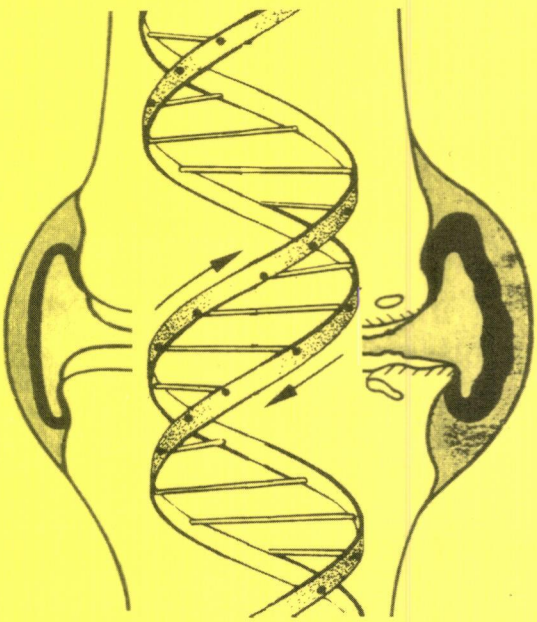
GENETIC PREDISPOSING FACTORS

OF

OSTEOARTHRITIS



Ingrid Meulenbelt

GENETIC PREDISPOSING FACTORS
OF
OSTEOARTHRITIS

Proefschrift
ter verkrijging van de graad van Doctor
aan de Rijksuniversiteit te Leiden,
op gezag van de Rector Magnificus Dr. W.A. Wagenaar,
hoogleraar in de faculteit der Sociale Wetenschappen,
volgens besluit van het College van Dekanen
te verdedigen op dinsdag 28 oktober 1997
te klokke 14.15 uur

door

Ingrid Meulenbelt
geboren te Weesp in 1967

Promotiecommissie:

Promotoren:	Prof. dr. F.C. Breedveld
	Prof. dr. D.L. Knook
Copromotor:	dr. P.E. Slagboom
Referent:	dr. H.G. Brunner (Academische Ziekenhuis Nijmegen, St Radboud)
Lid:	Prof. dr. R.R. Frants

ISBN 90-5412-040-1
Trefwoorden: osteoarthritis, genes, genetic linkage study, genetic association study

The studies presented in this thesis were performed at the Gaubius Laboratory of TNO-PG, Leiden. This work was financially supported by the Dutch League against Rheumatism and the Loosco foundation.

Financial support by TNO-PG Gaubius Laboratorium, and the Dutch League against Rheumatism for the publication of this thesis is gratefully acknowledged.

STELLINGEN

1 Aangezien radiologische osteoartrose (ROA) in de leeftijdscategorie 55-65 jaar frequent voorkomt, leidt het selecteren van een controlegroep vrij van ROA tot het vinden van een associatie met potentieel beschermende allelen tegen ROA (dit proefschrift).

2 Milde mutaties in genen die alleen tijdens de ontwikkeling van het skelet tot expressie komen, kunnen leiden tot ROA op latere leeftijd (dit proefschrift).

3 De hypothese dat het collageen type II (COL2A1) gen uitsluitend betrokken is bij ernstige gewrichtsaandoeningen wordt bevestigd door de associatie van het COL2A1 gen en gegeneraliseerde ROA in de leeftijdscategorie 55-65 jaar (Prockop and Kivirikko 1995, Ann Rev Biochem 64: 403-434; dit proefschrift).

4 Bij het uitvoeren van genetische associatiestudies bij een heterogene ziekte is de marge tussen het maken van fouten van de eerste en tweede soort klein (dit proefschrift).

5 Met een haplotype-analyse in een associatiestudie kan onderscheid gemaakt worden tussen het "linkage disequilibrium" en "susceptibility" model van associatie (Hodge 1993, Am J Hum Genet 53: 367-384; dit proefschrift).

6 Omdat in tandem herhaalde DNA motieven betrokken zijn bij het ontstaan van humane ziekten past men er tegenwoordig voor op deze als "junk" DNA te beschouwen (Willems 1994, Nature Genet 8: 213-215; Bennett et al. 1995, Nature Genet 9: 284-291).

7 Als de hypothese juist is dat door recente evolutionaire veranderingen gewrichten bij de mens verschillen in functionele reserve zal genetische variatie vooral tot osteoartrose leiden in gewrichten met weinig functionele reserve (Hutton 1987, The Lancet I: 1463-1465).

8 Bij het afzoeken van het genomische DNA naar nog onbekende ziektegenen kan, door intelligent gebruik te maken van markers gelegen in gebieden met een hoge gendichtheid, met het afzoeken van 17% van het genoom meer dan 37% van de aanwezige genen worden uitgesloten als ziektegen (Inglehearn 1997, Nature Genet 16: 15).

9 Succesvolle wetenschappers anno 1997 moeten even trend gevoelig zijn als geslaagde couturiers.

10 In de genetische epidemiologie bepaalt het geslacht van de onderzoeker of mannen of vrouwen nullen zijn.

11 Bij het spelen van squash is het krijgen van een "let" of "stroke" afhankelijk van de omvang van de tegenstander.

Stellingen behorende bij het proefschrift getiteld: *"Genetic predisposing factors of osteoarthritis"*.

Leiden, 28 oktober 1997 Ingrid Meulenbelt

Aan mijn ouders,
Voor Guido,

Table of contents

1 GENERAL INTRODUCTION

1 GENERAL INTRODUCTION

1.1 DEFINITION AND CLASSIFICATION OF OSTEOARTHRITIS

Osteoarthritis (OA) is a disease affecting articular cartilage of joints and underlying (subchondral) bone causing pain and disability. Clinically, OA is characterized by joint pain, tenderness, limitation of movement, crepitus, occasional effusion, and local inflammation occurring to a variable degree. Pathologically, OA is characterized by fibrillation due to irregularly distributed loss of cartilage especially in areas of increased mechanical load. Further characteristics are the occurrence of marginal osteophytes, subchondral cysts and sclerosis of subchondral bone. Except for large osteophytes these pathological changes can only be observed by X-ray analysis. The radiological diagnosis of OA relies on characteristic pathological features including 1) formation of osteophytes on the joint margins or in ligamentous attachments 2) periarticular ossicles, chiefly in relation to distal and proximal interphalangeal joints 3) narrowing of joint space associated with sclerosis of subchondral bone 4) cystic areas with sclerotic walls situated in the subchondral bone 5) altered shape of the bone ends. Based on these changes, Kellgren and Lawrence [1] proposed criteria for diagnosing radiographically OA with the severity graded from 0 to 4. Evidence of OA is usually defined as a Kellgren score grade 2 or higher. OA may thus be diagnosed either on the basis of radiographs or on the basis of clinical criteria. However, these different diagnoses have a low association [2, 3]. For population-based case-control studies it is, therefore, of particular importance that OA is defined identically in case and control groups by either clinical signs, radiographic signs or a combination in both groups. In this thesis we will use the term ROA for radiological studies of OA irrespective of clinical OA symptoms.

The classification criteria of Mankin [4] distinguish two categories of OA: primary or idiopathic (without a known cause) and secondary (associated with a known event or disease). Primary OA is divided in two groups A) localized and B) generalized.

Localized forms of OA vary in anatomic location whereas generalized OA (GOA) is a distinct clinical entity of patients with OA at three or more joint groups including the small joints of the hands as most characteristic feature [5]. A genetic disorder in the structure, biochemistry, or metabolism of cartilage is considered to contribute importantly to the pathogenic mechanism of GOA [6].

Secondary OA is the classification for patients with OA due to identifiable, congenital, traumatic, or systemic disease. All secondary forms of OA have similar

clinical, radiological, and pathological features as primary OA but in addition there are unique features that suggest a specific underlying cause. Several of these distinguishing features will be discussed as risk factors of OA (section 1.3).

1.2 PREVALENCE OF RADIOLOGICAL OSTEOARTHRITIS

The prevalence of OA has been investigated in different populations, races and geographic areas. OA increases in prevalence with increasing age suggesting a continuous development of new or incident cases as the population ages. The exact prevalence of OA depends on gender, joint site, and age. An extensive population-based survey of ROA is the EPOZ-study, which consisted of a random sample of 6585 inhabitants of Zoetermeer (The Netherlands) [7]. This and other studies revealed that the pattern of joint sites affected by ROA at different ages is sex specific. At all ages the prevalence of ROA in the small joints of the hands (distal interphalangeal (DIP), proximal interphalangeal (PIP), and metacarpophalangeal (MCP)) and both knees was higher in women than in men, whereas ROA in the lumbar and cervical spine joints (LS and CS, respectively) was more prevalent in men than in women. Between the ages of 55 and 65 ROA in the hip was more frequently present in men than women, whereas hip ROA occurs more frequently in women aged over 65 [7].

Considerable gender differences are also observed in the age of onset and incidence rate of ROA at specific joint sites. ROA in CS and LS occurs at the earliest ages and is, by the age of 50 years, more present than absent in both men and women. In contrast, the increase in prevalence of hip ROA occurs relatively late (over 55 years) but does not increase further after the age of 64 years. Overall, ROA in knees and hands becomes most prevalent with age.

Comparison of population-based studies reveals that the specific prevalence patterns of ROA depending on sex, joint site and age were generally very similar among various populations, suggesting common factors in the etiology of ROA [7]. The existing differences can be explained by 1) interobserver variations during the interpretation of radiographs or 2) by a different distribution of genetic and/or environmental risk factors such as the prevalence of obesity among populations.

The similarity in ROA prevalence may be explained by changes in joint function of various joints in the recent evolutionary development of man (< 1 million years) [8]. Joints most commonly affected (CMC1, MCP1 and DIP, knee and hip) underwent recent rapid evolutionary changes to new functions. An example of this is the extended leg which increased mechanical loading at the hip and knee. These joints

may be relatively underdesigned for their current functions and are likely to fail relatively early in the extended lifespan of today's modern men. Joints that are rarely affected (wrist, elbow, shoulder, ankle) have changed to a function with reduced loading. They have a large functional reserve and, therefore, may fail rarely and late in life.

In this thesis we have studied ROA in a 55-65 year-old cohort. The prevalence of ROA at most joint sites increases considerably in this specific age group and for some joint sites such as the hip has already reached its maximum.

1.3 RISK FACTORS FOR OSTEOARTHRITIS

As indicated in the previous section the etiology of OA is multifactorial. The most important risk factors are age, trauma, obesity, anatomical abnormalities, gender, chondrocalcinosis, and genetic predisposition. Each of these risk factors will be discussed shortly.

1.3.1 Age

The prevalence of OA increases rapidly with age and is almost universal over 70 years of age. Even in the absence of OA, gross anatomical as well as biochemical changes have been reported to occur in articular cartilage as a function of increasing age [9]. Although most of these changes are the reverse of what is seen in arthritic cartilage, they provide a basis explaining why occurrence of OA may increase with age. The collagen network, which is especially important for the tensile strength of cartilage, is affected by a linear rise in non-enzymatic cross links (NEC) with increasing age. NECs are irreversible end products of a biochemical reaction between sugars and collagen fibres that make the collagen network brittle and susceptible to breaks especially under high mechanical loading [10]. Proteoglycans, which are especially important for retaining osmotic pressure, become fragmented with age. The age-related fragmentation of proteoglycans does not decrease the mechanical properties of cartilage [10]. However, breaks in the collagen network that occur with age, especially under high mechanical stress, may eventually lead to leakage of the fragmented proteoglycans from the extracellular matrix (ECM). Depletion of proteoglycans may lead to loss of resilience capacity due to a decrease in osmotic pressure. Under these conditions chondrocytes, in an attempt to repair the ECM, may become hypertrophic and will inevitably produce proteolytic enzymes resulting in cartilage degradation. When the regulation of ECM synthesis and

proteolysis becomes increasingly out of balance, the mechanical properties of cartilage will decline and the arthritic process will progress. Thus, age-related changes make cartilage increasingly prone to fatigue failure and susceptible to the onset of OA.

1.3.2 Trauma

Experimental and observational evidence suggests that knee injury such as meniscal tears, meniscectomy and instability due to ligament disruption, increases articular surface stress and thus increases the risk of subsequent knee OA [11]. In animal models, OA is induced by controlled damage of meniscus and cruciate ligament tears [12]. Several population-based surveys investigated the influence of joint injury on the occurrence of OA. It was found that OA at the site of injury occurs in 35% of men and 15% of women [13]. In both sexes the knee (e.g. meniscectomy [11, 14]) most frequently showed an association with trauma, whereas, in men hip [15] and spine were frequently involved. In women injury did not explain the frequent occurrence of OA at multiple joints. Injury may thus be a prominent cause of knee OA but also of the less common circumstances of hip fracture which causes biomechanical alterations and OA.

1.3.3 Anatomical abnormalities

Gross abnormalities in the anatomy of the hip joint, such as congenital dislocation of the hip, Perthes' disease, or slipped capital femoral epiphysis are well-known for leading to hip OA [16, 17, 18]. These abnormalities are uncommon and unlikely to account for a large proportion of hip OA cases in the population. It has been shown, however, that a large proportion (79%) of cases with idiopathic hip OA may be caused by unrecognized mild developmental abnormalities [19]. Of these cases 40% were found to have acetabular dysplasia, and 39% had a characteristic shape of the femoral head/neck called "pistol grip deformity". The latter may be a result of mild unrecognized slipped capital femoral epiphysis. Both mild acetabular dysplasia and pistol grip deformity occur in a sex specific manner. Among men with hip OA 66% was caused by pistol grip deformity and only 10% among women. In contrast, mild acetabular dysplasia occurs in 68% of women with hip OA and in 15% of men. The presence of mild congenital anatomical abnormalities thus accounts for a substantial proportion of hip OA. It has even been postulated that primary OA of the hip does not exist or is quite uncommon [20, 21, 22].

1.3.4 Obesity

Obesity increases mechanical load on the joint. It has, therefore, often been hypothesized that ROA, especially in weight-bearing joints, will be more common among obese individuals. A number of large population-based studies (NHANES1 [11, 23, 24, 25], EPOZ [26, 27], the Framingham Study [28], New Haven survey [29] and the Leigh population study [13]) show that ROA in the knee is positively associated with obesity. The role of obesity for hip ROA, however, is controversial. Several studies show positive associations of obesity and hip ROA [13, 29], while others report no association [13, 30], weak association [31], or an association of only ROA in the right hip of males [27]. It has also been reported that obesity or peripheral body fat distribution is associated with ROA in the finger joints [13, 23, 29, 27], which cannot be explained by an increase of mechanical load on the joints. It is possible that specific metabolic conditions (e.g. diabetes, hyperuricemia, hypercholesterolemia) or certain elements of an obesity-producing diet (i.e. high fat) accelerate cartilage degradation. It was found, however, that various metabolic factors such as blood pressure, serum uric acid, cholesterol, diabetes, and adipose tissue distribution did not contribute to the association of obesity with knee ROA [24].

Furthermore, it has been shown that obesity causes an increased impulse loading on articular cartilage. This may result in microfractures of the trabeculae in bone which then remodels and becomes stiffer. Stiff bone transmits more load to overlying cartilage, making it more vulnerable to degradation. Alternatively the association of OA with obesity may be mediated by high bone mineral density (BMD) due to high levels of circulating oestrogens generated by the peripheral aromatization of androstenedione by fatty tissue [32].

It has also been suggested that obesity is a consequence rather than a cause of ROA because of a sedentary lifestyle that results from ROA-associated disability. It has however been reported by several studies that obesity is associated with both symptomatic and non-symptomatic knee ROA [25, 28, 31]. In addition, longitudinal studies showed a strong and consistent association between obesity at baseline and the development of knee ROA, respectively, 12 and 36 years later [26, 28]. Taken together, these data suggest that obesity increases the risk of ROA through a mechanical effect on some joints (e.g. knees) and through metabolic pathways on other joints (e.g. fingers).

1.3.5 Gender

Gender-specific differences in prevalence, location, and severity of ROA have repeatedly been reported in population-based surveys (section 1.2) [7]. Factors which may account for sex differences in ROA include obesity, trauma, occupational stress, as well as hormonal milieu, genetic factors, and metabolic disorders.

ROA increases among women after the menopause. The question is, however whether the increase is just due to increasing age or endocrine factors. Oestrogen treatment prevents bone loss but does not prevent the development of ROA in the hands [32]. Furthermore, no difference in hand ROA was found when premenopausal women and women with a oophorectomy or hysterectomy were compared [33]. Women with a relatively long fertile period ($\geq$ 36 years) showed a weak positive association with the occurrence of Heberden's nodes [33].

Sex hormones (especially oestrogen) may thus have little effect on cartilage structures. The inverse relation of ROA with osteoporosis among (postmenopausal) women may be the only indication for a hormonal role in the pathogenesis of ROA. Oestrogen deficiency causes bone loss in premenopausal and postmenopausal women. Postmenopausal women with ROA at multiple joint sites may have an increased bone mineral density not related to obesity but to high oestrogen exposure. Such high bone mineral density causes increased biomechanical stress on cartilage with as result an increased occurrence of ROA [32].

1.3.6 Chondrocalcinosis

Chondrocalcinosis, radiological evidence of deposition of calcium in cartilage, is usually due to calcium pyrophosphate dihydrate (CPPD) crystals. When present in large quantity, these crystals are visible as linear cartilage calcification on x-ray photos. Chondrocalcinosis is associated with a variety of metabolic diseases, but most often it occurs idiopathically in older people. Population-based studies have estimated that between 6% [34] and 35% [35] of elderly persons have evidence of chondrocalcinosis in their knees. Chondrocalcinosis increases in prevalence at ages above 60 years [35], and has also been identified in familial clusters [36, 37, 38, 39]. Chondrocalcinosis is frequently associated with OA. CPPD crystals have been found in a large percentage of patients with ROA of the knees [40]. On the other hand, the pattern of joint involvement in the two disorders is somewhat different. In the chronic polyarticular type of chondrocalcinosis ("pseudo-osteoarthritis"), wrists, metacarpophalangeal joints, shoulders, and elbows are often affected, thus differing

from the general pattern of ROA [41, 42]. The association of chondrocalcinosis with ROA may, at least in part, be explained by the increased prevalence with age of both diseases.

1.3.7 Genetic predisposition

A genetic influence on the etiology of OA was initially recognized in individuals with generalized OA (GOA), defined as OA in at least three different joints groups with involvement of the small joints of the hand as the most characteristic feature [5]. In these individuals a significantly increased risk was observed for first degree relatives [6]. Heberden's nodes as a single entity of the disease occur three times more frequently in sisters of affected women than in the general population [43]. It has been hypothesized that Heberden's nodes are transmitted as a simple, autosomal dominant genetic defect in females, and as a recessive one in males [43]. Other evidence for genetic predisposition in GOA was provided by observations in a group of patients 19 years after unilateral meniscectomy of the knee. ROA changes in the operated knee only occurred in patients who had other signs of GOA [44].

In the Framingham study it was shown that the heritability of ROA is joint-site specific, with the highest heritability estimate for DIP, CMC, and PIP joints, and a modest heritability for MCP and knees [45]. Further evidence for genetic influences on ROA of specific joint sites was provided by a number of studies. In a British female twin study, a heritability estimate of 39-65% was reported for different combinations of scores for osteophytes and joint space narrowing at the hand and knee [46]. In a Finnish twin study it was shown that OA (in any joint) in females, diagnosed by a physician, was explained for 44% by genetic effects. In male twin pairs this effect was not observed [47]. Furthermore, a family history (first degree relative) of medically diagnosed knee ROA was found to be a risk factor for knee ROA in women [48]. Finally, among siblings a significant increased risk of symptomatic OA of the hip and knees was observed, measured by the prevalence of total hip and knee replacement [49]. In conclusion, genetic influences clearly contribute not only to GOA, but also to more commonly occurring ROA phenotypes of localized joints.

Genetic influences are also clearly present in families with early-onset GOA (before 30 years of age) transmitted as an autosomal dominant trait [50, 51]. Clinical and radiological investigation of such families indicated the presence of ROA as the basic disease process without signs of CPDD and chondrodysplasias. These particular families may be used for genetic linkage analysis to identify genes

involved in the pathogenesis of the GOA disease phenotype. Numerous other hereditary skeletal disorders associated with OA are known, for example hereditary chondrocalcinosis, multiple epiphyseal dysplasia, pseudoachondroplasia, and the various types of Ehlers-Danlos syndrome. These syndromes may all be transmitted by a single gene, usually autosomal, sometimes dominant, and occasionally recessive. OA associated with these diseases may be caused by deformity or malfunction of the joint or results from some primary alteration in articular cartilage biochemistry.

The etiology of the genetic influence in OA remains to be unravelled and may involve either a structural cartilage defect, alterations in cartilage or bone metabolism, or alternatively genetic influences on known risk factors for OA such as obesity.

The literature described above shows that genetic predisposition may contribute importantly to OA occurring at single joint sites (such as hands), a combination of joint sites (such as hand and knee in women), and multiple joint sites. It is as yet not clear for which subgroups of patients the genetic component is the main risk factor and which genes are involved.

To study the genetics of OA in this thesis assumptions were made as to which functional defects in the joint may be the basis of OA development. The earliest OA changes are observed in extracellular matrix components of cartilage. The biochemical and cellular changes that occur in OA cartilage form the basis for a choice of OA candidate genes that are studied in this thesis.

1.4 BIOCHEMISTRY OF CARTILAGE AND OSTEOARTHRITIS

Cartilage is composed of a functional extracellular matrix (ECM) which contains collagens, proteoglycans, glycoproteins and a small number of chondrocytes. The chondrocytes synthesize, organize, and regulate the deposition of their surrounding matrix and in normal mature tissue they actively maintain a stable equilibrium between synthesis and degradation of matrix molecules [52]. The collagens form a dense network of fine fibres that provide the overall shape of the tissue. Within this network proteoglycans are retained. Proteoglycans have a branched structure and form negatively charged aggregates. These negatively charged groups create a large osmotic pressure, drawing water into the tissue. The swelling pressure of proteoglycans is limited by the collagen network. The collagen network acts as a sponge. During mechanical stress, the collagen network is compressed forcing water and waste products out of the ECM. At relaxation, water containing nutrients

for chondrocytes are drawn into the tissue from the synovial fluid [53].

During the osteoarthritic process a progressive depletion of proteoglycans is observed. Chondrocytes in their effort to compensate for this loss will become hypertrophic and, inevitably, produce proteolytic enzymes (metalloproteases) that affect both proteoglycans and the collagen network. During the osteoarthritic process the total collagen network content remains unchanged but the integrity is impaired as a result of degradative lesions [54]. These lesions allow the tissue to swell and proteoglycans to leak from the ECM [55]. Furthermore, a dedifferentiation of chondrocytes is observed in arthritic cartilage causing the expression of non-cartilaginous collagens (such as type I, III and X). Expression of these collagens has an important functional impact, as this results in inappropriate repair leading to further cartilage malfunction. For example, expression of collagen type X during the osteoarthritic process is observed in particular areas of osteophytes and of subchondral bone sclerosis [56]. The mineralization and formation of bone on calcified cartilage represent key phases in the bone remodelling of OA [57]. Since collagen type X is normally expressed only during the mineralization process in the growth plate in specific stages of bone development, collagen type X may be important in the progression of OA.

Two main metalloproteases involved in cartilage matrix degradation are collagenase and stromelysin [58]. Collagenase is responsible for the degradation of collagen, while stromelysin can degrade the proteoglycan monomer core protein, including the hyaluronic acid binding region. Increased collagenase and stromelysin levels were identified in human osteoarthritic cartilage probably as result of an imbalance between the synthesis of metalloproteases and the inhibition by tissue inhibitor metallo proteases (TIMP). Stromelysin has been reported to efficiently cleave the type IX collagen into two pieces. Stromelysin in osteoarthritic cartilage could act as a selective depolymerizer of type IX collagen and break the restraints between type II collagen fibrils in cartilage which will lead to an increase in hydration and swelling [54].

This thesis focusses mainly on genes encoding structural matrix proteins including: collagen genes (COL2A1, COL9A1, COL9A2, COL9A3, COL11A1, COL11A2), genes encoding non-collagenous glycoproteins (DCN, CRTL-1, CRTM, COMP), and genes encoding proteoglycans (AGN). These genes may play a role in GOA and ROA since they determine the composition of cartilage architecture and thereby the strength and integrity of articular cartilage during life, especially the collagen network which has almost no turnover. In addition to the genes encoding structural matrix proteins, genes were studied involved in postranslational

modification which ensures correct building of the collagen network (PLOD, LOX) as well as, genes involved in the process of ECM remodelling such as encoding metalloproteinases (MMP3), growth factors (IGF-1 and TGF-ß) and cytokines (IL-1). Mutations in these genes may affect the integrity of the network or can interfere with normal cartilage repair. The following sections summarizes the main characteristics and functions of these gene products (section 1.5, 1.6, 1.7).

1.5 CARTILAGE EXTRACELLULAR MATRIX COMPONENTS

1.5.1 Collagens

The collagen network is the major constituent of cartilage and is crucial for resisting mechanical stress of the joint. The network consists of collagen type II, IX and XI. Collagen type II represents over 80% of the collagenous protein and approximately 50% of the overall protein content of cartilage [59]. Collagen type II is composed of three identical polypeptide α-chains. The three α-chains are assembled tightly into a helix to form a rope-like structure. The excess of collagen type II in cartilage and its crucial functional role in resisting mechanical stress allocate collagen type II as a major candidate gene for OA.

The collagen type II fibril is located around type XI collagen and contains collagen type IX on its surface (Figure 1). Collagen type IX is a minor collagen and represents approximately 10% of the collagenous protein. It is a heterotrimer composed of three genetically distinct polypeptide α-chains. These α-chains contain alternating non-collagenous (NC1-4) and collagenous domains (COL1-3). The type IX collagen molecules decorate the surface of type II collagen in an antiparallel relationship [60]. On the exterior fibril position, type IX collagen molecules covalently cross link both the surface of type II collagen fibrils and other type IX molecules (Figure 1). Collagen type IX thus provides effective interactions between different collagen type II fibrils, while permitting limited, but necessary deformation under compression [60, 61]. Synthesis of abnormal collagen type IX due to a genetic alteration may affect the matrix integrity in a similar way and may decrease lateral strength of the network.

Collagen type XI is also a minor collagen and forms a microfibril around which type II collagen molecules are located. This collagen may therefore be important for the regulation of the diameter of collagen fibrils [62]. It is a fibril-forming collagen with three distinct α-chains α1(XI), α2(XI), and α3(XI). The α3(XI) chain is transcribed from the gene encoding the α1(II) chain of type II collagen and the lysine residues of the α3 chain in collagen are extensively overmodified to glycosylated hydroxylysine

residues. It was also reported that type XI collagen binds with high affinity to proteoglycans [63]. Such binding may be important *in vivo* for anchoring cartilage proteoglycans to the collagen fibrillar network. Together the collagen network with its highly cross-linked fibrillar structure, is crucial in providing tensile strength and in resisting shear forces in cartilage [10].

1.5.2 Proteoglycans

Proteoglycans consist of highly sulphated glycosaminoglycan (GAG) side chains, and N- and O-linked oligosaccharides covalently bound to a relatively invariant core protein. The core proteins are subsequently noncovalently attached to hyaluronic acid via the link protein (Figure 1). Proteoglycans, especially aggrecan (AGN), play a crucial role in maintaining cartilage matrix volume. It is likely that defects in proteoglycans will be subtle and associated with GAG chains. Defects in the core proteins that alter hyaluronan-aggrecan association or decorin-collagen associations are likely to cause extensive disruption of matrix structure or stability. Minor defects in any one of the enzymes associated with elongation, sulphation, and epimerization of the GAG chains could alter the susceptibility of PGs to degradation, change the amount of water they bind, or influence their interactions with other matrix components.

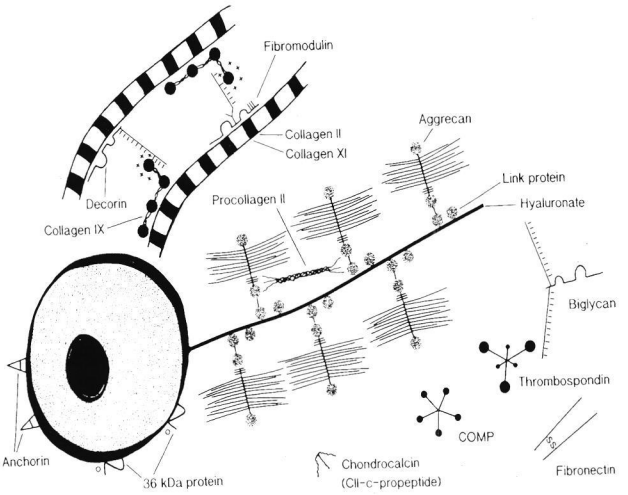


Figure 1. Cartilage ECM components

1.5.3 Glycoproteins

Non-collagenous, non-proteoglycan glycoproteins, referred to as cartilage glycoproteins, constitute a small but significant portion of cartilage (Figure 1). Although functions of most of these proteins are unknown, except for a few such as cartilage link protein (CRTL1), each of these components may contribute to obtaining or maintaining cartilage integrity. The three most relevant glycoproteins are CRTL1, cartilage matrix protein (CRTM) and cartilage oligomeric protein (COMP). CRTL1 is a 41-48 kDa glycoprotein which binds non-covalently to a specific site near the N-terminus of the PG core protein and to hyaluronic acid [64]. In this way CRTL1 stabilizes proteoglycan-hyaluronic acid aggregates [65] and protects PGs from proteolysis [66]. CRTL1 also binds non-covalently with type II collagen [67] and may thus contribute to further stability of the cartilage matrix. The absence of functional CRTL1 will lead to a deficiency of aggrecan immobilization in cartilage, which is then able to leak from the ECM.

CRTM is a 148 kDa trimeric glycoprotein and a component of the cartilage collagen fibrils [68], but not detected in articular cartilage [69]. The CRTM protein is expressed especially in the epiphyseal growth plate, during the process of endochondral bone formation [70, 71]. In this process the length and shape of the bone are determined. Within the growth plate the CRTM protein is a product of and a marker for mature and hypertrophic chondrocytes (postmitotic stages) [71]. The presence of CRTM may interfere with the weight-bearing capacity of articular cartilage and the intervertebral disc. This may be related to the function of CRTM in the assembly of the cartilaginous extracellular matrix.

COMP represents 1% of the wet weight in articular cartilage [72]. It consists of five subunits of 100 kDa each which constrain small amounts of chondroitin sulphate. Homology among both the EGF-like and calmodulin-like domains of COMP and similar regions that bind calcium in other proteins, including the thrombospondins, suggests that COMP binds calcium. Additionally, morphologic studies demonstrated dramatic alterations in the structure of thromobspondin upon removal of calcium indicating that the structure of the thromobospondin family of proteins is calcium dependent. Like other members of the thrombospondin gene family, COMP may participate in calcium-dependent interactions with other proteins. In this context COMP would contribute to the structure of cartilage by interactions with structural proteins in the cartilage matrix [73].

1.6 COLLAGEN SYNTHESIS

Collagen molecules consist of three chains of a tripeptide sequence (Gly-X-Y) assembled into a helix. The folding of collagens and the formation of fibrils occur in a process called nucleated growth. Intracellularly, biosynthesis of collagen involves two important characteristic features that ensure correct registration and folding of the three α-chains into a triple helix [74, 75]. Firstly, collagen is synthesized as a precursor procollagen composed of pro α-chains. After a correct positioning of these α chains (nucleus) the triple helix is formed, propagated from the COOH to the NH_2 terminus of the molecule in a zipper-like fashion [76].

Secondly, correct folding of the triple helix is ensured by a series of posttranslational enzymes that hydroxylate and glycosylate the protein [74, 75]. The posttranslational enzyme prolyl-4-hydroxylase converts prolyl residues in the sequence Gly-X-Pro to 4-hydroxyproline. Hydroxylation of proline is essential for the three pro α-chains to form a helix, since folding is delayed until the pro α-chains acquire sufficient residues of hydroxyproline [77]. This delay ensures an orderly manner of the nucleation and propagation of the triple helix and allows posttranslational modifications of lysyl residues. The enzyme lysylhydroxylase converts lysyl residues into hydroxylysine [78]. The hydroxylysines serve important functions in providing attachment sites for glycosyl residues and in the formation of interchain covalent cross-links. All posttranslational modifications cease as soon as the protein folds into a triple helix, as the posttranslational enzymes cannot interact with the triple-helical protein. Thus folding in itself limits all the posttranslational (over)modifications.

The assembly of collagen monomers into fibrils occurs extracellularly. Extracellular procollagen precursor protein is relatively soluble. Enzymatic cleavage of the N- and C- propeptides, however, decreases solubility dramatically [79] and the collagen monomers spontaneously polymerize into fibrils. This assembly is accompanied by the formation of inter- and intramolecular covalent cross-links between α-chains catalysed by a single enzyme, lysyloxidase. Lysyloxidase oxidates lysyl and hydroxylysyl residues to a variety of di-, tri- and tetrafunctional cross-links [80, 81], which provide the fibrils with a high tensile and mechanical strength.

Assembly of collagen fibrils is probably influenced by other constituents of the ECM. Some fibrils are copolymers of more then one type of collagen (see above), while others may have proteoglycans or other ECM macromolecules on their surface. Because nucleated growth requires perfect structures and enzyme activity,

collagen fibril assembly is one of the most mutation sensitive systems in biology [82]. Genes encoding enzymes involved in nucleated growth as well as the formed structural proteins are regarded as potential OA candidate genes.

1.7 CARTILAGE EXTRACELLULAR MATRIX METABOLISM

Remodelling of joint tissues is a normal process that is most rapid during growth, but it persists throughout life and involves the coordinated degradation and re-synthesis of molecules of the extracellular matrix. Core proteins of the Pgs, especially in contrast to collagens which are very long-lived, are extremely sensitive to proteolysis and have a high turnover throughout life. Collagen turnover is observed in pathological conditions. The polypeptides which mediate the chondrocyte in maintaining a balance between synthesis and degradation are described here. Fibroblast growth factor (FGF), insulin-like growth factor 1 (IGF-1), and transforming growth factor ß (TGFß) play an important role in matrix anabolism. Cytokines and proteinases are involved in matrix catabolism.

1.7.1 Growth factors

Fibroblast growth factor (FGF) plays an important role in cartilage development during embryogenesis. In the growth plate FGF is a known mitogen for chondrocytes and is able to suppress the differentiation of proliferative chondrocytes to a hypertrophic stage with subsequent mineralization. A decrease in active FGF permits mineralisation [83, 84]. The stimulatory effect of FGF, chondrocyte proliferation without differentiation, is also observed in mature articular cartilage [85].

Insulin-like growth factor 1 (IGF-1) increases synthesis of the aggrecan-hyaluronic acid complex responsible for the resiliency of cartilage [86]. Transforming growth factor-ß (TGF-ß) increases general protein synthesis and the specific synthesis of small interstitial proteoglycans of cartilage, including decorin and biglycan. The effect of TGF-ß on protein synthesis in human cartilage is more pronounced than of IGF-1. IGF-1 and TGF-ß are stored as latent forms in large amounts in the cartilage matrix and their activity is probably subject to specific regulatory mechanisms.

Epidemiological studies suggest that serum concentration of IGF-1 is correlated to OA, bone density, and inversely to adiposity [87, 88, 89]. Since obesity, bone mineral density and adiposity are strongly associated with OA (see section 1.3.4), IGF-1 might be an important mediator for the role of these conditions in OA. The

direct relation between serum IGF-1 concentrations and OA, however, is far from clear since high [90], low [91], and normal [92] levels of IGF-1 have been reported in OA. Recently, a modest association of high IGF-1 serum concentration was observed in cases with ROA in the small joints of the finger and with more severe bilateral grades of knee OA [93]. Furthermore, the growth and size of osteophytes were associated with IGF-1 serum concentration, in a study where other characteristics of ROA (such as cartilage degradation) were not associated [94]. Together these data show an attractive role for IGF-1 in the development of OA, however, the exact mechanisms (cartilage degradation, osteophyte development, obesity, bone density, or cartilage loss) remain unclear.

1.7.2 Cytokines

There is much evidence to suggest that IL-1 and TNFα are involved in modulating the chondrocyte metabolism during normal tissue turnover and repair. IL-1 alone is a more potent suppressor/stimulator than TNFα, however, a synergistic enhancement is observed [98]. IL-1 and TNFα stimulate the chondrocyte to synthesize and secrete matrix metallo proteinases (MMPs), particularly collagenase and stromelysin [95, 96, 97].

IL-1 and TNFα are considered to be important mediators of connective tissue destruction in arthritic conditions. In addition to the secretion of MMPs, IL-1 and TNFα suppress the chondrocyte to synthesize collagens and proteoglycans characteristic of hyaline cartilage [98, 99], while they promote synthesis of collagens characteristic of fibroblasts [100]. Such a change of collagen species in cartilage will lead to successive impairment of cartilage function. Finally, IL-1 and TNFα may be involved in proliferative, secondary events of OA (e.g. osteophyte formation) by stimulating the proliferation of human osteoblast-like cells.

1.7.3 Matrix metallo proteinases

Matrix metallo proteinases (MMP) execute degradation of ECM components in articular cartilage. These MMPS are locally produced by chondrocytes as inactive proenzymes.

The MMP family can be divided into three main groups, stromelysins, collagenases, gelatinases. Stromelysins consist of two highly homologous enzymes stromelysin-1 (MMP3) and stromelysin 2 (MMP10) and a third smaller enzyme matrilysin (MMP7). The natural substrate of these enzymes is broad and consists of

PGs, fibronectin, laminin, link protein and type IX collagen. Collagenases consist of three distinct types, interstitial collagenase (MMP1), neutrophil collagenase (MMP8), and the recently cloned collagenase MMP13 [101]. These enzymes are distinguished by their capacity to cleave triple-helical collagen at a single site resulting in fragments corresponding to 75% and 25% of its initial length [102]. Once the initial cleavage of collagen takes place, the two fragments of collagen are no longer stable at body temperature, after which the polypeptide chains are further degraded by other proteinases. The gelatinases, have substrate specificity for denatured collagens, for type IV basement membrane collagen, type V collagen and elastin. Gelatinase A (MMP2) is the most widespread of all MMPs.

The action of MMPs is regulated in at least four different ways, transcription of proMMP genes is regulated by a variety of biologically active agents, the pro-MMPs are activated by plasmin, stromelysin can superactivate collagenase, and the activation of MMPs can be inhibited by the presence of specific tissue inhibitors of MMP (TIMPs) [103]. The proteolytic activity within cytokine-activated cartilage is thus carefully regulated at several stages. Considering all these data, it will be obvious that an imbalance in the regulatory mechanisms of MMP can contribute to the pathogenesis of OA.

1.8 GENETIC FAMILY STUDIES AND OSTEOARTHRITIS

1.8.1 Genetic linkage analysis

Genetic linkage analysis is a powerful tool for mapping genes which cause genetic disorders with a Mendelian pattern of inheritance. Linkage analysis is primarily based on the coinheritance of a disease phenotype and an allele of a polymorphic marker. The extent of coinheritance depends on the genetic distance, measured as the recombination fraction, between polymorphic marker and disease gene.

Important features of linkage analysis are the correct diagnosis of affected and unaffected family members and the choice of a model which explains the mode of transmission of the disease trait in the pedigree. Linkage analysis is thus very suitable for simple Mendelian transmitted traits with distinct clinical phenotypes. In contrast, for the identification of genes involved in complex age-related diseases linkage analysis can be more problematic. In these cases linkage analysis may reject a true candidate gene or sometimes accept a false one. Linkage analysis can be used either to identify unknown genes in the genome by randomly screening

polymorphic markers in search for a locus linked to the disease phenotype of family subjects or by investigating polymorphic markers in close vicinity to candidate genes for the disease.

A large number of genes causing heritable skeletal and cartilage disorders associated with OA have thus far been identified by genetic linkage analysis. Further investigations revealed disease causing mutations in most of these genes. Determining their genetic basis has provided insight into the pathogenetic mechanisms of these disorders and resulted in increased molecular understanding of skeletal development and pathology.

1.8.2 Mutations in the gene encoding collagen type II

In humans the gene encoding the α1-chain of the procollagen type II gene (COL2A1) is an important candidate gene for skeletal heritable disorders (Table 1). It is the major component of articular cartilage and provides tensile strength to the collagen network. These disorders are often associated with early onset of OA at multiple joint sites. A large number of mutations in the COL2A1 gene was identified with distinct pathogenic end points (Table 1) [104]. The spectrum of chondrodysplasias caused by 21 different mutations in the COL2A1 gene has been reviewed [105]. Most of these mutations are dominant and include replacement of the codon for glycine. Interruption of glycine in the Gly-X-Y sequence obviously leads to severe cartilage abnormalities probably because glycine is the only amino acid small enough to occupy the crowded interior of the helix.

Familial osteoarthritis (FOA) is characterized by an autosomal dominant transmitted form of early onset OA at multiple joint sites [5]. Genetic studies of families and unrelated patients with FOA accompanied by mild spondylo-epiphyseal-dysplasia (SED), demonstrated linkage to COL2A1 and mutations in the gene in approximately 25% of the cases [106]. One specific mutation causing an amino acid substitution Arg_{519}->Cys, was found in several apparently unrelated families [107, 108, 109, 110, 111]. Articular cartilage of these patients revealed an abnormal chain incorporated in the triple helical molecule. The protein also appeared to have undergone post-translational overmodification. It was suggested that the FOA mild chondrodysplasia phenotype may represent a subset of heritable OA distinguished by the presence of this specific COL2A1 mutation. It is of special interest that the mutation did not affect the codon for a glycine but a Y position amino acid. Mutations at X and Y positions in general can be expected to be milder than those of glycine substitutions. This may explain why these patients had OA with only mild

chondrodysplasia rather than the severe chondrodysplasias as a consequence of a glycine substitution.

Only a few studies have been focused on the role of COL2A1 in families where OA is the only and primary disease process without any dysplasia. Genetic studies of such families revealed positive linkage to COL2A1 in 2 families [50, 112]. Mutation analysis of all 54 exons of the COL2A1 gene in 45 unrelated FOA patients without dysplasia revealed only one COL2A1 mutation [113]. By using an exon orientated screening, however, possible splicing defects localized within introns and mutations in the promoter region of the gene could not be detected. Sequence variations in regulatory sequences 5' of the gene could exert some influence on the transcriptional activity of the gene with a less severe OA phenotype in which dysplasia is absent [114]. Furthermore, these particular studies were restricted to the analysis of the COL2A1 gene. Mutations in genes other than the COL2A1 gene may be responsible for the occurrence and development of the primary FOA phenotype [106], this thesis Chapter 3].

Table 1. Chromosomal location and associated disease of candidate genes with polymorphic markers used for linkage analysis.

Candidate Gene	Location	Associated disease
COL2A1	12q12-q13.2	FOA*, chondrodysplasia, Stickler syndrome [105]
COL9A1	6q12-q14	OA†, mild chondrodysplasia in mice [119, 120]
COL9A2	1p32.3-p33	EDM2‡ [117, 118]
COL9A3	20q13.3	
COL11A1	1p21	chondrodyslasia in mice [124], Stickler syndrome [123]
COL11A2	6p21.3	Stickler/Kniest- [121] and osteochondro-dysplasia [122]
DCN	12q21.3-q23	
CRTL1	5q13-q14.1	
COMP	19q12	PSACH¶, EDM1§ [138]
CRTM	1p35	
CPPD	8q	familial CPDD ¥ [133]
LOX	5q23.3-q31.2	Ehlers Danlos type IX [144]
PLOD	1p36.3-p36.2	
MMP3	11q22-q23	

*FOA = familial osteoarthritis †OA = osteoarthritis ‡EDM2 = multiple epiphyseal dysplasia type 2 ¶PSACH = Pseudoachondrodysplasia §EDM1 = Multiple epiphyseal dysplasia type 1 ¥CPDD = Calcium pyrophosphatedihydrate deposition disease

Transgenic mice carrying a partially deleted version of the human collagen type II express a phenotype similar to some human chondrodysplasias with dwarfism, a short snout, a cranial bulge, left palate, and delayed mineralization of bone [115]. Similar phenotypes were observed in transgenic mice expressing mouse collagen type II with a deletion elimination exon 16 to 27. This deletion was, however, in frame in terms of the coding sequences and the requirement of GLY as every third amino acid in the triple helix. The phenotypes of these latter mice were, however, highly variable in severity. The mice with the mildest phenotype develop degenerative changes in joints resembling OA at older ages [116].

1.8.3 Mutations in the gene encoding collagen type IX

The human collagen type IX is a heterotrimer composed of three genetically distinct polypeptide chains encoded by different genes (Table 1). Mutations in the human COL9A2 gene cause multiple epiphyseal dysplasia (MED) and the gene is known as the second disease locus (EDM2) [117, 118]. The first locus (EDM1) is caused by a mutation within the cartilage oligomeric matrix gene (COMP) [73, section 1.8.7]. Affected individuals showed variable degrees of a waddling gait and stiffness and/or pain in the knee. Other joints were only rarely involved and no shoulder or hip complaints were registered. There were no spine abnormalities and only some patients had mild short stature and/or stubby hands. X-rays of the knees demonstrated flattened, irregular epiphyses and progressive OA. Thus, mutations in COL9A2 appear to have a particularly strong effect in the knee. Some forms clinically resemble another chondrodysplasia phenotype, the mild form of pseudoachondroplasia (PSACH). On the basis of their clinical similarities as well as similar ultrastructural and biochemical features of cartilage, it has been proposed that MED and PSACH belong to a single bone dysplasia family. Thus far, linkage with hereditary diseases or mutations in the COL9A1 and COL9A3 genes have not been reported.

Transgenic mice with either a dominant negative mutation or a homozygous null mutation in COL9A1 [119, 120] had unexpectedly mild phenotypes, suggesting that collagen type IX is not essential for the assembly of cartilage ECM during development. Interestingly, these mice developed a degenerative joint disease resembling human osteoarthritis. These data suggest that collagen type IX does not play a critical limiting rol during skeletal development, but instead may be important for the physical integrity of articular cartilage or participate in cartilage homeostasis.

1.8.4 Mutations in the gene encoding collagen type XI

The human collagen type XI is a fibril-forming collagen with three genetically distinct α-chains, one of which is transcribed from the α1(II) chain of type II collagen gene (Table 1). A mutation in the human COL11A2 gene was reported in a family with autosomal dominant transmitted Stickler syndrome characterized by mild spondyloepiphyseal dysplasia and osteoarthritis [121, 122]. The phenotype differs from the Stickler cases (COL2A1 defects) by the absence of severe myopia and vitreo-retinal degeneration. A recessive mutation in the COL11A2 gene was reported in a family with autosomal recessive ostospondylomegaepiphyseal dysplasia (OSMED) [122]. Heterozygous carrier parents for this mutation are asymptomatic possibly because the mutation is located close to the amino terminus of the molecule which may still allow incorporation of mutated molecules into collagen fibrils without severe loss of function. Although both heterozygous parents did not show clinical abnormalities in articular cartilage, there was a history of OA on the paternal side of the family, raising the possibility that the COL11A2 mutation is a predisposing factor for OA. Recently, a mutation in the COL11A1 gene was found in a family with the full type II Stickler syndrome including vitreous and retinal abnormalities [123].

Autosomal recessive chondrodysplasia (*cho*) in mice is caused by a mutation in the COL11A1 gene [124]. Histopathological and biochemical analysis of *cho* mice indicated that the absence of the α1(XI) chain results in severe disruption of the columnar arrangement and maturation of the growth plate chondrocytes and to unusually thick collagen fibrils in cartilage. Thick collagen fibrils in mutant cartilage of *cho* mice [125], provide strong support for the function of collagen type XI in determining the diameter of collagen fibrils. This was also indicated by the observation of thicker fibrils in the cartilage of transgenic mice overexpressing type II collagen at a normal expression of collagen type XI [126]. Transgenic mice lacking the expression of the α2(XI) collagen chain developed a mild phenotype with a short snout, prominent forehead, shortened limbs, and shortened tail [127].

1.8.5 Mutations in the gene encoding collagen type X

Collagen type X consists of three identical α1-chains encoded by the COL10A1 gene (Table 1). Several mutations in the COL10A1 gene are associated with the autosomal dominant disorder metaphyseal chondrodysplasia type Schmid [128, 129, 130]. These mutations alter the carboxyl-terminal, non-triple helical domain of

collagen X homotrimers, which is highly conserved across species and may participate in the initial trimerization and alignment of the triple helix. The likely consequence of a mutation in this region is impairment of chain association and folding into triple helices. As a result, a reduction in the level of normal functional type X collagen molecules is observed with consequent morphological changes in growth plates of developing long bones. Affected individuals are short in stature, but not severely dwarfed.

Evidence supporting the structural role of collagen type X in hypertrophic cartilage matrices undergoing degradation was established with transgenic mice that express a mutant chicken type X collagen. The growth plates were compressed in these mice and contained smaller hypertrophic chondrocytes. The number and size of bony trabeculae, composed of calcified hypertrophic cartilage cores with newly deposited bone, were reduced although the mineralization of trabeculae and periosteal bone were normal [131]. Collagen type X may fulfil its support function by forming pericellular polymeric networks as observed *in vivo* [132]. Surprisingly, mutant mice that lacked type X collagen developed normally which may indicate that other collagens may compensate for its absence.

1.8.6 Chromosomal loci involved in CPPD

The disease gene for familial CPDD was genetically mapped to a region on chromosome 8q [133]. A second location of the gene causing CPDD was mapped to chromosome 5p [134]. Candidate genes have not yet been identified in these regions.

1.8.7 Mutations in genes encoding non-collagenous glycoproteins

Although the function of most of the ECM non-collagenous proteins is not known, they may support the integrity of cartilage. The genes encoding the most investigated non-collagenous proteins that may be candidates for FOA are the cartilage link protein gene (CRTL1), the decorin gene (DCN), the cartilage matrix gene (CRTM), and the cartilage oligomeric matrix gene (COMP) (Table 1). The CRTL-1, CRTM, and DCN genes were, thus far, excluded as the disease loci in several heritable chondrodysplasias by genetic linkage analysis [135, 136, 137, this thesis].

The first non-collagenous gene and the first gene identified to cause dominantly inherited mild and severe pseudoachondroplasia (PSACH) and multiple epiphyseal

dysplasia (MED), was COMP, and now known as the EDM1 locus [73, 138]. Mild PSACH and MED are characterized by mild short stature and early onset OA. The identification of this first MED locus was quickly followed by the localization of the second locus (EDM2) [117, 118] caused by a mutation in the COL9A2 (see section 1.8.3) and demonstrates that MED is genetically heterogeneous.

1.8.8 Mutations in genes encoding proteoglycans

A number of PG loci were investigated in human skeletal disorders and showed no linkage [139]. Mice homozygous for the *cmd* mutation had a deficient synthesis of cartilage proteoglycan core protein [140]. These mice are characterized by disproportionate dwarfism and cleft palate. Their cartilage lacked the high molecular weight proteoglycan characteristic of cartilage, however, smaller proteoglycans were present in normal amounts.

1.8.9 Mutations affecting post-translational modification enzymes

Genes encoding enzymes involved in procollagen post-translational modification are lysyl-hydroxylase (PLOD) and lysyl-oxidase (LOX) (Table 1). Mutations in the PLOD gene were associated with Ehlers Danlos type VI [141, 142, 143], characterized by joint hypermobility, skin hyper-extensibility, skin fragility, and ocular abnormalities. Of these characteristics the joint hypermobility especially is often associated with the onset of OA (see section 1.3.3).

Mutations in the LOX gene have been associated with type IX Ehlers-Danlos [144] (Table 1). The specific clinical characteristics of type IX Ehlers Danlos include bladder diverticula with spontaneous ruptures, inguinal hernias, slight skin laxity and hyperextensibility, and a number of skeletal abnormalities.

Because PLOD and LOX act on the lysyl and prolyl residues of several types of collagen (including collagen type I and II), mutations within these genes show a phenotypic effect in multiple tissues (skin, bone, and cartilage). Since collagen type I predominates in skin and bone and consists of a much larger number of prolyl and lysyl residues than collagen type II, mutations in PLOD and LOX show severe functional impairment of skin and bone especially and a less clear effect on cartilage.

1.8.10 Most important candidate genes for osteoarthritis

In mice mutations within the COL2A1 and COL9A1 have so far been most clearly involved in phenotypes resemble human OA. As shown in section 1.8.1-1.8.9, genes identified in these disorders are often structural components of the ECM (Table 1). In mice mutations within the COL2A1 and COL9A1 were most clearly involved in phenotypes that resemble human OA (section 1.8.1 and 1.8.2). The STR/Ort mice develop naturally-occurring animal osteoarthritis and are considered a well-characterized model of idiopathic OA [145]. The loci responsible, however, have not yet been identified.

In humans the COL2A1 gene was involved in families with early onset primary OA in multiple joints (FOA) with or without a mild spondyloepiphyseal dysplasia (section 1.8.2). The COL9A1 gene is until now not identified as a disease gene in FOA families without dysplasia. The COL9A2 gene, however, is known to cause MED, which has a relatively mild phenotype with a particular strong OA effect in the knee but not in other joints. Together it seems that mutations identified thus far in the genes encoding collagen type II and collagen type IX do not necessarily cause a critical limiting role during skeletal development. Instead these proteins may be important for the physical integrity of articular cartilage or participate in cartilage homeostasis.

Although, the function of most noncollagenous cartilage proteins is not (yet) known, the identification of COMP as a mutated gene in mild PSACH, and MED associated with early onset OA indicates that these noncollagenous genes potentially may be involved in milder types of heritable disorders.

Apart from these most obvious and investigated candidate genes for OA, genes encoding enzymes involved in cartilage ECM matrix metabolism (section 1.7) may play an important role in the onset of OA. Mutations that influence for example para- or endocrine levels of these enzymes may cause subtle changes in the balance of synthesis and degredation of cartilage in response to repair of small cartilage lesions. The effect of these genes may thus determines especially the clinical severity or progression of OA.

1.9 GENETIC POPULATION-BASED STUDIES

Mild mutations in the genes discussed in section 1.8 may predispose to the onset of OA at later ages without a clear Mendelian inheritance pattern. The investigation of genetic influences on the etiology of sporadic OA in the population and at later

ages is more complex, since the effects of other risk factors may interfere with the genetic etiology of OA. A few characteristic problems arise when identifying "susceptibility genes" to a complex disease such as OA. Individuals that inherit a predisposing allele may not manifest the disease at all or very late in life (incomplete penetrance), whereas others who inherit no predisposing genetic allele may nevertheless develop the disease as a result of non genetic causes (phenocopies). Thus the genotype at a given locus affects only the probability of the disease but, however, does not determine the outcome. It may also be that the onset of OA at later ages may be genetic (or locus) heterogeneous, with mutations in any of several candidate genes that result in identical phenotypes. Finally, OA may manifest a polygenic inheritance pattern, with the requirement of the simultaneous presence of mutations in multiple susceptibility genes.

Methods which are suitable to detect genetic factors predisposing to more complex diseases are not based on the construction of a model of inheritance but are nonparametric (independent of a model of inheritance). Two methodst of detecting genetic predisposing or protective alleles for OA are analysis of disease association in unrelated individuals and allele-sharing methods in relatives (such as sibling pairs).

1.9.1 Disease association studies and osteoarthritis

Association studies are based on a comparison of unrelated affected and unaffected individuals from a population (case-control). A locus is associated with the disease if allele(s) of the locus occur at a significantly higher frequency among affected rather than control individuals. The power to detect association using a marker depends on several factors: 1) the strength of the linkage disequilibrium between the marker and the disease, 2) the frequency of the disease mutation, 3) the increase in risk attributable to the particular disease-susceptibility locus under consideration and 4) the penetrance of the different disease locus genotypes.

Several association studies aimed to detect association of type II procollagen gene (COL2A1) alleles with the occurrence of ROA. Inconsistent associations were observed (Table 2) which may be due to substantial variation between the studies concerning the markers used and the definition of cases and controls. Allele frequencies of COL2A1 RFLP polymorphims (*Eco*R1, *PsT*1, *Hind*III, and *Bam*H1) were investigated in British women with symptomatic and radiologically confirmed OA in more than one joint before the age of 60 as compared to controls from a hospitalized population without known musculoskeletal symptoms. A significant

Table 2. OA association studies

Gene	origin	Marker	♂/♀	Disease phenotype		Association	Reference
				cases (n)	controls (n)		
COL2A1	British	BamHI + other RFLPs	♀	>1 joint ROA before 60 yrs (n=172)	no symptoms (n=182)	predisposing allele ($p<0.04$)	[146]
COL2A1	British	BamHI + other RFLPs	♂+♀	Heberden' nodes before 60 yrs + ≥3 joints ROA (n=120)	no symptoms (n=74-334)	no association	[147]
COL2A1	British	MaeII	♂+♀	Heberden's nodes before 60 yrs + >3 joints ROA (n=152)	GOA phenotype undetermined (n=146)	predisposing allele ($p<0.01$)	[148]
COL2A1	Finnish	PvuII+VNTR	♂+♀	>1 weight-bearing joint ROA before 50 yrs (n=82)	no joint symptoms before age 60 yrs+no ROA (n=96)	no association	[149]
COL2A1	Finnish	PvuII+VNTR	♂+♀	finger ROA (n=98)	no joint symptoms before age 60 yrs+no ROA (n=96)	no association	[149]
AGN	American	CA-repeat exon	♂	bilateral hand ROA ≥ 60 yrs (n=62)	no bilateral hand ROA (n=84)	predisposing allele OR=5.3 (P=0.01)	[151]
A1ACT	British	Taq 1	♂+♀	GOA (n=35)	normal (n=125)	predisposing genotype RR = 2.9 (P= 0.01)	[152]

($p<0.04$) predisposing association for OA with the *Bam*HI marker was observed [146]. In a second study, British GOA patients defined by the presence of Heberden's nodes before the age of 60 years and ROA involvement in at least three other joint groups, showed no association with the *Bam*HI RFLP, nor with 4 other RFLPs (*Hind*III and 3 different *Pvu*II sites) [147]. In these patients a significant increase of a MaeII allele was observed, which was also associated with a reduced expression of the COL2A1 gene in arthritic knee cartilage [148].

An association study with a variable number of tandem repeat (VNTR) polymorphism located 1.35 kb to the 3` end of the COL2A1 gene [149] was performed in Finnish cases with ROA in more than one joint and in a subgroup of cases with OA of the finger joints as compared to cases without ROA in these joints. An association with any of the VNTR alleles in these cases was not observed [150]. A significant association with the human aggrecan (AGN) gene was observed in men with bilateral hand OA (OR = 5.3, $p=0.007$) [151]. Finally, a predisposing association was observed with a homozygous *TaqI* genotype of the α1-antichemotrypsin (A1ACT) gene was observed in GOA patients [152].

These studies have shown that genetic predisposition may contribute importantly to ROA but to a varying extent in men and women depending on the joint site(s). It is far from clear for which subgroup of patients the genetic component is relevant. Moreover the studies summarized here and in section 1.3.7 indicate that genetic factors may not only contribute to the onset of GOA.

1.9.2 Allele-sharing methods and osteoarthritis

Allele-sharing methods involve affected relatives (sibling pairs, or more distant relatives) to find out whether a particular chromosomal region is inherited "identical by descent" (IBD) (inherited from a common ancestor) more often than expected under random Mendelian segregation. Under random segregation, two siblings are expected to show IBD sharing of zero, one or two copies of any locus with a 25%-50%-25% distribution. Whenever the chromosomal region investigated contributes to the disease phenotype of the affected relatives the observed distribution of alleles will tend to swift to a higher percentage of sharing.

A limited association ($0.05<p<0.1$) was observed (Table 3) in a small sibling pair study (N=21) using 5 RFLP markers (*Bam*HI, *Hind*III and three different *Pvu*II sites) within the COL2A1 gene and siblings with GOA defined by the presence of Heberden's nodes before the age of 60 and the involvement of at least three other joints groups with OA [147]. In a second study, the IBD sharing of these siblings plus

Table 3. OA allele sharing studies

Gene	origin	Marker	♂/♀	Disease phenotype siblings	Association	Reference
COL2A1	British	BamHI + other RFLPs	♂+♀	Heberdens before 60 yrs + ≥ 3 joints ROA (N=21)	limited (0.05<p<0.1)	[147]
COL2A1	British	VNTR	♂+♀	Heberdens before 60 yrs + ≥3 joints ROA (N=21+17)	no association	[153]
CRTM	British	3' UTR	♂+♀	Heberdens before 60 yrs + ≥3 joints ROA (N=21+17)	no association	[153]
CRTL1	British	CA-repeat promoter	♂+♀	Heberdens before 60 yrs + ≥3 joints ROA (N=21+17)	no association	[153]

an additional number of siblings (N=17) was investigated using the highly informative COL2A1 VNTR marker, and multi-allelic markers within the CRTM and CRTL-1 gene [153]. This study did not show significant allele-sharing IBD for any of the three loci. Since the number of siblings in this study (N=38) was low, these results do not exclude the possibility of mutations in the COL2A1, CRTL-1, and CRTM gene causing OA. In addition the relatively severely affected individuals who were included in the study may provoke the identification of rare mutations rather than susceptibility alleles for common OA or ROA.

1.10 OUTLINE OF THIS THESIS

This thesis represents a search for genetic factors predisposing to the onset of OA at an early age in families and in unrelated individuals between 55-65 years of age. To study genetic factors in both families and populations, human genomic DNA should be collected. Usually, genomic DNA is isolated from peripheral blood samples which is expensive and an invasive procedure to which, for ethical reasons, objections may be raised (especially in studies involving older individuals). To simplify participation for individuals and the collection of genomic DNA the possibility of developing a non-invasive DNA sampling and isolation method involving oral samples taken with cotton swabs was investigated (Chapter 2).

Thus far genetic linkage studies to identify genes involved in the OA process were most often performed in families with a FOA phenotype accompanied by mild spondylo-epiphyseal-dysplasia (SED). These studies demonstrated linkage to COL2A1 and mutations in the gene in approximately 25% of the cases [106]. Only a few studies have focused on the role of COL2A1 in families where OA is the only and primary disease process without any dysplasia. Genetic studies of such families demonstrated positive linkage to COL2A1 in 2 families [154, 155]. Studies on other important candidate genes have not been reported for FOA. To investigate the "pure" familial OA process without another underlying disease a Dutch family with primary FOA without mild epiphyseal dysplasia was investigated. The COL2A1 and other 13 other important candidate gene loci were tested for their involvement in the disease process (Chapter 3).

In addition to FOA families, genetic factors play a role in the onset of OA at later ages in the population. A suitable method for investigating the role of candidate genes in the onset of OA at later ages is offered by disease association studies. Association studies described in this thesis were performed in a population-based cohort study of determinants and prognosis of chronic disease in the elderly, The Rotterdam study (ERGO). Subjects (N=1040) were selected between 55-65 years and their which radiographic characteristics of knee, hip, hands, wrists and the thoraco-lumbar spine were determined. Furthermore, information on sex, age (in years), body mass index (BMI; measured as weight in kg divided by height^2 in metres), and bone mineral density (BMD; measured as grams of mineral divided by area in cm^2) was assessed. From this population-based subset cases were selected on the basis of ROA in one or both knee and/or hip joints, whereas controls were selected on the basis of absence of ROA in all radiographed joints. In addition, this study design provided the possibility of investigating the genetic influences on the occurrence of ROA of specific subgroups of cases (generalized or localized pattern of ROA) and the effect of gender, age, BMI and BMD. The first and most extensively investigated candidate gene was the collagen type II gene. Collagen type II is the most abundant collagen protein and crucial in resisting mechanical stress of the joint. Within the ERGO population the aim was to detect a relationship between COL2A1 VNTR alleles and the occurrence of ROA in the knee and/or hip. Then the effects of generalization of the osteoarthritic process, age, gender, Heberden's nodes and BMI on the association between the COL2A1 VNTR polymorphism and ROA were examined (Chapter 4).

Disease associations may be explained by a linkage disequilibrium model in which the mutation causing OA is in linkage disequilibrium with the VNTR allele. By

defining haplotypes of this allele including additional COL2A1 polymorphisms it is possible to delineate the risk allele, resulting in an optimized disease association with a particular ancestral haplotype, and excluding all other alleles. Furthermore, insight into the evolutionary relations of the gene provides important information on the number of founder genes that occur in the population and the number of recombinations and mutations that have occurred since. For the COL2A1 gene pairwise linkage disequilibrium analysis of the COL2A1 gene was performed with 7 intragenic RFLP sites in a population of randomly selected Dutch individuals. Furthermore, haplotypes of these RFLPs were determined and an evolutionary tree was constructed (Chapter 5). This provided an indication of the number of COL2A1 haplotypes that occurred in the Dutch population and its evolutionary relations.

In chapter 6 haplotype analysis for the ROA association with the COL2A1 VNTR alleles was performed. Linkage disequilibrium analysis using a maximum likelihood estimation method [156], was performed with three COL2A1 RFLP polymorphisms (within intron 9, intron 33, and at 3'UTR) in addition to the VNTR polymorphism in an extended population of cases and controls. Estimated frequencies of pairwise and three way haplotypes were used to delineate specific COL2A1 haplotypes putatively predisposing the development of ROA.

The contribution of other, non-collagenous, proteins of the extracellular matrix (ECM) to the etiology of ROA is an area of increasing interest. Genes encoding such proteins are the cartilage link protein gene (CRTL1) and the cartilage matrix protein gene (CRTM). The cartilage link protein stabilizes the proteoglycans [157] and the cartilage matrix protein is a component of cartilage collagen fibrils with an as yet unknown function [158]. Intragenic polymorphisms in the 3'UTR of CRTM [159], and in the promoter region of CRTL1 [160] are associated with ROA in ERGO subjects of 55-65 years irrespective of possible manifestation of clinical OA (Chapter 7).

By using these studies an investigation was made of candidate genes contribute to the onset of OA and for which subgroups of patients (sex and/or specific joints sites) the influence of these genes was most relevant.

1.11 REFERENCES

1. Kellgren JH, Lawrence JS. Atlas of standard radiographs; the epidemiology of chronic rheumatism. Vol 2 Oxford: Blackwell Scientific Publications, 1963.

2. Claessens AAMC, Schouten JSAG, Ouweland van den FA, Valkenburg HA (1990) Do clinical findings associate with radiographic osteoarthritis of the knee? Ann Rheum Dis 49:771-774.

3. Hart DJ, Spector TD, Brown P, Wilson P, Doyle DV, Silman AJ (1991) Clinical signs of early osteoarthritis: reproducibility and relation to X-ray changes in 541 women in the

general population. Ann Rheum Dis 50:467-470.

4. Mankin HJ, Brandt KD, and Shulman LE (1986) Workshop on etiopathogenesis of OA. J Rheumatol 13:1130-1160.

5. Kellgren JH, and Moore R. (1952) Generalized osteoarthritis and Heberden's nodes. Br Med J 1:181-187.

6. Kellgen JH, Lawrence S, and Bier F (1963) Genetic factors in generalized osteo-arthrosis. Ann Rheum Dis 22:237-253.

7. Saase van JLCM, Romunde van LKJ, Cats A, Vandenbroucke JP, and 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 48:271-280.

8. Hutton CW (1987) Generalised osteoarthritis: An evolutionary problem? The Lancet 1463-1465.

9. Brandt KD, Fife RS (1986) Ageing in relation to the pathogenesis of osteoarthritis Clin Rheum Dis 12:117-130.

10. Mayne R and Brewton RG (1993) Extracellular matrix of cartilage: Collagen. In: Woessner JF Jr. and Howel DS eds. Dekker, Inc. New York 81-108.

11. Davis MA Ettinger WH, Neuhaus JM, Sangsook AC and Hauck WW (1989) The association of knee injury and obesity with unilateral and bilateral osteoarthritis of the knee. Am J Epidem 130: 278-288.

12. Troyer H (1982) Experimental models of osteoarthritis: a review. Semin Arthritis Rheum 11:362-374.

13. Kellgren JH and Lawrence JS (1958) Osteo-arthrosis and disk degeneration in an Urban population. Ann Rheum Dis. 17: 388-397.

14. Cooper C, Mc Alindon T, Snow S, Vines K, Young P, Kirwan J, and Dieppe P (1994) Mechanical and constitutional risk factors for symptomatic knee osteoarthritis: Differences between medial tibiafemoral and patellofemoral disease. J Rheumatol 21:307-313.

15. Heliövaara M, Mäkelä M, Impivaara O, Knekt P, Aromaa A, and Sievers K (1993) Association of overweight, trauma and workload with coxarthrosis. Acta Orhtop Scand 64:513-518.

16. Wedge JH (1978) The natural history of congenital dislocation of the hip: a critical review. Clin Orhtop 137:154-162.

17. Gower W, Johnston RC (1971) Leg-Perthes disease: long-term follow-up of thirty- six patients. J Bone Joint Surg (AM) 53:759-768.

18. Ordeberg G, Hansson LI, Sandstrom S (1984) Slipped capital femoral epiphysis in southern Sweden: a long-term result with no treatment or symptomatic primary treatment. Clin Orhtop 191:95-104.

19. Stulberg SD Cordell LD, Harris WH, Ramsey PL, and Mac Ewen GD: Unrecognized childhood hip disease: A major cause of idiopathic osteoarthritis of the hip. In: Amstutz HC (ed): The Hip, Proceedings of the third open scientific meeting of the Hip Society. St. Louis, CV Mosby, 1975, p212.

20. Harris W (1986) Etiology of osteoarthritis of the hip. Clin Orthop 213:20-33.

21. Murray RO (1965) The aetiology of primary osteoarthritis of the hip. Br J Radiol 36:810-824.

22. Solomon L (1976) Patterns of osteoarthritis of the hip. J Bone Joint Surg (Br) 58:176-183.

23. Davis MA, Neuhaus JM, Ettinger WH and Mueller WH (1990) Body fat distribution and osteoarthritis. Am J Epidemiol 132: 701-707.

24. Davis MA, Ettinger WH and Neuhaus JM (1988) The role of metabolic factors and blood pressure in the association of obesity with osteoarthritis of the knee. J Rheumatol 15: 1827-1832.

25. Davis MA, Ettinger WH, Neuhaus JM, and Hauck WW (1988) Sex differences in osteoarthritis of the knee; the role of obesity. Am J Epidemiol 127:1019-1030.

26. Schouten JSA, Ouweland van den FA, and Valkenburg HA (1992) A 12-year follow-up study in the general population on prognostic factors of cartilage loss in osteoarthritis of the knee. Ann Rheum Dis 51:932-937.

27. Saase van JLCM, Vandenbroucke, JPI, Romunde, LKJ, and Valkenburg, HA (1988) Osteoarthritis and obesity in the general population. A relationship calling for an explanation. J Rheum 15:1152-1158.

28. Felson DT, Anderson JJ, Naimark A, Walker AM, Meenan RF (1988) Obesity and knee osteoarthritis; The Framingham study. Ann Intern Med 109: 18-24.

29. Acheson RM and Collart AB (1975) New Haven survey of joint diseases XVII Relationship between some systemic characteristics and osteoarthrosis in a general population. Ann Rheum Dis 34: 379-387.

30. Bjorntorp P (1985) Regional patterns of fat distribution. Ann Intern MEd 103: 994-995.

31. Hartz AJ, Fischer ME, Bril G, Kelber S, Rupley D, Oken B, and Rimm AA (1986) The association of obesity with joint pain and osteoarthritis in the HANES data. J Chron Dis 39:311-319.

32. Dequeker J (1986) The relationship between osteoporosis and osteoarthritis. Clin Rheum Dis 11:271-296.

33. Schouten JSA, Ouweland van den FA, and Valkenburg HA (1992) Natural menopause, oophorectomy, hysterectomy and the risk of osteoarthritis of the DIP joints. Scan J Rheumatol 21:196-200.

34. Cabanel G, Phelip X, Verdier J-M, Gras J-P (1970) Frequence des calcifications meniscales et leur signification pathologique. Rhumatologie 7:255-262.

35. Ellman MH, Brown NL, Bertram L (1981) Prevalence of knee chondrocalcinosis in hospital and clinical patients aged 50 or older. J Am Geriatr Soc 29:189-192.

36. Reginato AJ (1976) Articular chondrocalcinosis in the Chiloe islanders. Arthritis and Rheum 19:395-404.

37. Korst van der JK, Geerards J, Driessens FCM (1974) A hereditary type of idiopathic articular chondrocalcinosis. Am J Med 56:307-15.

38. Doherty M, Hamilton E, Henderson J, Misra H, and Dixey J (1991) Familial chondrocalcinosis due to calcium pyrophosphate dihydrate crystal deposition in English families. Brit J Rheum 30:10-15.

39. Balsa A, Martin-Mola E, Gonzalez T, Cruz A, Ojeda S, Gijon-Banos J (1990) Familial articular chondrocalcinosis in Spain. Ann Rheum Dis 49:531-535.

40. Mitrovic Dr (1983) Pathology of articular deposition of calcium salts and their relationship to osteoarthritis. Ann Rheum Dis (suppl) 42:19-26.

41. McCarty DJ (1977) Pyrophosphate dihydrate crystal deposition disease (pseudogout syndrome). Clinical aspects. Clin Rheum Dis 3:61-89.

42. Riestra JL, Sanchez A, Rodriquez-Valverde V, Castillo E, Calseron J (1985) Roentgenographic features of the arthropathy associated with CPPD crystal deposition disease. A comparative study with primary osteoarthritis. J Rheumatol 12:1154-1158.

43. Stretcher RM (1955) Heberden's nodes a clinical description of osteo-arthritis of the finger joints. Ann Rheum Dis 14:1-10.

44. Doherty M, Watt I, and Dieppe P. (1983) Influence of primary generalised osteoarthritis on development of secondary osteoarthritis. Lancet ii: 8-11.

45. Myers RH, Couropmitree NN, Chaissson CE, Hannan MT, Zhang Y, McAlindon T, LaValley M, Felson DT (1996) Relative heritability of Osteoarthritis (OA) of different hand and knee joint groups: The Framingham study. Arthritis Rheum 39(suppl):S170.

46. Spector TD, Cicuttini F, Baker J, Hart DJ (1995) Genetic influences on osteoarthritis: in women: a twin study. BMJ 312: 940-943.

47. Kaprio J, Kujala UM, Peltonen L, Koskenvuo M (1996) Genetic liability to osteoarthritis may be greater in women than men. BMJ 313:232.

48. Cooper C, Mc Alindon T, Snow S, Vines K, Young P, Kirwan J, and Dieppe P (1994) Mechanical and constitutional risk factors for symptomatic knee osteoarthritis: Differences between medial tibiafemoral and patellofemoral disease. J Rheumatol 21:307-313.

49. Chitnavis J, Clipsham K, Burge PD, and Carr AJ (1996) Familial tendency to severe symptomatic osteoarthritis of the hip and knee. Arthritis and Rheum 39(suppl):S170.

50. Palotie A, Väisänen P, Ott J, Ryhänen, Elima K, Vikkula M, Cheah K, Vuorio E, Peltonen L (1989) Predisposition to familial osteoarthrosis linked to type II collagen gene. Lancet I:924-927.

51. Knowlton RG, Katzenstein PL, Moskowitz RW, Weaver EJ, Malemud CJ, Pathria MN, Jimenez SA, Prockop DJ (1990) Genetic linkage of a polymorphism in the type II procollagen gene (COL2A1) to primary osteoarthritis associated with mild chondrodysplasia. New Eng J Med 322:526-530.

52. Bullough PG (1992) The pathology of osteoarthritis. In: Moskowitz RW, Howell DS, Goldberg VM, Mankin HJ eds. Osteoarthritis; Diagnosis and Medical/Surgical Management. WB Saunders company (1992) 2nd edition: 39-69.

53. Hardingham TE, Venn G, and Bayliss MT (1991) Chondrocyte responses in cartilage and in experimental osteoarthritis. Br J Rheum 30(suppl 1):32-37.

54. Hwang WS, Li B, Jin LH, Ngo, K, Schachar NS, and Hughes GNF (1992) Collagen fibril structure of normal, aging and osteoarthritic cartilage. J Pathol 167:425-433.

55. Maroudas A, Venn M (1977) Chemical composition and swelling of normal and osteoarthritic femoral head cartilage II. Swelling. Ann Rheum Dis 36:399-406.

56. Hoyland JA, Thomas JT, Donn R, Marriott A, Ayad S, Boot-Handford RP, Grant ME and Freemont AJ (1991) Distribution of type X collagen mRNA in osteoarthritic human cartilage. Bone Miner 15:151-164.

57. Sokoloff L (1987) Osteoarthritis as a remodelling process. J Rheumatol 14(suppl):7-10.

58. Cawston TE (1994) Metalloproteinases and connective tissue breakdown. Rheumatology Review 3:147-154.

59. Gerstenfeld LC, Finer MH, and Boedtker H (1989) Quantitative analysis of collagen expression in embryonic chick chondrocytes having different development rates J Biol Chem 264:5112-5120.

60. Wu JJ Woods PE and Eyre DR (1992) Identification of cross-linking sites in bovine cartilage type IX collagen reveals an antiparallel type II-type XI molecular relationship and type IX to type IX bonding. J Biol Chem 267:23007-23014.

61. van der Rest M, Mayne R (1988) Type IX collagen proteoglycan from cartilage is covalently cross-linked to type II collagen. J Biol Chem 263:1615-1618.

62. Thom JR, Morris NP (1991) Biosynthesis and proteolytic processing of type XI collagen in embryonic chick sterna. J Biol Chem 266:7262-7269.

63. Smith GN, Williams JM and Brandt KD (1985) Interaction of proteoglycans with the pericellular (1α, 2α, 3α) collagens of cartilage. J Biol Chem 260:10761-10767.

64. Franzén A, Björnrnsson S, Heinegård D (1981) Cartilage proteoglycan aggregate formation. Role of link protein Biochem J 197:669-674.

65. Perin JP, Bonnet F, Thurieau C, Jollès P (1987) Link protein interactions with hyaluronate and proteoglycans. Characterization of two distinct domains in bovine cartilage link protein. J Biol Chem 262:13269.

66. Muir IHM (1980) The chemistry of the ground substance of joint cartilage. In Sokoloff L, ed. The joints and synovial fluid. Vol II New York: Academic Press:27-94.

67. Chandresekhar S, Kleinman Hk, Hassell JR (1983) Interaction of link protein with collagen. J Biol Chem 258: 6226-6231.

68. Winterbottom N, Tondravi MM, Harrington TL, Klier FG, Vertel BM, Goetinck PRF (1992) Cartilage matrix protein is a component of the collagen fibril of cartilage. Dev Dyn 193:266-276.

69. Paulsson M, and Heinegård D (1982) Radioimmunoassay of the 148-kilodalton cartilage protein. Biochem J 207:201-213.

70. Mundlos S, Zabel B (1994) Developmental expression of human cartilage matrix protein. Dev Dynam 199:241-252.

71. Chen Q, Johnson DM, Haudenschild Dr, Goetinck PF (1995) Progression and recapitulation of the chondrocyte differentiation program: cartilage matrix protein is a marker for cartilage maturation. Dev Biol 172:293-309.

72. Newton G, Weremowicz S, Morton CC, Copeland NG, Gilbert DJ, Jenkins NA, Lawler J (1994) Characterization of human and mouse cartilage oligomeric matrix protein Genomics 24:435-439.

73. Cohn DH, Briggs MD, King LM, Rimoin DL, Wilcox WR, Lachman RS, Knowlton RG (1996) Mutations in the cartilage oligomeric matrix protein (COMP) gene in

pseudoachondroplasia and multiple epiphyseal dysplasia. Ann NY Acad Sci 785:188-194.

74. Prockop DJ (1990) Mutations that alter the primary structure of type I collagen. The perils of a system for generating large structures by the principle of nucleated growth. J Biol Chem 265:15349-15352.

75. Engel J, and Prockop DJ (1991) The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. Annu Rev Biophys Biophys Chem 20:137-152.

76. Kadler, KE, Hoijima Y, and Prockop DJ (1990) Collagen fibrils in vitro grow from pointed tips in the C- to N-terminal direction. Biochem J 268:339-343.

77. Chopra RD and Ananthanarayanan VS (1982) Conformational implications of enzymatic proline hydroxylation in collagen. Proc Natl Acad Sci USA 79:7180-7184.

78. Popenoe EA, Aronson RB, Van Slyke DD (1966) Hydroxylysine formation from lysine during collagen biosynthesis. Proc Natl Acad Sci USA 55:393-397.

79. Bornstein P, Ehrlich HP, and Wyke AW (1972) Procollagen: conversion of the precursor to collagen bya neutral protease. Science 175:544-546.

80. Han S, and Tanzer ML (1979) Collagen cross-linking. Purification of lysyl oxidase in solvents containing nonionic detergents. J Biol Chem 254:10438-10442.

81. Wu JJ, and Eyre DR (1984) Identification of hydroxypyridinium cross-linking sites in type II collagen of bovine articular cartilage. Biochemistry 23:1850-1857.

82. Kuivaniemi H, Tromp G, and Prockop DJ (1991) Mutation in collagen genes: cause of rare and some common diseases in humans. The FASEB journal 5:2052-2060.

83. Iwamoto M, Shimazu A, Nakashima K, Suzudi F, Kato Y (1991) Reduction in basic fibroblast growth factor receptor is coupled with terminal differentiation of chondrocytes J Biol Chem 266:461-467.

84. Kato Y and Iwamoto M (1990) fibroblast growth factor is an inhibitor of chondrocyte terminal differentiation. J Biol Chem 265:5903-5909.

85. Jentzsch KD, Wellmitz G, Heder G, Petzold E, Buntrock P, and Oehme P (1980) A bovine brain fraction with fibroblast growth factor activity inducing articular cartilage regeneration in vivo. Acta Biol Med Ger 39:967-971.

86. Tyler JA (1989) Insulin-like growth factor 1 can decrease degradation and promote synthesis of proteoglycan in cartilage exposed to cytokines. Biochem J 260:543-548.

87. Florini JR (1987) Hormonal control of muscle bulk. Muscle Nerve 10:577-598.

88. Kelly PF, Eisman JA, Stuart MC, Pocock NA, Sambrook PN, Gwinn TH (1990) Somatomedin-C, physical fitness and bone density. J Clin Endocrinol Metab 70:718-723.

89. Rudman D, Feller AG, Hoskote SN, Gergans GA, Lalitha PY, Goldberg AF (1990) Effects of growth hormone in men aged over 60 years old. N Engl J Med 323:1-6.

90. Schouten JSAG, Van den Ouweland FA, Valkenburg HA, Lamberts SWJ (1993) Insulin-like growth factor-1: a prognostic factor of knee osteoarthritis. Br J Rheumatol 32:274-280.

91. Denko CW, Boja B, Moskowitz RW (1990) Growth promoting peptides in osteoarthritis: insulin, insulin-like growth factor-1, growth hormone. J Rheumatol 17:1217-1221.

92. Hochberg MC, Lethbridje-Cejku M, Scott WW, Reichle R, Plato CC, Tobin JD (1994) Serum concentrations of insulin-like growth factor 1 in subjects with osteoarthritis of the knee. Data from the Baltimore Longitudinal Study of Ageing. Arthritis Rheum 37:1177-1180.

93. Lloyd ME, Hart DJ, Nandra D, McAlindon TE, Wheeler M, Doyle DV, and Spector TD (1996) Relation between insulin-like growth factor-1 concentration, osteoarthritis, bone density, and fractures in the general population: the Chingford study. Ann Rheum Dis 55:870-874.

94. Schouten JSAG, Van den Ouwenland FA, Valkenburg HA, Lamberts SWJ (1993) Insulin-like growth factor-1L: a prognostic factor of knee osteoarthritis. Br J Rheumatol 32:274-280.

95. Tyler JA (1985) Chondrocyte-mediated depletion of articular cartilage *in vitro.* Biochem J 225:493-507.

96. Gowen M, Wood DD, Ihrie EF, Meats JE, Russell RGG (1984) Stimulation by human IL1 of cartilage breakdown and production of collagenase and proteoglycanase by human chondrocytes but not by human osteoblast *in vitro.* BBA 797:186.

97. Pettipher ER, Higgs GA, Henderson B (1986) Interleukin 1 induces leucocyte infiltration and cartilage proteoglycan degradation in the synovial joint. Proc Natl Acad Sci USA 83:8749-8753.

98. Saklatvala J (1986) Tumour necrosis factor α stimulates resorption and inhibits synthesis of proteoglycan in cartilage. Nature 322:547-459.

99. Tyler JA (1985) Articular cartilage cultured with catabolic (pig interleukin 1) synthesizes a decreased number of normal proteoglycan molecules. Biochem J 227:869-878.

100. Goldring MB, Birkhead J, Sandell LJ, Kimura T, and Krane SM (1988) Interleukin 1 suppresses expression of cartilage-specific types II and IX collagens and increases types i and III collagens in human chondrocytes. J Clin Invest 82: 2026-2037.

101. Freije JMP, Diez-Itza I, Balbib M, Sanchez LM, Blasco R, Tolivia J, Lopez-Otin C (1994) Molecular cloning and expression of collagenase-3, a novel human matrix metalloproteinase produced by breast carcinomas. J Biol Chem 269:16766-16773.

102. Miller EJ, Harris ED, Chung E, Finch JE, McCroskery PA and Butler WT (1976) Cleavage of type II and III collagens with mammalian collagenase: site of cleavage and primary structure at the NH_2-terminal portion of the smaller fragment released from both collagens. Biochemistry 15:787-792.

103. Matrisian LM (1990) Metalloproteoinases and their inhibitors in matrix remodeling. TIG 6: 121-125.

104. Spranger J, Beighton P, Giedion A, Gorlin R, Hall J, Horton B, Kozlowski K, Lachman R, Langer LO, Maroteaux P, Poznanski A, Rimoin DL, Sillence D (1992) International classification of osteochondrodysplasias. Eur J Pediatr 151:407-415.

105. Spranger J, Winterpacht A, Zabel B (1994) The type II collagenopathies: a spectrum of chondrodysplasias. Eur J Pediatr 153:56-65.

106. Williams CJ, Jimenez SA (1995): Heritable diseases of cartilage caused by mutations in collagen genes. J Rheum 22:1(supplement) 28-33.

107. Ala-Kokko L, Baldwin, CT, Moskowitz, RW, and Prockop, DJ (1990) Single base mutation in the type II procollagen gene (COL2A1) as a cause of primary osteoarthritis associated with a mild chondrodysplasia. Proc Natl Acad Sci USA 87:6565-6568.

108. Holderbaum D, Malemud CJ, Moskowitz RW, & Haqqi TM (1993) Human cartilage from late stage familial osteoarthritis transcribes type II collagen mRNA encoding a Cysteine in position 519. BBRC 192:1169-1174.

109. Pun YL, Moskowitz RW, Lie S, Sundstrom WR, Block SR, Mcewen V, Williams HJ, Bleasel JF, Holderbaum D, Haqqi TM (1994): Clinical correlations of osteoarthritis associated with a single-base mutation (arginine^{519} to cysteine) in type II procollagen gene. Arthritis Rheum 37:264-269.

110. Williams CJ, Rock M, Considine E, Mccarron S, Gow P, Ladda R, Mclain D, Michels VM, Murphy W, Prockop DJ, Ganguly A (1995): Three new point mutations in type II procollagen (COL2A1) and identification of a fourth family with the COL2A1 Arg519->Cys base substitution using conformation sensitive gel electrophoresis. Hum Mol Genet 4:309-312.

111. Eyre DR, Weis MA, and Moskowitz RW (1991) Cartilage expression of a type II collagen mutation in an inherited form of Osteoarthritis associated with a mild chondrodysplasia J Clin Invest 87:357-361.

112. Vikkula M, Palotie A, Ritvaniemi P, Ott J, Ala-Kokko L, Sievers U, Aho K, Peltonen L (1993) Early-onset Osteoarthritis linked to the type II procollagen gene. Arthritis Rheum 36:401-409.

113. Ritvaniemi P, Körkkö J, Bonaventure J, Vikkula M, Hysland J, Paassilta P, Kaitila I, Kääriäinen H, Sokolov BP, Hakala M, Mannismäki P, Meerson EM, Klemola T, Williams C, Peltonen L, Kivirikko KI, Prockop DJ, Ala-Kokko L (1995): Identification of COL2A1 gene mutations in patients with chondrodysplasias and familial osteoarthritis. Arthritis Rheum 38:999-1004.

114. Cooper DN, and Krawczak M (1995) Human gene mutation. BIOS scientific Publishers Limited, p274-276.

115. Vandenberg P, Khillan JS, Prockop DJ, Helminen H, Kontusaari S, and Ala-Kokko L (1991) Expression of a partially deleted gene of human type II procollagen (COL2A1) in transgenic mice produces a chondrodysplasia. Proc Natl Acad Sci USA 88:7640-7644.

116. Helminen HJ, Kiraly K, Pelttari A, Tammi MI, Vandenberg P, Pereira R, Dhulipala R, Khillan JS, Ala-Kokko L, Hume EL, Sokolov BP, and Prockop DJ (1993) An inbred line of transgenic mice expressing an internally deleted gene for type II procollagen (COL2A1). J Clin Invest 92:582-595.

117. Briggs MD, Choi HC, Warman ML, Loughlin JA, Wordsworth P, Sykes B, Irven CMM, Smith M, Wynne-Davies R, Lipson MH, Biesecker LG, Garber AP, Lachman R, Olsen BR, Rimoin DL, Cohn DH (1994): Genetic mapping of a locus for multiple epiphyseal dysplasia (EDM2) to a region of chromosome 1 containing a type IX collagen gene. Am J Hum Genet 55:678-684.

118. Muragaki Y, Mariman ECM, van Beersum SEC, Perälä M, van Mourik JB, Warman ML, Olson BR, Hamel BCJ (1996) A mutation in the gene encoding the α2 chain of the fibril-associated collagen IX, COL9A2, causes multiple epiphyseal dysplasia (EDM2). Nature Genet 1:103-105.

119. Nakata K, Ono K, Miyazaki J-I, Olsen BR, Murakgaki Y, Adachi E, Yamamura K-I et al. (1993): Osteoarthritis associated with mild chondrodysplasia in transgenic mice expressing α1(IX) collagen chains with a central deletion. Proc Natl Acad Sci USA 90:2870-2874.

120. Fassler R, Schnegelsberg PNJ, Dausman J, Shinya T, Muragaki Y, McCarthy MT, Olsen BR (1994) Mice lacking α1(IX) collagen develop non-inflammatory degenerative joint disease. Proc Natl Acad Sci USA 91:5070-5074

121. Brunner HG, Beersum SEC, Warman ML, Olsen BR, Ropers HH, Mariman ECM (1994): A Stickler syndrome gene is linked to chromosome 6 near the COL11A2 gene. Hum Mol Genet 3:1561-1564.

122. Vikkula M, Mariman ECM, Lui VCH, Zhidkova NI, Tiller GE, Goldring MB, Beersum van SEC, de Waak Malefijt MC, Hoogen van den FHJ, Ropers HH, Mayne R, Cheah KSE, Olsen BR, Warman ML, and Brunner HG (1995) Autosomal dominant and recessive osteochondrodysplasias associated with the COL11A2 locus. Cell 80: 431-437.

123. Richard AJ, Yates JRW, Williams R, Payne SJ, Pope FM, Scott JD, and Snead MP (1996) A family with Stickler syndrome type 2 has a mutation in the COL11A1 gene resulting in the substitution of glycine 97 by valine in α1(XI) collagen. Hum Mol Genet 5:1339-1343.

124. Li Y, Lacerda DA, Warman ML, Beier DR, Yoshioka H, Ninomiya Y, Oxford JT, Morris NP, Andrikopoulos K, Ramirez F, Wardell BB, Lifferth GD, Teuscher C, Woodward SR, Taylor BA, Seegmiller RE, Olsen BR (1995): A fibrillar collagen gene COL11A1 is essential for skeletal morphogenesis. Cell 80:423-430.

125. Seegmiller RE, Fraser FC and Sheldon H (1971) A new chondrodystrophic mutant in mice. J Cell Biol 48:580-593.

126. Garofalo S, Vuorio E, Metsaranta M, Rosati R, Toman D, Vaughan J, Lozano G, Mayne R, Ellard J, Horton W, and Crombrugghe de B (1991) Reduced amounts of cartilage collagen fibrils and growth plate anomalies in transgenic mice harboring a glycine-to-cysteine mutation in the mouse type II procollagen α1-chain gene. Proc Natl Acad Sci USA 88:9648-9652.

127. Vandenberg P, Khillan J, Prockop DJ (1994) Minigene and gene knock-out experiments of the COL11A2 gene in transgenic mice. Matrix Biol 14:415 (Abstract).

128. Warman ML Abbott M, Apte SS, Hefferon T, McIntosh I, Cohn DH Hecht JT, Olsen FB and Francomano CA (1993) A type X collagen mutation causes Schmid metaphyseal chondrodysplasia. Nat Genet 5: 79-82.

129. Wallis GA, Rash B, Sweetman WA, Thomas JT, super M, Evans G, Grant ME, and Boot-Handford (1994) Amino acid substitutions of conserved residues in the Carboxyl-terminal domain of the α1(X) chain of type X collagen occur in two unrelated families with metaphyseal chondrodysplasia type Schmid. Am J Hum Genet 54:169-178.

130. McIntosh I, Abbott MH, Warman ML, Olsen GR, Francomano CA (1994) Additional mutations of type X collagen confirm COL10A1 as the Schmid metaphyseal chondrodysplasia locus. Hum Mol Genet 3:303-307.

131. Jacenko O, LuValle PA, and Olsen BR (1993) Spondylometaphyseal dysplasia in mice carrying a dominant negative mutation in a matrix protein specific for cartilage-to-bone transition. Nature 365: 56-61.

132. Kwan APL, Cummings CE, Chapman JA, and Grant ME (1991) Macromolecular organization of chicken type X collagen in vitro. J Cell Biol 114: 597-604.

133. Baldwin CT, Farrer LA, Adair R, Dharmavaram R, Jimenez, S, Anderson L (1995): Linkage of early-onset osteoarthritis and chondrocalcinosis to human chromosome 8q. Am J Hum Genet 56:692-697.

134. Hughes AE, McGibbon D, Woodward E, Dixey J, and Doherty M (1995) Localisation of a gene for chondrocalcinosis to chromosome 5p. Hum Mol Genet 4:1225-1228.

135. Rimoin DL, Rasmussen IM, Briggs MD, Roughley PJ, Gruber HE, Warman ML, Olsen BR, Hsia YE, Yuen J, Reinker K, Garber AP, Grover J, Lachman RS, Cohn DH (1994): A large family with features of pseudoachondroplasia and multiple epiphyseal dysplasia: exclusion of seven candidate gen loci that encode proteins of the cartilage extracellular matrix. Hum Genet 93:236-242.

136. Stanescu V, Pham T, Caminade F, Maroteaux P, and Stanescu R (1994) Non-collagenous protein screening in the human chondrodysplasias: Link proteins, cartilage oligomeric matrix protein (COMP) and Fibromodulin. Am J Med Genet 51:22-28.

137. Loughlin J, Irven C, Sykes B (1994) Exclusion of the cartilage link protein and the cartilage matrix protein as the mutant loci in several heritable chondrodysplasias Hum Genet 94: 698-700.

138. Briggs MD, Hoffman SMG, King LM, Olsen AS, Mohrenweiser H, Leroy JG, Mortier GR, Rimoin DL, Lachman RS, Gaines ES, Cekleniak JA, Knowlton RG, Cohn DH (1995): Pseudoachondroplasia and multiple epiphyseal dysplasia due to mutations in the cartilage oligomeric matrix protein gene. Nature Genet. 10:330-336.

139. Finkelstein JE, Doege K, Yamada Y, Pyeritz RE, Graham JM jr, Moeschler JB< Pauli RM, Hecht JT< and Francomano CA (1991) Analysis of the chondroitin sulfate proteoglycan core protein (CSPGCP) gene in achondroplasia and pseudoachondroplasia. Am J Hum Genet 48:97-102.

140. Kimata K, Barrach H-J, Brown KS, and Pennypacker JP (1981) Absence of proteoglycan core protein in cartilage from the cmd/cmd (Cartilage Matrix Deficiency) mouse. J Biol Chem 256:6961-6968.

141. Hyland J, Ala-Kokko L, Royce P, Steinmann B, Kivirikko KI, and Myllylä R (1992) A homozygous stop codon in the lysyl hydroxylase gene in two siblings with Ehlers-Danlos syndrome type VI. Nature Genet 2:228-231.

142. Pousi B, Hautala T, Heikkinen J, Pajunen L, Kivirikko KI, and Myllylä R (1994) Alu-Alu recombination results in a duplication of seven exons in the Lysyl Hydroxylase gene in a patient with the type VI variant of Ehlers-Danlos Syndrome. Am J Hum Genet 55:899-906.

143. Ha VT, Marshall MK, Elsas LJ, Pinnell SR, and Yeowell HN (1994) A patient with Ehlers-Danlos syndrome type VI is a compound heterozygote for mutations in the Lysyl Hydroxylase gene. J Clin Invest 93:1717-1721.

144. Kuivaniemi H, Peltonen L, and Kivirikko KI (1985) Type IX Ehlers-Danlos syndrome and Menkes syndrome: The decrease in Lysyl Oxidase activity is associated with a corresponding deficiency in the enzyme protein. Am J Hum Genet 37:789-808.

145. Walton M (1977) Degenerative joint disease in the mouse knee: histological observations. J Pathol 123:109-122.

146. Hull R, and Pope FM (1989) Osteoarthritis and cartilage collagen genes. The Lancet 1:1337-1338.

147. Priestley L, Fergusson C, Ogilvie D, Wordsworth P, Smith R, Pattric M, Doherty M, and Sykes B (1991) A limited association of generalized osteoarthritis with alleles at the type II collagen locus: COL2A1. Brit J Rheum 30:272-275.

148. Loughlin J, Irven C, Athanasou N, Carr A, and Sykes B (1995) Differential allelic expression of the type II collagen gene (COL2A1) in osteoarthritic cartilage. Am J Hum Genet 56:1186-1193.

149. Berg ES, Olaisen B (1993) Characterization of the COL2A1 VNTR polymorphism. Genomics 16:350-354.

150. Vikkula M, Nissilä M, Hirvensalo E, Nuotio P, Palotie A, Aho K, Peltonen L. (1993): Multiallelic polymorphism of the cartilage collagen gene: no association with osteoarthrosis. Ann Rheum Dis 52:762-764.

151. Horton W, Balakir R, Precht P, Plato CC, Tobin JD, Lethbridge-Cejku M Hochberg MCD Meek L, and Doege K (1996) Association between an aggrecan polymorphic allele and bilateral hand osteoarthritis in elderly white men: data from the Baltimore longitudinal study of Aging (BLSA). Arthritis Rheum 39(suppl):S167.

152. Sakkas LI, Macfarlane DG, Bird H, Welsh KI and Panayi GS (1990) Association of osteoarthritis with homozygosity for a 5.8 kb Taq I fragment of the α1-antichemotrypsine gene. Brit J Rheum 29: 245-248.

153. Loughlin J, Irven C, Fergusson C, and Sykes B. (1994) Sibling pair analysis shows no linkage of generalized osteoarthritis to the loci encoding type II collagen, cartilage link protein, or cartilage matrix protein. Brit J Rheum 33:1103-1106.

154. Palotie A, Väisänen P, Ott J, Ryhänen, Elima K, Vikkula M, Cheah K, Vuorio E, Peltonen L. Predisposition to familial osteoarthrosis linked to type II collagen gene. *Lancet* 1989;**1**:924-927.

155. Vikkula M, Palotie A, Ritvaniemi P, Ott J, Ala-Kokko L, Sievers U, Aho K, Peltonen L (1993) Early-onset osteoarthritis linked to the type II procollagen gene. Arthritis Rheum 36:401-409.

156. Long JC, Wiliams RC and Ubanek M (1995) An E-M Algorithm and testing strategy for multiple-locus haplotypes. Am J Hum Genet 56:799-810.

157. Franzén A, Björnsson S, Heinegård D: Cartilage proteoglycan aggregate formation. Role of link protein Biochem J 197:669-674, 1981

158. Winterbottom N, Tondravi MM, Harrington TL, Klier FG, Vertel BM, Goetinck PRF: Cartilage matrix protein is a component of the collagen fibril of cartilage. Dev Dyn 193: 266-276, 1992

159. Wang Y, Sadler L, Hecht JT: Polymorphic dinucleotide repeat in a cartilage matrix protein (CRTM) Hum Mol Genet 1:780, 1992

160. Hecht JT, Wang Y, Rhodes C, Yamada Y: GT repeat polymorphism in the human proteoglycan link gene promoter region Nucl Acid Res 19:6666, 1991

2 HIGH YIELD NONINVASIVE HUMAN GENOMIC DNA ISOLATION METHOD FOR GENETIC STUDIES IN GEOGRAPHICALLY DISPERSED FAMILIES AND POPULATIONS

Ingrid Meulenbelt[1], Simone Droog[1], Gerjan J.M. Trommelen[2], Dorret I. Boomsma[3] and P. Eline Slagboom[1].

[1]TNO-PG, Gaubius laboratory, Dept. of Vascular and Connective Tissue Research, P.O box 2215, 2301 CE Leiden, The Netherlands. [2]Gene Pharming Europe bv., Niels Bohrweg 11-13, 2333 CA Leiden, The Netherlands. [3]Dept. of. Physiology, Free University, De Boelelaan 1111, 1081 HV Amsterdam, The Netherlands.

Reproduced from

The American Journal of Human Genetics, 1995, 57:1252-1254

To the editor:

Human genomic DNA is commonly isolated from peripheral blood samples for genetic studies of families and populations. Blood sampling, however, is expensive and an invasive procedure to which, for ethical reasons, objections may be raised, especially in studies involving old individuals and babies. We have developed a new noninvasive DNA sampling and isolation method using mouth swabs with cotton bud sticks. Participants can take mouth swabs themselves, and can send these to the research centre by mail where DNA can be isolated at least up to 3 weeks after sampling. DNA isolation from 20 cotton bud sticks resulted in an average yield of 40 µg of high molecular weight DNA per individual, sufficient for complete genome searches with approximately 800 polymorphic DNA markers using PCR. Compared to blood sampling, involving clinically trained personnel, this procedure is fast, less expensive, and suitable especially for DNA collection from geographically scattered subjects.

We have isolated human genomic DNA of family members (ages between 4 and 72 years), young twins (ages between two months and 5 years), and random controls by using mouth swabs. Mouth swabs of the family members and random controls were taken by the participants themselves, and parents took mouth swabs of their twins, following a written protocol. Although twin-pairs were sometimes only a few months old, problems were not encountered by parents obtaining samples from their children. At least 10 consecutive samples can be taken from one subject without a significant loss in yield per cotton bud. A second round of 10 cotton buds can be taken after approximately 4 hours, giving a maximal result of 20 samples in one day. The mouth swab sample should be taken in a clean mouth without food remains. After rubbing, the cotton bud, sample-end first, should be placed in a Falcon tube, containing 0.5 ml of STE buffer (100 mM NaCl, 10 mM TrisHCL (ph 8,0) and 10 mM EDTA) with proteinase K (0.2 mg/ml) and SDS (0.5%) per cotton bud stick. In this way samples can be kept at least three weeks without decrease in yield or quality of DNA.

High molecular weight human genomic DNA (> 50 kb) was isolated with an average yield of 2.1±0.2 µg per cotton bud stick from 81 individuals using mouth swab samples which were placed in lysis buffer immediately after wiping. DNA isolations of another 181 individuals were performed with mouth swab samples which were kept dry after wiping, that is without lysis buffer. High molecular weight human genomic DNA was also isolated from these individuals. The average DNA yield of these samples, however, was 1.3±0.05 µg per cotton bud stick. Mouth swab samples which were kept dry for more than 7 days before isolation gave an extra

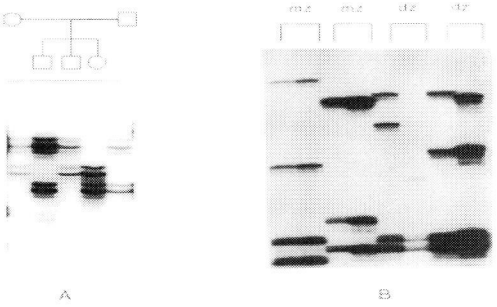


Figure 1 To isolate human genomic DNA, Falcon tubes containing the cotton bud sticks and STE buffer with proteinase K and SDS, are placed in a 65°C waterbath for two hours. To collect a maximal amount of buffer from the soaked cotton buds after lysis, sticks are placed in a syringe which is placed upside-down in the Falcon tube. Falcon tubes containing the syringes are centrifuged for 5 minutes at 1000 rpm in a Beckman centrifuge. Genomic DNA is subsequently isolated from the collected lysis buffer using phenol/ chloroform/ isoamylalcohol (24:24:1) and chloroform/ isoamylalcohol (24:1) extractions followed by isopropanol precipitation as described by Sambrook et al. (1990). RNAase treatment was necessary to remove low molecular weight RNA bands and smears. PCR reactions were performed in 25 µl containing 25-50 ng genomic DNA, 2.5 pmoles of each primer, 1xTaq buffer (Sphaero Q), 2 µCi $\alpha[^{32}P]$-dCTP, 200 µM each of dCTP, dGTP, dTTP, and dATP, and 0.05 U of Super Taq DNA polymerase. Amplification was initiated with 3 min denaturation at 94°C followed by 35 cycles of 15 sec at 94°C, 30 sec at 60°C, 30 sec at 72°C. The amplification was finished by a final incubation at 72°C for 3 min. Alleles were separated by electrophoresis through a denaturing 6% polyacrylamide gel, and analyzed by autoradiography. A, Autoradiograph showing dinucleotide polymorphism D6S276 screened in a subset of a family using genomic DNA isolated from mouth swabs. B, Autoradiograph showing multiplex PCR of four twin pairs (two monozygotic, and two dizygotic) using dinucleotide repeats ACTBP2 with D21S11.

decrease in DNA yield. The storage of cotton bud sticks in lysis buffer after wiping is therefore the better procedure. Variation of DNA yield within each method is primarily caused by the difference in pressure exerted during the mouth swab sampling.

Of 262 DNA samples isolated using mouth swabs (both dry and wet) 257 were successfully used in PCR reactions of 20 different human loci. Five DNA samples, kept without lysis buffer, did not work in the PCR. Figure 1A shows dinucleotidepolymorphism D6S276 (Gyapay et al. 1994) which was screened in a family. DNA samples of baby twin pairs were used in multiplex PCR reactions (Kimpton et al. 1993) to determine zygosity. Two different multiplex PCR reactions, ACTBP2 (Polymeropoulos et al. 1992) with D21S11 (Sharma et al. 1992) (Fig. 1B) and D15S221 (Allamand et al. 1994) with D11S898 (Gyapay et al. 1994) and CSF1R (Hastbacka et al. 1992), were used. Together, these two multiplex PCR reactions determine the zygosity of each twin pair with 99,96% certainty.

We have used phenol/chloroform extractions to isolate uncontaminated genomic DNA, without yeast sporule or bacteria that can be stored for many years and can be used also for other DNA analysis methods than PCR. Other DNA isolation procedures eliminating phenol/ chloroform

extractions as described, for example, by Richards et al. (1993) might be more efficient and cheap. These methods may be tested in the near future.

The presence of contamination with foreign genomic DNA such as yeast or bacteria might result in extra bands or completely non-matching bands between monozygotic twin pairs. Using genomic DNA isolated from mouth swab samples in PCR reactions of 20 known human loci and in the zygosity determination of 50 twin pairs, we found no evidence of contamination.

This fast and cost effective method is being used in our laboratory in a genetic linkage study, various genetic population studies, and in zygosity determination of twin pairs.

Acknowledgement

We would like to thank Saskia de Wildt and Daphne van Leeuwen for technical support. This work was supported by the Dutch Prevention Fund and "Het Nationaal Reumafonds" of the Netherlands.

References

- Allamand V, de-Souza AP, Richard I, Brenguier L, Beckmann JS (1994) Dinucleotide repeat polymorphism at D15S221. Hum Mol Genet 3:382
- Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, Bernardi G, Lathrop M, and Weissenbach J (1994) The 1993-94 Généthon human genetic linkage map. Nature Genet 7:246-339
- Hastbacka J, Chapelle de la A, Kaitila I, Sistonen P, Weaver A, Lander E(1992) Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nature Genet 2:204-211
- Kimpton CP, Gill P, Walton A, Urquhart A, Millicam ES, and Adams M (1993) Automated DNA profiling employing multiplex amplification of short tandem repeat loci. PCR methods and Applications 3:13-22
- Polymeropoulos MH, Rath DS, Xiao H, and Merril CR (1992) Tetranucleotide repeat polymorphism at the human beta-actin related pseudogenes H-beta-Ac-psi-2 (ACTBP2). Nucleic Acids Res 20:1432
- Richards B, Skoletsky J, Shuber AP, Balfour R, Stern RC,Korkin HL, Parad RB, Witt D, and Klinger KW (1993) Multiplex PCR amplification from the CFTR gene using DNA prepared from buccal brushes/swabs. Hum Mol Genet 2:159-163
- Sambrook J, Fritsch EJ, Maniatis T (ed) (1990) Molecular Cloning: A laboratory Manual. Vol 2, 2nd Cold Spring Harbor, Cold Spring Harbor Laboratory Press 1989
- Sharma V and Litt M (1992) Tetranucleotide repeat polymorphism at the D21S11 locus. Hum Mol Genet 1:67

3 GENETIC LINKAGE ANALYSIS OF 14 CANDIDATE GENE LOCI IN A FAMILY WITH AUTOSOMAL DOMINANT OSTEOARTHRITIS WITHOUT DYSPLASIA

I Meulenbelt[1,3], C Bijkerk[1,2], FC Breedveld[3] and PE Slagboom[1].

[1] TNO-PG, Gaubius Laboratory, Dept. of Vascular and Connective Tissue Research, P.O box 2215, 2301 CE Leiden, The Netherlands [2] Dept. of Epidemiology and Biostatistics, Erasmus University Medical School Rotterdam, The Netherlands [3] Dept. of Rheumatology and Dept. of Internal Medicin, University Hospital Leiden, The Netherlands.

Accepted for publication

Journal of Medical Genetics

Abstract

The role of various gene loci was investigated in a family in which familial osteoarthritis (FOA), occurring at an early age of onset, is transmitted as an autosomal dominant mendelian trait. The absence of clinical and roentgenographic signs of dysplasia and calcium pyrophosphate deposition disease (CPDD) indicates that the basic disease process in this family is osteoarthritis (OA). Genetic linkage analysis of 14 candidate genes resulted in the exclusion of 10 important genes (COL2A1, COL9A1, COL9A2, COL11A1, COL11A2, COMP, the CPDD region, CRTL-1, CRTM and MMP3). Other relevant genes were not informative in this family. The candidate loci previously identified in FOA and heritable skeletal disorders associated with OA are clearly not involved in the development of the primary FOA phenotype of the family investigated, indicating genetic heterogeneity.

Osteoarthritis (OA) is a degenerative disease of the joints, characterized by degradation of the hyaline articular cartilage and remodelling of the subchondral bone with sclerosis. Genetic factors play a role in the etiology of familial OA (FOA) occurring at an early age of onset (20-40 year) in multiple joints.

Research into the genetic loci involved in FOA has thusfar mainly been focused on the type II procollagen gene (COL2A1). The COL2A1 gene encodes the major stress resisting element and the most abundant structural collagenous protein in cartilage [1]. Genetic studies of families and unrelated patients with an FOA phenotype accompanied by mild spondylo-epiphyseal-dysplasia (SED), have demonstrated linkage to COL2A1 and mutations in the gene in approximately 25% of the cases [2]. However, only a few studies have been focused on families where OA is the only and primary disease process without any dysplasia. Genetic studies demonstrated positive linkage to COL2A1 in 2 such FOA families [3] [4]. Mutation analysis of the COL2A1 gene in 45 unrelated FOA patients showed only one COL2A1 mutation [5]. Studies on other candidate genes have not been reported for FOA. Genes, however, identified in heritable skeletal disorders associated with generalized OA may play a possible role also in FOA (Table 1).

We have investigated the role of 14 candidate genes for FOA in a four-generation Dutch family of Jewish descent in which the pedigree included 21 persons (Figure 1). Informed consent and complete medical history was obtained for all family members. The study was approved by the Medical Ethics Committee of the Academic Hospital Leiden. Physical examinations were performed on all members of the third generation and on 7 our of 14 direct descendants in the fourth

Table 1: candidate genes with associated disease used for linkage analysis.

Gene	*Associated disease*	*Reference*
COL2A1	several types chondrodysplasia, FOA*, Stickler	[2-7]
COL9A1	OA, mild chondrodysplasia in mice	[8] [9]
COL9A2	EDM2†	[10] [11]
COL9A3	-	-
COL11A1	Stickler, chondrodysplasia in mice	[12] [13]
COL11A2	Stickler/Kniest and OSMED‡	[14] [15]
DCN	-	-
CRTL1	-	-
COMP	PSACH¶, EDM1§	[16] [17]
CRTM	-	-
CCAL2	familial CPDD	[18]
LOX	Ehlers Danlos type IX	[19]
PLOD	Ehlers Danlos type VI	[20] [21]
MMP3	-	-

COL2A1 = α1-collagen type II gene. COL9A1 = α1-collagen type IX gene. COL9A2 = α2-collagen type IX gene. COL9A3 = α3-collagen type IX gene. COL11A1 = α1-collagen type XI gene. COL11A2 = α2-collagen type XI gene. DCN = decorine gene. CRTL-1 = cartilage link protein gene. COMP = cartilage oligomeric protein gene. CRTM = cartilage matrix gene. CCAL2 = locus of familial chondrocalcinosis. LOX = lysyl oxidase gene. PLOD = lysyl hydroxylase gene. MMP3 = stromelysin I gene. *FOA = familial osteoarthritis. †EDM2 = multiple epiphyseal dysplasia type 2. ‡OSMED = osteochondrodysplasias ¶PSACH = pseudoachondrodysplasia. §EDM1 = multiple epiphyseal dysplasia type 1. ¥CPDD = calcium pyrophosphate dihydrate deposition disease.

generation. All clinical evaluations and diagnostic decisions were made prior to the genetic linkage analyses.

The family included multiple members affected with OA at an age of onset between 20 and 40 years. Symptoms began with intermittent acute pain and swelling in one or both knee joints with subsequent OA development in other joints. The hip was only rarely affected (Figure 1 and Table 2). Roentgenographic signs of chondrodysplasia, spinal dysplasia or abnormal development of theepiphyses of the peripheral joints were absent (Figure 2). Some individuals had marked Heberden's nodes (Table 2; Figure 3). The mean ratio of upper to lower segment of affected members was 0.95 (range 0.89 to 1.02), indicating the absence of a short trunk form of dysplasia. All family members had a normal stature (Table 2). The diagnosis was "pure" familial OA transmitted as an autosomal dominant mendelian trait (Figure 1 and Table 2).

Table 2. Roentgenographic and clinical abnormalities in members of a family with autosomal dominant FOA

Patient (age)	♂/♀	DIP*	PIP†	MCP‡	CMC1¶	Elbow	Shoulder	Hip	Knee	Foot	Ankle	CS§	LS¥	No.¤	BMI○	Height	U/L□
II-2(-)	♂	-	-	-	-	OA	-	N	OA,C	OA	OA			4			
III-1(73)	♂	P,H	-	-	OA	OA,C	OA	-	OA,C	OA	OA	OA	OA	9	26.8	1.76	1.02
III-3(70)	♂	OA,H	OA,B		OA	P	-	-	OA,C	-	OA	OA	-	9	24.3	1.64	0.96
III-5(67)	♂	-	-	-	-	-	-	N	-	-	-	OA	OA	2			
IV-1(42)	♀	N	N	N	OA	-	-	-	OA	-	N	OA	OA	4	26.3	1.70	0.83
IV-2(41)	♀				N	N	-	-	OA,C	-	P	OA	-	3	23.4	1.68	0.95
IV-3(39)	♂	-	-	-	-	-	-	-	-	-	-	-	-	0		1.76	
IV-4(37)	♂	N	N	N	N	-	-	N	N	-	-	-	-	0		1.80	
IV-5(35)	♂	-	-	-	-	-	-	N	N	-	-	-	-	0		1.76	
IV-6(42)	♀	N	N	N	OA	-	-	N	OA	-	-	OA		3			
IV-7(40)	♀	-	-	-	-	-	-	-	-	-	-	OA	-	1			
IV-8(38)	♂	-		-	-	N-	-	-	OA,C	OA	N	OA	N	3	27.5	1.71	0.91
IV-9(27)	♂	-		-	-		-	-	-	-	-	-	-	0	22.4	1.83	0.91

*DIP = distal interphalangeal joints †PIP = proximal interphalangeal joints ‡MCP = metacarpophalangeal joints ¶CMC1 = first carpometacarpal joints §CS = cervical spine joints ¥LS = lumbar spine joints ¤No.=Number of joints affected with roentgenological and/or clinical OA; ○BMI=Body mass index □U/L=upper/lower segment ratio. - = X-ray photo has not been made, no clinical signs. OA=roentgenological osteoarthritis according to Kellgren (grade 2 or higher) in one or both joints N=roentgenological normal joint on both sides (Kellgren 0 or 1). Ç=roentgenological chondrocalcinosis. P=clinical signs of OA (bony enlargements and/or joint deformity and/or limited range of motion) H=Heberden's nodes. B=Bouchard's nodes.

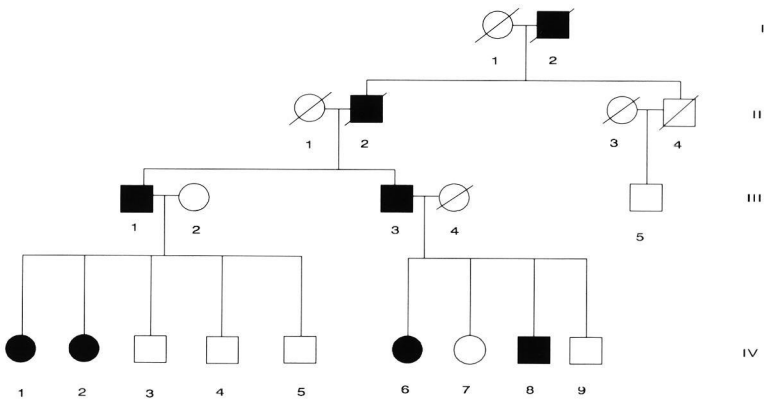


Figure 1. Pedigree of the FOA family used for linkage analysis. The diagonally filled symbol represents the individual with diagnostic uncertainty, the filled symbols represent affected individuals.

Genetic linkage analysis was performed using either intragenic or closely linked markers to the candidate genes (Table 3). Genomic DNA samples of family members were collected by mouth swabs [37]. All markers, except D20S19, were analyzed by genomic PCR containing $\alpha[^{32}P]$-dCTP [38]. Alleles were separated by standard electrophoresis through a denaturing polyacrylamide gel (3.5-6%) and visualized by autoradiography [38]. Alleles of marker D20S19 were analysed by Southern blotting analysis, and hybridization with the clone pCMM20 radioactively labelled with $\alpha[^{32}P]$-dCTP [38].

Two point LOD scores between the disease phenotype of family members and the markers, were calculated using *MLINK* from the LINKAGE package version 5.1 [39]. Multipoint LOD score analysis was performed using the *LINKMAP* program [40]. The disease locus was modeled as an autosomal dominant trait. Individuals with clinical and roentgenographic evidence of OA in two or more joints were considered affected. Penetrance was modelled to rise linearly from 0% at age 15 to 100% at 40 years, based on the onset age of

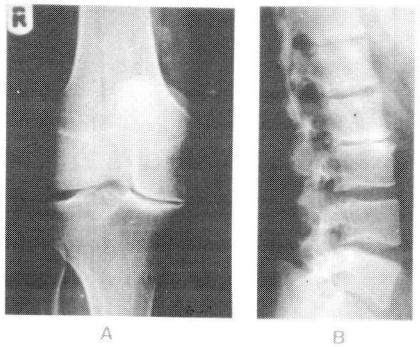


Figure 2. Roentgenographs of individual III-1 at age 53 years A) right knee joint demonstrating marked OA medially and chondrocalcinosis laterally, without evidence of epiphyseal dysplasia. B) lumbosacral spine joints (lateral view) demonstrating osteophytosis and discus degeneration at level L2-L3. No dysplastic features.

OA similarly as described previously [3]. Individuals with ages of onset above 40 years were considered as phenocopies.

LOD scores and the recombination fractions yielding a LOD score of -2 are shown in Table 4. Since the LOD-scores over the genetic region covered by the COL2A1, COL9A1, COL9A2, COL11A1, CRTL-1, COMP, CPDD, and MMP3 genes are $\leq$ -2 (odds 100:1 against linkage) these genes were excluded from being involved in causing FOA in this family. The COL9A3, COL11A2, DCN, CRTM, LOX and PLOD genes could not be excluded by two point linkage analysis (Table 4). The COL11A2 and CRTM genes were excluded (LOD scores $\leq$ -2; Table 5) by using multipoint LOD score analysis with fixed markers and known recombination fractions (Table 3). For the COL9A3 and PLOD gene the excluded region surrounding the highly polymorphic markers was too small to exclude the genes.

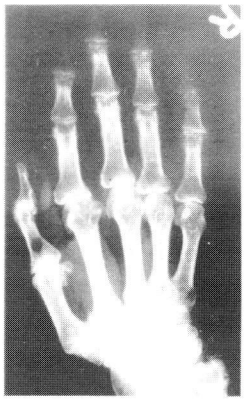


Figure 3. Roentgenograph of individual III-3 at age 65 years of hand joints demonstrating marked OA in DIP, PIP, CMC1, MCP-I and II. Heberden and Bouchard nodes are visible.

Table 3: Chromosomal location of candidate genes with polymorphic markers used for linkage analysis.

Gene	*Location*	*Marker*	*Distance (Kosambi cM) between marker and gene*	*Reference*
COL2A1	12q12-q13.2	VNTR	0 KcM	[22]
COL9A1	6q12-q14	8B1	intragenic	[23]
COL9A2	1p32.3-p33	MYCL1	0.2 KcM	[24] [25]
COL9A3	20q13.3	D20S19	5 KcM	[26] [27]
COL11A1	1p21	7B1	intragenic	-
COL11A2	6p21.3	TNF locus; D6S291	map element 3.5 KcM	[28-30]
DCN	12q21.3-q23	dinucleotide intron 1A	intragenic	[31]
CRTL1	5q13-q14.1	dinucleotide promoter	intragenic	[32]
COMP	19q12	D19S212	0.8 KcM	[29]
CRTM	1p35	dinucleotide 3' UTR D1S247; D1S513	intragenic map element 2 KcM	[33]
CCAL2	8q	D8S545	0 KcM	[29]
LOX	5q23.3-q31.2	RFLP exon 1 (Pst I)	intragenic	[34]
PLOD	1p36.3-p36.2	FRG (1p36.2-p36.1)	15 KcM	[35]
MMP3	11q22-q23	D11S35	2 KcM	[36]

*personal communication M. Warman.

Table 4: Two point LOD scores calculated between familial osteoarthritis (FOA) and markers within or flanking the candidate gene loci.

Genes	*Recombination fraction (Θ)*								*Exclusion (KcM)1**
	0.00	*0.001*	*0.01*	*0.05*	*0.1*	*0.2*	*0.3*	*0.4*	
COL2A1	- ∞	-2.01	-1.02	-0.39	-0.17	-0.03	-0.00	-0.01	0.1
COL9A1	- ∞	-1.96	-0.97	-0.31	-0.08	0.05	0.04	-0.00	0.1
COL9A2	- ∞	-4.86	-2.86	-1.47	-0.88	-0.35	-0.11	-0.01	2.7
COL9A3	- ∞	-1.86	-0.87	-0.20	0.04	0.18	0.15	0.06	0.1
COL11A1	- ∞	-2.85	-1.85	-1.14	-0.81	-0.44	-0.20	-0.06	0.7
COL11A2	- ∞	-1.81	-0.81	-0.14	0.10	0.21	0.24	0.24	0.1
DCN	0.41	0.41	0.40	0.37	0.31	0.20	0.10	0.03	-
CRTL1	- ∞	-4.29	-2.31	-0.99	-0.49	-0.11	0.01	0.03	1.4
COMP	- ∞	-4.90	-2.90	-1.51	-0.94	-0.41	-0.16	-0.04	2.8
CRTM	-0.01	-0.00	-0.00	0.01	0.01	0.01	0.01	0.00	-
8q	- ∞	-3.38	-2.33	-1.42	-0.92	-0.40	-0.14	-0.02	1.9
LOX	0.36	0.36	0.35	0.30	0.25	0.15	0.07	0.02	-
PLOD	- ∞	-5.05	-3.05	-1.63	-1.02	-0.45	-0.17	-0.04	3.3
MMP 3	- ∞	-4.85	-2.85	-1.47	-0.90	-0.39	-0.15	-0.04	2.7

* The excluded distance is calculated with odds of 100:1 against linkage in Kosambi centiMorgans (KcM).

Table 5: Multipoint LOD score analysis of disease locus (1) versus map element of polymorphic markers (2 and 3).

COL11A2 1=2=3	*LOD-score**	*COL11A2 2=1=3*	*LOD-score**
.500 .035	0.00	.000 .035	- ∞
.400 .035	0.05	.007 .028	-2.56
.300 .035	0.17	.014 .022	-2.08
.200 .035	0.26	.021 .015	-2.08
.100 .035	0.22	.028 .007	-2.26
.000 .035	-15.12	.035 .000	- ∞
CRTM 1=2=3	*LOD-score**	*CRTM 2=1=3*	*LOD score**
.500 .020	0.00	.000 .020	- ∞
.400 .020	0.00	.004 .016	-3.83
.300 .020	- 0.10	.008 .012	-3.23
.200 .020	- 0.36	.012 .008	-2.88
.100 .020	- 0.96	.016 .004	-2.63
.000 .020	-32.22	.020 .000	-2.43

*computed by division of the location scores with 2 ln (10)

The markers located within the DCN and LOX gene (Table 3) were not informative in important meioses. These genes could, therefore, not be excluded. The COL9A3 and DCN loci are mapped in regions containing too few alternative polymorphic markers and the genetic location of the LOX and PLOD gene is not well defined.

In summary, ten genes were excluded from involvement in FOA in this family. Among these loci were important candidate genes involved in several heritable skeletal disorders, mild chondrodysplasia and epiphyseal dysplasia associated with early onset OA in multiple joints (COL2A1, COL9A1, COL9A2, COL11A1, COL11A2, COMP, and the CPDD region). Other possible candidate genes encoding non-collagenous structural components, or genes involved in posttranslational modification and remodelling of cartilage were also excluded (CRTL-1, CRTM and MMP3). The COL9A3, DCN, LOX and PLOD genes could not be excluded as the cause of FOA in our family.

We have demonstrated once more that FOA is genetic heterogeneous and that ten important OA candidate genes are not involved in the pathogenesis of an FOA phenotype without dysplasia. Unidentified genes may be detected in future studies of this and other families for which FOA is the primary disease process.

We would like to thank Dr. JK van der Korst, Dr. PJJM Rompa, and Prof. Dr. JJ Rasker for providing us with the roentgenographs and medical histories of their patients. Saskia de Wildt for technical support and Dr. CJ Williams for helpfull discussions. This work was supported by the TNO-Health Organization, the "Dutch

League against Rheumatism" of the Netherlands and Looscofonds.

1 Prockop DJ, Kivirikko KI. Collagens: Molecular biology, diseases, and potentials for therapy. *Annu Rev Biochem* 1995;**64**:403-34.
2 Williams CJ, and Jimenez SA. Heritable diseases of cartilage caused by mutations in collagen genes. *J Rheumatol* 1995;**22(suppl 43)**:28-33.
3 Palotie A, Väisänen P, Ott J, Ryhänen, Elima K, Vikkula M, Cheah K, Vuorio E, Peltonen L. Predisposition to familial osteoarthrosis linked to type II collagen gene. *Lancet* 1989;**1**:924-7.
4 Vikkula M, Palotie A, Ritvaniemi P, Ott J, Ala-Kokko L, Sievers U, Aho K, Peltonen L. Early-onset osteoarthritis linked to the type II procollagen gene. *Arthritis Rheum* 1993;**36**:401-9.
5 Ritvaniemi P, Körkkö J, Bonaventure J, Vikkula M, Hysland J, Paassilta P, Kaitila I, Kääriäinen H, Sokolov BP, Hakala M, Mannismäki P, Meerson EM, Klemola T, Williams C, Peltonen L, Kivirikko KI, Prockop DJ, Ala-Kokko L. Identification of COL2A1 gene mutations in patients with chondrodysplasias and familial osteoarthritis. *Arthritis Rheum* 1995;**38**:999-1004.
6 Pun YL, Moskowitz RW, Lie S, Sundstrom WR, Block SR, Mcewen V, Williams HJ, Bleasel JF, Holderbaum D, Haqqi TM. Clinical correlations of osteoarthritis associated with a single-base mutation (arginine^{519} to cysteine) in type II procollagen gene. *Arthritis Rheum* 1994;**37**:264-9.
7 Williams CJ, Rock M, Considine E, McCarron S, Gow P, Ladda R, Mclain D, Michels VM, Murphy W, Prockop DJ, Ganguly A. Three new point mutations in type II procollagen (COL2A1) and identification of a fourth family with the COL2A1 Arg519->Cys base substitution using conformation sensitive gel electrophoresis. *Hum Mol Genet* 1995;**4**:309-12.
8 Nakata K, Ono K, Miyazaki J-I, Olsen BR, Murakgaki Y, Adachi E, Yamamura K-I et al. Osteoarthritis associated with mild chondrodysplasia in transgenic mice expressing α1(IX) collagen chains with a central deletion. *Proc Natl Acad Sci USA* 1993;**90**:2870-4.
9 Fassler R, Schnegelsberg PNJ, Dausman J, Shinya T, Muragaki Y, McCarthy MT, Olsen BR et al. Mice lacking α1(IX) collagen develop non-inflammatory degenerative joint disease. *Proc Natl Acad Sci USA* 1994;**91**:5070-4.
10 Briggs MD, Choi HC, Warman ML, Loughlin JA, Wordsworth P, Sykes B, Irven CMM, Smith M, Wynne-Davies R, Lipson MH, Biesecker LG, Garber AP, Lachman R, Olsen BR, Rimoin DL, Cohn DH. Genetic mapping of a locus for multiple epiphyseal dysplasia (EDM2) to a region of chromosome 1 containing a type IX collagen gene. *Am J Hum Genet* 1994;**55**:678-84.
11 Muragaki Y, Mariman ECM, van Beersum SEC, Perälä M, van Mourik JB, Warman ML, Olson BR, Hamel BCJ. A mutation in the gene encoding the α2 chain of the fibril-associated collagen IX, COL9A2, causes multiple epiphyseal dysplasia (EDM2). *Nature Genet* 1996;**1**:103-5.
12 Richard AJ, Yates JRW, Williams R, Payne SJ, Pope FM, Scott JD, and Snead MP. A family with Stickler syndrome type 2 has a mutation in the COL11A1 gene resulting in the substitution of glycine 97 by valine in α1(XI) collagen. *Hum Mol Genet* 1996;**5**:1339-43.
13 Li Y, Lacerda DA, Warman ML, Beier DR, Yoshioka H, Ninomiya Y, Oxford JT, Morris NP, Andrikopoulos K, Ramirez F, Wardell BB, Lifferth GD, Teuscher C, Woodward SR, Taylor BA, Seegmiller RE, Olsen BR. A fibrillar collagen gene COL11A1 is essential for skeletal morphogenesis. *Cell* 1995;**80**:423-30.
14 Brunner HG, Beersum SEC, Warman ML, Olsen BR, Ropers HH, Mariman ECM. A Stickler syndrome gene is linked to chromosome 6 near the COL11A2 gene. *Hum Mol*

Genet 1994;**3**:1561-4.
15 Vikkula M, Mariman ECM, Lui VCH, Zhidkova NI, Tiller GE, Goldring MB, Beersum SEC, de Waal Malefijt MC, van den Hoogen FHJ, Ropers HH, Mayne R, Cheah KSE, Olsen BR, Warman ML. Autosomal dominant and recessive osteochondrodysplasias associated with the COL11A2 locus. *Cell* 1995;**80**:431-7.
16 Briggs MD, Hoffman SMG, King LM, Olsen AS, Mohrenweiser H, Leroy JG, Mortier GR, Rimoin DL, Lachman RS, Gaines ES, Cekleniak JA, Knowlton RG, Cohn DH. Pseudoachondroplasia and multiple epiphyseal dysplasia due to mutations in the cartilage oligomeric matrix protein gene. *Nature Genet* 1995;**10**:330-6.
17 Cohn DH, Briggs MD, King LM, Rimoin DL, Wilcox WR, Lachman RS and Knowlton RG. Mutations in the cartilage oligomeric matrix protein (COMP) gene in Pseudoachondroplasia and Multiple Epiphyseal Dysplasia. *Ann N Y Acad Sci* 1996;**785**:188-94.
18 Baldwin CT, Farrer LA, Adair R, Dharmavaram R, Jimenez S, Anderson L. Linkage of early-onset osteoarthritis and chondrocalcinosis to human chromosome 8q. *Am J Hum Genet* 1995;**56**:692-7.
19 Kuivaniemi H, Peltonen L, and Kivirikko KI. Type IX Ehlers-Danlos syndrome and Menkes syndrome: The decrease in lysyl oxidase activity is associated with a corresponding deficiency in the enzyme protein. *Am J Hum Genet* 1985;**37**:798-808.
20 Heikkinen J, Hautala T, Kivirikko KI, and Myllylä R. Structure and expression of the human lysyl hydroxylase gene (PLOD): introns 9 and 16 contain Alu sequences at the sites of recombination in Ehlers-Danlos syndrome tyupe VI patients. *Genomics* 1994;**24**:464-71.
21 Hyland J, Ala-Kokko L, Royce P, Steinmann B, Kivirikko K, and Myllylä R. A homozygous stop codon in the lysyl hydroxylase gene in two siblings with Ehlers-Danlos syndrome type VI. *Nature Genet* 1992;**2**:228-31.
22 Berg ES, Olaisen B. Characterization of the COL2A1 VNTR polymorphism. *Genomics* 1993;**16**:350-4.
23 Warman ML, Tiller GE, Polumbo PA, Seldin MF, Rochelle JM Knoll JHM, Cheng S-D, et al. Physical and linkage mapping of the human and murine genes for the α1 chain of the type IX collagen (COL9A1). *Genomics* 1993;**17**:694-8.
24 Mäkelä TP, Hellsten E, Vesa J, Alitalo K, and Peltonen L. An Alu variable poly A repeat polymorphism upstream of L-Myc at 1p32. *Hum Mol Genet* 1992;**1**:217.
25 Hellsten E, Vesa J, Heiskanen M, Mäkelä TP, Jarvela I, Cowell JK, Mead S, Alitalo K, Palotie A and Peltonen L. Identification of YAC clones for human chromosome 1p32 and physical mapping of the infantile neuronal ceroid lipofuscinonsis (INCL) locus. *Genomics* 1995;**25**:404-12.
26 Brewton RG, Wood BM, Ren ZX, Gong Y, Tiller GE, Warman ML, Lee B, Horton WA, Olson BR, Baker JR and Mayne R. Molecular cloning of the α3 chain of human type IX collagen: linkage of the gene COL9A3 to chromosome 20q13.3. *Genomics* 1995;**30**:329-36.
27 Steinlein O, Fisher C, Keil R Smigrodzki R, and Vogel R. D20S19, linked to low voltage EEG, benign neonatal convulsions, and Fanconi anaemia, maps to a region of enhanced recombination and is localized between CpG islands. *Hum Molec Genet* 1992;**1**:325-9.
28 Nedospasov SA, Udalova IA, Kuuprash DV, Turetskaya RL. DNA sequence polymorphism at the human tumor necrosis factor (TNF) locus. *J of Immunology* 1991;**147**:1053-9.
29 Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, Bernardi G, Lathrop M, Weissenbach J. The 1993-94 Généthon human genetic linkage map. *Nature Genet* 1994;**7**:246-339.

30 Martin G, Mann D, Carington M. Recombination rates across the HLA complex: use of microsatellites as a rapid screen for recombinant chromosomes. *Hum Mol Genet* 1995;**4**:423-8.
31 Briggs MD and Cohn DH. Length polymorphism within a complex dinucleotide repeat in the human decorin (DCN) gene. *Hum Mol Genet* 1993;**7**:1087.
32 Hecht JT, Wang Y, Rhodes C, Yamada Y. GT repeat polymorphism in the human proteoglycan link gene promoter region. *Nucl Acid Res* 1991;**19**:6666.
33 Wang Y, Sadler L, Hecht JT. Polymorphic dinucleotide repeat in a cartilage matrix protein (CRTM). *Hum Mol Genet* 1992;**1**:780.
34 Csiszar K, Mariani TJ, Gosin JS, Deak SB and Boyd CD. A restriction fragment length polymorphism results in a nonconservative amino acid substitution encoded within the first exon of the human Lysyl Oxidase gene. *Genomics* 1993;**16**:401-6.
35 Patel MS, Mankoo BS, Bruckell PM. A polymorphic microsatellite repeat is located close to the promoter region of the c-fgr proto-oncogene (FGR) at chromosome 1p36.2-p36.1. *Hum Mol Genet* 1992;**1**:65.
36 Litt M, Sharma V, and Luty JA. Dinucleotide repeat polymorphism at the D11S35 locus. *Nucleic Acids Res* 1990;**18**:5921.
37 Meulenbelt I, Droog S, Trommelen GJM, Boomsma DI, Slagboom PE. High-yield noninvasive human genomic DNA isolation method for genetic studies in geographically dispersed families and populations. *Am J Hum Genet* 1995;**57**:1252-4.
38 Sambrook J, Fritsch EJ, and Maniatis T. 'Molecular Cloning: A Laboratory Manuals' Vols. I, II, III, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
39 Lathrop GM, Lalouel JM, Julier C, Ott J. Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. *Am J Hum Genet* 1985;**37**:482-98.
40 Ott J. Analysis of human genetic linkage, revised ed. Baltimore: John Hopkins University Press, 1991.
41 Hanson IM, Gorman P, Lui VCH, Cheah KSE, Solomon E, Trowsdale J. The human α2(XI) collagen gene (COL11A2) maps to the centromeric border of the major histocompatibility complex on chromosome 6. *Genomics* 1989;**5**:925-31.

4 ASSOCIATION OF THE COL2A1 GENE WITH RADIOLOGICAL OSTEOARTHRITIS IN A POPULATION BASED STUDY. THE ROTTERDAM STUDY

C Bijkerk, MD[1,2], I Meulenbelt, MSc[2], E Odding, MD, PhD[1], HA Valkenburg,MD, PhD[1], H Miedema, MD, PhD[2], FC Breedveld, MD, PhD[3], A Hofman, MD, PhD[1], JM Te Koppele PhD, HAP Pols PhD, CM Van Duijn, PhD[1], PE Slagboom, PhD[2]

[1]Department of Epidemiology & Biostatistics, Erasmus University Medical School, Rotterdam, The Netherlands.[2]TNO, Institute of Prevention and Health, Gaubius Laboratory, Leiden, The Netherlands.[3]Department of Rheumatology, University Hospital Leiden, The Netherlands.

Submitted for publication

Objective. **To examine the role of the COL2A1 gene in radiological OA (ROA).**

Methods. **A genetic association study was performed, using the COL2A1 VNTR (variable number of tandem repeats) polymorphism. Cases and controls, aged between 55 and 65 years, were derived from a prospective population based cohort study, the Rotterdam Study. ROA was assessed by means of Kellgren's grading system. In the cases the occurrence of Heberden's nodes was examined prospectively.**

Results. **The present study included 180 cases with at least ROA in the knee or hip joint and 285 controls without ROA in the knee, hip or hand. The frequency distribution of alleles of the COL2A1 VNTR polymorphism in cases differed statistically significantly from that in controls (p=0.03). The frequency of 13R1 homozygosity was 1.9 times increased (95% CI 1.1-3.5) in cases as compared to controls. Carriers of the 14R2 allele had a 2.3 times higher risk of developing Heberden's nodes (95% CI 1.0-5.2) then subjects without this allele.**

Conclusion. **The present study suggests that the common 13R1 allele of the COL2A1 VNTR locus is associated with ROA and that the rare allele 14R2 is related to the occurrence of Heberden's nodes.**

Osteoarthritis (OA) is the most common disease of the musculoskeletal system. The prevalence of OA rises from 4% in 18-24 year old individuals to 85% in subjects of ages over 74 years (1,2). The radiographic changes in OA reflect a progressive deterioration of articular (hyaline) cartilage of diarthrodial joints with narrowing of the joint space, formation of osteophytes and development of sclerosis and pseudocystic areas in subchondral bone.

An influence of genetic factors on the etiology of OA was initially recognized in individuals with generalized OA (GOA), defined as OA in at least three different joint groups, in the presence of Heberden's nodes (3). The exact pattern of joint involvement in subjects with GOA is unclear and the origin and role of Heberden's nodes in GOA is contentious (4-6). However, the clinical expression of the genetic predisposition to GOA appears to be modified by gender and age (4,5). Furthermore, genetic factors are involved in the rare forms of early-onset (before 30 years of age) familial OA, transmitted as an autosomal dominant trait (7,8). A recent twin study (9) suggested that genetic factors may play an important role in the occurrence of sporadic OA in the population.

Many different cartilage related genes, including COL2A1, COL9A1, COL9A2, COL11A1, COL11A2, CRTM and CRTL, have been proposed to be implicated in the

etiology of OA (10). The COL2A1 gene (12q13), encoding the predominant cartilage collagen, type II, is a major candidate gene for OA. Mutations in the COL2A1 gene give rise to various grades of matrix failure in articular cartilage resulting in OA (11,12). Over 40 mutations have indeed been identified in families with severe early-onset OA associated with osteochondrodysplasia (13-15).

However, findings on the role of the COL2A1 gene in OA in population studies have been controversial. Hull and Pope (16) found an increased frequency of a BamH1 COL2A1 allele (frequency 0.07) in British women with radiologically confirmed symptomatic OA in more than one joint as compared to a control group. A recent study by Loughlin et al. (17), among British patients with GOA, indicated the existence of a rare MaeII COL2A1 allele (frequency 0.03) that may be associated with an increased susceptibility for OA. In contrast, an association study and sib-pair study of limited size (18,19), also in British patients with primary GOA, failed to show a significant association with a highly informative VNTR (variable number of tandem repeats) polymorphism located at the 3' end of the COL2A1 gene. Neither did a study of patients with primary GOA or OA of the finger joints in a Finnish population (20). Overall, these findings demonstrate that the role of the COL2A1 gene in OA in the population is unclear and requires further study.

We conducted an association study on the relation between the COL2A1 gene and ROA, occurring before 65 years of age and irrespective of possible manifestations of clinical OA. Cases and controls were derived from a single population based study, the Rotterdam Study. Further, we examined whether the COL2A1 gene plays a role in the development of Heberden's nodes in individuals with radiographic signs of developing GOA.

Materials and Methods

Subjects. The present study was part of the Rotterdam Study, a prospective population based cohort study of determinants and prognosis of chronic diseases in the elderly (21). For this study, all inhabitants of a suburb of Rotterdam, aged 55 years or over, including institutionalized persons, were invited to participate. In total 7983 participants were examined between 1990 and 1993. The response rate was 78%. From all subjects informed consent was obtained and the study was approved by the medical ethics committee of Erasmus University Medical School.

In order to ascertain ROA of relatively early onset, the current study was restricted to non-institutionalized participants aged between 55 and 65 years (n=2593). In this age category a random subset of 385 men and 559 women was drawn. In this subset radiographs of four different joint groups, i.e. knees, hips, hands and thoracolumbar spine, were scored for the presence of ROA. Radiographs from other joints were not available.

Within the subset of 944 subjects, cases (n=188) with ROA in two or more joint groups, including at least ROA in one of the large weight bearing joints, i.e. the knee or hip, were selected for genetic studies. These subjects were included in more extensive follow-up

studies. In order to maximize the phenotypic contrast in the present association study, cases were compared with controls (n=336) who were free from ROA in the peripheral joints, i.e. knees, hips and hands.

Measurements Weight bearing anterior-posterior radiographs of the hips and knees, anterior-posterior radiographs of the hands and wrists and lateral radiographs of the spine (Th4-S1) were obtained. ROA was assessed by means of the grading system proposed by Kellgren et al. (22). All radiographs were scored by two independent readers, blinded to all other data of the participant. Whenever the scores differed more than 2 points or was 1 for one reader and 2 or more for the other, a consensus score was agreed upon. ROA of the hand was assessed in each inter- and metacarpalphalangeal joint individually. For the carpometacarpal and intercarpal joints only the first carpometacarpal (CMC I) and the trapezoscaphoideal (TS) joint were assessed. ROA of the wrist was assessed at the radiocarpal and distal radioulnar joints. For the spine each individual level from Th4 to S1 was scored with regard to osteophytes and disk space narrowing. In the analysis definite ROA was defined as Kellgren-score 2 or over in the left and/or right corresponding joint. Hand ROA was defined as Kellgren-score 2 or over in at least one of the 36 joints that were scored. For this purpose the joints of the wrist were included in the category hand ROA. ROA of the spine was defined as Kellgren-score 2 or over in at least one level from Th4 to S1. Heberden's nodes were scored at baseline by trained investigators. At follow-up of the cases with ROA in at least two joint groups, after approximately 5 years, measurements included Heberden's and Bouchard's nodes, joint complaints and a physical examination of the joints.

VNTR polymorphism DNA was obtained in 818 subjects (87%), including 180 cases (96%) and 285 controls (85%). In order to genotype the multiallelic VNTR locus, PCR reactions were performed in 25 µl containing 25-50 ng genomic DNA, 2.5 pmoles of each primer (forward primer: 5'-CAA CTG ATA AAA CAG AGA GC- 3' and reverse primer: 5'-CTC CTT TGT CAT GAA CTA GC- 3' [23]), 1*Super Taq buffer, 2 µCi $\alpha[^{32}P]$-dCTP, 200 µM each of dCTP, dGTP, dTTP, and dATP, and 0.05 U of superTaq DNA polymerase (HT Biotechnology, Cambridge, UK). Amplification was initiated with 3 min denaturation at 94°C followed by 35 cycles with 93°C (1 min) denaturation, 56°C (1 min) annealing step, and a 72°C (2 min) elongation step in a Hybaid OmniGene thermal cycler using tube control. The amplification was finished by a final incubation at 72°C for 4 min. Alleles were separated by high resolution gel electrophoresis through a denaturing 3.5% polyacrylamide gel, and analyzed by autoradiography. The nomenclature of alleles and allelic ladder was used from Berg and Olaisen (24).

Statistical analysis Demographic variables were compared by using the Student's t test and chi-square test. Allele frequencies for cases and controls were assessed by counting alleles and calculating sample proportions. In the statistical analyses, alleles with a frequency lower then 0.05 were pooled. A likelihood ratio test was used to compare allele frequencies between groups (25). Furthermore, a logistic regression model was used to examine the effects of age, gender, body mass index (BMI) and Heberden's nodes on the association between a VNTR genotype and the occurrence of ROA. The strength of association between a genotype and the occurrence of ROA was estimated as the odds ratio (OR). ORs are presented with 95% confidence intervals (95% CI). In the case of small sample size exact ORs and 95% confidence intervals were calculated using the Fisher exact test (26).

To examine the relationship between the occurrence of Heberden's nodes and alleles of the COL2A1 VNTR locus, a cumulative incidence rate for each allele was calculated. Next, a crude RR was obtained using, as a reference, subjects without the respective allele. Cox proportional hazard's model was used to obtain multivariate adjusted RRs, with the presence or absence of incident Heberden's nodes in each person as dependent variable.

Results

The baseline characteristics of our study population are shown in Table I. The overall case group can be subdivided into three groups: 112 (62.2%) cases with ROA in one or both knee joints, 51 (28.3%) cases with ROA in one or both hip joints and 17 (9.4%) cases with ROA in one or both knee and hip joints. Hip ROA was clearly more frequent in men, whereas knee ROA was more frequent in women. Heberden's nodes are more frequent in cases as compared to controls ($p < 0.0001$).

In this study population 14 out of 23 reported alleles of the COL2A1 VNTR polymorphism were detected. In Table II, allele frequencies of the five most frequent alleles in all 818 genotyped subjects and in cases and controls are shown. The frequency distribution of alleles in cases and controls is statistically significantly different ($p=0.03$), which is primarily caused by an increased frequency of allele 13R1 in cases as compared to controls. None of the other alleles, including allele 14R1, was statistically significantly associated with the presence of ROA. When adjusting for age, sex, BMI and Heberden's nodes, a statistically significant association was found between 13R1 homozygotes and ROA in the cases (OR=1.9; 95% CI 1.1-3.5, see Table III). The association was absent in cases with Heberden's nodes (OR=0.9; 95% CI 0.2-3.5). When considering the most severe cases, with either knee or hip ROA in combination with hand and spine ROA, the association

Table I: Description study population

	Cases			Controls
	Overall (n=180)	Knee (n=112)	Hip (n=51)	(n=285)
Age (SD) in years	60.9 (2.5)*	60.9 (2.5)*	60.8 (2.6)†	59.6 (2.8)
range	55.4-65.0	55.7-64.8	55.4-65.0	55.0-65.0
No. (%) of men	64 (35.6) ‡	28 (25.0)*	33 (64.7)	147 (51.6)
Body mass index (SD) in kg/m^2	27.6 (4.0)*	28.3 (4.2)*	26.0 (2.6)	25.6 (3.2)
No. with Heberden's nodes (%)	53/166* 31.9	37/102* 36.3	11/50 22.0	36/26113.8
No. of incident Heberden's nodes (%)	34/99 (34.0)	22/62 (35.5)	10/28 (35.7)	
No. of women (%)	25/58 (42.9)¶	19/44 (43.2)	4/8 (50.0)	
No. of men (%)	9/41 (22.0)	3/18 (16.7)	6/20 (30.0)	

17 cases had knee and hip ROA, these are not shown separately. *$p < 0.0001$, †$p < 0.001$, ‡$p < 0.01$ (cases compared to controls). ¶$p < 0.05$ (women compared to men)

Table II: Allele frequencies in cases and controls (the five most frequent alleles are shown; the other nine alleles, all with individual frequencies lower than 0.05 are summed in the category others).

Allele	11R1	13R1	13R2	14R1	14R2	Others	Total	LRT
Overall: No	172	689	101	440	98	136	1636	
(freq)	(0.11)	(0.42)	(0.06)	(0.27)	(0.06)	(0.08)		
Controls: No	62	222	45	160	36	45	570	
(freq)	(0.11)	(0.39)	(0.08)	(0.28)	(0.06)	(0.08)		
Cases: No	35	168	24	82	25	26	360	p=0.03
(freq)	(0.10)	(0.47)	(0.07)	(0.23)	(0.07)	(0.07)		

Numbers listed are number of alleles. ROA=radiological osteoarthritis, LRT=Likelihood ratio test.

with 13R1 homozygosity persisted in subjects with hip ROA (OR=3.0; 95% CI 1.0-9.0), but was absent in subjects with knee ROA (OR=1.0; 95% CI 0.4-2.7), see Table III.

Out of 188 cases at baseline, with ROA in two or more joint groups including at least the knee or hip joint, 174 were still available for the present study. Follow-up was complete for 147 (84%) of these eligible probands, with an average follow-up period of approximately 5 years. Of these subjects, 99 were free from Heberden's nodes at baseline. After 5 years, 34 had developed Heberden's nodes (see Table I).

Table III: Association between ROA and allele 13R1 (13R1 hetero- and homozygotes are simultaneously compared with subjects without the 13R1 allele).

	No. of subjects	OR 13R1 heterozygotes (95% CI)	OR 13R1 homozygotes (95% CI)
Cases overall	180	1.4 (0.8-2.2)	1.9 (1.1-3.5)*
Heberden's absent	113	1.2 (0.7-2.0)	2.1 (1.1-4.3)*
Heberden's present	53	1.5 (0.5-4.9)	0.9 (0.2-3.5)
ROA Hip + ROA Hand and Spine (ROA Knee -)	28	1.0 (0.4-2.9)	3.0 (1.0-9.0)*
ROA Knee + ROA Hand and Spine (ROA Hip -)	61	0.9 (0.4-1.7)	1.0 (0.4-2.7)

Information on Heberden's nodes was missing for 38 subjects (14 cases and 24 controls). All ORs are adjusted for age, sex and body mass index. ORs in the non-stratified case groups are also adjusted for Heberden's nodes. *$p < 0.05$

Table IV: Relationship of genotype of the COL2A1 VNTR polymorphism to incident Heberden's nodes (n=99).

Allele	Cumulative incidence of Heberden's nodes, No. (%)	Crude RR (95% CI)	Adjusted RR (95% CI)*
14R2	8/11 (72.7)	2.5 (1.5-4.0)†	2.3 (1.0-5.1)†
13R1	20/66 (30.3)	0.7 (0.4-1.2)	0.7 (0.4-1.4)
14R1	12/45 (26.7)	0.7 (0.4-1.2)	0.7 (0.3-1.4)
13R2	8/16 (50.0)	1.6 (0.9-2.9)	1.6 (0.7-3.5)
11R1	5/17 (29.4)	0.8 (0.4-1.8)	0.8 (0.3-2.1)

*Adjusted simultaneously for age, sex and body mass index. In each stratum carriers of an allele are compared with non-carriers of this allele. †$p < 0.05$. RR = risk ratio, CI = confidence interval, ROA = radiological osteoarthritis.

The relationship of the five most frequent alleles to the development of incident Heberden's nodes is presented in Table IV. The cumulative incidence of Heberden's nodes was highest in carriers of the 14R2 allele of the COL2A1 VNTR polymorphism. Carriers of the 14R2 allele (which were all heterozygous) had a 2.5 times higher risk ($p < 0.05$) of developing Heberden's nodes then subjects without this allele (see Table IV). After adjustment for other risk factors this remained statistically significant (RR=2.3; 95% CI 1.0-5.2). None of the other alleles was associated with incident Heberden's nodes. The frequency of the 14R2 allele did not differ between the overall case group and the control group (see Table II). However, among cases with Heberden's nodes, the frequency of the 14R2 allele was increased in subjects with generalized ROA in knee, hand and spine (adjusted OR=3.4; 95% CI 0.6-20.8). There was no statistically significant difference in the frequency distribution of alleles in cases with Heberden's nodes as compared to cases without Heberden's nodes.

Discussion

We found a statistically significant association between the most frequent allele of the COL2A1 VNTR locus, the 13R1 allele, and ROA in two or more joint groups including at least the knee or hip joint. The strongest effects were found in cases without Heberden's nodes and cases with hip, hand and spine ROA. The association with the 13R1 allele was due to 13R1 homozygotes. The effect of allele 13R1 could not be explained by differences between cases and controls in age, gender or BMI. Allele 14R2 was associated with the development of Heberden's nodes, with a 2.5

times increased risk of Heberden's nodes in the presence of this allele.

The number of alleles as well as the allele frequencies in our study population are comparable with those previously reported (23,24,27). The prevalence of ROA at different joint sites as observed in our study population is in keeping with earlier reports on caucasian populations, taking into account the considerable variance in the prevalence of ROA in different populations (2). Although all radiographs were scored by two trained readers, some misclassification may still have occurred. However, all radiographs were scored before genotyping of the VNTR locus and genotyping was performed blind to the disease status.

Even though the genetic component in OA may be substantial, both genetic and clinical heterogeneity in OA hamper the possibilities of identifying susceptibility genes in association studies. This may explain the fact that an association was found between the COL2A1 VNTR locus and ROA in our study, but not in two previous studies of the same locus. The first of these was a study in 61 British hospital patients with primary generalized OA, defined as the presence of Heberden's nodes before 60 years of age and the involvement of at least three other joint groups (18,19). Since this study was carried out in a subset of relatively rare and severe OA cases, the study does not exclude the possibility of an association of the COL2A1 VNTR locus with more commonly occurring ROA in the general population. Furthermore, cases were compared with controls of unknown OA status. Given the high frequency of ROA in the population this may have decreased the statistical power of the study substantially.

The second association study with a negative finding, involving the COL2A1 VNTR polymorphism, compared 41 Finnish hospital patients with primary GOA (defined as ROA in more than one weight bearing joint) and 49 patients with hand ROA with 48 controls confirmed as free of ROA (20). It is important to note that only 4 out of 23 reported alleles of the COL2A1 VNTR polymorphism were described in this study. It is conceivable that, due to genetic drift, the COL2A1 gene does not play a predominant role in the pathogenesis of ROA in the Finnish genetic isolate, but is an important factor involved in the development of ROA in other populations. Or, that mutations leading to ROA may be in linkage disequilibrium with alleles 13R1 and 14R2 in our population, but are not in linkage disequilibrium with any of the VNTR alleles detected in the Finnish population.

Divergent outcomes, and even spurious associations, in the comparison between association studies in different populations can be due to the effect of population admixture (28). However, in the present study cases and controls were derived from a well-defined caucasian population, making the occurrence of bias due to

admixture unlikely.

Our observations indicate that the common 13R1 allele (frequency 0.42) of the COL2A1 VNTR locus is associated with ROA and that the rare 14R2 allele (frequency 0.06) is associated with the occurrence of Heberden's nodes. This supports the findings of two separate studies of British patients (16,17), that the COL2A1 gene is involved in the etiology of OA in the population. These two studies used other polymorphic markers in the COL2A1 gene than the present study. Finally, the results of the present study suggest the existence of at least two separate subsets of individuals with GOA. Firstly, classical nodal GOA, which involves knee and hand OA in the presence of Heberden's nodes associated with COL2A1 allele 14R2. Secondly, individuals with hip and hand OA without Heberden's nodes associated with homozygosity for COL2A1 allele 13R1. This distinction might be an additional explanation for the controversial findings on the role of the COL2A1 gene in OA.

Acknowledgments

We wish to thank H Kornman for her supportive activities, S de Wildt for genotyping the VNTR locus and A Uitterlinden, PhD for helpful discussions. Prof C van Broeckhoven, PhD (University of Antwerp) is acknowledged for the contribution in DNA extraction for a subset of the participants. This study was supported by the Netherlands Organization for Scientific Research (NWO), the "Dutch League against Rheumatism", the NESTOR stimulation program for geriatric research in the Netherlands (Ministries of Health and Education), the Loosco Foundation and the Municipality of Rotterdam.

References

1 Masi AT, Medsger TA: Epidemiology of the rheumatic diseases. In: McCarthy DJ (ed) Arthritis and allied conditions, Lea and Febiger, London, 1989, pp 16-54
2 Van Saase JLCM, Van Romunde LKJ, Cats A, Vandenbroucke JP, Valkenburg HA: Epidemiology of osteoarthritis: Zoetermeer survey, comparison of radiological osteoarthritis in a Dutch population with that in 10 other populations. Ann Rheum Dis 1989;48:271-80
3 Kellgren JH, Moore R: Generalized osteoarthritis and Heberden's nodes. Br Med J 1952;1:181-7
4 Stecher RM: Heberden Oration. Heberden's nodes. A clinical description of osteoarthritis of the finger joints. Ann Rheum Dis 1955;14:1-10
5 Kellgren JH, Lawrence JS, Bier F: Genetic factors in generalized osteoarthrosis. Ann Rheum Dis 1963;22:237-55
6 Cooper C, Egger P, Coggon D, Hart DJ, Masud T, Cicuttini F, et al. Generalized

osteoarthritis in women: pattern of joint involvement and approaches to definition for epidemiological studies. J Rheum 1996;23:1938-42
7 Palotie A, Ott J, Elima K, Cheah K, Vaisanen P, Ryhanen L, Vikkula M, et al: Predisposition to familial osteoarthrosis linked to type II collagen gene. Lancet 1989;1:924-7
8 Knowlton RG, Katzenstein PL, Moskowitz RW, Weaver EJ, Malemud CJ, Pathria MN, Jiminez SA, et al: Genetic linkage of a polymorphism in the type II procollagen gene (COL2A1) to primary osteoarthritis associated with mild chondrodysplasia. New Engl J Med 1990;322:526-30
9 Spector TD, Cicutti F, Baker J, Loughlin J, Hart DJ: Genetic influences on osteoarthritis in women: a twin study. BMJ 1996;312:920-4
10 Williams CJ, Jiminez SA: Heredity, genes and osteoarthritis. Rheum Dis Clin North America 1993;19:523-43
11 Eyre DR, Weis MA, Moskowitz RW: Cartilage expression of a type II collagen mutation in an inherited form of osteoarthritis associated with a mild chondrodysplasia. J Clin Invest 1991;87:357-61
12 Kuivaniemi H, Tromp G, Prockop DJ: Mutations in collagen genes: causes of rare and some common diseases in humans. Faseb J 1991;5:2052-60
13 Ala-Kokko L, Baldwin CT, Moskowitz RW, Prockop DJ: Single base mutation in the type II procollagen gene (COL2A1) as a cause of primary osteoarthritis associated with a mild chondrodysplasia. Proc Natl Acad Sci USA 1990;87:6565-8
14 Ahmad NN, Ala-Kokko L, Knowlton RG, Jiminez SA, Weaver EJ, Maguire JL, Tasman W, et al: Stop codon in the procollagen II gene (COL2A1) in a family with the Stickler syndrome (arthro-ophthalmopathy). Proc Natl Acad Sci USA 1991;88:6624-7
15 Williams CJ, Considine EL, Knowlton RG, Reginato A, Neumann G, Harrison D, Buxton P, et al: Spondyloepiphyseal dysplasia and precocious osteoarthritis in a family with an Arg_{75}→Cys mutation in the procollagen type II gene (COL2A1). Hum Genet 1993;92:499-505
16 Hull R, Pope FM: Osteoarthritis and cartilage collagen genes. Lancet 1989;i:1337-8
17 Loughlin J, Irven C, Athanasou N, Carr A, Sykes B: Differential allelic expression of the type II collagen gene (COL2A1) in osteoarthritic cartilage. Am J Hum Genet 1995;56:1186-93
18 Priestley L, Fergusson C, Ogilvie D, Wordsworth P, Smith R, Pattrick M, Doherty M, et al: A limited association of generalized osteoarthritis with alleles at the type II collagen locus: COL2A1. Br J Rheumatol 1991;30:272-5
19 Loughlin J, Irven C, Fergusson C, Sykes B: Sibling pair analysis shows no linkage of generalized osteoarthritis to the loci encoding type II collagen, cartilage link protein or cartilage matrix protein. Br J Rheumatol 1994;33:1103-6
20 Vikkula M, Nissilä M, Hirvensalo E, Nuotio P, Palotie A, Aho K, Peltonen L: Multiallelic polymorphism of the cartilage collagen gene: no association with osteoarthritis. Ann Rheum Dis 1993;52:762-4
21 Hofman A, Grobbee DE, De Jong PTVM, Van den Ouweland FA: Determinants of disease and disability in the elderly. Eur J Epidemiol 1991;7:403-22
22 Kellgren JH, Jeffrey MR, Ball J (eds): The epidemiology of chronic rheumatism. Volume II: Atlas of standard radiographs of arthritis. Oxford: Blackwell Scientific Publications, 1963
23 Priestley L, Kumar D, Sykes B: Amplification of the COL2A1 3' variable region used for segregation analysis in a family with the Stickler syndrome. Hum Genet 1990;85:525-6
24 Berg ES, Olaisen B: Characterization of the COL2A1 VNTR polymorphism. Genomics 1993;16:350-4
25 Terwilliger JD: A powerful likelihood method for the analysis of linkage disequilibrium

between trait lociand one or more polymorphic marker loci. Am J Hum Genet 1995;56:777-87
26 Mehta CR, Patel NR, Gray R: Computing an exact confidence interval for the common odds ratio in several 2x2 contingency tables. JASA 1986;80:969-73
27 Rand S, Puers C, Skowasch K, Wiegand P, Budowle B, Brinkmann B: Population genetics and forensic efficiency data of 4 AMPFLP's. Int J Leg Med 1992;104:329-33
28 Lander ES, Schork NJ: Genetic dissection in complex traits. Science 1994;265:2037-48

5 POPULATION HAPLOTYPE ANALYSIS AND EVOLUTIONARY RELATIONS OF THE COL2A1 GENE

I Meulenbelt[1], CJ Williams[2], JM Te Koppele[1], GC Van de Giessen[1], & PE Slagboom[1].

[1] TNO-PG, Dept. of Vascular and Connective Tissue Research, PO box 2215, 2301 CE Leiden, The Netherlands; [2] Jefferson Institute of Molecular Medicine, Dept. of Biochemistry and Molecular Biology, 233 South 10th Street, Philadelphia, USA.

Reproduced from

Annals of Human Genetics, 1996, 60:189-199

Abstract

We have determined the allele frequencies and pairwise linkage disequilibria of restriction fragment length polymorphisms (RFLPs) distributed over the entire COL2A1 gene (spanning 23.6 kb) in a population of unrelated Dutch Caucasians. Pairwise linkage disequilibrium analysis of RFLP sites between exon 5B and 51 indicated a high degree of partly positive (the rare alleles of both loci are associated) and partly negative (the rare allele is associated with the common allele) linkage disequilibrium.

The high degree of linkage disequilibrium enabled the assignment of 13 out of 128 possible haplotypes with 7 RFLPs. An evolutionary tree of these haplotypes was derived using a minimum spanning tree approach, indicating at least two ancestral haplotypes. Our data indicate that disease related population studies involving the COL2A1 gene should include a minimum of 4 RFLPs (D9, A9, H33, P51) to obtain 98% of possible haplotypes occurring.

Collagen type II is the most abundant collagenous protein of epiphyseal and articular cartilage and consists of three α1(II)-chains in a helix. The amino acid structure of each chain consists of a repetitive Gly-X-Y structure [Baldwin *et al.* 1989] which corresponds with a high GC-content and a high degree of repetitiveness throughout the COL2A1 gene. COL2A1 mutations are involved in several types of heritable skeletal disorders such as different types of chondrodysplasias and Stickler/Kniest dysplasia [see e.g. Williams *et al.* 1993]. An Arg^{519}-Cys COL2A1 mutation was found in four apparently different families with an distinctive pattern of dominantly inherited generalized osteoarthrosis (GOA) associated with a mild chondrodysplasia [Ala-Kokko *et al.* 1990; Holderbaum *et al.* 1993, Pun *et al.* 1994, Williams *et al.* 1995]. It has been suggested that defects in the COL2A1 gene with a milder effect may be involved in sporadic GOA [Prockop *et al.* 1990]. Especially in a group of patients which develop GOA in combination with Heberden nodes before the age of 65, a significant role of genetic factors was demonstrated [Kellgren *et al.* 1963]. A relationship between the COL2A1 gene and sporadic GOA may be established by association studies in which the frequency of COL2A1 alleles and/or haplotypes is compared between distinctly affected individuals and controls. Thus far, association studies which aimed at the detection of common COL2A1 alleles in middle aged GOA patients, did not clearly demonstrate such a relation [Priestley *et al.* 1991; Vikkula *et al.* 1993]. Recently, Loughlin *et al.* (1995) showed a reduced expression of a rare MaeII(-) allele of the

COL2A1 gene in arthritic cartilage. Subsequently, a significant increase of the MaeII(-) allele was found in a clinically well defined OA group with generalized nodal osteoarthrosis. These results indicated the possible existence of a rare COL2A1 allele that predisposes to a GOA phenotype [Loughlin *et al.* 1995]. The chance of finding disease association is increased when the majority of existing alleles of the candidate gene in the population is known. The detection of existing alleles of the gene is increased by haplotype construction of a defined pattern of intragenic RFLP alleles.

As a first step towards an association study in Dutch GOA affected and control individuals we have determined allele frequencies, linkage disequilibria, and we estimated the expected haplotype frequencies of RFLP sites distributed over the COL2A1 gene (Fig. 1) in random Dutch Caucasians. The COL2A1 RFLP's are in tight (partly positive, partly negative) linkage disequilibrium. Of 128 possible haplotypes using 7 RFLPs, 13 different haplotypes were found in our population which might result from mutations, recombinations, and founder effects. These haplotypes may also be found in other populations and can be used in further disease association studies of the COL2A1 gene.

Materials and methods

Population sample

Peripheral blood samples were obtained from random unrelated Caucasian individuals of Dutch origin.

PCR analysis

PCR reactions covering exon 5B, intron 9, intron 33 were carried out as described by Williams *et al.* (1992). The PCR covering exon 51 is performed as described by Vissing *et al.* (1989). PCR products were purified with phenol/chloroform/isoamylalcohol (24:24:1) extractions. Subsequently, PCR products were digested to completion with the appropriate restriction enzyme under conditions recommended by the supplier (Gibco-BRL). Digestion products were separated on 7.5% or 15% acrylamide gels. Gel separation patterns were visualized by ethidium bromide staining and documented by Polaroid photography under 302 nm UV light.

Restriction fragment length polymorphisms (RFLPs)

In total 7 different RFLPs within the COL2A1 gene were studied. The polymorphic sites within the gene were detected by different restriction enzymes creating different restriction fragment lengths (Table 1). The common alleles were designated with 1 and the rare alleles with 2. The allele frequency of each RFLP was estimated by allele counting of the independent chromosomes with standard error $\sqrt{(p(1-p)/2N)}$. Tests for goodness of fit into Hardy Weinberg equilibrium (HWE) were calculated using standard methods (see eg. Chakravarti *et al.* (1984).

Statistical analysis

Pairwise linkage disequilibrium statistics

The linkage-disequilibrium coefficient (Δ) was determined for each RFLP pair. The coefficient Δ is positive when the rare alleles of both loci are associated and is negative when a common allele is associated with a rare allele. Δ is calculated from the disequilibrium measure $D^* = h_{22}-pq$ where h_{22} is the frequency of the haplotype with the rarer allele at each locus and p and q are frequencies of the rarer alleles at loci 1 and 2, respectively, where $p \leq q \leq 0.5$ [Thompson *et al.* 1988]. For double heterozygous individuals for which haplotype 2-2 could not be determined unambiguously, the haplotype probabilities were estimated by the maximum likelihood procedure, using the iterative process outlined by Hill (1974). The procedure uses a starting value for iteration assuming that the genotype frequencies of the double heterozygous class is exactly that computed from the other classes. The iterative process is continued until stability is reached (criteria 1.6E-9, empirically chosen by Hill, 1974) when D^* is obtained as $h_{22}-pq$. From this the linkage disequilibrium coefficient Δ is calculated as $D^*/\sqrt{(p(1-p)q(1-q))}$. Although Δ is defined to be within the range -1.0-1.0, its true range is determined by the allele frequencies. It can thus be misleading to compare Δ values when one Δ is close to its maximum while another is not (Hedrick 1987). The maximum and minimum possible value of D^* is determined as $D_{max}=p(1-q)$ and $D_{min}=-pq$. Therefore, the maximum and minimum possible values of Δ varies between $D_{max}/\sqrt{(p(1-p)q(1-q))}$ and $D_{min}/\sqrt{(p(1-p)q(1-q))}$. The disequilibrium coefficient D' is the fraction of Δ_{max} and/or Δ_{min} achieved by Δ. D' is, therefore, less dependent on allele frequencies. To test the hypothesis $\Delta = 0$ a χ^2 test is given by $N\Delta^2$ with 1 df where N is the number of individuals in the population sample [Hill 1974].

Relationship between Δ and physical distance

The theoretically inverse relationship between Δ and the physical map distance is given by: $\Delta \approx 1/(1+4N_e c)$ where N_e is the effective population size and c is the recombination fraction. For distances within tens of kb units, $c = kd$, where k is recombination frequency per kb [Hill & Roberston 1968, Sved 1971], and d is the physical distance between markers. While this relationship may not be strictly applicable to RFLPs, and is subject to sampling error [Golding 1984, Hudson 1985, Weir & Hill 1986], it has been used in the context of other gene regions for which it has been supported by direct observations [Chakravarti *et al.* 1984, Chakravarti *et al.* 1986]. $4N_ek$ is a standardized measure of recombination frequency per kb that allows the comparison of values between different RFLP pairs of the COL2A1 gene. Since the absolute value of Δ is relatively low in linkage disequilibria between RFLPs with low frequency alleles [see this article and e.g Leitersdorf *et al.* 1989], we used D' to estimate $4N_ek$.

Evolutionary relations

The evolutionary tree was derived using a minimum spanning tree approach, where a graph is drawn connecting all haplotypes such that the number of differences measured over the graph is minimized. A distance matrix was made based on the number of RFLP loci varying between two haplotypes. This was used as a dissimilarity matrix input in a multidimensional scaling analysis in two dimensions, using linear instead of monotonic regression. Dissimilarities were explained for 99% of the variance from euclidean distances. The minimum spanning tree is drawn in the configuration as computed by the multidimensional scaling procedure. The statistical package SYSTAT was used for multidimensional scaling, while the minimum spanning tree program was written in PROLOG (courtesy dr. te Meerman).

Results

Frequencies and genotypes of RFLPs in the COL2A1 gene

The location and fragment sizes of 7 RFLPs investigated are depicted in Fig. 1 and Table 1. The RFLPs extended from exon 5B to exon 51 and span a total distance of 23.6 kb. The allele frequencies of these RFLPs were determined in a population of 83 to 124 Dutch Caucasians (Table 2). The arrangement of alleles in genotypes for each polymorphic site was not significantly different from those expected for populations in Hardy Weinberg equilibrium (HWE). RFLP P51 which converts a glycine to serine at amino acid position 191 in the C-propeptide of the COL2A1 gene [Vissing *et al.* 1989], was not previously included in a population study. The allele frequencies of the remaining 6 polymorphic sites did not differ significantly from other Caucasian populations [Williams *et al.* 1992, Eng & Strom 1985, Nunez *et al.* 1985, Sykes *et al.* 1985, Väisänen *et al.* 1988, Ogilvie *et al.* 1987, Hull & Pope 1989, Sokolov *et al.* 1991] (Results not shown).

Pairwise linkage disequilibrium analysis at the COL2A1 locus

Population linkage disequilibrium data for the entire COL2A1 gene have not been published previously. Pairwise disequilibria of alleles of RFLPs H5B, A9, D9, H33, N33 and P51 were estimated from the diploid genotypes. Table 3 shows the disequilibrium statistics Δ, D', and the number of haplotypes assigned. For double

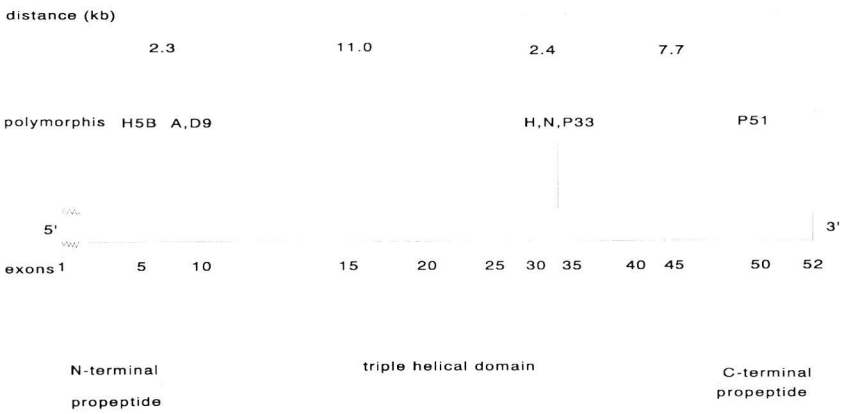


Figure1. Schematic representation of the COL2A1 gene, showing the location of the seven RFLPs analysed.Exons are indicated by the vertical lines.

Table 1:*Collagen type II polymorphisms*

RFLP	Location	enzyme	Allele sizes in kb		References
			1	2	
H5B	exon 5B	*Hae* III	0.103+0.018	0.121	(Williams *et al.* 1992)
A9	intron 9	*Acc* I	0.106+0.010	0.116	(Williams *et al.* 1992)
D9	intron 9	*Dra* I	0.330	0.252+0.078	(Williams *et al.* 1992)
H33	intron 33	*Hind* III	0.711	0.421+0.290	(Nunez *et al.* 1985)
N33	intron 33	*Nci* I	0.711	0.585+0.126	(Williams *et al.* 1992)
P33	intron 33	*Pvu* II	0.466	0.351+0.115	(Williams *et al.* 1992)
P51	exon 51	*Pvu* II	0.154+0.184	0.338	(Vissing *et al.* 1989)

Table 2:*Allele frequencies of RFLP sites in the COL2A1 gene in the Dutch Caucasian population.*

RFLP	Frequency rare allele (± SE)	2N	χ^2 (1 df)
H5B	.23 ± .027	246	3.33^{NS}
A9	.44 ± .039	166	1.73^{NS}
D9	.25 ± .027	248	0.00^{NS}
H33	.38 ± .003	248	1.15^{NS}
N33	.36 ± .003	244	1.27^{NS}
P33	.36 ± .003	244	1.92^{NS}
P51	.11 ± .002	244	1.92^{NS}

NS Not significant

homozygous or single heterozygous genotypes the haplotypes were determined unambiguously; for double heterozygous individuals haplotypes were determined according to the maximum likelihood procedure, using the iterative process outlined by Hill (1974).

A highly significant linkage disequilibrium was observed in 18 out of 21 pairs tested (Table 3) involving the regions H5B to P33 and H33 to P51. Negative linkage disequilibrium (the common allele is associated with the rare allele) was found in any pair involving A9 and P51. Although the disequilibrium values Δ of P51 with H33, N33, P33, are relatively low, they are highly significant and they have reached their maximum (D'=1).

The disability in our data to detect a significant negative linkage disequilibrium between P51 and H5B/D9 (table 3) may be caused by the low minor allele frequencies of the RFLPs and therefore, the insufficient number of individuals screened to establish such linkage disequilibrium [Thompson *et al.* 1988]. Linkage disequilibrium measurements between these loci requires investigation in a larger sample size. RFLP A9 and P51 are in linkage equilibrium (Table 3) which cannot be explained by the lack of power between these pairs.

Table 3: *Disequilibrium statistics of COL2A1 RFLP pairs.* (Pairs are listed from 5' to 3' end)

RFLP pair A,B (n)	Distance (kb)	# of haplotypes observed[1]				Δ	D'	$4N_ek'$	$\Delta\chi^2$ (1 df)
		11	12	21	22				
H5B, A9[2] (166)	2.318	53.00	73.00	40.00	0.00	-.499	1.000	0.000	20.683***
H5B, D9[2] (246)	2.345	184.00	5.00	0.00	57.00	.946	1.000	0.000	110.089***
H5B,H33[2] (246)	15.710	148.96	40.04	5.04	51.96	.610	0.859	0.023	45.785***
H5B,N33[2] (242)	15.875	151.19	33.81	4.81	52.19	.650	0.869	0.020	51.089***
H5B,P33[2] (242)	15.886	152.20	32.80	4.80	52.20	.657	0.870	0.020	52.158***
H5B,P51[3] (242)	23.567	161.85	24.15	53.15	2.85	-.106	0.544	0.101	1.352[NS]
A9, D9[2] (166)	0.027	48.00	45.00	73.00	0.00	-.540	1.000	0.000	24.230***
A9, H33[2] (166)	13.392	38.95	54.05	68.05	4.95	-.532	0.809	0.039	23.535***
A9,NP33[2,4] (166)	13.563	40.58	52.42	70.42	2.58	-.557	0.893	0.019	25.765***
A9, P51[2] (164)	21.249	81.15	11.85	64.85	6.15	-.065	0.211	1.010	0.343[NS]
D9, H33[2] (248)	13.365	142.77	42.23	11.23	51.77	.532	0.713	0.072	35.160***
D9, N33[2] (244)	13.530	145.27	35.73	10.73	52.27	.576	0.734	0.063	40.512***
D9, P33[2] (244)	13.541	146.28	34.72	10.72	52.28	.583	0.736	0.062	41.470***
D9, P51[3] (244)	21.222	157.48	24.52	59.52	2.48	-.132	0.639	0.068	2.113[NS]
H33,N33[2] (244)	0.165	149.97	1.03	6.03	86.97	.939	0.981	0.237	107.560***
H33, P33[2] (244)	0.176	151.00	0.00	6.00	87.00	.949	1.000	0.000	109.767***
H33, P51[2] (244)	7.857	124.00	27.00	93.00	0.00	-.277	1.000	0.000	9.349***
N33, P33[2] (244)	0.011	156.00	0.00	1.00	87.00	.991	1.000	0.000	119.845***
N33, P51[2] (240)	7.692	127.00	26.00	87.00	0.00	-.263	1.000	0.000	8.290***
P33, P51[2] (240)	7.681	128.00	26.00	86.00	0.00	-.260	1.000	0.000	8.142***

NS =Not significant ***$p \leq 0.005$ [1]Haplotypes of double heterozygous individuals were assigned, incorporating linkage disequilibrium, according to the maximum likelihood procedure, with the iterative process outlined by Hill (1974).[2]No. of individuals required for detection of strong positive and/or strong negative linkage disequilibrium is sufficient.[3]No. of individuals required for detection of strong negative linkage disequilibrium is not sufficient.[4]Disequilibrium statistics of RFLP-pairs A9-N33 and A9-P33 are similar and therefore combined as a single entry in the table.

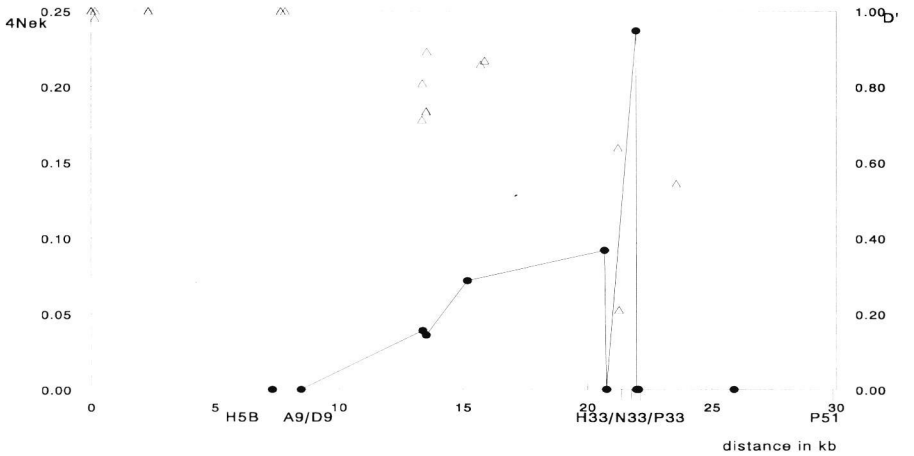


Figure 2. The graph shows in triangles the standardized linkage disequilibrium coefficient *D'* (right y-axis) against the physical distance (kb) between COL2A1 RFLP pairs (x-axis). Filled triangle indicates the A9-P51 pair which was in linkage equilibrium. The dots connected by a solid line give the standardized measures of recombination frequency per kb ($4N_ek$; left y-axis) estimated from the *D'* values of two adjacent RFLPs and, in order to smooth the graph, from the values of two RFLPs with one RFLP in between. The x-axis represents the physical distance (kb) calculated from the 3' end of the gene with the location of the RFLPs in the gene (see also Fig. 1). The resulting $4N_ek$ values (dots on the lines) were placed in the middle of the physical distances between the two RFLPs compared.

Relationship between Δ and physical distance

If significant linkage disequilibrium was solely determined by recombination, then *Δ* would decline with increasing physical distance. Since *D'* is less dependent on allele frequencies (see Leiterdorf *et al.* 1989 and linkage disequilibrium data on intron 33 RFLPs with P51) we plotted *D'* against the physical distance separating RFLP pairs. Fig. 2 shows (trianglesand right y-axis) that in general the RFLP pairs investigated follow the theoretically predicted negative relationship. The estimate of standardized recombination frequency per kb ($4N_ek$) for each RFLP pair was also calculated with *D'* (Table 3). The dots and line in Fig.2 (left y-axis) shows that recombinations in the COL2A1 gene have occurred between RFLP D9 and H33 at a low rate. In addition, a relatively high recombination rate is visible between H33 and N33, which might indicate a recombination hotspot at this site.

Haplotype assignment

The informativeness of RFLPs can markedly be improved by constructing haplotypes consisting of defined patterns of RFLP alleles linked together on a chromosome. Without family data haplotypes can only be determined unambiguously if individuals are homozygous at all RFLP loci or heterozygous at only one. The proportion of haplotypes constructed from 7 RFLPs that could be determined unambiguously is shown in table 4. Remaining haplotypes were derived assuming the lowest number of recombination and/or mutation events, and following the pairwise linkage disequilibrium results.

Table 3 shows that for RFLP-pairs H5B-D9, H33-P33 and N33-P33 which are in positive maximal linkage disequilibrium all double heterozygous individuals are assigned by the maximum likelihood procedure, using the iterative process outlined by Hill (1974) as 1-1 and 2-2. Based on these calculations individual double heterozygous for these loci and homozygous for the 5 remaining loci, haplotypes could be assigned on the basis of these calculations (Table 4).

For each RFLP-pair which is in negative maximal linkage disequilibrium (H5B-A9, A9-D9, H33-P51, N33-P51, P33-P51) haplotype 2-2 is not expected to be found in

Table 4: *Haplotype assignment of RFLPs in the COL2A1 gene.*

Haplotype	H5B	A9	D9	H33	N33	P33	P51	#unambiguously determined (% of total)	Total #	estimated haplotype frequency
A	1	2	1	1	1	1	1	39 (64%)	61	0.37
B	1	1	1	1	1	1	1	8 (42%)	19	0.12
C	2	1	2	2	2	2	1	4 (11%)	36	0.22
D	1	1	1	2	2	2	1	3 (18%)	17	0.10
E	1	2	1	2	2	2	1	1 (50%)	2	0.01
F	1	1	1	1	1	1	2	4 (40%)	10	0.06
G	1	2	1	1	1	1	2	3 (50%)	6	0.04
H	1	1	1	2	1	1	1	1 (50%)	2	0.01
I	1	2	1	2	1	1	1	0	2	0.01
J	1	1	2	1	1	1	1	1 (20%)	5	0.03
K	2	1	2	1	1	1	1	1 (50%)	2	0.01
L	2	1	2	1	1	1	2	0	2	0.01
M	1	-	1	1	2	1	1	0	1	0.01
Total								65 (40%)	165	

the population investigated. Following these calculations each individual which is double heterozygous for these loci and homozygous for the remaining loci were therefore assigned following the pairwise disequilibrium results (Table 4). Furthermore, for loci which are in high but not maximal linkage disequilibrium a (small) error is made in the haplotype assignment of 7 RFLPs. For example between RFLP-pair H5B-H33, with a significant positive linkage disequilibrium of 86% of its maximum value, haplotype 2-1 is not absent but rare in our population (Table 3). The rare haplotype 2-1 was unambiguously found in 4 individuals (not double heterozygous). The iterative process calculated that in the 41 double heterozygous individuals found, 1.04 extra haplotype 2-1 should be present (in total 5.04; Table 3). For this 1.04 haplotype (approximately 21% of total) we did not correct. This means that the rare haplotype 2-1 (H5B-H33) is visible both in haplotype K and L in Table 4, and that the double heterozygous individual between H5B-H33 were assigned as 1-1 and 2-2. For other RFLP-pairs with linkage disequilibrium results varying between 63 and 98 percent of themaximum values, comparable but usually smaller errors are made (varying between 22% and 0.5% of total). As a result of constructing haplotypes in this way an overestimation (rather small) in the common haplotype frequency calculation may exist and an underestimation in haplotype frequency of rare haplotypes (compare table 3 and 4). For RFLP pair A9-P51, which is in linkage equilibrium, the 8 double heterozygous individuals were assigned following the ratios calculated by the iterative process between haplotype 1-2 and 2-1. In total 13 different haplotypes out of 128 expected with 7 RFLPs could be recognized (Table 4), showing that only a fraction of the possible haplotypes is actually found in the sample, and that haplotype frequencies vary considerably.

Evolutionary relations

An evolutionary tree was derived using a minimum spanning tree approach. Fig. 3 shows the network of all haplotypes on euclidean reconstructed distances such that the number of RFLP differences between two haplotypes over the graph is minimized. A possible phylogeny is indicated with the haplotypes most parsimoniously connected following the most likely evolutionary history. Each link in Fig. 3 requires at least one mutation and the RFLP(s) involved in the mutation(s) are indicated at each line. The phylogenetic tree was further based on the assumption that the age of a haplotype is on average directly proportional to its frequency [Watlarson & Guess 1977]. Haplotype H for example, varying from both haplotypes B and I at one RFLP site (Table 4), is indicated as derived from common haplotype

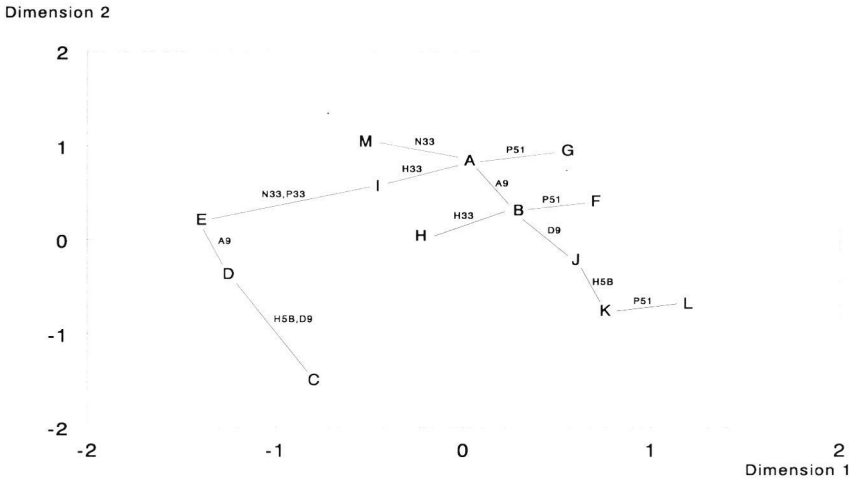


Figure 3. Minimum spanning tree on euclidean constructed distances such that the number of differences over the graph is minimized. Each letter represents an individual haplotype corresponding to the letters in Table 4. A possible phylogeny is indicated following the most likely evolutionary history. The RFLP sites associated with each link in the network represent the RFLP that is mutated.

B. At least two candidate haplotypes A and C, which are both common and almost completely opposite (Table 4), satisfy the properties of being ancestral. The remaining haplotypes might be the result of mutations, recombinations, and founder effects.

Discussion

RFLPs of the COL2A1 gene

Allele frequencies and pairwise linkage disequilibria were investigated at RFLP sites covering the entire COL2A1 gene. The allele frequency of RFLP P51 (0.11) has not yet been reported in other population studies. Although P51, introducing an amino acid substitution, could be under selective pressure, the arrangement of alleles in genotypes was not significantly different from those expected for populations in HWE (Table 2). The allele frequencies of 6 RFLP sites (H5B, A9, D9, H33, N33, P33) did not differ significantly from those reported for other Caucasian populations [Williams *et al.* 1992, Eng & Strom 1985, Nunez *et al.* 1985, Sykes *et al.*

1985, Väisänen *et al.* 1988, Ogilvie *et al.* 1987, Hull & Pope 1989, Sokolov *et al.* 1991].

Pairwise linkage disequilibrium analysis at the COL2A1 locus

Pairwise linkage disequilibrium analysis was performed using each RFLP pair combination. A highly significant linkage disequilibrium was observed in 18 out of 21 pairs tested (Table 3) involving the regions H5B to P33, and H33 to P51. A negative disequilibrium was found in any pair involving P51 and A9 (negative Δ-values). Although linkage disequilibrium values of intron 33-P51 RFLP pairs have reached their maximum (D'=1.00) and are highly significant, the Δ values are relatively low which probably reflects the dependence of Δ on low frequency alleles. The expected negative linkage disequilibrium between H5B and D9 with P51 is not significant which is also reflected in the relative low D' (0.54 and 0.64, respectively). This may be explained by a lack of power (low minor allele frequency and small number of individuals) and the presence of a rare haplotype which has then a substantial effect on negative linkage disequilibrium parameters and requires investigation in a larger sample size.

The linkage equilibrium between A9 and P51 cannot be explained by a lack of power. The possibility of a recombination hotspot between A9 and P51 is rather unlikely in view of the high significant linkage disequilibrium between A9 with H33 and P51 with H33. Alternatively this can be explained by the evolutionary history of these polymorphisms. In the relationship between Δ and physical distance and haplotype analysis this issue is further discussed.

Relationship between Δ and physical distance

Overall, the physical distance and disequilibrium parameters of the COL2A1 sites investigated follow the theoretically predicted negative relationship (Table 3). Since the absolute value of Δ is relatively low in linkage disequilibria between RFLPs with low frequency alleles (s.a. intron 33 RFLPs with P51 and Leitersdorf *et al.* 1989) we used D' to estimate the standardized recombination frequency per kb ($4N_ek$) for each RFLP pair (Table 3). Figure 2 shows recombinational events between D9 and H33 at a low rate and a recombination hotspot between H33 and N33. Recombination in this region is rather unlikely in view of the small distance (165 bp) and the high significance of the linkage disequilibria between P33 and D9/H5B (Table 3). This result shows that a monotonic relationship between disequilibrium and distance is only present when $4N_ek$ is large. In small genomic segments the effects of recombination may be overcome by factors such as drift, mutations, admixture, and

gene conversion. The possibility of point mutations occurring at H33 and/or N33, however, would complement the observation of a cluster of high frequency polymorphisms due to point mutations present in a 2 kb area of the COL2A1 gene containing intron 33 [Vikkula & Peltonen 1989].

Haplotype assignment and evolutionary relations

The informativeness of RFLPs can markedly be improved by constructing haplotypes consisting of defined pattern of RFLP alleles linked together on a chromosome. By using 7 RFLPs at least 13 different haplotypes of the COL2A1 could be recognized in our population 4 of which have frequencies estimated to be $\geq$ 10%.

An evolutionary tree was derived using a minimum spanning tree approach. Fig. 3 shows all haplotypes on euclidean reconstructed distances such that the number of differences measured over the graph is minimized. Based on the assumption that ancestral haplotypes may be more common then derived haplotypes [Watlarson & Guess 1977], a possible phylogeny is indicated with the haplotypes most parsimoniously connected following the most likely evolutionary relations. According to these data 14 mutations (that is on average 2 mutations at each RFLP site) must have occurred to explain all haplotypes. This large number of mutations would be consistent with each of these RFLPs being mutational hotspots. Another plausible explanation might be recombinational events, since this would put a RFLP on different haplotypes and simulate the effect of multiple mutation [Leitersdorf *et al.* 1989]. The occurrence of haplotype D, for example, might have arisen by a recombination event of common haplotype B and C between RFLP D9 and H33 instead of 3 mutational events. Finally, these results might be explained by founder effects of several ancestral haplotypes. At least two candidate haplotypes A and C which are both common and almost completely opposite satisfy the properties of being ancestral (Table 4).

Together, these results increase the knowledge on structural aspects of the current COL2A1 haplotypes in a Caucasian population and evolutionary events that have most likely created these haplotypes. Our data indicate that disease related population studies involving the COL2A1 gene should include minimal 4 RFLPs (D9, A9, H33, P51) to obtain 98% of possible haplotypes occurring. The panel of different haplotypes observed in this study will facilitate future disease association studies aimed at the identification of common COL2A1 mutations.

Acknowledgements

We thank Dr. P de Knijff (TNO-PG Leiden) for help with the statistics, Dr. G.J. te Meerman for suggesting and computing the minimum spanning tree representation (Dept. of Medical Genetics, University of Groningen) and Marianne Rots for technical support. This work was supported by the TNO-Health organization and "Het Nationaal Reumafonds" of the Netherlands.

References

- ALA-KOKKO, L., BALDWIN, C.T., MOSKOWITZ, R.W., & PROCKOP, D.J. (1990) Single base mutation in the type II procollagen gene (COL2A1) as a cause of primary osteoarthritis associated with a mild chondrodysplasia. *Proc. Natl. Acad. Sci. USA* **87,** 6565-6568.
- BALDWIN, C.T., REGINATO, A.M., SMITH C., JIMENEZ, S.A., & PROCKOP, D.J. (1989) Structure of cDNA clones coding for human type II procollagen. The α1(II) chain is more similar to the α1(I) chain than two other α chains of fibrillar collagens. *Biochem. J.* **262**, 521-528.
- CHAKRAVARTI, A., BUETOW, K.H., ANTONARAKIS, S.E., WABER, P.G., BOEHM, C.D., KAZAZIAN, H.H. (1984) Nonuniform recombination within the human ß-globin gene cluster. *Am. J. Hum. Genet.* **36**, 1239-1258.
- CHAKRAVARTI, A., ELBEIN, S.C., & PERMUTT, M.A. (1986) Evidence for increased recombination near the human insulin gene: Implication for disease association studies. *Proc Natl Acad Sci USA* **83,** 1045-1049.
- ENG, C.E.L., & STROM, C.M. (1985) Analysis of three restriction fragment length polymorphisms in the human type II procollagen gene. *Am. J. Hum. Genet.* **37,** 719-732.
- GOLDING, G.B. (1984) The sampling distribution of linkage disequilibrium. *Genetics* **108**, 257-274.
- HEDRICK, P.W. (1987) Gametic disequilibrium measures:proceed with caution. *Genetics* **117,** 331-341.
- HILL W.G., & ROBERTSON A. (1968) Linkage disequilibrium in finite populations. *Theor Appl Genet* **38**, 226-231.
- HILL, W.G. (1974) Estimation of linkage disequilibrium in randomly mating populations. *Heredity* **33,** 229-239.
- HOLDERBAUM, D., MALEMUD, C.J., MOSKOWITZ, R.W., & HAQQI, T.M. (1993) Human cartilage from late stage familial osteoarthritis transcribes type II collagen mRNA encoding a Cysteine in position 519. *BBRC* **192**, 1169-1174.
- HUDSON, R.R. (1985) The sampling distribution of linkage disequilibrium under infinite allele model without selection. *Genetics* **109**, 611-631.
- HULL, R., & POPE, F.M. (1989) Osteoarthritis and cartilage collagen genes. *The Lancet* **1,** 1337-1338.
- KELLGREN, J.H., LAWRENCE, J.S., BIER, F. (1963) Genetic factors in generalized osteoarthrosis. *Ann. Rheum. Dis.* **22**, 237-255.
- LEITERSDORF, E., CHAKRAVARTI, A., & HOBBS, H.H. (1989) Polymorphic DNA Haplotypes at the LDL receptor locus. *Am J Hum Genet* **44**, 409-421.
- LOUGHLIN, J., IRVEN, C., ATHANASOU, N., CARR, A., &SYKES, B. (1995) Differential Allelic expression of the type II collagen gene (COL2A1) in osteoarthritic cartilage. *Am J Hum Genet* **56**, 1186-1193.

- NUNEZ, A.M., FRANCOMANO, C., YOUNG, M.F., MARTIN, G.R., & YAMADA, Y. (1985) Isolation and partial characterization of genomic clones coding for a human Pro-α1(II) Collagen chain and demonstration of restriction fragment length polymorphism at the 3' end of the gene. *Biochemistry* **24,** 6343-6348.
- OGILVIE, D.J., WORDSWORTH, B.P., PRIESTLEY, L.M., DALGLEISH, R., SCHMIDTKE, J., ZOLL, B., SYKES, B.C. (1987) Segregation of all four major fibrillar collagen genes in the Marfan syndrome. *Am. J. Hum. Genet.* **41,** 1071-1082.
- PRIESTLEY, L., FERGUSSON, C., OGILVIE, D., WORDSWORTH, P., SMITH, R., PATTRICK, M., DOHERTY, M., & SYKES, B. (1991) A limited association of generalized osteoarthritis with alleles at the type II collagen locus: COL2A1. *Brit. J. Rheum.* **30,** 272-275.
- PROCKOP, D.J., OLSEN, A., KONTUSAARI, S., HYLAND, J., ALA-KOKKO, L., VASAN, N.S., BARTON, E., BUCK, S., HARRISON, K., & BRENT, R. (1990) Mutations in Human Procollagen genes. *Ann. New York Acad. Sci.* **580**, 330-339.
- PUN, Y.L., MOSKOWITZ, R.W., LIE, S., SUNDSTROM, W.R., BLOCK, S.R., MCEWEN, C., WILLIAMS, H.J., BLEASEL, J.F. HOLDERBAUM, D., & HAQQI, T.M. (1994) Clinical correlations of osteoarthritis associated with a single-base mutation (arginine^{519} to cysteine) in type II procollagen gene. *Arthritis Rheum.* **37**, 264-269.
- SOKOLOV, B.P., DZEMELINSKI, V.V., & KALININ, V.N. (1991) Restriction fragment length polymorphisms of three collagen genes in Russians and Byrats. *Hum Hered* **41**, 57-60.
- SYKES, B., SMITH, R., VIPOND, S., PATERSON, C., CHEAH, K., SOLOMON, E. (1985) Exclusion of the α1(II) cartilage collagen gene as the mutant locus in type IA osteogenesis imperfecta. *J. Med. Genet.* **22,** 187-191.
- SVED, J.A., (1971) Linkage disequilibrium and homozygosity of chromosomal segments in finite populations. *Theor Popul Biol* **2**, 125-141.
- THOMPSON, E.A., DEEB, S., WALKER, D., & MOTULSKY, A.G. (1988) The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII Apolipoprotein genes. *Am. J. Hum. Genet.* **42,** 113-124.
- VÄISÄNEN, P., ELIMA, K., PALOTIE, A., PELTONEN, L., & VUORIO, E. (1988) Polymorphic restriction sites of type II Collagen gene: their location and frequencies in the Finnish population. *Hum. Hered.* **38,** 65-71.
- VIKKULA, M., & PELTONEN, L. (1989) Structural analyses of the polymorphic area in the II collagen gene. *FEBS* **250** 171-174.
- VIKKULA, M., NISSILÄ, M., HIRVENSALO, E., NUOTIO, P., PALOTIE, A., AHO, K., & PELTONEN, L. (1993) Multiallelic polymorphism of the cartilage collagen gene: no association with osteoarthrosis. *Ann. Rheum. Dis.* **52**, 762-764.
- VISSING, H., D'ALESSIO, M., LEE, B., RAMIREZ, F., GODFREY, M., & HOLLISTER, D.W. (1989) Glycine to Serine substitution in the triple helical domain of pro-α1(II) collagen results in a lethal perinatal form of short-limbed Dwarfism. *J. Biol. Chem.* **264,** 18265-16267.
- WATLARSON, G.A., & GUESS, H.A. (1977) Is the most frequent allele the oldest? *Theor. Popul. Biol.* **11,** 141-160.
- WEIR, B.S., & HILL, W.G. (1986) Non-uniform recombination within the human ß-globin gene cluster. *Am J Hum Genet* **38**, 776-778.
- WILLIAMS, C.J., HARRISON, D.A., HOPKINSON, I., BALDWIN, C.T., AHMAD, N.N., ALA-KOKKO, L., KORN, R.M., BUXTON, P.G., DIMASCIO, J., CONSIDINE, E,L,, & PROCKOP, D.J. (1992) Detection of sequence variants in the gene for human type II procollagen (COL2A1) by direct sequencing of polymerase chain reaction-amplified genomic DNA. *Human Mutation* **1,** 403-416.

- WILLIAMS, C.J., & JIMENEZ, S.A. (1993) Heredity, genes and osteoarthritis. *Rheum. Dis. Clin. North America* **19**, 523-543.
- WILLIAMS, C.J., ROCK, M., CONSIDINE, E., MCCARRON, S., GOW, P., LADDA, R., MCLAIN, D., MICHELS, V.M., MURPHY, W., PROCKOP, D.J. & GANGULY, A. (1995) Three new point mutations in type II procollagen (COL2A1) and identification of a fourth family with the COL2A1 Arg519->Cys base substitution using conformation sensitive gel electrophoresis. *Hum. Mol. Genet.* **4**, 309-312.

6 ASSOCIATION OF A COL2A1 HAPLOTYPE WITH RADIOLOGICAL OSTEOARTHRITIS IN A POPULATION BASED STUDY: THE ROTTERDAM STUDY

I Meulenbelt[1,2], C Bijkerk[1,2], SCM de Wildt[1], HS Miedema[1], HA Valkenburg[2], FC Breedveld[3], HAP Pols[2], DL Knook[3], A Hofman PhD[2], CM van Duijn PhD[2], and PE Slagboom PhD[1]

[1]TNO Prevention and Health, Dept. of Vascular and Connective Tissue Research, Gaubius Laboratory, Leiden; [2]Dept. of Internal Medicine III and Dept. of Epidemiology & Biostatistics, Erasmus University Medical School, Rotterdam; [3]Dept. of Rheumatology, and Dept. of Internal Medicine, University Hospital, Leiden, The Netherlands.

In preparation

It was investigate whether radiographic osteoarthritis (ROA) in 55-65-year subjects is associated with specific haplotypes of the COL2A1 gene. Radiographs of knees, hips, hands, and spine were scored in subjects from a population-based cohort study on the presence of ROA. Cases had ROA in ≥ 3 joints. Controls, selected from the same population, had ROA in < 3 joints. Allele frequencies of 3 dimorphisms and a VNTR polymorphism of the COL2A1 gene were determined in the case and control group. The VNTR allele 14R2 and 2 dimorphisms seperately showed a significant association in cases with ROA ≥ 3 joints. When haplotype analysis of these 3 polymorphisms was performed a specific haplotype (1-2-14R2) was observed which showed a strong association with ROA ≥ 3 joints with a distinctly increased odds ratio (OR=5.5, 95% CI 2.3-13.1) as compared to the single allele associations. These results suggest that a specific haplotype of the COL2A1 locus may play a role in the generalization of the ROA process.

Collagen type II is the most abundant collagenous protein of epiphyseal and articular cartilage and consists of three α1(II)-chains in a helix. The amino acid structure of each chain consists of a repetitive Gly-X-Y structure [Baldwin et al. 1989] (1). A large number of mutations in the COL2A1 gene have been identified causing a spectrum of chondrodysplasias [Spranger et al. 1994] (2). Most of these mutations are dominant and generate substitution of glycine, interrupting the Gly-X-Y sequence. One mutation affecting the Y position was found in several apparently different families with a distinctive pattern of dominantly inherited generalized osteoarthrosis (GOA, defined as OA in ≥ 3 joint groups) associated with a mild chondrodysplasia [Ala-Kokko et al. 1990 (3); Holderbaum et al. 1993 (4), Pun et al. 1994 (5), Williams et al. 1995 (6), Eyre et al. 1991(7)]. Mutations at X and Y positions, in general, are expected to be milder than those of glycine substitutions and may be involved in sporadic GOA [Williams and Jiminez 1995] (8).

A significant effect of genetic influences on OA was demonstrated in patients which develop GOA in combination with Heberden nodes before the age of 65 [Kellgren et al. 1963] (9). In a study of mono- and dizygotic female twins, a heritability estimate of 0.39-0.65 was reported for radiological characteristics of OA (ROA) at hand and knee joints in women [Spector et al. 1996] (10). The role of COL2A1 in the etiology of sporadic OA has been studied in population samples of subjects affected by GOA or ROA. Thus far, these genetic association studies did not unequivocally demonstrate such a role for the COL2A1 gene. An increased frequency of a Bam HI COL2A1 allele (frequency 0.07) was reported in British

females with symptomatic, radiologically confirmed OA in more than one joint before the age of 60 [Hull and Pope 1989] (11). Secondly, a rare MaeII allele (frequency 0.03) in the 3' untranscribed region (3'UTR) was reported to be associated with GOA in combination with Heberden's nodes again in British subjects [Loughlin et al. 1995] (12). This allele was also associated with a decreased level of expression of the COL2A1 gene in osteoarthritic cartilage. In contrast, association and small sibling pair studies using several COL2A1 RFLP polymorphism [Priestley et al. 1991] (13) and a VNTR polymorphism located 1.35 kb to the 3' end of the COL2A1 gene did not show association in British GOA patients [Loughlin et al. 1994] (14) and neither diid the latter polymorphism in Finnish OA patients (> 1 joint ROA) or with OA specifically in the finger joints [Vikkula et al. 1993] (15). By studying the same polymorphism in a Dutch population based sample, we found [Bijkerk et al. submitted] (16) a significant predisposing association for subjects homozygous for the common 13R1 allele and ROA especially in hip, hand, and spine. Among cases with ROA in two or more joint groups including at least the knee or hip joint, carriers of the rare allele 14R2 had a 2.5 times higher risk of developing Heberden nodes. Furthermore, the 14R2 allele was increased in subjects with generalized ROA in knee, hand and spine in combination with Heberden's nodes. Together rather common COL2A1 alleles may predispose to a generalization of the ROA process.

These associations may be explained by a linkage disequilibrium model in which an OA causing mutation is in linkage disequilibrium with the VNTR allele. By defining haplotypes of this allele including additional COL2A1 polymorphisms it is possible to delineate the risk allele, resulting in an optimized disease association with a particular ancestral haplotype, and excluding all other alleles. Pairwise linkage disequilibrium analysis of the COL2A1 gene with 7 intragenic RFLP sites in a population of randomly selected Dutch individuals showed at least 13 different haplotypes of the COL2A1 gene [Meulenbelt et al. 1996] (17). By genotyping four of these RFLP sites 98% of all possible haplotypes occurring in the Dutch population could be distinguished, 6 of which had an estimated frequency large enough to be informative in association studies of this gene. These haplotypes were most importantly distinghuised by using RFLP markers in intron 9 and 33.

The current study is an expanded analysis of the associations we have previously found for the VNTR locus in ROA cases [Bijkerk et al. submitted] (16). Linkage disequilibrium analysis using a maximum likelihood estimation method [Long et al. 1995] (18), was performed with three COL2A1 RFLP polymorphisms (within intron 9, intron 33, and the Mae II polymorphism at 3'UTR) in addition to the VNTR polymorphism in a somewhat extended population. Estimated frequencies of

pairwise and three way haplotypes were used to delineate specific COL2A1 haplotypes putatively predisposing the development of generalised ROA.

Materials and methods

Subjects and measurements

Cases and controls were derived from a prospective population based cohort study of determinants and prognosis of chronic diseases in the elderly, the Rotterdam study [Hofman et al. 1991] (19). For the current study, we used a random subset of 385 men and 559 women drawn from the age category 55-65 years (N=944). Furthermore, subjects were added from a subset of 200 individuals, aged 55-70 years, selected from the Rotterdam study, on the basis of a positive (N=100) or negative (N=100) answer to the single question if probands had siblings affected with OA (not confirmed in any other way). Radiographs of the knees, hips, hands, and thoraco-lumbar spine were scored in these 200 additional subjects for the presence of ROA as described by Bijkerk et al. [submitted] (16). In this way 32 additional subjects with ROA in knee and/or hip joints and 61 subjects free of ROA in all investigated joints were [Bijkerk et al. submitted] (16).

Since haplotype analysis will ensue low frequency alleles and to investigate the generalization process of ROA, we compared alleles and haplotypes of unrelated subjects with ROA in $\geq$ 3 joint groups (knee, hip, hand and thoraco-lumbar spine) to those with ROA in $<$ 3 joint groups.

Genotyping

Genomic DNA was obtained in 849 subjects (90%). Genotypes of the VNTR polymorphism of the COL2A1 gene were determined as described by Bijkerk et al. [submitted] (16) and the nomenclature of alleles was used as previously described [Berg et al. 1993] (20). Genotypes of the Dra I polymorphism in intron 9 (D9) and the Hind III polymorphism within intron 33 (H33) were determined as described by Meulenbelt et al. (17). Genotypes of the Mae II polymorphism at the 3' UTR region were determined as described by Louglin et al. (1995) (12). The common alleles of the RFLP polymorphisms are designated as 1 and the rare alleles as 2.

Statistical analysis

Demographic variables (gender, age, and body mass index (BMI)) were compared between cases and controls by using t-tests for independent samples. Allele frequencies were assessed by counting alleles and calculating sample proportions. Tests for goodness of fit into Hardy Weinberg equilibrium were calculated using HWE-program (LINKUTIL package) [Ott, 1991] (21).

Haplotype frequencies for subgroups were determined by using the maximum likelihood estimation of Long et al. [1995] (18). In the analysis we performed the recommended "forward-selection strategy", whereby the test for global equilibrium is followed by a forward-selection, adding sets containing different (pairwise and three-way) disequilibrium coefficients to the most restricted model. In this way the number of necessary hypothesis tests will be minimized while at the same time the structure of disequilibrium will be described.

Fisher's exact test was used for comparisons between groups with an expected allele and/or haplotype number $\leq$ 5. The strength of association was estimated as the odds ratio (OR) using the statistical package SPSS. ORs are presented with 95% confidence intervals (95% CI). When possible, adjusted ORs were calculated by logistic regression models that

adjusted for gender, age (in years), and BMI (kg/m^2). Age and BMI were used as continuous variables after assessing that they led only to trivial differences for the estimators of interest as categorized dummy variables.

Results

Population description

Baseline characteristics of the population studied and the number of individuals in each group that was genotyped are shown in Table 1. Mean age, and body mass index (BMI) of subjects with ROA in ≥ 3 joints and subjects with ROA < 3 joints differed significantly (P=0.000). Gender specific differences in the ROA joint site pattern is shown by the relatively low percentage of men with ROA in ≥ 3 joint sites (31%).

Allele frequencies

Table 2 shows the frequency of allele 13R1 and 14R2 of the VNTR polymorphism and those of the dimorphism Dra I in intron 9 (D9), Hind III in intron 33 (H33), Mae II at 3'UTR (M2) of the COL2A1 gene in the population investigated. The arrangement of alleles into genotypes for the H33, M3, and VNTR polymorphism of the COL2A1 gene in subjects and subgroups thereof, did not show significant deviation from Hardy Weinberg expectations [Ott, 1991] (21). The D9 dimorphism in the overall cohort and in subjects with ROA in < 3 joints, however, was not in Hardy Weinberg equilibrium (for both P ≤ 0.001).

The 13R1 allele or homozygous 13R1 carriers did not show a significant association with ROA in ≥ 3 joint site as compared to subjects with ROA in < 3 joint sites. A significant association was, however, found for the 14R2 allele with ROA in

Table 1: Baseline population characteristics of unrelated individuals of 55 to 65 year from the Rotterdam study. Subjects were included on the basis of ROA in ≥ 3 joint sites (knee, hip, hand, spine). Controls were included on the basis of ROA at < 3 assessed joints.

Groups	< 3 ROA joints affected (N=724)	≥ 3 ROA joints affected (N=124)
no. of men (%)	301 (42)	38 (31)
mean age in years (SE)	60.3(0.1)*	62.0(0.3)*
mean BMI† (SE)	26.1(0.1)*	28.0(0.4)*

*P<0.001, †BMI=body mass index in kg/m^2

Table 2: Frequencies of COL2A1 alleles among subjects with ROA in < 3 joint sites and subjects with ROA in ≥ 3 joint sites

ROA groups	Alleles					n
	13R1	14R2	D9	H33	M2	
< 3 joints OA	0.41	0.06	0.22	0.38	0.09	1448
≥ 3 joints OA	0.46	0.11*	0.28†	0.46‡	0.05	248
≥ 3 joints OA + Heberden's nodes	0.49	0.14‡	0.34†	0.53‡	0.04	74

* significantly different from frequency of subjects with ROA < 3 joints ($P \leq 0.001$)
† significantly different from frequency of subjects with ROA < 3 joints ($P \leq 0.05$)
‡ significantly different from frequency of subjects with ROA < 3 joints ($P \leq 0.01$)

≥ 3 joint sites as compared to subjects with ROA in < 3 joint sites (Table 2). The crude odds ration of allele 14R2 with the remaining alleles as reference was 2.1, 95% CI 1.3-3.4. Subsequent adjustment for gender (OR=2.1, 95% CI 1.3-3.4), age (OR=1.7, 95% CI 1.1-2.8), and BMI (OR=2.2 95% CI 1.3-3.4) did not alter this effect largely. Interaction of the variables tested was not observed. A significant association was also observed for the dimorphisms H33 (crude OR=1.4, 95% CI 1.1-1.9) and D9 (crude OR=1.4, 95% CI 1.0-1.9) with ROA in ≥ 3 joints (Table 2). Adjustment for gender, age, and BMI did not alter these effects (results not shown). The strength of the observed association increased when subjects with ROA in ≥ 3 joints were selected for the presence of Heberden's nodes to an ORs for 14R2 of 2.7 95% CI 1.3-5.5, for H33 of 1.8 95%CI 1.2-2.9, and for D9 of 1.8 95% CI 1.1-3.0 (Table 2). The allele frequency of the M2 dimorphism did not show significant deviation.

Linkage disequilibrium analysis

Linkage disequilibrium analysis was performed with polymorphism D9, H33, and the VNTR polymorphism using the maximum likelihood estimation of Long et al. 1995 (18) and following their recommended hypothesis testing strategy. Estimated pairwise haplotype frequencies of RFLP polymorphism D9 with H33 and H33 with M2 were determined. The rare allele of M2 was in complete negative linkage disequilibrium with polymorphism H33. Polymorphism M2 was, therefore, not further used in the haplotype analysis. The significant ($p \leq 0.001$) non-random association of D9 and H33 alleles is demonstrated by a low frequency of haplotype 2-1 (freq. 0.02).

The measure of general linkage disequilibrium of the three other polymorphism of the COL2A1 gene (D9, H33 and the VNTR), was analyzed in subjects with ROA in ≥ 3 joints and in subjects with ROA in < 3 joints. In both groups a highly significant global linkage disequilibrium was observed between the three loci ($p<0.0001$). To identify subsequent locus pairs in disequilibrium, the analysis was performed by contrasting the three loci with each other. In both groups a strong and highly significant ($p<0.0001$) linkage disequilibrium was observed between all possible pairs (D9-H33, D9-VNTR, H33-VNTR). Finally, the three way disequilibrium coefficient was added to the test, which, showed a significant deviation among subjects with ROA in < 3 joints ($P=0.002$), however was not significant in subjects with ROA in ≥ 3joints ($P=0.26$).

ROA association haplotype analysis

To investigate whether the observed associations could be subscribed to one particular haplotype, haplotype frequencies of the 14R2 allele in combination with the dimorphisms D9 and H33 were calculated by using the maximum likelihood estimation method. Estimated haplotype frequencies of subjects with ROA in < 3 joint groups and of subjects with ROA in ≥ 3 joints are shown in Table 3. A highly significant ($P<0.001$) increase of haplotype 1-2-14R2 (D9-H33-VNTR) observed in subjects with ROA in ≥ 3 joint groups as compared to ones with ROA in < 3 joints. The strength of association of the 1-2-14R2 haplotype in subjects with ROA in ≥ 3 joint measured as the OR demonstrated a distinct increase (5.5, 95% CI 2.3-13.1) as compared to the separate odds ratio of allele 14R2 (2.1, 95% CI 1.3-3.4) and of allele 2 of the H33 dimorphism (1.4, 95% CI 1.1-1.9). A haplotype analysis comparable to that for 14R2 did not reveal specific sub haplotypes for the 13R1 allele to be associated with ROA in ≥ 3 joints.

Table 3: Three way haplotype frequencies

ROA groups	Haplotypes (D9-H33-14R2)				n
	1-1-14R2	1-2-14R2	2-1-14R2	2-2-14R2	
< 3 joints OA	0.02	0.01*	0.01	0.03	1448
≥ 3 joints OA	0.03	0.04*	0.00	0.04	248

*P<0.001

Discussion

It was investigated whether the reported associations of COL2A1 VNTR allele 14R2 and 13R1 with generalized ROA (at ≥ 3 joint groups) in 55-65 year old individuals could be subscribed to specific haplotypes [Bijkerk et al. submitted] (16). The 14R2 allele showed significant association in cases with ROA in ≥ 3 joint sites and an increased OR was observed when cases were stratified by the presence of Heberden's nodes. These results of the current study indicate that the presence of Heberden's nodes contributes to the strength of the association, however, is not necessary nor sufficient. Dimorphism H33 and D9 showed significant association with the occurrence of ROA in ≥ 3 joints independent of gender, age, and BMI. The strength of these associations was, however, relatively small with a crude OR of 1.4, 95% CI 1.1-1.9 and 1.4 95% CI 1.0-1.9, respectively. In contrast to all other polymorphisms (H33, M2, VNTR), genotypes of the D9 dimorphism in the cohort and in subjects with ROA < 3 joint sites did not show Hardy Weinberg equilibrium, especially due to relatively high number of homozygotes D9 (2/2) and less heterozygous subjects (1/2).

Linkage disequilibrium analysis in the population under study showed a strong and highly significant ($p<0.0001$) pairwise linkage disequilibrium between the polymorphisms (D9-H33, D9-VNTR, H33-VNTR). The subsequent addition of the three way linkage disequilibrium coefficient was, however, not significant for subjects with ROA in ≥ 3 joints, indicating that the pairwise disequilibria observed are independent of the third locus. A specific haplotype (1-2-14R2) was observed when polymorphism D9 and H33 were combined with the VNTR allele 14R2 which showed a strong association with ROA ≥ 3 joints with a distinctly increased odds ratio (OR=5.5, 95% CI 2.3-13.1) as compared to the single allele 14R2 and H33 allele 2 association (OR=2.1, 95% CI 1.3-3.4 and 1.4, 95% CI 1.1-1.9, respectively). These results indicate that the putative mutation causing the observed association with allele 14R2 and H33 may be on one particular haplotype. The observed association with allele 2 of dimorphism D9 alone may be of other origin. Since haplotype frequencies were assigned by the maximum likelihood method, logistic regression models that adjust for age, BMI, and gender could not be performed. Logistic regression analysis showed that the effect of the 14R2 allele separately was not explained by differences between cases and controls in age, gender, or BMI.

The haplotyping results suggests that the reported 14R2 allele association may be based on a linkage disequilibrium with a mutation causing generalised ROA (≥ 3 joint groups). The increased OR (2.1 to 5.5) for the 1-2-14R2 haplotype (only

present in heterozygous form) indicates the presence of a putative dominant or codominant mutation on this specific COL2A1 allele. Although the estimated frequency of this haplotype is low (0.01), the putative mutation could very well be present also on other COL2A1 alleles and, therefore, be more frequent.

Association with the 13R1 allele or the homozygous genotype 13R1/13R1 was not observed when subjects with ROA in $\geq$ 3 joints and those with ROA in < 3 joints were compared. A significant association was, however, observed for the 13R1 allele and homozygous genotype and cases with ROA in hip, hand, and spine when compared to controls free of ROA in the peripheral joints (i.e. knees, hips and hands) (results not shown). This allele and/or genotype may thus predispose specifically to a subgroup of cases with ROA in hip, hand, and spine. Furthermore, estimated haplotype frequencies of the 13R1 allele did not show significant deviations neither when individuals with ROA $\geq$ 3 joints were analyzed, nor for those with ROA in hip, hand, and spine. The underlying putative mutation of the specific hip, hand, spine ROA association may be distributed over many haplotypes (since 13R1 is a frequent allele), or the observed association with the 13R1 allele may be spurious.

Together COL2A1 gene associations with generalized ROA have been reported for the COL2A1 polymorphisms, Bam HI (frequency 0.07) [Hull and Pope 1989] (11), M2 (frequency 0.03) [Loughlin et al. 1995] (12) and the D9-H33-VNTR 1-2-14R2 haplotype (frequency 0.02; this study). Furthermore, it was shown that polymorphism H33 was in complete positive linkage disequilibrium with polymorphism Bam HI in a British study [Priestley et al. 1991] (13) and in negative linkage disequilibrium with M2 in the current Dutch population. The observation that the 1-2-14R2 haplotype and the Bam HI allele predispose to OA in Dutch and British individuals, respectively in addition to the observed complete (BamHI-H33) linkage disequilibrium, may indicate that these association may be subject to one particular functional COL2A1 variant. This hypothesis needs conformation in both studies preferably by haplotype analysis.

The association of haplotype 1-2-14R2 (D9-H33-VNTR) in the Dutch and that with M2 in the British in combination with the complete negative linkage disequilibrium observed between H33-M2 (absence of the haplotype H33-M2 2-2), suggests that these associations may be of different origin. Secondly, the allele frequency of the Mae II allele in both our case and control group appeared to be significantly larger (frequency 0.07 and 0.09, respectively) as compared to the British study and did not show any association. The gene variant detected by Mae II associated with a decreased COL2A1 expression and an increased susceptibility to GOA in the British

population is either not present in the Dutch population or resides, due to a different linkage disequilibrium, on another haplotype.

In summary the current study shows that the 1-2-14R2 (D9-H33-VNTR) allele of the COL2A1 is strongly associated to a generalized pattern of ROA in 55-65 year old individuals. Furthermore, this study shows that determination of haplotype frequencies by using linkage disequilibrium analysis provide more insight into the origin of detected associations.

References

1 Baldwin CT, Reginato AM, Smith C, Jimenez SA, and Prockop DJ (1989) Structure of cDNA clones coding for human type II procollagen. The α1(II) chain is more similar to the α1(I) chain than two other α chains of fibrillar collagens. Biochem J 262:521-528.

2 Spranger J, Winterpacht A, Zabel B (1994) The type II collagenopathies: a spectrum of chondrodysplasias. Eur J Pediatr 153:56-65.

3 Ala-Kokko L, Baldwin CT, Moskowitz RW, and Prockop DJ (1990) Single base mutation in the type II procollagen gene (COL2A1) as a cause of primary osteoarthritis associated with a mild chondrodysplasia. Proc Natl Acad Sci USA 87:6565-6568.

4 Holderbaum D, Malemud CJ, Moskowitz RW, & Haqqi TM (1993) Human cartilage from late stage familial osteoarthritis transcribes type II collagen mRNA encoding a Cysteine in position 519. BBRC 192: 1169-1174.

5 Pun YL, Moskowitz RW, Lie S, Sundstrom WR, Block SR, Mcewen V, Williams HJ, Bleasel JF, Holderbaum D, Haqqi TM (1994) Clinical correlations of osteoarthritis associated with a single-base mutation (arginine^{519} to cysteine) in type II procollagen gene. Arthritis Rheum 37:264-269.

6 Williams CJ, Rock M, Considine E, Mccarron S, Gow P, Ladda R, Mclain D, Michels VM, Murphy W, Prockop DJ, Ganguly A (1995) Three new point mutations in type II procollagen (COL2A1) and identification of a fourth family with the COL2A1 Arg519->Cys base substitution using conformation sensitive gel electrophoresis. Hum Mol Genet 4:309-312.

7 Eyre DR, Weis MA, and Moskowitz RW (1991) Cartilage expression of a type II collagen mutation in an inherited form of Osteoarthritis associated with a mild chondrodysplasia J Clin Invest 87: 357-361.

8 Williams CJ, Jimenez SA (1995) Heritable diseases of cartilage caused by mutations in collagen genes. J Rheum 22:1(supplement) 28-33.

9 Kellgen JH, Lawrence S, and Bier F (1963) Genetic factors in generalized osteo-arthrosis. Ann Rheum Dis 22:237-253.

10 Spector TD, Cicuttini F, Baker J, Loughlin J, Hart DJ: Genetic influences on osteoarthritis in women: a twins study. BMJ 312:490-944, 1996.

11 Hull R, and Pope FM (1989) Osteoarthritis and cartilage collagen genes. Lancet:1337-1338.

12 Loughlin J, Irven C, Athanasou N, Carr A, and Sykes B (1995) Differential allelic expression of the type II collagen gene (COL2A1) in osteoarthritic cartilage. Am J Hum Genet 56:1186-1193.

13 Priestley L, Fergusson C, Ogilvie D, Wordsworth P, Smith R, Pattric M, Doherty M, and Sykes B (1991) A limited association of generalized osteoarthritis with alleles at the type II collagen locus: COL2A1. Brit J Rheum 30:272-275.

14 Loughlin J, Irven C, Fergusson C and Sykes B (1994) Sibling pair analysis shows no linkage of generalized osteoarthritis to the loci encoding type II collagen, cartilage link protein or cartilage matrix protein. Br J Rheum 33:1103-1106.
15 Vikkula M, Nissilä M, Hirvensalo E, Nuotio P, Palotie A, Aho K, Peltonen L. (1993) Multiallelic polymorphism of the cartilage collagen gene: no association with osteoarthrosis. Ann Rheum Dis 52:762-764.
16 Bijkerk C, Meulenbelt I, Odding E, Valkenburg H, Miedema H, Breedveld P, Hofman A, Te Koppele J, Pols H, Van Duijn C, and Slagboom PE. Association of two alleles of the COL2A1 gene with radiological osteoarthritis in a population based study. The Rotterdam Study. Submitted for publication.
17 Meulenbelt I, Williams CJ, Te Koppele JM, Van de Giessen GC, and Slagboom PE (1996) Population haplotype analysis and evolutionary relations of the COL2A1 gene. Ann Hum Genet 60:189-199.
18 Long JC, Wiliams RC and Ubanek M (1995) An E-M Algorithm and testing strategy for multiple-locus haplotypes. Am J Hum Genet 56:799-810.
19 Hofman A, Grobbee DE, De Jong PTVW, Van den Ouweland FA (1991) Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidem 7:403-422.
20 Berg ES, Olaisen B (1993) Characterization of the COL2A1 VNTR polymorphism. Genomics 16:350-354.
21 Ott J: Analysis of human genetic linkage, revised ed. Baltimore: John Hopkins University Press, 1991

7 INVESTIGATION OF THE ASSOCIATION OF THE CRTM AND CRTL1 GENES WITH RADIOLOGRAPHICALLY EVIDENT OSTEOARTHRITIS IN SUBJECTS FROM THE ROTTERDAM STUDY

I Meulenbelt[1], C Bijkerk[1,2], SCM de Wildt [1], HS Miedema[1], HA Valkenburg[2], FC Breedveld [3], HAP Pols[2], JM Te Koppele[1], VFG Sloos[1], A Hofman[2], PE Slagboom[1] and CM van Duijn [2]

[1]TNO Prevention and Health, Dept. of Vascular and Connective Tissue Research, Gaubius Laboratory, Leiden; [2]Dept. of Internal Medicine III and Dept. of Epidemiology & Biostatistics, Erasmus University Medical School, Rotterdam; [3]Dept. of Rheumatology, University Hospital, Leiden, The Netherlands.

Accepted for publication

Arthritis and Rheumatism

Objective. **To investigate whether radiographically evident osteoarthritis (ROA) in 55-65-year old men and women is associated with specific alleles or genotypes of the cartilage matrix protein (CRTM) and cartilage link protein (CRTL1) genes.**

Methods. **Cases were selected from a population-based study on the presence of ROA of the knee or hip. Radiographic analysis further included studies for spine and hand ROA. Controls, selected from the same population, were free of ROA in all joints.**

Results. **The CRTM locus was significantly associated with hip ROA in men (odds ratio 0.50, 95% confidence interval 0.26-0.95). A significant association between ROA and the CRTL1 gene was not observed.**

Conclusions. **These results suggest that the CRTM locus may play a role in the sex- and joint-site specific pattern of ROA development.**

Osteoarthritis (OA) is a degenerative disease of the joints which is clinically characterized by local joint pain, stiffness, limitation in movement, and deformity. Pathologic changes, visible on radiographs include osteophyte formation, joint space narrowing, bony sclerosis, and subchondral cysts. Although the relation between clinical signs and radiographic findings of OA is relatively poor (1,2), determination of the presence of radiographic characteristics of OA is a widely used indication of OA pathogenesis in population-based studies.

The occurrence of radiographically evident OA (ROA) and the pattern of joint involvement is sex dependent. In general, ROA in multiple joints occurs more frequently in females than in males. Between the ages of 55 and 64 years, 47% of women and only 29% of men show ROA in 4 or more joints (3). Men more often have ROA in the hip joints, whereas in women the joints of the hands and knees are most frequently affected (4,5).

A genetic influence in the onset of OA was first demonstrated by the finding of a significantly increased risk of OA in first-degree relatives of women with generalized OA (GOA; determined based on the presence of radiological signs of OA with Kellgren grade [6] $\geq$ 2 in 3 or more joint groups) in combination with Heberden's nodes (7). In a study of mono- and dizygotic female twins, Spector et al (8) reported a heritability estimate of 0.39-0.65 for different combinations of scores for osteophytes and joint space narrowing at hand and knee sites in women indicating a genetic effect in the development of OA at these sites. These findings indicate that genetic predisposition may contribute importantly to ROA, but to a varying extent in men and women depending on the joint site(s) considered. The relevance of the

genetic component varies among subgroups of patients, and it is as yet not clear which genes are involved.

Several association and sibling pair studies have investigated the possibility of predisposing type II procollagen gene (COL2A1) alleles in middle-aged patients with GOA (9,10,11,12,13). Results obtained in these studies, however, remain controversial and can explain only a fraction of the genetic component in OA. The contribution of other, noncollagenous, proteins of the extracellular matrix (ECM) to the etiology of ROA is an area of increasing interest. Genes encoding such proteins are the cartilage link protein (CRTL1) gene and the cartilage matrix protein (CRTM) gene. The cartilage link protein stabilizes proteoglycan aggregates by binding both to the proteoglycan core protein and to hyaluronic acid (14). The cartilage matrix protein is a component of cartilage collagen fibrils with an as-yet-unknown function (15). A study of these genes in a small number of sibling pairs (n = 38 pairs) expressing GOA and Heberden's nodes before the age of 60 years did not support any association between GOA and the CRTL1 or CRTM gene (12).

In the present study, we have investigated whether alleles and/or genotypes of the dinucleotide repeat polymorphism in the 3'-untranslated region of CRTM (16), and in the promoter region of CRTL1 (17) are associated with ROA in 55-65-year-old male and female subjects irrespective of possible manifestations of clinical OA. Cases (73 men and 107 women) had ROA in knee or hip joints and were stratified on the basis of ROA in hand joints. To maximize the difference between this case group and the control group, cases were compared to a control group (63 men and 72 women) consisting of subjects without any ROA at the joints assessed, i.e., knee, hip, hand, and spine.

Patients and methods

Subjects. We studied the association of the CRTM and CRTL1 genes with ROA in unrelated cases with ROA of the knee or hip joints and unrelated controls free of ROA in the knees, hips, hands, wrists, and the thoraco-lumbar spine. Cases and controls were derived from a prospective population-based cohort study of determinants and prognosis of chronic diseases in the elderly (the Rotterdam study) (18). All residents of a suburb of Rotterdam who were age $\geq$ 55 years were invited to participate in the Rotterdam study. In total, 7,983 participants were examined. The response rate was 78%. Informed consent was obtained from all subjects and the study was approved by the Medical Ethics Committee of the Erasmus University Medical School.

To distinguish genetic predisposition to ROA from other determinants, the current study was restricted to non-institutionalized participants ages 55-65 years. Radiographs of the knees and hips had previously been scored in a random subset of 1,040 individuals in this age category (425 men and 615 women) (19). From this subset, the radiographs of the hands and thoraco and lumbar spine were scored for the presence of ROA. The present study utilized a case-control design in which the contrast between cases and controls was

maximized with regard to ROA status. Cases included subjects with ROA in at least 1 or both knee joints (36 men and 90 women) or 1 or both hip joints (37 men and 17 women). Controls (63 men and 72 women) from the subset of 1,040 individuals were included in the study on the basis of absence of ROA in all radiographed joints. Given the sex differences in prevalence and site specific expression of ROA, the CRTM and CRTL1 gene associations with ROA were investigated in men and women separately. Furthermore, information on age (in years) and body mass index (BMI; measured as kg/m^2) was recorded.

Measurements. At the research center, weightbearing anterioposterior radiographs of the pelvis and knees, anteroposterior radiographs of the hands and wrists, and lateral radiographs of the spine (T-S1) were obtained. ROA was assessed by the grading system proposed by Kellgren and Lawrence (6). All radiographs were scored by 2 independent readers, who were blinded to all data of the participant. When the scores differed by >2 points or one reader assigned a score of 1 and the other assigned a score of ≥2, a consensus score was agreed upon. For each individual joint, definite ROA was defined as Kellgren-Lawrence score of ≥ 2. ROA of the hand was assessed indvidually in each inter- and metacarpophalangeal joint. For the carpometacarpal and intercarpal joints, only the first carpometacarpal and the trapezioscaphoidal joint were assessed. ROA of the wrist was assessed at the radiocarpal and distal radioulnar joints. Hand ROA was defined as a Kellgren-Lawrence score of ≥ 2 in at least 1 of the 36 joints that were scored (for this purpose the joints of the wrist were included in the hand ROA category). For the spine each individual level from T4 to S1 was scored with regard to osteophytes and disc space narrowing.

Genotyping. Genotypes of the dinucleotide repeat polymorphism in the 3'-untranslated region of the CRTM gene were determined as described by Wang et al (16 159, ?). Genotypes of the CA-repeat polymorphism in the promoter region of the CRTL1 gene were determined as described by Hecht et al (17) using the nomenclature and allelic ladder also describe by Hecht et al (17). Polymerase chain reaction PCR was performed in a reaction volume of 25 µl containing 25-50 ng genomic DNA, 2.5 pmoles of each primer, 1x Super *Taq* buffer (Sphaero Q; The Netherlands), 2 µCi α^{32}P-dCTP, 200 µmoles each of dCTP, dGTP, dTTP, and dATP, and 0.05 U of super Taq DNA polymerase (Sphaero Q). Amplification was initiated with 3 minutes denaturation at 94°C followed by 35 cycles of 15 seconds at 94°C, 30 seconds at appropriate annealing temperature, and 30 seconds at 72°C. The amplification was completed by a final incubation at 72°C for 3 minutes. Alleles were separated by electrophoresis through a denaturing polyacrylamide gel (6%) and analyzed by autoradiography.

One of the extra alleles detected in our population (CRTL1 A9.3; see below) showed a small length variation to A10. This was confirmed by heteroduplex analysis of genotypes containing CRTL1 A9.3 by denaturation of PCR products at 100°C for 5 minutes and re-annealing at room temperature for 1 hour. Heteroduplex patterns were analyzed by electrophoresis through 4-6% polyacrylamide gels and visualized by staining with ethidium bromide (20).

Statistical analysis. Demographic variables (age and BMI) were compared between cases and controls by using t-tests for independent samples. Allele frequencies for cases and controls were assessed by counting alleles and calculating sample proportions. Tests for goodness of fit into Hardy Weinberg equilibrium were calculated using the HWE program (LINKUTIL package) (21). For the multiallelic CRTL1 marker, the alleles with an observed allele frequency of < 0.05 were pooled and designated as allele AX. The likelihood ratio test was used to investigate association of CRTM and CRTL1 alleles with the occurrence of ROA (22). As an extension to this likelihood ratio test, a logistic regression model was used to adjust for covariables in the association between CRTM and CRTL1 alleles and ROA. The strength of association between an allele and the occurrence of ROA was estimated using

the odds ratio (OR). Adjusted OR were calculated by logistic regression models with adjustment for age (in years) and BMI as continuous variables, after determining that age and BMI as categorized dummy variables in the model led only to trivial differences for the estimators of interest. OR are presented with 95% confidence intervals (95% CI). The SPSS statistical package was used, and *P* values les than 0.05 were considered significant.

Results

Population description. Allele frequencies of dinucleotide repeat polymorphisms identified in the 3'-0untranslated region of the CRTM gene (16) and in the 5' promoter region of the CRTL1 gene (17) were determined in 73 men and 107 women with knee or hip ROA (cases) and 63 men and 72 women without ROA in the knee, hip, hand, or spine (controls).

Table 1. Baseline population characteristics of a random subset of 55-65 year-old subjects from the Rotterdam study*

Groups	ROA cases		ROA controls
	Knee (n=126)†	Hip (n = 54)‡	(n=135)
Men			
Age, years	61.0±0.3§	60.7±0.4	59.6±0.3
BMI, kg/m^2	26.8±0.3¶	26.5±0.4¶	25.3±0.3
No. (% of total)	36(29)	37(69)	63 (47)
No. with hand ROA (% of knee or hip ROA)#	15(42)	24(65)	-
Women			
Age, years	60.7±0.3§	60.9±0.5¶	59.3±0.3
BMI, kg/m^2	28.7±0.4§	25.1±0.5	24.9±0.4
No. (% of total)	90(71)	17(31)	72 (53)
No. with hand ROA (% of knee or hip ROA)#	68(76)	13(76)	-

*Cases had radiographically evident osteoarthritis (ROA) of the knee or hip. Controls were included on the basis of absence of ROA at the knee, hip, hand, and spine. Unless otherwise indicated, values are the mean ± SEM. BMI = body mass index. †Individuals with ROA exclusively in 1 or both knees. ‡Individuals with ROA exclusively in 1 or both hips. § $P < 0.01$ versus controls. ¶ $P < 0.05$ versus controls. # Men or women with hand ROA in addition to ROA of the knee or hip.

Table 1 shows the mean age and BMI of the men and women studied, and the number of individuals in each subgroup.The mean age and BMI of men and women with ROA differed significantly from the age and BMI of controls, except for the mean age of men with hip ROA and the mean BMI of women with hip ROA (Table 1). A difference in the frequency of ROA in specific joint sites among male and female cases is reflected in Table 1 as an excess of men in the hip ROA group (69%), with an OR (adjusted for age and BMI) of 2.6, 95% CI 1.3-5.3. In contrast, an excess of women in the knee ROA group (71%) and in the group with knee ROA in combination with hand ROA (76%) was observed (adjusted OR 1.9, 95% CI 1.1-3.4 and 4.0, 95% CI 1.9-8.3, respectively). The number of controls in the study population was small because only 135 individuals (of 1,040 55-65 year old subjects) were free of ROA in all jointsinvestigated.

CRTM gene association. Allele frequencies of the dinucleotide repeat polymorphism in the 3'-untranslated region of the CRTM gene were determined as described by Wang et al (16). Table 2 shows overall allele frequencies, allele frequencies among men and women with ROA in the knee or hip (cases), and allele frequencies among those without ROA in any of the joint groups investigated

Table 2. Allele frequencies of dinucleotide repeat polymorphism in the 3'-untranslated region of the cartilage matrix gene (CRTM) in ROA cases and in controls*

Group	Alleles (length in bp)			No. of alleles	*P*†
	1 (110)	2 (108)	3 (106)		
Overall	0.59	0.34	0.07	630	
Men					
Controls	0.51	0.40	0.09	126	
Knee ROA	0.56	0.40	0.04	72	≥0.50
Hip ROA	0.66	0.26	0.08	74	0.04
Women					
Controls	0.61	0.35	0.04	144	
Knee ROA	0.62	0.31	0.07	180	≥0.50
Hip ROA	0.59	0.35	0.06	34	≥0.50

* See Table 1 for explanations and definitions.
† By likelihood ratio association test ($\Lambda = -2Ln[L(H_0)/L(H_1)]$), cases versus controls.

(controls). In total, 3 alleles (A1-A3), similar to those described by Fujimori et al (23) were identified in 315 subjects. CRTM allele A4, previously described by Wang et al (16), was not observed in this population. Overall allele frequencies corresponded to those reported by Fujimori et al (23) and differed, especially for alleles A1 and A3, from those reported by Wang et al (16). The overall arrangement of alleles in genotypes in men and women was not significantly different from that expected for a population in Hardy-Weinberg equilibrium (P = 0.30 and P = 0.38, respectively). In addition, the subsequent Hardy-Weinberg equilibria tested for male and female cases and controls were not significant (male cases and controls P = 0.60 and P =0.61, respectively; female cases and controls P = 0.29 and P = 0.76, respectively).

As seen in Table 2, there was an association of the CRTM polymorphism in men with hip ROA as compared with controls (P = 0.04). This association was not observed in women. The contrast in men was caused by a decreased frequency of A2 in male cases with hip ROA and an increase of A2 in male controls, as compared with the overall frequency (which resembled the overall frequency reported by Fujimori et al [23]). The opposite was found for allele A1.

To estimate the strength of the association of the CRTM locus in men with hip ROA, the OR was calculated by entering the CRTM alleles into a logistic regression model. The corresponding crude OR for CRTM allele A2, with the most frequent allele (A1) as reference, for men with hip ROA was 0.50 (95% CI 0.26-0.95). Subsequent adjustment for age (OR 0.50, 95% CI 0.26-0.97) and BMI (OR 0.51, 95% CI 0.26-0.99) did not alter this effect, and there was no observed interaction of the variables tested. Since the ROA status of the control group was relatively rare in this age category of the general population (of whom only 13% had no ROA), the association found was also tested by comparing men with hip ROA with a less stringently defined control group, men without hip ROA (i.e., controls plus cases with knee ROA). The corresponding OR (adjusted for age and BMI) was 0.53 (95% CI 0.29-0.97). The effect of the CRTM genotype A2/A2, as compared with A1/A1, among men with hip ROA was not significant (adjusted OR 0.32, 95% CI 0.07-1.41).

CRTL1 gene association. Allele frequencies of the CA-repeat polymorphism in the promoter region of the CRTL1 gene were determined, using nomenclature similar to that described by Hecht et al (17). Table 3 shows the frequencies of most common alleles (> 5%) for men and women in partitioned case and control groups. Overall allele frequencies were similar to those reported by Hecht et al (17). Five extra low-frequency alleles (CRTL1 alleles A0, A9.2, A9.3, A11, and A12; not shown in Table 3) were identified in our study population, whereas CRTL1 allele A9 was not

Table 3. Allele frequencies of the dinucleotide repeat polymorphism in the promoter region of the cartilage link protein (CRTL1) gene in ROA cases and controls*

Group	Alleles (length in bp)							No. of alleles	P‡
	3 (236)	4 (234)	5 (230)	7 (226)	8 (225)	10 (222)	X†		
Overall	0.13	0.09	0.07	0.15	0.27	0.22	0.08	630	
Men									
Controls	0.13	0.12	0.06	0.10	0.33	0.18	0.07	126	
Knee ROA	0.15	0.14	0.10	0.06	0.26	0.22	0.07	72	≥0.50
Hip ROA	0.08	0.11	0.07	0.15	0.28	0.22	0.10	74	≥0.50
Women									
Controls	0.12	0.06	0.07	0.15	0.25	0.27	0.08	144	
Knee ROA	0.15	0.07	0.07	0.18	0.26	0.20	0.07	180	≥0.50
Hip ROA	0.12	0.09	0.00	0.21	0.29	0.27	0.03	34	≥0.50

* See Table 1 for explanations and definitions.
† Pooled low-frequency alleles.
‡ By likelihood ratio association test ($\Lambda = -2\ln[L(H_0)/L(H_1)]$), cases versus controls.

detected. Overall Hardy-Weinberg equilibrium was calculated for CRTL1 genotypes with an expected value of ≥ 5 and was not statistically significant ($P = 0.87$). The findings of tests for Hardy-Weinberg equilibrium in male and female case and control groups were also not significant (male cases and controls $P = 0.22$ and $P = 0.80$, respectively; female cases and controls $P = 0.22$ and $P = 0.47$, respectively). The likelihood ratio disequilibrium test investigating nonrandom association of CRTL1 alleles with the occurrence of ROA did not show evidence for association of CRTL1 alleles with ROA in either men or women ($P \geq 0.50$; Table 3).

Discussion

We investigated the role of the CRTM and CRTL1 gene loci in a population-based study of cases (age 55-65 years) with knee or hip ROA as compared with controls from the same population, without ROA in the knee, hip, hands, or spine. The allele frequencies of the dinucleotide repeat polymorphisms in the 3'-untranslated region of the CRTM gene in our study corresponded to those reported by Fujimori et al (23) and differed especially with regard to alleles A1 and A3, from those reported by

Wang et al (16). This may suggest that the Caucasian population described by Wang et al and colleagues was of a different origin (for example, with recent admixture) than the Caucasian populations in the present study and that of Fujimori and coworkers.

A significant association was found for the CRTM locus in the comparison of male controls and cases with hip ROA (OR 0.50, 95% CI 0.26-0.95). The fact that alleles A2 and A1 inversely contribute to the association may indicate that CRTM A2 is associated with a decreased risk for the development of hip ROA in men and A1 is associated with an increased risk, or that only one of the associations is true and the other is the consequence of compensating allele frequencies. The independent effect of these alleles could not be tested by exclusion of individuals with either CRTM A1 or A2 since the remaining number of individuals (carrying CRTM allele A3) was not sufficient. The likelihood ratio disequilibrium test did not show evidence for an association of CRTL1 alleles with ROA in either men or women ($P \geq 0.50$; Table 3).

A sib-pair analysis of British siblings expressing GOA and Heberden's nodes before the age of 60 years (12), did not reveal associations between ROA and the CRTL1 or CRTM genes. That study, however, included a small number of sibpairs (n = 38 pairs) with a relatively severe disease subset, and the subjects were not stratified by sex. Furthermore, the control group was not selected based on absence of ROA, but consisted of randomly selected individuals without clinical joint disability. This in contrast to the reference group in our study, which comprised 55-65 year-old subjects without ROA. The relatively strict selection criteria for our controls may have attenuated the association. In this age group the absence of ROA in all joint sites investigated is uncommon (only 13% of the cohort of 1,040 individuals).

In view of the presence of the CRTM protein in cartilage ECM and the close proximity of the dinucleotide repeat polymorphism to the gene, the reported association may be due to a role of this gene, and not another locus in linkage disequilibrium with the markers, in the etiology of ROA. However, since disease association studies are sometimes subject to false-positive results, the observed association requires confirmation in a second population-based sample. The association we have found follows the sex- and joint site-specific pattern of ROA development in the population. Typically, hip ROA is most frequently present in men in the 55-65-year age group (4,5). The observed association for this specific subgroup may be due to the larger number of individuals in this case group or it may reflect a sex- and/or joint-specific effect of the allele on the development of ROA.

An effect of genetic variation at the CRTM locus on, especially, the development

of hip ROA may be explained by the fact that the CRTM protein is expressed within the epiphyseal growth plate (postmitotic stages), during the process of endochondral bone formation (24,25). During this process, the length and shape of the bone are determined. Although the function of CRTM in epiphyseal cartilage remains unclear, it may affect the overall shape of the joint. Since ROA particularly of the hip is often considered to arise due to anatomic abnormalities (26), the possible effect of CRTM alleles on hip ROA might be exhibited in this way.

The results of this population-based study suggest that genotypic variation in the genes encoding matrix protein may play a role in the etiology of ROA. This gene may contribute to a sex- and joint-site specific pattern of ROA, but the nature of the effect of CRTM remains to be elucidated.

References

1 Claessens AAMC, Schouten JSAG, van den Ouweland FA, Valkenburg HA: Do clinical findings associate with radiographic osteoarthritis of the knee? Ann Rheum Dis 49:771-774, 1990
2 Hart DJ, Spector TD, Brown P, Wilson P, Doyle DV, Silman AJ: Clinical signs of early osteoarthritis: reproducibility and relation to x ray changes in 541 women in the general population. Ann Rheum Dis 50:467-470, 1991
3 Lawrence JS, Bremner JM, Bier F: Osteoarthrosis: prevalence in the population and relationship between symptoms and x-ray changes. Ann Rheum Dis 25:1-24, 1966
4 Lawrence JS, Degraaf R, Laine VAI: Degenerative joint disease in random samples and occupational groups. In, The Epidemiology of Chronic Rheumatism. Edited by JH Kellgren, MR Jeffrey, J Ball. Oxford, Blackwell Scientific, 1963
5 Van Saase JLCM, van Romunde LKJ, Cats A, Vandenbroucke JP, Valkenburg HA: Epidemiology of osteoarthritis: Zoetermeer survey. Comparison of radiological osteoarthritis in a Dutch population with that in 10 other populations. Ann Rheum Dis 48:271-280, 1989
6 Kellgren JH, Lawrence JS: Atlas of Standard Radiographs: The Epidemiology of Chronic Rheumatism. Vol. 2. Oxford Blackwell Scientific, 1963
7 Kellgren JH, Lawrence S, Bier F: Genetic factors in generalized osteoarthrosis. Ann Rheum Dis 22:237-253, 1963
8 Spector TD, Cicuttini F, Baker J, Loughlin J, Hart DJ: Genetic influences on osteoarthritis in women: a twins study. BMJ 312:940-944, 1996
9 Hull R, Pope FM: Osteoarthritis and cartilage collagen genes. Lancet 1:1337-1338, 1989
10 Priestley L, Fergusson C, Ogilvie D, Wordsworth P, Smith R, Pattric M, Doherty M, Sykes B: A limited association of generalized osteoarthritis with alleles at the type II collagen locus: COL2A1. Br J Rheumatol 30:272-275, 1991
11 Vikkula M, Nissilä M, Hirvensalo E, Nuotio P, Palotie A, Aho K, Peltonen L: Multiallelic polymorphism of the cartilage collagen gene: no association with osteoarthrosis. Ann Rheum Dis 52:762-764, 1993
12 Loughlin J, Irven C, Fergusson C, Sykes B: Sibling pair analysis shows no linkage of generalized osteoarthritis to the loci encoding type II collagen cartilage link protein or cartilage matrix protein. Br J Rheumatol 33:1103-1106, 1994
13 Loughlin J, Irven C, Athanasou N, Carr A, Sykes B: Differential allelic expression of the

type II collagen gene (COL2A1) in osteoarthritic cartilage. Am J Hum Genet 56:1186-1193, 1995
14 Franzén A, Björnsson S, Heinegård D: Cartilage proteoglycan aggregate formation: role of link protein. Biochem J 197:669-674, 1981
15 Winterbottom N, Tondravi MM, Harrington TL, Klier FG, Vertel BM, Goetinck PRF: Cartilage matrix protein is a component of the collagen fibril of cartilage. Dev Dyn 193:266-276, 1992
16 Wang Y, Sadler L, Hecht JT: Polymorphic dinucleotide repeat in a cartilage matrix protein (CRTM). Hum Mol Genet 1:780, 1992
17 Hecht JT, Wang Y, Rhodes C, Yamada Y: GT repeat polymorphism in the human proteoglycan link gene promoter region. Nucleic Acids Res 19:6666, 1991
18 Hofman A, Grobbee DE, De Jong PTVW, van den Ouweland FA: Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol 7:403-422, 1991
19 Odding E, Valkenburg HA, Grobbee DE, Hofman A, Pols HA: Locomotor disability in the elderly; the ERGO study (Erasmus Rotterdam Health and the Elderly). Ned Tijdschr Geneeskd 139:2096-3100, 1995
20 White MB, Carvalho M, Derse D, O'Brien SJ, Dean M: Detecting single base substitutions as heteroduplex polymorphisms. Genomics 12:301-306, 1992
21 Ott J: Analysis of human genetic linkage. Revised edition. Baltimore, John Hopkins University Press, 1991
22 Terwilliger JD: A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. Am J Hum Genet 56:777-787, 1995
23 Fujimori M, White PS, Marshall HN, Brodeur GM: Simple sequence repeat polymorphism in the cartilage matrix protein (CRTM) gene at 1p35. Hum Mol Genet 2:824, 1993
24 Chen Q, Johnson DM, Haudenschild DR, Goetinck PF: Progression and recapitulation of the chondrocyte differentiation program: cartilage matrix protein is a marker for cartilage maturation. Dev Biol 172:293-309, 1995
25 Mundlos S, Zabel B: Developmental expression of human cartilage matrix protein. Dev Dyn 199:241-252, 1994
26 Harris W: Etiology of osteoarthritis of the hip. Clin Orthop 213:20-33, 1986

8 GENERAL DISCUSSION

8 GENERAL DISCUSSION

Etiology of osteoarthritis

Osteoarthritis (OA) is a disease of complex etiology characterised in its early stages by degradation of cartilage and formation of new bone during the progression of the disease. Two major theories on the etiology of OA exist [1]. The first theory is based on cartilage failure (under normal or excessive stress) and the second hypothesis on the failure of chondrocytes to maintain the balance between cartilage synthesis and degradation. Failure of cartilage may be caused by abnormal stresses, due to obesity, anatomical abnormalities of bone, joint instability, trauma, or bone remodelling. Alternatively, failure of cartilage under normal stress may occur due to an abnormal cartilage structure. This may be caused by biochemical changes as a consequence of the ageing process or by genetic or metabolic disorders. OA is thought to arise by the occurrence of lesions in cartilage (especially in the collagen network) followed by a failure of the chondrocyte to repair the lesions. This may lead to an imbalance of cartilage synthesis and degradation with a net loss of cartilage (Figure 1).

Even small failures of cartilage integrity or a small imbalance between chondrocyte synthesis and breakdown may account for the onset of OA. With respect in particular to the high proteoglycan (PG) turnover and the highly organized collagen network which has no turnover, a subtle decrease in extracellular matrix (ECM) strength or a sub-optimal turnover balance could result, over many years, in cartilage degradation. When enough matrix is damaged or lost, the mechanical properties of cartilage will become insufficient to withstand mechanical load which inevitable leads to the onset of OA. In this thesis it an investigation was made of gene variants encoding ECM components contribute to the onset of generalised familial OA and to the presence of radiological signs of OA (ROA) in middle- aged persons.

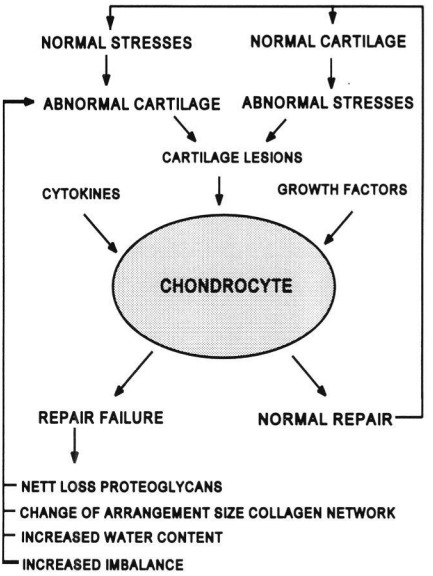


Figure 1. Osteoarthritic process

Genetic influence and OA

In several studies it was shown that genetic factors may contribute importantly to OA but to a varying extent in men and women and depending on the joint site(s) [2, 3]. It is not yet clear for which subgroups of OA patients the genetic influences are most relevant. One particular phenotype is a generalized pattern of OA (GOA) occurring at an early age of onset. This may be caused by relatively severe gene defects that make each joint susceptible to OA development even under normal mechanical stress [2]. Milder gene defects, on the other hand, must contribute to the onset of OA in hand and knee in women [3], the development of Heberden's nodes [4] or that of OA due to mild hip dysplasia [5]. Such mild mutations may contribute to the lesion proneness of cartilage. This phenotypic outcome may depend on environmental influences and may thus act in a gender, age, and lifestyle dependant manner. Gender specificity of the genetic effect may be caused by hormonal influences, differences in occupational stress or body mass index. Genetic defects affecting the length and shape of bone may contribute to OA in specific joint sites. In this thesis the study of genetic factors in age cohorts with a high prevalence of ROA revealed common alleles associated with common phenotypes as well as rare alleles associated with subgroups of ROA or even with ROA negative subgroups.

Family studies of OA

Mutations identified thus far in heritable skeletal disorders form a selection of defects with a clearly diagnosable phenotype and an obvious pattern of inheritance. The genes identified in these disorders are often structural components of the ECM [6]. Interestingly, some of the mutations identified show a relatively mild phenotype with a particularly strong effect in specific joints (EDM2 locus; [7, 8]), in contrast to disease mutations which result in a systemic phenotype such as GOA [9].

Because of the focus on severely affected phenotypes, hardy any mutations were identified with relatively mild effects, such as 'pure' familial OA (FOA) without dysplasia. To identify genes involved in the primary OA process we have investigated 14 candidate genes in such an FOA family, using genetic linkage analysis (Chapter 3). The family investigated included multiple members affected by OA at an age of onset between 20 and 40 years. Symptoms began with intermittent acute pain and swelling in one or both knee joints with subsequent OA development in other joints. The hip was only rarely affected. Ten genes were excluded from involvement in FOA in this Dutch family. Among these loci were important candidate

genes involved in several heritable skeletal disorders, mild chondrodysplasia and epiphyseal dysplasia associated with early onset OA in multiple joints (COL2A1, COL9A1, COL9A2, COL11A1, COL11A2, COMP, and the CPDD region). Other possible candidate genes encoding non-collagenous structural components, or genes involved in posttranslational modification and remodelling of cartilage were also excluded (CRTL-1, CRTM and MMP3). The COL9A3, DCN, LOX and PLOD genes could not be excluded as the cause of FOA in our family. The study demonstrates that these candidate genes were not involved in the pathogenesis of FOA in our study subjects. The study further indicated genetic heterogeneity for the FOA phenotype and the likely existence of yet unidentified OA genes which may be detected in future studies of families for which FOA is the primary disease process. Although theoretically many more candidate genes could be tested, a full genome search for the unidentified FOA gene may be more effective at this point. One would need, however, many more family members to perform such a search. In the age-related penetrance model used in this study, individuals are definitively diagnosed around the age of 40 years. Individuals of the fifth generation (aged 1-25 years) could, therefore, not be incorporated into the study. Our current study design was thus restricted to a candidate gene approach. In order to perform a genome-wide search to localize an as yet unknown disease gene in this family a horizontal expansion of affected individuals is necessary to increase the power.

In contrast to our association studies of ROA described in chapter 5 and 6 the important COL2A1 candidate gene was not involved in the onset of the Dutch FOA family. This indicates that although the phenotype of our FOA family resembles OA at later ages, the mechanism by which OA arises may differ. Therefore, linkage studies in FOA families do provide candidate genes involved in the OA process, however, the pathogenic mechanisms may differ for different OA phenotypes.

Association of the COL2A1 gene with ROA

Genetic association studies were performed to investigate the role of several candidate genes in more commonly occurring forms of OA at later ages. These association studies were performed in the ERGO population, a prospective cohort study of determinants and prognosis of chronic diseases in the elderly. It consists of men and women of ages over 55 years of age whose radiographic characteristics of OA were determined in hip, knee, hand and spine. We chose to investigate subjects at ages between 55-65 years of age. In this specific age group the genetic factors influencing the onset of OA may still be detected among other environmental factors.

Radiographic signs of OA, independent of clinical symptoms, provides a quantifiable reflection of cartilage degradation and dysfunction. A case-control design was chosen aimed at maximizing the contrast between cases and controls. Cases had ROA (Kellgren score ≥ 2) in at least one large joint (knee and/or hip) whereas controls were free of ROA in knee, hip, hand, and spine. The absence of ROA in each of the joint groups investigated was relatively rare (13% of the 1040 individuals investigated) as well as the presence of ROA in multiple joint sites (13 % had ROA in ≥ 3 joints) or with Kellgren score 3 (7% for knee ROA, 10% for hip ROA, 15% for hand ROA, and 24% for spine ROA).

As a first marker the multiallelic VNTR located at the 3' end of the COL2A1 was investigated. Recognition of multiple existing COL2A1 alleles in the population increases the chance of finding disease association. A significant predisposing association was observed between homozygous subjects for the common 13R1 allele and hip ROA in combination with hand and spine (OR = 3.0, 95% CI 1.0-9.0). Among cases with ROA in two or more joint groups including at least the knee or hip joint, carriers of the rare 14R2 allele had a 2.5 times higher risk of developing Heberden nodes. Furthermore, the 14R2 allele was increased in subjects with generalized ROA in knee, hand and spine in combination with Heberden's nodes (OR = 3.4, 95% CI 0.6-20.8). The effect of these alleles may thus contribute especially to the onset of ROA at several joints (generalized ROA) and the presence or absence of Heberden's nodes may influence these associations. The observation that both a common (13R1) and a rare (14R2) allele of the COL2A1 locus are associated with an increased risk, suggests that the etiology of a more generalized pattern of ROA may involve allelic heterogeneity. The increased association observed especially in subjects with ROA in several joints suggests a gene effect causing error-prone cartilage which may eventually fail under normal stress in an increasing number of joint sites. The sites that become affected differ between 13R1 and 14R2 alleles.

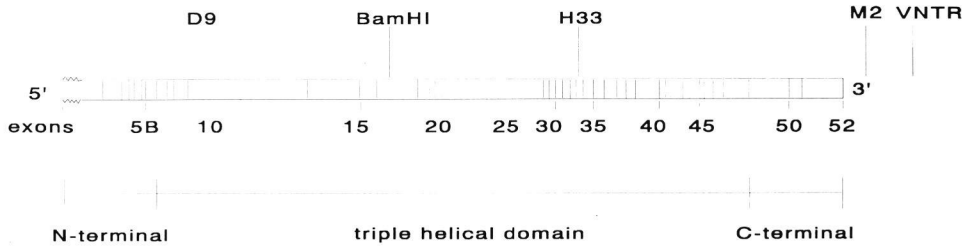
The association of the VNTR polymorphic marker locus with the presence of ROA may be explained either by an effect of the VNTR marker itself or by a predisposing mutation in the COL2A1 gene in linkage disequilibrium with the marker locus. Alternatively, the association may be caused by any effect of a gene in the vicinity of COL2A1 in linkage disequilibrium with the VNTR marker. To investigate whether the VNTR alleles themselves predispose to ROA or if these alleles are in linkage disequilibrium with the functional COL2A1 gene variant, haplotype analysis of additional polymorphisms in the gene may be performed. Before applying haplotype analysis for ROA association first the number of COL2A1 haplotypes residing in the

Dutch population was estimated by analysis of 7 intragenic (diallelic) RFLP polymorphisms in a population of random Dutch individuals (Chapter 4). Pairwise linkage disequilibrium analysis of these RFLPs showed a highly significant disequilibrium in 18 out of 21 RFLP pairs tested. When linkage disequilibria were compared to physical distance, a possible recombination hotspot was observed in intron 33. In view of the small distance between the RFLPs in intron 33 (165 bp), a high frequency of recombination is rather unlikely. A mutation hotspot in intron 33 was, therefore, assumed which complements a previous observation by others that a cluster of high frequency polymorphisms occur in the area of the COL2A1 gene containing intron 33 [10].

By using the pairwise linkage disequilibria data, at least 13 different COL2A1 haplotypes could be recognized in our population. An evolutionary tree was derived using the constructed haplotypes, based on the assumption that ancestral haplotypes may be more common than derived haplotypes [11]. According to these data both mutations (intron 33) and recombinations (especially in the region between intron 9 and intron 33) are necessary to explain all haplotypes observed in the Dutch population. Furthermore, at least two candidate ancestral haplotypes were observed which were both common and almost completely opposite. Our data further indicated that with a minimum of four RFLPs 98% of possible haplotypes were detected. Among them were at least 6 haplotypes with frequencies estimated from 0.04 to 0.37 which could be informative in the ROA association study of the COL2A1 gene.

Haplotype analysis of the COL2A1 locus with the occurrence of ROA was performed in an extended population using three COL2A1 RFLP polymorphisms (Dra I within intron 9 (D9), Hind III within intron 33 (H33), and Mae II in the 3'UTR

Figure 2. The collagen type II gene polymorphisms.



(M2)) in addition to the VNTR polymorphism (Figure 2). It was expected that especially using intragenic di-allelic polymorphisms in combination with the multi-allelic VNTR would be sufficiently powerful and informative for the haplotype analysis. Since the COL2A1 VNTR alleles 13R1 and 14R2 reported in Chapter 5 contributed especially to the generalization of the OA process, for haplotype analysis the population was divided into subjects with ROA in < 3 joint sites and subjects with ROA in ≥ 3 joint sites (hip, knee, hand, and spine). The individual RFLP polymorphism M2 did not show association with ROA, however, a significant association was observed for the dimorphism H33 (OR = 1.4, 95% CI 1.1-1.9) and D9 (OR = 1.4, 95% CI 1.0-1.9) with ROA in ≥ 3 joints. Linkage disequilibrium analysis showed that RFLP M2 and H33 were in complete negative linkage disequilibrium. RFLP M2 was, therefore, not further used in the haplotype analysis.

The association of the homozygous genotype 13R1/13R1 in this extended population with the presence of ROA in hip, hand and spine was confirmed. The 13R1 allele, however, was not associated with the occurrence of ROA in 3 or more joint sites, nor did we observe any association with the presence of ROA with any of the haplotypes containing allele 13R1. The potential functional variant of 13R1 may thus be predisposing to a specific subgroup of ROA cases, which may be distributed over the extra alleles detected with the haplotypes. Alternatively, the previously observed association with the 13R1 allele may be spurious. The association of the 14R2 allele was confirmed with the occurrence of ROA in 3 or more joints sites with a crude OR of 2.1, 95% CI 1.3-3.4. When subjects with ROA in 3 or more joints were stratified by the presence of Heberden's nodes an increased OR was observed for 14R2 (OR = 2.7, 95% CI 1.3-5.5), H33 (OR = 1.8, 95% CI 1.2-2.9) and D9 (OR = 1.8, 95% CI 1.1-3.0). When haplotypes of the 14R2 allele with D9 and H33 were associated with ROA in 3 or more joints, a highly significant association of haplotype 1-2-14R2 was observed (OR = 5.5, 95% CI 2.3-13.1) which extended the association with the VNTR marker locus.

These results suggest that the reported 14R2 allele (Chapter 5) and especially the 1-2-14R2 haplotype (Chapter 7) is associated with "generalized" ROA (ROA in 3 or more joint sites). Furthermore, the marked increase in strength of association when allele 14R2 is compared to haplotype 1-2-14R2, indicates strongly that this association is based on a gene variant marked by haplotype 1-2-14R2. This haplotype is expected to be in high linkage disequilibrium with the potential OA mutation, meaning that on each 1-2-14R2 haplotype the mutation occurs but that the mutation does not necessarily occur only on the 1-2-14R2 haplotype.

Together with previous reports, positive associations with generalized ROA have

now been described for the COL2A1 polymorphisms, Bam HI [12], M2 [13], and haplotype D9-H33-VNTR 1-2-14R2 of the VNTR (Figure 2; Chapter 5 and 6). In contrast, a number of studies failed to detect predisposing COL2A1 alleles [14, 15, 16] (see also General Introduction; Table 2). Several factors may cause these inconsistent results. Firstly, definitions (age, number of joint sites, gender) of affected cases and controls varies between the studies, while genetic susceptibility may be gender- and joint site-specific. Control groups were not always radiographically examined [12, 15, 16] which may result in a substantial number of false negative subjects in the control group and also lowers the contrast between cases and controls. Finally, an association based on a linkage disequilibrium may differ between populations. It may, therefore, be important to have a study design which enables the selection of cases and controls on one OA criteria and a second population sample of similar origin to confirm association.

The observation that several different alleles show association with ROA either means that allelic heterogeneity may be involved in the etiology of ROA, or that these associations represent the effect of one particular allele of the COL2A1 gene. It was shown previously that polymorphism H33 was in complete positive linkage disequilibrium with Bam HI [12, 15] and in negative linkage disequilibrium with M2 (Chapter 6). The observation that H33 and Bam HI are in complete linkage disequilibrium may indicate that the predisposing association with the occurrence of generalized ROA in both British and Dutch individuals may be subject to one particular functional COL2A1 variant. The observation that RFLP H33 alone in our population did not clearly demonstrate such an association may be due to the higher allele frequency of H33 as compared to the Bam HI site which may have decreased power. This hypothesis needs confirmation in both British and Dutch population samples.

The associations with haplotype 1-2-14R2 (D9-H33-VNTR) in the Dutch study and that with M2 in the British study may be of different origin. This is suggested by the complete negative linkage disequilibrium observed between H33-M2, which means absence of the haplotype H33-M2 2-2. Secondly, the M2 polymorphism did not show association in our population. In summary, the COL2A1 gene is highly likely to be involved in ROA development in multiple joints (present thesis) and in GOA with clinical symptoms in middle- aged women (British study). The haplotype analysis of the COL2A1 gene shows that the use of several markers within the candidate gene by performing an association study followed by linkage disequilibrium analysis has several advantages, 1) it increases the knowledge and detection of the number of existing alleles in the population, 2) it provides the possibility of estimating the

number of founder alleles that occur in the population, 3) it gives an indication of whether a disease association is based on a linkage disequilibrium with a functional variant in the candidate genes itself, and 4) it provides the tool for the delineation of the etiological mutation.

CRTM and CRTL1 gene associations with ROA

An ROA association study was also performed with the cartilage matrix (CRTM) and cartilage link (CRTL-1) genes (Chapter 5). A sex specific negative association by the putative effect of a protective allele was observed with a dinucleotide repeat polymorphism within the 3'UTR of the CRTM gene and with a dinucleotide within the promoter region of the CRTL-1 gene. A significantly decreased allele frequency of CRTM allele A2 was observed especially with ROA of the hip in men (OR = 0.5, 95% CI 0.3-1.0). The opposite was found for allele A1. The observation that alleles A2 and A1 inversely contribute to the association may indicate that CRTM A2 is associated with a decreased risk of men developing hip ROA, and A1 is associated with increased risk. Alternatively, only one of the associations is true and the other is the consequence of compensating allele frequencies. The independent effect of these alleles could not be tested by the exclusion of individuals with either CRTM A1 or A2 since the remaining number of individuals (carrying CRTM allele A3) was not sufficient.

For the CRTL1 alleles significant evidence of association was not observed. The frequency of allele A10, however, tended to be decreased in women with knee ROA and particular in those women with additional ROA in the hand as compared to female controls. Although the small number of individuals requires cautious interpretation, a strong and significant negative association was observed for women carrying the A10/A10 genotype and knee ROA or knee and hand ROA (OR = 0.07, CI 0.01-0.62 and 0.06, 95% CI 0.00-0.86, respectively).

These associations follow the sex and joint site specific pattern of ROA development in the population. Typically at ages between 55-65 years, hip ROA is most frequently present in men and ROA in knee and hand in women [17]. In this age group the prevalence of hip ROA in men has reached its maximum, whereas the incident rate of knee and hand ROA in women is high. The observed positive and/or negative associations for these specific subgroups may be due to the larger numbers of individuals in these case groups or it may reflect gender- and/or joint site-specific effects of the alleles on ROA.

An effect of genetic variation at the CRTM locus on especially the development of

hip ROA may be explained by the fact that the CRTM protein is expressed within the epiphyseal growth plate (postmitotic stages) during the process of endochondral bone formation [18]. During this phase the length and shape of the bone are determined. Although the function of CRTM in epiphyseal cartilage remains unclear, it may affect the overall shape of the joint. Since ROA of the hip especially is often considered to arise due to anatomical abnormalities [5], the possible effect of CRTM alleles on hip ROA might be exhibited in this way.

Genetic variation at the CRTL1 gene could contribute to or protect from ROA by affecting stabilization of the proteoglycan aggregates in the extracellular matrix in cartilage. A genetic variation of the CRTL1 protein may also change its sensitivity to degrading enzymes (stromelysin) during the arthritic process. The CRTM and CRTL-1 genes may thus contribute to a gender- and joint-site specific pattern of ROA, however, the nature these effects remains to be unravelled.

Future perspectives

The studies in this thesis provide a basis for further research into the genetics of OA. The study of (new) Dutch FOA families with pure OA without dysplasia may result in the identification of other OA genes. To identify these genes, it is of particular importance that "pure" FOA families are collected from orthopaedic and rheumatological clinics. Such FOA families may not be as easily recognized as the severe heritable skeletal disorders with early ages of onset. Analysis of either large FOA families or separate sibling pairs are necessary for genomic wide scans to identify the genomic location of new OA genes. Such a study may identify genes even more specifically involved in the OA process than the currently investigated ECM components identified in families with skeletal disorders. Other candidate genes for ROA to be investigated in the future may be genes involved in cartilage maintenance (e.g. IGF-1 or TGF-ß) and/or degradation (e.g. MMP3).

The current associations of COL2A1, CRTM, and CRTL1 genes with the occurrence of ROA at later ages need to be confirmed in a second population-based sample. For the CRTM and CRTL-1 genes additional markers may be used to delineate specific haplotypes involved in the association. The ROA association of the 1-2-14R2 haplotype can be investigated with more efficiently in both linkage analysis or allele-sharing methods using either affected (nuclear) families or (distant) relatives or may be used directly for mutation analysis in cases carrying the predisposing haplotype. Once a mutation is identified, the frequency of the mutation carriers can be determined in the population, the association with ROA examined

and the environmental influences affecting the ROA susceptibility of the carriers investigated.

Another interesting question is whether the COL2A1 association with generalized ROA is also detected in older age groups where the prevalence of this generalized pattern of ROA is much higher. Also in older age groups gene relations can be studied of more severe ROA cases with Kellgren scores higher than 2. In our study design, an increased severity resulted automatically in the selection of subjects with a higher number of joints affected since the number of individuals with a Kellgren score 3 or higher was very rare.

This thesis focused on associations of genes with radiographic characteristics of OA. For further analysis it may also be relevant to analyse clinical features of OA. In this respect, the association may be replicated in a GOA-patient population to determine whether the COL2A1 haplotype and CRTM and CRTL-1 alleles predispose to clinical endpoints. In addition, the relevance of the candidate genes to the progression of ROA may be investigated in the follow-up study of the ERGO population (ERGO III) combined with information on occupation, history of trauma and other lifestyle factors. This may provide insight into the questions of which of the (genetically predisposed) individuals actually develop symptomatic OA and at what rate.

References

1. Howell DS. Etiopathogenesis of osteoarthritis, in Mc Carty DM (ed): Arthritis and Allied conditions. Philadelphia PA, Leae and Febiger (1989) pp1595-1604.
2. Kellgen JH, Lawrence S, and Bier F (1963) Genetic factors in generalized osteo-arthrosis. Ann Rheum Dis 22:237-253.
3. Spector TD, Cicuttini F, Baker J, Hart DJ (1995) Genetic influences on osteoarthritis in women: a twin study. BMJ 312: 940-943.
4. Stretcher RM (1955) Heberden's nodes a clinical description of osteo-arthritis of the finger joints. Ann Rheum Dis 14:1-10.
5. Harris W (1986) Etiology of osteoarthritis of the hip. Clin Orthop 213:20-33.
6. Spranger J, Winterpacht A, Zabel B (1994) The type II collagenopathies: a spectrum of chondrodysplasias. Eur J Pediatr 153:56-65.
7. Briggs MD, Hoffman SMG, King LM, Olsen AS, Mohrenweiser H, Leroy JG, Mortier GR, Rimoin DL, Lachman RS, Gaines ES, Cekleniak JA, Knowlton RG, Cohn DH (1995) Pseudoachondroplasia and multiple epiphyseal dysplasia due to mutations in the cartilage oligomeric matrix protein gene. Nature Genet 10:330-336.
8. Muragaki Y, Mariman ECM, van Beersum SEC, Perälä M, van Mourik JB, Warman ML, Olson BR, Hamel BCJ (1996) A mutation in the gene encoding the α2 chain of the fibril-associated collagen IX, COL9A2, causes multiple epiphyseal dysplasia (EDM2). Nature Genet 1:103-105.

9. Knowlton RG, Katzenstein PL, Moskowitz RW, Weaver EJ, Malemud CJ, Pathria MN, Jimenez SA, Prockop DJ (1990) Genetic linkage of a polymorphism in the type II procollagen gene (COL2A1) to primary osteoarthritis associated with mild chondrodysplasia. New Eng J Med 322:526-530.

10. Vikkula M. and Peltonen L (1989) Structural analyses of the polymorphic area in the II collagen gene. FEBS 250:171-174.

11. Watlarson GA, and Guess HA (1977) Is the most frequent allele the oldest? Theor Popul Biol 11:141-160.

12. Hull R, and Pope FM (1989) Osteoarthritis and cartilage collagen genes. Lancet:1337-1338.

13. Loughlin J, Irven C, Athanasou N, Carr A, and Sykes B (1995) Differential allelic expression of the type II collagen gene (COL2A1) in osteoarthritic cartilage. Am J Hum Genet 56:1186-1193.

14. Vikkula M, Nissilä M, Hirvensalo E, Nuotio P, Palotie A, Aho K, Peltonen L (1993) Multiallelic polymorphism of the cartilage collagen gene: no association with osteoarthritis. Ann Rheum Dis 52:762-764.

15. Priestley L, Fergusson C, Ogilvie D, Wordsworth P, Smith R, Pattrick M, Doherty M, (1991) A limited association of generalized osteoarthritis with alleles at the type II collagen locus: COL2A1. Br J Rheumatol 30:272-275.

16. Loughlin J, Irven C, Fergusson C, Sykes B (1994) Sibling pair analysis shows no linkage of generalized osteoarthritis to the loci encoding type II collagen, cartilage link protein or cartilage matrix protein. Br J Rheumatol 33:1103-1106.

17. Saase van JLCM, Van Romunde LKJ, 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 48:271-280.

18. Chen Q, Johnson DM, Haudenschild DR, Goetinck PF (1995) Progression and recapitulation of the chondrocyte differentiation program: cartilage matrix protein is a marker for cartilage maturation. Dev Biol 172:293-309.

SUMMARY

SUMMARY

The influence of genetic factors in the onset of OA has been demonstrated in families with dominantly inherited early onset (20-40 years) OA, and at later ages (55-65 years) due to a clustering of OA in twin and sibling pairs. The nature of such a genetic influence in OA may involve either a structural cartilage defect or alterations in cartilage and/or bone metabolism. Alternatively, a genetic defect may influence a known risk factor for OA such as obesity. This thesis is a search for genetic factors predisposing to the onset of OA at early ages in FOA families and at later ages in unrelated individuals in the population.

To study genetic factors in both families and populations, human genomic DNA must be collected. Usually, genomic DNA is isolated from peripheral blood samples. This is an expensive and invasive procedure to which, for ethical reasons, objections may be raised, especially in studies involving older individuals. A noninvasive DNA sampling and isolation method was, therefore, developed involving oral samples taken by using cotton swabs. Participants can sample mouth swabs themselves, and can send samples by mail to the research centre. This fast and cost effective method is now increasingly being used for the collection of human genomic DNA (Chapter 2).

To identify genes that are able to cause early onset generalized OA a genetic linkage study was performed of 14 candidate genes. Generalised OA is defined as OA in 3 or more joint groups. In a Dutch family with dominantly inherited primary OA (FOA) without mild epiphyseal dysplasia, ten important candidate gene loci for OA were excluded to be involved in the onset of FOA among which genes encoding collagens, non-collagenous glycoproteins and enzymes involved in cartilage post-translational modification, or remodelling. As yet unknown disease genes may be identified by performing a genome wide search in this and other FOA families (Chapter 3).

Thus far, the most widely studied candidate genes for OA encode collagenous components of cartilage. The collagen network is crucial in resisting sheared and tensile forces at the joints. Mutations in these genes are known to cause heritable skeletal disorders associated with OA. Mutations in collagen genes with a relatively mild effect may predispose to the onset of OA at later ages. These mild mutations may be detected by comparing the DNA of unrelated individuals with and without OA. Such a genetic association study was performed by comparing unrelated individuals with knee and/or hip ROA (cases) with individuals free of ROA in the knees, hips, hands, wrists and the thoraco-lumbar spine (controls). Cases and

controls were derived from a prospective population based cohort study of determinants and prognosis of chronic disease in the elderly, The Rotterdam study.

By using polymorphic DNA markers, allele frequencies of the collagen type II (COL2A1), cartilage matrix (CRTM), and the cartilage link (CRTL1) genes were studied. For the COL2A1 locus association was performed with a variable number of tandem repeat (VNTR) marker of the COL2A1 gene. A significant predisposing association was observed between homozygous subjects for a common VNTR allele and hip ROA in combination with hand and spine. Among cases with ROA in two or more joint groups including at least the knee or hip joint, carriers of a rare allele had a 2.5 times higher risk of developing Heberden nodes. Furthermore, an increased frequency of allele 14R2 was observed in subjects with generalized ROA in knee, hand and spine in combination with Heberden's nodes. The effect of these alleles may thus contribute especially to the onset of ROA at several joints (generalized ROA) (Chapter 4).

To investigate whether these VNTR alleles a) directly predispose to ROA or b) if these alleles are in linkage disequilibrium with a functional COL2A1 gene variant, haplotype analysis of additional polymorphisms in the gene can be performed. Before applying haplotype analysis for ROA association studies, first the number of COL2A1 haplotypes residing in the Dutch population was estimated. By genotyping 7 intragenic RFLP polymorphisms of the COL2A1 gene in a population of randomly selected Dutch individuals the existence of at least 13 different haplotypes was observed. By analysis of four of these RFLP polymorphisms 98% of all possible COL2A1 haplotypes occurring in the Dutch population could be distinguished (Chapter 5). Two of the RFLP's were most suitable to used in addition to the VNTR.

Haplotype association analysis with the predisposing alleles of the COL2A1 VNTR marker was performed in an extended population, using two RFLP polymorphisms in addition to the VNTR. In this way, a specific predisposing haplotype of the rare VNTR allele was observed with an OR of 5 for developing ROA in three or more joints before the age of 65 years. Individuals carrying this specific haplotype may reveal a functional mutation predisposing potentially to the generalized ROA phenotype (Chapter 6).

For both the CRTM and the CRTL1 locus a sex specific negative association (protective allele) was observed. An allele of the polymorphisms within the 3' untranscribed region of the CRTM locus was significantly associated with hip ROA in men. In women a homozygous CRTL1 genotype of the dinucleotide repeat polymorphism in the promoter region was significantly associated with ROA in only the knee as well as in combination with hand (Chapter 7).

Several genes contribute to the onset of ROA and each individual gene may contribute to a specific pattern of affected joints. The COL2A1 gene is involved in the onset of generalized OA both at early and at later ages. The CRTM gene appears to be involved in the protection against hip ROA in men. Further research may identify the functional gene variants which may provide further insight into the relevance of these genes in the ROA process. Unknown genes involved in the onset of FOA may be detected by performing genome wide scans of specific FOA families.

SAMENVATTING

SAMENVATTING

De invloed van genetische factoren op het ontstaan van osteoartrose (OA) is aangetoond in families waarin OA op jonge leeftijd (20-40 jaar) voorkomt en in tweeling- en broer/zuster paren waarbij OA voorkomt op middelbare leeftijd (55-65 jaar). Een erfelijke invloed op het ontstaan van OA kan worden veroorzaakt door een genetische afwijking in de kraakbeenstructuur of door genetische veranderingen in het kraakbeen- of botmetabolisme. Het is ook mogelijk dat de genetische aanleg risicofactoren voor het ontstaan van OA beïnvloed, zoals overgewicht. In dit proefschrift werd de invloed van een aantal genen onderzocht die mogelijk een rol spelen bij het ontstaan van familiaire gegeneraliseerde OA op jonge leeftijd en OA op latere leeftijd in ongerelateerde personen uit de populatie.

Om de genetische factoren in families en populaties te onderzoeken moet genomisch DNA worden verzameld. Genomisch DNA wordt vaak geïsoleerd uit perifeer bloed. Afname van bloed is een relatief dure en invasieve procedure waartegen ethische bewaren kunnen ontstaan, vooral in studies waarbij oudere personen zijn betrokken. Daarom werd een non-invasieve DNA isolatiemethode ontwikkeld gebaseerd op afname van wanguitstrijkje met behulp van wattenstaafjes. Deelnemers kunnen bij zichzelf een wanguitstrijkje afnemen en deze opsturen naar het onderzoekscentrum. Deze snelle en goedkope methode wordt nu algemeen toegepast voor de verzameling van genomisch DNA (Hoofdstuk 2).

Om genen te identificeren die betrokken zijn bij het ontstaan van vroege gegeneraliseerde OA werd koppelingsonderzoek uitgevoerd in een familie waarin OA op jonge leeftijd voorkomt en overerft met een dominant Mendeliaans patroon (FOA). In deze Nederlandse familie met primaire FOA zonder aanwezigheid van milde epifysiële dysplasie werd aangetoond dat 10 bekende kandidaatgenen voor OA niet betrokken zijn bij het ontstaan van artrose in deze familie (Hoofdstuk 3). Door een volledige screening van het genomisch DNA van deze en andere FOA families kunnen onbekende genen worden opgespoord die het vroege FOA ziektebeeld kunnen verklaren.

De meest bestudeerde kandidaatgenen voor OA coderen voor de collageen-componenten van kraakbeen omdat deze cruciaal zijn voor het opvangen van de mechanische belasting van het gewricht. Ernstige mutaties in deze genen geven aanleiding tot skeletaandoeningen. Relatief milde mutaties in de collageengenen kunnen ten grondslag liggen aan aanleg (predispositie) voor het ontstaan van OA op middelbare of latere leeftijd. Deze milde erfelijke afwijkingen kunnen worden opgespoord door DNA van ongerelateerde personen met en zonder OA met elkaar

te vergelijken. Een dergelijke associatie studie werd verricht met een groep ongerelateerde personen met radiologische OA (ROA) in de knie en/of heup (cases) en met een groep ongerelateerde personen die geen ROA hadden in de knie, heup, hand, wervelkolom (controles). De cases en controles werden verzameld uit een populatieonderzoek naar ziekten bij personen ouder dan 55 jaar het Erasmus Rotterdam Gezondheid en Ouderen (ERGO) onderzoek.

Met behulp van polymorfe DNA merkers werd bekeken of cases en controles verschillende allelen bezaten van genen coderend voor het collageen type II eiwit (COL2A1), kraakbeen matrix-eiwit (CRTM), en het kraakbeen link-eiwit (CRTL-1). Een significante associatie werd gevonden met de variabele in tandem herhaalde (VNTR) DNA merker van het COL2A1 gen. Voor personen met een homozygoot genotype van het veel voorkomende allele 13R1 (frequentie 0.44) werd een significante predisponerende associatie gevonden met ROA in heup, hand en wervelkolom. Voor personen met ROA in twee of meer gewrichten waaronder het knee en/of heup gewricht werd voor de dragers van het zeldzame allele 14R2 werd een significante associatie gevonden met het krijgen van Heberden's nodes. Het effect van deze collageen-allelen lijkt daarom vooral bij te dragen aan het ontstaan van ROA in meerdere gewrichten (gegeneraliseerde ROA). De aan- of afwezigheid van Heberden's nodes lijkt deze associaties te beinvloeden (Hoofdstuk 4).

Om te onderzoeken of de VNTR allelen a) zelf predisponeren voor ROA in de populatie of b) fysiek gekoppeld (in linkage disequilibrium) zijn met een functionele COL2A1 genvariant, kan een haplotype analyse uitgevoerd worden. Voordat deze haplotype associatie-analyse werd toegepast, is onderzocht welke COL2A1 haplotypes aanwezig zijn in de Nederlandse populatie. Door genotypes van 7 intragene COL2A1 restrictie fragment lengte polymorfismen (RFLP) te bepalen in een random steekproef van Nederlandse personen werden minstens 13 verschillende COL2A1 haplotypes aangetoond. Met de analyse van 4 van deze RFLPs kan 95% van alle COL2A1 haplotypes aanwezig in de Nederlandse populatie worden gedetecteerd (Hoofdstuk 5).

Haplotype associatie analyse met de predisponerende COL2A1 VNTR allelen werd vervolgens uitgevoerd in een uitgebreide populatie, waarbij twee additionele RFLP polymorfismen werden gebruikt naast het VNTR polymorfisme. Een specifiek COL2A1 haplotype van het zeldzame VNTR allel bleek geassocieerd, met een OR van 5, met het ontstaan van ROA in 3 of meer gewrichten voor de leeftijd van 65 jaar. Bij personen met dit haplotype kon vervolgens worden gezocht naar de functionele sequentievariatie in het COL2A1 gen dat predisponeert voor gegeneraliseerde ROA (Hoofdstuk 6).

Voor het CRTM en het CRTL-1 gen werd een geslachtsspecifieke, negatieve (beschermende) associatie gevonden. Een allel van het polymorfisme in het "3' untranscribed region" van het CRTM gen was speciaal bij mannen significant geassocieerd met aanwezigheid van ROA in de heup. In vrouwen was een homozygoot CRTL-1 genotype van een dinucleotide "repeat" polymorfisme in het promotorgebied significant geassocieerd met ROA in een combinatie van hand- en kniegewrichten (Hoofdstuk 7).

Meerdere genen beïnvloeden het ontstaan van ROA. Verschillende genen dragen bij tot een verschillend patroon van aangedane gewrichten. Het COL2A1 gen lijkt zowel bij vroege OA als bij OA op middelbare leeftijd steeds geassocieerd met gegeneraliseerde OA. CRTM lijkt geassocieerd met een bescherming tegen ROA in de heup bij mannen. Meer onderzoek is nodig voor het opsporen van de functionele sequentievariaties waarmee het belang van deze genen voor OA verder kan worden bepaald. Daarnaast mag worden verwacht dat onbekende OA specifieke genen zullen worden opgespoord bij verder onderzoek naar FOA families.

LIST OF PUBLICATIONS

- Meulenbelt I, Wapenaar M, Patterson D, Vijg J, and Uitterlinden AG (1993) Isolation and mapping of human chromosome 21 cosmids using a probe for RTVL-H retrovirus-like elements. Genomics 15:492-9.

- Meulenbelt I, Droog S, Trommelen GJM, Boomsma DI and Slagboom PE (1995) High yield noninvasive human genomic DNA isolation method for genetic studies in geographically dispersed families and populations. Am J Hum Genet 57:1252-4.

- Meulenbelt I, Williams CJ, Te Koppele JM, van de Giessen GC, and Slagboom PE (1996) Population haplotype analysis and evolutionary relations of the COL2A1 gene. Ann Hum Genet 60:189-99.

- Droog S, Lakenberg N, Meulenbelt I, de Maat MPM, Huisman LGM, Jie AFH, Slagboom PE (1996) Isolation and storage of DNA for population studies. Fibrinolysis 10 (suppl2) 29-30.

- Slagboom PE, Bijkerk C, Meulenbelt I, Breedveld FC, Valkenburg HA, Odding E, Miedema H, Te Koppele JM, Pols HAP, Hofman A, and van Duijn CM (1996) Population based association studies of the COL2A1 gene locus with radiological osteoarthritis. Arthritis and Rheumatism 39 (suppl): S167.

- Meulenbelt I, Bijkerk C, Breedveld FC and Slagboom PE (1996) The role of genetic factors of osteoarthritis. Arthritis and Rheumatism 39 (suppl): S270.

- Meulenbelt I, Bijkerk C, de Wildt SCM, Odding E, Valkenburg HA, Miedema HS, Breedveld FC, Hofman A, Pols HAP, Sloos VGF, van Duijn CM and Slagboom PE. (1997) Disease association of the CRTM and CRTL1 genes with radiological osteoarhtritis in a population based study: The Rotterdam study. Arthritis and Rheumatism in press.

- Meulenbelt I, Bijkerk C, Breedveld FC, and Slagboom PE. (1997) Genetic linkage analysis with 14 candidate gene loci in a family with autosomal dominant osteoarthritis without dysplasia. J Med Genet in press.

- Meulenbelt I, Bijkerk C, Valkenburg HA, Miedema HS, Breedveld FC, Hofman A, Pols HAP, van Duijn CM and Slagboom PE. The IGF-1 gene with radiological osteoarthritis in a population based cohort study. Submitted for publication.

- Bijkerk C, Meulenbelt I, Odding E, Valkenburg HA, Miedema HS, Breedveld FC, Hofman A, Pols HAP, van Duijn CM and Slagboom PE. Association of a highly informative VNTR-polymorphism linked to the COL2A1 locus and generalized radiological osteoarthritis in a population based study: The Rotterdam Study. Submitted for publication.

- Meulenbelt I, Bijkerk C, de Wildt SCM, Odding E, Valkenburg HA, Miedema HS, Breedveld FC, Hofman A, Pols HAP, van Duijn CM and Slagboom PE. Disease association with COL2A1 haplotypes of radiological osteoarthritis in a population based study: The Rotterdam study. In preparation.

ABBREVIATIONS

AGN	aggrecan
BMD	bone mineral density
BMI	body mass index
CMC1	first carpometacarpa joint
COL2A1	gene encoding the α1-chain of the collagen type II
COL9A1	gene encoding the α1-chain of collagen type IX
COL9A2	gene encoding the α2-chain of the collagen type IX
COL9A3	gene encoding the α3-chain of the collagen type IX
COL10A1	gene encoding the α1-chain of the collagen type X
COL11A1	gene encoding the α1-chain of the collagen type XI
COL11A2	gene encoding the α2-chain of the collagen type XI
COMP	cartilage oligomeric protein gene
CPPD	calcium pyrophosfate dehydrate deposition disease
CRTL-1	cartilage link protein gene
CRTM	cartilage matrix protein gene
CS	cervical spine joint
DCN	decorin gene
DIP	distal interphalangeal joint
ECM	extracellular matrix
ERGO	determinants and prognosis of chronic disease in the elderly
FGF	fibroblast growth factor
FOA	familial osteoarthritis
GOA	generalized osteoarthritis
IBD	identity by descent
IGF-I	insuline like growth factor 1
LOX	lysyl oxidase gene
LS	lumbar spine joint
MCP	metacarpophalangeal joint
MED	multiple epihyseal dysplasia
MMP	metallo protease
MMP3	stromelysin 1 gene
NEC	non enzymatic cross links
OA	osteoarthritis
OSMED	ostospondylomegaepiphyseal dysplasia
PG	proteoglycans

PIP	proximal interphalangeal joint
PLOD	lysyl hydroxylase gene
PSACH	pseudoachondrodysplasia
RFLP	restriction fragment length polymorphism
ROA	radiological osteoarthritis
TGF-β	transforming growth factor β
TIMP	tissue inhibitor metallo protease
UTR	untranscribed region
VNTR	variable number of tandem repeats

CURRICULUM VITAE

Ingrid Meulenbelt is geboren op 18 oktober 1967 te Weesp. In 1986 behaalde zij haar VWO diploma aan het St. Ignatius College in Purmerend. In dat jaar ving zij aan met haar studie Biologie aan de Universiteit van Amsterdam (UVA). Na het behalen van haar propadeuse Biologie studeerde zij van 1987 tot 1990 Medische Biologie aan de UVA met als hoofdvak Moleculaire Biologie. In het kader van haar afstuderen doorliep zij een stage bij TNO-IVEG te Rijswijk. Van september 1990 tot juli 1992 was zij werkzaam als onderzoeksmedewerker bij Medscand INGENY.

In juli 1992 begon zij aan haar promotieonderzoek als assistent in opleiding aan de Rijksuniversiteit Leiden (RUL; professor dr. FC Breedveld en professor dr. DL Knook). Het onderzoek werd uitgevoerd op het Gaubius Laboratorium (TNO-Preventie en Gezondheid) onder begeleiding van dr. PE Slagboom. De resultaten van het promotieonderzoek staan beschreven in dit proefschrift.

Vanaf juli 1997 heeft zij een aanstelling als postdoc aan de RUL en blijft gestationeerd op het Gaubius Laboratorium (TNO-Preventie en Gezondheid).

NAWOORD

Lang heb ik uitgezien naar het moment waarop ik het nawoord van m'n proefschrift zou schrijven, nu is het zover. Een aantal van de mensen die een bijdrage hebben geleverd aan dit proefschrift wil ik met name bedanken.

Allereerst wil ik Eline Slagboom bedanken. Ik dank haar vooral voor de lange dagen met inspirerende en stimulerende discussies over de interpretatie van data. Ook wil ik haar bedanken voor het kritisch lezen van de, vaak nog ruwe versies, van manuscripten. Daarnaast was er altijd tijd voor een goed gesprek over andere zaken dan wetenschap. Zonder haar was het nooit tot een promotie gekomen.

Mijn twee promotoren, professor Ferry Breedveld en professor Dick Knook wil ik bedanken voor de bijdrage aan, vooral de laatste fase, van mijn promotieonderzoek.

Simone Droog en Saskia de Wildt wil ik bedanken voor alle hulp, maar vooral voor het luisteren naar mijn verhaal bij tegenslag en/of voorspoed in de afgelopen jaren. Het was/is ontzettend prettig en gezellig om met jullie samen te werken.

Casper Bijkerk wil ik bedanken voor het lezen van alle röntgenfoto's en voor het stellen van de klinische diagnose bij het familieonderzoek. Ook wil ik hem bedanken voor zijn kritische blik op resultaten en manuscripten. Harald Miedema en professor HA Valkenburg wil ik bedanken voor het lezen van röntgenfoto's van de ERGO populatie. Zij hebben hiermee mede bijgedragen aan de verzameling van data beschreven in dit proefschrift. Cock van Duijn wil ik bedanken voor al haar hulp tijdens het analyseren van resultaten en het lezen van mijn manuscripten. Ik vind het jammer dat ze niet bij m'n promotie aanwezig kan zijn. Ook wil ik Beate Böhm, Bert van der Wurff en Pascal Negre bedanken dat ze voor mij, tijdens het familieonderzoek, familieleden in Frankrijk hebben benaderd. Hierdoor heb ik, ondanks mijn tekort aan kennis van de Franse en Duitse taal, DNA van deze familieleden kunnen verzamelen.

De verschillende studenten Richard, Marianne, Saskia, Vincent, Nienke en Eka wil ik bedanken voor al het praktische werk waarmee ze mij de afgelopen jaren hebben geholpen. Daarnaast dank ik de andere "labgenoten" Bas, Marian, Nico en Karin, voor de nodige gezelligheid in het lab en in de koffiehoek.

Tenslotte wil ik mijn ouders bedanken, zij hebben me geleerd om door te zetten. Guido jij zorgde, met je nimmer aflatende steun, dat ik door kon gaan.

Ingrid

Leiden, september 1997