennari C, Reginster JY (2001) Genetic markers of osteoarticular disorders: facts and hopes. *Arthritis Res* 3:270-80
- Brandt KD, Myers SL, Burr D, Albrecht M (1991) Osteoarthritic changes in canine articular cartilage, subchondral bone, and synovium fifty-four months after transection of the anterior cruciate ligament. *Arthritis Rheum* 34:1560-70
- Brandt KD (2003) Chondrocalcinosis, osteophytes and osteoarthritis. *Ann Rheum Dis* 62:499-500
- Bruyere O, Dardenne C, Lejeune E, Zegels B, Pahaut A, Richy F, Seidel L, Ethgen O, Henrotin Y, Reginster JY (2003) Subchondral tibial bone mineral density predicts future joint space narrowing at the medial femoro-tibial compartment in patients with knee osteoarthritis. *Bone* 32:541-5
- Bullough PG, Jagannath A (1983) The morphology of the calcification front in articular cartilage. Its significance in joint function. *J Bone Joint Surg Br* 65:72-8
- Burr DB (1998) The importance of subchondral bone in osteoarthrosis. *Curr Opin Rheumatol* 10:256-62
- Buschmann MD, Gluzband YA, Grodzinsky AJ, Hunziker EB (1995) Mechanical compression modulates matrix biosynthesis in chondrocyte/agarose culture. *J Cell Sci* 108 (Pt 4):1497-508

- Cardon LR, Bell JI (2001) Association study designs for complex diseases. *Nat Rev Genet* 2:91-9
- Cardon LR, Abecasis GR (2003) Using haplotype blocks to map human complex trait loci. *Trends Genet* 19:135-40
- Carney SL, Muir H (1988) The structure and function of cartilage proteoglycans. *Physiol Rev* 68:858-910
- Cawston TE, Ellis AJ, Bigg H, Curry V, Lean E, Ward D (1996) Interleukin-4 blocks the release of collagen fragments from bovine nasal cartilage treated with cytokines. *Biochim Biophys Acta* 1314:226-32
- Chapman K, Mustafa Z, Irven C, Carr AJ, Clipsham K, Smith A, Chitnavis J, Sinsheimer JS, Bloomfield VA, McCartney M, Cox O, Cardon LR, Sykes B, Loughlin J (1999) Osteoarthritis-susceptibility locus on chromosome 11q, detected by linkage. *Am J Hum Genet* 65:167-74
- Chapman KL, Mortier GR, Chapman K, Loughlin J, Grant ME, Briggs MD (2001) Mutations in the region encoding the von Willebrand factor A domain of matrilin-3 are associated with multiple epiphyseal dysplasia. *Nat Genet* 28:393-6
- Chapman K, Mustafa Z, Dowling B, Southam L, Carr A, Loughlin J (2002) Finer linkage mapping of primary hip osteoarthritis susceptibility on chromosome 11q in a cohort of affected female sibling pairs. *Arthritis Rheum* 46:1780-3
- Chitnavis J, Sinsheimer JS, Clipsham K, Loughlin J, Sykes B, Burge PD, Carr AJ (1997) Genetic influences in end-stage osteoarthritis. Sibling risks of hip and knee replacement for idiopathic osteoarthritis. *J Bone Joint Surg Br* 79:660-4
- Christgau S, Mouritzen U, Cloosi P, Lehman HJ, Christiansen C, Henriksen D (2001) Effect of bisphosphonate on cartilage metabolism assessed by a newly developed assay for collagen type II degradation. *Osteoarthritis Cartilage* 9 ((Suppl B):S7):19
- Clark JM, Huber JD (1990) The structure of the human subchondral plate. *J Bone Joint Surg Br* 72:866-73
- Colige A, Sieron AL, Li SW, Schwarze U, Petty E, Wertelecki W, Wilcox W, Krakow D, Cohn DH, Reardon W, Byers PH, Lapiere CM, Prockop DJ, Nussgens BV (1999) Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis are caused by mutations in the procollagen I N-proteinase gene. *Am J Hum Genet* 65:308-17

- Concannon P, Gogolin-Ewens KJ, Hinds DA, Wapelhorst B, Morrison VA, Stirling B, Mitra M, Farmer J, Williams SR, Cox NJ, Bell GI, Risch N, Spielman RS (1998) A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus. *Nat Genet* 19:292-6
- Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 261:921-3
- Cowles CR, Hirschhorn JN, Altshuler D, Lander ES (2002) Detection of regulatory variation in mouse genes. *Nat Genet* 32:432-7
- Cox NJ, Frigge M, Nicolae DL, Concannon P, Hanis CL, Bell GI, Kong A (1999) Loci on chromosomes 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans. *Nat Genet* 21:213-5
- Cox NJ (2001) Challenges in identifying genetic variation affecting susceptibility to type 2 diabetes: examples from studies of the calpain-10 gene. *Hum Mol Genet* 10:2301-5
- Creamer P, Hochberg MC (1997) Osteoarthritis. *Lancet* 350:503-8
- Croft P, Coggon D, Cruddas M, Cooper C (1992) Osteoarthritis of the hip: an occupational disease in farmers. *Bmj* 304:1269-72
- Darvasi A, Weinreb A, Minke V, Weller JI, Soller M (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* 134:943-51
- Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, Reed PW, Gough SC, Jenkins SC, Palmer SM, et al. (1994) A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 371:130-6
- Deichmann K, Bardutzky J, Forster J, Heinzmann A, Kuehr J (1997) Common polymorphisms in the coding part of the IL4-receptor gene. *Biochem Biophys Res Commun* 231:696-7
- Demissie S, Cupples LA, Myers R, Aliabadi P, Levy D, Felson DT (2002) Genome scan for quantity of hand osteoarthritis: the Framingham Study. *Arthritis Rheum* 46:946-52
- Dennison EM, Arden NK, Kellingray S, Croft P, Coggon D, Cooper C (1998) Hormone replacement therapy, other reproductive variables and symptomatic hip osteoarthritis in elderly white women: a case-control study. *Br J Rheumatol* 37:1198-202

- Dequeker J (1997) Inverse relationship of interface between osteoporosis and osteoarthritis. *J Rheumatol* 24:795-8
- Devlin B, Roeder K (1999) Genomic control for association studies. *Biometrics* 55:997-1004
- Dieppe P, Cushnaghan J, Young P, Kirwan J (1993) Prediction of the progression of joint space narrowing in osteoarthritis of the knee by bone scintigraphy. *Ann Rheum Dis* 52:557-63
- Dreyer SD, Zhou G, Lee B (1998) The long and the short of it: developmental genetics of the skeletal dysplasias. *Clin Genet* 54:464-73
- Drysdale CM, McGraw DW, Stack CB, Stephens JC, Judson RS, Nandabalan K, Arnold K, Ruano G, Liggett SB (2000) Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. *Proc Natl Acad Sci U S A* 97:10483-8
- Duerr RH, Barmada MM, Zhang L, Pfutzer R, Weeks DE (2000) High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11-12. *Am J Hum Genet* 66:1857-62
- Elliott DM, Guilak F, Vail TP, Wang JY, Setton LA (1999) Tensile properties of articular cartilage are altered by meniscectomy in a canine model of osteoarthritis. *J Orthop Res* 17:503-8
- Felson DT (1990) The epidemiology of knee osteoarthritis: results from the Framingham Osteoarthritis Study. *Semin Arthritis Rheum* 20:42-50
- Felson DT, Zhang Y, Hannan MT, Naimark A, Weissman BN, Aliabadi P, Levy D (1995) The incidence and natural history of knee osteoarthritis in the elderly. The Framingham Osteoarthritis Study. *Arthritis Rheum* 38:1500-5
- Felson DT, Couropmitree NN, Chaisson CE, Hannan MT, Zhang Y, McAlindon TE, LaValley M, Levy D, Myers RH (1998) Evidence for a Mendelian gene in a segregation analysis of generalized radiographic osteoarthritis: the Framingham Study. *Arthritis Rheum* 41:1064-71
- Felson DT, Zhang Y (1998) An update on the epidemiology of knee and hip osteoarthritis with a view to prevention. *Arthritis Rheum* 41:1343-55
- Felson DT, Lawrence RC, Dieppe PA, Hirsch R, Helmick CG, Jordan JM, Kington RS, Lane NE, Nevitt MC, Zhang Y, Sowers M, McAlindon T, Spector TD, Poole AR, Yanovski SZ, Ateshian G, Sharma L, Buckwalter JA, Brandt KD, Fries JF (2000) Osteoarthritis: new insights. Part 1: the disease and its risk factors. *Ann Intern Med* 133:635-46

- Fernandes RJ, Hirohata S, Engle JM, Colige A, Cohn DH, Eyre DR, Apte SS (2001) Procollagen II amino propeptide processing by ADAMTS-3. Insights on dermatosparaxis. *J Biol Chem* 276:31502-9
- Fernandes JC, Martel-Pelletier J, Pelletier JP (2002) The role of cytokines in osteoarthritis pathophysiology. *Biorheology* 39:237-46
- Flugsrud GB, Nordsletten L, Espehaug B, Havelin LI, Meyer HE (2002) Risk factors for total hip replacement due to primary osteoarthritis: a cohort study in 50,034 persons. *Arthritis Rheum* 46:675-82
- Forster T, Chapman K, Loughlin J (2004) Common variants within the interleukin 4 receptor alpha gene (IL4R) are associated with susceptibility to osteoarthritis. *Hum Genet* 114:391-5
- Forster T, Chapman K, Marcelline L, Mustafa Z, Southam L, Loughlin J (2004) Finer linkage mapping of primary osteoarthritis susceptibility loci on chromosomes 4 and 16 in families with affected women. *Arthritis Rheum* 50:98-102
- Frederick MJ, Clayman GL (2001) Chemokines in cancer. *Expert Rev Mol Med* 2001:1-18
- Fukui N, McAlinden A, Zhu Y, Crouch E, Broekelmann TJ, Mecham RP, Sandell LJ (2002) Processing of type II procollagen amino propeptide by matrix metalloproteinases. *J Biol Chem* 277:2193-201
- Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. *Science* 296:2225-9
- Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD (2000) Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 97:11383-90
- Gillaspy E, Spreckley K, Wallis G, Doherty M, Spector TD (2002) Investigation of linkage on chromosome 2q and hand and knee osteoarthritis. *Arthritis Rheum* 46:3386-7
- Glazier AM, Nadeau JH, Aitman TJ (2002) Finding genes that underlie complex traits. *Science* 298:2345-9
- Goldring MB (2000) Osteoarthritis and cartilage: the role of cytokines. *Curr Rheumatol Rep* 2:459-65
- Goldring MB (2000) The role of the chondrocyte in osteoarthritis. *Arthritis Rheum* 43:1916-26

- Grabowski PJ (1998) Splicing regulation in neurons: tinkering with cell-specific control. *Cell* 92:709-12
- Graveley BR (2001) Alternative splicing: increasing diversity in the proteomic world. *Trends Genet* 17:100-7
- Green WT, Jr., Martin GN, Eanes ED, Sokoloff L (1970) Microradiographic study of the calcified layer of articular cartilage. *Arch Pathol* 90:151-8
- Gretarsdottir S, Sveinbjornsdottir S, Jonsson HH, Jakobsson F, Einarsdottir E, Agnarsson U, Shkolny D, et al. (2002) Localization of a susceptibility gene for common forms of stroke to 5q12. *Am J Hum Genet* 70:593-603
- Groman JD, Meyer ME, Wilmott RW, Zeitlin PL, Cutting GR (2002) Variant cystic fibrosis phenotypes in the absence of CFTR mutations. *N Engl J Med* 347:401-7
- Gudbjartsson DF, Jonasson K, Frigge ML, Kong A (2000) Allegro, a new computer program for multipoint linkage analysis. *Nat Genet* 25:12-3
- Hackstein H, Hofmann H, Bohnert A, Bein G (1999) Definition of human interleukin-4 receptor alpha chain haplotypes and allelic association with atopy markers. *Hum Immunol* 60:1119-27
- Hackstein H, Hecker M, Kruse S, Bohnert A, Ober C, Deichmann KA, Bein G (2001) A novel polymorphism in the 5' promoter region of the human interleukin-4 receptor alpha-chain gene is associated with decreased soluble interleukin-4 receptor protein levels. *Immunogenetics* 53:264-9
- Hage T, Reinemer P, Sebald W (1998) Crystals of a 1:1 complex between human interleukin-4 and the extracellular domain of its receptor alpha chain. *Eur J Biochem* 258:831-6
- Hakonarson H, Bjornsdottir US, Halapi E, Palsson S, Adalsteinsdottir E, Gislason D, Finnbogason G, Gislason T, Kristjansson K, Arnason T, Birkisson I, Frigge ML, Kong A, Gulcher JR, Stefansson K (2002) A major susceptibility gene for asthma maps to chromosome 14q24. *Am J Hum Genet* 71:483-91
- Hannan MT, Felson DT, Anderson JJ, Naimark A, Kannel WB (1990) Estrogen use and radiographic osteoarthritis of the knee in women. The Framingham Osteoarthritis Study. *Arthritis Rheum* 33:525-32
- Hanson EM, Dickensheets H, Qu CK, Donnelly RP, Keegan AD (2003) Regulation of the dephosphorylation of Stat6. Participation of Tyr-713 in the interleukin-4 receptor alpha, the tyrosine phosphatase SHP-1, and the proteasome. *J Biol Chem* 278:3903-11

- Haque SJ, Harbor P, Tabrizi M, Yi T, Williams BR (1998) Protein-tyrosine phosphatase Shp-1 is a negative regulator of IL-4- and IL-13-dependent signal transduction. *J Biol Chem* 273:33893-6
- Harada S, Rodan GA (2003) Control of osteoblast function and regulation of bone mass. *Nature* 423:349-55
- Hardingham T, Bayliss M (1990) Proteoglycans of articular cartilage: changes in aging and in joint disease. *Semin Arthritis Rheum* 20:12-33
- Hart DJ, Cronin C, Daniels M, Worthy T, Doyle DV, Spector TD (2002) The relationship of bone density and fracture to incident and progressive radiographic osteoarthritis of the knee: the Chingford Study. *Arthritis Rheum* 46:92-9
- Hartz AJ, Fischer ME, Bril G, Kelber S, Rupley D, Jr., Oken B, Rimm AA (1986) The association of obesity with joint pain and osteoarthritis in the HANES data. *J Chronic Dis* 39:311-9
- Hedbom E, Heinegard D (1993) Binding of fibromodulin and decorin to separate sites on fibrillar collagens. *J Biol Chem* 268:27307-12
- Heinegard D, Oldberg A (1989) Structure and biology of cartilage and bone matrix noncollagenous macromolecules. *Faseb J* 3:2042-51
- Heinegard DB, M and Lorenzo, P (1998) Biochemistry and metabolism of normal and osteoarthritic cartilage. In: Brandt KD, M and Lohmander, L (ed) *Osteoarthritis*. Oxford University Press, Oxford
- Hershey GK, Friedrich MF, Esswein LA, Thomas ML, Chatila TA (1997) The association of atopy with a gain-of-function mutation in the alpha subunit of the interleukin-4 receptor. *N Engl J Med* 337:1720-5
- Hirsch R, Lethbridge-Cejku M, Hanson R, Scott WW, Jr., Reichle R, Plato CC, Tobin JD, Hochberg MC (1998) Familial aggregation of osteoarthritis: data from the Baltimore Longitudinal Study on Aging. *Arthritis Rheum* 41:1227-32
- Homandberg GA, Meyers R, Xie DL (1992) Fibronectin fragments cause chondrolysis of bovine articular cartilage slices in culture. *J Biol Chem* 267:3597-604
- Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, Hinokio Y, Lindner TH, Mashima H, Schwarz PE, del Bosque-Plata L, Oda Y, Yoshiuchi I, Colilla S, Polonsky KS, Wei S, Concannon P, Iwasaki N, Schulze J, Baier LJ, Bogardus C, Groop L, Boerwinkle E, Hanis CL, Bell GI (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 26:163-75

- Hovatta I, Lichtermann D, Juvonen H, Suvisaari J, Terwilliger JD, Arajärvi R, Kokko-Sahin ML, Ekelund J, Lonnqvist J, Peltonen L (1998) Linkage analysis of putative schizophrenia gene candidate regions on chromosomes 3p, 5q, 6p, 8p, 20p and 22q in a population-based sampled Finnish family set. *Mol Psychiatry* 3:452-7
- Huebner JL, Hanes MA, Beekman B, TeKoppele JM, Kraus VB (2002) A comparative analysis of bone and cartilage metabolism in two strains of guinea-pig with varying degrees of naturally occurring osteoarthritis. *Osteoarthritis Cartilage* 10:758-67
- Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 411:599-603
- Hulth A, Lindberg L, Telhag H (1972) Mitosis in human osteoarthritic cartilage. *Clin Orthop* 84:197-9
- Hurskainen TL, Hirohata S, Seldin MF, Apte SS (1999) ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of zinc metalloproteases. General features and genomic distribution of the ADAM-TS family. *J Biol Chem* 274:25555-63
- Ingvarsson T (2000) Prevalence and inheritance of hip osteoarthritis in Iceland. *Acta Orthop Scand Suppl* 298:1-46
- Ingvarsson T, Stefansson SE, Hallgrimsdottir IB, Frigge ML, Jonsson H, Jr., Gulcher J, Jonsson H, Ragnarsson JJ, Lohmander LS, Stefansson K (2000) The inheritance of hip osteoarthritis in Iceland. *Arthritis Rheum* 43:2785-92
- Ingvarsson T, Stefansson SE, Gulcher JR, Jonsson HH, Jonsson H, Frigge ML, Palsdottir E, Olafsdottir G, Jonsdottir T, Walters GB, Lohmander LS, Stefansson K (2001) A large Icelandic family with early osteoarthritis of the hip associated with a susceptibility locus on chromosome 16p. *Arthritis Rheum* 44:2548-55
- Jeffreys AJ, Kauppi L, Neumann R (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat Genet* 29:217-22
- Jeffreys AJ, Neumann R (2002) Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. *Nat Genet* 31:267-71

- Jimenez SA, Williams C. J. (1999) Genetic and metabolic aspects. In: Reginster JYP, J. P., Martel-Pelletier, J. Henrotin, Y. (ed) Osteoarthritis: Clinical and Experimental Aspects. Springer, New York, pp 134-156
- Johnson GC, Todd JA (2000) Strategies in complex disease mapping. *Curr Opin Genet Dev* 10:330-4
- Jones A, Doherty M (1995) ABC of rheumatology. Osteoarthritis. *Bmj* 310:457-60
- Jonsson H, Manolescu I, Stefansson SE, Ingvarsson T, Jonsson HH, Manolescu A, Gulcher J, Stefansson K (2003) The inheritance of hand osteoarthritis in Iceland. *Arthritis Rheum* 48:391-5
- Jortikka MO, Inkinen RI, Tammi MI, Parkkinen JJ, Haapala J, Kiviranta I, Helminen HJ, Lammi MJ (1997) Immobilisation causes longlasting matrix changes both in the immobilised and contralateral joint cartilage. *Ann Rheum Dis* 56:255-61
- Jung T, Wagner K, Neumann C, Heusser CH (1999) Enhancement of human IL-4 activity by soluble IL-4 receptors in vitro. *Eur J Immunol* 29:864-71
- Jung T, Schrader N, Hellwig M, Enssle KH, Neumann C (1999) Soluble human interleukin-4 receptor is produced by activated T cells under the control of metalloproteinases. *Int Arch Allergy Immunol* 119:23-30
- Kalichman L, Kobylansky E, Malkin I, Yakovenko K, Livshits G (2003) Search for linkage between hand osteoarthritis and 11q 12-13 chromosomal segment. *Osteoarthritis Cartilage* 11:561-8
- Kaplan N, Morris R (2001) Issues concerning association studies for fine mapping a susceptibility gene for a complex disease. *Genet Epidemiol* 20:432-57
- Kaprio J, Kujala UM, Peltonen L, Koskenvuo M (1996) Genetic liability to osteoarthritis may be greater in women than men. *Bmj* 313:232
- Karvonen RL, Negendank WG, Teitge RA, Reed AH, Miller PR, Fernandez-Madrid F (1994) Factors affecting articular cartilage thickness in osteoarthritis and aging. *J Rheumatol* 21:1310-8
- Kashiwada M, Giallourakis CC, Pan PY, Rothman PB (2001) Immunoreceptor tyrosine-based inhibitory motif of the IL-4 receptor associates with SH2-containing phosphatases and regulates IL-4-induced proliferation. *J Immunol* 167:6382-7
- Kauppi L, Sajantila A, Jeffreys AJ (2003) Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region. *Hum Mol Genet* 12:33-40

- Keegan AD, Nelms K, White M, Wang LM, Pierce JH, Paul WE (1994) An IL-4 receptor region containing an insulin receptor motif is important for IL-4-mediated IRS-1 phosphorylation and cell growth. *Cell* 76:811-20
- Kellgren J (1963) Genetic factors in generalized osteo-arthritis. *Annals of the rheumatic diseases* 22:237
- Knight JC (2003) Functional implications of genetic variation in non-coding DNA for disease susceptibility and gene regulation. *Clin Sci (Lond)* 104:493-501
- Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 61:1179-88
- Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K (2002) A high-resolution recombination map of the human genome. *Nat Genet* 31:241-7
- Kruglyak L, Lander ES (1995) High-resolution genetic mapping of complex traits. *Am J Hum Genet* 56:1212-23
- Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 58:1347-63
- Kruglyak L, Nickerson DA (2001) Variation is the spice of life. *Nat Genet* 27:234-6
- Kruse S, Japha T, Tedner M, Sparholt SH, Forster J, Kuehr J, Deichmann KA (1999) The polymorphisms S503P and Q576R in the interleukin-4 receptor alpha gene are associated with atopy and influence the signal transduction. *Immunology* 96:365-71
- Kruse S, Forster J, Kuehr J, Deichmann KA (1999) Characterization of the membrane-bound and a soluble form of human IL-4 receptor alpha produced by alternative splicing. *Int Immunol* 11:1965-70
- Kuettner K, Goldberg, V.M, (1995) Osteoarthritic disorders. American Academy of Orthopaedic Surgeons, Rosemont
- Kuida S, Beier DR (2000) Genetic localization of interacting modifiers affecting severity in a murine model of polycystic kidney disease. *Genome Res* 10:49-54
- Kwok PY, Deng Q, Zakeri H, Taylor SL, Nickerson DA (1996) Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. *Genomics* 31:123-6
- Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for

- interpreting and reporting linkage results. *Nat Genet* 11:241-7
- Lane NE, Nevitt MC (2002) Osteoarthritis, bone mass, and fractures: how are they related? *Arthritis Rheum* 46:1-4
- Lanyon P, Muir K, Doherty S, Doherty M (2000) Assessment of a genetic contribution to osteoarthritis of the hip: sibling study. *Bmj* 321:1179-83
- Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci U S A* 81:3443-6
- Lawrence JS (1977) Rheumatism in populations. Heinemann, London
- Leiter EH, Reifsnyder PC, Flurkey K, Partke HJ, Junger E, Herberg L (1998) NIDDM genes in mice: deleterious synergism by both parental genomes contributes to diabetogenic thresholds. *Diabetes* 47:1287-95
- Leppavuori J, Kujala U, Kinnunen J, Kaprio J, Nissila M, Heliovaara M, Klinger N, Partanen J, Terwilliger JD, Peltonen L (1999) Genome scan for predisposing loci for distal interphalangeal joint osteoarthritis: evidence for a locus on 2q. *Am J Hum Genet* 65:1060-7
- Lesage S, Zouali H, Cezard JP, Colombel JF, Belaiche J, Almer S, Tysk C, O'Morain C, Gassull M, Binder V, Finkel Y, Modigliani R, Gower-Rousseau C, Macry J, Merlin F, Chamaillard M, Jannot AS, Thomas G, Hugot JP (2002) CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Genet* 70:845-57
- Li SW, Sieron AL, Fertala A, Hojima Y, Arnold WV, Prockop DJ (1996) The C-proteinase that processes procollagens to fibrillar collagens is identical to the protein previously identified as bone morphogenic protein-1. *Proc Natl Acad Sci U S A* 93:5127-30
- Lieverse AM, Bierma-Zeinstra SM, Verhagen AP, van Baar ME, Verhaar JA, Koes BW (2002) Influence of obesity on the development of osteoarthritis of the hip: a systematic review. *Rheumatology (Oxford)* 41:1155-62
- Liu H, McKenna LA, Dean MF (2000) An N-terminal peptide from link protein can stimulate biosynthesis of collagen by human articular cartilage. *Arch Biochem Biophys* 378:116-22
- Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 33:177-82
- Long P, Gassner R, Agarwal S (2001) Tumor necrosis factor alpha-dependent proinflammatory gene induction is inhibited by cyclic tensile strain in articular chondrocytes in vitro. *Arthritis Rheum* 44:2311-9

- Lotz M, Terkeltaub R, Villiger PM (1992) Cartilage and joint inflammation. Regulation of IL-8 expression by human articular chondrocytes. *J Immunol* 148:466-73
- Loughlin J, Mustafa Z, Irven C, Smith A, Carr AJ, Sykes B, Chapman K (1999) Stratification analysis of an osteoarthritis genome screen-suggestive linkage to chromosomes 4, 6, and 16. *Am J Hum Genet* 65:1795-8
- Loughlin J, Mustafa Z, Smith A, Irven C, Carr AJ, Clipsham K, Chitnavis J, Bloomfield VA, McCartney M, Cox O, Sinsheimer JS, Sykes B, Chapman KE (2000) Linkage analysis of chromosome 2q in osteoarthritis. *Rheumatology (Oxford)* 39:377-81
- Loughlin J, Dowling B, Mustafa Z, Smith A, Sykes B, Chapman K (2000) Analysis of the association of the matrillin-1 gene (CRTM) with osteoarthritis: comment on the article by Meulenbelt et al. *Arthritis Rheum* 43:1423-5
- Loughlin J (2001) Genetic epidemiology of primary osteoarthritis. *Curr Opin Rheumatol* 13:111-6
- Loughlin J (2002) Genome studies and linkage in primary osteoarthritis. *Rheum Dis Clin North Am* 28:95-109
- Loughlin J, Mustafa Z, Dowling B, Southam L, Marcelline L, Raina SS, Ala-Kokko L, Chapman K (2002) Finer linkage mapping of a primary hip osteoarthritis susceptibility locus on chromosome 6. *Eur J Hum Genet* 10:562-8
- Loughlin J, Dowling B, Mustafa Z, Southam L, Chapman K (2002) Refined linkage mapping of a hip osteoarthritis susceptibility locus on chromosome 2q. *Rheumatology (Oxford)* 41:955-6
- Loughlin J, Dowling B, Mustafa Z, Chapman K (2002) Association of the interleukin-1 gene cluster on chromosome 2q13 with knee osteoarthritis. *Arthritis Rheum* 46:1519-27
- Lubberts E, Joosten LA, van Den Bersselaar L, Helsen MM, Bakker AC, van Meurs JB, Graham FL, Richards CD, van Den Berg WB (1999) Adenoviral vector-mediated overexpression of IL-4 in the knee joint of mice with collagen-induced arthritis prevents cartilage destruction. *J Immunol* 163:4546-56
- Lubberts E, Joosten LA, van de Loo FA, van den Gersselaar LA, van den Berg WB (2000) Reduction of interleukin-17-induced inhibition of chondrocyte proteoglycan synthesis in intact murine articular cartilage by interleukin-4. *Arthritis Rheum* 43:1300-6
- MacGregor AJ, Spector TD (1999) Twins and the genetic architecture of osteoarthritis. *Rheumatology (Oxford)* 38:583-8

- MacGregor AJ, Antoniadou L, Andrew T, Barber MJ, Hunter D, Hart DJ, Spector TD (2002) Genetic Expression in Osteoarthritis is joint specific not generalised: evidence from a radiographic study of twins. Paper presented at BSR meeting
- Manek NJ, Hart D, Spector TD, MacGregor AJ (2003) The association of body mass index and osteoarthritis of the knee joint: an examination of genetic and environmental influences. *Arthritis Rheum* 48:1024-9
- Mankin HJ, Thrasher AZ (1975) Water content and binding in normal and osteoarthritic human cartilage. *J Bone Joint Surg Am* 57:76-80
- Mankin HJ, Brandt K.D, Shulman L.E, (1986) Workshop on etiopathogenesis of osteoarthritis: proceedings and recommendations. *Journal of Rheumatology* 13:1130-1160
- Mansell JP, Tarlton JF, Bailey AJ (1997) Biochemical evidence for altered subchondral bone collagen metabolism in osteoarthritis of the hip. *Br J Rheumatol* 36:16-9
- Mansell JP, Bailey AJ (1998) Abnormal cancellous bone collagen metabolism in osteoarthritis. *J Clin Invest* 101:1596-603
- March LM, Bagga H (2004) Epidemiology of osteoarthritis in Australia. *Med J Aust* 180:S6-10
- Marchini J, Cardon LR, Phillips MS, Donnelly P (2004) The effects of human population structure on large genetic association studies. *Nat Genet* 36:512-7
- Martin JA, Buckwalter JA (2003) The role of chondrocyte senescence in the pathogenesis of osteoarthritis and in limiting cartilage repair. *J Bone Joint Surg Am* 85-A Suppl 2:106-10
- McAlindon TE, Felson DT, Zhang Y, Hannan MT, Aliabadi P, Weissman B, Rush D, Wilson PW, Jacques P (1996) Relation of dietary intake and serum levels of vitamin D to progression of osteoarthritis of the knee among participants in the Framingham Study. *Ann Intern Med* 125:353-9
- McKenna LA, Gehrsitz A, Soder S, Eger W, Kirchner T, Aigner T (2000) Effective isolation of high-quality total RNA from human adult articular cartilage. *Anal Biochem* 286:80-5
- Mendell JT, Dietz HC (2001) When the message goes awry: disease-producing mutations that influence mRNA content and performance. *Cell* 107:411-4
- Meng Z, Zaykin DV, Xu CF, Wagner M, Ehm MG (2003) Selection of genetic markers for association analyses, using linkage disequilibrium and haplotypes. *Am J Hum Genet* 73:115-30

- Mente PL, Lewis JL (1994) Elastic modulus of calcified cartilage is an order of magnitude less than that of subchondral bone. *J Orthop Res* 12:637-47
- Meulenbelt I, Bijkerk C, Breedveld FC, Slagboom PE (1997) Genetic linkage analysis of 14 candidate gene loci in a family with autosomal dominant osteoarthritis without dysplasia. *J Med Genet* 34:1024-7
- Meulenbelt I, Bijkerk C, De Wildt SC, Miedema HS, Breedveld FC, Pols HA, Hofman A, Van Duijn CM, Slagboom PE (1999) Haplotype analysis of three polymorphisms of the COL2A1 gene and associations with generalised radiological osteoarthritis. *Ann Hum Genet* 63 (Pt 5):393-400
- Millward-Sadler SJ, Wright MO, Lee H, Nishida K, Caldwell H, Nuki G, Salter DM (1999) Integrin-regulated secretion of interleukin 4: A novel pathway of mechanotransduction in human articular chondrocytes. *J Cell Biol* 145:183-9
- Millward-Sadler SJ, Wright MO, Davies LW, Nuki G, Salter DM (2000) Mechanotransduction via integrins and interleukin-4 results in altered aggrecan and matrix metalloproteinase 3 gene expression in normal, but not osteoarthritic, human articular chondrocytes. *Arthritis Rheum* 43:2091-9
- Mira MT, Alcais A, Nguyen VT, Moraes MO, Di Flumeri C, Vu HT, Mai CP, Nguyen TH, Nguyen NB, Pham XK, Sarno EN, Alter A, Montpetit A, Moraes ME, Moraes JR, Dore C, Gallant CJ, Lepage P, Verner A, Van De Vosse E, Hudson TJ, Abel L, Schurr E (2004) Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature* 427:636-40
- Mirza MM, Fisher SA, King K, Cuthbert AP, Hampe J, Sanderson J, Mansfield J, Donaldson P, Macpherson AJ, Forbes A, Schreiber S, Lewis CM, Mathew CG (2003) Genetic Evidence for Interaction of the 5q31 Cytokine Locus and the CARD15 Gene in Crohn Disease. *Am J Hum Genet* 72:1018-22
- Mitsuyasu H, Yanagihara Y, Mao XQ, Gao PS, Arinobu Y, Ihara K, Takabayashi A, Hara T, Enomoto T, Sasaki S, Kawai M, Hamasaki N, Shirakawa T, Hopkin JM, Izuhara K (1999) Cutting edge: dominant effect of Ile50Val variant of the human IL-4 receptor alpha-chain in IgE synthesis. *J Immunol* 162:1227-31
- Modrek B, Lee C (2002) A genomic view of alternative splicing. *Nat Genet* 30:13-9
- Moo V, Sieper J, Herzog V, Muller BM (2001) Regulation of expression of cytokines and growth factors in osteoarthritic cartilage explants. *Clin Rheumatol* 20:353-8
- Moos V, Rudwaleit M, Herzog V, Hohlig K, Sieper J, Muller B (2000) Association of genotypes affecting the expression of interleukin-1beta or interleukin-1 receptor antagonist with osteoarthritis. *Arthritis Rheum* 43:2417-22

- Murata T, Puri RK (1997) Comparison of IL-13- and IL-4-induced signaling in EBV-immortalized human B cells. *Cell Immunol* 175:33-40
- Murillo L, Crusius JB, van Bodegraven AA, Alizadeh BZ, Pena AS (2002) CARD15 gene and the classification of Crohn's disease. *Immunogenetics* 54:59-61
- Murillo L, Crusius JB, van Bodegraven AA, Alizadeh BZ, Pena AS (2002) CARD15 gene and the classification of Crohn's disease. *Immunogenetics* 54:59-61
- Mustafa Z, Chapman K, Irven C, Carr AJ, Clipsham K, Chitnavis J, Sinsheimer JS, Bloomfield VA, McCartney M, Cox O, Sykes B, Loughlin J (2000) Linkage analysis of candidate genes as susceptibility loci for osteoarthritis-suggestive linkage of COL9A1 to female hip osteoarthritis. *Rheumatology (Oxford)* 39:299-306
- Nagase T, Ishikawa K, Nakajima D, Ohira M, Seki N, Miyajima N, Tanaka A, Kotani H, Nomura N, Ohara O (1997) Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro. *DNA Res* 4:141-50
- Neame PJ, Tapp H, Grimm DR (1999) The cartilage-derived, C-type lectin (CLECSF1): structure of the gene and chromosomal location. *Biochim Biophys Acta* 1446:193-202
- Nebert DW, Dalton TP, Stuart GW, Carvan MJ, 3rd (2000) "Gene-swap knock-in" cassette in mice to study allelic differences in human genes. *Ann N Y Acad Sci* 919:148-70
- Nemoto O, Yamada H, Kikuchi T, Shinmei M, Obata K, Sato H, Seiki M, Shimmei M (1997) Suppression of matrix metalloproteinase-3 synthesis by interleukin-4 in human articular chondrocytes. *J Rheumatol* 24:1774-9
- Neuhold LA, Killar L, Zhao V, Sung M-LA, Warner L, Kulik J (2001) Osteoarthritis develops in transgenic mice expressing constitutively active human collagenase-3 (MMP-13) in articular cartilage. *The Journal of Clinical Investigation* 107:35-44
- Neuman P, Hulth A, Linden B, Johnell O, Dahlberg L (2003) The role of osteophytic growth in hip osteoarthritis. *Int Orthop*
- Newman B, Wallis GA (2002) Is osteoarthritis a genetic disease? *Clin Invest Med* 25:139-49
- Nickerson DA, Taylor SL, Fullerton SM, Weiss KM, Clark AG, Stengard JH, Salomaa V, Boerwinkle E, Sing CF (2000) Sequence diversity and large-scale typing of SNPs in the human apolipoprotein E gene. *Genome Res* 10:1532-45
- Nordborg M, Tavare S (2002) Linkage disequilibrium: what history has to tell us.

Trends Genet 18:83-90

- Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, Britton H, Moran T, Karaliuskas R, Duerr RH, Achkar JP, Brant SR, Bayless TM, Kirschner BS, Hanauer SB, Nunez G, Cho JH (2001) A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 411:603-6
- Oliveria SA, Felson DT, Reed JI, Cirillo PA, Walker AM (1995) Incidence of symptomatic hand, hip, and knee osteoarthritis among patients in a health maintenance organization. *Arthritis Rheum* 38:1134-41
- Olson JM, Goddard KA, Dudek DM (2002) A second locus for very-late-onset Alzheimer disease: a genome scan reveals linkage to 20p and epistasis between 20p and the amyloid precursor protein region. *Am J Hum Genet* 71:154-61
- Page WF, Hoaglund FT, Steinbach LS, Heath AC (2003) Primary osteoarthritis of the hip in monozygotic and dizygotic male twins. *Twin Res* 6:147-51
- Palotie A, Vaisanen P, Ott J, Ryhanen L, Elima K, Vikkula M, Cheah K, Vuorio E, Peltonen L (1989) Predisposition to familial osteoarthrosis linked to type II collagen gene. *Lancet* 1:924-7
- Pan PY, Rothman P (1999) IL-4 receptor mutations. *Curr Opin Immunol* 11:615-20
- Patil N, Berno AJ, Hinds DA, Barrett WA, Doshi JM, Hacker CR, Kautzer CR, Lee DH, Marjoribanks C, McDonough DP, Nguyen BT, Norris MC, Sheehan JB, Shen N, Stern D, Stokowski RP, Thomas DJ, Trulson MO, Vyas KR, Frazer KA, Fodor SP, Cox DR (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 294:1719-23
- Pericak-Vance MA, Bebout JL, Gaskell PC, Jr., Yamaoka LH, Hung WY, Alberts MJ, Walker AP, Bartlett RJ, Haynes CA, Welsh KA (1991) Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. *Am J Hum Genet* 48:1034-50
- Pernis A, Witthuhn B, Keegan AD, Nelms K, Garfein E, Ihle JN, Paul WE, Pierce JH, Rothman P (1995) Interleukin 4 signals through two related pathways. *Proc Natl Acad Sci U S A* 92:7971-5
- Petes TD (2001) Meiotic recombination hot spots and cold spots. *Nat Rev Genet* 2:360-9
- Phillips MS, Lawrence R, Sachidanandam R, Morris AP, Balding DJ, Donaldson MA, Studebaker JF, et al. (2003) Chromosome-wide distribution of haplotype

- blocks and the role of recombination hot spots. *Nat Genet* 33:382-7
- Picciotto MR, Caldarone BJ, King SL, Zachariou V (2000) Nicotinic receptors in the brain. Links between molecular biology and behavior. *Neuropsychopharmacology* 22:451-65
- Poole AR, Ionescu M, Swan A, Dieppe PA (1994) Changes in cartilage metabolism in arthritis are reflected by altered serum and synovial fluid levels of the cartilage proteoglycan aggrecan. Implications for pathogenesis. *J Clin Invest* 94:25-33
- Poole AR (1997) Cartilage in health and disease. In: Koopman WJ (ed) *Arthritis and allied conditions: a textbook of rheumatology*. Williams and Wilkins, Baltimore, pp 255 - 308
- Poole AR (2001) Cartilage in health and disease. In: Koopman W (ed) *Arthritis and allied conditions. A textbook of rheumatology*. Lippincott, Williams and Wilkins, Philadelphia, pp 226-84
- Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. *Am J Hum Genet* 69:1-14
- Radin EL, Paul IL, Lowy M (1970) A comparison of the dynamic force transmitting properties of subchondral bone and articular cartilage. *J Bone Joint Surg Am* 52:444-56
- Redler I, Mow VC, Zimny ML, Mansell J (1975) The ultrastructure and biomechanical significance of the tidemark of articular cartilage. *Clin Orthop*:357-62
- Reginato AM, Olsen BR (2002) The role of structural genes in the pathogenesis of osteoarthritic disorders. *Arthritis Res* 4:337-45
- Risma KA, Wang N, Andrews RP, Cunningham CM, Ericksen MB, Bernstein JA, Chakraborty R, Hershey GK (2002) V75R576 IL-4 receptor alpha is associated with allergic asthma and enhanced IL-4 receptor function. *J Immunol* 169:1604-10
- Robberson BL, Cote GJ, Berget SM (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* 10:84-94
- Roberts SB, MacLean CJ, Neale MC, Eaves LJ, Kendler KS (1999) Replication of linkage studies of complex traits: an examination of variation in location estimates. *Am J Hum Genet* 65:876-84
- Roby P, Eyre S, Worthington J, Ramesar R, Cilliers H, Beighton P, Grant M, Wallis G (1999) Autosomal dominant (Beukes) premature degenerative osteoarthropathy of the hip joint maps to an 11-cM region on chromosome

4q35. *Am J Hum Genet* 64:904-8

Ryan JJ, McReynolds LJ, Keegan A, Wang LH, Garfein E, Rothman P, Nelms K, Paul WE (1996) Growth and gene expression are predominantly controlled by distinct regions of the human IL-4 receptor. *Immunity* 4:123-32

Saladin KS, Fox (2002) Human Development. In: *Anatomy and Physiology: The Unity of Form and Function*. McGraw-Hill, Boston, Mass, pp 1104

Salter DM, Millward-Sadler SJ, Nuki G, Wright MO (2002) Differential responses of chondrocytes from normal and osteoarthritic human articular cartilage to mechanical stimulation. *Biorheology* 39:97-108

Sambrook PN, MacGregor AJ, Spector TD (1999) Genetic influences on cervical and lumbar disc degeneration: a magnetic resonance imaging study in twins. *Arthritis Rheum* 42:366-72

Sandell LJ, Aigner T (2001) Articular cartilage and changes in arthritis. An introduction: cell biology of osteoarthritis. *Arthritis Res* 3:107-13

Sandell LJ, Aigner T (2001) Articular cartilage and changes in arthritis. An introduction: cell biology of osteoarthritis. *Arthritis Res* 3:107-13

Sawcer S, Jones HB, Judge D, Visser F, Compston A, Goodfellow PN, Clayton D (1997) Empirical genomewide significance levels established by whole genome simulations. *Genet Epidemiol* 14:223-9

Sawcer SJ, Maranian M, Singlehurst S, Yeo T, Compston A, Daly MJ, De Jager PL, Gabriel S, Hafler DA, Ivinson AJ, Lander ES, Rioux JD, Walsh E, Gregory SG, Schmidt S, Pericak-Vance MA, Barcellos L, Hauser SL, Oksenberg JR, Kenealy SJ, Haines JL (2004) Enhancing linkage analysis of complex disorders: an evaluation of high-density genotyping. *Hum Mol Genet* 13:1943-9

Schaal TD, Maniatis T (1999) Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. *Mol Cell Biol* 19:261-73

Shifman S, Kuypers J, Kokoris M, Yakir B, Darvasi A (2003) Linkage disequilibrium patterns of the human genome across populations. *Hum Mol Genet* 12:771-776

Smith CW, Valcarcel J (2000) Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem Sci* 25:381-8

Smith DJ, Lusk AJ (2002) The allelic structure of common disease. *Hum Mol Genet* 11:2455-61

- Spector TD, Hart DJ, Byrne J, Harris PA, Dacre JE, Doyle DV (1993) Definition of osteoarthritis of the knee for epidemiological studies. *Ann Rheum Dis* 52:790-4
- Spector TD, Cicuttini F, Baker J, Loughlin J, Hart D (1996) Genetic influences on osteoarthritis in women: a twin study. *Bmj* 312:940-3
- Spector TD, Snieder H, Keen R, Lewis C, MacGregor A (1999) Interpreting the results of a segregation analysis of generalized radiographic osteoarthritis: comment on the article by Felson et al. *Arthritis Rheum* 42:1068-70
- Stankovich J, Sale MM, Cooley HM, Bahlo M, Reilly A, Dickinson JL, Jones G (2002) Investigation of chromosome 2q in osteoarthritis of the hand: no significant linkage in a Tasmanian population. *Ann Rheum Dis* 61:1081-4
- Stecher R (1941) Heredity in hypertrophic arthritis of the finger joints. *American Journal of the Medical Sciences* 201:801
- Stewart A, Black AJ (2000) Bone mineral density in osteoarthritis. *Curr Opin Rheumatol* 12:464-7
- Stumpf MP (2002) Haplotype diversity and the block structure of linkage disequilibrium. *Trends Genet* 18:226-8
- Stumpf MP, Goldstein DB (2003) Demography, recombination hotspot intensity, and the block structure of linkage disequilibrium. *Curr Biol* 13:1-8
- Tabor HK, Risch NJ, Myers RM (2002) Opinion: Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet* 3:391-7
- Tang BL (2001) ADAMTS: a novel family of extracellular matrix proteases. *Int J Biochem Cell Biol* 33:33-44
- Tavtigian SV, Simard J, Teng DH, Abtin V, Baumgard M, Beck A, Camp NJ, et al. (2001) A candidate prostate cancer susceptibility gene at chromosome 17p. *Nat Genet* 27:172-80
- Thelin A, Vingard E, Holmberg S (2004) Osteoarthritis of the hip joint and farm work. *Am J Ind Med* 45:202-9
- Todd JA, Acha-Orbea H, Bell JI, Chao N, Fronek Z, Jacob CO, McDermott M, Sinha AA, Timmerman L, Steinman L (1988) A molecular basis for MHC class II--associated autoimmunity. *Science* 240:1003-9
- Tortorella MD, Burn TC, Pratta MA, Abbaszade I, Hollis JM, Liu R, Rosenfeld SA, Copeland RA, Decicco CP, Wynn R, Rockwell A, Yang F, Duke JL, Solomon K, George H, Bruckner R, Nagase H, Itoh Y, Ellis DM, Ross H, Wiswall BH, Murphy K, Hillman MC, Jr., Hollis GF, Arner EC (1999) Purification and

- cloning of aggrecanase-1: a member of the ADAMTS family of proteins. *Science* 284:1664-6
- van Saase JL, Vandenbroucke JP, van Romunde LK, Valkenburg HA (1988) Osteoarthritis and obesity in the general population. A relationship calling for an explanation. *J Rheumatol* 15:1152-8
- van Saase JL, van Romunde LK, Cats A, Vandenbroucke JP, Valkenburg HA (1989) Epidemiology of osteoarthritis: Zoetermeer survey. Comparison of radiological osteoarthritis in a Dutch population with that in 10 other populations. *Ann Rheum Dis* 48:271-80
- Varilo T, Paunio T, Parker A, Perola M, Meyer J, Terwilliger JD, Peltonen L (2003) The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories. *Hum Mol Genet* 12:51-9
- Vink JM, Boomsma DI (2002) Gene finding strategies. *Biol Psychol* 61:53-71
- Wang HY, Paul WE, Keegan AD (1996) IL-4 function can be transferred to the IL-2 receptor by tyrosine containing sequences found in the IL-4 receptor alpha chain. *Immunity* 4:113-21
- Wang HY, Shelburne CP, Zamorano J, Kelly AE, Ryan JJ, Keegan AD (1999) Cutting edge: effects of an allergy-associated mutation in the human IL-4R alpha (Q576R) on human IL-4-induced signal transduction. *J Immunol* 162:4385-9
- Wang X, Moylan B, Leopold DA, Kim J, Rubenstein RC, Togias A, Proud D, Zeitlin PL, Cutting GR (2000) Mutation in the gene responsible for cystic fibrosis and predisposition to chronic rhinosinusitis in the general population. *Jama* 284:1814-9
- Wang N, Akey JM, Zhang K, Chakraborty R, Jin L (2002) Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination, and mutation. *Am J Hum Genet* 71:1227-34
- Watson PJ, Hall LD, Malcolm A, Tyler JA (1996) Degenerative joint disease in the guinea pig. Use of magnetic resonance imaging to monitor progression of bone pathology. *Arthritis Rheum* 39:1327-37
- Weeks DE, Lathrop GM (1995) Polygenic disease: methods for mapping complex disease traits. *Trends Genet* 11:513-9
- Weeks DE, Conley YP, Mah TS, Paul TO, Morse L, Ngo-Chang J, Dailey JP, Ferrell RE, Gorin MB (2000) A full genome scan for age-related maculopathy. *Hum Mol Genet* 9:1329-49

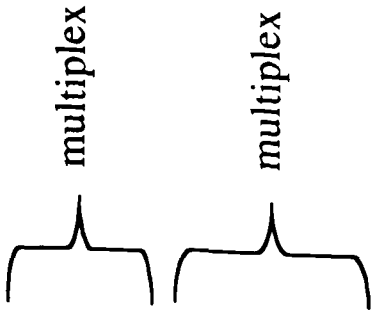
- Weiss KM, Terwilliger JD (2000) How many diseases does it take to map a gene with SNPs? *Nat Genet* 26:151-7
- Werner T (2002) Finding and decrypting of promoters contributes to the elucidation of gene function. *In Silico Biol* 2:249-55
- Westacott CI, Webb GR, Warnock MG, Sims JV, Elson CJ (1997) Alteration of cartilage metabolism by cells from osteoarthritic bone. *Arthritis Rheum* 40:1282-91
- Wiltshire S, Hattersley AT, Hitman GA, Walker M, Levy JC, Sampson M, O'Rahilly S, et al. (2001) A genomewide scan for loci predisposing to type 2 diabetes in a U.K. population (the Diabetes UK Warren 2 Repository): analysis of 573 pedigrees provides independent replication of a susceptibility locus on chromosome 1q. *Am J Hum Genet* 69:553-69
- Wiltshire S, Cardon LR, McCarthy MI (2002) Evaluating the results of genomewide linkage scans of complex traits by locus counting. *Am J Hum Genet* 71:1175-82
- Wright GD, Hughes AE, Regan M, Doherty M (1996) Association of two loci on chromosome 2q with nodal osteoarthritis. *Ann Rheum Dis* 55:317-9
- Xu J, Meyers DA, Ober C, Blumenthal MN, Mellen B, Barnes KC, King RA, Lester LA, Howard TD, Solway J, Langefeld CD, Beaty TH, Rich SS, Bleeker ER, Cox NJ (2001) Genomewide screen and identification of gene-gene interactions for asthma-susceptibility loci in three U.S. populations: collaborative study on the genetics of asthma. *Am J Hum Genet* 68:1437-46
- Zhang Y, Hannan MT, Chaisson CE, McAlindon TE, Evans SR, Aliabadi P, Levy D, Felson DT (2000) Bone mineral density and risk of incident and progressive radiographic knee osteoarthritis in women: the Framingham Study. *J Rheumatol* 27:1032-7
- Zhang K, Calabrese P, Nordborg M, Sun F (2002) Haplotype block structure and its applications to association studies: power and study designs. *Am J Hum Genet* 71:1386-94
- Zhang JL, Simeonowa I, Wang Y, Sebald W (2002) The high-affinity interaction of human IL-4 and the receptor alpha chain is constituted by two independent binding clusters. *J Mol Biol* 315:399-407
- Zhao JH, Curtis D, Sham PC (2000) Model-free analysis and permutation tests for allelic associations. *Hum Hered* 50:133-9
- Zhu Y, Oganessian A, Keene DR, Sandell LJ (1999) Type IIA procollagen containing the cysteine-rich amino propeptide is deposited in the extracellular matrix of

Bibliography

prechondrogenic tissue and binds to TGF-beta1 and BMP-2. J Cell Biol
144:1069-80

Appendix 1: Primer nomenclature and PCR conditions for microsatellite markers

Marker	Panel	Label ¹	Forward primer sequence (where available) ² 5' → 3'	Reverse primer sequence (where available) ² 5' → 3'	Primer conc (μM)	MgCl ₂ (mM)	Temp ³ (°C)
MD10							
D4S1575	1	Fam	ABI LMS v2	ABI LMS v2	0.13	2	58
D4S1592	1	Hex	ABI LMS v2	ABI LMS v2	0.20	2	58
D4S391	1	Hex	ABI LMS v2	ABI LMS v2	0.20	2	58
D16S3091	1	Ned	ABI LMS v2	ABI LMS v2	0.17	2	58
D4S1535	1	Ned	ABI LMS v2	ABI LMS v2	0.20	2	58
D4S414	1	Hex	ABI LMS v2	ABI LMS v2	0.20	2	58
D4S402	1	Fam	ABI LMS v2	ABI LMS v2	0.20	2	58
D16S503	1	Hex	ABI LMS v2	ABI LMS v2	0.20	2	58
D4S403	2	Fam	ABI LMS v2	ABI LMS v2	0.10	2	58
D4S415	2	Ned	ABI LMS v2	ABI LMS v2	0.15	2	58
D16S3046	2	Fam	ABI LMS v2	ABI LMS v2	0.08	2	58
D16S415	2	Fam	ABI LMS v2	ABI LMS v2	0.13	2	58
D16S423	2	Hex	ABI LMS v2	ABI LMS v2	0.20	2	58
D16S515	2	Fam	ABI LMS v2	ABI LMS v2	0.20	2	58
D16S404	2	Hex	ABI LMS v2	ABI LMS v2	0.20	2	58
D4S1534	3	Ned	ABI LMS v2	ABI LMS v2	0.05	2	58
D4S2935	3	Hex	ABI LMS v2	ABI LMS v2	0.10	2	58
D16S3136	3	Fam	ABI LMS v2	ABI LMS v2	0.05	2	58
D4S424	3	Hex	ABI LMS v2	ABI LMS v2	0.20	1	58
D4S413	3	Fam	ABI LMS v2	ABI LMS v2	0.20	2	58
D16S3075	3	Fam	ABI LMS v2	ABI LMS v2	0.20	3	58
D16S3068	3	Fam	ABI LMS v2	ABI LMS v2	0.20	1	58
D16S3103	3	Hex	ABI LMS v2	ABI LMS v2	0.20	1	58
D16S516	3	Ned	ABI LMS v2	ABI LMS v2	0.20	2	58



D4S2964	4	Hex	ABI LMS v2	ABI LMS v2	0.17	2	58
D4S412	4	Hex	ABI LMS v2	ABI LMS v2	0.22	2	58
D4S1572	4	Fam	ABI LMS v2	ABI LMS v2	0.25	2	58
D4S406	4	Fam	ABI LMS v2	ABI LMS v2	0.19	2	58
D4S1597	4	Ned	ABI LMS v2	ABI LMS v2	0.20	2	58
D4S405	4	Hex	ABI LMS v2	ABI LMS v2	0.20	2	58
D4S1539	4	Fam	ABI LMS v2	ABI LMS v2	0.05	2	58
D4S392	4	Fam	ABI LMS v2	ABI LMS v2	0.20	2	58
D4S426	4	Ned	ABI LMS v2	ABI LMS v2	0.20	2	58
D4S419	4	Ned	ABI LMS v2	ABI LMS v2	0.20	2	58
HD5							
D4S1615	1	Ned	ABI LMS v2.5	ABI LMS v2.5	0.20	2	58
D4S2460	1	Vic	ABI LMS v2.5	ABI LMS v2.5	0.20	2	60
D4S2920	1	Fam	ABI LMS v2.5	ABI LMS v2.5	0.20	2	58
D4S2924	1	Fam	ABI LMS v2.5	ABI LMS v2.5	0.20	2	58
D4S2952	1	Fam	ABI LMS v2.5	ABI LMS v2.5	0.20	2	58
D4S2986	1	Fam	TCA CTC TAC AGC ACT CCA GC	TTA CTT TGT TTA TTA GTT CTG ACC G	0.20	2	58
D4S3046	1	Vic	ABI LMS v2.5	ABI LMS v2.5	0.20	3	58
D4S1579	2	Ned	ABI LMS v2.5	ABI LMS v2.5	0.20	3	60
D4S1586	2	Ned	ABI LMS v2.5	ABI LMS v2.5	0.20	2	60
D4S1595	2	Fam	ABI LMS v2.5	ABI LMS v2.5	0.20	2	58
D4S2930	2	Ned	ABI LMS v2.5	ABI LMS v2.5	0.20	2	60
D4S2962	2	Vic	ABI LMS v2.5	ABI LMS v2.5	0.20	2	58
D4S3051	2	Vic	TTT GTT TGG TCT ATC AAA AGT C	CAA GAT GTG CAG TGG G	0.20	2	58
D4S2460	3	Vic	ABI LMS v2.5	ABI LMS v2.5	0.20	2	60
D4S3004	3	Fam	ABI LMS v2.5	ABI LMS v2.5	0.20	2	58
D4S3042	3	Fam	ABI LMS v2.5	ABI LMS v2.5	0.20	2	58
D4S398	3	Fam	ABI LMS v2.5	ABI LMS v2.5	0.20	1	58
D4S428	3	Ned	ABI LMS v2.5	ABI LMS v2.5	0.20	2	60
D4S3022	4	Ned	ABI LMS v2.5	ABI LMS v2.5	0.20	1	58
D4S2994	4	Fam	ABI LMS v2.5	ABI LMS v2.5	0.20	2	60

D4S2912	4	Fam	ABI LMS v2.5	ABI LMS v2.5	ABI LMS v2.5	0.20	2	58
D4S1587	4	Vic	ABI LMS v2.5	ABI LMS v2.5	ABI LMS v2.5	0.20	2	58
D4S2986	4	Fam	ABI LMS v2.5	ABI LMS v2.5	ABI LMS v2.5	0.20	2	58
D4S2971	4	Vic	ABI LMS v2.5	ABI LMS v2.5	ABI LMS v2.5	0.20	3	60
D16S3102	5	Fam	ABI LMS v2	ABI LMS v2	ABI LMS v2	0.20	2	60
D16S3100	5	Fam	ABI LMS v2	ABI LMS v2	ABI LMS v2	0.20	3	58
D16S3040	5	Hex	ABI LMS v2	ABI LMS v2	ABI LMS v2	0.20	2	58
D16S505	5	Ned	ABI LMS v2	ABI LMS v2	ABI LMS v2	0.20	2	60
D16S514	5	Fam	ABI LMS v2	ABI LMS v2	ABI LMS v2	0.20	2	58
D16S403	5	Hex	ABI LMS v2	ABI LMS v2	ABI LMS v2	0.20	2	58
D16S3066	5	Hex	ABI LMS v2	ABI LMS v2	ABI LMS v2	0.20	2	58
D16S3140	5	Hex	ABI LMS v2	ABI LMS v2	ABI LMS v2	0.20	1	58
D16S500	5	Ned	ABI LMS v2	ABI LMS v2	ABI LMS v2	0.20	2	58
D16S3041	5	Ned	ABI LMS v2	ABI LMS v2	ABI LMS v2	0.20	3	58
D16S3049	5	Hex	GCA ATG AAG GCA ACA AAG T	TTA AAA GAC CTG GGG GAA T	TTA AAA GAC CTG GGG GAA T	0.20	2	55
AFM								
D4S1583	1	Fam	GCC TAG GTA GGG CCA TAG AA	AAC CTA TTA ATT TTG GTA GGG AGA	AAC CTA TTA ATT TTG GTA GGG AGA	0.20	1	60
D4S1600	1	Fam	TGA AAA TGT TGT GGA GGC	AAC TAG TGT TTG AAG AAA CAG AAA T	AAC TAG TGT TTG AAG AAA CAG AAA T	0.20	3	58
D4S2987	1	Ned	TTA GGT TTA TAC TGA NCT TTG C	TGG AAC ATG TAC TAT CGT TCT TAA T	TGG AAC ATG TAC TAT CGT TCT TAA T	0.20	3	Touch down ⁴
D4S3253	1	Fam	CTC AGG GAA TCT TAG GGA GC	GTG ACA CTT GCC TCC CAT AT	GTG ACA CTT GCC TCC CAT AT	0.20	3	55
D4S1519	1	Ned	AGA TAT CCG ATA CTT GTG GG	TGA AGA ATG GAC TAA TAC ACC	TGA AGA ATG GAC TAA TAC ACC	0.20	2	58
D16S3018	2	Ned	GGA TAA ACA TAG AGC GAC AGT TC	AGA CAG AGT CCC AGG CAT T	AGA CAG AGT CCC AGG CAT T	0.20	2	55
D16S3073	2	Fam	CTT CCC AGA GGA TTC AAC AA	GCC TTT AGA TGG GGC TGA	GCC TTT AGA TGG GGC TGA	0.20	2	58
D16S3113	2	Ned	TCA GGC TCC CAA AGT GC	GAG GCC CCA CAT ACT TGT C	GAG GCC CCA CAT ACT TGT C	0.20	1	55

D16S3138	2	Fam	ATG TTA GCC AGG ACG G	TCT GCA TGT ATG AGT GTG TG	0.20	2	58
D16S409	2	Hex	TGA ATC TTA CAT CCC ATC CC	AGT CAG TCT GTC CAG AGG TG	0.20	1	58
D16S411	2	Hex	TCA TCT CCA AAG GAG TTT CT	GTG CAT GTG TTC GTA TCA AC	0.20	2	58
D16S420	2	Ned	ATT TCC TGA GGT CTA AAG CAC CC	TTA GGC CCA GTC CAC ACT CAA G	0.20	1	58
D16S518	2	Fam	GGC CTT TTG GCA GTC A	ACC TTG GCC TCC CAC C	0.20	1	55
D16S769	3	Fam	GAG CCT GAG CAG TAG TCA GC	GAC ATT GTC AGT GTC AGC CA	0.20	2	58
D16S3036	3	Hex	CCA AAA CAG CAT GGT ATT G	GGT GCC TTC ATG GAA AA	0.20	1	60
D16S3097	3	Fam	TGA TAG CCA AAG AAG TTG GT	CTT GTG GGT CAA TAT AGA TTA AAA A	0.20	2	55
D16S3129	3	Hex	AGC TGA ATA CAC ACA CTC CC	TCA CAG ATT ATT GAT GCT CTA TTT	0.20	1	58
D16S3132	3	Ned	ATG CTT TGT GGG CTG T	CCT TGG TTA ATG TAT TTG GA	0.20	2	55
D16S3133	3	Ned	TGA TCC ACC TGC TTT G	GTT TTG TGT CTG GCT TAG AA	0.20	2	58
D16S412	3	Ned	AGC CTG GTT GTT AGA GTG AG	AAC AGG TTG ATT ATT TGG CA	0.20	2	58
D16S501	3	Fam	CTG NCC CTA CTC TGC CAC AC	GGA ACC TGC CCA TGG AGT GT	0.20	1	58
D16S3031	4	Ned	ACT TAC CAC TGT GCC AGT TG	ATA CAT GGG TCC TTA AAC CG	0.20	2	55
D16S3056	4	Fam	CCA AGG GGT TCT ACA AGA TG	GTA ACA CTA TGA AAC CGA CAA TGA	0.20	3	55
D16S3057	4	Ned	CCT GTG TGT ATA ACT ATG TCA AAA T	GCC CTT GAA ACT AGG CAA TA	0.20	2	58
D16S3071	4	Hex	AGC TCT CTG ATG GGC AGT G	TGG AAG ATA GCC CCC AAA T	0.20	2	58

D16S3095	4	Hex	TCA GTT GGA AGA TGA GTT GG	TAT AGT TTG TGT CCC CCG AC	0.20	1	55
D16S499	4	Hex	GGT GGA CCC TAA TTG CAT AGG ATT G	TCT CAC AGT TCT GGA GGC TGG AAG	0.20	2	60
D16S504	4	Fam	AGC TTG TTC AGG GAA ACC	CAG GGA TGT AGG ACG TAG G	0.20	3	58

¹ Fluorescent dye labelling primer

² ABI LMSv2 = primers ordered from Applied Biosystems linkage mapping set version 2, ABI LMSv2.5 = primers ordered from

Applied Biosystems linkage mapping set version 2.5. No primer sequence available

³ Annealing temperature

⁴ 60 °C to 50 °C touchdown cycling conditions: 94 °C for 10 minutes, then 20 cycles of 94 °C for 45 seconds, 60 °C for 45 seconds, -5 °C/cycle, 72 °C for 45 seconds, finally 20 cycles of 94 °C for 45 seconds, 50 °C for 45 seconds, 72 °C for 45 seconds.

Code

N=A+C+G+T

Appendix 2: Family NPL scores for the regions showing maximum evidence for linkage on chromosomes 4q, 16p and 16q in the 146 female-THR families

Family	Chromosome 4q	Chromosome 16p	Chromosome 16q
1	-0.816497	-0.279116	0.816222
2	-1.413983	-0.000984	-0.001007
3	0.000146	0.002201	0
4	1.413746	-0.019447	-0.015239
5	-1.414214	-0.002206	-0.002059
6	2.75484	0.995442	1.154807
7	1.273337	-0.003124	-1.388869
8	-0.015593	0.028564	-1.404111
9	-0.326477	1.410179	1.367691
10	1.413118	1.412691	-0.72874
11	-0.000115	1.414214	0.451987
12	-0.003355	-1.414214	0.20443
13	-0.004585	-1.372663	-0.202107
14	0	-0.001855	-1.411729
15	-1.414	-1.410718	0
16	1.414214	-0.007163	-1.413359
17	-0.816497	2.444427	0.793415
18	1.366782	-0.023137	-0.012854
19	-0.009051	-0.000636	1.40878
20	0.000498	-0.001528	-1.413738
21	-0.566691	-0.011286	-0.024083
22	1.378856	-0.001085	-1.402563
23	0	1.413308	1.412606
24	-0.00057	-1.41126	0.001308
25	-0.025576	1.391752	1.412354
26	-0.006816	-1.406378	-1.413264
27	0.816497	0.816497	1.193441
28	-0.000361	0.000466	1.411094
29	1.409296	-1.403548	-0.672247
30	0.816323	0.816497	2.331163
31	1.216069	-0.015889	1.41054
32	0.000184	1.412125	0
33	0.816497	-0.816497	0.798532
34	1.414125	-1.343254	-0.132955
35	1.41396	-0.015477	-1.411687
36	1.413694	1.408734	-0.056592
37	-0.000326	1.32713	-1.377781
38	-0.000115	0	-1.41244
39	0.817742	2.444507	0.81494
40	-0.005229	1.410791	1.410478
41	0	1.414214	1.413262
42	0.000009	1.412691	0.003229
43	0	0.001534	-0.001301
44	0.007837	-0.016398	-0.023056
45	-0.044871	0.850032	-0.048363

46	-0.000144	-0.01117	-0.003249
47	-0.000116	-0.000423	0
48	1.354891	1.413784	1.411297
49	-0.002404	-0.00312	1.41359
50	-0.006898	1.411673	1.413039
51	-0.468597	1.410439	1.411872
52	0	1.413224	1.411425
53	-0.677802	-0.769923	-0.414874
54	0	0	-0.022003
55	1.413941	1.410632	1.379799
56	1.413791	-0.029237	-1.22075
57	-0.001192	-0.146133	0.000154
58	-0.003731	1.408616	0.424039
59	-0.000101	-0.001199	1.412946
60	0.000388	-1.414214	-1.401632
61	1.406143	0	0
62	0.000088	-1.408622	0
63	-0.954738	-0.009286	-1.383407
64	-0.206134	-0.009302	-1.14346
65	1.414214	-0.000466	1.27111
66	-1.105635	-0.027673	1.149847
67	-1.413867	-1.412417	-0.000002
68	1.413671	0	1.413739
69	1.381026	-0.002218	1.409612
70	1.414214	0.221898	-0.000449
71	-0.007304	0.333132	0.260515
72	0.002809	1.412939	0.028654
73	1.413703	-0.00005	0
74	-1.414214	1.413882	-0.004146
75	1.412157	1.400471	0.06348
76	1.329566	-0.004804	-0.03001
77	1.413874	1.188719	-0.015967
78	1.412211	1.258191	-0.006918
79	-0.002319	1.214752	1.409295
80	-1.413857	0.948914	1.367943
81	1.339451	-0.845089	0.004328
82	-1.412125	-1.38249	-1.389481
83	1.413601	1.412277	1.410119
84	1.413463	-0.068727	-1.077608
85	-0.034688	-0.010698	-0.487545
86	-1.414214	-0.003911	-1.355976
87	-0.016661	1.413381	1.410382
88	-0.018718	-1.043633	-0.003588
89	-1.323621	1.413538	1.281913
90	1.414214	0.009407	-0.003944
91	-0.033828	0.077102	1.391697
92	-0.01011	-0.008898	-1.099946
93	1.413964	1.401862	1.40991
94	-0.005887	-1.400982	-1.399436
95	-0.319863	-0.004396	-0.028813
96	1.413834	-0.00716	-0.065518

97	1.413991	-0.464629	-0.009639
98	1.413849	-0.014942	1.398806
99	0.811591	2.448685	0.815165
100	2.449097	-0.812523	-0.787984
101	-0.073427	1.412874	0.061286
102	-0.026118	1.411877	-0.001464
103	-0.008325	-0.025277	-0.506558
104	-1.41396	-0.002569	-0.719537
105	1.414125	-0.027153	-0.866904
106	-0.017333	-1.402392	-0.026954
107	-0.004272	-0.010576	-1.353786
108	-1.412681	1.407823	-0.959677
109	1.413709	1.406465	1.409323
110	-0.014168	-1.363571	-0.386589
111	1.413713	-1.060612	-0.00356
112	-0.002271	1.253531	0.466382
113	1.411393	-0.696926	1.394926
114	0.00341	-0.000473	1.411591
115	-0.193626	-0.179657	1.938829
116	-0.049477	1.167784	0.004861
117	1.331228	-1.414214	0.012512
118	-0.006179	-0.299636	1.298434
119	-0.009109	-1.400855	-0.927197
120	-1.15988	-0.022682	1.358076
121	-0.001998	-1.347206	-0.009363
122	-0.004889	-0.718198	-0.104189
123	-0.925272	-0.004089	-1.413264
124	-0.002454	-1.406119	1.401729
125	-0.369054	1.172346	-0.515315
126	-1.413065	1.410351	1.410543
127	0.000536	-1.069648	-0.282463
128	-0.001848	-1.408211	1.064482
129	-0.000409	1.412856	-0.923373
130	-0.075188	-0.507332	1.126089
131	1.343888	-0.88023	-0.441622
132	1.412678	-1.412977	-1.407106
133	0.000395	0	0.004588
134	-0.002588	-0.006964	-1.293561
135	1.414098	-0.010923	0.026638
136	0.591158	0.794542	0.811715
137	1.413123	0.016973	1.198367
138	-0.100995	1.408913	-0.004947
139	1.414214	-0.006528	1.293709
140	-0.005219	-0.016636	0.251235
141	-1.414214	-0.941499	-0.128688
142	-1.413284	1.41187	0.200054
143	1.41322	0.007352	0.967684
144	-0.004361	-1.282783	0.097399
145	-0.010394	0.007942	0.002558
146	1.411112	1.408627	-1.175497

Appendix 3: Family NPL scores for the regions showing maximum evidence for linkage on chromosomes 16p and 16q in the 218 female families

Family	Chromosome 16p	Chromosome 16q
1	-0.816462	0.816497
2	-0.000475	-0.001259
3	-0.816497	0.816497
4	-0.015765	-0.015907
5	-0.002502	0.003383
6	0.9954	1.153958
7	-1.295592	-1.414214
8	-0.001466	-1.385155
9	-0.00677	-1.404998
10	1.413808	1.357126
11	1.41395	-0.702246
12	0	0
13	-0.014181	1.252339
14	-0.002694	1.409678
15	-0.176004	0.998138
16	-1.413905	0.145944
17	1.414111	0.004062
18	-1.414123	0.026571
19	-0.005185	-1.411712
20	-1.414214	0
21	-0.816497	-0.816497
22	2.44941	0.799503
23	-0.040904	-0.0139
24	-0.000133	1.409595
25	-0.00167	1.403238
26	0	-1.414214
27	1.100339	-0.024943
28	-0.001102	-1.39982
29	0.816335	2.447289
30	1.403782	0.821404
31	-1.413848	-0.466357
32	1.410845	1.403521
33	-0.127661	-0.036858
34	-1.414214	0.001635
35	1.412785	1.412513
36	1.40507	-0.001306
37	-1.414214	-1.414214
38	0.816497	1.287919
39	0	1.411584
40	-1.414214	-0.760882
41	0.816497	2.30922

42	-0.002698	1.411211
43	1.413859	0
44	-0.816462	0.80086
45	-1.414214	-0.160068
46	-0.000512	-1.411666
47	0.111868	1.308168
48	1.413965	1.413155
49	1.413861	-1.320331
50	1.389345	-0.052697
51	1.412615	-1.414214
52	0	-1.41259
53	1.372456	-0.005756
54	2.437877	0.815234
55	1.414214	1.410023
56	1.414042	1.413711
57	1.394425	1.404272
58	0.047996	1.413116
59	0	1.414214
60	1.414154	0.004241
61	0	-0.001626
62	-0.064422	-0.024249
63	-0.002376	-0.035547
64	-0.070695	-0.004062
65	0	0
66	1.414155	1.411361
67	-0.002351	1.414094
68	-0.000477	-0.016684
69	1.414214	1.413585
70	1.410601	1.412223
71	1.413977	1.411616
72	-0.429862	-0.505087
73	0	-0.022845
74	1.413992	1.371887
75	-0.034024	-1.241059
76	-1.411058	-0.010068
77	0.319205	-1.016691
78	-1.408719	-1.414214
79	-1.414214	0.649366
80	0	0.000856
81	1.414115	1.413097
82	-1.414214	0
83	1.413793	0.881355
84	0	1.413571
85	0	-1.414214
86	-0.844241	-1.400951
87	2.444375	0.777929
88	0	0

89	-1.414214	0
90	-0.003588	-1.377462
91	-0.001664	-1.067884
92	-0.000099	1.246797
93	-0.022967	1.09471
94	-0.816497	-0.816497
95	-0.000222	1.414214
96	0.816497	0.827542
97	1.414214	-0.197882
98	-0.000593	1.410001
99	0.23929	-0.000436
100	-0.001943	-1.414214
101	-0.01995	-0.00374
102	2.449164	0.852074
103	-0.002627	1.41375
104	0	0
105	1.414195	-0.004677
106	0	1.412251
107	1.413772	0.067003
108	0.338831	-0.030573
109	-0.000283	-0.016271
110	-1.284467	-1.414214
111	-0.002174	-0.044141
112	1.411301	-0.007637
113	-0.135249	1.38667
114	-0.014136	-0.17329
115	1.412988	1.409957
116	-1.414214	-1.403071
117	-0.001427	1.356569
118	0.289663	-0.814074
119	-1.413915	-1.383349
120	1.414023	1.410661
121	0.932167	-0.013634
122	-0.07339	-1.131601
123	-0.001735	-0.54941
124	0.016412	-1.352392
125	-0.003264	1.399203
126	1.413883	1.410865
127	-0.982374	-0.003017
128	1.413665	1.263001
129	1.413727	1.413005
130	-0.73594	1.387616
131	-0.000946	-0.003574
132	-0.00186	1.389294
133	-0.000543	-1.122479
134	0.000955	-0.004443
135	1.414087	1.410412

136	-1.414214	-1.395655
137	-0.104488	-0.04219
138	-0.000702	-1.414214
139	-0.493436	-0.406641
140	-0.001687	-0.048539
141	-0.008603	-1.410394
142	-0.071117	-0.01081
143	-1.273065	0.18001
144	1.25602	1.407458
145	2.449385	0.815417
146	-0.816497	-0.777258
147	-1.414214	1.41195
148	1.413994	0.06233
149	1.41407	-0.002101
150	-0.001473	-0.063266
151	-1.414214	-1.316427
152	-0.00299	-0.560674
153	-0.001579	-0.633285
154	-0.019616	-0.893676
155	-1.409583	-0.034431
156	-0.056351	-1.360663
157	1.414032	-0.899291
158	0.814847	-0.816497
159	-0.00354	1.293346
160	1.40767	1.410431
161	-1.414214	-0.363029
162	-1.414214	-0.003489
163	1.408108	0.507438
164	-0.267581	1.391529
165	-0.013799	1.411914
166	-0.000144	1.411877
167	-0.179612	1.950683
168	-0.020657	1.411701
169	-0.001216	-0.004253
170	-1.408067	0.016145
171	0.204803	1.411243
172	-0.841186	1.271789
173	-1.413737	1.248464
174	-1.414214	-0.893881
175	-0.206844	-1.394637
176	-0.018155	1.367421
177	-1.366796	-0.008911
178	-0.465487	-1.192825
179	-1.414214	-1.414214
180	-1.413977	-0.110414
181	-0.002562	-1.414214
182	-1.414214	1.403453

183	-0.023158	-0.00794
184	1.002476	-0.475206
185	1.414018	1.411093
186	-1.413506	0
187	-1.414214	0.998665
188	1.414145	-0.808561
189	-1.414214	1.0716
190	-1.413915	-0.465857
191	1.414028	-1.410394
192	-0.816497	-0.816497
193	1.391684	0.00744
194	-1.413126	-0.013479
195	-0.000066	-1.409672
196	-0.014528	-1.311752
197	-0.964489	0.679819
198	0.816772	0.816497
199	-0.007062	0.031182
200	-1.413848	-0.137442
201	0.813229	0.810864
202	-0.003143	-1.400235
203	-0.013511	1.266473
204	1.414139	-0.006774
205	-0.000805	1.339544
206	1.413685	1.405385
207	-0.009343	0.250495
208	-0.020518	-1.414214
209	0.001595	1.042144
210	1.390928	0.063686
211	-0.42169	-0.155695
212	1.414073	-0.006232
213	1.414046	0.038765
214	1.414096	1.414214
215	-0.00337	1.20778
216	-1.239444	-0.080463
217	-0.002616	0.007352
218	1.413921	-1.106335

Appendix 4: Primer nomenclature and PCR conditions for ADAMTS3 and IL4R DNA sequencing

Gene	Exon ¹	For primer sequence ² 5' → 3'	Rev primer sequence ³ 5' → 3'	MgCl ₂ (mM)	Temp (°C) ⁴	Extra ⁵	Sequencing primer(s) ⁶	Internal primer sequence ⁷ 5' → 3'
ADAMTS3	1	AGA GGG GAG GAA CTT TTC G	GAA GCA AAG CAC AGA AG	2	52		2 x Internal	Int Rev: GGT GGG TTT GCT CTG Int For: TCC ACT TAC TTG TCC
ADAMTS3	2	CCA GTG GGA AAC ATT TCA G	GCT CTT GGA AAC CAT TAA TGG	1	58		For/Rev	
ADAMTS3	3a	TTT CCA CCT CCA CCT TCC	GAG AGT ATG AGC TGG TG	3	58		For/ Internal	Int For: GGA ATG TCC ACG ATG
ADAMTS3	3b	AAG ACA CGT CCC TCG CTG AC	ATT AAG GGC ATG CCT TG	2	58		For	
ADAMTS3	4	CAT GGC AAT CAC ATT CTC	CAC AAG TCA TAT TCT TGA C	3	58		Rev	
ADAMTS3	5	CAT GGA AGT TCT CAC ACA G	ATA GGC CCC TGA GAA GAT TC	2	58		Rev	
ADAMTS3	6	GCC TTA AGT TAG ATG TAG C	ACT CAC CTA TTT GTG GTG	2	58		For	
ADAMTS3	7	AGA GCA GAA CAT TTG AAC AG	CTT TAA TCC CAA ATT CAG C	2	60		For	
ADAMTS3	8	TTG TAT GCT GGC ATG CAT TCT GC	TAC TTT TTT CCC CCT AGG	1	58		For	
ADAMTS3	9	TTG CTA TGA TCA AGT CC	TAC GTT ATT TCC CAG GGC	2	60		For	
ADAMTS3	10	TAC TTG CAC CAG TGA GTC TG	CTG GGA TGA AAT TCT AAA G	2	58		Rev	
ADAMTS3	11	ACA GTG CAG CTG TGT TTA CC	CAG GAC AGA AGA AAC TG	2	58		For	
ADAMTS3	12	GCC CTG TGT TTT	GAA AGC CCT TCA	2	58		Rev	

		ACA TGC AC	ACT AGA TG					
ADAMTS3	13	ACC ACA TAG GGA TCA AGA G	GAA ACT TTG GCT TTG TGC C	2	58		Rev	
ADAMTS3	14	GGC TAC TAA AAT ACC CGA	CAG GAA ATC ACA GTG AAA CC	1	52		Rev	
ADAMTS3	15	TAT AGA GAA AAT GCT GC	TGG GCC ATG ACT GAA CAA AG	3	60		Rev	
ADAMTS3	16	CGT ATG CAC ATA TAT GTG	TTC TGC ACG CAG AAG CAG TAA G	3	52	5% DMSO	Rev	
ADAMTS3	17	GTA CAC TTA ATG TTC AAG G	GGT CAT TGA TTT GTA GAA AGG	2	60		For	
ADAMTS3	18	GAT TGT GGC CTC GAT TAG AG	ATA AGA GCA ATA TTC TG	2	52	5% DMSO	For	
ADAMTS3	19	CTA TAT AGA GCT TTG ACT G	ATG GAT GAA TAA ATG TAT G	3	58		Rev	
ADAMTS3	20	CAC ACA AAG CCT AAC TGA TG	GAG AGC AGA AAA GTC TAG GC	2	58		For	
ADAMTS3	21	TCC TTT GGT TAT TTA TCA G	AAC CCA CTC CAG ATC TC	2	58		For	
ADAMTS3	22 a	CAA GCA ACT ACA AGT AGC	TTA ACC TCT CAA AGC ACA G	2	58		For/Rev	
ADAMTS3	22 b	AGC AGC AAT TCT TTA GAG G	TGG TAA ACT AGA GAG AC	3	58		For	
ADAMTS3	22 c	GTT CAG TTT TAG CTT TTG AG	CAG TTA TAC TCA ATG TAC TG	2	58		For/Rev	
ADAMTS3	22 d	CCC AAA GCT CTA CTA TAC TC	GTG CCT GTC GTT GAG TAT TG	1	52		Rev	
ADAMTS3	22 e	GTC AGT AGC TGT AAT TC	CTA GTT GCT TTC ATC CTC C	2	52		For	

ADAMTS3	22 f	CCA ACA CAA TCT TAG CAA C	CCA ACA CAA TCT	GAT GGT GCT AAT TTA CGC C	2	58		For	
ADAMTS3	22 g	AAG CTG AAC TGA GGT GGA C	AAG CTG AAC TGA GGT GGA C	CCG TTA GTG TCA GAT TGC AC	2	58		For	
IL4R	2	TGG AAA AGG GAG AGC	TGG AAA AGG GAG AGC	CTA CAG GCA AGC CAC	2	58		For	
IL4R	3	ATC TCC CAG GAC AGT TG	ATC TCC CAG GAC AGT TG	AGA GAG AGG AAG CAG G	2	60		For	
IL4R	4	ATA AAT GGC CAG GGC	ATA AAT GGC CAG GGC	ACT GCA GTG TAA TCT CC	3	58		For	
IL4R	5	ACT TTG CCT GCA CTG TG	ACT TTG CCT GCA CTG TG	AAC CAT ACT CAG CAT CC	2	55		For	
IL4R	6	TAC AGG TGA CCA GCC TAA CC	TAC AGG TGA CCA GCC TAA CC	GCA GCT GTG GGA ACA CAC C	1	58		Rev	
IL4R	7	CTC TGA ATC TGT GTG GT	CTC TGA ATC TGT GTG GT	GAA GCA AAG GGA CTT GC	2	58		For	
IL4R	8	GGT GGC TCT TAA ACA TGG	GGT GGC TCT TAA ACA TGG	ATT GCC ATT AGA GGC ATC C	3	58		Rev	
IL4R	9	TGG CAC TGA GCT TTG	TGG CAC TGA GCT TTG	CAG CTC TAC TCT CAA GG	2	58		For	
IL4R	10	CCA TGT CAC CAG GTT AG	CCA TGT CAC CAG GTT AG	TTT GCC ATT CCA GAA GCC	2	58		Rev	
IL4R	11	GTA AGT GAC CTG TTG G	GTA AGT GAC CTG TTG G	GAA GGC TAA TGT CAC AGG	2	58		Rev	
IL4R	12 a	AAC GTT AGA AGG CAT G	AAC GTT AGA AGG CAT G	CAC TTG GGA ACT CAT C	2	58		For/Rev/ Internal	Int For: TGT CTT ACC AAG CTC
IL4R	12 b	CTT TTG CCA GCA GGA CAT G	CTT TTG CCA GCA GGA CAT G	CCA GGA ATG AGG TCT TGG AA	2	58		For	
IL4R	12 c	GAA ATG TCC TCC AGC ATG G	GAA ATG TCC TCC AGC ATG G	AGG TGG CAG GTA AGG	3	60	5% DMSO	For	

IL4R	12 d	AGG AGC AGG CCA CAG AC	TCT CTG CAG CCC AGT GTT AG	1	55		For	
IL4R	12 e	TGG GAA ATG CCA CCT C	CCT AAG GAA GCT CTG	2	55		For	
IL4R	12 f	TGG GAA ATC GAT GAG	CTG GAG GTA TGT TTG G	3	58		For	

¹ a,b,c etc denote more than one PCR product used to cover exon

² All primers had M13 forward sequence on 5’ end of primer, sequence TGT AAA ACG ACG GCC AGT

³ All primers had M13 reverse sequence on 5’ end of primer, sequence CAG GAA ACA GCT ATG ACC

⁴ Annealing temperature

⁵ Three PCR reactions required 5% dimethylsulphoxide (DMSO) to facilitate amplification

⁶ Direction of primer(s) used for sequencing reaction. For = M13 Forward sequencing primer, Rev = M13 Reverse sequencing primer, Int = internal primer

⁷ Sequence of internal primer if used

Appendix 5: Primer nomenclature and PCR conditions for ADAMTS3 and IL4R restriction digests

Gene	SNP	For primer sequence ¹ 5' → 3'	Rev primer sequence ¹ 5' → 3'	MgCl ₂ (mM)	Temp (°C) ²
ADAMTS3	Intron1indel	GAA ATG TCC AAC TGC AGG	AGG TGA TAA GAT GAA ACC CAG TAA CTT TCT TTT TCT TGC TTT CCC CCG CC	2	52
ADAMTS3	Intron2A/T	GCA GAA AGA GTA TGG GAG	AAA TGC ATT TGA GAA ATG ATT TTT GTT CTC AAA ATT TAA TAT TTA TCA GC	2	58
ADAMTS3	R138K	CAT AAC AAT GAA GCC TCC	CGA TCC CAT TAA CAA CCA TCA ACC AGG AAG TGC TAC GTA TAG AAT CCC GA	3	58
ADAMTS3	Intron12A/C	TAA CCA TCC ACA CAG CAG G	AGA GAA GGG TGC TTT GG	2	60
ADAMTS3	Intron15C/T	CAG CAG CTA AAT TTT GAT TAT GTT TAT AGA GAA AAT GCT GCA TTT CTT AA	GAC ATC TTC TCT CTC CAG	2	58
ADAMTS3	3'UTRG/A	CGA TTA TTA CAG TAC ATT GAG TAT AAC TGT TTG TAT CTG TCA ACA ACG	CAG ATA TCA GCT TAC C	2	52
IL4R	C-3223T	ACT CTC AGT ACT ATC GGC TC	GTA CTC TTC TCC CAA CTT TCC	3	58
IL4R	T-1914C	AGC TCA CGA CTT GGC ATA TGG	TTG TGA CAA GCG CTG ATG ACG	2	58
IL4R	I50V	TAA GAG GCT GTG GCC AGC AAG	ATC CAT GAG CAG GTG GCA CAC GCA CCC CGC GCC TCC GTT GTT CTC AGA GA	3	58
IL4R	N142N	TGG TGG CTG TGG TTT GTA GG	CAG AGG TCA CAG AAA AAC GTC AAA GCA TGC CCA CTC ACA TCT GCC CGG TC	3	58
IL4R	E375A	AAG ATC CTC ACA AGG CTG C	CAC TTG GGA ACT CAT CCC AG	2	58

IL4R	C406R	CGA TGT GTG GAG TTG TTT GAG G	AAC TCA TCC CAG GGC ATG TGA GCA CTC GTA CTT CCC GAA GGT GGA AGA <u>ACG C</u>	2	58
IL4R	S411L	AAG GGA GCT TCT GTG	GCC CTG CAC TTG GGA ACT CAT CCC AGG GCA TGT GAG CAC TCG TAC TTC <u>TC</u>	3	55
IL4R	S478P	AGA CGC CCC TCG TCA TCG CAG GCA ACC CTG CTT ACC GCA GCT TCA GCG AC	CCA TGC TGG AGG ACA TTT CG	2	58
IL4R	Q551R	TCC AGC ATG GGG CAG CTG CAG CCC CCG TCT CGG CCC CCA CCA GTG GCT A <u>CC</u>	CCA GGA ATG AGG TCT TGG AA	2	58
IL4R	S727A	GTA GAG GAC ATG CCA AAG C	AGG ATG GAA GGA TGA TGA GG	1	58

¹ The sequence underlined is an artificial change introduced to generate a restriction site at the SNP

² Annealing temperature

Appendix 6: Enzymes and conditions used for ADAMTS3 and IL4R restriction digests

Gene	SNP	Enzyme	N° of Units/ml	N° of units/ reaction	NEB ¹ buffer	Digestion temp °C	Inactivation temp °C
ADAMTS3	Intron1indel	BglI	10,000	4	3	37	65
ADAMTS3	Intron2A/T	BsoBI	10,000	2	2	37	None
ADAMTS3	R138K	DraI	20,000	2	4	37	65
ADAMTS3	Intron12A/C	DraI	20,000	4	4	37	65
ADAMTS3	Intron15C/T	AflII	20,000	2	2	37	65
ADAMTS3	3'UTRG/A	BstUI	10,000	2	2	60	None
IL4R	C-3223T	DrdI	5,000	2	4	37	65
IL4R	T-1914C	MspI	20,000	10	2	37	65
IL4R	I50V	BsmAI	5,000	5	3	55	80
IL4R	N142N	BsiEI	10,000	2	2	60	80
IL4R	E375A	AciI	10,000	6	3	37	65
IL4R	C406R	MluI	10,000	5	3	37	65
IL4R	S411L	TaqI	20,000	4	3	65	80
IL4R	S478P	HinfI	10,000	5	2	37	80
IL4R	Q551R	MspI	20,000	5	2	37	65
IL4R	S727A	AvaII	10,000	4	4	37	65

¹New England Biolabs

1X NEBuffer 2:

- 10 mM Tris-HCl
- 50 mM NaCl
- 10 mM MgCl₂
- 1 mM dithiothreitol
- pH 7.9 @ 25°C

1X NEBuffer 3:

- 50 mM Tris-HCl
- 100 mM NaCl
- 10 mM MgCl₂
- 1 mM dithiothreitol
- pH 7.9 @ 25°C

1X NEBuffer 4:

- 20 mM Tris-acetate
- 50 mM potassium acetate
- 10 mM magnesium acetate
- 1 mM dithiothreitol
- pH 7.9 @ 25°C

Appendix 7

Summary of ADAMTS3 sequencing. Sequence covered, positions of primers and SNP locations are shown in relation to the genomic sequence.

Key

aaaa	Forward primer
aaaa	Reverse primer
aaaa	Internal primer
aaaaa	3'UTR
aaaaa	Exon
aaaaa	Sequence covered by sequencing
aaaaa	Sequence not covered by sequencing
a	Common SNP (minor allele frequency ≥ 0.05)
a	Rare SNP (minor allele frequency † 0.05)
CAT	Start codon
TCA	Termination codon

Gene is on antisense strand.

Exon 22

caagcaactacaagtagc aaataaattgatgccatttactgtgtaggaggaaatacacgttccattacccc
atgatgaacctaatgataaaatctagatagagccagatgtatatttggtaacattttattggtacgtgctctcag
tacaacaaacagcatcagtagtgtacactttgataaaaaggaatttttagcttagtagaaaagaaagccca
aaggtcagaagtataatgaatatgtacatctttatggaaactgtttgtgtgaccatctttatctcccctgtggatg
agatgtatgcacacacaagtaaaactagagagaattttattctatgggaatttttaggaaaaaaaagactta
aggcttagggattgtttactatctggaatttgccaaatttgacctcaaagttctgcacaaagttagcagcaattc
tttagagggtatagagtcaatagtttaagctgttagttacaacgataaaatgcaaaagttctgcacattctgtg
ctttgagagggttaattcatctacatttttagttgtataatatatttttagacagctcaggtactgacctcaaaatcatt
atgctgatatcactatctcctctcatgtacatgtaattcatgatgcacttatctcttcaaaaacttttaaaaatgttcc
taattcacaatttcattgaacagctagctgagtcacatacaatttaattgttccctgaagggaacaagtcacaa
cttaatatcaactactaaaaatgcaagatgtttgcataacatcaaatttctcagtttttagcttttaggttgctgta
gtagtttgcttaaagggtctctatgcacagaacaaagcatgcactcttaattatgtctctctagttaccaaaatat
gtaattgatattagtgtaaatattagtacataaaaatactatcaagaaaaaaaatcaactttaataaatatctt
cagttctgaattctatgtacaagtactttacatctactccctggagagagcatttctggcttgcgtgtgagcaagg
gactactgacataaattattagcacattattgcagtcagatgcattgtggtgtgaacattgaacatgaaccat
atggaaaataaataatctcaaccatcgacatatgcttctgaacctgccttcaaagcccaaaagctctactata
ctc aaatatcaattctagactgatgatattttctcagctaccaacatacaaaataagttatcccaaaacttttcctt
tgcgattatta cagtacattgagtataactgtttgtatctgtcaacaatgcgaaaattaaaaaagacacacac
acacaaaaacaaaggacacctcaacggtgcctactgaggtcctggaaagagggttcagtttgcgtgtgcac
gagttcattatttactattcggtgatccctgtgaccacaagtgtactaagacaccaatccttaaaattataagtata
gaaggcaataggtaagctgatatctgtatatctatagctttgtgtttcttttcttttctataattacgtatcacattgtt
gattcaacaaaaccctttatattgtcagtagctgttaattc taacttctctttgttaatactctgcttgaataaacct
atagaatgtcaaacaaggcaatactcaacgacaggcacctagaacatccagaaatgtaatatagtaaac
aatcatttgagaacacatctcttaaaataatactgattttgcaatatacagtatttacacaacagctgcaaattt

Exon 22 continued

gctcatacaatatatttctaatatagataaataggaatacaagctctaagaattctttgttccatatccaaaatttctgt
taaagtgtaaaccattggtatctcttttaatttttactgtaaccttttctaaaacaaaacaaaagtcagctgatatgc
acaattgggtccagtccttttctgcttcagagagcgacccaacacaattcttagcaacatatatttttcttttgtaataat
ctttggtgatttcttccaatgaattctggtaaatgagtcfaatgcagaacaaaaaggaggatgaaagcaactag
aaagtgagccagcactttctgttcttccaattacagataaaatgagctgacgatctatagagatttccactttaa
acaagcatatgcaccatgggaagagagagagggtgtccagggtttcctctggtttctagccttttggttcaacttct
CActtttctaagggtgatgatctgtcgacgtctgtgtcaatgatctttccatctttctttgaggttcttgccctgag
aagaagcacctattgaatcactggctgcaaagaaggaagcagcagccattgtgaagctgaactgaggtg
gacccctcttgggtgggtggggaggatggtacggtgaccagtctcacagtcttacttctgcttgctgagcactcc
tctggcgtaaatagcaccatcagggttaactgtttggcctgaaagcagcatatgcatttgacctcccactgaa
gagatgctactcaaagacatcttcttgcaggggtctctgaatgataaggaaccaaaagatgtaggcatcact
agagatctagggagggtcactaggggttagagatgacatcatcatgagtttcagcagcttctagaaggatggt
ggtggcaggggtgctactgctgtgctgcaggactcacaacataactgttataacctggatggagcagtatc
gtgccaacacttccatttgacagaatatggactgtctccaaacatggttcattctgaaaaataaaaagaaa
ataaactaacactagcatctaaacataaaaataaaaggaggaaaaaaattaaaatttttaatgaatgtaaaa
aqatttaaagtgaactctgacactaacgg

Exon 21

[illegible]

Exon 20

cacacaaagcctaactgatgtttaagccaaattaaaatatctacacatgcttacacacacactttacttatgca
tgcacggaattcacacacctacctcactccagggctctgtttccactgtgcagggcagggcactctgttaca
gggccggcggctctcgggacggtcacccatgcagtatttgctgtgcacagagcggttggtgccatcaagga
gtggctgaaggcagcgtagtggaagctgatagccagaacttccacaggtttggtgcagtgttccattc
ttctgctaccagctgtgggtgcggggatttaatttgcaattatcactggtatgtatagtgtacacatttctacct
qcaaaagccttagacttttctgctctc

Exon 19

Ctatatagagctttgactgaaaataaaatcaaaactaaaaagagctaaaggcagtgaaagtccctgatttga
cctccagatatgatagttaaaatagatgcttacagtggatgtgtacactcttgaatattgcacattcgtctaatag
gtttcggcttttgttggcctcacagaagctgcatggaccattttattatcacttttcctacggcatccatatttagt
gtactggaaacctggaaaaggaaataaagttctttggatttaatcacttcttcatgctgatagttacttgattatc
acttactcattcatacattattcatccat

Exon 18

Gattgtggcctcgattagagacagcaataaaggtagaagcaagattggttttatatgcattacattttaatata
ataaatgtgttcaatggttacctccaccacagggtttgaacactgagaccagctcttcaaagcccactcaa
aagtatctaattcttcttgatgacattgttgctgttgattgtaggtacagagcttcatggatgatgtacttatatgt
caggctagagcgggtatcattttcttgaggataatctaacatacacatgaaatcaatatgtaaatcacctgag
ggttttaaatgtgagtgtaaaattacaaatattcagaatattgctcttat

Exon 17

gtacacttaatgttcaaggtaaaagtgtttcagataaattaattataaagtaagctgagctaactataccatga
gcccataatgccacataccaaaacaataacaggatcatgtaaagggtccatcggtgtgaagactttcaatgtc
atcttcaatgttataatcccactccacaccaagatctatgaagggtccgcgacttggcttctccccttgccattt
aaaatataatggcctgtagcctggttcttaatagctaaaggggagaaaaatgagtaaccagcatacattttattt
tatcatattaagtgtgaagattcacagcaagtatatcctttctacaaatcaatgacc

Exon 16

Cgtatgcacatatatgtgtacatatatacatatctatatttaggatttctttgacatttcaatgtttcagtgcattgtaa
agcagagacagcagacagaaacacttaccaagaatatgaggagaagcctcgtcttcttgattaacacat
gtctagccccaggggttatcaaacatcttaagggtaccctttgcatgtgtagtaaattgttaggaagcaaa
gaagaaaacaaaagcaacaacatgggtagttgagcactacattcttactgcttctgcgtgcagaa

Exon 15

Tatagagaaaatgctgcatttctaagaagaagcctcaaaaagtacaaatcacagagaagaataactaa
atccatgagagacctcatcttaccaagcttctgggagttctggtaaattgtccccttcacgggtcggcagtggg
aattatctctccacagacaccacacttatctcaaccttattagaaccaatttctttatcacagcccactttctgg
agagagaagatgtcaaatgtggctaattttagtggtgccagtagtggtcaacatcccagagaatgatgaacc
ctttgttcagtcatggccca

Exon 14

ggctactaaaaataccggaagcaaatgcattattaccaatttgatttcccaacctagaggcctcagagttactc
agaaaaggcatttggctggaaattaggaaaattatttaaaccttgattctgccagtgaccaaagtttgatttga
gaaagtacaatgtctctcaaatttttagctattgaactattcatagacacttactataatgtgatttcaaagata
aaatagtttaaaaaattaaaaatgtgcagatgaatgacagtaaacgtagataacgtagaaagcctttgggg
aagcaatttgaactaccacacactctctcgcacacatatgctatatggatctttgtaagaacagtgcggtcc
atcatgcaccagttgtttcatgtaagcaacatctccagctctccttgactgacagtaaagggtggcatctttcttg
gctgcataagatggaggataaaaattaactatttacataaaatggaatgtttgaacagcatagtttattttatta
aatataagggttcactgtgatttctg

Exon 13

accacatagggatcaagagtcctatgggggtcgtgctggcaaagcttaaactgtgagtaacctccacacag
caggaaggagagcacgaggctactcacggtcaggatgttcatatggcaaccagtgggtgtttggtattctggt
attcaaagtgggagtttcgctgctgacactgctgtgctctgaagtcctcaaagtgttttggcattcttctgtgttac
aaagctggtactcaaaattaacaccaggacaatcctgaccaccaattgatgggcctaaagaaaagacaac
atttaaaaaggccttttggcttctgtctagcagttgaagctgcagacacagtgcctgagggcacaaagccaa
agtttc

Exon 12

Gccctgtgtttacatgcactttataaaacagaataactaataacataatattgaagtatttctggtctctccaaa
aataacaagagttacaaggatatactcacatgggattattgcactggcgtgttctgaaacgaacaccagttc
cacatgtccgagaacaggagccaaatttagtccatgacccccaattgccatcttgttttgctgattagcattctt
ccacatgcaatgacccttatagcaccacttagaaaaatgacaaaaggtaattattaaaatcacgcatacact
aaagctgaagttgtataacaagaaccatctagttgaagggctttc

Exon 11

acagtgcagctgtgtttaccattttccagacctttaagatgataattatgcaacatatcttacatatttattgtga
aataatataactcacttttccagcagcacattcagtcctcatcaagtggagggtcccttttagtcttcaaaaagta
gggattatcaggatggctacaccacagctgtttacatgggtcaaagggtcggaaactggaagatagataatca
aattgtcagtgaactctcagctagcatgacatgcaccaaaaatatattctatacagtttctctgtcctg

Exon 10

tacttgcaccagtgaagtctgtaaatctaccaagcaggagctacacaaatagatttcgagctgcaaaatcta
catcaactgtgagcaacataccgcggtgcacattttatagccaacaccaaaaatcaaaacgacattgctcatc
catagaataattgattccaggaagttctgggagtttaggccaatcatgatcaaaagggcatcaaggagac
agtcataggaactgtaggaagaaaaaatatgcatgtagttaaatgttcaacttaaaaaaatcatgtcctgattta
aaaaaaagtagtcttagcttatatactttagaatttcacccag

Exon 9

ttgctatgatcaagtcctaaagagtcagtttcatgtctgtaggctataaaataaatactttcatgacatgcagcag
gggacatactggatatacttttcagttcttgaccactgcacgaggaccagtggtaacgatggaatgctgcttgt
accaagggagccatgacacttccatagcagtcctcatcaccacacctgttgccctgtccatcatgctccattcc
caacctagaacacaaagagacctcagtggtgaagaatcaggggctactgccttcaacttaattttggttttgaaa
ataagccctgggaaataacgta

Exon 8

ttgtatgctggcatgcattctgccaagaggaaacaatactgggagaagcaggaaagaaaagatcttttt
tggctttatagctgtcagcagtgacttacacatggcccgtttcatgggctactacaaaagcagatgaaaaacc
atcctcatgattcaggggtacaacttctcactggatgacacatgccggtgactggagcatatcctgtagagaat
aacaattacttttattttcattttatgagaaaaccattaagaattacaactggactttgcctaggggggaaaaa
agta

Exon 7

Agagcagaacattgaacagactgcctgattaggattactttttaaagtcatgccctcaagttttcttctaca
tattgacagcaataccttgcatccagcagggtccaaagtcttgctgggttaaaaaaattgcatggatcatgggtt
cagagtgggtgagatcagatctttgctgtgggacgccagcgacacacattctccaagcttctggatgggtt
cccctttctatgaggctgatggactaagtgaaaacaatatgttaaagacacacctgacaattcccaaaggct
tcgtttgataattcctctgaagctgaattgggattaaag

Exon 6

gccttaagttagatgtagcttgaacaatttagtcttctatttgctaaccagtcctaatgtttataattcatgtttta
gacattttacctttgcataatccagcattatcatgcgcaccaggaccacatttatatgcactccgagggactcat
catggtaaatttcattcacctagcaaaaaaaaaagataattgctaaaagaaaggaaacaccgaggatttca
tgagatgtacatataagctgtgtaatcaccacaaatagggtgagt

Exon 5

catggaagtctcacacagattgccaaagggtagcaacaataagagaagtgaattaaaagatcaaaaa
gctttaaaaaggagtcactactacaatgttcattaggggtcaggaggtagtttggacgtgctctttgccatgga
aacggaccacagagtcactccagcagtagctcgatattgtaatcgtttctccgcgtgtctgcggcgt
ctcattgtttcattcagctgctgggtggatgttgccataaacagtagcatcaaggccttcagggtccgac
tctaataagagacaaaaggaaccaggaatttcagttccctaacaggccttccaatcacttcttagaaaaaa
agaatcttctcaggggcctat

Exon 4

catggcaatcacattcttacttgatgggagagtgtttcatcatatcttaaaattatattcattattaagctttgtc
aagacgcctacctctgtagtggaagtctttggacatgtctatgggagcctgttctacagctgatctctttagac
aacatgaatccttcccttttcttccatctgtttacctctttccaaggggtcaatgaaatactcttcattatcactttt
atcattccagcctagagaaaagcacaataatgtgttaatttaaaaaagaaatatctatcatcttcagctcacaag
cacatcttttaaatgtcaagaatatgacttg

Exon 3

Tttccacctccaccttccaaactatgcagaagccatggctgcactcccagcacctgcaaacatacgcacaa
aacagaaggagtcttaccagaccatcacagttgctgatggcaacagagggttcctggaatgtccacgagtc
accaacataagcacagttagtctgcaaaggctctgtttccggattctatacgtagcacttcctggttgatggttg
ttaatgggatcgggttatattcccaggcaccagagatgtctcatgccactccacaacagccccaggagctact
agttgagtgttgggctttagtcgcagatgaaaatctttccaaatgccgtgatgttaaagaacaactgctcagg
gttggaaagacacgtccctcgctgacctcttttgtgactcgcagaaagagtatgggagagatagcgtccttcta
gatttgtgctgactggagtaccagctcatactctctatatctctttattggtaaactctgtggggtgaaaaacaga
actgtcaaaccctgtacaataagagaacacaggagaaacataacacagacaaaagctgccacctcaca
agaccatacttcaatgtgcataatatagacattcctgaatcagctcataaatattaaattttgagaacaaaaat
catttctcaaatgcatttaagtaaccctgtgttcaatgaatcatttttacttaaaaatagtattaaaattacattgtaa
atatctacatctcattaatacatttgcaaacacgaatataatgtgtccataaactataaatggcttgcttaaaatc
aaggcatgcccttaat

Exon 2

ccagtgggaaacatttcagaaacaatatatttgcttagctcacttcattcccttactttcaactaagataagagtg
acatctgagaacaaagcttacctatttgaccatttctcattaccagcctggaaaggaggggggaaagcaag
aaaaagaaagttactgggttcatcttatcacctgtgagttttatttctaccattaatgggttccaagagc

Exon 1

agagggggaggaactttcgggaaaagggtgggttgagtttttttattgttattattccactactgtgatcag
ctgaagtcctaacctctaccagagcggctgctatcaaccaaagtgcaggagaaacCATcacgagtcga
gttactttccaactactgggcaaagcaaatgccagacaaaccaccccccgcccaaaaataagtttc
ttaagaaaaaaaggaaaagggaataatgcgaaatagaaaaaaatgatcttctgtgctttgcttc

Appendix 8

Summary of IL4R sequencing. Sequence covered, positions of primers and SNP locations are shown in relation to the genomic sequence.

Key

aaaa	Forward primer
aaaa	Reverse primer
aaaa	Internal primer
aaaaa	3'UTR
aaaaa	Exon
aaaaa	Sequence covered by sequencing
aaaaa	Sequence not covered by sequencing
a	Common snp (minor allele frequency ≥ 0.05)
a	Rare snp (minor allele frequency $\uparrow 0.05$)
ATG	Start codon
TAG	Termination codon

Exon 2

tggaaaagggagagctggggagtgtatcccagggctgaggccaatggaggattgcagaagagggagag
gggataggaggggaatgtagagggataagcaggggaattggctctttcttcattttcttaaattaaacttttat
tttcagataattaaagatttacacacagctggaagaaatcatagagaagccgggctggtggctcatgccta
taatcccagcacttttggaggctgaggcgggcagatcacttgagatcaggagttcgagaccagcctgggtca
gtatgatgaaaccccatcttactaaaaatacaaaaattagttgggcatggtggcttgccctgtag

Exon 3

atctcccaggacagttgtgttagctcacctctggactttgtgggtctgggtgtgtttcatgattatagagaga
gctctgtgaacgtggaggacctgtgtcggcagagacacaaatggccagggcatggctgggcagccgca
gtggctcaggcctgtaatcccagcacttcgagaagaccagaggggcagatcatgaggtcagaagtcaa
gaccagcctggccaacatggtgaaaccccgcttctactaaaaatacaaaaattagccaggtgtggtgggtg
gcacctgtaatcccagctactcgggaggctgaggcagaagaatcgcttgaaccgggaggtggaggttgc
agtgagctgagattgcaccactgcactccagccttgagacagagcgagactctgtctcggaaaaacaaa
caaacaagcaaacaacaaacaaataaatggccagggcaggggaggggtgcatattgaataagatga
gctctgctggaagcacagggtcagcactaacctgcttcctctctct

Exon 4

ataaatggccagggcaggggaggggtgcatattgaataagatgagctctgctggaagcacagggtcagcac
taacctgcttcctctctctcgcagggtgccttggcatctcccaatgggtggccttgctctgggctcctgttcctgt
gagctgcctggctcctgctgcagggtggcaagctctggttaagtcaccacttctcaatcattcattgttggtattaa
tggcgtgccagggtcctgcagtatgtcacctggccttatggagattacactgcagt

Exon 5

actttgccctgcactgtgctttgtgtattccccttggctctgttgggtgcaagtccccctcacgcattgagttcctg
ggccctcagggtgctcctgtgtctccccaggggaacatgaaggctctgcaggagcccacctgcgtctcggac
tacatgagcatctctacttgcgagtggaagatgaatgggtcccaccaattgcagcaccgagctcggcctgtgt
accagctgggttttctgctctccgagtaagcctgcgctggagctggaggttggggaggtgtgccc aaagggt
ttgccc aaagagtgaagctgggtccagggtggtgcgctggagtgcaaggatgctgagtatgggt

Exon 6

acaggtgaccagcctaaccagccccctgtgtctgcagagcccacacgtgtatccctgagaacaacggag
gcgcgggggtgcgtgtgccacctgctcatggatgacgtggtcagtgcgataactatacactggacctgtggg
ctgggcagcagctgctgtggaagggtccttcaagcccagcgagcatggtagcagggcggagtgggc
aggggtggctgggtgtgttcccacagctgc

Exon 7

ctctgaatctgtgtggtgctccagaagtatgctatggaggtttgtcgacaaaaatctgggtgggtggctgtggt
ttagggccggggctgggctgggtgatgggggagtcactgcatagatcctcacatagaggccgcttctccc
cagtgaaaccagggtcccaggaaacctgacagttcacaccaatgtctccgacactctgctgctgacctg
gagcaaccctgatccccctgacaattacctgtataatcatctcacctatgcagtcaacatttgagtgaaaa
gaccggcagatgtgagtgggcatgctttgacgttttctgtgacctctggggaacagggtgggtgaccagc
agaggcccagtccttgagccaggagcctgggaggcaagccctggggctggatagcaaatcccagga
gctagagacctggcttctcacctggctctgcactagcaagtcctttgcttc

Exon 8

gggtggctcttaaacatggtgggtcagctaacgacagcaaccagggtgggcactggtgcccctcgcccc
ggctggtgccctaacatctcccttttctctaccagttcagaatctataacgtgacctacctagaaccctccctc
gcatcgcagccagcaccctgaagtctgggatttctacagggtcacgggtgagggcctgggctcagtgtat
aacaccacctggagtgagtggagccccagcaccaagtggcacaactgtgagtatcaagaggcctaagc
aatggtaatctccactctccattcttcccctgtggccagacacttcccctggctgagttctctgggcttttatcat
aggatgcctctaatggcaat

Exon 9

tggcactgagccttgaggagcagggggcaatgaggtggccggtgtgacgaggactcaatgtgcaagag
ggagagtggggggagatgaggtggaggggtggtcggcggtcagatcgtggaggggtctcggacgaggggt
cctgaccctgggtctccagtctctgggaaatggagcccaggctgtaccatgggtgacctcagctcatggcttc
ccctcccacttccagcctacagggagcccttcgagcagcacctcctgctgggcgtcagcgtttcctgcattgt
catcctggccgtctgctgtgtgtgctatgtcagcatcaccaagtgtgctgggcccagtgctgccgagcagt
ccctctggagtgcaggggtggcagggacttgcccctctagtctgccccttgcagtcctctcagtcaataatcgcg
cattactgagcagctactacacaacctgagagtagagctg

Exon 10

ccatgtcaccagggttagaatgatgagtcagtggttgacctccagaaactaatcagcctgatctctgatgcc
aataagtattggtgataacgaccacttttatgggaggagcgttcacctgtcaataattcagagatcaacacctt
ttcctttgttttcaggattaagaaagaatggtgggatcagattccaacccagcccgcagccgcctcgtggc
tataataatccaggatgctcaggtaggagtaggcgtggatgaggacatgtgggactgtgtacatgaagaag
tgtggttcagaacacctgggctgttaaggaccttactggccttctggaatggcaaa

Exon 11

gtaagtacctgttggacttctgatctgtgtgatgtcgaggctgtaccttctctgagcattgccgtactccagg
ccgggctgcaaggccactctgctctttcattggctgtctctgtattttaggggtcacagtgggagaagcgggtcc
cgaggccaggaaccagccaagtgccgctatgtatctgaacttaggtcacagcctgcatgcattgggaagggt
gatagaattggagaggcaagcccctagctccatgtctgccttctcttccctgcattcggttaattgcctgtgaca
ttagccttc

Exon 12

aacgttagaaggcatgtctgaagtagacagccatcaggacatggtgatttcaggctgggctttgaagaatg
aataggagtttttcaagtgtcgaaactgaaccctgaccaaactttgcttttcgagacactggaagaattgtctta
ccaagctcttgcctgtttctggagcacaacatgaaaagggtgaagatcctcacaaggctgccaaagag
atgcctttccagggtcttgaaaaatcagcatggtgccagtgagatcagcaagacagtcctctggccaga
gagcatcagcgtgggtgcatgtgtggagtgtttgaggccccgggtggagtgtgaggaggaggaggagga
gaggaagaaaaaggagcttctgtgcatcgctgagagcagcagggatgacttccaggagggaaggg
agggcattgtggcccggtaacagagagcctgttctggacctctcggagaggagaatgggggctttgc
cagcaggaatgggggagtcagcttctccaccttgggaagtacgagtgctcacatgccctgggagga
gttcccaagtgcagggcccaaggaggcacctccctggggcaaggagcagcctctccacctggagccaa
gtcctcctgccagcccagccagagtcagacaacctgacttgacagagacgcccctcgtcatcgagg
caacctgcttaccgcagcttcagcaacccctgagccagtcaccgtgtcccagagagctgggtccagacc
cactgtgcccagacacctggaggaagtagaaccgcagatgcctgtgtccccagctctctgagccaac
cactgtgccccaacctgagccagaaaacctgggagcagatcctccgccaatgtcctccagcatgggc
agctgcagccccgtctcgccccaccagtggtatcaggagtgtgtacatgcggtggagcaggggtggca
cccaggccagtgcggtggtgggcttgggtccccagagaggctggttacaaggccttctcaagcctgcttg
ccagcagtgctgtgtccccagagaaaatgtgggtttggggctagcagtggggaagaggggtataagccttc

caaggacctcattccctggctgccctggggaccctgccccagtcacctgtcccctgttcacctttggactggacag
ggagccacctcgagtcgcagagctcacatctcccaagcagctccacagagcacctgggtctggagcc
gggggaaaaggtagaggacatgccaaagccccacttccccaggagcaggccacagacccccctgtgg
acagcctgggcagtggtgtctactcagcccttacctgccacctgtgcggccacctgaaacagtgtcatg
gccaggaggatggtggccagacccctgtcatggccagtccttgctgtggctgctgctgtggagacaggcct
cgccccctacaacccccctgagggccccagacccctctccaggtggggttccactggaggccagtcgtgt
ccggcctccctggcacccctcgggcatccagagaagagtaaatacctcatcatcctccatcctgccccctggc
aatgtcagagctcaagccagacccccaaaatcgtgaactttgtctccgtgggacccacatacatgagggt
ctcttaggtgcatgctcttgttgcagatgaggactagggcttatccatgctgggaaatgccac
ctcctggaaggcagccagggtggcagattccaaaagactgaagaacctatggtatgaagggtattggcc
ccactgacgttggcctaacactgggctgcagagaactggaccccgcccagcattgggctgggctcgccaca
tcccatgagagtagagggcactgggtcgccgtgccccacggcagggccctgcaggaaaactgaggccct
tgcgacctcgacttgtgaacgagttgttggtgctccctccacagcttctgcagcagactgtccctgttgtaac
tgcccaaggcatgtttgccaccagatcatggcccacgtggaggccacctgcctctgtctactgaactag
aagccgagcctagaaaactaacacagccatcaagggaatgactgggcggcctgggaaatcgatgaga
aattgaacttcagggagggtggtcattgcctagagggtgctcattcatttaacagagcttccctagggtgatgctg
gaggcagaatcccggctgtcaaggggtgttcagttaaggggagcaacagaggacatgaaaaattgctat
gactaaagcagggacaatttgcgtccaaacacccatgccagctgtatggctgggggctcctcgatgcat
ggaacccccagaataaataatgctcagccaacctgtgggccgggcaatccagacagcaggcataaggca
ccagttacctgcatgttgcccagacctcaggtgctaggggaaggcgggaaccttggttgagtaatgctcg
tctgtgtgttttagtttcatcacctgttatctgtgttgcgtgaggagagtggaacagaaggggtggagtttgtataa
ataaagtttcttctcttatttttatgtatttaaccaaacataacctccag