Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/33065 holds various files of this Leiden University dissertation

Author: Ramos, Y.F.M.
Title: Osteoarthritis, a degenerative disease of the articular joints : towards the implementation of functional genomics in OA
Issue Date: 2015-05-26

Osteoarthritis, a degenerative disease of the articular joints

Towards the implementation of functional genomics in OA

Yolande Francisca Maria Ramos

Financial support for the publication of this thesis was kindly provided by the Anna Foundation|NOREF and Artrose Zorg.

PhD thesis with summary in Dutch
ISBN: 978-94-6295-160-0

Osteoarthritis, a degenerative disease of the articular joints © 2015 YFM Ramos
All rights reserved. No part of this thesis may be reproduced, stored in a retrieval system or transmitted in any other form by any means, without permission of the author or when appropriate of the publisher of the published articles.

Cover	Immunohistochemistry of hip and knee joint sections
Cover design and layout	YFM Ramos
Printed and published by	Uitgeverij BOXPress, 's Hertogenbosch

Osteoarthritis, a degenerative disease of the articular joints

Towards the implementation of functional genomics in OA

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 26 mei 2015
klokke 16:15 uur

door

Yolande Francisca Maria Ramos

geboren te Eindhoven in 1970

Promotiecommissie

Promotor	Prof. dr. P. E. Slagboom
Co-promotor	Dr. I. Meulenbelt
Commissieleden	Prof. dr. J.V.M.G. Bovée
	Prof. dr. M. Karperien *MIRA Institute for Biomedical Technology & Technical Medicine; University of Twente*
	Dr. J.B.J. van Meurs *Department of Internal Medicine; Genetic Laboratory Erasmus Medical Center*

Contents

Chapter 1	Introduction	7
Chapter 2	Markers of the osteoarthritis process identified through genome wide expression analysis in articular cartilage; the RAAK study	25
Chapter 3	Genes expressed in blood link osteoarthritis with apoptotic pathways	49
Chapter 4	Meta-analysis identifies loci affecting levels of the potential OA biomarkers sCOMP and uCTX-II with genome wide significance	71
Chapter 5	Genome-wide association study: identifying new genes and pathways conferring risk to susceptibility of osteoarthritis at multiple joint locations as defined in the GARP study	103
Chapter 6	A gain of function mutation in TNFRSF11B causes osteoarthritis with chondrocalcinosis	121
Chapter 7	Underlying molecular mechanisms of DIO2 susceptibility in symptomatic osteoarthritis	147
Chapter 8	Discussion	177
Appendix	Summary	204
	Summary in Dutch	208
	Curriculum Vitae	213
	List of Publications	214
	Dankwoord	216

Introduction

OSTEOARTHRITIS, AN OVERVIEW

Osteoarthritis (OA) is a degenerative disease of the joints showing increased incidence with aging. While the incidence of the disease is steadily increasing due to an increasing population of elderly and rising numbers of people suffering from obesity, OA has become part of the top 5 most common diseases especially in Europe and the United States.[2-4] Of 291 conditions included in the Global Burden of Disease 2010 study (GBD 2010) OA ranked 11th highest in terms of so-called 'years of life lived with disability' (YLDs;[1]) and it is anticipated that OA will become a major problem for health systems globally with large socio-economic burden on society. Occurrence of OA is rare before the age of 40 years, with prevalence of knee OA peaking around 50 years of age while prevalence of hip OA continues to increase with age (Figure 1A). Similar to hip OA, prevalence of hand OA increases with age and percentages are joint-specific (Figure 1B).[5, 6]

Since OA is a chronic and painful disability and a major cause of morbidity, development of drugs to delay onset and/or progression is essential. This requires however more knowledge of the etiology, improved mechanistic insight into the disease, and sensitive biomarkers to diagnose OA at an early stage before irreversible damage becomes apparent in X-ray, manifested by the presence of osteophytes, joint space narrowing, and subchondral cyst formation of the bone. Up until now, treatments that could cure or stop disease progression are still poor and osteoarthritis is not diagnosed until late stages when damage is irreversible. Therapeutic approaches are aimed at pain relief, controlling symptoms, and minimizing disability, until in the end stage total joint arthroplasty is required.[7] Main impediments in OA management are the lack of biomarkers that detect OA at an early stage when damage is still reversible, or that allow stratification of subsets of OA patients with different disease modes. This considerably hampers development of effective therapies and requires better insight into the heterogeneous etiology of OA.

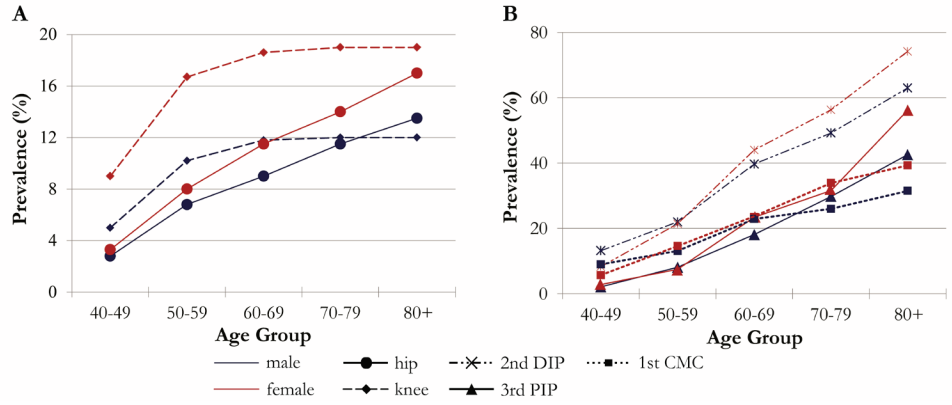


Figure 1 Prevalence osteoarthritis, stratified by joint-site and sex. A Hip and knee OA (data for North American population[1]). **B** Hand OA stratified for 3 joints (data from an American community-based, longitudinal study[5]). Legend: red line: females; blue line: males; diamonds: knee OA; filled circles with dashed line: hip OA; asterisks with dashed line: hand OA 2nd DIP; triangles: hand OA 3rd PIP; squares with dotted line: hand OA 1st CMC.

ONE NAME, MULTIPLE DISEASES

Even though the most common form of arthritis has a single name, osteoarthritis, in fact it has become manifest that it is a heterogenic disease. In general, two categories of OA are distinguished (reviewed by Johnson and Hunter[8]): primary or idiopathic OA which has no clear cause, and secondary OA which results from a known event (e.g. traumatic events[9]) or diseases such as acromegaly[10] or pattelofemoral joint osteoarthritis[11]). Although OA is most prevalent at the joints of hand, hip, and knee, it can occur at any joint.[2]

For cases of primary OA, depending on the occurrence at a specific joint-site or the affection of 3 or more joints, respectively localized and generalized OA can be distinguished. Furthermore, different subtypes can be discriminated, for example: hypertrophic (when only osteophytosis is present), atrophic (only joint space narrowing), and classic or normotrophic OA (when both, joint space narrowing and osteophytosis, are present).[12]

Diagnosis of OA relies on the combination of clinical, radiographical, and physical findings. Clinical symptoms of osteoarthritis are pain, joint stiffness and crepitus, resulting in an increasing loss of mobility. Radiographically, OA becomes manifest by osteophytes, joint space narrowing (JSN), and subchondral cysts of the bone. However, the patient's perception of pain is often not proportional to the clinical and radiographical stage of the disease thereby hampering consistent and analogous diagnoses. In addition, substantial discordance exists between clinical and radiographical OA.[13]

Osteoarthritis as whole joint disorder. Originally, degradation of articular cartilage was emphasized most in OA pathology. However, based on their studies already around 1970 Radin and colleagues proposed for the first time that cartilage degeneration may in fact result from changes in the (subchondral) bone.[14, 15] At present, it is generally accepted that OA is a whole joint disease involving not only bone and cartilage but also ligaments, synovial tissues, and the joint musculature,[16, 17] although it is most likely that OA originates in bone and/or cartilage subsequently affecting the other tissues.

Epidemiological studies have shown that the risks for osteoarthritis and osteoporosis (OP) are inversely correlated, with higher systemic bone mass being a risk factor for OA which was recently confirmed in a study by Hardcastle and colleagues.[18] Unfavorable mechanical forces imposed on the cartilage during movement due to suboptimal (higher) bone mineral density (BMD) could result in the accumulation of microtraumas throughout life, eventually leading to OA. Despite the fact that bone and cartilage seem of solid, impermeable composition it has been shown that molecules can diffuse into the extracellular matrix of both, (calcified) cartilage and subchondral bone.[19] Catabolic enzymes or other molecules released in the bone following subchondral bone remodeling may diffuse to the cartilage and activate chondrocytes or vice versa.

GENETIC EPIDEMIOLOGY OF OSTEOARTHRITIS

The multifactorial origin of OA is reflected

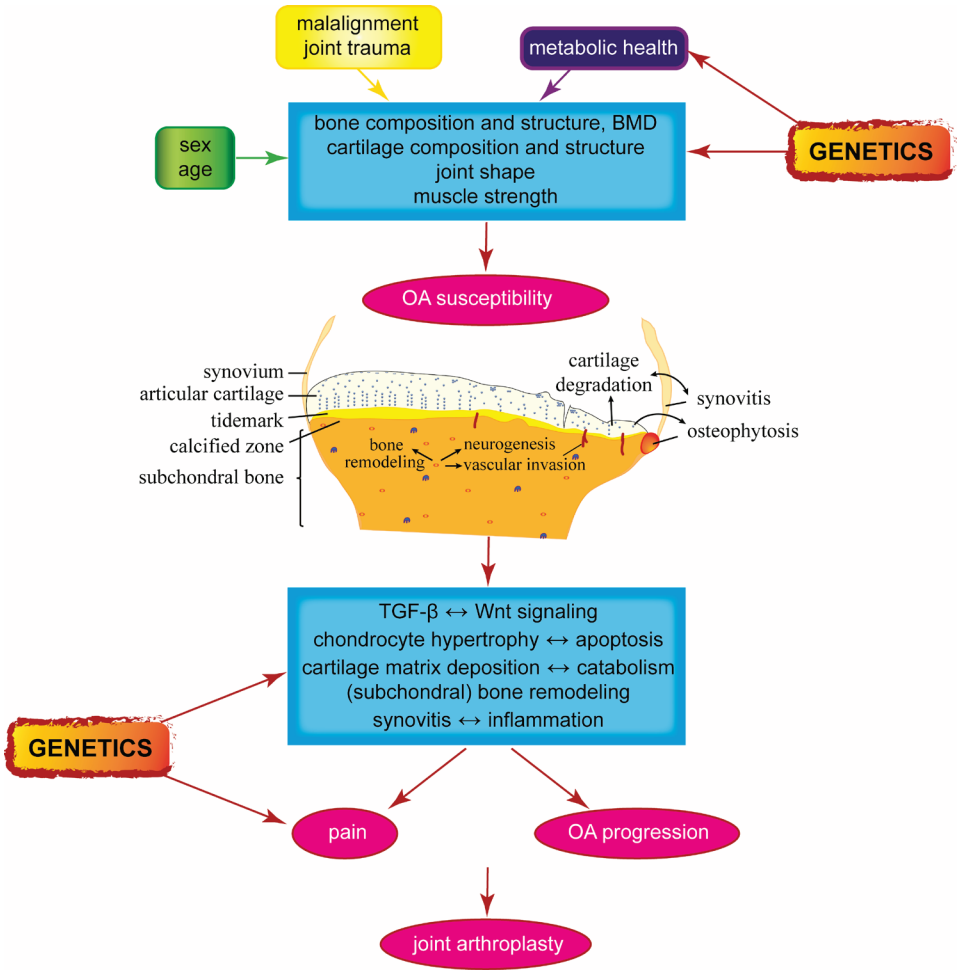


Figure 2 Role of genetic variation in susceptibility to osteoarthritis and effects for joint tissues.

by the observation that it is transmitted in a non-Mendelian manner. Major risk factors for development and progression of OA are joint injury, aging, poor metabolic health (e.g. obesity), and genetic predisposition.[16, 20, 21] Importantly, genetic factors influencing OA can apply via metabolic health[22, 23] as well as via skeletal features: bone and cartilage structure and composition, and aspects of shape[24-26] (Figure 2). The heritability of osteoarthritis was already hypothesized in the beginning of last century and it was firstly confirmed in family-based studies showing that 39-60% of OA incidences can be attributed to heritability.[27, 28] In fact, Dieppe *et. al.*[29] stated that development of OA is the result of the dynamics between genetics and environment, with a continuous distribution between the extremes of predominantly genetic and predominantly environmental influences.

Since clear insight into the etiology of OA would greatly contribute to early

diagnosis as well as to the development of disease modifying treatments that are still lacking and to the possibility to better predict and monitor OA progression, major efforts are made to unravel the genetic secrets of OA. As a complex disease, the genetic component of OA is highly polygenic, with multiple risk alleles conferring small effects. This adds up to the mentioned clinical and radiological heterogeneity among patients. Therefore, genetic studies in OA require careful consideration of the study design and patient selection to increase the chance to identify susceptibility genes. Whereas large-scale cohorts of random samples (population-based or case-control studies) are more likely to identify common variants with small effect sizes, family-based studies of more severe OA phenotypes are expected to result in the identification of rare variants, with larger effects.[30]

Digging into the genetics underlying osteoarthritis. Genetic research in OA originally included mainly candidate gene association studies and genome wide linkage analyses. Especially the findings of the candidate gene studies appeared difficult to confirm in replication studies due to problems such as small sample size and the small numbers of polymorphic variants tested.[27, 31] The HapMap Project has shifted genetic epidemiology of complex inheritance away from linkage into genome wide association mapping of loci affecting disease and response to therapy.[32] Although the resolution of genome wide association studies (GWAS) is limited, the use of single nucleotide polymorphisms (SNPs) instead of microsatellites which were used in linkage studies has substantially improved the resolution.[33] Similar to genome wide linkage studies, GWA studies are hypothesis-free. However, GWAS require large sample sizes and are based on high throughput genotyping that typically ascertain at least 300,000 SNPs per sample nowadays. Before going on with the analysis of each of these variants for association with the trait of interest, genotyping is mostly followed by imputation of SNPs from the 1000 Genome Project (http://www.1000genomes.org/). The imputation methodology increases the genetic information per study subjects through the prediction of genotypes at genomic locations that have not actually been measured based on their known linkage disequilibrium (LD) with the genotyped SNPs.

A problem that scientists encounter in GWA studies is the so-called 'missing heritability':[34] although heritability estimates of OA are relatively high (39-60%[27, 28]), the common susceptibility loci identified thus far only explain a small part of the heritability. To some extend this could be explained by a possible overestimation of the heritability based on twin and family studies. On the other hand, it could suggest that many loci still remain to be identified. OA susceptibility loci identified by GWAS have small odds ratios. As explained by Manolio *et al*, this is due to a trade-off between higher risk allele frequencies and stronger genetic effects,[30] typical for the identification of genetic variants predisposing to complex diseases. To overcome this problem and to increase power, researchers have tried to apply GWAS with ever increasing sample size: large scale genome wide initiatives were established to pile up substantial amounts of genome wide association data for meta-analyses such as the GWAS in arcOGEN,[35] and the joint-stratified meta-

analyses of GWAS performed within the Treat~OA consortium for knee[36] and for hip OA.[37] Together, the GWAS approach has resulted in the highest number of consistent OA susceptibility loci (Table 1A).

Next generation sequencing (NGS) has greatly augmented research possibilities and enhanced the pace of discovery of disease-associated genetic variants. Especially with the current increasing speed and decreasing costs, most likely NGS will shift its focus from sequencing of exomes to sequencing of the whole genome. The expectations that this may solve 'the mysterious case of the missing heritability' are high but to identify causal variants by NGS, well-selected (severely affected) patients and their families are required as well as extensive computational tools to process the information. Although generalizability of results from mutations private to a particular family is uncertain, it is expected that more common variants in the same gene or in the same pathways in which the genes function will affect susceptibility in a larger population.[38] In OA this paradigm has been proven its validity for example for the *SMAD3*,[39, 40] *COL11A1*,[41-43] and *ALDH1A2*[44] genes.

Established osteoarthritis susceptibility loci. At present, based on genome wide significance in GWAS ($p \leq 5x10^{-8}$) and/or proven functional involvement in OA by follow-up studies, 21 independent OA susceptibility loci were established from candidate gene studies, linkage studies, and GWAS (Table 1A and reviewed in references 31 and/or 45). As a comparison, in the first half of 2013 the number of loci known to predispose to other complex diseases such as type 2 diabetes and coronary heart disease were respectively 65[46] and 46.[47] For anthropometric traits such as body mass index (BMI), waist-hip ratio, and height the numbers of associated genetic loci identified in May 2013 were 36, 14, and 184, respectively.[48] Identified OA susceptibility loci resulted in a variety of compelling OA candidate genes, however, GWA studies also identified loci within gene deserts or within regions with multiple genes that have not yet been implicated in OA etiology.[35, 49] Multiple SNPs in a region can be in strong LD with each other (Table 1A), making it often challenging to determine the causal variant. This illustrates the typical problem in mapping complex traits and exposes the existing gap between genetic evidence and the molecular mechanism underlying the disease. In order to defeat these problems multiple molecular determinants (genetic, transcriptomic, proteomic, and epigenetic) will have to be integrated on a more regular base.[50, 51] Based on signals found to associate with hip OA Evangelou *et al* calculated that, together, the identified OA loci contribute only 2-3% to the heritability (H^2) of hip OA.[37] Using the same formula that was used by Evangelou and that was described by Park and colleagues,[52] the overall heritability is estimated to be around 13% based on the 17 independent OA susceptibility loci identified with GWAS (Table 1A below 'Genome Wide Association Study'). This seemingly high percentage as compared to the H^2 of hip OA alone is mainly explained by a few loci with larger odds ratios such as rs12907038 annotated to the *ALDH1A2* gene and associated with hand OA, and rs7639618 located within *DVWA* or *COL6A4P1* and associated with knee OA. However, as compared to the 39-60% estimated from family-based studies it suggests that still many of the loci involved in

Table 1. Established OA susceptibility loci (A) and pathways involved (B) (EA: effect allele; EAF: allele frequency of effect allele; OR: odds ratio; CI: confidence interval; F: female; M: male; FM: both female and male; FF: functional follow-up; *signal found in independent studies in Asians[53] and Europeans;[54] **SNPs in strong linkage disequilibrium; #SNPs in strong linkage disequilibrium; ‡SNPs in strong linkage disequilibrium; †signal decreased when adjusting for body mass index suggesting that effect on OA is applied via obesity, possibly targeting the iroquois homeobox 3 (*IRX3*) gene and not *FTO*[55]).

A

SNP	CHR	POS	Annotated Gene	Gene Symbol	EA	EAF	OR (95% CI)	*p*	Joint	Sex	Ref	FF
CANDIDATE GENE STUDY:												
Triple D repeat	9		Asporin	*ASPN*			2 (1.4-2.8)	6.60E-06	Knee	M	[56-59]	[56, 60]
rs12901499	15	67370695	Mothers against decapentaplegic homolog 3	*SMAD3*	G	0.57	1.22 (1.12-1.34)	7.5E-6	Hip & knee	FM	[39]	[61]
rs143383*	20	34025983	Growth differentiation factor 5	*GDF5*	T	0.74	1.79 (1.53-2.09)	2.0E-13	Hip	FM	[62]	[63-66]
LINKAGE STUDY:												
rs7775	2	183699584	Frizzled motif associated with bone development	*FRZB*	G	0.11	4.1 (1.6-10.7)	4.0E-03	Hip	F	[67]	[68-70]
rs225014	14	80669580	Deiodenase iodothyronine type 2	*DIO2*	C	0.34	1.79 (1.37-2.34)	2.0E-05	Hip	F	[71]	[70, 72, 73]
GENOME WIDE ASSOCIATION STUDY:												
rs2615977	1	103452642	Collagen type 11A1	*COL11A1*	C	0.22	1.10 (1.05-1.16)	3.2E-05	Hip	FM	[43, 74]	[75]
rs7639618	3	15216429	Double Von Willebrand Factor A	*DVWA*	G	0.63	1.43 (1.28-1.59)	7.3E-11	Knee	FM	[53]	
rs11177**	3	52721305	**3p21.1 locus**: e.g. Guanine nucleotide binding protein-like 3	*GNL3*	A	0.38	1.12 (1.08-1.16)	1.3E-10	Hip & knee	FM	[35]	[76]
rs6976**	3	52728804	**3p21.1 locus**: e.g. Glycosyltransferase 8 domain containing 1	*GLT8D1*	T	0.37	1.12 (1.08-1.16)	7.2E-11	Hip & knee	FM	[35]	
rs12107036	3	189600160	Tumor protein p63	*TP63*	G	0.52	1.21 (1.13-1.29)	6.7E-08	Knee	F	[35]	
rs10947262#	6	32373312	Butyrophilin-like 2	*BTNL2*	C	0.58	1.31 (1.20-1.44)	5.1E-09	Knee	FM	[77]	
rs7775228#	6	32658079	Major histocompatibility complex, class II, DQ beta 1	*HLA-DQB1*	T	0.62	1.34 (1.21-1.49)	2.4E-08	Knee	FM	[77]	
rs10948172	6	44777691	Suppressor of Ty 3 homolog and CD5 molecule-like	*SUPT3H; CDC5L*	G	0.29	1.14 (1.09-1.20)	7.9E-08	Hip & knee	M	[35]	

rs9350591	6	76241527	Filamin A interacting protein 1 and SUMO1/sentrin specific peptidase 6	FILIP1; SENP6	T	0.11	1.18 (1.12-1.25)	2.4E-09	Hip	FM	[35]	
rs3815148‡	7	106938420	**7q22 locus**: e.g. Component of oligomeric golgi complex 5 and HMG-box transcription factor 1	*COG5; HBP1*	C	0.23	1.14 (1.09-1.19)	8.0E-08	Knee & hand	FM	[49]	[78]
rs4730250‡	7	107207695	**7q22 locus**: e.g. Dihydrouridine synthase 4-like	*DUS4L*	G	0.17	1.17 (1.11-1.24)	9.2E-09	Knee	FM	[36]	
rs4836732	9	119266695	Astrotactin 2 and Pregnancy-associated plasma -protein A	*ASTN2; PAPPA*	C	0.47	1.2 (1.13-1.27)	6.1E-10	Hip	F	[35]	
rs10492367	12	28014970	Kelch domain containing 5 and Parathyroid hormone-like hormone	*KLHDC5; PTHLH*	T	0.19	1.14 (1.09-1.20)	1.5E-08	Hip	FM	[35]	
rs835487	12	105060767	Carbohydrate (chondroitin 4) sulfotransferase 11	*CHST11*	G	0.34	1.13 (1.09-1.18)	1.6E-08	Hip	FM	[35]	
rs11842874	13	113694509	MCF.2 cell line derived transforming sequence-like	*MCF2L*	A	0.93	1.17 (1.11-1.23)	2.1E-08	Knee & hip	FM	[79]	
rs12907038	15	58244032	Aldehyde dehydrogenase 1 family, member A2	*ALDH1A 2*	G	0.507	1.51	4.0E-10	Hand	FM	[44]	
rs8044769†	16	53839135	Fat mass and obesity associated	*FTO*	C	0.5	1.11 (1.07-1.15)	6.9E-08	Hip & knee	F	[35]	
rs12982744	19	2177193	DOT1-like histone H3K79 methyltransferase	*DOT1L*	C	NA	1.17 (1.11-1.23)	7.8E-09	Hip	M	[80]	
rs143383*	20	34025983	Growth differentiation factor 5	*GDF5*	T	NA	1.16 (1.11-1.22)	8.3E-09	Knee	FM	[54]	
rs6094710	20	46095649	Nuclear receptor coactivator 3	*NCOA3*	A	0.04	1.28 (1.18-1.39)	7.9E-09	Hip	FM	[37]	

B

Term	Count	Pct	*p*	p_{adj}	Pop Hits	Fold Enr.	Bonf.	Benj.	FDR
GO:0030326~embryonic limb morphogenesis	5	19.23	4.56E-06	1.19E-04	87	40.92	2.01E-03	2.01E-03	0.01
GO:0035113~embryonic appendage morphogenesis	5	19.23	4.56E-06	1.19E-04	87	40.92	2.01E-03	2.01E-03	0.01
GO:0035107~appendage morphogenesis	5	19.23	7.63E-06	1.98E-04	99	35.96	3.37E-03	1.69E-03	0.01
GO:0035108~limb morphogenesis	5	19.23	7.63E-06	1.98E-04	99	35.96	3.37E-03	1.69E-03	0.01
GO:0048736~appendage development	5	19.23	8.93E-06	2.32E-04	103	34.56	3.94E-03	1.32E-03	0.01
GO:0060173~limb development	5	19.23	8.93E-06	2.32E-04	103	34.56	3.94E-03	1.32E-03	0.01
GO:0048598~embryonic morphogenesis	6	23.08	3.92E-05	1.02E-03	307	13.92	1.72E-02	4.32E-03	0.06

development of OA remain to be identified.

Given the fact that 21 established susceptibility loci were identified, the number of publications reporting on their functional follow-up is highly disappointing. In fact, to the best of our knowledge as by August 2014 only 8 of the loci have been characterized in more detail following their identification (Table 1A). Some of the studies were aimed at determining the particular gene affected by the genetic variation.[75, 80] Other studies have determined the function of the gene expected to be targeted by the SNP,[37, 44, 81] and several studies intended to determine the mechanism underlying OA susceptibility via the identified locus: either at the molecular level or at an organismal level. With respect to the latter, in particular Growth differentiation factor 5 (*GDF5*[63-66]), Frizzled motif associated with bone development (*FRZB*[68-70]), and Deiodenase iodothyronine type 2 (*DIO2*[70, 72, 73]) have received full attention.

Pathway analysis of the genes annotated to the OA susceptibility loci with the online Database for Annotation, Visualization and Integrated Discovery tool (DAVID[82]) shows enrichment for genes involved in limb development (Table 1B). Since several of the identified OA susceptibility genes encode for proteins involved in endochondral ossification during early development such as *DIO2*, *FRZB*, and *GDF5* (see Figure 3), OA has been hypothesized to be an early developmental disease with late-life onset.[24] The developmental origin of OA is twofold: First, genes involved in endochondral ossification determine both composition of the bone and cartilage and the shape of the joint. While a suboptimal joint shape will make the tissues more prone to damage and thus to development of OA,[26] bone and cartilage strength determine the capacity to persist mechanical stresses. Secondly, the maturational arrested phenotype of articular chondrocytes is changed during OA and chondrocytes retrieve characteristics from hypertrophic growth plate chondrocytes.[83] Given this hypothesis, and thus the need to investigate early as well as late effects of genetic variation, the best would be to apply *in vivo* animal models in which developmental effects as well as late life effects can be studied, and where mechanical stresses can be applied in a controlled manner for example by using a running model. Alternatively, experiments can be performed *in vitro* to study genetic effects on chondrogenic potential of mesenchymal stem cells and matrix deposition of primary chondrocytes.

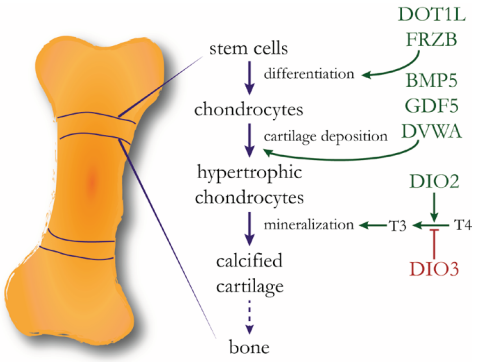


Figure 3 OA susceptibility genes involved in endochondral ossification.

BIOBANKING IN OA RESEARCH

To achieve the aims of the work presented in this thesis and in general for OA research, the assembly of a biobank with relevant tissues stored tailored to isolate RNA and DNA, possibly also to make protein lysates

for analysis by Western blotting, and to apply immunohistochemistry is required. In addition, the collection of relevant cell types (mesenchymal stem cells, chondrocytes, osteoblasts) would facilitate the setup of experimental cell models. Together, this will allow analyses at multiple molecular levels.

Human genome wide gene expression datasets are available for OA cartilage from non-OA and OA-affected joints[84, 85] as well as from unaffected and affected cartilage areas of the same joints[86-88]), OA affected and unaffected subchondral bone,[89] and OA synovial membrane.[90] However, these are collected in different studies. Ideally, the different datasets should be derived from the same study group which could provide detailed information on the dynamics of gene expression profiles during OA pathophysiological processes and also include expression profiles of blood which would permit detection of biomarkers in the blood that reflect ongoing OA. Besides gene expression, RNA can also be used to study differential allelic expression of genetic variants. The isolated DNA can be used for genotyping and to perform whole genome or exome sequencing and to establish epigenetic patterns of the different tissues.

For the studies performed in this thesis we used biomaterials of the ongoing RAAK study (Research Arthritis and Articular Cartilage), in which joint materials (cartilage of healthy and diseased (areas of the) joint, bone, and where available ligaments), mesenchymal stem cells (hip joints only), and primary chondrocytes of controls and patients undergoing joint replacement surgery in the Leiden University Medical Center and collaborating outpatient clinics in the Leiden area are collected. In addition to the RAAK study, studies were performed in the Genetics osteoARthritis and Progression (GARP) study. This study consists of sibling pairs with symptomatic and radiographic OA in at least two joints[91] and is tailored for molecular epidemiological research. Of all participants a complete collection of biomaterials such as DNA, RNA (blood), serum, plasma, and urine for biomarker detection is available in addition to extended clinical (X-rays, MRI) and demographic data in a follow-up design at 2 and 5 years. In parallel (including follow-up) data on 30 age-matched controls without OA in the indicated joints has been collected, the so-called NORREF study.

To summarize, genetic studies have identified many OA susceptibility loci but most of the heritability is still missing. Several issues warrant further investigation. In the first place, treatment of OA is substantially impaired by its late diagnosis, and therapies would significantly benefit from the availability of biomarkers that can detect the disease at an early stage, monitor disease progression, and distinguish between different OA subtypes. Secondly, further insight into genetic factors causal to OA will increase insight into the etiology of OA. With respect to GWA studies, more severe selection of samples to decrease heterogeneity is required. Current availability of next generation sequencing is likely to contribute to the identification of genetic variants predisposing to OA. However, this will require cutting-edge study design and careful selection of samples. Thirdly, functional genomic research to generate biological understanding of genetic susceptibility beyond that of knowing the genomic location is lacking way behind the mapping of OA loci.

RATIONALE AND OUTLINE OF THIS THESIS

In this thesis, we set out to characterize in detail the ongoing pathophysiological processes in arthritic cartilage within the RAAK study (**chapter 2**). Genome wide gene expression profiles were generated from OA affected and macroscopically intact cartilage of the same joint and gene enrichment analysis was performed to study the major pathways involved. Also genome wide expression profiles were determined from blood of patients with generalized OA from the GARP study and compared to those of healthy controls (**chapter 3**). Results were used to identify genes that may serve as biomarkers for OA, and overlap between both databases (cartilage and blood) was investigated in order to learn to what extent ongoing OA processes may be reflected in the blood.

Although selection of biomarkers in OA is promising, the distinction between individual patients and controls is not yet possible due to relatively high inter-individual variation. Variation in basal levels of biomarkers may originate from genetic influences, and taking this variation into account by stratification for genotype or by using other biomarkers for persons with a specific genotype could improve the informativity of the biomarkers on OA onset and progression detection. Therefore, in **chapter 4** we aimed at the identification of genetic loci affecting levels of the promising biomarkers serum COMP and urinary CTX-II. Subsequently, we explored a meta-analysis of GWAS to determine whether any of the identified loci were associated to OA.

To identify genetic loci associated with OA, 2 different approaches were taken. As an alternative to the ever increasing number of samples used in GWAS we investigated whether selection of OA patients with familial and relatively severe phenotype could be an alternative (**chapter 5**). This was done by investigating the GARP study, in which siblings were selected based on their relatively severe generalized OA phenotype. Furthermore, exome sequencing was done for 2 members of a family characterized by early-onset OA (**chapter 6**). By applying exome sequencing we aimed to identify a genetic variant with large effect that could hint towards general pathways involved in OA. Finally, in **chapter 7**, a 3-dimensional *in vitro* chondrogenesis method was set up that can be applied in functional genomic research. Here, it was used to functionally characterize the effects of variation in thyroid signaling during chondrogenesis as a model for variation in the OA susceptibility gene *DIO2*.

REFERENCES

1 Cross M, Smith E, Hoy D, Nolte S, Ackerman I, Fransen M, *et al.* The global burden of hip and knee osteoarthritis: estimates from the Global Burden of Disease 2010 study. *Ann Rheum Dis* 2014.

2 Bijlsma JW, Berenbaum F, Lafeber FP. Osteoarthritis: an update with relevance for clinical practice. *Lancet* 2011;**377**:2115-26.

3 Schaap LA, Peeters GM, Dennison EM, Zambon S, Nikolaus T, Sanchez-Martinez M, *et al.* European Project on OSteoArthritis (EPOSA): methodological challenges in harmonization of existing data from five European population-based cohorts on aging

BMC Musculoskelet Disord 2011;**12**:272.

4 Valdes AM, Spector TD. The clinical relevance of genetic susceptibility to osteoarthritis. *Best Pract Res Clin Rheumatol* 2010;**24**:3-14.

5 Wilder FV, Barrett JP, Farina EJ. Joint-specific prevalence of osteoarthritis of the hand. *Osteoarthritis Cartilage* 2006;**14**:953-7.

6 Kloppenburg M, Kwok WY. Hand osteoarthritis--a heterogeneous disorder. *Nat Rev Rheumatol* 2012;**8**:22-31.

7 Seed SM, Dunican KC, Lynch AM. Treatment options for osteoarthritis: considerations for older adults. *Hosp Pract (Minneap)* 2011;**39**:62-73.

8 Johnson VL, Hunter DJ. The epidemiology of osteoarthritis. *Best Pract Res Clin Rheumatol* 2014;**28**:5-15.

9 Schenker ML, Mauck RL, Ahn J, Mehta S. Pathogenesis and prevention of posttraumatic osteoarthritis after intra-articular fracture. *J Am Acad Orthop Surg* 2014;**22**:20-8.

10 Colao A, Pivonello R, Scarpa R, Vallone G, Ruosi C, Lombardi G. The acromegalic arthropathy. *J Endocrinol Invest* 2005;**28**:24-31.

11 Mills K, Hunter DJ. Patellofemoral joint osteoarthritis: An individualised pathomechanical approach to management. *Best Pract Res Clin Rheumatol* 2014;**28**:73-91.

12 Castano-Betancourt MC, Rivadeneira F, Bierma-Zeinstra S, Kerkhof HJ, Hofman A, Uitterlinden AG, *et al.* Bone parameters across different types of hip osteoarthritis and their relationship to osteoporotic fracture risk. *Arthritis Rheum* 2013;**65**:693-700.

13 Hannan MT, Felson DT, Pincus T. Analysis of the discordance between radiographic changes and knee pain in osteoarthritis of the knee. *J Rheumatol* 2000;**27**:1513-7.

14 Radin EL, Paul IL, Tolkoff MJ. Subchondral bone changes in patients with early degenerative joint disease. *Arthritis Rheum* 1970;**13**:400-5.

15 Radin EL, Rose RM. Role of subchondral bone in the initiation and progression of cartilage damage. *Clin Orthop Relat Res* 1986;34-40.

16 Goldring MB. Chondrogenesis, chondrocyte differentiation, and articular cartilage metabolism in health and osteoarthritis. *Ther Adv Musculoskelet Dis* 2012;**4**:269-85.

17 Loeser RF, Goldring SR, Scanzello CR, Goldring MB. Osteoarthritis: a disease of the joint as an organ. *Arthritis Rheum* 2012;**64**:1697-707.

18 Hardcastle SA, Dieppe P, Gregson CL, Hunter D, Thomas GE, Arden NK, *et al.* Prevalence of radiographic hip osteoarthritis is increased in high bone mass. *Osteoarthritis Cartilage* 2014.

19 Pan J, Zhou X, Li W, Novotny JE, Doty SB, Wang L. In situ measurement of transport between subchondral bone and articular cartilage. *J Orthop Res* 2009;**27**:1347-52.

20 Chapman K, Valdes AM. Genetic factors in OA pathogenesis. *Bone* 2012;**51**:258-64.

21 Bay-Jensen AC, Slagboom E, Chen-An P, Alexandersen P, Qvist P, Christiansen C, *et al.* Role of hormones in cartilage and joint metabolism: understanding an unhealthy metabolic phenotype in osteoarthritis. *Menopause* 2013;**20**:578-86.

22 Elliott KS, Chapman K, Day-Williams A, Panoutsopoulou K, Southam L, Lindgren CM, *et al.* Evaluation of the genetic overlap between osteoarthritis with body mass index and height using genome-wide association scan data. *Ann Rheum Dis* 2013;**72**:935-41.

23 Panoutsopoulou K, Metrustry S, Doherty SA, Laslett LL, Maciewicz RA, Hart DJ, *et al.* The effect of FTO variation on increased osteoarthritis risk is mediated through body mass index: a mendelian randomisation study. *Ann Rheum Dis* 2013.

24 Bos SD, Slagboom PE, Meulenbelt I. New insights into osteoarthritis: early developmental features of an ageing-related disease. *Curr Opin Rheumatol* 2008;**20**:553-9.

25 Pitsillides AA, Beier F. Cartilage biology in osteoarthritis--lessons from developmental biology. *Nat Rev Rheumatol* 2011;**7**:654-63.

26 Sandell LJ. Etiology of osteoarthritis: genetics and synovial joint development. *Nat Rev*

Rheumatol 2012;**8**:77-89.

27 Valdes AM, Spector TD. Genetic epidemiology of hip and knee osteoarthritis. *Nat Rev Rheumatol* 2011;**7**:23-32.

28 Ikegawa S. The genetics of common degenerative skeletal disorders: osteoarthritis and degenerative disc disease. *Annu Rev Genomics Hum Genet* 2013;**14**:245-56.

29 Dieppe PA, Lohmander LS. Pathogenesis and management of pain in osteoarthritis. *Lancet* 2005;**365**:965-73.

30 Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, *et al.* Finding the missing heritability of complex diseases. *Nature* 2009;**461**:747-53.

31 Panoutsopoulou K, Zeggini E. Advances in osteoarthritis genetics. *J Med Genet* 2013;**50**:715-24.

32 Morton NE. Into the post-HapMap era. *Adv Genet* 2008;**60**:727-42.

33 Sellick GS, Longman C, Tolmie J, Newbury-Ecob R, Geenhalgh L, Hughes S, *et al.* Genomewide linkage searches for Mendelian disease loci can be efficiently conducted using high-density SNP genotyping arrays. *Nucleic Acids Res* 2004;**32**:e164.

34 Maher B. Personal genomes: The case of the missing heritability. *Nature* 2008;**456**:18-21.

35 Zeggini E, Panoutsopoulou K, Southam L, Rayner NW, Day-Williams AG, Lopes MC, *et al.* Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* 2012;**380**:815-23.

36 Evangelou E, Valdes AM, Kerkhof HJ, Styrkarsdottir U, Zhu Y, Meulenbelt I, *et al.* Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. *Ann Rheum Dis* 2011;**70**:349-55.

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

38 Goldstein DB, Allen A, Keebler J, Margulies EH, Petrou S, Petrovski S, *et al.* Sequencing studies in human genetics: design and interpretation. *Nat Rev Genet* 2013;**14**:460-70.

39 Valdes AM, Spector TD, Tamm A, Kisand K, Doherty SA, Dennison EM, *et al.* Genetic variation in the SMAD3 gene is associated with hip and knee osteoarthritis. *Arthritis Rheum* 2010;**62**:2347-52.

40 Hilhorst-Hofstee Y, Scholte AJ, Rijlaarsdam ME, van HA, Kroft LJ, Reijnierse M, *et al.* An unanticipated copy number variant of chromosome 15 disrupting SMAD3 reveals a three-generation family at serious risk for aortic dissection. *Clin Genet* 2013;**83**:337-44.

41 Mio F, Chiba K, Hirose Y, Kawaguchi Y, Mikami Y, Oya T, *et al.* A functional polymorphism in COL11A1, which encodes the alpha 1 chain of type XI collagen, is associated with susceptibility to lumbar disc herniation. *Am J Hum Genet* 2007;**81**:1271-7.

42 Tompson SW, Bacino CA, Safina NP, Bober MB, Proud VK, Funari T, *et al.* Fibrochondrogenesis results from mutations in the COL11A1 type XI collagen gene. *Am J Hum Genet* 2010;**87**:708-12.

43 Panoutsopoulou K, Southam L, Elliott KS, Wrayner N, Zhai G, Beazley C, *et al.* Insights into the genetic architecture of osteoarthritis from stage 1 of the arcOGEN study. *Ann Rheum Dis* 2011;**70**:864-7.

44 Styrkarsdottir U, Thorleifsson G, Helgadottir HT, Bomer N, Metrustry S, Bierma-Zeinstra S, *et al.* Severe osteoarthritis of the hand associates with common variants within the ALDH1A2 gene and with rare variants at 1p31. *Nat Genet* 2014;**46**:498-502.

45 Reynard LN, Loughlin J. The genetics and functional analysis of primary osteoarthritis susceptibility. *Expert Rev Mol Med* 2013;**15**:e2.

46 Yaghootkar H, Frayling TM. Recent progress in the use of genetics to understand links between type 2 diabetes and related metabolic traits. *Genome Biol* 2013;**14**:203.

47 Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, *et al.* Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*

2013;**45**:25-33.

48 Berndt SI, Gustafsson S, Magi R, Ganna A, Wheeler E, Feitosa MF, *et al.* Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet* 2013;**45**:501-12.

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

50 Iliopoulos D, Malizos KN, Oikonomou P, Tsezou A. Integrative microRNA and proteomic approaches identify novel osteoarthritis genes and their collaborative metabolic and inflammatory networks. *PLoS One* 2008;**3**:e3740.

51 Westra HJ, Franke L. From genome to function by studying eQTLs. *Biochim Biophys Acta* 2014.

52 Park JH, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, *et al.* Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet* 2010;**42**:570-5.

53 Miyamoto Y, Shi D, Nakajima M, Ozaki K, Sudo A, Kotani A, *et al.* Common variants in DVWA on chromosome 3p24.3 are associated with susceptibility to knee osteoarthritis. *Nat Genet* 2008;**40**:994-8.

54 Valdes AM, Evangelou E, Kerkhof HJ, Tamm A, Doherty SA, Kisand K, *et al.* The GDF5 rs143383 polymorphism is associated with osteoarthritis of the knee with genome-wide statistical significance. *Ann Rheum Dis* 2011;**70**:873-5.

55 Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* 2014;**507**:371-5.

56 Kizawa H, Kou I, Iida A, Sudo A, Miyamoto Y, Fukuda A, *et al.* An aspartic acid repeat polymorphism in asporin inhibits chondrogenesis and increases susceptibility to osteoarthritis. *Nat Genet* 2005;**37**:138-44.

57 Mustafa Z, Dowling B, Chapman K, Sinsheimer JS, Carr A, Loughlin J. Investigating the aspartic acid (D) repeat of asporin as a risk factor for osteoarthritis in a UK Caucasian population. *Arthritis Rheum* 2005;**52**:3502-6.

58 Valdes AM, Loughlin J, Oene MV, Chapman K, Surdulescu GL, Doherty M, *et al.* Sex and ethnic differences in the association of ASPN, CALM1, COL2A1, COMP, and FRZB with genetic susceptibility to osteoarthritis of the knee. *Arthritis Rheum* 2007;**56**:137-46.

59 Kaliakatsos M, Tzetis M, Kanavakis E, Fytili P, Chouliaras G, Karachalios T, *et al.* Asporin and knee osteoarthritis in patients of Greek origin. *Osteoarthritis Cartilage* 2006;**14**:609-11.

60 Kou I, Nakajima M, Ikegawa S. Expression and regulation of the osteoarthritis-associated protein asporin. *J Biol Chem* 2007;**282**:32193-9.

61 Raine EV, Reynard LN, van de Laar IM, Bertoli-Avella AM, Loughlin J. Identification and analysis of a SMAD3 cis-acting eQTL operating in primary osteoarthritis and in the aneurysms and osteoarthritis syndrome. *Osteoarthritis Cartilage* 2014;**22**:698-705.

62 Miyamoto Y, Mabuchi A, Shi D, Kubo T, Takatori Y, Saito S, *et al.* A functional polymorphism in the 5' UTR of GDF5 is associated with susceptibility to osteoarthritis. *Nat Genet* 2007;**39**:529-33.

63 Dodd AW, Syddall CM, Loughlin J. A rare variant in the osteoarthritis-associated locus GDF5 is functional and reveals a site that can be manipulated to modulate GDF5 expression. *Eur J Hum Genet* 2013;**21**:517-21.

64 Syddall CM, Reynard LN, Young DA, Loughlin J. The identification of trans-acting factors that regulate the expression of GDF5 via the osteoarthritis susceptibility SNP rs143383. *PLoS Genet* 2013;**9**:e1003557.

65 Ratnayake M, Ploger F, Santibanez-Koref M, Loughlin J. Human chondrocytes respond discordantly to the protein encoded by the osteoarthritis susceptibility gene GDF5. *PLoS*

One 2014;**9**:e86590.

66 Reynard LN, Bui C, Syddall CM, Loughlin J. CpG methylation regulates allelic expression of GDF5 by modulating binding of SP1 and SP3 repressor proteins to the osteoarthritis susceptibility SNP rs143383. *Hum Genet* 2014.

67 Loughlin J, Dowling B, Chapman K, Marcelline L, Mustafa Z, Southam L, *et al.* Functional variants within the secreted frizzled-related protein 3 gene are associated with hip osteoarthritis in females. *Proc Natl Acad Sci U S A* 2004;**101**:9757-62.

68 Lories RJ, Peeters J, Bakker A, Tylzanowski P, Derese I, Schrooten J, *et al.* Articular cartilage and biomechanical properties of the long bones in Frzb-knockout mice. *Arthritis Rheum* 2007;**56**:4095-103.

69 Lories RJ, Peeters J, Szlufcik K, Hespel P, Luyten FP. Deletion of frizzled-related protein reduces voluntary running exercise performance in mice. *Osteoarthritis Cartilage* 2009;**17**:390-6.

70 Waarsing JH, Kloppenburg M, Slagboom PE, Kroon HM, Houwing-Duistermaat JJ, Weinans H, *et al.* Osteoarthritis susceptibility genes influence the association between hip morphology and osteoarthritis. *Arthritis Rheum* 2011;**63**:1349-54.

71 Meulenbelt I, Min JL, Bos S, Riyazi N, Houwing-Duistermaat JJ, van der Wijk HJ, *et al.* Identification of DIO2 as a new susceptibility locus for symptomatic osteoarthritis. *Hum Mol Genet* 2008;**17**:1867-75.

72 Bos SD, Bovee JV, Duijnisveld BJ, Raine EV, van Dalen WJ, Ramos YF, *et al.* Increased type II deiodinase protein in OA-affected cartilage and allelic imbalance of OA risk polymorphism rs225014 at DIO2 in human OA joint tissues. *Ann Rheum Dis* 2012;**71**:1254-8.

73 Bomer N, den HW, Ramos YF, Bos SD, van der Breggen R, Lakenberg N, *et al.* Underlying molecular mechanisms of DIO2 susceptibility in symptomatic osteoarthritis. *Ann Rheum Dis* 2014.

74 Rodriguez-Fontenla C, Calaza M, Evangelou E, Valdes AM, Arden N, Blanco FJ, *et al.* Assessment of osteoarthritis candidate genes in a meta-analysis of 9 genome-wide association studies. *Arthritis Rheum* 2013.

75 Raine EV, Dodd AW, Reynard LN, Loughlin J. Allelic expression analysis of the osteoarthritis susceptibility gene COL11A1 in human joint tissues. *BMC Musculoskelet Disord* 2013;**14**:85.

76 Gee F, Clubbs CF, Raine EV, Reynard LN, Loughlin J. Allelic expression analysis of the osteoarthritis susceptibility locus that maps to chromosome 3p21 reveals cis-acting eQTLs at GNL3 and SPCS1. *BMC Med Genet* 2014;**15**:53.

77 Nakajima M, Takahashi A, Kou I, Rodriguez-Fontenla C, Gomez-Reino JJ, Furuichi T, *et al.* New sequence variants in HLA class II/III region associated with susceptibility to knee osteoarthritis identified by genome-wide association study. *PLoS One* 2010;**5**:e9723.

78 Raine EV, Wreglesworth N, Dodd AW, Reynard LN, Loughlin J. Gene expression analysis reveals HBP1 as a key target for the osteoarthritis susceptibility locus that maps to chromosome 7q22. *Ann Rheum Dis* 2012;**71**:2020-7.

79 Day-Williams AG, Southam L, Panoutsopoulou K, Rayner NW, Esko T, Estrada K, *et al.* A Variant in MCF2L Is Associated with Osteoarthritis. *Am J Hum Genet* 2011;**89**:446-50.

80 Evangelou E, Valdes AM, Castano-Betancourt MC, Doherty M, Doherty S, Esko T, *et al.* The DOT1L rs12982744 polymorphism is associated with osteoarthritis of the hip with genome-wide statistical significance in males. *Ann Rheum Dis* 2013;**72**:1264-5.

81 Raine EV, Wreglesworth N, Dodd AW, Reynard LN, Loughlin J. Gene expression analysis reveals HBP1 as a key target for the osteoarthritis susceptibility locus that maps to chromosome 7q22. *Ann Rheum Dis* 2012;**71**:2020-7.

82 Castano Betancourt MC, Cailotto F, Kerkhof HJ, Cornelis FM, Doherty SA, Hart DJ, *et al.* Genome-wide association and functional

studies identify the DOT1L gene to be involved in cartilage thickness and hip osteoarthritis. *Proc Natl Acad Sci U S A* 2012;**109**:8218-23.

83 Huang dW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;**4**:44-57.

84 Sun MM, Beier F. Chondrocyte hypertrophy in skeletal development, growth, and disease. *Birth Defects Res C Embryo Today* 2014;**102**:74-82.

85 Karlsson C, Dehne T, Lindahl A, Brittberg M, Pruss A, Sittinger M, *et al.* Genome-wide expression profiling reveals new candidate genes associated with osteoarthritis. *Osteoarthritis Cartilage* 2010;**18**:581-92.

86 Xu Y, Barter MJ, Swan DC, Rankin KS, Rowan AD, Santibanez-Koref M, *et al.* Identification of the pathogenic pathways in osteoarthritic hip cartilage: commonality and discord between hip and knee OA. *Osteoarthritis Cartilage* 2012;**20**:1029-38.

87 Sato T, Konomi K, Yamasaki S, Aratani S, Tsuchimochi K, Yokouchi M, *et al.* Comparative analysis of gene expression profiles in intact and damaged regions of human osteoarthritic cartilage. *Arthritis Rheum* 2006;**54**:808-17.

88 Ijiri K, Zerbini LF, Peng H, Otu HH, Tsuchimochi K, Otero M, *et al.* Differential expression of GADD45beta in normal and osteoarthritic cartilage: potential role in homeostasis of articular chondrocytes. *Arthritis Rheum* 2008;**58**:2075-87.

89 Tsuritani K, Takeda J, Sakagami J, Ishii A, Eriksson T, Hara T, *et al.* Cytokine receptor-like factor 1 is highly expressed in damaged human knee osteoarthritic cartilage and involved in osteoarthritis downstream of TGF-beta. *Calcif Tissue Int* 2010;**86**:47-57.

90 Chou CH, Wu CC, Song IW, Chuang HP, Lu LS, Chang JH, *et al.* Genome-wide expression profiles of subchondral bone in osteoarthritis. *Arthritis Res Ther* 2013;**15**:R190.

91 Lambert C, Dubuc JE, Montell E, Verges J, Munaut C, Noel A, *et al.* Gene expression pattern of cells from inflamed and normal areas of osteoarthritis synovial membrane. *Arthritis Rheumatol* 2014;**66**:960-8.

92 Riyazi N, Meulenbelt I, Kroon HM, Ronday KH, Hellio le Graverand MP, Rosendaal FR, *et al.* Evidence for familial aggregation of hand, hip, and spine but not knee osteoarthritis in siblings with multiple joint involvement: the GARP study. *Ann Rheum Dis* 2005;**64**:438-43.

Markers of the osteoarthritis process identified through genome wide expression analysis in articular cartilage; the RAAK study

PLoS ONE 9(7): e103056; 2014. (doi:10.1371/journal.pone.0103056)

Markers of the osteoarthritis process identified through genome wide expression analysis in articular cartilage; the RAAK study

Yolande F.M. Ramos[1,2*‡], Wouter den Hollander[1*], Judith V.M.G. Bovée[3], Nils Bomer[1], Ruud van der Breggen[1], Nico Lakenberg[1], J. Christiaan Keurentjes[4], Jelle. J. Goeman[5], P. Eline Slagboom[1,2], Rob G.H.H. Nelissen[4], Steffan D. Bos[1,2**], Ingrid Meulenbelt[1,2**]

[1]Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; [2]The Netherlands Genomics Initiative, sponsored by the NCHA, Leiden-Rotterdam, The Netherlands; [3]Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands; [4]Department of Orthopeadics, Leiden University Medical Center, Leiden, The Netherlands; [5]Department of Biostatistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands.

* Authors contributed equally to this work

** Shared last author position

ABSTRACT

Objective. **Identify gene expression profiles associated with OA processes in articular cartilage and determine pathways changing during the disease process.**

Methods. **Genome wide gene expression was determined in paired samples of OA affected and preserved cartilage of the same joint using microarray analysis for 33 patients of the RAAK study. Results were replicated in independent samples by RT-qPCR and immunohistochemistry. Profiles were analyzed with the online analysis tools DAVID and STRING to identify enrichment for specific pathways and protein-protein interactions.**

Results. **Among the 1717 genes that were significantly differently expressed between OA affected and preserved cartilage we found significant enrichment for genes involved in skeletal development (e.g. *TNFRSF11B* and *FRZB*). Also several inflammatory genes such as *CD55*, *PTGES* and *TNFAIP6*, previously identified in within-joint analyses as well as in analyses comparing preserved cartilage from OA affected joints versus healthy cartilage were among the top genes. Of note was the high up-regulation of *NGF* in OA cartilage. RT-qPCR confirmed differential expression for 18 out of 19 genes with expression changes of 2-fold or higher, and immunohistochemistry of selected genes showed a concordant change in protein expression. Most of these changes associated with OA severity (Mankin score) but were independent of joint-site or sex.**

Conclusion. **We provide further insights into the ongoing OA pathophysiological processes in cartilage, in particular into differences in macroscopically intact cartilage compared to OA affected cartilage, which seem relatively consistent and independent of sex or joint. We advocate that development of treatment could benefit by focusing on these similarities in gene expression changes and/or pathways.**

Keywords: osteoarthritis, gene expression profiling, cartilage, skeletal development

INTRODUCTION

Osteoarthritis (OA) is a degenerative disease of the joints causing pain and disability for an increasing proportion of the population thereby imposing a large patient and socio-economic burden [1,2]. Risk factors for OA include age, sex, joint injury, obesity, and mechanical stresses. In addition, predisposition to OA has a considerable genetic component and it has been proposed that OA can be viewed as a continuum resulting from the interaction between genetics affecting cartilage extracellular matrix composition and joint shape and sensitivity to the other factors mentioned [3,4]. Major efforts are made to identify loci associated with OA susceptibility to elucidate underlying mechanisms [5]. Treatment options to slow down or reverse the OA process are still very limited and at the time of diagnosis the damage is already irreversible. Together, this emphasizes the importance to increase insight into the disease process and to identify genes and pathways involved in development of OA. A way to achieve this is by investigating the pathophysiological processes in articular cartilage by means of gene expression analyses.

Initially, expression profiles were established for cartilage from knee OA joints in comparison to healthy joints using only a limited number of genes [6]. More recently, exploratory genome wide expression profiling has been performed for the intact cartilage of hip and knee OA joints of patients undergoing joint replacement surgery compared to non-OA joints either derived from autopsies or from neck of femur fractures [7,8]. These studies showed that many genes involved in extracellular matrix (ECM) production as well as genes involved in ECM degradation or in inflammation were changed. Together, this resulted in significant enrichment for genes involved in skeletal development and response to external stimuli. Although studies that compare healthy cartilage with the preserved cartilage of joints from OA patients are very useful to acquire insight into the pathogenetic differences, the findings are likely biased by confounding factors such as innate differences, age, and stratification by joint. Moreover, due to the study design distinction between age-related changes and early or late changes of OA pathophysiology is hampered.

One of the characteristics of OA is focal loss of articular cartilage, resulting in areas of degradation as well as areas with a relative preservation of cartilage thickness and appearance in the joint. Insight into gene expression specific for the focal areas of cartilage degradation compared to those in preserved areas can provide clues towards dynamic changes of genes and pathways involved in OA pathophysiology independent of confounding factors such as age. Gene expression profiles of cartilage from OA affected and macroscopically preserved areas of the same joint have been determined before, however, in most of these studies limited numbers of donors (4-5 knee joints) were included [9-11].

As part of the ongoing Research Arthritis and Articular Cartilage (RAAK) study we set out to perform genome wide analysis of differential gene expression by comparing 33 pairs of matched OA affected and preserved cartilage samples, originating from the same joint of patients that underwent total joint replacement of either hip or knee. Results provide further insights in the ongoing OA

disease processes in cartilage, in particular into differences in macroscopically intact cartilage compared to OA affected cartilage.

MATERIALS AND METHODS

Ethics statement. Participants of the RAAK study provided written informed consent. The ongoing RAAK study and its consent procedure is approved by the institutional ethics review committee (Commissie Medische Ethiek of the Leiden University Medical Center; protocol no. P08.239).

Discovery cohort. The RAAK study is aimed at the biobanking of joint materials as well as mesenchymal stem cells and primary chondrocytes from patients and controls in the Leiden University Medical Center and collaborating outpatient clinics in the Leiden area. In the current study we used paired preserved and OA affected cartilage samples for 33 donors undergoing joint replacement surgery for primary OA (22 hips, 11 knees). Characteristics of the donors are shown in Supplementary Table S1.

At the moment of collection (within 2 hours following surgery) tissue was washed extensively with phosphate buffered saline (PBS) to decrease the risk of contamination by blood. Cartilage was classified macroscopically and collected separately from OA affected and preserved regions around the weight-bearing area of the joint (Supplementary Figure S1). Classification was done based on predefined features of OA related damage as described previously [9,10]: color/whiteness of the cartilage, surface integrity as determined by visible fibrillation/crack formation, and depth and hardness of the cartilage upon sampling with a scalpel. Care was taken to avoid contamination with bone or synovium. Collected cartilage was snap frozen in liquid nitrogen and stored at -80°C prior to RNA extraction.

Validation and replication cohort. Validation was performed by RT-qPCR in 8 sample pairs of the discovery cohort (3 knee and 5 hip) and for replication of the results we included 28 additional matched sample pairs (20 knee, 8 hip) of similar mean age (shown in Supplementary Table S1). Sampling procedures were according to the discovery cohort.

RNA isolation. Cartilage samples were pulverized using a Retsch MM200 under cryogenic conditions. On average 150 mg of pulverized cartilage was dissolved in 1 ml of Trizol reagent, and mixed vigorously. After addition of 200 µl of chloroform the sample was mixed and centrifuged for 15 minutes (16,000 g). The clear aqueous layer was transferred to a new vial and 1 volume of 70% ethanol/DEPC-treated water was added to precipitate RNA. RNA was collected using Qiagen mini columns according to the manufacturers protocol and quality was assessed using a Bioanalyzer lab-on-a-chip. RNA integrity numbers above 8 were considered suitable for microarray analysis.

Microarrays. After *in vitro* transcription, amplification, and labeling with biotin-labeled nucleotides (Illumina® TotalPrep™ RNA Amplification Kit) Illumina HumanHT-12 v3 microarrays were hybridized. Sample pairs were randomly dispersed over the microarrays, however each pair was measured on a single chip. Microarrays were read using

an Illumina Beadarray 500GX scanner and after basic quality checks using Beadstudio software data were analyzed in R statistical programming language. Intensity values were normalized using the "rsn" option in the Lumi-package and absence of large scale between-chip effects was confirmed using the Globaltest-package in which the individual chip numbers were tested for association to the raw data [12]. After removal of probes that were not reliably detected (detection p>0.05 in more than 50% of the samples) a paired t-test was performed for the remaining 13277 probes comprising 11421 unique genes on all sample pairs while adjusting for chip (to adjust for possible batch effects) and using multiple testing correction as implemented in the "BH" (Benjamini and Hochberg) option in the Limma-package. Analyses for differential expression between OA and healthy and between preserved and healthy cartilage was performed likewise, adjusting in addition for sex and for age.

Quantitative reverse transcription PCR (RT-qPCR). 0.5 μg of total RNA was processed with the First Strand cDNA Synthesis Kit according to the manufacturer's protocol (Roche Applied Science) and RT-qPCR was performed for the 19 genes showing at least 2-fold expression differences in the microarray analysis (Taqman gene expression assays used are listed in Supplementary Table S2) using the Biomark™ 96.96 Dynamic Arrays Fluidigm RT-qPCR platform [13]. Relative gene expressions were calculated with the $2^{-\Delta\Delta Ct}$ method [14], using household gene Beta Actin (*ACTB*) expression as internal standard.

Immunohistochemistry and staining analysis. For histological examination, joints were fixed using a 4% formaldehyde solution. Subsequently, samples were decalcified using a 10% EDTA solution and embedded in paraffin. Sections (5 μm) adjacent to the collected area were stained using hematoxilin and eosin (H&E) and toluidine blue. Immunohistochemistry was performed for SERPINE1 (mouse monoclonal antibody from American Diagnostica Inc.) without antigen retrieval and for CD55 (rabbit polyclonal antibody from Santa Cruz Biotechnology Inc.) with heat antigen retrieval (0,01M Citratebuffer pH=6.0) as described previously [15].

Quantification of OA related cartilage damage was scored by 2 observers (JVMGB and YFMR) according to Mankin *et al* [16]. Quantification of SERPINE1 expression was performed by scoring staining of chondrocytes in the superficial, middle, and deep cartilage layer with a score of 0 (no staining), 1 (moderate staining), or 2 (strong staining). Using Generalized Estimating Equations, scores were summed and used as a predictor variable with Mankin score as outcome whilst correcting for sex and age of each donor.

Pathway analysis and protein-protein interaction. Gene enrichment among the 1717 genes showing significant differential expression was performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID) tool [17] selecting for biological processes identified in the Gene Ontology database (GOTERM_BP_FAT in the options menu implemented in DAVID), selecting for cell compartment (GOTERM_CC_FAT), or selecting for molecular function (GOTERM_MF_FAT)

and using the microarray background (HumanHT-12_V3_0_R2_11283641_A). Pathways with $p \leq 0.05$ after correction for multiple testing according to Bonferroni were considered significant (Bonferroni corrections were performed by multiplying the raw P-values with the number of genes included in the analysis).

Enrichment in protein-protein interactions among the significant genes was analyzed with the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) 9.0 [18] available online.

Literature based candidate genes. Based on selected publications of genome wide association study meta-analyses [19-28] we investigated our expression dataset for evidence of differential gene expression of the reported significant candidate genes.

RESULTS

Differential expression between preserved and OA affected cartilage. To identify genes with changed expression in response to ongoing OA processes genome wide expression profiles were generated for preserved and OA affected cartilage of the same joint of 33 donors. Characteristics of the donors are shown in Supplementary Table S1. Males (N=13) and females (N=20) included in the study were aged between 54

Table 1 Genes significantly differently expressed in OA affected cartilage. Shown are results of the microarray analysis (Discovery) and the replication for genes with at least 2-fold change (Dir: direction of effects relative to OA; FC: fold change).

		DISCOVERY			REPLICATION		
Gene Symbol	**Function**	**FC**	**Dir**	***p***	**FC**	**Dir**	***p***
TNFAIP6	cartilage development	2.90	Up	4.40×10^{-8}	4.01	Up	3.61×10^{-4}
CRLF1	immune response	2.97	Up	4.40×10^{-8}	4.01	Up	8.09×10^{-6}
LEPREL1	cartilage development	2.37	Up	4.40×10^{-8}	2.88	Up	9.49×10^{-4}
CD55	complement cascade	2.39	Up	4.40×10^{-8}	3.00	Up	5.18×10^{-5}
RARRES2	immune response	2.26	Dn	6.06×10^{-8}	3.32	Dn	1.23×10^{-4}
NGF	neurogenesis	2.26	Up	3.36×10^{-7}	4.54	Up	2.50×10^{-5}
FRZB	bone development	2.54	Dn	1.28×10^{-6}	3.04	Dn	4.39×10^{-4}
SERPINE1	complement cascade	2.40	Up	1.30×10^{-6}	3.13	Up	1.54×10^{-4}
PTGES	immune response	2.34	Up	1.46×10^{-6}	3.29	Up	1.76×10^{-5}
SPP1	ECM-receptor interaction	2.53	Up	3.40×10^{-6}	2.56	Up	7.05×10^{-3}
CXCL14	immune response	2.16	Up	3.54×10^{-6}	4.97	Up	1.01×10^{-4}
TNFRSF11B	bone development	2.10	Up	4.23×10^{-6}	2.69	Up	6.58×10^{-4}
FN1	ECM-receptor interaction	2.31	Up	1.13×10^{-5}	1.52	Up	3.28×10^{-2}
HBA2	oxygen transport	2.26	Up	2.02×10^{-5}	2.97	Up	6.15×10^{-3}
HBB	oxygen transport	2.00	Up	6.97×10^{-5}	2.50	Up	1.69×10^{-2}
PAPPA	wound healing	2.08	Up	7.69×10^{-5}	2.25	Up	4.05×10^{-3}
CRISPLD1	--	2.50	Dn	1.42×10^{-4}	1.05	--	7.54×10^{-1}
COL9A1	cartilage development	2.18	Dn	2.67×10^{-4}	3.14	Dn	2.27×10^{-4}
CHRDL2	cartilage development	2.39	Dn	3.55×10^{-4}	5.33	Dn	5.71×10^{-4}

and 80 years (mean age: 66.2). In total, 22 patients received a hip replacement and 11 patients underwent total knee replacement. Among all OA joints included in this study (61 in total), 28 pairs were randomly selected to assess the Mankin scores of preserved and affected areas. Mankin scores were significantly higher in the samples macroscopically designated as 'OA affected' as compared to sections distinguished as 'preserved' (mean Mankin score 7.8 vs. 4.7, respectively, p=4x10^{-4}, paired t-test) and as a result gene expression differences can be directly linked to these differences in Mankin scores.

After normalization and correction for multiple testing, significant differential expression between the OA affected and preserved cartilage was identified for 1893 probes, representing 1717 unique genes (Supplementary Table S3A). Among the 1717 unique genes 19 were differently expressed with fold-changes of 2 and higher (Table 1). Notably, 14 of these were up-regulated in OA as compared to preserved cartilage and only 5 were down-regulated. Overall, 748 (44%) of the differentially expressed genes were up-regulated. Larger fold changes were observed in expression of genes well known for their association with OA cartilage such as tumor necrosis factor alpha-induced protein 6 (*TNFAIP6* also known as *TSG-6*, 2.9-fold up in OA cartilage; p=4.4x10^{-8}), cytokine receptor-like factor 1 (*CRLF1*, 3-fold up in OA cartilage; p=4.4x10^{-8}), and *Wnt*-inhibitor frizzled related protein beta (*FRZB*, 2.5-fold down in OA cartilage; p=1.3x10^{-6}). A notable gene highly up-regulated in OA cartilage was neuronal growth factor (*NGF*, 2.3-fold up; p = 3.4 x 10^{-7}).

Validation of the 19 genes with fold-changes of 2 or higher in 8 sample pairs used in the microarray analyses by means of RT-qPCR showed similar effect sizes and directions as those found in the microarray analysis (Supplementary Table S4). Replication performed in an additional set of 28 independent preserved and affected cartilage sample pairs also showed comparable effect sizes and directions and, except for cysteine-rich secretory protein LCCL domain-containing 1 precursor (*CRISPLD1*), all genes were significantly different expressed (Table 1). Individual expression boxplots of the replicated genes are shown in Supplementary Figure S2.

Expression profiles of genes with fold-changes of 2 and higher were analyzed for association with Mankin score as a grade of disease severity (Table 2). Almost all genes associated with Mankin score, except for *COL9A1*, *HBA2* and *HBB*. To further characterize expression of the 19 genes with highest fold changes in OA affected cartilage, we investigated whether the observed changes were either joint or sex specific. As shown in Table 2, for most of the genes fold changes of the (joint or sex) stratified analyses were highly comparable and not statistically different from those of the discovery analysis. However, increased expression of pregnancy-associated plasma protein A (*PAPPA*) was significantly less pronounced in knee OA (1.3-fold increase) than in hip OA (2.6-fold increase).

In addition to the gene expression profiles of preserved and OA affected cartilage, gene expression profiles were also generated for 7 healthy cartilage (characteristics of the donors are shown in Supplementary Table 1C) and explored for the 19 genes. For most of these 19 genes we did not find significant differences between

healthy and preserved cartilage. However, when analyzing the trend of the differences between healthy, preserved and OA affected cartilage we did find a significant linear effect on the expression of most genes. In contrast, expression changes of *CRISPLD1* and *COL9A1* in healthy versus preserved cartilage were not significant and appeared to be increased while the expression in preserved versus OA affected cartilage was found to be decreased (Supplementary Figure S3A).

Functional annotation of genes differently expressed in OA affected cartilage. To investigate whether the genes differently expressed between preserved and OA affected cartilage belonged to specific pathways, we used the online functional annotation tool DAVID. Seven GO-terms referring to 6 independent pathways were identified (Table 3). The most significant GO-term was observed for "skeletal system development". This term captured several of the genes with fold-changes of 2 or higher (e.g. *FRZB* and *TNFRSF11B*). Furthermore, of note was the GO-term referring to "extracellular matrix organization" including decorin (*DCN*) and several collagens (e.g. *COL1A2*, *COL2A1*, *COL3A1*). When analyzing for enrichment using the cellular compartment option, most significant GO-term was "extracellular matrix", and analyzing for molecular function showed genes involved in "copper ion binding" and "glycosaminoglycan binding" to be significantly enriched (Table 3B and C).

Among the 19 genes that were highly changed in OA affected cartilage (at least 2-fold), 3 pathways were significantly enriched with lowest P-value for GO-term "response to wounding" (Table 3D), which included the genes *TNFAIP6*, *SERPINE2*, and *CD55*. When analyzing for interaction among proteins encoded by these genes using STRING, we found significant enrichment for protein-protein interactions (p=4.7x10^{-10}) in which fibronectin-1 (*FN1*) seemed to play

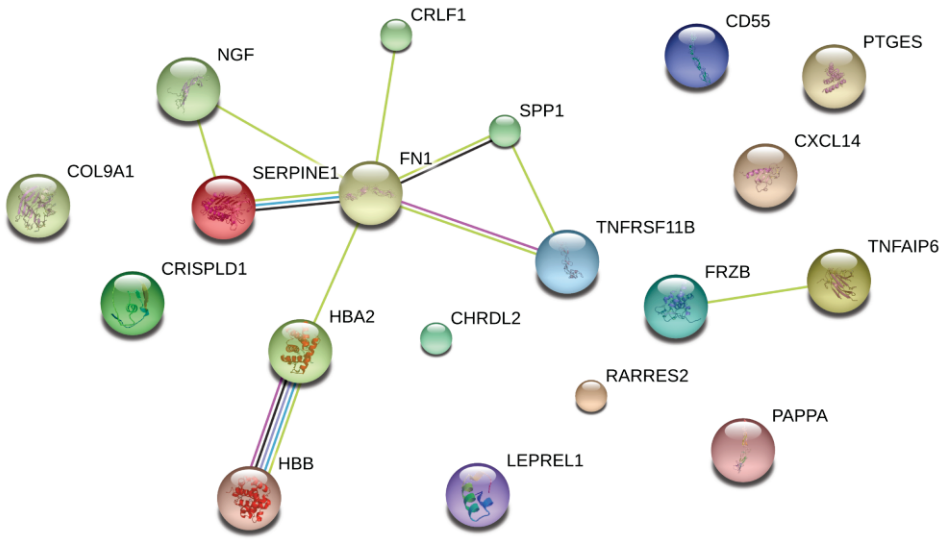


Figure 1. Protein-protein interaction between the genes with expression changes of at least 2-fold (Table 1) as determined with STRING.

a central role considering that 6 of the 19 proteins were found to relate to *FN1* (Figure 1).

Immunohistochemical assessment of proteins encoded by genes identified in the microarray analysis. In addition to differential expression of proteins encoded by genes found in the microarray analysis, immunohistochemical (IHC) staining provides insight in expression pattern and localization of differentially expressed genes in the different cartilage layers. Therefore, as a proof of principal, IHC was performed for SERPINE1 and CD55. Figure 2 shows representative sections for the staining of preserved (P) versus OA affected cartilage, with Figure 2A and B showing an example of the H&E and Toluidine blue staining respectively (upper panel: 4x magnification, lower panel 20x magnification). Immunohistochemistry for SERPINE1 showed that the protein is expressed in chondrocytes and differential expression between OA and preserved cartilage at the protein level was very pronounced (Figure 2C). In OA affected cartilage, SERPINE1 was not only expressed in the superficial layer, but also in the middle layer. In the most affected parts, we even observed SERPINE1 protein expression in chondrocytes residing in the deep zone. In addition, increasing matrix staining of SERPINE1 was observed with increasing OA affection state. We performed a quantification of the staining as described in Materials and Methods, and statistical analysis showed a significant difference between protein abundance in OA and preserved cartilage (p=2.4x10^{-4}). The expression difference seemed to correlate mostly with toluidine blue staining, and thus with the level of proteoglycan constituents of chondromucin aggregates in the samples (p=2.9x10^{-9}). CD55 protein expression was most pronounced in the superficial layer, with higher levels in the more OA affected zones of the cartilage, while hardly any CD55 positive cells were detected in the deep layer (Figure 2D). The differences, however, were more subtle than for SERPINE1 and the range in the quantification did not allow for statistical analysis.

Prioritization of genes residing in compelling genome wide association signals. In order to explore whether genes identified by genome wide association studies are active in cartilage and/or change in response to the OA process, we screened for differential expression of genes originating from recently published large scale meta analyses on OA (Table 4). Sixteen of the 29 genes selected were well detected in the microarray ($p_{detection} \leq 0.05$) and from these, 8 were significantly different between OA and preserved cartilage. Most genes showed only modest expression changes. Of note was differential expression of the HMG-box transcription factor 1 (*HBP1*) gene, identified in the Rotterdam study [23], which showed 1.1-fold up-regulation in OA cartilage (p=2.0x10^{-3}).

DISCUSSION

As part of the RAAK study we compared genome wide expression levels between preserved and OA affected cartilage of the same joint from 33 donors. Such a paired study design allows the detection of genes specifically involved in the OA pathophysiological process, independent of

Table 2 Characteristics of the replicated genes. Results of the analysis for association of the genes with Mankin score and for joint- and sex-stratified analyses (FC: fold change).

		Hip		Knee		Male		Female	
Gene Symbol	p_{Mankin}	**FC**	***p***	**FC**	***p***	**FC**	***p***	**FC**	***p***
TNFAIP6	6.1E-06	2.80	3.89E-05	3.10	6.19E-03	3.09	7.88E-04	2.78	2.11E-04
CRLF1	3.1E-07	2.93	1.49E-04	3.06	1.47E-03	3.51	1.22E-03	2.67	2.34E-04
LEPREL1	5.3E-15	2.50	3.85E-05	2.12	1.67E-02	2.90	7.88E-04	2.07	2.52E-04
CD55	7.8E-11	2.67	3.85E-05	1.92	6.86E-03	2.63	1.53E-03	2.25	2.34E-04
RARRES2	5.9E-08	0.42	3.85E-05	0.48	3.18E-02	0.44	5.51E-03	0.44	2.11E-04
NGF	2.1E-09	2.39	1.49E-04	2.03	7.26E-03	2.68	5.48E-03	2.02	2.69E-04
FRZB	1.0E-05	0.35	1.30E-04	0.50	3.67E-02	0.36	3.20E-03	0.41	1.95E-03
SERPINE1	1.1E-03	2.40	6.10E-04	2.41	6.19E-03	2.64	8.36E-03	2.26	1.15E-03
PTGES	5.0E-05	2.40	3.54E-04	2.21	1.70E-02	2.81	5.05E-03	2.07	1.84E-03
SPP1	2.8E-05	2.56	1.94E-04	2.46	5.87E-02	2.64	9.09E-03	2.45	2.36E-03
CXCL14	2.2E-04	2.49	1.87E-04	1.63	3.16E-02	2.73	9.09E-03	1.85	1.29E-03
TNFRSF11B	8.3E-06	2.27	1.87E-04	1.80	8.65E-02	2.63	8.10E-03	1.82	2.71E-03
FN1	9.2E-05	2.64	1.49E-04	1.83	1.40E-01	2.89	9.35E-03	2.00	4.48E-03
HBA2	9.9E-02	2.46	1.01E-03	1.91	5.04E-02	1.96	2.46E-02	2.49	2.87E-03
HBB	2.3E-01	2.17	1.50E-03	1.70	1.30E-01	1.72	2.86E-02	2.21	5.56E-03
PAPPA	1.6E-05	2.59	4.68E-04	1.34	1.98E-01	2.22	4.79E-02	1.99	6.10E-03
CRISPLD1	5.8E-02	0.38	5.98E-03	0.44	4.28E-02	0.36	2.56E-03	0.43	3.47E-02
COL9A1	5.5E-01	0.42	6.24E-03	0.54	6.46E-02	0.42	4.95E-03	0.49	4.52E-02
CHRDL2	2.8E-04	0.37	3.56E-03	0.52	2.09E-01	0.34	2.70E-02	0.48	3.50E-02

Table 3 DAVID pathway analysis. Analysis considering the biological processes option (GOTERM_BP_FAT option in DAVID), the cellular compartment option (GOTERM_CC_FAT), or the molecular function option (GOTERM_MF_FAT) using medium classification stringency for all genes significantly different expressed between OA affected and preserved cartilage. Legend: Term: GO-terms within the different clusters; Count: number of genes identified for the respective GO-term; Pct: percentage of genes from total number of genes tested; Enr: fold enrichment of indicated pathway; FDR: false discovery rate).

Analysis option	GO-Term	Count	Pct.	Enr.	p	p_{adj}	FDR
Biol. Process	GO:0001501~skeletal system development	59	3.86	2.12	5.21×10^{-8}	7.97×10^{-5}	9.53×10^{-5}
Biol. Process	GO:0007167~enzyme linked receptor protein signaling pathway	56	3.66	1.88	5.74×10^{-6}	8.78×10^{-3}	1.05×10^{-2}
Biol. Process	GO:0008285~negative regulation of cell proliferation	57	3.73	1.81	1.52×10^{-5}	2.32×10^{-2}	2.77×10^{-2}
Biol. Process	GO:0005996~monosaccharide metabolic process	40	2.61	2.07	1.64×10^{-5}	2.51×10^{-2}	3.00×10^{-2}
Biol. Process	GO:0030198~extracellular matrix organization	24	1.57	2.67	2.10×10^{-5}	3.22×10^{-2}	3.85×10^{-2}
Biol. Process	GO:0007155~cell adhesion	93	6.08	1.53	3.11×10^{-5}	4.76×10^{-2}	5.69×10^{-2}
Biol. Process	GO:0022610~biological adhesion	93	6.08	1.53	3.30×10^{-5}	5.06×10^{-2}	6.04×10^{-2}
Cell Compart.	GO:0031012~extracellular matrix	69	4.51	2.26	1.49E-10	2.29E-07	2.19E-07
Cell Compart.	GO:0005578~proteinaceous extracellular matrix	65	4.25	2.03	2.63E-10	4.02E-07	3.86E-07
Cell Compart.	GO:0044421~extracellular region part	133	8.69	1.56	1.10E-07	1.68E-04	1.61E-04
Cell Compart.	GO:0044420~extracellular matrix part	30	1.96	2.90	2.16E-07	3.31E-04	3.17E-04
Cell Compart.	GO:0005581~collagen	14	0.92	4.45	5.20E-06	7.95E-03	7.62E-03
Mol. Function	GO:0005507~copper ion binding	19	1.24	3.36	6.69E-06	1.02E-02	1.08E-02
Mol. Function	GO:0005539~glycosaminoglycan binding	29	1.90	2.40	1.99E-05	3.04E-02	3.20E-02
Mol. Function	GO:0005198~structural molecule activity	84	5.49	1.58	2.52E-05	3.85E-02	4.06E-02

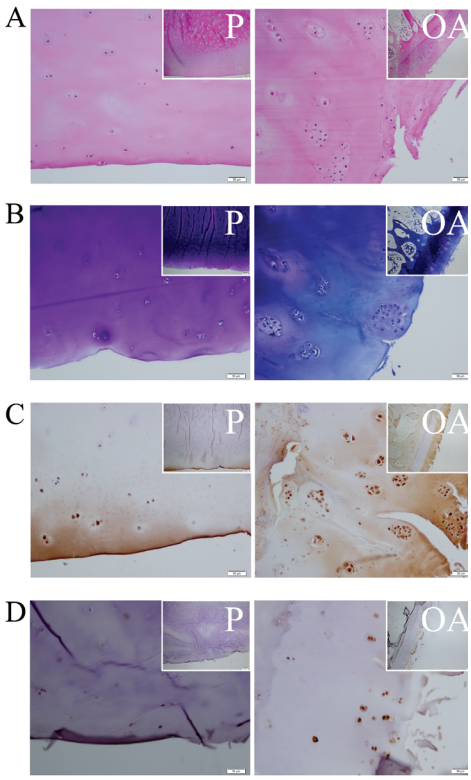


Figure 2. Representative slides of immunohistochemical staining. A) H&E staining. **B**) Toluidine blue staining. **C**) SERPINE1. **D**) CD55 (magnification 20x; insets show larger overview at magnification 4x; white scale bars indicate 50 µm and 200 µm, respectively). The left panels show preserved cartilage area (P) and the right panels show the OA affected cartilage area (OA).

inter-individual or age-related confounding factors as also reflected by the highly comparable differential gene expression patterns when stratifying according to joint and sex. After correction for multiple testing 1717 unique genes showed significant differential expression, of which 19 genes had a fold-change of 2 or higher. In an independent paired cartilage sample set, differential expression was confirmed for 18 genes by RT- qPCR. For most of these genes, except *HBA2*, *HBB*, and *COL9A1*, expression associated with disease severity as determined by scoring according to Mankin [16], and OA-associated increase in protein expression for 2 genes (*CD55* and *SERPINE1*) was demonstrated by immunohistochemistry.

We confirmed several genes previously identified in within-joint analyses for OA affected versus preserved cartilage as well as analyses comparing preserved cartilage from OA affected joints versus healthy cartilage such as the inflammatory genes *CD55* [8], *PTGES* and *TNFAIP6* [11]. This overlap is noteworthy since in our analysis considerably more samples were included. A large sample size increases power to detect replicable findings and allows detection of differences that were previously missed or more subtle. Our data thus indicate that at least a number of genes are consistently involved in the OA disease process despite the appreciated heterogeneous pathophysiology. Another gene present among the top genes and highly up-regulated in OA affected cartilage was the tumor necrosis factor receptor superfamily 11b (*TNFRSF11B*) gene encoding osteoprotegerin. Very recently we reported for this protein on a gain of function mutation likely causal in a family with early onset OA with chondrocalcinosis [29]. In this respect, the up-regulated expression, could contribute to respective mineralization of the cartilage and eventually formation of bone, a major hallmark of the ongoing osteoarthritis disease process.

Studies comparing intact cartilage with OA affected cartilage of the same joint allow detection of gene expression changes specific to the ongoing OA pathophysiological processes independent of confounding

Table 4 Gene enrichment analysis considering the biological processes option (GOTERM_BP_FAT option in DAVID) using the genes with at least 2-fold expression difference between OA affected and preserved cartilage.

Term	Count	Pct	Enr.	p	p_{adj}	FDR
GO:0009611~response to wounding	6	31.58	8.87	2.96E-04	5.63E-03	3.94E-01
GO:0001501~skeletal system development	5	26.32	12.10	4.95E-04	9.40E-03	6.58E-01
GO:0006954~inflammatory response	5	26.32	12.06	5.01E-04	9.51E-03	6.65E-01

factors such as sex and age and joint as was also demonstrated by the highly comparable results of our stratified analysis. Identification of such genes commonly changing during OA independent of joint site or sex could be very useful with respect to drug development. On the other hand, differences identified between the intact cartilage derived from patients undergoing joint replacement surgery and healthy cartilage of independent joints are of a cross-sectional nature and provide information on innate differences among OA patients as well as genes changing during OA. Therefore, genes overlapping among the different studies may be of interest to better understand dynamic changes during onset and ongoing OA. A notable example was the expression of the *COL9A1* gene that was higher in preserved as compared to healthy cartilage (3.6-fold), but was subsequently decreased in the OA affected cartilage (Supplementary Figure S3A). Although we acknowledge the fact that the included 7 healthy cartilage samples had a large age-range, our results are in line with the findings of Karlsson *et al* [8] and Xu *et al* [7] showing increased expression of *COL9A1* in cartilage from patients undergoing joint replacement surgery in comparison to healthy cartilage. This altered direction of effect in ongoing OA may explain the fact that *COL9A1* was found not to be associated with Mankin score and suggests that it is mainly involved in the initial response of the chondrocyte to cartilage damage. Gene enrichment analyses performed with all significant genes showed especially that genes involved in the skeletal development were changed in OA affected as compared to preserved cartilage. Notably, this is in accordance with observations from Xu *et al* [7] who found enrichment of genes involved in skeletal development by comparing healthy cartilage versus cartilage of OA affected joints, suggesting that this is a pathway commonly affected in OA cartilage, both in the initiation phases as well as in ongoing OA. The fact that genes involved in skeletal development (e.g. *FRZB* and *TNFRSF11B*, but also *OSTF1*, *FGFR3*, and *IGFBP3*; Supplementary Table S3A) change during ongoing OA processes confirms the hypothesis that OA chondrocytes lose their maturational arrested phenotype, specific for articular cartilage, towards their end-stage differentiation, resembling growth plate during skeletal development [3].

As reviewed by Barter and Young [30], gene expression differences in OA affected tissues may originate from changes in epigenetic control mechanisms. More recently, a comparison between the methylome of hip OA cartilage with cartilage of non-OA hips indeed showed more than 5000 differentially methylated loci whereas the annotated genes were mainly involved in

pathways related to skeletal development [31] similar to the current and previous transcriptomic analyses [7]. Although direct association between such changes in DNA methylation and respective gene expression remains to be demonstrated, the skeletal developmental processes appear to consistently mark ongoing OA pathophysiology.

Recently, a GWAS for hand OA identified a locus in the aldehyde dehydrogenase 1 family, member A2 (*ALDH1A2*) gene [32]. Expression of *ALDH1A2* was shown to be allele dependent and with decreased expression in OA affected cartilage. Despite this and other recent successes of genome wide association studies [23,27] a variety of the identified signals indicate chromosomal regions without obvious OA candidate genes or regions of high linkage disequilibrium with many relative unknown genes [23,27]. Here, we provide a means of exploring the overall expression and behavior during disease in cartilage. Although OA should be considered a 'whole joint disease' [2] and expression profiles of other OA affected joint tissues such as those performed recently in subchondral bone [33] are highly valuable, expression profiles in OA cartilage could serve as one of the selection criteria to prioritize genes for functional follow-up studies and research directed at understanding pathophysiological mechanisms of OA and drug design. In our cartilage dataset, we found differential expression for several of the genes, among which *PAPPA* was most significant (p=1.1x10^{-6}), positionally localized in close neighborhood of one the arcOGEN genome wide hits: rs4836732 within the *ASTN2* gene. The exact linkage disequilibrium across this locus needs to be further explored. We also found *HBP1,* at the chr7q22 locus, to be differently expressed, although with small effect size in the OA versus preserved comparison (1.1-fold higher in OA affected cartilage). When comparing diseased cartilage (OA affected as well as macroscopically intact cartilage) with healthy cartilage we observed a much stronger and opposite direction of effects: healthy versus OA and healthy versus preserved both showed 1.4-fold lower expression in accordance with a previous study by Raine *et al.* showing increased expression of *HBP1* in OA affected cartilage [34]. Given that *HBP1* resides in the 7q22 gene cluster [23] results mark this gene as most likely candidate for further functional follow-up investigations.

Although *MCF2L* (MCF.2 cell line derived transforming sequence-like), a gene previously identified in GWA as an OA susceptibility gene [21], was not well-detected in the microarray analysis, the significant increased expression of neuronal growth factor (*NGF*) is worth mentioning in this respect. Neurotrophin-3 (NT3), another member of the NGF-family of proteins, enhances migration of premyelinating Schwann cells via Dbs/MCF2L [35], possibly implicating nociception in OA. Interestingly, antibodies generated against NGF or its receptor have been used successfully to treat OA patients and effectively reduced their pain [36]. The fact that *NGF* was not identified previously by comparing healthy with OA affected cartilage [7,8] suggests that *NGF* may be more specific for the "late" OA process. Alternatively, selection of druggable targets from early-responsive genes that start changing before damage is irreversible could be more eligible to effectively slow-down or stop the OA process.

The sample collection is performed by well-trained lab personnel, however, we cannot exclude the possibility of minor contamination with bone tissue. In this respect, it is of note that several cartilage-specific genes (e.g. decorin or *DCN*, collagen type 2 A1 (*COL2A1*), cartilage intermediate layer protein (*CILP*), and cartilage oligomeric matrix protein (*COMP*) were amongst the 100 genes with highest levels of expression in

Table 5 Genes identified in robust genome wide approaches with fold-changes and P-values for OA versus preserved cartilage (OA vs P). (Ref: reference, where indicated gene was published as OA susceptibility gene; FC: fold change; --: not detected on microarray).

			OA vs P	
Gene Symbol	**Ref**	**Joint published**	**FC**	***p***
ASTN2	[27]	Hip&Knee	--	--
BCAP29	[23]	Knee	1.1	$1.7x10^{-2}$
BTNL2	[25]	Knee	--	--
C6ORF130	[26]	Hip&Knee	0.88	$1.6x10^{-3}$
CDC5L	[27]	Hip&Knee	--	--
CHST11	[27]	Hip&Knee	--	--
COG5	[23]	Knee	0.98	$2.1x10^{-1}$
COL11A1	[26]	Hip&Knee	0.94	$4.5x10^{-1}$
DOT1L	[19]	Hip	--	--
DUS4L	[23]	Knee	--	--
DVWA	[24]	Knee	--	--
FILIP1	[27]	Hip&Knee	--	--
FTO	[27]	Hip&Knee	1.0	$8.5x10^{-1}$
GDF5	[20]	Hip&Knee	1.1	$4.3x10^{-2}$
GLT8D1	[27]	Hip&Knee	1.0	$3.0x10^{-1}$
GNL3	[27]	Hip&Knee	1.1	$4.6x10^{-2}$
GPR22	[23]	Knee	--	--
HBP1	[23]	Knee	1.1	$2.0x10^{-3}$
HLA-DQB1	[25]	Knee	--	--
KLHDC5	[27]	Hip&Knee	1.0	$4.8x10^{-1}$
MCF2L	[21]	Hip&Knee	--	--
MICAL3	[26]	Hip&Knee	--	--
NCOA3	[22]	Hip	0.93	$7.9x10^{-3}$
PAPPA	[27]	Hip&Knee	2.1	$1.1x10^{-6}$
PRKAR2B	[23]	Knee	1.0	$9.9x10^{-1}$
PTHLH	[27]	Hip&Knee	1.4	$1.8x10^{-3}$
SENP6	[27]	Hip&Knee	1.1	$3.3x10^{-1}$
SUPT3H	[27]	Hip&Knee	1.0	$6.1x10^{-1}$
TP63	[27]	Hip&Knee	--	--
VEGF	[28]	Hip	1.0	$4.5x10^{-1}$

the dataset while no bone-specific genes (e.g. *COL1A1, COL1A2*, *TNFRSF11B*, and bone sialoprotein II or *IBSP*) were identified here.

In conclusion, our results add to the insight into the ongoing pathological processes in OA cartilage by the identification of different gene expression patterns depending on OA severity as determined by Mankin score. This large scale analysis of joint-matched OA affected and preserved cartilage seems to hint at relatively consistent changes in gene expression during OA development. We think research and development of OA treatment could benefit by focusing on these similarities in gene expression changes and/or pathways.

ACKNOWLEDGMENTS

We thank I. Briaire-de Bruin for expert technical assistance with the immunohistochemistry.
This study was supported by the Leiden University Medical Center, the Dutch Arthritis Association, the Centre of Medical System Biology and the Netherlands Consortium for Healthy Ageing both in the framework of the Netherlands Genomics Initiative (NGI). We also acknowledge support by TreatOA which is funded by the European Commission framework 7 program (PF7/2007) under grant agreement no. 200800, and the European Union's Seventh Framework Program (FP7/2007-2011) under grant agreement n° 259679.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions:
All authors have made substantial contributions to the completion of this study:
Study concept and design: YR, PS, SB, IM.
Acquisition of material and data: YR, WH, NB, SB, RB, NL, CK, RN.
Microarray analysis: SB, WH, JG.
Immunohistochemistry: YR, SB, JB.
Preparation of manuscript: YR, WH, SB, IM.
Critical reviewing and approval of the manuscript: all authors.

Conflict of interest
The authors have declared that no competing interests exist.

Data sharing statement
Gene expression profiles of the samples have been deposited in NCBI's Gene Expression Omnibus [37] and are accessible through GEO Series accession number GSE57218.

REFERENCES

1 Goldring MB, Marcu KB. Cartilage homeostasis in health and rheumatic diseases. *Arthritis Res Ther* 2009;**11**:224.
2 Loeser RF, Goldring SR, Scanzello CR, Goldring MB. Osteoarthritis: a disease of the joint as an organ. *Arthritis Rheum* 2012;**64**:1697-707.
3 Bos SD, Slagboom PE, Meulenbelt I. New insights into osteoarthritis: early developmental features of an ageing-related disease. *Curr Opin Rheumatol* 2008;**20**:553-9.
4 Sandell LJ. Etiology of osteoarthritis: genetics and synovial joint development. *Nat Rev Rheumatol* 2012;**8**:77-89.
5 Gonzalez A. Osteoarthritis year 2013 in review: genetics and genomics. *Osteoarthritis*

Cartilage 2013;**21**:1443-51.

6 Aigner T, Fundel K, Saas J, Gebhard PM, Haag J, Weiss T, *et al.* Large-scale gene expression profiling reveals major pathogenetic pathways of cartilage degeneration in osteoarthritis. *Arthritis Rheum* 2006;**54**:3533-44.

7 Xu Y, Barter MJ, Swan DC, Rankin KS, Rowan AD, Santibanez-Koref M, *et al.* Identification of the pathogenic pathways in osteoarthritic hip cartilage: commonality and discord between hip and knee OA. *Osteoarthritis Cartilage* 2012;**20**:1029-38.

8 Karlsson C, Dehne T, Lindahl A, Brittberg M, Pruss A, Sittinger M, *et al.* Genome-wide expression profiling reveals new candidate genes associated with osteoarthritis. *Osteoarthritis Cartilage* 2010;**18**:581-92.

9 Geyer M, Grassel S, Straub RH, Schett G, Dinser R, Grifka J, *et al.* Differential transcriptome analysis of intraarticular lesional vs intact cartilage reveals new candidate genes in osteoarthritis pathophysiology. *Osteoarthritis Cartilage* 2009;**17**:328-35.

10 Tsuritani K, Takeda J, Sakagami J, Ishii A, Eriksson T, Hara T, *et al.* Cytokine receptor-like factor 1 is highly expressed in damaged human knee osteoarthritic cartilage and involved in osteoarthritis downstream of TGF-beta. *Calcif Tissue Int* 2010;**86**:47-57.

11 Sato T, Konomi K, Yamasaki S, Aratani S, Tsuchimochi K, Yokouchi M, *et al.* Comparative analysis of gene expression profiles in intact and damaged regions of human osteoarthritic cartilage. *Arthritis Rheum* 2006;**54**:808-17.

12 Goeman JJ, van de Geer SA, de KF, van Houwelingen HC. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics* 2004;**20**:93-9.

13 Citri A, Pang ZP, Sudhof TC, Wernig M, Malenka RC. Comprehensive qPCR profiling of gene expression in single neuronal cells. *Nat Protoc* 2012;**7**:118-27.

14 Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001;**25**:402-8.

15 Bos SD, Bovee JV, Duijnisveld BJ, Raine EV, van Dalen WJ, Ramos YF, *et al.* Increased type II deiodinase protein in OA-affected cartilage and allelic imbalance of OA risk polymorphism rs225014 at DIO2 in human OA joint tissues. *Ann Rheum Dis* 2012;**71**:1254-8.

16 Mankin HJ, Dorfman H, Lippiello L, Zarins A. Biochemical and metabolic abnormalities in articular cartilage from osteo-arthritic human hips. II. Correlation of morphology with biochemical and metabolic data. *J Bone Joint Surg Am* 1971;**53**:523-37.

17 Huang dW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;**4**:44-57.

18 Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 2011;**39**:D561-D568.

19 Castano Betancourt MC, Cailotto F, Kerkhof HJ, Cornelis FM, Doherty SA, Hart DJ, *et al.* Genome-wide association and functional studies identify the DOT1L gene to be involved in cartilage thickness and hip osteoarthritis. *Proc Natl Acad Sci U S A* 2012;**109**:8218-23.

20 Chapman K, Takahashi A, Meulenbelt I, Watson C, Rodriguez-Lopez J, Egli R, *et al.* A meta-analysis of European and Asian cohorts reveals a global role of a functional SNP in the 5' UTR of GDF5 with osteoarthritis susceptibility. *Hum Mol Genet* 2008;**17**:1497-504.

21 Day-Williams AG, Southam L, Panoutsopoulou K, Rayner NW, Esko T, Estrada K, *et al.* A Variant in MCF2L Is Associated with Osteoarthritis. *Am J Hum Genet* 2011;**89**:446-50.

22 Evangelou E, Kerkhof HJ, Styrkarsdottir U, Ntzani EE, Bos SD, Esko T, *et al.* A meta-

analysis of genome-wide association studies identifies novel variants associated with osteoarthritis of the hip. *Ann Rheum Dis* 2013.

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

24 Miyamoto Y, Shi D, Nakajima M, Ozaki K, Sudo A, Kotani A, *et al.* Common variants in DVWA on chromosome 3p24.3 are associated with susceptibility to knee osteoarthritis. *Nat Genet* 2008;**40**:994-8.

25 Nakajima M, Takahashi A, Kou I, Rodriguez-Fontenla C, Gomez-Reino JJ, Furuichi T, *et al.* New sequence variants in HLA class II/III region associated with susceptibility to knee osteoarthritis identified by genome-wide association study. *PLoS One* 2010;**5**:e9723.

26 Panoutsopoulou K, Southam L, Elliott KS, Wrayner N, Zhai G, Beazley C, *et al.* Insights into the genetic architecture of osteoarthritis from stage 1 of the arcOGEN study. *Ann Rheum Dis* 2011;**70**:864-7.

27 Zeggini E, Panoutsopoulou K, Southam L, Rayner NW, Day-Williams AG, Lopes MC, *et al.* Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* 2012;**380**:815-23.

28 Rodriguez-Fontenla C, Calaza M, Evangelou E, Valdes AM, Arden N, Blanco FJ, *et al.* Assessment of osteoarthritis candidate genes in a meta-analysis of 9 genome-wide association studies. *Arthritis Rheum* 2013.

29 Ramos YF, Bos SD, van der Breggen R, Kloppenburg M, Ye K, Lameijer EW, *et al.* A gain of function mutation in TNFRSF11B encoding osteoprotegerin causes osteoarthritis with chondrocalcinosis. *Ann Rheum Dis* 2014.

30 Barter MJ, Young DA. Epigenetic mechanisms and non-coding RNAs in osteoarthritis. *Curr Rheumatol Rep* 2013;**15**:353.

31 Rushton MD, Reynard LN, Barter MJ, Refaie R, Rankin KS, Young DA, *et al.* Characterization of the cartilage DNA methylome in knee and hip osteoarthritis. *Arthritis Rheumatol* 2014.

32 Styrkarsdottir U, Thorleifsson G, Helgadottir HT, Bomer N, Metrustry S, Bierma-Zeinstra S, *et al.* Severe osteoarthritis of the hand associates with common variants within the ALDH1A2 gene and with rare variants at 1p31. *Nat Genet* 2014;**46**:498-502.

33 Chou CH, Wu CC, Song IW, Chuang HP, Lu LS, Chang JH, *et al.* Genome-wide expression profiles of subchondral bone in osteoarthritis. *Arthritis Res Ther* 2013;**15**:R190.

34 Raine EV, Wreglesworth N, Dodd AW, Reynard LN, Loughlin J. Gene expression analysis reveals HBP1 as a key target for the osteoarthritis susceptibility locus that maps to chromosome 7q22. *Ann Rheum Dis* 2012;**71**:2020-7.

35 Yamauchi J, Chan JR, Miyamoto Y, Tsujimoto G, Shooter EM. The neurotrophin-3 receptor TrkC directly phosphorylates and activates the nucleotide exchange factor Dbs to enhance Schwann cell migration. *Proc Natl Acad Sci U S A* 2005;**102**:5198-203.

36 Seidel MF, Lane NE. Control of arthritis pain with anti-nerve-growth factor: risk and benefit. *Curr Rheumatol Rep* 2012;**14**:583-8.

37 Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002;**30**:207-10.

Supplementary Table S1. A) Characteristics of donors included in the microarray analyses. **B)** Characteristics of donors included in the replication. **C)** Characteristics of healthy donors included in the microarray analysis (F: females).

A

	DISCOVERY		
	all	**hip**	**knee**
N (% F)	33 (57.6)	22 (68.2)	11 (36.4)
Mean age (range)	66.2 (53.8-79.7)	68.5 (54.6-79.7)	61.5 (53.8-71.4)

B

	REPLICATION		
	all	**hip**	**knee**
N (% F)	28 (78.6)	8 (50.0)	20 (90)
Mean age (range)	66.0 (44.1-82.8)	64.2 (44.1-80.6)	66.7 (46.5-82.8)

C

	HEALTHY		
	all	**hip**	**knee**
N (% F)	7 (71.4)	2 (50)	5 (80)
Mean age (range)	39.9 (13.9-80)	77.8 (75.5-80.0)	24.8 (13.9-62.9)

Supplementary Table S2. Taqman probes used in the fluidigm RT-qPCR experiment.

Gene Symbol	assay
ACTB	Hs99999903_m1
CD55	Hs00892618_m1
CHRDL2	Hs01060234_m1
COL9A1	Hs00932129_m1
CRISPLD1	Hs00230322_m1
CRLF1	Hs00191064_m1
CXCL14	Hs01557413_m1
FN1	Hs00365052_m1
FRZB	Hs00173503_m1
HBA2	Hs00361191_g1
HBB	Hs00758889_s1
LEPREL1	Hs00216998_m1
NGF	Hs00171458_m1
PAPPA	Hs01032307_m1
PTGES	Hs01115610_m1
RARRES2	Hs01123775_m1
SERPINE1	Hs01126606_m1
SPP1	Hs00959010_m1
TNFAIP6	Hs01113602_m1
TNFRSF11B	Hs00900358_m1

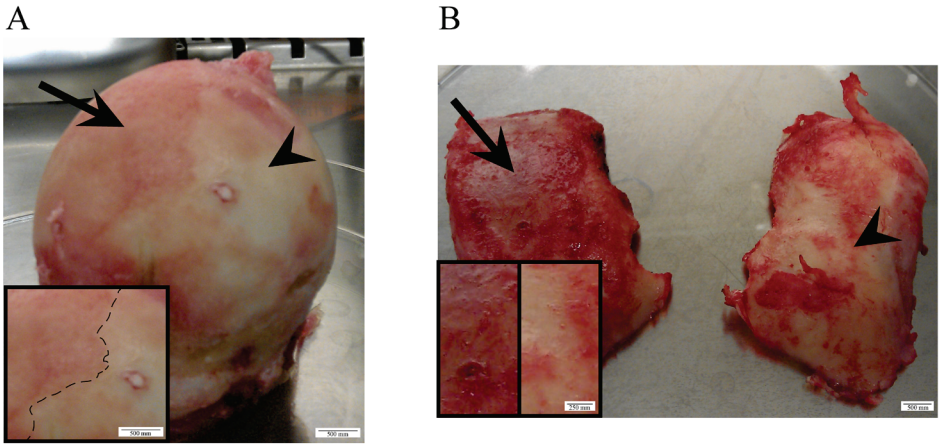
Supplementary Table S3: can be downloaded from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0103056

A) Genes significantly different expressed between OA and preserved cartilage in microarray analysis of 33 paired OA affected and preserved samples (highlighted in yellow the genes that are significantly different in the healthy versus preserved cartilage comparison). **B)** Genes significantly different expressed between preserved and healthy cartilage (FC: fold change; Pval: P-value).

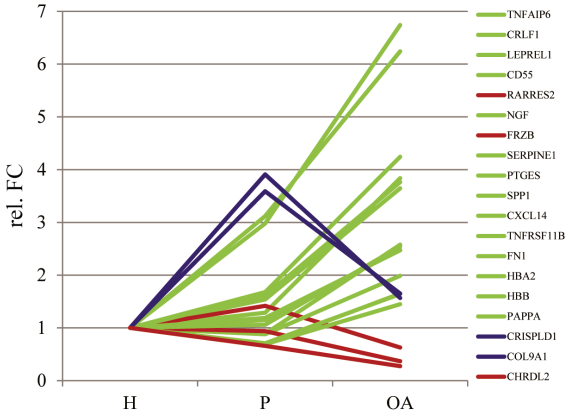
Supplementary Table S4. Results of the validation of the genes with at least 2-fold significant differential expression between OA affected and preserved cartilage in the microarray analyses (Dir: direction of effects; FC: fold change).

	DISCOVERY			VALIDATION		
Gene Symbol	**FC**	**Dir**	***p***	**FC**	**Dir**	***p***
TNFAIP6	2.90	Up	4.40E-08	3.66	Up	3.00E-03
CRLF1	2.97	Up	4.40E-08	2.60	Up	3.10E-02
LEPREL1	2.37	Up	4.40E-08	2.61	Up	2.00E-02
CD55	2.39	Up	4.40E-08	2.29	Up	1.33E-01
RARRES2	2.26	Dn	6.06E-08	3.18	Dn	2.00E-02
NGF	2.26	Up	3.36E-07	3.18	Up	9.20E-02
FRZB	2.54	Dn	1.28E-06	2.59	Dn	5.60E-02
SERPINE1	2.40	Up	1.30E-06	2.16	Up	1.70E-02
PTGES	2.34	Up	1.46E-06	2.50	Up	5.10E-02
SPP1	2.53	Up	3.40E-06	2.30	Up	6.80E-02
CXCL14	2.16	Up	3.54E-06	6.52	Up	4.00E-03
TNFRSF11B	2.10	Up	4.23E-06	1.86	Up	1.25E-01
FN1	2.31	Up	1.13E-05	1.37	Up	3.36E-01
HBA2	2.26	Up	2.02E-05	1.51	Up	4.28E-01
HBB	2.00	Up	6.97E-05	1.69	Up	3.29E-01
PAPPA	2.08	Up	7.69E-05	3.80	Up	5.60E-02
CRISPLD1	2.50	Dn	1.42E-04	1.31	Dn	3.62E-01
COL9A1	2.18	Dn	2.67E-04	2.35	Dn	9.50E-02
CHRDL2	2.39	Dn	3.55E-04	3.26	Dn	2.03E-01

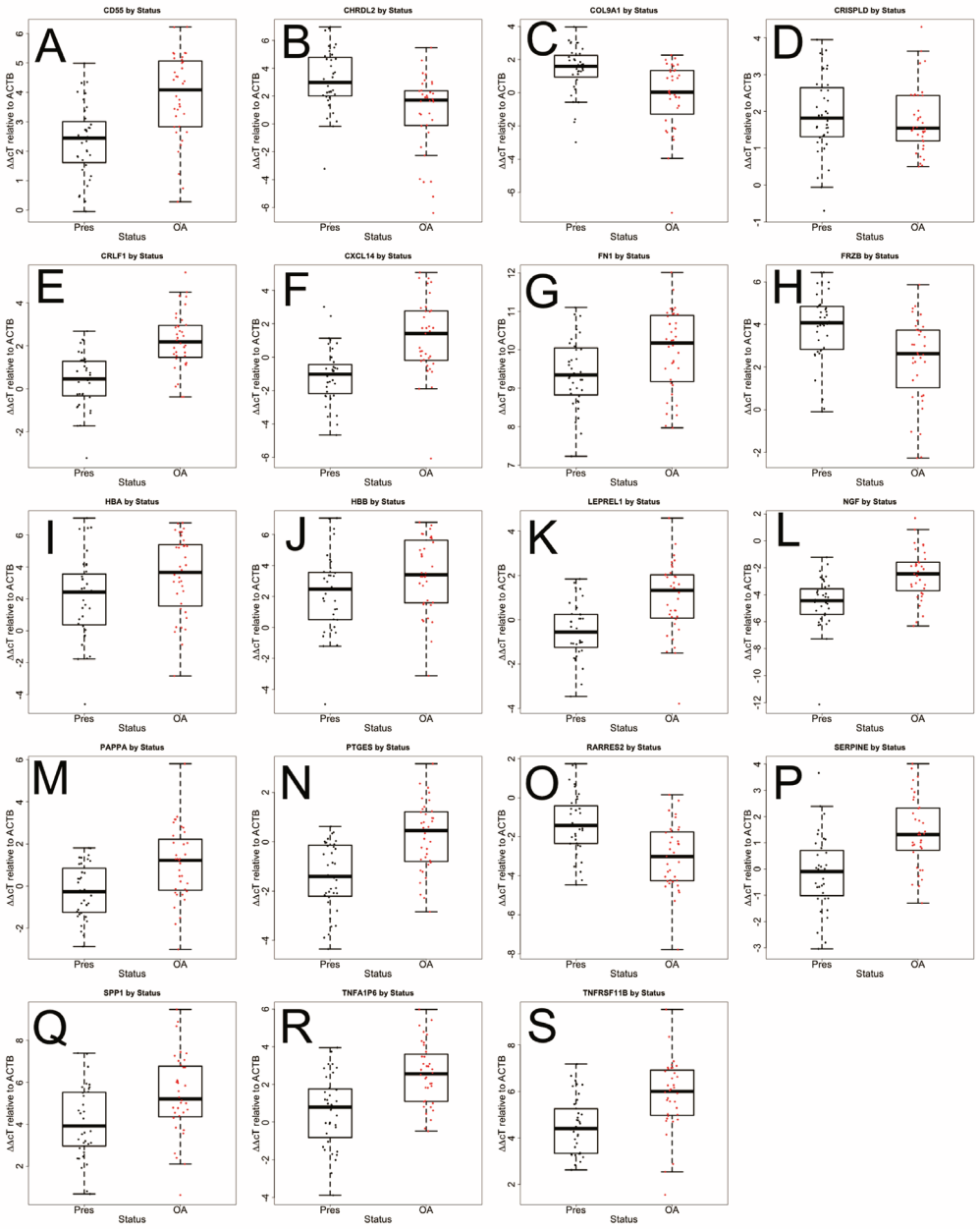
Supplementary Figure S1. Typical example of hip **(A)** and knee **(B)** joint with areas of macroscopically preserved (arrow head) and OA affected cartilage (arrow; white scale bars indicate 500 mm). Insets show detail of preserved (right) and OA affected area (left), in A separated by a dashed line (scale bar inset in B: 10 mm).



Supplementary Figure S3. Relative changes in gene expression levels in preserved and OA affected cartilage relative to healthy cartilage for the 19 genes with at least 2-fold difference in the OA versus preserved analysis (note that the line does not imply continues changes given the fact that the healthy cartilage was derived from independent donors).



Supplementary Figure S2. Individual box plots per status for genes validated by RT-qPCR.



3

Genes expressed in blood link osteoarthritis with apoptotic pathways

Ann Rheum Dis 2014;73:1844-1853 (doi:10.1136/annrheumdis-2013-203405)

Genes expressed in blood link osteoarthritis with apoptotic pathways

Yolande F. M. Ramos[1*], Steffan D. Bos[1,2], Nico Lakenberg[1], Stefan Böhringer[3], Wouter den Hollander[1], Margreet Kloppenburg[4], P. Eline Slagboom[1,2], Ingrid Meulenbelt[1,2]

[1]Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; [2]The Netherlands Genomics Initiative, sponsored by the NCHA, Leiden-Rotterdam, The Netherlands; [3]Department of Biostatistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands; [4]Depts. of Clinical Epidemiology and Rheumatology, Leiden University Medical Center, Leiden, The Netherlands

ABSTRACT

Objective **To identify novel gene expression networks in blood of osteoarthritis (OA) patients as compared to controls.**

Methods **A comprehensive exploration of gene expression in peripheral blood was performed by microarray analysis (Illumina HT12_v3_BeadChips) for a subset of OA patients from the GARP study in comparison with sex- and age-matched healthy controls. To identify pathways, we performed gene enrichment analyses (DAVID and STRING). Quantitative PCR analysis in overlapping and in additional OA samples was performed for prioritized genes to validate and replicate findings. Classification of cases and controls was explored by applying statistical models.**

Results **A total of 741 probes representing 694 unique genes were differentially expression between cases and controls, including 86 genes expressed with at least 1.5-fold difference. *ATF4*, *GPR18* and *H3F3B* were among the top genes identified ($p<4.5x10^{-8}$). We found that in blood of OA patients the apoptosis pathway, including the well-known gene *CASP3*, was significantly enriched among the differentially expressed genes. Our findings were validated in independent samples and when using a small subset of the validated genes, we could accurately distinguish patients from controls (AUC=98%).**

Conclusions **In the current study, we have identified specific gene expression networks, in the easily accessible tissue blood, that associated consistently to OA among GARP study cases. Our data further hint at the relevance of apoptosis as etiological factor in OA onset, thereby qualifying expression profiling of blood as useful tool to understand underlying molecular mechanisms of OA.**

Keywords: osteoarthritis, microarray, apoptosis, biomarkers

INTRODUCTION

Osteoarthritis (OA) is a prevalent, complex, disabling disease affecting the articular joints. It has a considerable genetic component and the architecture is of a polygenic nature.[1] Currently, adequate therapies to reverse or slow down the disease apart from pain relieve are very poor[2] and joint replacement surgery at the end stage is commonly applied. Diagnosis of OA occurs not before changes in the joint tissues are visible on radiographs. Such changes, however, reflect irreversible damage. Sensitive markers to monitor early changes or changes in active disease are not yet available and little is known about the early molecular processes marking OA onset. Such insights however are important to prevent irreversible damage and are considered crucial for the development of effective disease modifying treatments.

Initial genetic studies have resulted in a few consistent OA susceptibility genes, such as *GDF5*[3], *FRZB*[4], *SMAD3*[5], and *DIO2*[6]. Together, these genes denominated skeletogenesis as a common pathway that may be involved in OA susceptibility thereby contributing to the hypothesis that OA may have an early developmental origin that is expressed later in life.[7, 8] More recently, large consortia (such as TREAT~OA and arcOGEN) started compiling genome wide association (GWA) data on OA to perform powerful meta-analyses. These efforts have contributed to the discovery of a variety of genome wide significant signals in OA, some of which contained compelling OA candidate genes[9, 10] whereas other loci were found in gene deserts or in chromosomal regions with high linkage disequilibrium containing many genes without obvious candidates. These results point towards a typical problem in the genetic mapping of complex traits and expose the gap between genetic evidence and the underlying molecular mechanism of the disease. In addition, functional effects of genetic variation with a vast influence on many different biological processes but occurring at unknown (trans-acting) regulatory elements remain undetected.

Extensive large scale exploration of gene expression networks may be more easily established by comparing well selected OA cases and controls in a readily accessible tissue, such as blood. Moreover, it could be advocated that such expression profiles in blood reflect genetic background predisposing to OA which could be undetectable in the affected cartilage tissue itself as a result of ongoing disease. Expression profiles in blood have successfully been used as molecular biomarker for classification of disease subgroups in other complex diseases.[11, 12]

We set out to identify gene expression networks associated with OA by comparing single gene as well as pathway based blood expression profiles of 106 patients of the Genetics osteoARthritis and Progression (GARP) study, consisting of sibling pairs with symptomatic and radiographic OA in at least two joints, to those of 33 controls. Subsequently, we tested whether these specific OA signatures in blood reflected ongoing disease processes in articular cartilage or could discriminate patients from controls. Finally, we questioned whether the distinctive gene expression signature of blood could provide clues in the prioritization of OA susceptibility loci detected by large scale GWA meta analysis.

MATERIALS AND METHODS

Study populations. The GARP study consists of 191 sibling pairs (N=382) of white, Dutch ancestry. All participants (age-range: 40-70 years; mean age: 60 years) are clinically and radiographically diagnosed with primary, symptomatic OA (as defined by the American College of Rheumatology, ACR) at multiple joint sites in the hand, or in at least two joints of the following locations: hand, spine (cervical or lumbar), knee, or hip.[13] In this study, a subset of 106 (from which 68 are unrelated individuals) participants of the GARP study was selected for gene expression analysis ('Discovery Cohort'). OA patients were compared with age- and sex-matched healthy controls without symptomatic and clinical OA according to ACR criteria. In addition, similar to participants of the GARP study, controls were subjected to radiographic examination at the indicated joints, and were free of signs of OA (KL<1) at the indicated joints except for 5 with KL≤2 at a single joint (Table 1).

Technical validation of the microarray analysis was performed in a random selection of 28 of the 106 previous GARP participants ('Validation Cohort') in comparison to 25 of the controls, and validated genes were replicated in an additional set of 36 unrelated individuals of the GARP study ('Replication Cohort') in comparison to the same 25 controls supplemented with 5 additional controls (characteristics of the study participants are shown in Table 1).

For the comparison of our blood expression profiles with gene expression in cartilage we used the ongoing Research Arthritis and Articular Cartilage (RAAK) study, of which gene expression data is available for cartilage of 33 donors that underwent joint replacement surgery.[14]

RNA isolation. Blood was collected and peripheral blood mononuclear cells (PBMCs) were separated with Histopaque® gradient. Cells were lyzed in 1 ml of TRIzol reagent (Life Technologies, Bleiswijk, The Netherlands), and stored at -80ºC until isolation. After thawing, 200µl of chloroform was added before centrifugation (15 min at 16.000g). The aqueous upper layer was separated, 1 v/v 70% ethanol/DEPC-treated water was added, and total RNA was isolated with the RNeasy mini kit (Qiagen, Venlo, The Netherlands) following the manufacturer's recommendations. To assess the RNA quality, 36 random samples were analyzed on the 2100 Bioanalyzer (Agilent Technologies, Amstelveen, The Netherlands) which had an RNA integrity number (RIN) of at least 8.3 and an average 28S/18S ratio of 2.2 (range: 1.8-2.7). The RNA concentration was measured using a Nanodrop spectrophotometer.

Microarray analysis. For first- and second-strand reverse transcription 500ng of total RNA served as a template. To perform cDNA synthesis, amplification, biotin labeling, and hybridization onto the microarrays the Ambion® TotalPrep™-96 RNA amplification kit (Life Technologies, Bleiswijk, The Netherlands) was used according to the manufacturer's protocol. After hybridization on Illumina Human HT-12_v3_BeadChips (Illumina, Eindhoven, The Netherlands) slides were scanned with Illumina Beadscanner 500GX.

Raw data from the beadscanner were prepared with GenomeStudio Software

(Gene Expression Module v1.0 from Illumina), without normalization or background substraction and with imputation of missing data. Quality control was performed in R statistical programming language using the lumi-package,[15] and 2 out of 108 GARP samples analyzed on the Illumina BeadChips were excluded from the analysis based on their aberrant positions in the boxplot, the density plot, and the sample-relation analysis. Subsequently, high quality data of 106 cases and 33 controls were prepared for analysis by applying robust spline normalization (rsn) and variance-stabilizing transformation (vst). Differential expression between GARP-participants and healthy controls was analyzed using the limma-package for linear regression analysis,[16] correcting for sex, age, body mass index (BMI) and chip (to adjust for possible batch effects). The "BH" (Benjamini and Hochberg) option was used to correct for multiple testing. Probes with detection P-values higher than 0.05 in more than 50% of the samples were not included in subsequent analysis.

Pathway analysis. Gene enrichment was studied using either all significant differentially expressed genes (N=694 of which 619 were covered by the functional annotation algorithms in DAVID) or the 86 genes (of which 76 were covered) showing an expression difference of ≥1.5 fold between OA patients and healthy controls. Pathway analysis was performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID) tool,[17, 18] selecting for biological processes identified in the Gene Ontology database (GOTERM_BP_FAT in the options menu implemented in DAVID) and for the microarray background (HumanHT-12_V3_0_R2_11283641_A). Genes with $p \leq 0.05$ after correction for multiple testing according to Bonferroni were considered significant.

Analysis of protein interaction networks. To investigate protein interactions among the genes identified in our microarray, we used the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) 9.0[19] available online (data are thus analyzed from source 'previous knowledge' which means it is derived from text-mining and therefore not all data are determined experimentally). With the option 'Enrichment' we analyzed for enrichment of protein-protein interactions, and also 'GO Biological Processes' were analyzed with STRING.

Quantitative RT-PCR analysis. Quantitative reverse transcription PCR (RT-qPCR) was used to validate and replicate the results of the microarray analysis. Expression levels of 16 genes selected based on P-value and fold difference, and 14 genes involved in apoptosis, were determined in the 'Validation Cohort' (N_{GARP}=28) and the 'Replication Cohort' (N_{GARP}=36) with gene-specific and pre-validated Taqman® assays (Life Technologies, Bleiswijk, The Netherlands; Taqman assays are shown in Supplementary Table S1). Synthesis of cDNA using 1µg RNA was performed with the First Strand cDNA Synthesis Kit (Roche Applied Science, Almere, The Netherlands) following the manufacturer's protocol. The cDNA was amplified using the DNA Engine Tetrad® 2 Peltier Thermal Cycler (Bio-Rad, Veenendaal, The Netherlands) under standard PCR conditions and qPCR was performed in

triplicate with the Taqman® method using BiomarkTM 96.96 Dynamic Arrays (Fluidigm). Relative gene expression was calculated using the average of the absolute threshold cycle values (Ct values) in the $2^{-\Delta\Delta Ct}$ method,[20] and levels were normalized using the mean expression of two housekeeping genes (*YKT6* and *ACTB*). Commercially available human total reference RNA (Clontech Laboratories, Mountain View, CA, USA) served as reference RNA.

Statistical analysis of RT-qPCR data. Gene expression data were analyzed for normal distribution and, if necessary, log transformed before analysis in R statistical programming language using a generalized linear mixed model. P-values ≤0.05 were considered statistically significant. For technical validation, Spearman correlation between expression levels of the RT-qPCR and expression levels as determined using the microarrays were calculated with SPSS.

Classification models. Classification models were constructed using multiple, penalized logistic regression.[21] Models were evaluated by a receiver operating characteristic (ROC) analysis using the Z-scores of the expression for all genes and an area under the curve (AUC) criterion. Optimism in this procedure was controlled by applying leave-one-out cross-validation.

Analysis strategy. Figure 1 shows a schematic overview of the study strategy. Following microarray analysis, data were statistically analyzed. Genes significantly different between patients and controls were subjected to pathway analysis with the DAVID tool. A selection of 30 genes (16 selected based on P-value and fold difference and 14 selected from the apoptosis pathway) was used for technical validation of the results in the 'Validation Cohort', and 27 genes passing this control were replicated in a group of independent participants of the

Table 1. General information of the participants included in the analysis. **A)** Characteristics of samples included in the microarray analysis ('Discovery Cohort'). **B)** Characteristics of samples included in the validation with RT-qPCR ('Replication Cohort'). (Legend: F: females; M:males; ROA: radiographic OA; Pain: percentage of cases with pain according to the WOMAC-pain score; Pain>20: WOMAC pain-scores higher than 20; NSAID: percentage of cases taking NSAIDs; age given in years).

A	**controls**	**cases**
N	33	106
F /M	24/9	105/1
(% F)	(72.7)	(99.1)
Mean Age	58.1	60.9
(range)	(44.4-75.6)	(43.2-79.4)
Mean BMI	24.4	27.2
(range)	(12.7-32.8)	(20.0-46.5)
ROA Hip (%)	3.0	42.5
ROA Knee (%)	3.0	61.3
ROA Hip&Knee (%)	0.0	22.6
Pain/Pain>20 (%)	0/0	91.5/54.7
NSAID (%)	0.0	51.9

B	**controls**	**cases**
N	30	36
F /M	26/4	26/10
(% F)	(86.7)	(72.2)
Mean Age	58.1	61.0
(range)	(46.0-75.6)	(49.0-75.0)
Mean BMI	24.5	28.48
(range)	(20.0-31.9)	(20.2-46.5)
ROA Hip (%)	3.0	27.8
ROA Knee (%)	0.0	88.9
ROA Hip&Knee(%)	0.0	25.0
Pain/Pain>20 (%)	0/0	100/61.1
NSAID (%)	0.0	61.1

GARP study (the 'Replication Cohort'). In addition, using a classification model expression of the 27 genes was used for OA prediction in the 'Replication Cohort'.

Gene expression data available from the microarray analysis were used as a reference, to search for OA susceptibility genes known from the literature on the one hand, and to look up the significant genes in a dataset of cartilage expression profiles available in our laboratory on the other hand.[14]

RESULTS

Gene expression profiles of PBMCs, accessible at GEO (Gene Expression Omnibus) public repository under the accession GSE48556, were generated from a subset of 106 participants of the GARP study (mean age: 61 years and mean BMI: 27 kg/m^2) and 33 sex- and age -matched healthy controls (mean age: 58 years and mean BMI: 24 kg/m^2; see Table 1A). For an overview of the steps taken in the analysis of the data, see Figure 1.

Expression profiles of OA patients. In the microarray analysis we identified 741 probes comprising 15 non-coding RNA's and 679 unique genes (N=694 total) that were expressed significantly different between OA patients and healthy controls (Supplementary Table S2). From these, 89 probes (86 genes) differed at least 1.5-fold between patients and healthy controls (Table 2; for gene definitions see Table 4 or Supplementary Table S2) of which 46 were expressed higher in blood of

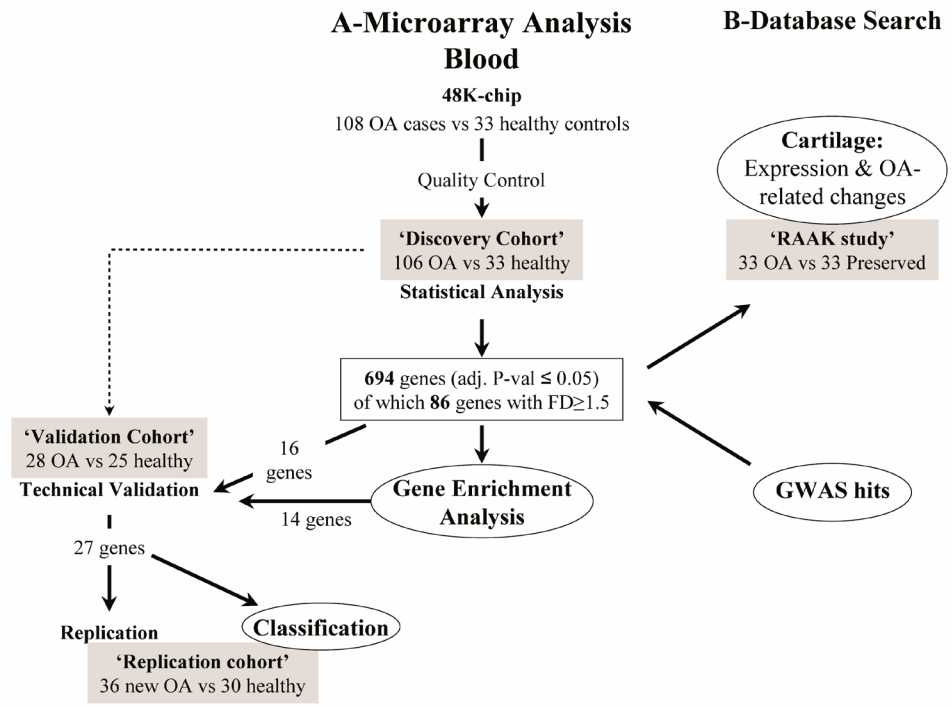


Figure 1. Flowchart of the gene expression analysis.

OA patients (52%). The lowest P-value was found for ubiquitin conjugating enzyme E2L 3 (*UBE2L3*; p=9.26x10^{-12}), followed by activating transcription factor 4 (*ATF4*; p=4.14x10^{-8}). Gene expression differences range from 2.6-fold higher as detected for G protein-coupled receptor 18 (*GPR18*; p=4.31x10^{-8}) to 5.3-fold lower in OA patients for early growth response 1 (*EGR1*; p=2.13x10^{-2}).

Expression profiles of apoptosis-related genes significantly different in OA patients. To gain more insight in the pathways most predominantly involved in the differential expression, gene enrichment analysis was performed for all biological processes using DAVID. When analyzing all significant genes including those with relatively small differences (N=694), we identified 8 GO-terms comprising 5 different biological process that were significantly enriched: protein transport and localization (GO:0015031, GO:0045184, GO:0008104, GO:0046907), macromolecular complex subunit organization (GO:0034621), cell cycle (GO:0007049), RNA processing (GO:0006396), and apoptosis (GO:0006917; Table 3A).

Gene enrichment analysis for the 86 genes differing at least 1.5-fold (Table 2), showed significant enrichment specifically for networks involved in apoptosis (Table 3B). The majority of the genes involved in cell death (13 out of the 86 genes tested according to the DAVID analysis) showed higher expression in OA patients. Among these genes, we identified increased expression (1.6-fold) of the well-known apoptosis-inducing gene Caspase 3 (*CASP3*; p=8.94x10^{-4}; Table 2).

Using STRING, significantly more protein-protein interactions were detected between the 86 genes than expected based on chance (56 interactions total; p=1.05x10^{-5}). Also the apoptosis pathway was found in STRING (p=5.49x10^{-2}), which was visualized by virtue of the multiple interactions between proteins involved in this pathway (Figure 2). As shown, 6 out of the 13 proteins with 3 or more interaction-partners are involved in apoptosis. Apparent is also a cluster of histone-proteins (HIST1H3H, HIST1H2AC, and HIST2H2BE) involved in the above mentioned pathway 'macromolecular complex subunit organization'.

Validation and replication. To validate our results RT-qPCR was carried out. Genes were selected based on P-value and fold difference (N=16), or based on a possible role in apoptosis (N=14). Technical validation was performed in the 'Validation Cohort' (53 samples previously included in the microarray analyses). Here, 27 out of the 30 genes tested showed similar effect sizes and direction as the original data (Supplementary Table S3). Next, we tested these 27 genes in the 'Replication Cohort' (36 independent individuals of the GARP study as compared to 30 controls). In the unrelated GARP patients, 22 out of the 27 genes tested (82%) were significantly different (Table 4). Note that, except for the lunatic fringe gene (*LFNG*, indicated with a red asterisk), the 4 genes that were not significant in the replication (indicated with a black asterisk) still had fold differences comparable to those as measured in the microarray analysis.

Classification capability of differentially expressed genes. Figure 3A shows hierarchical clustering based on the correlation between the differentially

expressed genes across the samples. Separation among OA patients and controls was quite good (more than 80%), although some patients were clustered among the healthy controls and vice versa.

To investigate the possibility to use the set of validated genes (depicted in Table 4) eventually as a diagnostic tool, a classification model was constructed and tested, first in our 'Discovery Cohort', and subsequently in the 'Replication Cohort'. As shown in Figure 3B and 3C, ROC curves of the baseline model containing sex, age, and BMI resulted in a fair discriminative capacity with an area under the curve (AUC) of 73% and 72%, respectively. However, when expression levels of the genes were added to the model the AUC increased to 95% and 97%, respectively, indicating an excellent classification of OA patients as compared to controls. According to our classification model expression of the serine/arginine-rich splicing factor 5 (*SFRS5*) gene contributed most in this model (data not shown). The possibility was examined to reduce our model to a smaller set of genes with which we could still obtain good classification between patients and controls, and we found a combination of 4 genes (*SFRS5*, together with *CASP3*, cyclin-dependent kinase inhibitor 2D (*CDKN2D*), and RAB GTPase-binding effector protein 1 (*RABEP1*) with which we obtained ROC curves with excellent AUCs as high as 98% in the 'Replication Cohort' (Figure 3C, dashed line).

Dataset searches for genes differentially expressed in blood of OA patients. Differential gene expression profiles in blood of OA patients as compared to controls may either reflect different setpoints of innate systemic processes taking place in OA patients and/or it may indicate that blood could reflect pathological processes occuring specifically in joint tissues. To look into these possibilities we set out to assess the cartilage characteristics of the genes we identified as being differentially expressed in blood and we investigated both the levels of expression and their respective differential expression in cartilage of another study sample, namely the RAAK study. Within the RAAK study, a cartilage gene expression dataset is available in our laboratory.[14] The median relative gene expression level of these cartilage samples is 7.1 (range: 6.6-14.9) as measured on the microarrays. We found that 24 out of the 86 genes expressed with at least 1.5-fold difference in blood (depicted in Table 2), also showed nominal significant changes when comparing cartilage from macroscopically unaffected (preserved) with the corresponding OA affected sites. The median relative expression level of these genes was higher than the overall median level (8.6, with range: 7.2-11.5; data not shown). Additional selection for genes with expression changes in the same direction in cartilage and in blood resulted in 12 genes with fold changes ranging between 1.2 up- and 1.3 down in cartilage (Supplementary Table S4). It is of note that among these genes we found

Table 2 (next page). List of genes significantly different expressed (≥1.5-fold difference). *p*: P-value after Bonferroni correction for multiple testing; Fold Diff: fold difference of individual probes in which values below 1 indicate a decreased expression in OA patients compared to healthy controls;
Fnc.: function of the respective gene; TF= transcription factor, E= enzyme, R= receptor, H= histone protein, C= cytokine, O= other.

Gene Symbol	Accession nr	p	Fold Diff.	Fnc.
INCREASED EXPRESSION				
GPR18	NM_001098200.1	4.31E-08	2.62	R
ZNF564	NM_144976.2	5.22E-08	1.59	TF
LFNG	NM_001040167.1	1.02E-07	1.79	E
ADRB2	NM_000024.4	4.27E-07	2.23	R
C16ORF72	NM_014117.2	4.27E-07	1.70	O
PELO	NM_015946.4	1.75E-06	1.65	E
SLAH1	NM_001006610.1	2.64E-06	1.56	E
HCP5	NM_006674.2	2.78E-06	1.76	O
SLA	NM_001045557.1	1.90E-05	1.68	O
LCMT2	NM_014793.3	1.90E-05	1.52	E
PLAGL2	NM_002657.2	2.18E-05	1.59	TF
FAM43A	NM_153690.4	2.26E-05	1.92	O
STK39	NM_013233.2	5.81E-05	1.56	E
USP24	NM_015306.1	1.49E-04	1.67	E
BCDIN3D	NM_181708.1	1.69E-04	1.51	E
TINF2	NM_012461.1	1.83E-04	1.54	O
PLEKHF1	NM_024310.4	2.67E-04	1.78	E
AP3M1	NM_207012.2	2.79E-04	1.53	O
TM2D3	NM_078474.2	3.11E-04	1.53	O
CX3CR1	NM_001337.3	4.18E-04	1.71	R
ZNF217	NM_006526.2	5.83E-04	1.53	TF
PVRIG	NM_024070.3	6.48E-04	1.53	R
PPM1B	NM_177968.2	7.02E-04	1.68	E
CASP3	NM_032991.2	8.94E-04	1.59	E
SLA	NM_006748.1	1.19E-03	1.65	O
NFE2L2	NM_006164.2	1.40E-03	1.51	TF
ID3	NM_002167.2	1.48E-03	2.04	O
RABEP1	NM_001083585.1	1.93E-03	1.75	E
KCNJ2	NM_000891.2	2.01E-03	1.56	O
RSBN1	NM_018364.3	4.98E-03	1.61	O
CLK1	NM_004071.2	5.34E-03	2.26	E
RNMT	NM_003799.1	5.54E-03	1.59	E
P2RY13	NM_023914.2	5.54E-03	1.53	R
EBI2	NM_004951.3	5.72E-03	1.66	R
BIRC3	NM_001165.3	8.02E-03	2.01	E
EDG1	NM_001400.3	8.83E-03	1.52	R
PTPN4	NM_002830.2	1.13E-02	1.60	E
HSPA1B	NM_005346.3	1.34E-02	1.68	O
RAB22A	NM_020673.2	1.43E-02	1.56	E
ALS2CR13	NM_173511.2	1.61E-02	1.52	O
PSMC6	NM_002806.2	2.39E-02	1.52	E
FRYL	NM_015030.1	2.53E-02	1.51	O
CSE1L	NM_001316.2	2.68E-02	1.50	R
IL1B	NM_000576.2	2.84E-02	1.76	C
TGFBR3	NM_003243.2	3.41E-02	1.66	R
EBI2	NM_004951.3	3.74E-02	1.63	R

Gene Symbol	Accession nr	p	Fold Diff.	Fnc.
DECREASED EXPRESSION				
UBE2L3	NM_198157.1	9.26E-12	1.82	E
ATF4	NM_001675.2	4.14E-08	2.07	TF
H3F3B	NM_005324.3	4.31E-08	2.19	H
SNORD68	NR_002450.1	1.54E-07	1.76	O
NFKBIA	NM_020529.1	6.02E-06	1.75	TF
IRF2BP2	NM_182972.2	7.74E-06	1.63	O
DAB2	NM_001343.1	2.16E-05	1.68	O
SFRS5	NM_006925.3	2.80E-05	1.87	O
STK17B	NM_004226.2	3.12E-05	1.63	E
CDKN2D	NM_079421.2	4.64E-05	1.62	E
SFRS17A	NM_005088.2	2.29E-04	1.55	O
SFRS5	NM_006925.3	4.61E-04	1.70	O
LOC338799	NR_002809.1	5.02E-04	1.70	O
ABCA7	NM_019112.3	5.14E-04	1.52	O
MAF	NM_005360.3	8.94E-04	1.66	TF
HS.333764	BG194895	1.18E-03	2.51	O
FLJ45337	NM_207465.1	1.29E-03	1.76	O
HS.578214	DA957648	1.45E-03	1.63	O
HIST2H2BE	NM_003528.2	2.38E-03	1.80	H
MAX	NM_145113.1	2.40E-03	1.52	TF
USP36	NM_025090.2	4.60E-03	1.54	E
LOC644422	XR_019449.1	4.60E-03	1.67	O
LEPROT	NM_017526.2	4.80E-03	1.59	O
SNORD13	NR_003041.1	5.09E-03	1.96	O
PPP1R15A	NM_014330.2	5.12E-03	1.65	E
NBPF10	NM_001039703.1	5.39E-03	1.81	O
ACTN1	NM_001102.2	5.54E-03	1.55	O
HS.534061	CR612552	5.54E-03	1.63	O
NBPF20	NM_001037675.1	5.72E-03	1.93	O
OSM	NM_020530.3	6.78E-03	1.54	O
HIST1H3H	NM_003536.2	7.42E-03	1.96	H
HIST1H2AC	NM_003512.3	8.64E-03	1.78	H
AXUD1	NM_033027.2	8.96E-03	1.51	TF

(continued 'genes with decreased expression')

VCL	NM_003373.3	1.39E-02	1.59	O
EGR1	NM_001964.2	2.13E-02	5.35	TF
LIMS1	NM_004987.3	2.19E-02	1.77	O
TNNC2	NM_003279.2	2.60E-02	1.56	O
IL8	NM_000584.2	2.98E-02	3.07	C
MGC13057	NM_032321.1	3.84E-02	1.64	O
ATHL1	NM_025092.3	3.98E-02	1.81	E
MMD	NM_012329.2	4.19E-02	1.55	O
MPL	NM_005373.1	4.19E-02	1.78	R
C7ORF41	NM_152793.2	4.97E-02	1.68	O

SFRS5, the gene that contributed most to our prediction model in blood of OA patients. *SFRS5* is highly expressed in cartilage (a relative expression of 11.5 is in the highest quartile).

DISCUSSION

In the current paper, we have established a distinctive gene expression profile of 694 genes in blood of OA patients of the GARP study. Of these, 86 appeared to be expressed with at least 1.5-fold difference. The most significant gene, which was validated and replicated in unrelated samples, was *ATF4* (2-fold lower; p=4.14x10^{-8}). Moreover, of note is the *H3F3B* gene that was significantly down-regulated in blood (2.17-fold; p=4.31x10^{-8}) and in articular cartilage (1.15-fold; p=2.05x10^{-3}). Subsequent pathway analyses revealed that genes involved in apoptosis were significantly enriched in our expression profile and included the renowned, OA relevant, apoptosis-inducing gene *CASP3* (p=8.94x10^{-4}). In addition, protein-protein interactions appeared to be condensed around apoptosis-related genes and *CASP3* for example was found to interact with the protein products of 6 other genes identified in the microarray analysis. The microarray platform was validated for 27 out of 30 genes tested, providing high confidence in the data. When using these 27 genes in a classification model a highly sensitive distinction between patients and healthy controls (AUC=97%) was reached. Finally, of the genes identified in our analysis, we found that 12 were also differentially expressed in cartilage and with the same effect directions. Expression of these genes in cartilage was higher than average, and especially *SFRS5*, an RNA splice factor contributing significantly to our classification model, was abundantly expressed in cartilage. It may not be surprising that we found relatively small overlap in the differentially expressed genes between OA cartilage and mononuclear blood cells of OA patients since gene expression in PBMCs can be the result of several processes. As mentioned by Attur *et al.*[22], blood cells passing through the different tissues of the joint may be activated by the OA process resulting in an 'inflammatory profile', as found by Attur *et al.* in a subset of symptomatic knee OA patients, or, as found here, in an 'apoptotic profile'. In most instances changes are subtle and we do not expect them to lead to a general pro-inflammatory state or to direct apoptosis of the blood cells of the patients. However, this remains to be investigated. Alternatively, differential gene expression in PBMCs may reflect genetically determined dysregulation of processes.

Our top gene *ATF4* encodes a potent transcription factor that specifically regulates

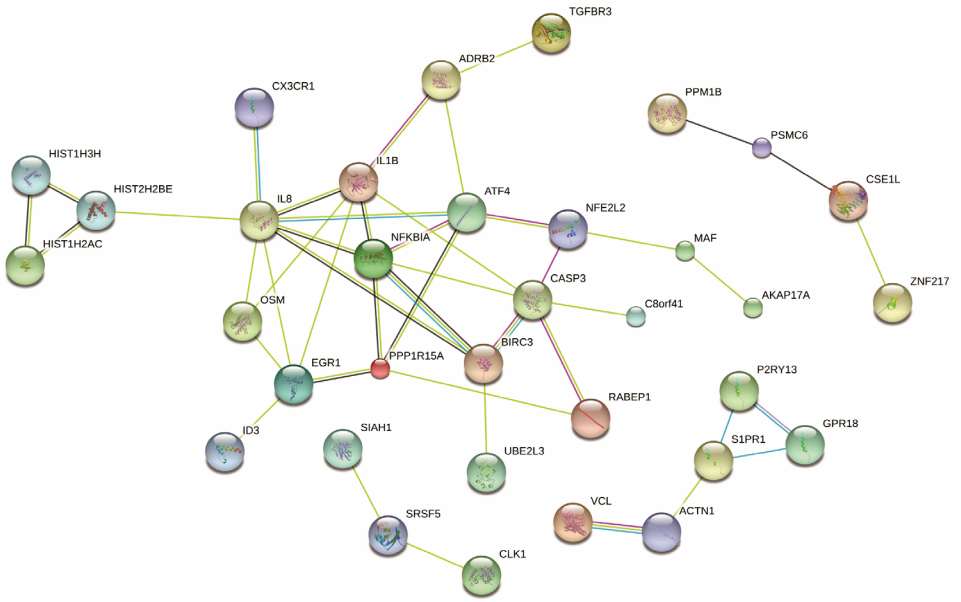


Figure 2. Gene networks in STRING. Screenshot of the protein-protein interactions determined with STRING for the 89 genes differing at least 1.5-fold between patients and controls (disconnected proteins are hided).

Indian hedgehog (*IHH)* expression during endochondral ossification.[23] Since we could not detect significant differential expression of *ATF4* in our and other published articular cartilage datasets,[24-27] the difference in expression of *ATF4* in blood might be a reflection of distinct innate characteristic of OA patients as compared to controls. In growth plate cartilage, where IHH ensures growth plate chondrocytes to proliferate, IHH is considered 'a master regulator in the process of chondrogenesis and osteogenesis'.[28] Given the fact that *ATF4* is significantly lower expressed in blood of OA patients, we hypothesize that, during endochondral ossification, the *IHH* levels may not optimally be assured resulting in suboptimal chondrocyte proliferative potential. Furthermore, these results support the hypotheses that late-onset OA could have a developmental origin.[7]

By applying pathway analysis (Table 3) among the differentially expressed genes in blood of OA patients as compared to controls we showed enrichment of compelling apoptotic genes (e.g. *CASP3* and FAS-associated death domain (*FADD*)) but also of genes less well-known for their involvement in the regulation of apoptosis (such as *GPR18*,[29] and *SLAH1*[30]). Apart from genetic and experimental evidence, the incidence of apoptosis/necrosis in OA cartilage is also supported by the observation of large numbers of empty lacunae and hypocellularity.[31-33] However, it remains to be established whether chondrocyte death is a cause or a consequence of OA.[34] Since the apoptotic pathway was not found to be significantly enriched in the cartilage specific gene expression dataset of the RAAK study,[14] we hypothesize that upon the occurrence of micro-traumas during life,

Table 3. Gene enrichment analysis using DAVID. Biological processes significantly enriched when testing all significant genes (A) or 86 genes with an adjusted $p \leq 0.05$ and ≥ 1.5 fold difference (B) in DAVID using gene ontology term GOTERM_BP_FAT (biological processes). Ngenes: number of genes in this GO-term; Pct.: percentage of genes from the total number of genes tested; p: P-value for enrichment in the indicated pathway; p_{corr}: P-value corrected for multiple testing; Fold Enr.: fold enrichment of the indicated pathway.

A

GO-Term	N_{genes}	Pct.	p	p_{corr}	Fold Enr.
GO:0015031~protein transport	51	8.24	1.22E-05	8.31E-03	1.91
GO:0045184~establishment of protein localization	51	8.24	1.61E-05	1.09E-02	1.89
GO:0034621~cellular macromolecular complex subunit organization	30	4.85	2.36E-05	1.60E-02	2.39
GO:0008104~protein localization	55	8.89	3.95E-05	2.68E-02	1.78
GO:0007049~cell cycle	50	8.08	4.60E-05	3.13E-02	1.83
GO:0006396~RNA processing	39	6.30	4.87E-05	3.31E-02	2.02
GO:0046907~intracellular transport	44	7.11	6.18E-05	4.19E-02	1.90
GO:0006917~induction of apoptosis	27	4.36	7.06E-05	4.79E-02	2.38

B

GO-Term	N_{genes}	Pct	p	p_{corr}	Fold Enr.
GO:0008219~cell death	13	17.11	8.52E-05	7.33E-03	3.88
GO:0012501~programmed cell death	12	15.79	8.92E-05	7.67E-03	4.22
GO:0016265~death	13	17.11	9.11E-05	7.83E-03	3.85
GO:0006915~apoptosis	11	14.47	3.69E-04	3.18E-02	3.93

systemic factors present in blood may contribute to the onset of apoptosis in the joint cells, eventually leading to irreversible damage and subsequent onset of OA.

Among the differentially expressed genes, we also found cytokines such as *IL1B* (up-regulated) and *IL8* (down-regulated) that were identified previously in a study with peripheral blood cells of knee OA patients.[22] Both proteins showed a considerable fold change in our analysis (1.8 and 3, respectively), and had a considerable number of interacting partners in the network analysis (6 and 7, respectively) suggesting that expression changes are likely to have a substantial impact on cellular processes including apoptosis. Also *MAFB* and *EGR1*, identified previously as biomarkers in blood expression profiles of patients with mild knee OA[35] were differentially expressed in our dataset. Taken together, the overlap between our analysis and other studies confirms that OA specific gene expression profiles of PBMCs exists. Besides the overlap we find differences as well most likely due to the fact that patients with generalized OA in at least 2 joints were included, comprising knee OA, but also OA at other joints (hip, hand, and/or spine). In addition, it was found previously that different subgroups of OA patients exist,[22] that differ in their gene expression profiles.

We showed that 4 of the genes identified in our analysis (*CASP3*, *CDKN2D*, *RABEP1*, and *SFRS5*), were sufficient to accurately distinguish patients from controls logically in

Table 4. Genes replicated by RT-qPCR. Expression of genes was determined by RT-qPCR on the fluidigm platform for 36 participants of the GARP study and 30 healthy controls (*: indicates genes that were not significantly different in the replication; Fold Diff: fold difference, relative to OA).

		ARRAY		REPLICATION	
Gene SYMBOL	**Annotated Gene**	***p***	**Fold Diff.**	***p***	**Fold Diff.**
ATF4	activating transcription factor 4 (tax-responsive enhancer element B67; transcript variant 1)	4.14E-08	0.48	3.59E-04	0.67
H3F3B *	H3 histone, family 3B (H3.3B)	4.31E-08	0.46	2.82E-01	0.95
GPR18	G protein-coupled receptor 18 (transcript variant 2)	4.31E-08	2.62	1.63E-02	1.39
ZNF564	zinc finger protein 564	5.22E-08	1.59	1.65E-05	1.91
ZNF20	zinc finger protein 20	9.14E-08	1.35	4.11E-06	1.96
LFNG *	LFNG O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase (transcript variant 1)	1.02E-07	1.79	2.14E-01	0.76
RIPK1	receptor (TNFRSF)-interacting serine-threonine kinase 1	4.27E-07	1.38	4.08E-03	1.16
ADRB2	adrenergic, beta-2-, receptor, surface	4.27E-07	2.23	3.40E-02	1.25
PELO	pelota homolog (Drosophila)	1.75E-06	1.65	1.92E-04	1.50
SIAH1	seven in absentia homolog 1 (Drosophila) (transcript variant 2)	2.64E-06	1.56	4.41E-06	2.60
NFKBIA *	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha	6.02E-06	0.57	3.14E-01	0.70
SLA	Src-like-adaptor (transcript variant 2)	1.90E-05	1.68	4.22E-03	1.19
SFRS5	splicing factor, arginine/serine-rich 5 (transcript variant 2)	2.80E-05	0.54	9.85E-08	0.56
CDKN2D	cyclin-dependent kinase inhibitor 2D (p19, inhibits CDK4) (transcript variant 2)	4.64E-05	0.62	5.75E-04	0.65
PLEKHF1	pleckstrin homology domain containing, family F (with FYVE domain) member 1	2.67E-04	1.78	1.42E-03	1.69
CASP3	caspase 3, apoptosis-related cysteine peptidase, transcript variant beta	8.94E-04	1.59	2.96E-04	1.59
ID3 *	inhibitor of DNA binding 3, dominant negative helix-loop-helix protein	1.48E-03	2.04	5.06E-01	3.86
RABEP1	rabaptin, RAB GTPase binding effector protein 1 (transcript variant 2)	1.93E-03	1.75	1.91E-03	1.79
FADD	Fas (TNFRSF6)-associated via death domain	3.20E-03	1.44	1.16E-02	1.33
PPP1R15A	protein phosphatase 1, regulatory (inhibitor) subunit 15A	5.12E-03	0.61	4.51E-05	0.42
CLK1	CDC-like kinase 1	5.34E-03	2.26	2.65E-02	1.25
OSM	oncostatin M	6.78E-03	0.65	2.06E-03	0.31
GSTM2	glutathione S-transferase M2 (muscle)	6.97E-03	0.80	5.06E-03	0.57
EGR1	early growth response 1	2.13E-02	0.19	4.10E-04	0.33
IL1B *	interleukin 1, beta	2.84E-02	1.76	6.40E-02	1.77
IL8	interleukin 8	2.98E-02	0.33	4.43E-05	0.27
E2F2	E2F transcription factor 2	3.48E-02	1.44	6.34E-03	1.48

the 'Discovery Cohort', but also in the 'Replication Cohort'. The classification models computed in our analysis combined clinical risk factors with expression levels of genes. In this way, cross-validated AUCs reached values as high as 98%. The AUC has the interpretation of a concordance index, meaning that the AUC represents the chance that for any case-control pair a higher disease probability is correctly predicted for the case. Therefore, when confirmed in independent cases 98% is a clinically relevant value. Moreover, expression levels represent risk independent of clinical variables (98% vs.

72%). It should be noted however that our study has a cross sectional design. As a result, the ROC curves merely indicate classification (diagnosis) but not prediction, and the result remains to be tested in independent study cohorts. Predictive value needs to be assessed and validated in a longitudinal study design. In addition, it remains to be investigated whether *CASP3*, *CDKN2D*, *RABEP1*, and *SRFS5* would be useful as early markers, to diagnose OA in a population based setting.

Although the results presented in this paper appear very convincing and are in agreement with previous studies, it should be mentioned that a draw-back of our analyses was the fact that the majority of the controls (25 of the 33 used) across our validation and classification model were overlapping with the controls used in our microarray analysis. As a result, part of our data could be based merely on the severe selection of controls (clinically and radiographically confirmed healthy joints at age ~60 years) marking rather a healthy blood expression profile instead of presence of OA. The latter could be explored further by replication using population based controls. Also, our 'Discovery Cohort' included 38 siblings which will have contributed to the robustness of the results. However, differential expression of 22 out of 27 genes has been confirmed by RT-qPCR in the 'Replication Cohort' consisting in independent cases (Table 4), and results of the analysis including only the probands of the 'Discovery Cohort' are generally comparable (data not shown). As shown in Table 1, most participants suffered from pain due to OA (92%) and more than half of our cases were using NSAIDs (52%). Adjusting for these phenotypes did not significantly affect our results (data not shown). It is of note that in this study mainly females were included. Therefore, although we have adjusted for gender we cannot exclude the possibility that results are more applicable to females. It would be interesting for future research to investigate whether sex-differences exist for gene expression profiles in PBMCs of OA patients. Finally, despite various efforts, we were not able to validate by RT-qPCR the observed highly significant differential expression of *UBE2L3* as detected by the microarray expression (Table 2).

Taken together, this study in patients with generalized OA in a familial background provides a valuable tool for a better understanding of the pathology of this complex disease and for the classification of OA. In addition, the gene networks may facilitate selection of targets for drug development and biomarkers.

ACKNOWLEDGEMENTS

The authors want to thank all members of the lab for suggestions and helpful discussions. Dr. J.M. Boer and Dr. J.J. Goeman are acknowledged for initial help with the analysis of the microarrays. This study was supported by the Leiden University Medical Center, the Dutch Arthritis Association and the Center of Medical System Biology. Pfizer Inc., Groton, CT, USA supported the inclusion of the GARP study.

Licence for Publication: The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted)

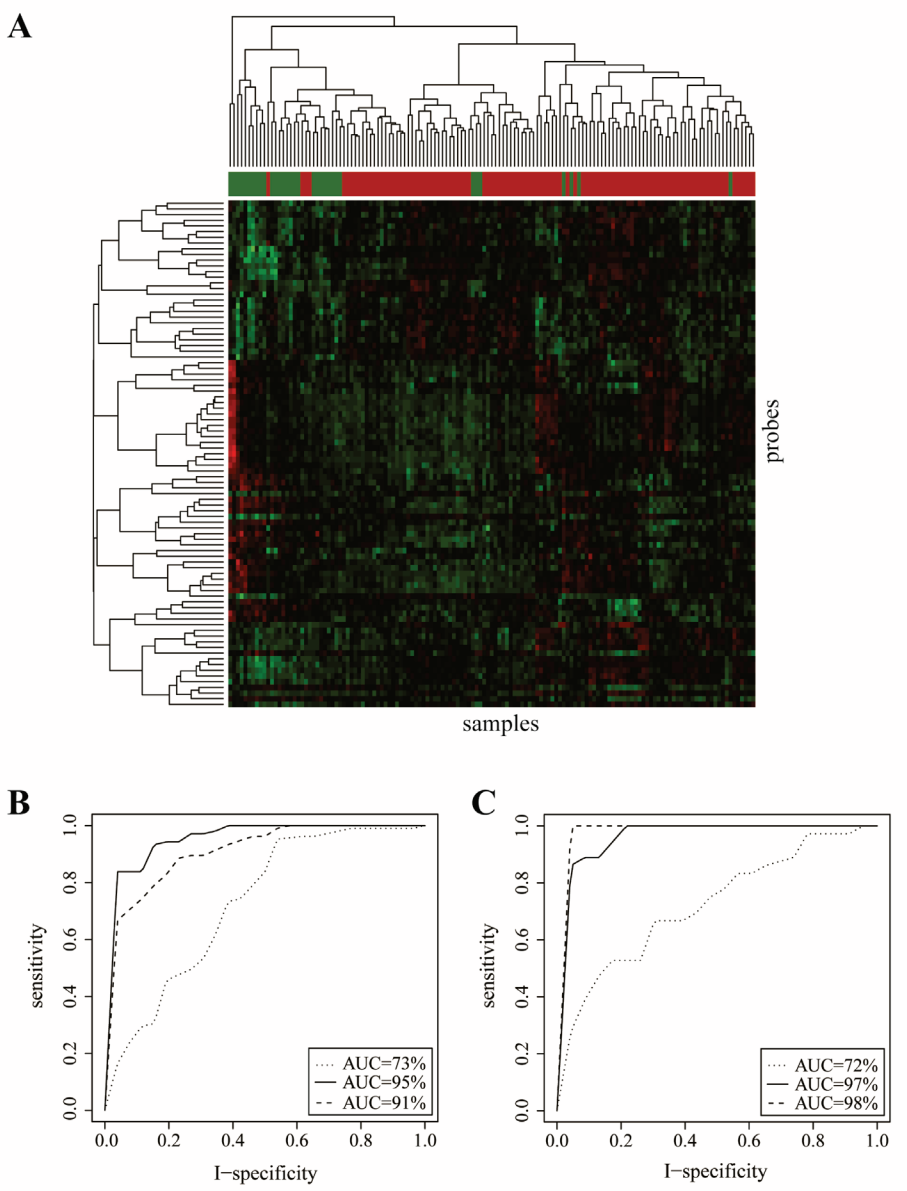


Figure 3. Clustering and classification of the study population. A) Clustering of OA patients of the GARP study (N=106; indicated in red in dendrogram) and age-matched healthy controls (N=33; indicated in green in dendrogram) based on euclidean distance. Each row represents a probe (N=89), and each column represents the expression of the probes for an individual (range from bright red [expression in OA higher than control mean] to bright green [expression in OA lower than control mean]). **B)** Receiver operating curve (ROC) for the 'Discovery Cohort', based on sex, age, and bmi alone (dotted line) or with the inclusion of the 27 validated genes (black line) or 4 genes (*CASP3*, *SFRS5*, *RABEP1*, and *CDKN2D*; dashed line) in our classification model. **C)** ROC as in B, for the 'Replication Cohort' (AUC: area under the curve).

to be published in ARD and any other BMJPGL products, and sublicenses such use and exploit all subsidiary rights, as set out in the license.

Competing Interest: None declared

REFERENCES

1 Panoutsopoulou K, Southam L, Elliott KS, Wrayner N, Zhai G, Beazley C, *et al.* Insights into the genetic architecture of osteoarthritis from stage 1 of the arcOGEN study. *Ann Rheum Dis* 2011;**70**:864-7.

2 Seed SM, Dunican KC, Lynch AM. Treatment options for osteoarthritis: considerations for older adults. *Hosp Pract (Minneap)* 2011;**39**:62-73.

3 Chapman K, Takahashi A, Meulenbelt I, Watson C, Rodriguez-Lopez J, Egli R, *et al.* A meta-analysis of European and Asian cohorts reveals a global role of a functional SNP in the 5' UTR of GDF5 with osteoarthritis susceptibility. *Hum Mol Genet* 2008;**17**:1497-504.

4 Loughlin J, Dowling B, Chapman K, Marcelline L, Mustafa Z, Southam L, *et al.* Functional variants within the secreted frizzled-related protein 3 gene are associated with hip osteoarthritis in females. *Proc Natl Acad Sci U S A* 2004;**101**:9757-62.

5 Valdes AM, Spector TD, Tamm A, Kisand K, Doherty SA, Dennison EM, *et al.* Genetic variation in the SMAD3 gene is associated with hip and knee osteoarthritis. *Arthritis Rheum* 2010;**62**:2347-52.

6 Meulenbelt I, Min JL, Bos S, Riyazi N, Houwing-Duistermaat JJ, van der Wijk HJ, *et al.* Identification of DIO2 as a new susceptibility locus for symptomatic osteoarthritis. *Hum Mol Genet* 2008;**17**:1867-75.

7 Bos SD, Slagboom PE, Meulenbelt I. New insights into osteoarthritis: early developmental features of an ageing-related disease. *Curr Opin Rheumatol* 2008;**20**:553-9.

8 Sandell LJ. Etiology of osteoarthritis: genetics and synovial joint development. *Nat Rev Rheumatol* 2012;**8**:77-89.

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

10 Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* 2012;**380**:815-23.

11 Batliwalla FM, Li W, Ritchlin CT, Xiao X, Brenner M, Laragione T, *et al.* Microarray analyses of peripheral blood cells identifies unique gene expression signature in psoriatic arthritis. *Mol Med* 2005;**11**:21-9.

12 Runne H, Kuhn A, Wild EJ, Pratyaksha W, Kristiansen M, Isaacs JD, *et al.* Analysis of potential transcriptomic biomarkers for Huntington's disease in peripheral blood. *Proc Natl Acad Sci U S A* 2007;**104**:14424-9.

13 Riyazi N, Meulenbelt I, Kroon HM, Ronday KH, Hellio le Graverand MP, Rosendaal FR, *et al.* Evidence for familial aggregation of hand, hip, and spine but not knee osteoarthritis in siblings with multiple joint involvement: the GARP study. *Ann Rheum Dis* 2005;**64**:438-43.

14 Ramos YF, den HW, Bovee JV, Bomer N, van der Breggen R, Lakenberg N, *et al.* Genes Involved in the Osteoarthritis Process Identified through Genome Wide Expression Analysis in Articular Cartilage; the RAAK

Study. *PLoS One* 2014;**9**:e103056.

15 Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* 2008;**24**:1547-8.

16 Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004;**3**:Article3.

17 Huang W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 9 A.D.;**37**:1-13.

18 Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;**4**:44-57.

19 Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res* 2011;**39**:D561-D568.

20 Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001;**25**:402-8.

21 Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2011;**73**:273-82.

22 Attur M, Belitskaya-Levy I, Oh C, Krasnokutsky S, Greenberg J, Samuels J, *et al.* Increased interleukin-1beta gene expression in peripheral blood leukocytes is associated with increased pain and predicts risk for progression of symptomatic knee osteoarthritis. *Arthritis Rheum* 2011;**63**:1908-17.

23 Wang W, Lian N, Li L, Moss HE, Wang W, Perrien DS, *et al.* Atf4 regulates chondrocyte proliferation and differentiation during endochondral ossification by activating Ihh transcription. *Development* 2009;**136**:4143-53.

24 Aigner T, Fundel K, Saas J, Gebhard PM, Haag J, Weiss T, *et al.* Large-scale gene expression profiling reveals major pathogenetic pathways of cartilage degeneration in osteoarthritis. *Arthritis Rheum* 2006;**54**:3533-44.

25 Sato T, Konomi K, Yamasaki S, Aratani S, Tsuchimochi K, Yokouchi M, *et al.* Comparative analysis of gene expression profiles in intact and damaged regions of human osteoarthritic cartilage. *Arthritis Rheum* 2006;**54**:808-17.

26 Geyer M, Grassel S, Straub RH, Schett G, Dinser R, Grifka J, *et al.* Differential transcriptome analysis of intraarticular lesional vs intact cartilage reveals new candidate genes in osteoarthritis pathophysiology. *Osteoarthritis Cartilage* 2009;**17**:328-35.

27 Xu Y, Barter MJ, Swan DC, Rankin KS, Rowan AD, Santibanez-Koref M, *et al.* Identification of the pathogenic pathways in osteoarthritic hip cartilage: commonality and discord between hip and knee OA. *Osteoarthritis Cartilage* 2012;**20**:1029-38.

28 Kronenberg HM, Chung U. The parathyroid hormone-related protein and Indian hedgehog feedback loop in the growth plate. *Novartis Found Symp* 2001;**232**:144-52.

29 Takenouchi R, Inoue K, Kambe Y, Miyata A. N-arachidonoyl glycine induces macrophage apoptosis via GPR18. *Biochem Biophys Res Commun* 2012;**418**:366-71.

30 Xu Z, Sproul A, Wang W, Kukekov N, Greene LA. Siah1 interacts with the scaffold protein POSH to promote JNK activation and apoptosis. *J Biol Chem* 2006;**281**:303-12.

31 Blanco FJ, Rego I, Ruiz-Romero C. The role of mitochondria in osteoarthritis. *Nat Rev Rheumatol* 2011;**7**:161-9.

32 van den Berg WB. Osteoarthritis year 2010 in review: pathomechanisms. *Osteoarthritis Cartilage* 2011;**19**:338-41.

33 Valdes AM, Spector TD. The clinical relevance of genetic susceptibility to osteoarthritis. *Best Pract Res Clin Rheumatol* 2010;**24**:3-14.

34 Zamli Z, Sharif M. Chondrocyte apoptosis: a cause or consequence of osteoarthritis? *Int J Rheum Dis* 2011;**14**:159-66.

35 Marshall KW, Zhang H, Yager TD, Nossova N, Dempsey A, Zheng R, *et al.* Blood-based biomarkers for detecting mild osteoarthritis in the human knee. *Osteoarthritis Cartilage* 2005;**13**:861-71.

Supplementary Table S1. Taqman assays used for RT-qPCR.

Gene Symbol	Taqman assay	Assay Type
ACTB	Hs99999903_m1	inventoried
ADRB2	Hs00240532_s1	inventoried
ATF4	Hs00909569_g1	inventoried
BIRC3	Hs00154109_m1	inventoried
CASP3	Hs00234387_m1	inventoried
CDKN2D	Hs00176481_m1	inventoried
CLK1	Hs00269734_m1	inventoried
E2F2	Hs00231667_m1	inventoried
EGR1	Hs00152928_m1	inventoried
FADD	Hs04187499_m1	inventoried
GPR18	Hs01921463_s1	inventoried
GSTM2	Hs03044640_gH	made to order
H3F3B	Hs00287906_s1	inventoried
ID3	Hs00171409_m1	inventoried
IL1B	Hs01555410_m1	inventoried
IL8	Hs00174103_m1	inventoried
LFNG	Hs00385436_g1	inventoried
NFKBIA	Hs00153283_m1	inventoried
OSM	Hs00171165_m1	inventoried
PELO	Hs00560251_g1	inventoried
PLEKHF1	Hs00759096_s1	inventoried
PPP1R15A	Hs00169585_m1	inventoried
RABEP1	Hs00388983_m1	inventoried
RIPK1	Hs00169407_m1	inventoried
SFRS5	Hs00951031_m1	inventoried
SIAH1	Hs00361785_m1	inventoried
SLA	Hs00277129_m1	inventoried
STK17B	Hs00892456_m1	inventoried
UBE2L3	Hs00748530_s1	inventoried
YKT6	Hs00559914_m1	inventoried
ZNF20	Hs00397296_m1	inventoried
ZNF564	Hs00793693_g1	inventoried

Supplementary Table S2. List of all significant probes can be downloaded from http://ard.bmj.com/content/73/10/1844.full

P-value and fold difference of the 741 significant probes (694 unique genes) that were identified in the analysis using R statistical programming language with lumi-package.

Supplementary Table S4. Genes significantly different expressed between OA and preserved cartilage. Depicted are genes significantly different expressed with OA in blood (patients versus healthy controls) and in cartilage (OA versus preserved; Fold Diff.=fold difference; AveExpr means the average relative gene expression as determined with the microarray analysis).

	BLOOD			CARTILAGE		
Gene SYMBOL	***p***	**Fold Diff.**	**AveExpr**	***p***	**Fold Change**	**AveExpr**
H3F3B	4.31E-08	0.46	9.18	2.05E-03	0.87	9.83
NFKBIA	6.02E-06	0.57	11.45	6.79E-03	0.91	11.14
DAB2	2.16E-05	0.59	7.89	4.99E-02	0.90	9.80
FAM43A	2.26E-05	1.92	8.15	1.13E-02	1.15	8.39
AP3M1	2.79E-04	1.53	8.85	5.00E-03	1.11	8.60
SFRS5	4.61E-04	0.59	10.99	3.81E-02	0.93	11.47
LOC338799	5.02E-04	0.59	7.65	6.14E-04	0.88	7.65
MAF	8.94E-04	0.60	8.50	2.98E-03	0.83	9.23
MAX	2.40E-03	0.66	7.99	2.63E-03	0.95	7.23
ATHL1	3.98E-02	0.55	8.44	2.39E-02	0.91	8.00
MPL	4.19E-02	0.56	8.17	2.82E-02	0.95	7.22
C7ORF41	4.97E-02	0.60	9.76	2.71E-05	0.79	9.30

Supplementary Table S3. Genes technically validated by RT-qPCR. Expression of genes was determined by RT-qPCR on the fluidigm platform for 28 participants of the GARP study and 25 healthy controls analyzed in the microarray. Genes that were not significantly different in the validation are indicated with an asterisk and these were not included in further analysis. Spearman correlation values and P-values are shown (corr.$_{Spear}$ and p_{Spear} respectively), Fold Diff.= fold difference, relative to OA.

	ARRAY		VALIDATION			
Gene SYMBOL	***p***	**Fold Diff.**	***p***	**Fold Diff.**	**corr.$_{Spear.}$**	**$p_{Spear.}$**
*UBE2L3 **	9.26E-12	0.55	2.42E-01	1.56	-0.36	7.36E-03
ATF4	4.14E-08	0.48	3.59E-04	0.81	0.48	2.42E-04
H3F3B	4.31E-08	0.46	2.82E-01	0.57	0.70	5.09E-09
GPR18	4.31E-08	2.62	1.63E-02	1.36	0.59	3.86E-06
ZNF564	5.22E-08	1.59	1.65E-05	1.40	0.34	1.20E-02
ZNF20	9.14E-08	1.35	4.11E-06	1.83	0.69	9.03E-09
LFNG	1.02E-07	1.79	2.14E-01	1.46	0.68	3.04E-08
RIPK1	4.27E-07	1.38	5.16E-02	1.24	0.27	4.80E-02
ADRB2	4.27E-07	2.23	3.40E-02	1.75	0.75	9.82E-11
PELO	1.75E-06	1.65	1.92E-04	2.00	0.68	1.92E-08
SIAH1	2.64E-06	1.56	9.02E-08	1.50	0.51	1.10E-04
NFKB1A	6.02E-06	0.57	9.56E-02	0.64	0.56	1.39E-05
SLA	1.90E-05	1.68	4.22E-03	1.68	0.59	2.81E-06
SFRS5	2.80E-05	0.54	9.85E-08	0.30	0.81	1.77E-13
*STK17B **	3.12E-05	0.61	5.33E-01	0.90	0.15	2.72E-01
CDKN2D	4.64E-05	0.62	4.37E-01	0.91	0.56	1.53E-05
PLEKHF1	2.67E-04	1.78	7.50E-09	2.46	0.73	4.50E-10
CASP3	8.94E-04	1.59	2.27E-01	1.23	0.60	2.30E-06
ID3	1.48E-03	2.04	5.06E-01	3.00	0.88	6.91E-18
RABEP1	1.93E-03	1.75	3.01E-01	1.37	0.67	3.75E-08
FADD	3.20E-03	1.44	1.15E-09	2.36	0.62	6.51E-07
PPP1R15	5.12E-03	0.61	1.15E-03	0.53	0.72	1.04E-09
CLK1	5.34E-03	2.26	2.65E-02	1.26	0.87	3.75E-17
OSM	6.78E-03	0.65	1.12E-01	0.40	0.57	9.91E-06
GSTM2	6.97E-03	0.80	5.06E-03	0.83	0.64	1.99E-07
*BIRC3 **	8.02E-03	2.01	7.78E-01	0.92	0.82	1.08E-13
EGR1	2.13E-02	0.19	4.10E-04	0.16	0.92	9.07E-19
IL1B	2.84E-02	1.76	2.63E-01	1.28	0.83	2.88E-14
IL8	2.98E-02	0.33	4.43E-05	0.04	0.89	2.87E-16
E2F2	3.48E-02	1.44	2.30E-07	2.19	0.32	1.87E-02

Meta-analysis identifies loci affecting levels of the potential OA biomarkers sCOMP and uCTX-II with genome wide significance

J Med Genet 2014;51:596-604 (doi:10.1136/jmedgenet-2014-102478)

Meta-analysis identifies loci affecting levels of the potential OA biomarkers sCOMP and uCTX-II with genome wide significance

Yolande F. M. Ramos[1,2*#], Sarah Metrustry[3*], Nigel Arden[4,5], Anne C. Bay-Jensen[6], Marian Beekman[1,2], Anton J. M. de Craen[7,2], L. Adrienne Cupples[8,9], Tõnu Esko[10,11,12], Evangelos Evangelou[13,3], David T. Felson[14], Deborah J. Hart[3], John P.A. Ioannidis[13,15], Morten Karsdal[6], Margreet Kloppenburg[16], Floris Lafeber[17], Andres Metspalu[10], Kalliope Panoutsopoulou[18], P. Eline Slagboom[1,2], Tim D. Spector[3], Erwin W. E. van Spil[17], Andre G. Uitterlinden[19,2], Yanyan Zhu[20], arcOGEN consortium, TreatOA collaborators, Ana M. Valdes[21‡], Joyce B.J. van Meurs[19,2‡] and Ingrid Meulenbelt[1,2‡]

[*] authors contributed equally

[‡] authors contributed equally

Author affiliations are shown at the end of this chapter

ABSTRACT

Background **Research for the use of biomarkers in osteoarthritis is promising, however, adequate discrimination between patients and controls may be hampered due to innate differences. We set out to identify loci influencing levels of serum cartilage oligomeric protein (sCOMP) and urinary C-telopeptide of type II collagen (uCTX-II).**

Methods **Meta-analysis of genome-wide association studies (GWAS) was applied to standardized residuals of sCOMP (N=3,316) and uCTX-II (N=4,654) levels available in respectively 6 and 7 studies from TreatOA. Effects were estimated using a fixed-effects model. Six promising signals were followed-up by *de novo* genotyping in the CHECK study (N=964). Subsequently, their role in OA susceptibility was investigated in large-scale GWAS meta-analyses for OA. Differential expression of annotated genes was assessed in cartilage.**

Results **Genome-wide significant association with sCOMP levels was found for a SNP within *MRC1* (rs691461, p=1.7x10^{-12}) and a SNP within *CSMD1* associated with variation in uCTX-II levels with borderline genome-wide significance (rs1983474, p=8.5x10^{-8}). Indication for association with sCOMP levels was also found for a locus close to the *COMP* gene itself (rs10038, p=7.1x10^{-6}). The latter SNP was subsequently found to be associated with hip OA whereas *COMP* expression appeared responsive to the OA pathophysiology in cartilage.**

Conclusions **We have identified genetic loci affecting either uCTX-II or sCOMP levels. The genome wide significant association of *MRC1* with sCOMP levels was found likely to act independent of OA-subtypes. Increased sensitivity of biomarkers with OA may be accomplished by taking genetic variation into account.**

Key words: meta-analysis, biomarkers, sCOMP, uCTX-II, osteoarthritis

INTRODUCTION

Osteoarthritis (OA) is a prevalent, complex, disabling disease affecting articular joints. Age and unfavorable metabolic states, for example high BMI, are strong risk factors of OA,[1] however, development of OA is also determined by a considerable genetic component of a polygenic nature.[2] Commonly, OA is diagnosed clinically and radiographically at a stage in which joint damage is already significant, and non-surgical therapies to cure or halt progress of OA are not yet available. Research directed to identify clinical biochemical markers to sensitively monitor OA disease activity in an individual over time or to assess quantitative joint tissue remodeling is currently in progress.[3-6] So far, these studies have suggested that serum cartilage oligomeric protein (sCOMP), and urinary C-telopeptide of type II collagen (uCTX-II) are promising candidates and they are listed among the biomarkers that were recommended to focus on in future research efforts.[4]

Initially, both sCOMP and uCTX-II were considered as markers for cartilage degradation: sCOMP would be representative of cartilage turnover, and uCTX-II is a collagen type II cleavage product representative of hyaline (articular) cartilage degradation. However, some reports indicate that uCTX-II may not merely be a cartilage specific marker but rather a reflection of joint tissue remodeling including calcified cartilage and subchondral bone.[7, 8] Moreover, Meulenbelt *et al*[9] showed that in patients with generalized OA uCTX-II is a sensitive marker of whole body cartilage degeneration as reflected by radiographic signs of OA but association with a specific joint site could not be demonstrated. This was recently confirmed in a large-scale meta-analyses showing that uCTX-II associated with risk for hand, knee, and hip OA.[10] In addition, the analyses showed that both, sCOMP as well as uCTX-II, can be useful to monitor OA disease activity. However, the study also highlighted that the use of sCOMP and uCTX-II as biomarkers in the clinic still remains largely unsatisfactory.

Importantly, investigations have shown that in subjects with familial OA at multiple joint sites besides clustering of sCOMP together with hand OA and with age, sCOMP has a high familial aggregation.[8, 11] At present, little is known about genetic variants affecting basal levels of sCOMP and uCTX-II. Identification of genes that contribute to these levels may provide mechanistic insight into factors that influence cartilage tissue turnover, dependent or independent of the OA pathophysiology, which in the latter case would thus perturb the use of these molecules as biochemical markers of OA. The objective of the current work is, therefore, to identify genetic variants involved in the variation in basal levels of sCOMP and uCTX-II by means of a large-scale hypothesis-free meta-analysis of genome-wide association studies (GWAS) across 3 English and 4 Dutch Caucasian studies. Additional replication occurred in the Dutch Cohort Hip and Cohort Knee study (CHECK) and signals were further tested for association with hip and knee OA. Finally, given that functional follow-up studies of susceptibility loci have demonstrated that they frequently result in expression differences of positional genes within the relevant diseased tissue,[12] we explored the behavior of mRNA expression of the annotated genes in preserved and osteoarthritic articular cartilage.

Table 1. Overview of the discovery and replication studies (Chingford: Chingford 1000 Women Study; GARP: Genetics OsteoArthritis and Progression study Leiden; LLS: Leiden Longevity Study; RS-I and RS-II: Rotterdam study I and II, respectively; TwinsUK: the United Kingdom Adult Twin Registry; VIDEO: Vitamin D evaluation in osteoarthritis; CHECK= Cohort Hip and Cohort Knee; Avg (SD): average of the log10-transformed values with the standard deviation).

			sCOMP				uCTX-II			
Study	**Ref.**	**Study design**	**# samples**	**# SNPs**	**Lamda**	**Avg (SD)**	**# samples**	**# SNPs**	**Lamda**	**Avg (SD)**
Chingford	[18]	Population Based Cohort	159	536,984	1.002	0.98 (0.15)	159	544,135	1.003	2.28 (0.22)
GARP	[13]	Sibling pair study	351	2,162,385	1.145	1.05 (0.12)	353	2,165,144	1.008	2.31 (0.26)
LLS	[14]	Offspring of long lived subjects & partners	620	2,267,101	1.033	0.98 (0.11)	639	2,225,148	1.026	2.35 (0.25)
RS-I	[15]	Population Based Cohort	-	-	-	-	1,100	2,543,887	1.02	2.25 (0.22)
RS-II	[16]	Population Based Cohort	1,161	2,543,887	1.01	1.03 (0.10)	1,044	2,543,887	0.997	2.32 (0.23)
TwinsUK	[17]	Twins Based Cohort	754	3,044,064	0.987	1.02 (0.13)	1,096	2,441,760	1.008	2.29 (0.24)
VIDEO	[19]	Vitamin D evaluation in OA	271	537,298	0.977	1.05 (0.15)	263	543,415	1.008	2.46 (0.22)
Total # samples:			**3,316**				**4,654**			
CHECK	[24]	Early OA Cohort	**964**		NA	0.93 (0.11)	**964**		NA	2.27 (0.27)

MATERIALS AND METHODS

Studies. A full detailed description of each study cohort is presented in the Supplementary Methods section. Baseline characteristics and sample sizes are shown in Table 1. We included studies from the Netherlands: GARP,[13] Leiden Longevity (LLS),[14] Rotterdam Study I and II (RS-I and RS-II),[15, 16] and the United Kingdom: TwinsUK,[17] Chingford,[18] VIDEO.[19] All studies were approved by their institutional ethics review committees and all participants provided written informed consent.

Study design. To identify genes associated with uCTX-II and sCOMP levels a 2-stage design was applied (Figure 1 shows a schematic representation of the work-flow). In the first (discovery) stage, a meta-analysis of GWAS was performed by combining the summary statistics of 7 independent studies (4 Caucasian studies in the Netherlands and 3 in United Kingdom) for genome-wide association with uCTX-II and sCOMP levels. In the second stage, loci (2 for uCTX-II and 3 for sCOMP) that reached suggestive evidence for genome wide association ($p \leq 1x10^{-6}$) and a locus within the *COMP* gene itself ($p=2.3x10^{-5}$) were genotyped *de novo* in the CHECK cohort. Subsequently, meta-analysis of all results was performed in R statistical language similar to the analysis of the discovery stage by combining the summary statistics of 8 independent studies (5 Caucasian studies in the Netherlands and 3 in United Kingdom) for genome-wide association with uCTX-II and sCOMP levels. In addition, data for association of the variants with hip and knee OA were derived from a meta-analysis of GWAS for OA associated loci among 6 and 7 studies, respectively,[20, 21] and OA-related expression changes of annotated genes in cartilage were derived from an available dataset generated within the ongoing Research Arthritis and Articular Cartilage study (RAAK study[22]).

An overview of the populations and available measurements is shown in Table 1. Detailed description of the populations included in this study is provided in Supplementary Methods, and Supplementary Table S1 shows quality control thresholds applied in each study.

Urinary CTX-II and serum COMP measurements. All measurements for uCTX-II in samples of Chingford, Rotterdam (studies I and II), TwinsUK, and VIDEO were performed as part of the TreatOA consortium with the use of an enzyme-linked immunosorbent assay based on a monoclonal antibody raised against the linear 6-amino acid EKGPDP epitope of the CII C-telopeptide (CartiLaps; Immunodiagnostic systems, Paris, France). Intra- and inter-assay variation was lower than 9% and 11%, respectively. Serum COMP levels for Chingford, RS-II, Twins-UK, and VIDEO were measured using a two-site immunoassay (COMP™ ELISA kit, AnaMar Medical, Lund, Sweden). Both, intra- and the inter-assay variation were below 10%.

Assessment of uCTX-II and sCOMP levels in samples of the GARP study and Leiden Longevity study was performed previously by Synarc Laboratories Lyon, France and has been described in detail elsewhere.[8, 9] Measurement of the biomarkers in the CHECK cohort was done with commercially available ELISA kits

according to the manufacturers protocol as described previously.[7]

Genotyping, quality controls and imputation. Genotyping of the cohorts was done by Illumina Infinium HumanHap550 Beadchip (RS-I and RS-II), the Illumina Infinium HumanHap610 (GARP, LLS and TwinsUK), or the Illumina Infinium HumanHap300 (VIDEO and CHINGFORD). Criteria for quality control (QC) are shown in Supplementary Table S2. For all studies, normalized intensity data of the GWAS were pooled, and genotypes were called on the basis of the Illuminus algorithm. No calls were assigned if the most likely call was less than a posterior probability of 0.95. Validation of pooling was done by visual inspection of 100 random, shared SNPs for overt batch effects (none were observed). SNPs that had a low call rate (≤ 90%), Hardy-Weinberg p-values < 10^{-6} and minor allele frequencies (MAF) < 1% were excluded, and samples with call rates < 95% were removed.

Genotype imputations were performed to increase the coverage and to evaluate association of one and the same SNP across samples typed on different genotyping platforms. Genotypes were imputed for all polymorphic SNPs (MAF > 0.01) using either MaCH[23] or IMPUTE[24] software, based upon phased autosomal chromosomes of the HapMap CEU Phase II panel (release 22, build 36), orientated on the positive strand. Imputation QC metrics from MACH and IMPUTE were used to filter out SNPs with low-quality data. SNPs that successfully passed QC criteria (MaCH r^2-hat ≥0.30 or IMPUTE proper_info ≥0.40) were considered for the analysis. Detailed description of QC and imputation procedures are provided in Supplementary Table S2.

De novo genotyping of SNPs in CHECK[25] was performed by mass spectrometry using the homogeneous MassARRAY iPlex Gold from Sequenom (San Diego, CA) under standard conditions. SNP rs691461 was genotyped by means of a TaqMan Genotyping Assay (Applied Biosystems, Foster City, CA) since this SNP did not fit into the iPlex design.

Analysis of GWAS. After selection and QC, data of sCOMP and uCTX-II were analyzed separately for each study. For both traits $\log_{10}$ transformations were performed and standardized residuals from regression were extracted with mean = 0 and standard deviation = 1. All studies performed regression analysis, adjusting for age, age squared, sex and population specific co-variables such as familial dependencies or population stratification. Multiplicative (Cochran-Armitage)/additive allele-dose were carried out for imputed SNPs where possible using a linear regression framework as implemented in MACH2DAT,[23] QTASSOC (in-house analysis program, developed to correct for familial dependence; H. W. Uh and Q. Helmer[26, 27]) and PLINK[28] (Supplementary Table S3). The estimated inflation factors in the sCOMP analyses were 1.002, 0.986, 1.033, 1.01, 0.987, and 0.977 for Chingford, GARP, LLS, RS-II, TwinsUK and VIDEO, respectively. The estimated inflation factors in the uCTX-II analyses were 1.003, 1.008, 1.026,1.020, 0.997, 1.008, 1.008 for Chingford, GARP, LLS, RS-I, RS-II, TwinsUK and VIDEO, respectively (Table 1).

Meta-analyses. Summary statistics of the

individual GWAS were combined by means of a fixed-effects inverse variance-weighted meta-analysis, as implemented in METAL[29] across 1,961,964 autosomal SNPs. In the meta-analysis the genomic control (GC) method[30] was used as implemented in METAL which corrects for any residual population stratification or relatedness that is not accounted for by the four most important PCs.[29] The genomic control inflation factor of the meta-data was than calculated again to assess resulting inflation of the test statistics in R statistical programming language using the GenABEL-package.[31]

$P<5x10^{-8}$ was considered genome-wide significant. Variants with a $p\leq1x10^{-6}$ in the meta-analyses were selected for follow-up in the replication stage of this study, and for these SNPs meta-analyses were performed in R (http://www.r-project.org) using the meta package. Here, summary regression coefficients (effect sizes, betas) were estimated using the fixed-effects model of DerSimonian and Laird.[32] Heterogeneity was quantified using the I2 statistic for inconsistency[33] and its statistical significance was tested with the $\chi2$ distributed Cochran Q statistic.[34] I2 describes the proportion of variation due to heterogeneity that is unlikely to be due to chance and is considered large for values over 50%.[33] Regional association plots of the meta-analysis results were obtained with LocusZoom.[35]

Exploratory OA phenotype check. To explore whether the identified variants in our study may share common pathophysiological paths and are associated with OA we tested for potential associations in previously performed meta-analyses of GWAS for hip OA (a total of 4,349 cases versus 46,903 controls, including the following studies: arcOGEN stage 1, deCODE, EGCUT, GARP, RS-I, RS-II, RS-II, and TwinsUK[20]) and knee OA (a total of 5,755 cases versus 48,578 controls, including the same studies as for hip OA and in addition the Framingham study[21]). Specifically, associations were assessed using a fixed-effects inverse variance meta-analysis of GWAS for the aforementioned phenotypes to obtain the summary effect size and the magnitude of the association along with the direction of the effect. A full detailed description of included studies on recruitment with radiographic and clinical assessment has been published previously.[20, 21]

Gene expression analyses. OA-related changes in expression levels of positional genes were assessed in the RAAK study aimed at the biobanking of joint materials (cartilage, bone and where available ligaments), mesenchymal stem cells, and primary chondrocytes of patients and controls in the Leiden University Medical Center and collaborating outpatient clinics in the Leiden area. A gene expression dataset of 33 paired samples of OA-affected versus preserved cartilage of the same joint derived from patients that underwent joint replacement due to end stage OA was used (described previously[20, 22, 36] and explained in more detail in Supplementary methods). In short, a total of 0.5 μg of RNA was used as template. Subsequently, first- and second-strand reverse transcription steps and labeling with biotin-labeled nucleotides was performed using an Ambion RNA amplification kit (TotalPrep™ RNA Amplification Kit). Microarray analysis

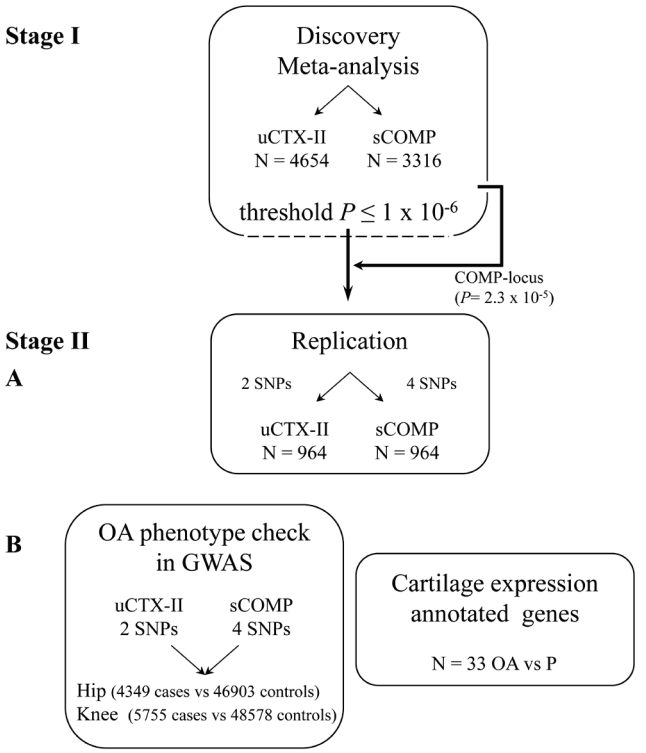


Figure 1. Schematic representation of the work-flow.

(Illumina HT-12 V3) was performed and analyzed in R statistical programming language using the Limma package to compare gene expression in the 33 sample pairs with a paired T-Test. Mean expression of all genes analyzed was 7.4 (6.6-14.9), expressed as the overall relative light intensity signals in respectively the preserved and OA affected cartilage as provided by Illumina™.

RESULTS

A meta-analysis of GWAS was performed across Caucasian samples to identify loci associated with serum levels of COMP available in 3,316 subjects across 6 studies and of uCTX-II available in 4,654 subjects across 7 studies (Figure 1). Of these, two studies had only genotyped SNPs (N ~ 530,000) whereas the other studies used additional imputed SNPs (N ~ 2.5 million; Table 1 and Supplementary Methods).

GWAS meta-analysis for sCOMP. In the sCOMP meta-analysis a total of 1,961,964 SNPs (directly genotyped or imputed in at least 4 studies) were tested for association. The overall genomic control lambda (λ_{GC}) was 1.0131, indicating no significant population stratification (Supplementary Figure S1). A genome wide significant signal was found for SNP rs691461 (p=1.9x10^{-12}), located at chromosome 10 (17.9 Mb) within intron 7 of the Mannose receptor C type 1 (*MRC1*) gene (Figure 2). Two other signals

that reached the pre-specified follow-up threshold, were rs10429475 with p=7.6x10^{-8} and rs17635227 with p=1.4x10^{-6} located, respectively, at chromosome 9 (120.2 Mb) between the Deleted in breast cancer 1 (*DBC1*) and Toll-like receptor 4 (*TLR4*) genes and at chromosome 14 (83.3 Mb) in a gene desert (Supplementary Figure S2 and Table S4A). Furthermore, of note was a SNP identified within the top 15 of independent loci of the meta-analyses, located at chromosome 19 (18.8 MB), annotated 1 kb downstream of the Cartilage Oligomeric Protein (*COMP)* gene (rs10038; p=2.3x10^{-5}) that was as positional candidate included in the replication stage of our study.

Replication of the 4 identified SNPs within the Dutch early OA study CHECK (N=964) indicated suggestive evidence for independent association with sCOMP levels only for SNP rs691461 (p=0.066; Supplementary Table S4B). When we added this result to the meta-analysis the association for the G allele (allele frequency: 0.502) of rs691461 with lower sCOMP levels became stronger (beta: -0.16, p=1.7x10^{-12}) without evidence for significant heterogeneity (I^2=23.8%; p=0.3; Figure 2A). Although the independent effect of rs10038, located close to the *COMP* gene, was not significant in the CHECK cohort (p=0.106), when we added this result to the meta-analysis the associationfor the A allele (allele frequency: 0.334) with increased levels of sCOMP became stronger (beta: 0.11; p=7.1x10^{-6}).

GWAS meta-analysis for uCTX-II. In the uCTX-II meta-analysis a total of 1,836,485 SNPs (directly genotyped or imputed in at least 4 studies) were tested for association. The genomic control inflation factor showed no systematic inflation (λ_{GC}=1.0132; Supplementary Figure S3). Two independent variants reached the pre-specified threshold for replication, respectively, rs1983474 (p=1.0x10^{-6}) located at chromosome 8 (4.4 Mb) within the CUB and sushi domain-containing protein 1 (*CSMD1*; Figure 3B)

Table 2: Tophits for meta-analyses of GWAS (Legend: CHR: chromosome; POS: chromosomal position; EA: Effect Allele; EAF: Effect Allele Frequency; BETA: effect of the indicated Effect Allele; SE: Standard Error; Dir.: Direction of the effect for the indicated effect allele in the respective studies (Chingford, GARP, LLS, RS-I –only for uCTX-II-, RS-II, TwinsUK, VIDEO, and CHECK).
*In CHECK rs8104411 was genotyped (pairwise LD (r2) between rs8104411 and rs10038 = 0.9).

SNP	CHR	POS	EA	EAF	BETA	*p*	Dir.	Gene
sCOMP								
rs10429475	9	120,165,567	T	0.454	0.11	1.42 x 10^{-6}	+++++++	*TLR4* *DBC1*
rs691461	10	17,933,215	G	0.502	-0.16	1.66 x 10^{-12}	- - - - - - -	*MRC1*
rs17635227	14	83,280,425	G	0.262	-0.09	1.49 x 10^{-4}	- - - - - - +	
rs10038*	19	18,753,728	A	0.334	0.11	7.11 x 10^{-6}	+++++++	*COMP*
uCTX-II								
rs4518370	5	82,216,062	G	0.750	-0.11	1.23 x 10^{-6}	- - - - - - + -	*TMEM1*
rs1983474	8	4,390,994	C	0.770	0.13	8.45 x 10^{-8}	++++++++	*CSMD1*

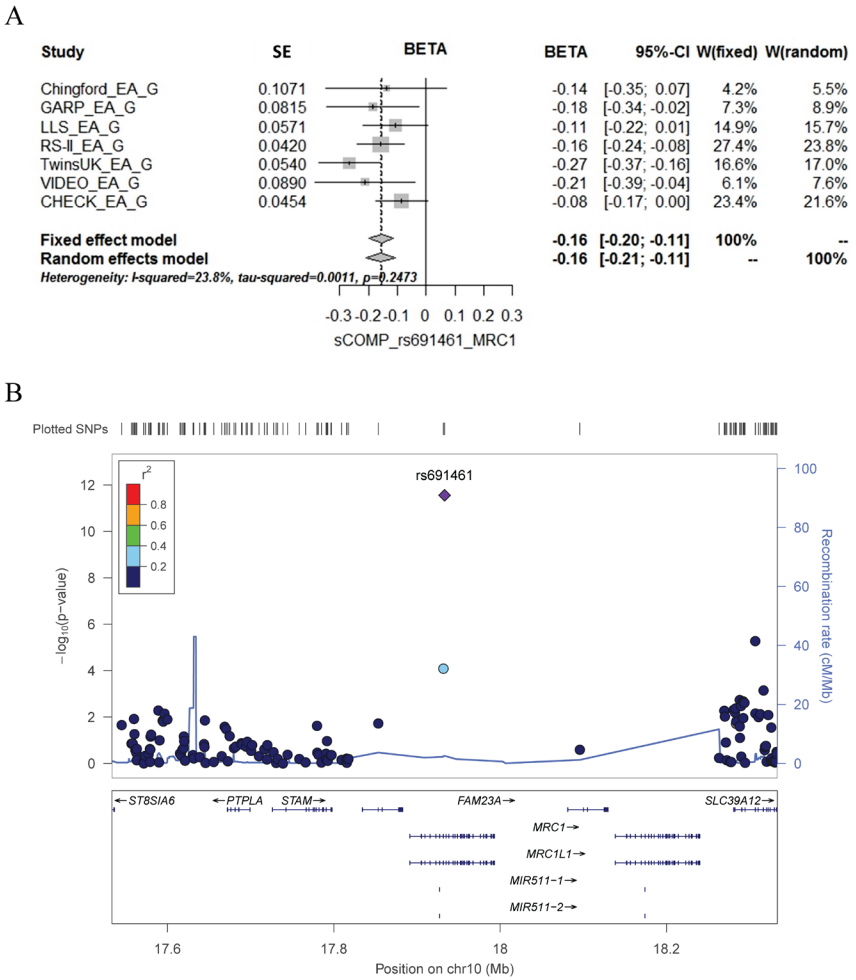
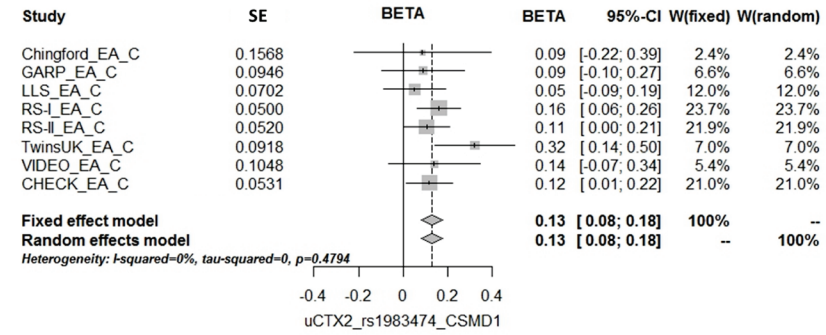


Figure 2. Association of sCOMP levels with rs691461 localized within the MRC1 gene. A. Forest plot of sCOMP meta-analysis for rs691461 in the MRC1 gene (p=1.7x10^{-12}). EA: Effect Allele; Chingford: Chingford 1000 Women Study; GARP: Genetics OsteoArthritis and Progression study Leiden; LLS: Leiden Longevity Study; RS-II: Rotterdam study cohort 2; VIDEO: randomized, placebo controlled trial of vitamin D replacement study; TwinsUK: the United Kingdom Adult Twin Registry; CHECK: Cohort Hip and Cohort Knee. **B.** LocusZoom plot showing positional genes of the sCOMP QTL SNP rs691461. Beta refers to standardized effect.

and rs4518370 (p=1.3x10^{-6}) located at chromosome 5 (82.2 Mb) close to 3 different genes (Supplementary Figure S4). In CHECK we observed significant independent association for rs1983474, (p=0.032; Supplementary Table S4B). When combining discovery and replication results of rs1983474, the association with uCTX-II became borderline genome-wide significant (beta: 0.13; p=8.5x10^{-8}) without evidence for heterogeneity (I^2=0%; p=0.5; Figure 3A).

Analysis for association with OA phenotypes and expression of annotated genes in cartilage. We explored whether the 6 loci identified in this study by virtue of

A



B

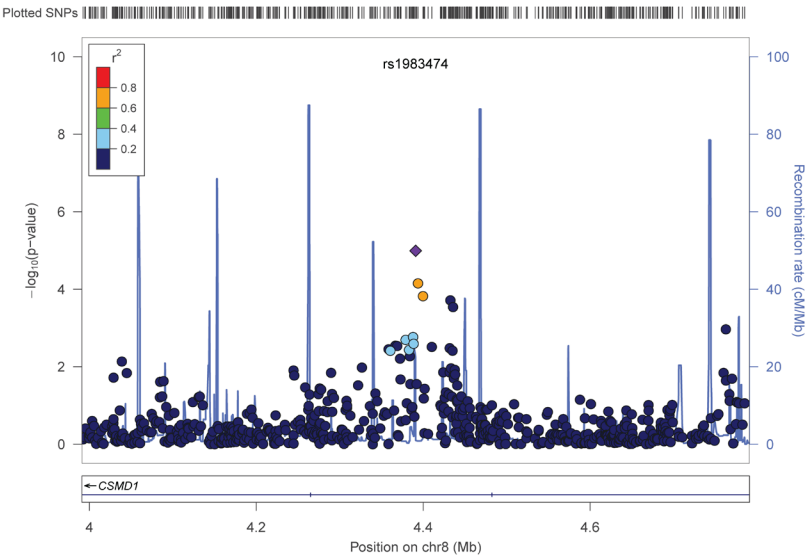


Figure 3. Suggestive evidence for association of rs1983474 with uCTX-II levels. A. Forest plot of uCTX-II meta-analysis for rs1983474 within the CSMD1 gene (p=8.5x10^{-8}). EA: Effect Allele; Chingford: Chingford 1000 Women Study; GARP: Genetics OsteoArthritis and Progression study Leiden; LLS: Leiden Longevity Study; RS-I: Rotterdam study cohort 1; RS-II: Rotterdam study cohort 2; VIDEO: randomized, placebo controlled trial of vitamin D replacement study; TwinsUK: the United Kingdom Adult Twin Registry; CHECK: Cohort Hip and Cohort Knee. **B** LocusZoom plot showing positional genes of the uCTX-II QTL SNP rs1983474. Beta refers to standardized effect.

their association with either sCOMP or uCTX-II (Table 2) also associated with OA. For this, the TreatOA genome wide association scan dataset including all participants stratified for hip and knee OA was examined. As shown in Table 3, evidence for nominal significant association (p<0.05) was only found for rs10038, within the *COMP* gene, for hip OA (p=8.9x10^{-3}; odds ratio (OR): 1.07 per allele, 95% confidence interval (CI) 1.02-1.13). Expression of the identified annotated genes of Table 2 in articular cartilage and their responsiveness to OA pathophysiology was examined in a

microarray mRNA expression dataset of human matched OA and preserved articular cartilage obtained from 33 individuals that underwent joint replacement surgery due to end stage OA. As expected, very high expression of *COMP* was observed in cartilage tissues (relative levels of expression are within the highest quartile of the dataset; Table 3) reflecting its relative importance herein. Furthermore, expression of *COMP* appeared subtle but significantly increased (1.1-fold) in cartilage harvested near the OA lesion (OA affected cartilage) as compared to cartilage distal to the OA lesion (preserved cartilage; p=3.6x10^{-3}). In contrast, low overall expression in articular cartilage and no differential expression between OA affected and preserved cartilage was detected for the other genes (Table 3).

DISCUSSION

In the current study, we have established novel genetic variants influencing levels of the well-known OA biochemical markers sCOMP and uCTX-II by means of a large-scale GWAS meta-analysis design. A genome-wide significant locus within the *MRC1* gene (rs691461) was associated with variation in the levels of sCOMP (p = 1.7 x10^{-12} in the combined analysis of the discovery and follow-up stage). The effect allele G with a frequency of 0.502 was associated with a decrease of sCOMP in a fixed-effect model (beta: -0.16). *MRC1*, encoding the Mannose Receptor C type 1, is a membrane receptor highly expressed by macrophages. It has been shown that *MRC1* mediates specific recognition of sugar residues to allow endocytosis of glycoproteins.[37] Since carriers of the risk allele have decreased levels of sCOMP this could reflect enhanced phagocytosis of serum glycoproteins such as sCOMP. Such a mechanism, likely independent of OA, could also explain the fact that *MRC1* is only lowly expressed in articular cartilage and that the *MRC1* SNP was not associated to OA phenotypes (Table 3). As such, this association could negatively affect the sensitivity of sCOMP as biomarker of OA in a clinical setting. On the other hand, of note was the signal detected with rs10038 located 1 kb downstream of the *COMP* gene with the sCOMP levels (Table 2). In this case, we showed that the relative high *COMP* mRNA expression in articular cartilage was responsive to the OA process and that the SNP rs10038 was associated to hip OA (p=8.9x10^{-3} with OR: 1.07 for the effect allele A). Given these data, the sCOMP levels in this respect appear to act as an endophenotype of OA and, therefore, the *COMP* gene itself may be considered as a potential new OA susceptibility gene which warrants further investigation. These results highlight the virtue and power of our study design.

For uCTX-II levels SNP rs1983474 located in *CSMD1* showed borderline genome wide significance in the combined analyses of the discovery and replication stage (p=8.5x10^{-8}). The C-allele, with a frequency of 0.770, was associated with increased uCTX-II (beta: 0.13). We did not find association between rs1983474 and hip or knee OA. A previous study identified association between OA and another SNP in *CSMD1*, rs9657371,[38] with negligible linkage disequilibrium with rs1983474 (r^2=0.03). However, exploration of the TreatOA GWAS dataset for hip and for knee OA showed no association for rs9657371

Table 3: Association of identified loci with hip and knee OA and expression of annotated genes in cartilage. (Legend: CHR: chromosome; POS: chromosomal position; OR: odds ratio; CI_{95}: 95% confidence interval; Dir: Direction of the effect for the indicated effect allele; AveExpr: Average expression in cartilage (mean=7.4 (6.6-14.9)); p_{expr}: P-value for differential expression between OA affected and preserved cartilage; FC: fold change in OA affected cartilage respective to preserved; NA: not available).

				HIP OA				KNEE OA					CARTILAGE		
SNP	**CHR**	**POS**	**EA**	**OR**	**CI_{95}**	***p***	**Dir.**	**OR**	**CI_{95}**	***p***	**Dir**	**Gene**	**AveExpr**	**p_{expr}**	**FC**
sCOMP															
rs10429475	9	120165567	T	0.96	[0.91-1.01]	8.33×10^{-2}	-	0.99	[0.95-1.04]	6.22×10^{-1}	-	*TLR4*	7.0	0.372	0.98
												DBC1	6.9	0.872	1.00
rs691461	10	17933215	G	0.97	[0.92-1.02]	2.22×10^{-1}	-	1.04	[0.99-1.09]	1.20×10^{-1}	+	*MRC1*	6.9	0.172	1.02
rs17635227	14	83280425	G	0.99	[0.94-1.04]	6.34×10^{-1}	-	1.02	[0.97-1.08]	3.57×10^{-1}	+	-			
rs10038	19	18753728	A	1.07	[1.02-1.13]	8.92×10^{-3}	+	1.03	[0.98-1.09]	1.84×10^{-1}	+	*COMP*	14.7	0.004	1.10
uCTX-II:															
rs4518370	5	82216062	G	0.95	[0.90-1.01]	9.70×10^{-2}	-	1.03	[0.98-1.08]	3.08×10^{-1}	+	*TMEM167A*	7.4	0.678	1.01
												SCARNA18	NA	NA	NA
												XRCC4	7.0	0.589	0.99
rs1983474	8	4390994	C	0.99	[0.94-1.05]	8.15×10^{-1}	-	1.01	[0.96-1.07]	6.17×10^{-1}	+	*CSMD1*	6.7	0.114	0.98

(data not shown). The rat homolog of *CSMD1* was identified in a search for proteins of the 'regulators of complement activation' (RCA) family,[39] and it was shown that CSMD1 blocked the classical complement pathway activation.[39, 40] It may therefore be speculated that innate dysregulated expression of *CSMD1* results in a (slight) activation of the complement cascade in articular cartilage independent of OA, leading to changes in the release of CTX-II. It should be noted, however, that in articular cartilage *CSMD1* gene expression was relatively low (Table 3). Altogether, thorough additional investigation is necessary to obtain mechanistic insight in this locus with respect to uCTX2 levels.

Our study has some limitation. Despite the large number of samples that we have assembled, small genetic effects could well have been missed in our analyses with the biomarker levels or OA phenotypes. Power may have eroded due to measurement errors in biomarker determinations or due to heterogeneity in the marker levels that may have been introduced by the OA phenotype of some of the subjects in participating studies. However, we expect this effect to be negligible since the meta-analyses were performed using the extracted standardized residuals from regression with mean = 0 and SD = 1 for all studies included. Alternatively, results may have been affected by other biological differences such as variation in urinary clearance of the biomarkers, body mass index, and previous physical exercise. Despite applied quality control (e.g. HWE), the region surrounding the *MRC1* locus appears different from the average genomic structure which may have affected our association (Figure 2B). In the follow-up phase, however, *de novo* genotyping was performed on a different genotype platform (Taqman assay) and confirmed association with borderline significance (p=6.6x10^{-2}) making such influences less likely. With respect to the gene expression analysis, we studied only expression of the genes annotated to the identified locus. We, therefore, cannot exclude the possibility that the identified SNPs affect the expression of more distant genes. In addition, the expression of *COMP* was not validated by an independent assay such as RT-qPCR.

In conclusion, by applying meta-analysis of GWAS we are the first to identify a locus within the *CSMD1* gene influencing innate levels of uCTX-II with borderline genome-wide significance. Moreover, we identified 2 loci influencing the sCOMP levels, located within respectively the *MRC1* and the *COMP* gene. The data and function of *MRC1* indicate that, apart from OA pathophysiology, also OA independent processes are likely to affect sCOMP levels. Taking such effects into account when applying sCOMP as OA biomarker increased sensitivity of sCOMP may be accomplished. In this respect, we advocate that the recently published meta-analysis on uCTX-II and sCOMP as OA biomarkers[10] should now be analyzed by taking into account the identified genetic factors.

ACKNOWLEDGEMENTS

The authors want to thank all the TreatOA participants and collaborating teams for their contribution to the study. We thank Dr. E. Zeggini as head of the arcOGEN analysis team. The biomarker work was funded by the European Commission framework 7 programme grant nr. 200800, TreatOA.

Chinford and VIDEO A proportion of the Chingford and all of the VIDEO DNAs were GWAS genotyped as part of the arcOGEN study which was funded by Arthritis Research UK, grant number 18030.

GARP (The Genetics osteoarthritis and Progression Study), LLS (Leiden Longevity Study) and RAAK (Research Arthritis and Articular Cartilage) The Leiden University Medical Centre, the Dutch Arthritis Association and Pfizer Inc., Groton, CT, USA support the GARP study, whilst RAAK and LLS was supported by the Netherlands Organization of Scientific Research (MW 904-61-095, 911-03-016, 917-66-344 and 911-03-012), Leiden University Medical Centre, and by the 'Centre of Medical System Biology' and the 'Netherlands Consortium of Healthy Aging' (NCHA; project nr. 050-060-810) in the framework of the Netherlands Genomics Initiative (NGI) and by BBMRI-NL, a Research Infrastructure financed by the Dutch government (NWO 184.021.007). Furthermore, the research leading to these results has received funding from the Dutch Arthritis Association (DAA 2010_017) and the European Union's Seventh Framework Programme (FP7/2007-2011) under grant agreement nr. 259679**.**

RS (The Rotterdam Study) The GWA study was funded by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the NGI / NCHA (project nr. 050-060-810). We thank Pascal Arp, Mila Jhamai, Dr Michael Moorhouse, Marijn Verkerk, and Sander Bervoets for their help in creating the GWAS database. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are very grateful to the participants and staff from the Rotterdam Study, the participating general practitioners and the pharmacists. We would like to thank Dr. Tobias A. Knoch, Luc V. de Zeeuw, Anis Abuseiris, and Rob de Graaf as well as their institutions the Erasmus Computing Grid, Rotterdam, The Netherlands, and especially the national German MediGRID and Services@MediGRID part of the German D-Grid, both funded by the German Bundesministerium fuer Forschung und Technology under grants nr. 01 AK 803 A-H and nr. 01 IG 07015 G for access to their grid resources.

TwinsUK (The UK Adult Twin Registry) This study was funded by the Wellcome Trust (Grant ref. 079771); European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F2-2008-ENGAGE and the European Union FP-5 GenomEUtwin Project (QLG2-CT-2002-01254) and Framework 6 Project EUroClot. The study also receives support from the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London. We thank the staff from the TwinsUK, the DNA Collections and Genotyping Facilities at the Wellcome Trust Sanger Institute for sample preparation; Quality Control of the Twins UK cohort for

genotyping (in particular Amy Chaney, Radhi Ravindrarajah, Douglas Simpkin, Cliff Hinds, and Thomas Dibling); Paul Martin and Simon Potter of the DNA and Genotyping Informatics teams for data handling; Le Centre National de Génotypage, France, led by Mark Lathrop, for genotyping; Duke University, North Carolina, USA, led by David Goldstein, for genotyping; and the Finnish Institute of Molecular Medicine, Finnish Genome Center, University of Helsinki, led by Aarno Palotie. Nicole Soranzo acknowledges financial support from the Wellcome Trust (Grant 091746/Z/10/Z).

Author contributions

All authors have made substantial contributions to the completion of this study: Study concept and design: JI, IM, AMV, and JvM. Acquisition of material and data: all authors.

Data analysis: AMV, JvM, EE, YR, and IM. Interpretation of the data: all authors. Preparation of the manuscript: all authors. Critical reviewing and approval of the manuscript: all authors.

Conflict of interest

Authors have no conflict of interest to declare.

Author affiliations

1. Department of Molecular Epidemiology, LUMC, Leiden, The Netherlands;
2. The Netherlands Genomics Initiative-Sponsored Netherlands Consortium for Healthy Aging, Leiden and Rotterdam, The Netherlands;
3. Department of Twin Research & Genetic Epidemiology, King's College London, London, UK;
4. NIHR Musculoskeletal Biomedical Research Unit, University of Oxford, Oxford, UK;
5. Arthritis Research UK, Sport, exercise and Osteoarthritis Centre of excellence;
6. Department of Rheumatology, Nordic Bioscience, Herlev, Denmark;
7. Department of Gerontology and Geriatrics, LUMC, Leiden, The Netherlands;
8. Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA;
9. The Framingham Heart Study, Framingham, MA, USA;
10. Institute of Molecular and Cell Biology and Estonian Genome Center, University of Tartu, Tartu, Estonia;
11. Department of Endocrinology, Children's Hospital Boston, Boston, MA, USA;
12. Broad Institute, Cambridge, MA, USA;
13. Department of Hygiene & Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece;
14. Clinical Epidemiology Unit, Boston University School of Medicine, Boston, MA, USA;
15. Stanford Prevention Research Center, Department of Medicine, and Department of Health Research and Policy, Stanford University School of Medicine, and Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, USA
16. Department of Rheumatology & Clinical Epidemiology, LUMC, Leiden, The Netherlands;
17. Department of Rheumatology & Clinical Immunology, UMC Utrecht, Utrecht, The Netherlands;
18. Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom;
19. Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands;
20. Global Analytical Science, Takeda Pharmaceuticals International Co., Cambridge, MA, USA;
21. Academic Rheumatology, University of Nottingham, Nottingham, United Kingdom

REFERENCES

1 Bay-Jensen AC, Slagboom E, Chen-An P, Alexandersen P, Qvist P, Christiansen C, *et al.* Role of hormones in cartilage and joint metabolism: understanding an unhealthy metabolic phenotype in osteoarthritis. *Menopause* 2013;**20**:578-86.

2 Valdes AM, Spector TD. The genetic epidemiology of osteoarthritis. *Curr Opin Rheumatol* 2010;**22**:139-43.

3 Lotz M, Martel-Pelletier J, Christiansen C, Brandi ML, Bruyere O, Chapurlat R, *et al.* Republished: Value of biomarkers in osteoarthritis: current status and perspectives. *Postgrad Med J* 2014;**90**:171-8.

4 Attur M, Krasnokutsky-Samuels S, Samuels J, Abramson SB. Prognostic biomarkers in osteoarthritis. *Curr Opin Rheumatol* 2013;**25**:136-44.

5 Mobasheri A. Osteoarthritis year 2012 in review: biomarkers. *Osteoarthritis Cartilage* 2012;**20**:1451-64.

6 Kraus VB, Burnett B, Coindreau J, Cottrell S, Eyre D, Gendreau M, *et al.* Application of biomarkers in the development of drugs intended for the treatment of osteoarthritis. *Osteoarthritis Cartilage* 2011;**19**:515-42.

7 van Spil WE, Drossaers-Bakker KW, Lafeber FP. Associations of CTX-II with biochemical markers of bone turnover raise questions on its tissue origin: data from CHECK, a cohort study of early osteoarthritis. *Ann Rheum Dis* 2013;**72**:29-36.

8 Meulenbelt I, Kloppenburg M, Kroon HM, Houwing-Duistermaat JJ, Garnero P, Hellio-Le Graverand MP, *et al.* Clusters of biochemical markers are associated with radiographic subtypes of osteoarthritis (OA) in subject with familial OA at multiple sites. The GARP study. *Osteoarthritis Cartilage* 2007;**15**:379-85.

9 Meulenbelt I, Kloppenburg M, Kroon HM, Houwing-Duistermaat JJ, Garnero P, Hellio le Graverand MP, *et al.* Urinary CTX-II levels are associated with radiographic subtypes of osteoarthritis in hip, knee, hand, and facet joints in subject with familial osteoarthritis at multiple sites: the GARP study. *Ann Rheum Dis* 2006;**65**:360-5.

10 Valdes AM, Meulenbelt I, Chassaing E, Arden NK, Bierma-Zeinstra S, Hart D, *et al.* Large scale meta-analysis of urinary C-terminal telopeptide, serum cartilage oligomeric protein and matrix metalloprotease degraded type II collagen and their role in prevalence, incidence and progression of osteoarthritis. *Osteoarthritis Cartilage* 2014.

11 Chen HC, Kraus VB, Li YJ, Nelson S, Haynes C, Johnson J, *et al.* Genome-wide linkage analysis of quantitative biomarker traits of osteoarthritis in a large, multigenerational extended family. *Arthritis Rheum* 2010;**62**:781-90.

12 Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, *et al.* Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet* 2011;**43**:513-8.

13 Riyazi N, Meulenbelt I, Kroon HM, Ronday KH, Hellio le Graverand MP, Rosendaal FR, *et al.* Evidence for familial aggregation of hand, hip, and spine but not knee osteoarthritis in siblings with multiple joint involvement: the GARP study. *Ann Rheum Dis* 2005;**64**:438-43.

14 Schoenmaker M, de Craen AJ, de Meijer PH, Beekman M, Blauw GJ, Slagboom PE, *et al.* Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* 2006;**14**:79-84.

15 Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol* 1991;**7**:403-22.

16 Hofman A, Breteler MM, van Duijn CM, Janssen HL, Krestin GP, Kuipers EJ, *et al.* The Rotterdam Study: 2010 objectives and design update. *Eur J Epidemiol* 2009;**24**:553-72.

17 Spector TD, Williams FM. The UK Adult Twin Registry (TwinsUK). *Twin Res Hum Genet* 2006;**9**:899-906.

18 Spector TD, McCloskey EV, Doyle DV, Kanis JA. Prevalence of vertebral fracture in women and the relationship with bone density and symptoms: the Chingford Study. *J Bone Miner Res* 1993;**8**:817-22.

19 Roemer FW, Kassim JM, Guermazi A, Thomas M, Kiran A, Keen R, *et al.* Anatomical distribution of synovitis in knee osteoarthritis and its association with joint effusion assessed on non-enhanced and contrast-enhanced MRI. *Osteoarthritis Cartilage* 2010;**18**:1269-74.

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

21 Evangelou E, Valdes AM, Kerkhof HJ, Styrkarsdottir U, Zhu Y, Meulenbelt I, *et al.* Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. *Ann Rheum Dis* 2011;**70**:349-55.

22 Ramos YF, den HW, Bovee JV, Bomer N, van der Breggen R, Lakenberg N, *et al.* Genes Involved in the Osteoarthritis Process Identified through Genome Wide Expression Analysis in Articular Cartilage; the RAAK Study. *PLoS One* 2014;**9**:e103056.

23 Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 2010;**34**:816-34.

24 Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007;**39**:906-13.

25 Wesseling J, Dekker J, van den Berg WB, Bierma-Zeinstra SM, Boers M, Cats HA, *et al.* CHECK (Cohort Hip and Cohort Knee): similarities and differences with the Osteoarthritis Initiative. *Ann Rheum Dis* 2009;**68**:1413-9.

26 Uh HW, Deelen J, Beekman M, Helmer Q, Rivadeneira F, Hottenga JJ, *et al.* How to deal with the early GWAS data when imputing and combining different arrays is necessary. *Eur J Hum Genet* 2012;**20**:572-6.

27 Wallace C, Chapman JM, Clayton DG. Improved power offered by a score test for linkage disequilibrium mapping of quantitative-trait loci by selective genotyping. *Am J Hum Genet* 2006;**78**:498-504.

28 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;**81**:559-75.

29 Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;**26**:2190-1.

30 Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999;**55**:997-1004.

31 Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* 2007;**23**:1294-6.

32 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;**7**:177-88.

33 Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;**21**:1539-58.

34 Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. *Ann Intern Med* 1997;**127**:820-6.

35 Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010;**26**:2336-7.

36 Styrkarsdottir U, Thorleifsson G, Helgadottir HT, Bomer N, Metrustry S, Bierma-Zeinstra S, *et al.* Severe osteoarthritis of the hand associates with common variants within the ALDH1A2 gene and with rare variants at 1p31. *Nat Genet* 2014;**46**:498-502.

37 Largent BL, Walton KM, Hoppe CA, Lee YC, Schnaar RL. Carbohydrate-specific adhesion of alveolar macrophages to mannose-derivatized surfaces. *J Biol Chem* 1984;**259**:1764-9.

38 Elliott KS, Chapman K, Day-Williams A, Panoutsopoulou K, Southam L, Lindgren CM, *et al.* Evaluation of the genetic overlap between osteoarthritis with body mass index and height using genome-wide association scan data. *Ann Rheum Dis* 2013;**72**:935-41.

39 Kraus DM, Elliott GS, Chute H, Horan T, Pfenninger KH, Sanford SD, *et al.* CSMD1 is a novel multiple domain complement-regulatory protein highly expressed in the central nervous system and epithelial tissues. *J Immunol* 2006;**176**:4419-30.

40 Escudero-Esparza A, Kalchishkova N, Kurbasic E, Jiang WG, Blom AM. The novel complement inhibitor human CUB and Sushi multiple domains 1 (CSMD1) protein promotes factor I-mediated degradation of C4b and C3b and inhibits the membrane attack complex assembly. *FASEB J* 2013.

Supplementary methods

STUDIES:

Chingford Study:

This study is a prospective population-based longitudinal cohort, which includes women derived from the age/sex register of a large general practice in North London[1,2]. The study design and rationale have been described elsewhere in detail[3]. The Guy's St. Thomas' Trust and the Waltham Forest Trust ethics committees approved the study protocol. After study procedures were explained to participants, written informed consent was given by each participant.

Genetics OsteoARthritis and Progression (GARP) Study:

The ongoing GARP study, which consists of Caucasian sib pairs of Dutch ancestry affected predominantly with symptomatic OA at multiple sites, is primarily aimed at the identification of genetic determinants of OA susceptibility and progression. Details of the ascertainment has been described elsewhere.[1] Symptomatic OA of a joint within the study was defined as having symptoms of OA in addition to radiographic signs. Informed consent was obtained from all patients. The GARP study consists of 382 subjects (312 females and 70 men) which meet all inclusion criteria, in this paper we included 10 additional subjects with OA at least at one major joint site which had marker levels available.

Leiden Longevity Study (LLS):

Between 2002 and 2006, the LLS included 421 Caucasian families were consisting of long-lived siblings with their offspring and partners of their offspring. Families were eligible to participate in the study when at least two long-lived siblings were alive and over 89 years old for males and over 91 years for females, representing less than 0.5% of the Dutch population in 2001. There were no selection criteria on health or demographic characteristics. Blood samples were taken at baseline, non-fasted between 0930 and 1700h. Urine samples were collected at the time of drawing blood and stored at -80°C. In this analysis we included a total of 329 offspring of long-lived siblings and 329 of their partners.

Rotterdam Study I and II (RS-I and RS-II):

The Rotterdam Study is a population-based prospective cohort study ongoing since 1990 to study determinants of chronic disabling disease [24]. Two subpopulations of the Rotterdam Study were used in the current manuscript, RS-I and RS-II. The Rotterdam Study I (RSI) is the first cohort of 7,983 persons living in the Ommoord district of Rotterdam in the Netherlands. All subjects were aged 55 years and older and recruitment started in 1990. The Rotterdam Study II (RSII) started in 1999 when 3,011 participants moved into the study since they became 55 years of age or moved into the study district. The medical ethics committee of Erasmus University Medical School approved the study and written informed consent was obtained from each participant.

TwinsUK:

The study participants were white monozygotic and dizygotic twin pairs from the TwinsUK adult twin registry, a group used to study the heritability and genetics of age-related diseases.

These unselected twins were recruited from the general population through national media campaigns in the United Kingdom. Ethics approval was obtained from the Guy's and St. Thomas' Hospital Ethics Committee. Written informed consent was obtained from every participant

VIDEO:

This study is a placebo-controlled randomized controlled trial of 800 unit's cholecalciferol in men and women with knee OA. In total, 477 cases are included in this trial which are classified as knee OA patients according to the Kellgren and Lawrence grading system[13] (OA defined as one definite osteophyte) and patients had to suffer from knee pain and have a medial joint space width > 1mm. The study was approved by the local research ethics committee, and all subjects provided written, informed consent prior to participation.

CHECK:

Cohort Hip and Cohort Knee (CHECK) is a long-term follow-up study (10 years) of early OA subjects with pain and/or stiffness of one and/or both knee(s) and/or hip(s), that had never or not longer than 6 months ago visited a general practitioner for these complaints for the first time. At baseline, CHECK subjects (79% female) were aged 56±5 years (mean ±SD) and had a median body mass index (BMI) of 25.5 kg/m^2.[2]

RAAK study:

The Research Arthritis and Articular Cartilage (RAAK) study is an ongoing study in which joint materials are collected (cartilage, bone and where available ligaments) and mesenchymal stem cells (hip joints only) of patients and controls in the Leiden University Medical Center and collaborating outpatient clinics in the Leiden area. In the current study we used sample pairs for 33 donors undergoing joint replacement surgery for primary OA, 22 hips and 11 knees were included in the study. At the moment of collection (within 2 hours following surgery) tissue was washed extensively with phosphate buffered saline (PBS) to decrease the risk of contamination by blood. Joints were assessed macroscopically for OA related damage such as the loss (depth and hardness) of articular cartilage, surface integrity as determined by visible fibrillation or crack formation and color/whiteness of the cartilage,[3, 4] and collected separately from OA affected and preserved regions around the weight-bearing area of the joint. Care was taken to avoid contamination with bone or synovium. Collected cartilage was snap frozen in liquid nitrogen and stored at -80°C prior to RNA extraction.

References

1 Riyazi N, Meulenbelt I, Kroon HM, Ronday KH, Hellio le Graverand MP, Rosendaal FR, *et al.* Evidence for familial aggregation of hand, hip, and spine but not knee osteoarthritis in siblings with multiple joint involvement: the GARP study. *Ann Rheum Dis* 2005;**64**:438-43.

2 Wesseling J, Dekker J, van den Berg WB, Bierma-Zeinstra SM, Boers M, Cats HA, *et al.* CHECK (Cohort Hip and Cohort Knee): similarities and differences with the Osteoarthritis Initiative. *Ann Rheum Dis* 2009;**68**:1413-9.

3 Tsuritani K, Takeda J, Sakagami J, Ishii A, Eriksson T, Hara T, *et al.* Cytokine receptor-like factor 1 is highly expressed in damaged human knee osteoarthritic cartilage and involved in osteoarthritis downstream of TGF-beta. *Calcif Tissue Int* 2010;**86**:47-57.

4 Geyer M, Grassel S, Straub RH, Schett G, Dinser R, Grifka J, *et al.* Differential transcriptome analysis of intraarticular lesional vs intact cartilage reveals new candidate genes in osteoarthritis pathophysiology. *Osteoarthritis Cartilage* 2009;**17**:328-35.

Supplementary Table S1. Study design, number of cases and controls, and sample quality control for the GWAS cohorts (Legend: Chingford: Chingford 1000 Women Study; GARP: Genetics OsteoArthritis and Progression study Leiden; LLS: Leiden Longevity Study; RS: Rotterdam study (I and II); VIDEO: randomized, placebo controlled trial of vitamin D replacement study; TwinsUK: the United Kingdom Adult Twin Registry)

Study Short name	Sample QC Call rate	Other exclusions	References
Chingford	NA	1) Missing sCOMP or uCTX2 levels	Spector *et al.* 1993[1]
GARP	> 98%	1) Missing sCOMP or uCTX2 levels 2) MAF ≥ 0.05	Riyazi *et al.* (2005)[2]
LLS	≥ 95%	1) Missing sCOMP or uCTX2 levels 2) MAF ≥ 0.01 2) Latent clustering of genotypes because of population substructure was assessed by pairwise identity-by-state (IBS) distance using Graphical Relationship Representation (GRR) [http://bioinformatics.well.ox.ac.uk/GRR, (Abecasis et al., 2001)].	Schoenmaker *et al.* 2006[3]
RS-I	≥ 97.5%	1) Gender mismatch with typed Xlinked markers; 2) Excess autosomal heterozygosity > 0.336~FDR > 0.1%; 3) Duplicates and/or 1st or 2nd degree relatives using IBS probabilities > 97% from PLINK; 4) Ethnic outliers using IBS distances > 3SD from PLINK; 5) Missing age, and BMI information. 1) Missing sCOMP or uCTX2 levels	Hofman *et al.* 1991[4]
RS-II	≥ 97.5%	1) Gender mismatch with typed X-linked markers; 2) Excess autosomal heterozygosity (F < -0.055); 3) Duplicates and/or 1st degree relatives using IBD piHAT > 40% from PLINK; 4) Ethnic outliers using IBS distances > 4SD mean HapMap CEU cluster from PLINK; 5) Missing sCOMP and uCTX2	Hofman *et al.* 2009[5]
TwinsUK	≥ 95%	1) Heterozygosity < 33% or > 67%; 2) Ethnic outliers; 3) Related individuals and duplicates; 4) Missing sCOMP or uCTX2 levels	Spector *et al* 2006[6]
VIDEO	NA	1) Missing sCOMP or uCTX2 levels	Roemer *et al* 2010[7]

Supplementary Table S2. Information on genotyping methods, quality control of SNPs, and imputation for GWAS cohorts.

Cohort	Genotyping					Imputation		
	Platform	Genotype Calling Algorithm	Inclusion Criteria			Imputation Software	Inclusion Criteria	
			MAF	Callrate	p-value for HWE		MAF	Imputation Quality
GARP	Illumina Human660W	Genome studio	> 5%	> 98%	> 0.0001	IMPUTE	> 5%	proper-info ≥ 0.40
Chingford	Human 610-Quad Beadchips as part of the arcOGEN GWA	Customised genotype calling algorithm[8]	NA	≥ 95%	> 0.05	NA	NA	NA
LLS	Illumina Human660W	Genome studio	≥ 1%	> 95%	> 0.0001	IMPUTE	> 5%	proper-info ≥ 0.40
RS-I	Illumina / HumanHap 550K V.3, Illumina / HumanHap 550K V.3 DUO;	BeadStudio Genecall	≥ 1%	≥ 97.5%	$>10^{-06}$	MACH	≥ 0%	(O/E)σ2 ratio ≥ 0.1 r^2-hat ≥ 0.30
RS-II	Illumina / HumanHap 550 V.3 DUO; Illumina / HumanHap 610 QUAD	Genomestudio Genecall	≥ 1%	≥ 97.5%	$>10^{-06}$	MACH	≥ 1%	(O/E)σ2 ratio ≥ 0.1 r^2-hat ≥ 0.30
TwinsUK	Illumina / HumanHap 300 & 550	Illuminus	≥ 1%	≥ 95.0%	$>10^{-06}$	IMPUTE	> 0%	proper-info ≥ 0.40
VIDEO	Human 610-Quad Beadchips as part of the arcOGEN GWA	Customised genotype calling algorithm[8]	NA	≥ 95%	> 0.05	NA	NA	NA

Supplementary Table S3. Information on statistical analysis for the GWAS cohorts.

Cohort	Genome wide Association Analyses	
	Analysis Software	**References Analysis Software**
Chingford	PLINK	Purcell *et al.* (2007)[9]
GARP	QTASSOC	In-house analysis program, developed to correct for familial dependence; H. W. Uh and Q. Helmer[10, 11]
LLS	QTASSOC	In-house analysis program, developed to correct for familial dependence; H. W. Uh and Q. Helmer[10, 11]
RS-I	GRIMP (MACH2DAT)	Estrada *et al.*(2009 [12] Li *et al.* (2006)[13]
RS-II	GRIMP (MACH2DAT)	Estrada *et al.*(2009)[14] Li *et al.* (2006)[13]
TwinsUK	PLINK	Purcell *et al.* (2007)[9]
VIDEO	PLINK	Purcell *et al.* (2007)[9]

Supplementary Table S4. Results of tophits for association in the discovery Meta-Analysis (A) and for replication in CHECK (B) (Legend: CHR: chromosome; POS: chromosomal position; EA: Effect Allele; EAF: Effect Allele Frequency; BETA: effect of the indicated Effect Allele; SE: Standard Error; Dir.: Direction of the effect for the indicated effect allele in the respective studies (A: Chingford, GARP, LLS, RS-I –only for uCTX2-, RS-II, TwinsUK, VIDEO, and B: CHECK)).
*Pairwise LD (r2) between rs8104411 and r10038 = 0.91

A

SNP	CHR	POS	EA	EAF	BETA	*p*	Dir.	Gene
sCOMP:								
rs10429475	9	120,165,567	T	0.454	0.14	7.61×10^{-8}	++++++	*TLR4, DBC1*
rs691461	10	17,933,215	G	0.502	-0.17	1.93×10^{-12}	- - - - - -	*MRC1*
rs17635227	14	83,280,425	G	0.262	-0.14	1.43×10^{-6}	- - - - - -	
rs10038*	19	18,753,728	A	0.334	0.12	2.30×10^{-5}	++++++	*COMP*
uCTX2:								
rs4518370	5	82,216,062	G	0.750	-0.13	1.27×10^{-6}	- - - - - - +	*TMEM167A,SCARNA18, XRCC4*
rs1983474	8	4,390,994	C	0.770	0.13	1.04×10^{-6}	+++++++	*CSMD1*

B

SNP	CHR	POS	EA	EAF	BETA	*p*	Dir.	Gene
sCOMP:								
rs10429475	9	120,165,567	T	0.462	0.01	0.850	+	*TLR4, DBC1*
rs691461	10	17,933,215	G	0.517	-0.08	0.066	-	*MRC1*
rs17635227	14	83,280,425	G	0.239	0.06	0.280	+	
rs8104411*	19	18,750,702	A	0.339	0.08	0.106	+	*COMP*
uCTX2:								
rs4518370	5	82,216,062	G	0.752	-0.06	0.220	-	*TMEM167A,SCARNA18, XRCC4*
rs1983474	8	4,390,994	C	0.764	0.12	0.032	+	*CSMD1*

References

1 Spector TD, McCloskey EV, Doyle DV, Kanis JA. Prevalence of vertebral fracture in women and the relationship with bone density and symptoms: the Chingford Study. *J Bone Miner Res* 1993;**8**:817-22.

2 Riyazi N, Meulenbelt I, Kroon HM, Ronday KH, Hellio le Graverand MP, Rosendaal FR, *et al.* Evidence for familial aggregation of hand, hip, and spine but not knee osteoarthritis in siblings with multiple joint involvement: the GARP study. *Ann Rheum Dis* 2005;**64**:438-43.

3 Schoenmaker M, de Craen AJ, de Meijer PH, Beekman M, Blauw GJ, Slagboom PE, *et al.* Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* 2006;**14**:79-84.

4 Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol* 1991;**7**:403-22.

5 Hofman A, Breteler MM, van Duijn CM, Janssen HL, Krestin GP, Kuipers EJ, *et al.* The Rotterdam Study: 2010 objectives and design update. *Eur J Epidemiol* 2009;**24**:553-72.

6 Spector TD, Williams FM. The UK Adult Twin Registry (TwinsUK). *Twin Res Hum Genet* 2006;**9**:899-906.

7 Roemer FW, Kassim JM, Guermazi A, Thomas M, Kiran A, Keen R, *et al.* Anatomical distribution of synovitis in knee osteoarthritis and its association with joint effusion assessed on non-enhanced and contrast-enhanced MRI. *Osteoarthritis Cartilage* 2010;**18**:1269-74.

8 Teo YY, Inouye M, Small KS, Gwilliam R, Deloukas P, Kwiatkowski DP, *et al.* A genotype calling algorithm for the Illumina BeadArray platform. *Bioinformatics* 2007;**23**:2741-6.

9 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;**81**:559-75.

10 Uh HW, Deelen J, Beekman M, Helmer Q, Rivadeneira F, Hottenga JJ, *et al.* How to deal with the early GWAS data when imputing and combining different arrays is necessary. *Eur J Hum Genet* 2012;**20**:572-6.

11 Wallace C, Chapman JM, Clayton DG. Improved power offered by a score test for linkage disequilibrium mapping of quantitative-trait loci by selective genotyping. *Am J Hum Genet* 2006;**78**:498-504.

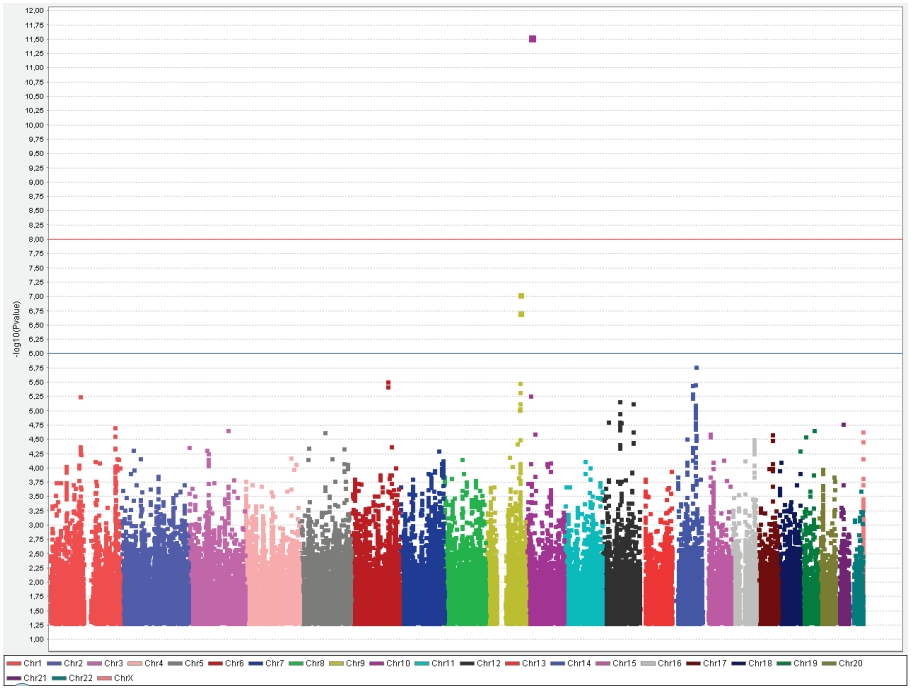
12 Estrada K, Abuseiris A, Grosveld FG, Uitterlinden AG, Knoch TA, Rivadeneira F. GRIMP: a web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. *Bioinformatics* 2009;**25**:2750-2.

13 Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 2010;**34**:816-34.

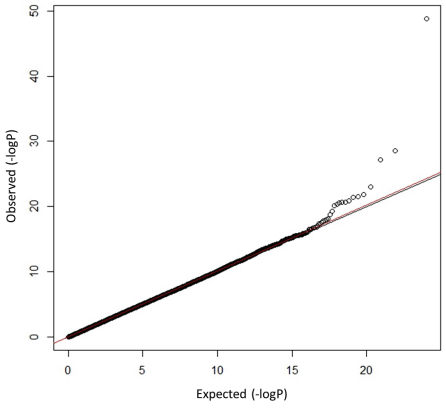
14 Estrada K, Abuseiris A, Grosveld FG, Uitterlinden AG, Knoch TA, Rivadeneira F. GRIMP: a web- and grid-based tool for high-speed analysis of large-scale genome-wide association using imputed data. *Bioinformatics* 2009;**25**:2750-2.

Supplementary Figure S1. A. Manhattan plot showing the significance of association for 1.9 million SNPs in the meta-analysis with sCOMP. SNPs are plotted on the x-axis according to their position on each chromosome. On the y-axis, the association p-values with sCOMP are shown (as –log 10 p-values). The blue solid horizontal line represents the preset p-Value threshold of 1 x 10^{-6}. The red horizontal line represents the p-Value threshold of 1 x 10^{-8} (the level for genome wide evidence of association). **B.** Overall Quantile-Quantile (QQ) plot of SNPs (estimated lamda: 1.0131).

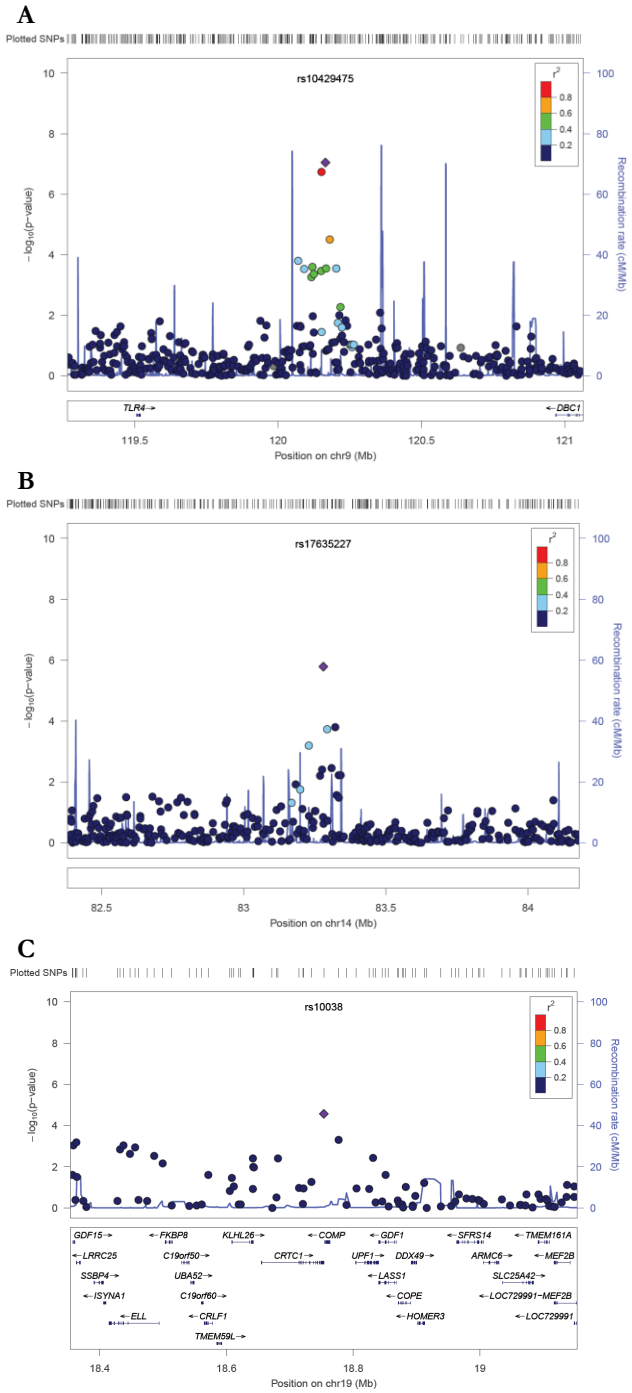
A



B

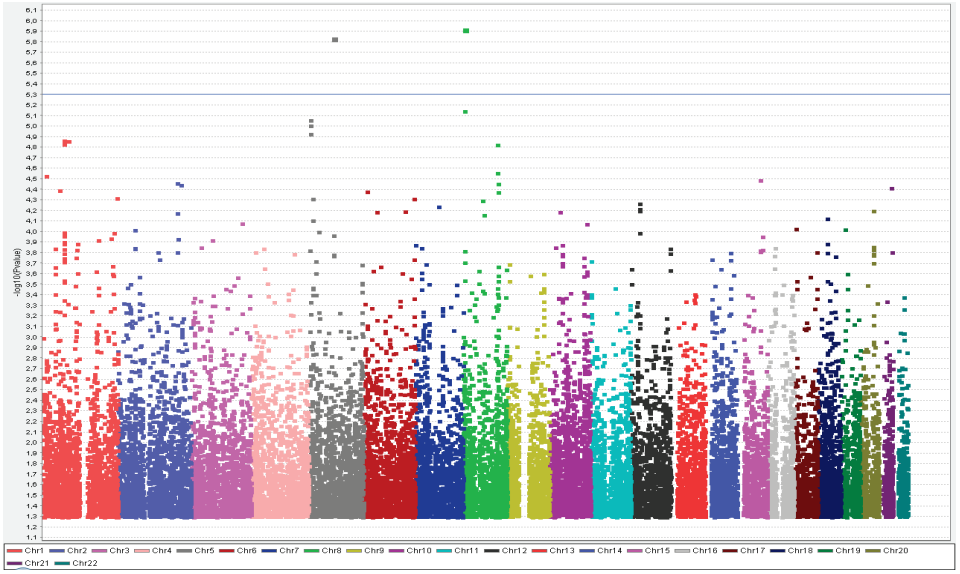


Supplementary Figure S2. LocusZoom plots showing positional genes of the sCOMP QTL SNPs rs10429475 (A), rs17635227 (B), and rs10038 (C).

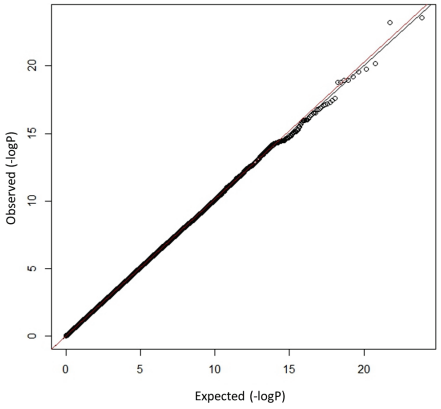


Supplementary Figure S3. A. Manhattan plot showing the significance of association for 1.8 million SNPs in the meta-analysis with uCTX2. SNPs are plotted on the x-axis according to their position on each chromosome. On the y-axis, the association p-values with uCTX2 are shown as –log 10 p-Value. The blue solid horizontal line represents the preset p-Value threshold of $1x10^{-6}$. **B.** Overall Quantile-Quantile (QQ) plot of SNPs (estimated lamda: 1.0132).

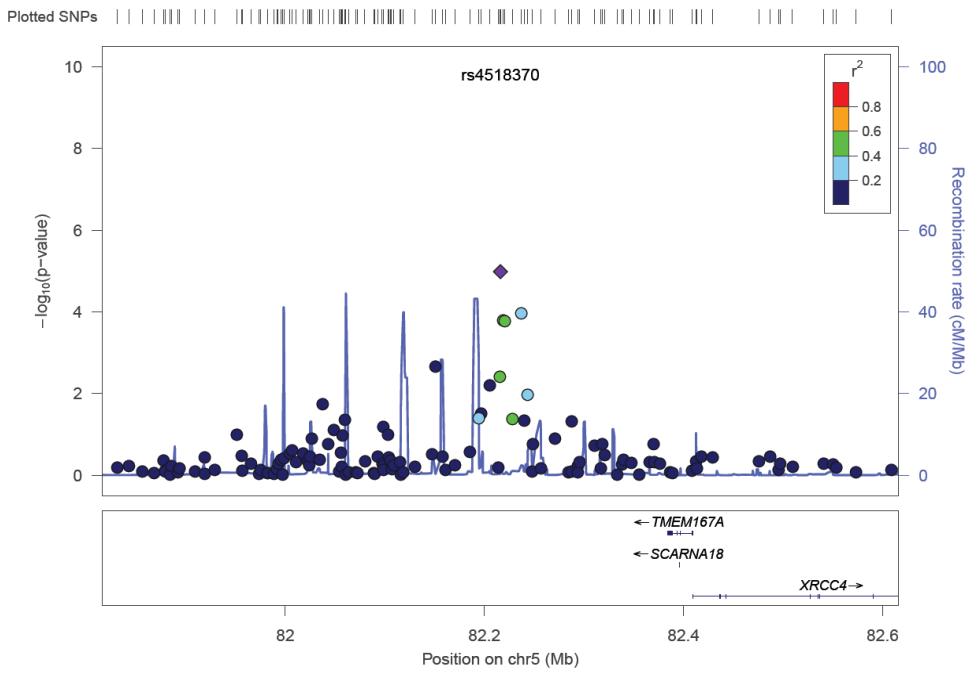
A



B



Supplementary Figure S4. LocusZoom plots showing positional genes of the uCTX-II QTL SNP rs4518370.



Genome-wide association study: identifying new genes and pathways conferring risk to susceptibility of osteoarthritis at multiple joint locations as defined in the GARP study

5

(Unpublished results)

Genome-wide association study: identifying new genes and pathways conferring risk to susceptibility of osteoarthritis at multiple joint locations as defined in the GARP study

YFM Ramos[1*], SD Bos[1,2], U Styrkarsdottir[3], K Panoutsopoulou[4], N Lakenberg[1], RGHH Nelissen[5], G Kloppenburg[6], H Kroon[7], HJM Kerkhof[8,2], AG Uitterlinden[8,2], J Loughlin[4,9], A Gonzalez[10], E Zeggini[4], I Jonsdottir[3], JB van Meurs[8,2], PE Slagboom[1,2] and I Meulenbelt[1,2]

[1]Dept of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; [2]The Netherlands Genomics Initiative, sponsored by the NCHA, Leiden-Rotterdam, The Netherlands; [3]deCODE Genetics, Reykjavik, Iceland; [4]arcOGEN Consortium, Cambridge, UK; [5]Dept of Orthopaedics, [6]Dept of Rheumatology, and [7]Dept. of Radiology, Leiden University Medical Center, Leiden, The Netherlands; [8]Dept. of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands; [9]Newcastle University, Musculoskeletal Research Group, Institute of Cellular Medicine, Newcastle, UK; [10]Hospital Clinico Universitario Santiago, Santiago de Compostela, Spain

ABSTRACT

Objective **Several compelling loci associated to osteoarthritis (OA) have been identified by genome wide association studies (GWAS), however, with small effects. Here, GWAS was conducted in the GARP study characterized by cases with familial, symptomatic generalized OA thereby aiming to discover susceptibility loci with larger effects.**

Methods **Following discovery GWAS, replication was performed in a 2-phase approach: loci with pre-defined threshold (p=1x10^{-5}; N=18) were analyzed in comparable cases with OA at ≥3 joints, followed by confirmation of 9 replicating SNPs in 4 additional studies with (mainly) single-site OA. Gene expression of annotated genes was explored in blood and OA affected and unaffected cartilage. Pathway analysis was performed for annotated genes of the discovery GWAS.**

Results **The strongest statistical evidence for association was found for rs347842 located within *SLC12A6*, a gene related to peripheral neuropathies (odds ratio OR=1.13 with 95% confidence interval CI$_{95}$=1.06–1.20 p=2.56x10^{-4}). Of note was a variant within *CDKAL1* showing small but consistent direction of effects in all studies (OR=1.12 with CI$_{95}$=1.03-1.21 p=4.8x10^{-3}). *CDKAL1* was found to be significantly differently expressed in blood of OA patients. Pathway analysis in the discovery stage showed significant enrichment of genes involved in neurogenesis (p=6.84x10^{-7}).**

Conclusions **Compelling loci with suggestive evidence for association were identified, and results warrant further investigation for a role of neurogenesis in OA etiology. However, replication was complicated due to the specific phenotype of the discovery cohort; it remains to be established whether trade-off between more specific phenotypes and increasing sample-numbers may result in new consistent susceptibility loci.**

Keywords: GWAS, meta-analysis, GARP study, osteoarthritis, heterogeneity

INTRODUCTION

Osteoarthritis (OA) is a degenerative disease of the joint characterized by wear and tear of cartilage, subchondral bone remodeling, and inflammation of synovial tissues. OA is most common in knee, hip, and/or hand joints, and prevalence increases with age.[1] Care is concentrated on controlling symptoms, minimizing disability, and slowing down progression until joint replacement is indispensable given that no treatment is available to cure OA (reviewed in [2]). More insight into OA pathophysiological processes and the etiology will greatly enhance possibilities for therapy and prospects for patients.

A variety of factors such as genetics, hormones, metabolism, and environment contribute to development of OA.[3] Heritability of OA differs per joint site and based on twin and sibling studies it has been estimated between 40 and 60% depending on the site affected.[4, 5] By compiling large numbers of cases and controls, consortia have recently succeeded to identify robust OA susceptibility loci with genome wide significance.[6-8] Typically, these loci exert small effects[9] partly since the OA phenotype is highly heterogeneous thereby accounting only for a small part of the OA heritability.

To discover OA genes with larger effect sizes we collected sibling pairs, affected by symptomatic OA at 2 or more joint locations with KL≥2, the GARP study.[10] By applying a genome wide linkage to GARP sibling pairs and subsequent replication in population based cases we previously identified the OA susceptibility gene *DIO2*. Here, we initiated a genome wide association scan in the GARP study to identify additional OA susceptibility genes. For this, we applied a 2-stage follow-up design: replication in similar severely affected OA patients, followed by large scale confirmation to population based cases by meta-analysis.

MATERIALS AND METHODS

Study populations. In the work presented, the Genetics osteoARthritis and Progression (GARP) study consisting in sibling pairs with symptomatic OA at 2 or more joint locations at relative young age (40-70 yr)[10] was used as discovery population in a genome wide association scan (N=378). For replication of our results, we included cases affected in 3 or more joints of the deCODE study (N=310).[11, 12] From the arcOGEN consortium,[13] no information about hand OA was available, but all cases with bilateral total joint replacement (TJR) due to OA at hip or knee were included (N=380). The Rotterdam study has been described previously,[14] and cases with radiographic hip or knee OA were included in the analysis (N=1,095). For *de novo* genotyping, cases with TJR for hip or knee OA from a Spanish study[15] (N=359) and a Dutch study (the Paprika study, N=1,016; a long-term follow-up study of patients that have undergone TJR at hip or knee[16]) were used. **Table 1** shows characteristics of the different studies, all with subjects from Western European descent. More details on the studies can be found in the corresponding references. All studies were approved by their respective institutional review boards and informed consent was obtained from all participants involved.

Genotyping. The GARP study (N=380) was

genotyped using Illumina Infinium HD Human660W-Quad BeadChips (Illumina, San Diego, CA, USA). Two subjects were removed based on low-quality genotyping resulting in a total of 378 cases for the analysis. A random sample of 1670 inhabitants of Rotterdam[14] was used as population control. Genotyping of these controls using Illumina Infinium II HumanHap 550K Beadchips and Illumina Infinium II HumanHap550-Duo BeadChips (Illumina) for the RS-I and the RS-II cohorts, respectively, has been described previously.[17] A total of 516,834 SNPs, genotyped for both GARP and controls, was used for the GWAS and analyzed with cc_assoc (in-house analysis program, developed to correct for familial dependence; H. W. Uh and Q. Helmer). Quality control was performed selecting for SNPs with call rate ≥0.99, $p_{HWE}>10^{-4}$, and MAF≥0.05, and Illumina clusterplots were visually inspected to confirm quality of the genotyping. Subsequently, imputation for all polymorphic SNPs (MAF >0.01) was performed with HapMap2 using IMPUTE,[18] and data were subjected to a second round of quality control as described before, resulting in analysis of 1,902,224 SNPs total. There was no overlap between the samples used in this analysis and controls used for the replication in Rotterdam.

Genotyping of the Paprika study and of the Spain cohort was performed using the Sequenom MassARRAY iPLEX Gold or TaqMan SNP Genotyping assays following the manufacturer's instructions. Controls for the Spain cohort were determined like-wise, and as a control for the Paprika study we used existing GWAS data in the Leiden Lang Leven Study after quality control screening described by Deelen *et al.*[19]

Study design and meta-analysis. Discovery GWAS was performed using the GARP study. Estimations for this analysis using Quanto software (http://biostats.usc.edu/software) suggested that for this study we would be able to identify loci with 80% power for odds ratios (OR) as high as 1.5 and allele frequencies of at least 10% with genome wide significance (Supplementary Table 1; note that results shown are most likely overestimated given the fact that the data are not corrected for sibling ship).

To replicate results from the discovery GWAS a 2-step approach was taken. First, since from all populations available a selection of the deCODE study was relatively similar to our study, 18 unique loci with a preset threshold of $p\leq1x10^{-5}$ were replicated in this study. Subsequently, SNPs with significant association after correction for multiple testing $p\leq2.8x10^{-3}$ in the meta-analysis GARP-deCODE (8 SNPs) were selected for replication in available additional studies. Meta-analyses were performed in R using the meta package to obtain p-value, 95% confidence interval (CI_{95}), odds ratio (OR), and forrest plots. Figure 1 shows the flow chart of the study design.

Exploration of gene expression. OA-related changes in expression levels of positional genes were explored as described previously in in-house available gene expression datasets of cartilage (the RAAK study[20, 21]) and blood (the GARP study[22]). Mean expression of all genes was 7.4 (range: 6.6-14.9), expressed as the overall relative light intensity signals in respectively the preserved and OA affected cartilage as provided by Illumina.

Pathway analysis. Gene enrichment analysis

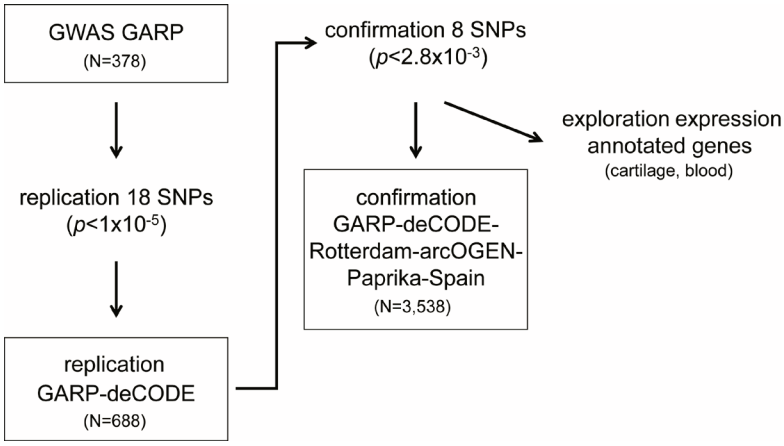


Figure 1. Flow chart of the experimental procedure.

Table 1. Description of the studies included in the genome wide association meta-analysis.
(GOA: generalized OA; TJR: total joint replacement; F: females)

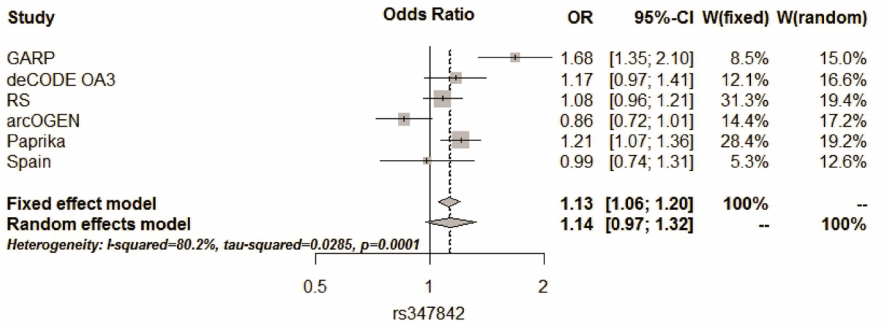
STUDY	N	Phenotype	Design	Mean Age	pct F	Platform
GARP	378	familial GOA (KL>2)	case-control	60	82%	Illumina Human660W
deCODE	310	GOA (≥ 3 joints)	case-control	69	58%	Infinium Human300
arcOGEN	380	TJR both sides	case-control	NA	60%	Illumina Human610
Rotterdam	1,095	OA	population-based	67	57%	Illumina Human550v3
Paprika	1,016	TJR	case-control	62	65%	Sequenome
Spain	359	TJR	case-control	66	77%	Sequenome

was performed with the online available Database for Annotation, Visualization and Integrated Discovery (DAVID; [23, 24]) selecting for all biological processes in the gene ontology database (GOTERM_BP_FAT). For the analysis, genes within 10 kb up- or downstream of the SNPs identified in the discovery GWAS with p<1x10^{-3} (N=547) were selected.

RESULTS

Discovery genome wide association. Baseline characteristics of the GARP and follow-up studies are shown in Table 1. After imputation and quality controls data were analyzed for 378 cases versus 1670 controls. Supplementary Figure 1 provides the Manhattan and QQ-plots with an overall genomic control lambda (λ_{GC}) of 1.0525, indicating no significant population stratification. Table 2 provides all unique locithat reached the pre-specified threshold of p≤1x10^{-5} for replication (N=18). The strongest statistical evidence for association with symptomatic OA at multiple joint sites was obtained for rs10488327 located at chromosome 7p12 with odds ratio OR=2.39

A



B

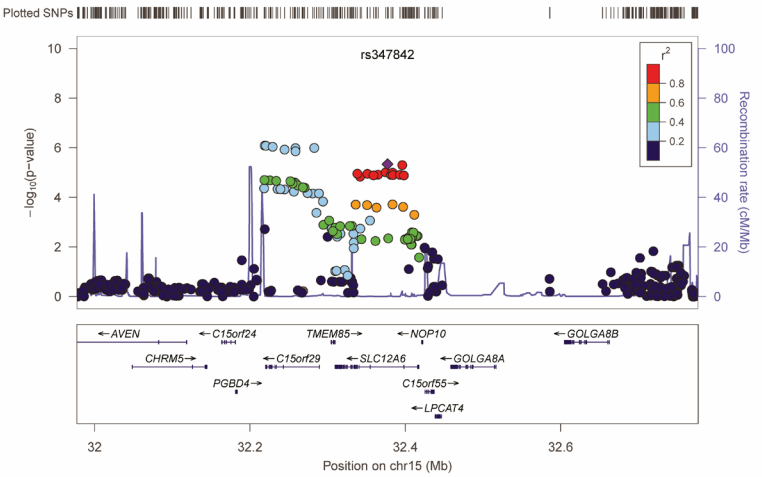


Figure 2. Forrest plot (A) and LocusZoom plot (B) for rs347842. (legend: OR: odds ratio; 95%-CI: 95% confidence interval; W: contribution to the meta-anlaysis)

(95% confidence interval CI_{95}=1.73-3.30 and p=1.23x10^{-7}). Closest gene, located 500 kb upstream, is the POM121 Membrane Glycoprotein-Like 12 (*POM121L12*) gene.

Pathway analysis. In order to investigate whether the results of our GWAS could link OA to a specific biological pathway, gene enrichment analysis for biological pathways was performed including genes within 10 kb up- or downstream of all SNPs with $p \leq 1x10^{-3}$ as identified in the discovery GWAS. Remarkably, many of the significant GO-terms appeared to be related to neuron differentiation and strongest statistical evidence was for GO:0007399 (nervous system development; p=6.84x10^{-7}; see Table 3).

Replication and confirmation. Following

the discovery stage, a total of 18 loci were tested *in silico* in independent cases with OA in 3 or more joints of the deCODE study. Of these, 9 were found to be significant after correction for multiple testing in the combined analysis of GARP and deCODE (N=688; Table 2A). Subsequently, these were tested for association with OA by comparing a total of 3,538 cases with 49,713 controls in a meta-analysis of 6 studies including GARP. Although none of the loci were genome wide significant, the strongest statistical evidence in the meta-analysis was found for rs347842 on chromosome 15q14 (p=3.3x10^{-4}, with OR=1.13 and CI_{95}=1.06-1.20; see Table 2A and Figure 2A). As shown in the LocusZoom plot (Figure 2B), rs347842 is located within the solute carrier family 12 potassium/chloride transporter member 6 gene (*SLC12A6*). Sensitivity analyses by excluding the GARP discovery cohort indicated an OR=1.09 with CI_{95}=1.01-1.16 (p=1.9x10^{-2}). Furthermore, of note was the association with SNP rs201328 on chromosome 6p22 (confirmation: p=4.8x10^{-3}, with OR=1.12 and CI_{95}=1.03-1.21) located within the *CDKAL1* gene (CDK5 regulatory subunit associated protein 1-like 1), showing consistent effects for all studies included in the meta-analysis. However, results of the sensitivity analyses were not significant.

Exploring gene expression of annotated genes. Next, we explored a dataset of gene expression profiles of blood and cartilage (unaffected and OA affected cartilage of the same joint) for OA-related changes of the nearest gene for each SNP. As shown in Table 2C, only *CDKAL1* was found to be expressed at moderate levels in cartilage, however, without response to OA pathophysiology (Table 2B). In contrast, gene expression profiles showed that in blood of OA patients expression of *CDKAL1* is decreased with nominal significance (1.1-fold) when compared to healthy controls.

DISCUSSION

To identify genes and pathways associated to relatively severe OA, a genome wide association study was performed within the GARP study consisting of sibling pairs with familial, relatively severe symptomatic OA. The strongest statistical evidence in the confirmation meta-analysis was found for the A allele of rs347842 within the *SLC12A6* gene which mediates electroneutral potassium-chloride cotransport (p=3.3x10^{-4}; OR=1.13 with CI_{95}=1.06-1.20 in the meta-analysis). Given that rs347842 is located within a linkage disequilibrium (LD) block with multiple signals and comprising several genes that were well detected in our cartilage and blood datasets such as the transmembrane domain protein 85 gene (*TMEM85*) or the gene encoding NOP10 ribonucleoprotein (Figure 2B), OA susceptibility could also be caused by rs347842 via other genes located in this region (Supplementary Table 2). Analysis of annotated genes from loci of the discovery phase however indicated highly significant enrichment of genes involved in neurogenesis (p=6.8x10^{-7}), which may confirm a role for neuronal genes in OA.[20, 25] This could either unexpected link nociception and OA[26] or aberrant nervous system development and OA due to failures in the locomotor system resulting in an accumulation of joint damage and concurrent disposition to incidence of OA during aging. In line with the latter, targeted deletion of the

Table 2. Susceptibility loci identified in the GARP study. A Results of unique loci identified in the discovery GWAS with $p \leq 1x10^{-5}$ and meta-analysis of GARP and deCODE (replication). **B** Results of meta-analysis in GARP, deCODE, Rotterdam, arcOGEN, the Paprika study, and Spain (confirmation) for loci with significant association after correction for multiple testing in the replication in deCODE. **C** Expression of nearest gene in cartilage and in blood; p-values refer to comparison of gene expression in OA-affected and unaffected cartilage or in blood of OA patients and healthy controls (legends: OR: odds ratios with 95% confidence interval; MAF: minor allele frequency; Dir: direction of effects; AveExpr: relative gene expression; FC: fold change, with negative values indicating decreased expression; NA: not significantly detected).

A				**DISCOVERY**		**REPLICATION**		
SNP	**CHR**	**POS**	**MAF**	p_{GARP}	OR_{GARP}	p	**OR**	**Dir**
rs347842	**15**	**32377099**	**0.33**	**4.63E-06**	**1.68[1.35-2.10]**	**2.37E-05**	**1.359**	**+ +**
rs11711534	3	50659815	0.12	3.24E-06	2.30[1.62; 3.27]	3.60E-03	1.419	+ -
rs17236749	15	32219072	0.23	8.18E-07	1.90[1.47-1.37]	**1.23E-04**	**1.401**	+ +
rs2523710	6	31558888	0.10	1.43E-05	0.58[0.45; 0.74]	7.20E-03	0.788	- +
rs11157970	14	52959740	0.12	6.47E-06	2.14[1.54-2.99]	**2.15E-04**	**1.443**	+ +
rs201328	6	21154069	0.16	1.03E-05	1.76[1.37-2.26]	**1.00E-03**	**1.323**	+ +
rs12148483	15	90881925	0.24	2.29E-06	1.60[1.32-1.94]	**8.58E-07**	**1.398**	+ +
rs10488327	7	52343239	0.11	1.23E-07	2.39[1.73; 3.30]	**1.40E-03**	**1.408**	+ -
rs7929217	11	40835136	0.16	4.58E-06	1.72[1.36-2.17]	**9.00E-04**	**1.323**	+ +
rs13051863	21	23759301	0.19	8.00E-07	1.85[1.45; 2.36]	**2.10E-03**	**1.276**	+ -
rs4678188	3	123429045	0.23	5.81E-06	1.74[1.37; 2.21]	1.29E-02	1.206	+ -
rs250513	5	78337952	0.24	1.11E-05	1.65[1.32; 2.07]	6.30E-03	1.224	+ -
rs12995756	2	28778518	0.09	1.16E-05	1.79[1.38; 2.33]	3.20E-03	1.333	+ -
rs2114620	2	203404761	0.25	1.29E-05	1.68[1.33; 2.11]	1.38E-02	1.205	+ -
rs7710039	5	80528723	0.37	1.73E-06	1.58[1.31; 1.90]	1.95E-02	1.175	+ -
rs8102824	19	46247047	0.20	7.70E-06	1.70[1.35; 2.14]	9.07E-02	1.146	+ -
rs13212606	6	4629361	0.13	6.70E-06	0.54[0.41; 0.70]	9.33E-02	0.860	- +
rs9694810	8	95730636	0.46	5.45E-06	0.65[0.53; 0.78]	9.10E-03	0.845	- +

B

	CONFIRMATION			SENSITIVITY	
SNP	***p***	**OR**	**Dir**	***p***	**OR**
rs347842	**3.30E-04**	**1.126[1.06-1.20]**	**+ + + - + -**	**1.87E-02**	**1.085 [1.01-1.16]**
rs17236749	6.39E-02	1.075[1.00-1.16]	+ + + - - +	7.02E-01	1.016 [0.94-1.10]
rs11157970	2.32E-01	1.067[0.96-1.19]	+ + - - + -	7.86E-01	0.985 [0.88-1.10]
rs201328	**4.80E-03**	**1.117[1.03-1.21]**	**+ + + + + +**	1.21E-01	1.070 [0.98-1.15]
rs12148483	**4.84E-04**	**1.122[1.05-1.20]**	+ + - - + +	**4.27E-02**	**1.074 [1.00-1.15]**
rs10488327	5.82E-02	1.098[1.00-1.21]	+ - + +	7.42E-01	1.017 [0.92-1.13]
rs7929217	4.65E-01	1.029[0.95-1.11]	+ + - - + +	4.04E-01	0.966 [0.89-1.05]
rs13051863	5.79E-01	1.027[0.93-1.13]	+ - - +	1.30E-01	0.923 [0.83-1.02]

C

		CARTILAGE			BLOOD		
SNP	**Gene Symbol**	**AveExpr**	**p_{expr}**	**FC**	**AveExpr**	**p_{expr}**	**FC**
rs347842	***SLC12A6***	**NA**			**6.4**	**1.98E-01**	**-1.03**
rs17236749	*C15orf29*	NA			NA		
rs11157970	*DDHD1*	NA			6.5	1.56E-01	-1.02
rs201328	***CDKAL1***	**7.6**	**4.80E-01**	**1.00**	**7.2**	**4.40E-02**	**-1.09**
rs12148483	***C15orf32***	**NA**			**NA**		
rs10488327	*POM121L12*	NA			NA		
rs7929217	*NGL1*	NA			NA		
rs13051863		--			--		

gene in mice leads to a locomotor and a sensorimotor gating deficit, suggesting a critical role for *SLC12A6* in development of the nervous system.[27, 28] In addition, a 17-base deletion within exon 4 of the *SLC12A6* gene was identified by Jiao *et al*[29] to be the cause of giant axonopathy, a neuromuscular disease in gaxp mice. Genetic variation at rs347842 might therefore dispose to OA via aberrations in the neuro-muscular system rather than by affecting structure or composition of cartilage and/or bone directly. The identified association of rs347842 with OA remains to be established since the signal was not genome wide significant and direction of effects in the replication were not unequivocal. The fact that results were replicated relatively well in the deCODE selection suggests that the use of more comparable cases in the meta-analysis could have increased the power to detect association with genome wide significance.

Although with small effects, most consistent directions of effects for all studies included in the meta-analysis were found for rs201328 located within the *CDKAL1* gene (p=4.8x10^{-3}; OR=1.12 with CI_{95}=1.03-1.21 in the confirmation meta-analysis). *CDKAL1* is expressed well in cartilage but its expression is not responsive to OA pathophysiology (Table 2C). In blood of OA patients however, *CDKAL1* expression was found to be significantly decreased. Genome-wide association studies have linked SNPs in *CDKAL1* with type 2 diabetes,[30] with body mass index (BMI) in east Asians,[31] and with intrauterine growth, adult height, and metabolism.[32] Metabolism and obesity are established risk factors for OA[3] and another locus associated with BMI was also found to establish risk to OA.[33, 34]

Despite the negligible linkage disequilibrium between the previously identified loci in *CDKAL1* and the here identified locus (LD <0.1), we cannot exclude the possibility that the association found in our analysis is due to association with either BMI or height which is in consonance with expression differences of *CDKAL1* in blood of OA patients pointing towards a more systemic effect. Since

Table 3. Pathway-analysis with DAVID. The grey shading indicates significant GO-term (Legend: N: number of genes in this GO-term; Pct: percentage of genes from the total number of genes tested; Fold Enr: fold enrichment of the indicated pathway; p_{adj}: P-value adjusted for multiple testing.

GO-Term	N	Pct	p	$p_{adj.}$	Fold Enr
GO:0007399~nervous system development	66	12.07	1.25E-09	6.84E-07	2.22
GO:0007275~multicellular organismal development	128	23.40	2.25E-09	1.23E-06	1.63
GO:0048731~system development	107	19.56	2.33E-08	1.27E-05	1.68
GO:0048856~anatomical structure development	113	20.66	3.33E-08	1.82E-05	1.64
GO:0032502~developmental process	132	24.13	6.23E-08	3.41E-05	1.53
GO:0032501~multicellular organismal process	155	28.34	2.66E-05	1.46E-02	1.32
GO:0048699~generation of neurons	33	6.03	6.29E-05	3.44E-02	2.16
GO:0030182~neuron differentiation	28	5.12	6.99E-05	3.82E-02	2.34
GO:0009653~anatomical structure morphogenesis	56	10.24	8.39E-05	4.59E-02	1.71

direction of effects for this locus were consistent but small, increased power and concurrent lower p-values could possibly be accomplished by increasing the number of samples.

The current study has some limitations that should be taken into account. Despite the severe selection of familial cases to find loci with larger effect sizes, the identified loci were not genome wide significant. Most likely our discovery analysis lacked power due to inclusion of siblings and the relatively small number of cases, thereby highlighting the difficult trade-off between the specific phenotype facilitating detection of SNPs potentially with larger effect sizes and large sample numbers to allow significant p-values. Another drawback of our study design was the lack of comparable cohorts to replicate results of the discovery GWAS, which is also demonstrated by the significant heterogeneity in the meta-analysis (p=1x10^{-4}). This heterogeneity may have corroded power to detect strong associations with OA. In fact, when taking this heterogeneity into account for rs347842 by applying a random effects model instead of the fixed effect model used here, result of the meta-analysis is not statistically significant. However, results for rs201328 and rs12148483 remained significant even when random effects models are used (Supplementary Figures 2A and C).

Taken together, in this study using a small OA cohort with familial, relatively severe phenotype we identified compelling susceptibility loci albeit without genome wide significance. Many of the identified loci are localized close to genes involved in neurogenesis, however, due to the lack of studies with comparable phenotype our findings warrant further investigation by replication in additional severe familial cases before additional conclusions can be drawn. Again, this stresses the importance to improve homogeneity in selection procedures of (subsets of) OA patients.

ACKNOWLEDGMENTS

This study was supported by a grant from the National Institute for Healthy Ageing (Grant 05060810) in the framework of the Netherlands Genomics Initiative/ Netherlands Organization for Scientific Research and of the Centre for Medical Systems Biology (CMSB). We thank R. van der Breggen, E. Suchiman, and D. Kremer for help with the genotyping.

REFERENCES

1 Johnson VL, Hunter DJ. The epidemiology of osteoarthritis. *Best Pract Res Clin Rheumatol* 2014;**28**:5-15.

2 Seed SM, Dunican KC, Lynch AM. Treatment options for osteoarthritis: considerations for older adults. *Hosp Pract (Minneap)* 2011;**39**:62-73.

3 Bay-Jensen AC, Slagboom E, Chen-An P, Alexandersen P, Qvist P, Christiansen C, *et al.* Role of hormones in cartilage and joint metabolism: understanding an unhealthy metabolic phenotype in osteoarthritis. *Menopause* 2013;**20**:578-86.

4 Valdes AM, Spector TD. Genetic epidemiology of hip and knee osteoarthritis. *Nat Rev Rheumatol* 2011;**7**:23-32.

5 Ikegawa S. The genetics of common degenerative skeletal disorders: osteoarthritis and degenerative disc disease. *Annu Rev Genomics Hum Genet* 2013;**14**:245-56.

6 Evangelou E, Valdes AM, Kerkhof HJ, Styrkarsdottir U, Zhu Y, Meulenbelt I, *et al.* Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. *Ann Rheum Dis* 2011;**70**:349-55.

7 Evangelou E, Valdes AM, Castano-Betancourt MC, Doherty M, Doherty S, Esko T, *et al.* The DOT1L rs12982744 polymorphism is associated with osteoarthritis of the hip with genome-wide statistical significance in males. *Ann Rheum Dis* 2013;**72**:1264-5.

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

9 Panoutsopoulou K, Zeggini E. Advances in osteoarthritis genetics. *J Med Genet* 2013;**50**:715-24.

10 Riyazi N, Meulenbelt I, Kroon HM, Ronday KH, Hellio le Graverand MP, Rosendaal FR, *et al.* Evidence for familial aggregation of hand, hip, and spine but not knee osteoarthritis in siblings with multiple joint involvement: the GARP study. *Ann Rheum Dis* 2005;**64**:438-43.

11 Ingvarsson T, Hagglund G, Lohmander LS. Prevalence of hip osteoarthritis in Iceland. *Ann Rheum Dis* 1999;**58**:201-7.

12 Stefansson SE, Jonsson H, Ingvarsson T, Manolescu I, Jonsson HH, Olafsdottir G, *et al.* Genomewide scan for hand osteoarthritis: a novel mutation in matrilin-3. *Am J Hum Genet* 2003;**72**:1448-59.

13 Panoutsopoulou K, Southam L, Elliott KS, Wrayner N, Zhai G, Beazley C, *et al.* Insights into the genetic architecture of osteoarthritis from stage 1 of the arcOGEN study. *Ann Rheum Dis* 2011;**70**:864-7.

14 Hofman A, Breteler MM, van Duijn CM, Krestin GP, Pols HA, Stricker BH, *et al.* The Rotterdam Study: objectives and design update. *Eur J Epidemiol* 2007;**22**:819-29.

15 Rodriguez-Lopez J, Pombo-Suarez M, Liz M, Gomez-Reino JJ, Gonzalez A. Lack of association of a variable number of aspartic acid residues in the asporin gene with osteoarthritis susceptibility: case-control studies in Spanish Caucasians. *Arthritis Res Ther* 2006;**8**:R55.

16 Keurentjes JC, Fiocco M, So-Osman C, Onstenk R, Koopman-Van Gemert AW, Poll RG, *et al.* Patients with severe radiographic osteoarthritis have a better prognosis in physical functioning after hip and knee replacement: a cohort-study. *PLoS One* 2013;**8**:e59500.

17 Teichert M, Eijgelsheim M, Rivadeneira F, Uitterlinden AG, van Schaik RH, Hofman A, *et al.* A genome-wide association study of acenocoumarol maintenance dosage. *Hum Mol Genet* 2009;**18**:3758-68.

18 Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007;**39**:906-13.

19 Deelen J, Beekman M, Uh HW, Helmer Q, Kuningas M, Christiansen L, *et al.* Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell* 2011;**10**:686-98.

20 Ramos YF, den HW, Bovee JV, Bomer N, van der Breggen R, Lakenberg N, *et al.* Genes Involved in the Osteoarthritis Process Identified through Genome Wide Expression Analysis in Articular Cartilage; the RAAK Study. *PLoS One* 2014;**9**:e103056.

21 Ramos YF, Metrustry S, Arden N, Bay-Jensen AC, Beekman M, de Craen AJ, *et al.* Meta-analysis identifies loci affecting levels of the potential osteoarthritis biomarkers sCOMP and uCTX-II with genome wide significance. *J Med Genet* 2014;**51**:596-604.

22 Ramos YF, Bos SD, Lakenberg N, Bohringer S, den Hollander WJ, Kloppenburg M, *et al.* Genes expressed in blood link osteoarthritis with apoptotic pathways. *Ann Rheum Dis* 2013.

23 Huang dW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009;**37**:1-13.

24 Huang dW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;**4**:44-57.

25 Day-Williams AG, Southam L, Panoutsopoulou K, Rayner NW, Esko T, Estrada K, *et al.* A Variant in MCF2L Is Associated with Osteoarthritis. *Am J Hum Genet* 2011;**89**:446-50.

26 Seidel MF, Wise BL, Lane NE. Nerve growth factor: an update on the science and therapy. *Osteoarthritis Cartilage* 2013;**21**:1223-8.

27 Howard HC, Mount DB, Rochefort D, Byun N, Dupre N, Lu J, *et al.* The K-Cl cotransporter KCC3 is mutant in a severe peripheral neuropathy associated with agenesis of the corpus callosum. *Nat Genet* 2002;**32**:384-92.

28 Sun YT, Lin TS, Tzeng SF, Delpire E, Shen MR. Deficiency of electroneutral K+-Cl-cotransporter 3 causes a disruption in impulse propagation along peripheral nerves. *Glia* 2010;**58**:1544-52.

29 Jiao Y, Jin X, Yan J, Zhang C, Jiao F, Li X, *et al.* A deletion mutation in Slc12a6 is associated with neuromuscular disease in gaxp mice. *Genomics* 2008;**91**:407-14.

30 Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, *et al.* A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet* 2007;**39**:770-5.

31 Wen W, Cho YS, Zheng W, Dorajoo R, Kato N, Qi L, *et al.* Meta-analysis identifies common variants associated with body mass index in east Asians. *Nat Genet* 2012;**44**:307-11.

32 Horikoshi M, Yaghootkar H, Mook-Kanamori DO, Sovio U, Taal HR, Hennig BJ, *et al.* New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nat Genet* 2013;**45**:76-82.

33 Zeggini E, Panoutsopoulou K, Southam L, Rayner NW, Day-Williams AG, Lopes MC, *et al.* Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* 2012;**380**:815-23.

34 Elliott KS, Chapman K, Day-Williams A, Panoutsopoulou K, Southam L, Lindgren CM, *et al.* Evaluation of the genetic overlap between osteoarthritis with body mass index and height using genome-wide association scan data. *Ann Rheum Dis* 2013;**72**:935-41.

Supplementary Table 1. Power estimates for different allele frequencies and effect sizes.

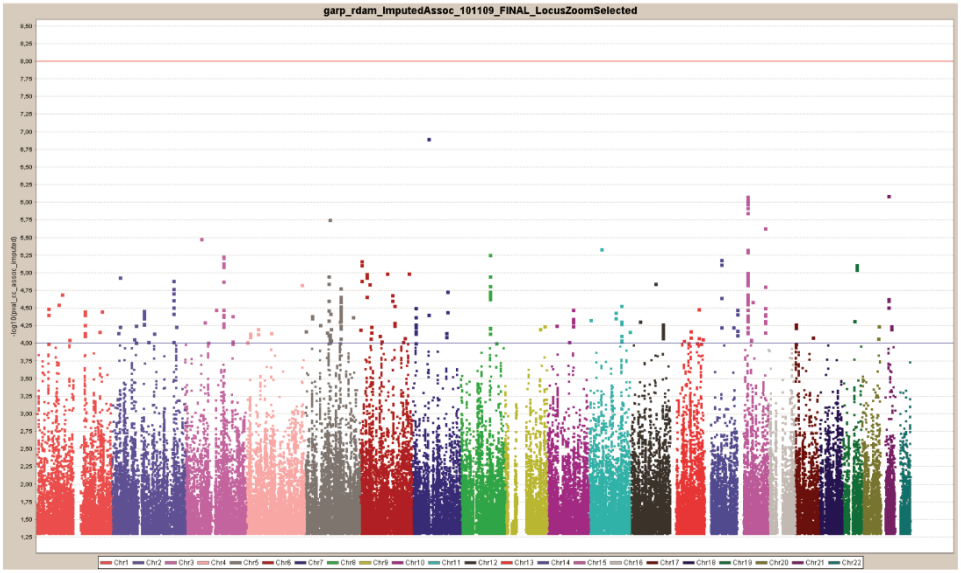
Quanto		
MAF	**OR**	**Power (%)**
0.05	1.2	16.8
	1.4	47.5
	1.6	78.0
	1.8	93.9
	2.0	98.8
0.10	1.2	25.4
	1.4	69.1
	1.6	93.9
	1.8	99.4
	2.0	100.0
0.15	1.2	31.1
	1.4	78.6
	1.6	97.3
	1.8	99.8
	2.0	100.0
0.2	1.2	34.4
	1.4	82.6
	1.6	98.3
	1.8	99.9
	2.0	100.0
0.25	1.2	35.8
	1.4	83.8
	1.6	98.4
	1.8	99.9
	2.0	100.0
0.3	1.2	35.8
	1.4	83.3
	1.6	98.2
	1.8	99.9
	2.0	100.0
0.35	1.2	34.6
	1.4	81.4
	1.6	97.6
	1.8	99.8
	2.0	100.0
0.4	1.2	32.7
	1.4	78.2
	1.6	96.4
	1.8	99.6
	2.0	100.0
0.45	1.2	30.0
	1.4	73.6
	1.6	94.4
	1.8	99.2
	2.0	99.9

Supplementary Table 2. Expression of genes localized in chromosomal region around rs347842 in cartilage and in blood (legends: p_{expr}: p-value for comparison of gene expression in OA-affected and unaffected cartilage or in blood of OA patients and healthy controls; AveExpr: relative gene expression; FC: fold change, with negative values indicating decreased expression; NA: not significantly detected).

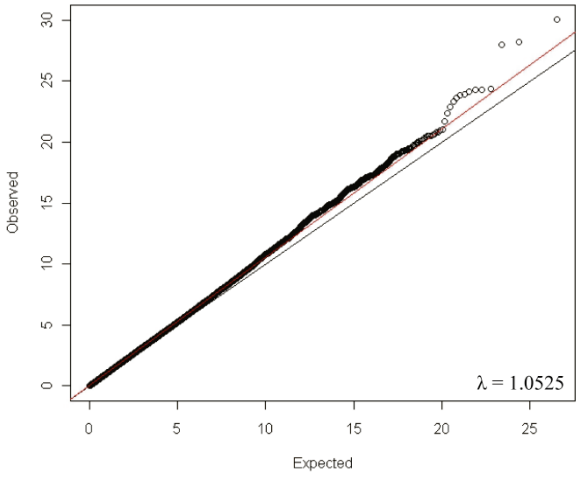
	CARTILAGE			BLOOD		
GeneID	**AveExpr**	**p_{expr}**	**FC**	**AveExpr**	**p_{expr}**	**FC**
C15ORF24	11.1	1.14E-02	1.09	10.2	6.75E-01	1.02
PGBD4	NA			6.6	1.24E-01	1.03
C15ORF29	NA			6.5	1.36E-01	-1.02
TMEM85	10.6	3.56E-01	1.03	10.0	7.49E-01	1.01
SCLC12A6	**NA**			6.4	1.98E-01	-1.03
NOP10	11.0	3.89E-01	1.04	11.4	5.90E-01	1.04
C15ORF55	NA			6.5	8.31E-01	1.00
LPCAT4	NA			6.9	6.12E-02	-1.08
GOLGA8A	7.7	1.02E-02	-1.10	6.8	1.41E-01	-1.05

Supplementary Figure 1. Visualization of discovery GWAS results. Manhattan plot (A) and QQ plot (B).

A

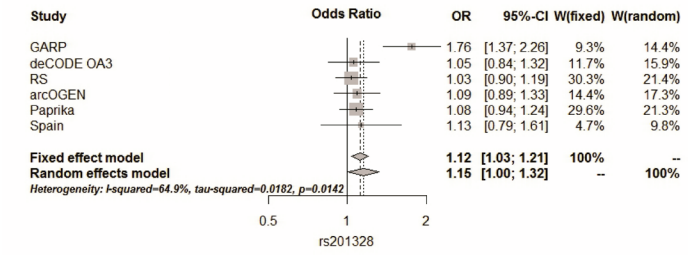


B

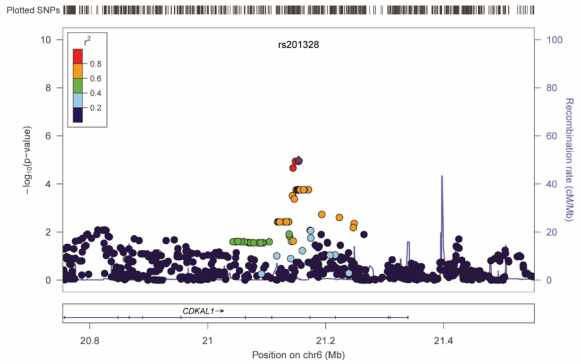


Supplementary Figure 2. Forrest plot and LocusZoom plot. For rs201328 (A and B, respectively) and for rs12148483 (C and D, respectively; legend: OR: odds ratio; 95%-CI: 95% confidence interval; W: contribution to the meta-analysis)

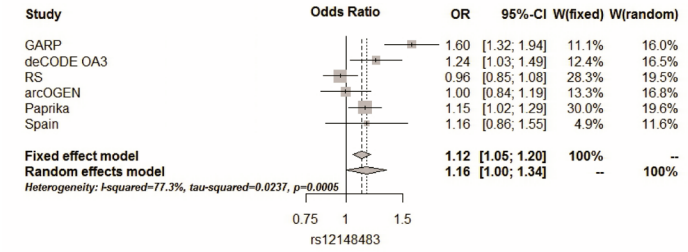
A



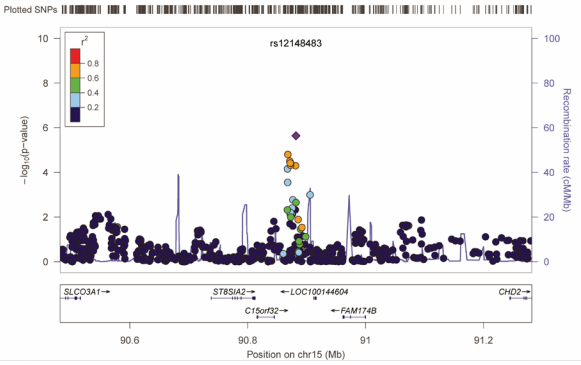
B



C



D



A gain of function mutation in *TNFRSF11B* causes osteoarthritis with chondrocalcinosis

6

Ann Rheum Dis (doi:10.1136/annrheumdis-2013-205149)

A gain of function mutation in *TNFRSF11B* causes osteoarthritis with chondrocalcinosis

Yolande F.M. Ramos[1,2], Steffan D. Bos[1,2], Ruud van der Breggen[1], Margreet Kloppenburg[3], Kai Ye[1], Eric-Wubbo E.M.W. Lameijer[1], Rob G.H.H. Nelissen[4], P. Eline Slagboom[1,2*] and Ingrid Meulenbelt[1,2*#]
[1]Leiden University Medical Center, Dept. Medical Statistics and Bioinformatics, section Molecular Epidemiology, Leiden, Netherlands, [2]Netherlands Consortium for Healthy Ageing, Leiden, Netherlands, [3]Dept of Rheumatology & Department of Clinical Epidemiology, [4]Leiden University Medical Center, Department of Orthopedics, Leiden, Netherlands

[*]Shared authorship position

6

ABSTRACT

Objective **To identify pathogenic mutations that reveal underlying biological mechanisms driving OA.**
Methods **Exome sequencing was applied to two distant family members with dominantly inherited early onset primary osteoarthritis at multiple joint sites with chondrocalcinosis (FOA). Confirmation of mutations occurred by genotyping, and linkage analyses across the extended family. The functional effect of the mutation was investigated by means of a cell-based assay. To explore generalizability, mRNA expression analysis of the relevant genes in the discovered pathway was explored in preserved and osteoarthritic articular cartilage of independent patients undergoing joint replacement surgery.**
Results **We identified a heterozygous, probably damaging, read-through mutation (c.1205A=>T; p.Stop402Leu) in *TNFRSF11B* encoding osteoprotegerin that is likely causal to the osteoarthritis phenotype in the extended family. In a bone resorption assay, the mutant form of osteoprotegerin showed enhanced capacity to inhibit osteoclastogenesis and bone resorption. Expression analyses in preserved and affected articular cartilage of independent osteoarthritis patients showed that up-regulation of *TNFRSF11B* is a general phenomenon in the pathophysiological process.**
Conclusions **Albeit that the role of the molecular pathway of osteoprotegerin has been studied in osteoarthritis, we are the first to demonstrate that enhanced osteoprotegerin function could be a directly underlying cause. We advocate that agents counteracting the function of osteoprotegerin could comply with new therapeutic interventions of osteoarthritis.**

Keywords: osteoarthritis, RANK-antagonism, *TNFRSF11B*, osteoclastogenesis, exome sequencing

INTRODUCTION

Osteoarthritis (OA) is a prevalent, complex, disabling disease of the articular joints, characterized by degradation of the hyaline articular cartilage and remodeling of the subchondral bone with sclerosis. As the population ages and obesity rates rise, OA is becoming more prevalent and current estimates of men and women aged over 60 years having symptomatic OA are 9.6% and 18.0%, respectively.[1] Yet, there is no effective therapy to reverse or slow down the disease, analgesia, physiotherapy and, at end stage, joint replacement surgery are the main treatment options. As a result, osteoarthritis has a large detrimental impact on the quality of life of the elderly and causes a major burden on health, and social care systems, with indirect costs being predominant.[1, 2] To allow the development of new therapies, there is an ongoing need for insight into the underlying biological mechanisms driving OA. Despite the detection of compelling OA susceptibility loci[3, 4], such insights, until now, remain limited. To detect rare mutations with large to moderate effect sizes, recently focus has been to apply next generation sequencing of exomes of severely affected family-based patients. The value of identified rare variants lies in the characterization of causal gene functions and underlying novel pathways in complex disease processes.[5, 6] Notably in this respect; among the identified genes with mutations causing early-onset families with OA associated syndromic skeletal phenotypes such as collagen type II (*COL2A1*),[7] the growth differentiating factor 5 (*GDF5*),[8] collagen type 11-alpha-1 (*COL11A1*),[9, 10] cartilage oligomeric protein (*COMP*)[11] and extracellular signaling molecule and member of the TGF-β superfamily (*SMAD3*)[12, 13] were three genes that appeared to additionally confer risk to common OA phenotypes, namely *SMAD3*,[15] *COL11A1*,[16, 17] and the *GDF5* gene which reached genome wide significance.[8, 14]

Previously, we reported on a single early onset family with familial generalized osteoarthritis (FOA) without any dysplasia, with radiographic chondrocalcinosis as reflected by calcifications within the joint- and fibro-cartilage.[18] Candidate gene screening, showed that none of the genes encoding the extracellular matrix structural proteins were involved in the pathogenesis of this osteoarthritic phenotype.[18] Furthermore, a genetic linkage search in this and 6 other FOA families revealed significant linkage on chromosome 2q32 albeit that the attribution of this particular family to the linkage was moderate and Sanger sequencing of positional exons failed to definitively indicate the causal mutation.[19] Recently, a relevant extension was established in the ascertainment of the family under study; additionally affected distant family members with the phenotype were detected whereas third generation family members definitively developed the FOA phenotype. Such extensions increases the power of successful identification of genes considerably. To identify pathogenic osteoarthritis mutations, we applied exome sequencing to two distant family members.

MATERIALS AND METHODS

Study design. To identify pathogenic mutations underlying development of OA exome sequencing was performed for 2 distantly related FOA family members. After applying an eligible prioritization scheme five candidates were selected for *de novo*

genotyping across the extended family (N=16). Linkage analysis was performed for the mutation that was shared among all affected members and the mutation was validated by *de novo* genotyping of independent OA cases (N=1,467) and healthy controls (N=744). Functionality of the mutation was investigated in an *in vitro* cell-based model, and generalizability was investigated by comparing gene expression in OA affected cartilage with preserved cartilage (N=33).

An overview of the workflow for the analyses, identification of the causal mutant, and follow-up on the mutation is provided in Supplementary Figure 1.

Study populations

Cases:

FOA family members. Previously, we reported on an early onset family with familial generalized osteoarthritis (FOA MIM 165720·1) without any dysplasia, and with radiographic chondrocalcinosis as reflected by calcifications within the joint- and fibro-cartilage (Figure 1A and Supplementary Table 1).[18] Family members were collected through Rheumatology clinics from different parts of the Netherlands and France. The age of onset of osteoarthritis in the family varies between 30 and 50 years and osteoarthritis arises in the absence of mild or severe chondrodysplasia. The phenotype within these families is characterized by distinct progressive osteoarthritis with symptoms and radiographic characteristics of osteoarthritis at multiple joint sites simultaneously, including involvement of the hands with noduli, knees, hips, and spine. More recently, relevant additional family members were ascertained (Figure 1B, subject 4, 5, and 16) with the unique FOA phenotype of this family. The study was approved by the medical ethic committee and written informed consent was obtained from all participants.

Controls:

Occurrence of the OPG read-through mutation in the population was tested by *de novo* genotyping in healthy subjects and OA cases. Middle-aged partners (N=744) of the offspring of nonagenarian siblings from the Leiden Longevity study (LLS)[20] were considered as controls for OA cases and are further named 'random controls'. OA cases included were selected from the Genetics osteoarthritis and Progression (GARP) study (N=177),[21] the Patients Prospectively Recruited in Knee and hip Arthroplasty (PAPRIKA) study performed at the Leiden University Medical Center (Dept. Orthopedics; N=1,137),[22] and the ongoing Research Arthritis and Articular Cartilage (RAAK) study (N=153) aimed at the biobanking of joint materials (cartilage, bone and where available ligaments) and bone marrow derived mesenchymal stem cells (hip joints only) of OA patients and controls in the Leiden University Medical Center and collaborating outpatient clinics in the Leiden area. More detailed description of the studies included can be found in the Supplementary methods.

Ethical permission for all studies described was obtained from the appropriate medical ethical committee. Written informed consent was obtained from all participants following detailed explanation of the study.

Exome sequencing. Exome sequencing of the FOA family members was performed by Illumina HiSeq2000 technology (Beijing

Genome Institute). The sequences of each individual were generated as 100 bp paired-end reads, after enrichment of 44 MB exonic sequences by NimbleGen EZ (Roche NimbleGen, Madison, WI). Raw image files were processed by Illumina base calling SoftwareV1.7 with default parameters. SOAPaligner/SOAP2.21 was used to map reads onto the hg19 reference genome (GRCh37) at UCSC. Approximately, 71% of bases originated from the targeted exome, resulting in a mean coverage of 55.93 and 55.81-fold for subjects one and four, respectively. More than 83% of the targeted exons were covered more than ten times. Single-nucleotide variants were subsequently called by the SOAPsnps. The final variant-filtering scheme, which focused on sharing of novel[23] pathogenic mutations between the distant related family members, is detailed in Supplementary Table 2.

Genotyping. Heterozygous deleterious gene variants shared among the two family members were chosen to fit efficiently in a Sequenom multiplex assay designed by the Assay Designer software version 3.1 (Sequenom, San Diego, CA). Single nucleotide polymorphisms (SNPs) were genotyped by mass spectrometry (the homogeneous MassARRAY system; Sequenom, San Diego, CA) using standard conditions and as described elsewhere.[24] In short, PCR reactions were carried out in a final volume of 5 µl and contained 2.5 ng of genomic DNA. Genotypes were assigned by using Genotyper version 3.0 software (Sequenom, San Diego, CA). Variants were genotyped across available FOA members (Supplementary Table 1 and Supplementary Table 3), osteoarthritis case studies (GARP, PAPRIKA study and RAAK), and random controls (LLS).

Linkage analysis. To assess the logarithm of the odds (LOD) score of the mutation in *TNFRSF11B* within the extended family, two point model-based linkage analysis was performed using the FASTLINK 2.2 version of the linkage program MLINK.[25] The disease locus was modeled as an autosomal-dominant trait with a disease frequency of 0.001, as symptomatic and radiographic osteoarthritis at multiple joint sites before the age of 50 years is rare. The *TNFRSF11B* allele frequency was set at 0.01. To perform a conservative analysis penetrance were assumed to be 100% among definitively phenotyped individuals (Patient 1, 2, 3, 4, 5, 6, 7, 8, 12, 14, and 16, Figure 1, and Supplementary Table 1). Family members with uncertain phenotype were considered unknown (Patient 9, 10, 13, and 15, in Figure 1 and Supplementary Table 1).

In vitro **mutagenesis.** A mammalian expression vector (pORF9) containing the human *TNFRSF11B* open reading frame was obtained from InvivoGen. The point mutation causing read-through was introduced using the QuickChange Site-Directed Mutagenesis kit (Agilent Technologies Inc.) according to the manufacturer's protocol with the following primers: Forward 5'-TAAGCTGCTTATCA CTGGAAATGGCCATTGAGC-3';
Reverse 5'-GCTCAATGGCCATTTCCAGT GATAAGCAGCTTA-3'. Introduction of the aimed mutation and exclusion of additional mutations by PCR artifacts was confirmed by sequencing from both strands.

Production of recombinant osteoprotegerin and Western blot analysis. For the production of recombinant osteoprotegerin (OPG), HEK293T cells were cultured in OPTI-MEM I-Reduced Serum Medium (Invitrogen). Cells were transfected either with empty vector or vector with wild-type or mutant *TNFRSF11B* using FuGENE 6 reagent (Roche). Tissue culture supernatants were collected after three days, cleared of debris by centrifugation, and stored frozen. To confirm expression and quality of the recombinant proteins in the conditioned media, Western blotting was performed on 45 µl tissue culture supernatant as described previously[26] using the rabbit monoclonal antibody EPR3592 (Epitomics; 1:5000 diluted) to detect OPG. Subsequently, concentration of OPG in the tissue culture supernatant was determined by ELISA (R&D Systems) following the manufacturer's instructions.

***In vitro* osteoclastogenesis.** Putative difference in the activity of wild-type and mutant OPG was investigated by applying an *in vitro* model of osteoclastogenesis on 96-wells plates with pieces of human bone. Poietics™ human osteoclast precursors (Lonza) were seeded onto the OsteoAssay Human Bone Plate (Lonza; 10^4 cells per well) in the absence (negative control) or in the presence (positive control) of RANKL. Either nothing, or equal volumes of tissue culture supernatant from the HEK293T cells was added to osteoclast growth and differentiation medium (Lonza) to a final concentration of 100ng/ml OPG. Osteoclast precursors were cultured for 6 days before adding fresh osteoclast growth and differentiation medium. Subsequently, cells were grown for 16-18 hours and media were harvested for analysis.

To analyze osteoclastogenesis, tartrate-resistant acid phosphatase staining (TRAP; Takara Bio Inc.) at day 7 of poietics human pre-osteoclast cultures was carried out following the manufacturer's protocol. Concentration of CTX-I in the media was determined by applying CTX-I CrossLaps® enzyme-linked immunosorbent assays (ELISAs; IDS Ltd.) according to the manufacturer's protocol. The amount of CTX-I released in the tissue culture supernatant reflects the amount of bone resorption by mature osteoclasts. The *in vitro* osteoclastogenesis assays were carried out 2 times in triplicate.

Gene expression analysis. To explore the generalizability of the discovered pathway towards common forms of osteoarthritis, mRNA expression was analyzed in an available dataset (Illumina HT-12 v3 microarray) of 33 paired preserved and osteoarthritic cartilage samples of the RAAK study. Following RNA isolation of the collected cartilage, samples were prepared for gene expression by microarray analysis as described in detail in the Supplementary methods. Subsequently, data were exported for analyses using Limma.[27] As implemented in Limma, a paired t-test was applied to all sample pairs. Overall mean normalized probe expression levels of the measured genes in cartilage ranged from 6.58 to 14.91 with a mean of 7.4 and a median of 7.1 with a strong right tailing.

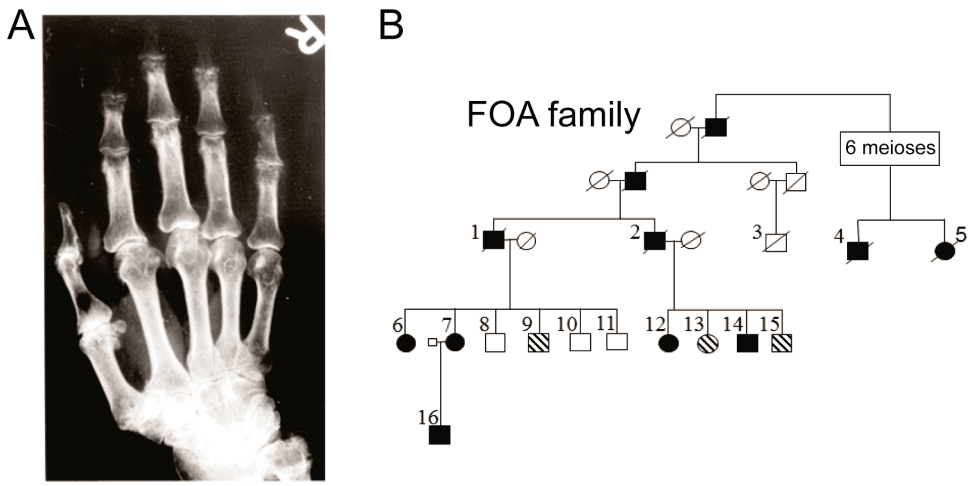


Figure 1. Early onset family with familial generalized osteoarthritis (FOA) without any dysplasia. A. Pedigree of the family with primary osteoarthritis at multiple joint locations without dysplasia with radiographic chondrocalcinosis. The hatched symbols represents diagnostic uncertainty, the filled symbols represent affected subjects. Indicated are phenotypic relevant family members. DNA of distant relatives one and four were subjected to exome sequencing. **B.** Radiograph of individual two at age 65 years showing hand joints with marked osteoarthritis at DIP, PIP, CMC1, MCP-I, MCP-II. Heberden and Bouchard nodes are visible. Figure as provided previously.[18]

RESULTS

Identification of read-through mutation in *TNFRSF11B*. Whole exome sequencing was applied to two distant FOA family members (Figure 1B; subject 1 and 4) and identified 57,018 and 60,652 variants, respectively, which fulfilled the quality criteria in each patient. To identify pathogenic mutations a prioritization scheme was applied in which we excluded variants present in dbSNPv132 (including 1000 Genomes Pilot project data), synonymous variants, tolerated missense variants as determined by SIFT (http://sift.jcvi.org/), intergenic variants and intron variants. In addition, we excluded variants detected in in-house analyzed human genomes (N=221), in-house analyzed human exomes (N=43), and analyzed independent human genomes by the Genome of the Netherlands project (GoNL N=473; www.nlgenome.nl).[28] This reduced the number of variants to, respectively, 102 and 105 variants per subject that were predicted to affect gene function (see Supplementary Table 2). Given that these two FOA patients were distantly related (Figure 1B subject 1 and 4; six meioses apart) they are expected to share the causal variant. We found in total five heterozygous deleterious gene variants that were shared among the two patients: 3 novel missense damaging, 1 splice site mutation, and 1 read-through mutation located in, respectively, the preferentially expressed antigen in melanoma family member 22 (*PRAMEF22*), dynein axonemal heavy chain 10 (*DNAH10*), tetraspanin 8 (*TSPAN8*), mitotic control homolog gene (*DIS3*), and the tumor necrosis factor receptor superfamily member 11b (*TNFRSF11B*; Table 1).

Genotyping of these variants in all

Table 1. Detected novel variants overlapping in family member 1 and 4. * Variants are novel, missense damaging, splice-sites or read-through, not present in dbSNPv132 (including 1000 Genomes Pilot project data), and not detected in in-house analyzed human genomes (LLS, N=221), in-house analyzed human exomes (GARP, N=43) and analyzed human genomes by means of the Genome of the Netherlands Project (GoNL, N=473).

Legend: CHR: chromosome; POS: chromosomal position; Base: base change; codon: codon change; AA subst.: amino acid substitution; Expr: relative expression level as measured in the microarray analyses of 33 osteoarthritic and preserved paired cartilage samples; *p*: nominal P-value of the differential expression between paired osteoarthritic and preserved samples (N=33 pairs); ND: not determined.

CHR	POS	Base*	Codon*	AA subst	Gene name	Expr	*p*
1	13036596	T/C	ATA-AcA	I223T	Preferentially expressed antigen in melanoma family member 22 (*PRAMEF22*)	ND	ND
12	124281300	C/A	TCT-TaT	S577Y	Dynein, axonemal, heavy chain 10 (*DNAH10*)	7.1	7.1×10^{-1}
12	71523146	C/A	GTT-tTT	V209F	Tetraspanin 8 (*TSPAN8*)	7.2	9.9×10^{-1}
13	73335781	C/T	splice site	splice site	Mitotic control homolog gene (*DIS3*)	7.3	9.8×10^{-1}
8	119936614	T/A	TAA=>TTA	'*=>L'	Tumor necrosis factor receptor superfamily, member 11b (*TNFRSF11B*)	10.9	2.9×10^{-8}

members of the extended FOA family (N=16, Figure 1) showed complete co-segregation with the phenotype across three generations only for the read-through mutation within the *TNFRSF11B* gene, resulting in a genome wide significant linkage analysis at this locus (LOD-score of 3.48; Supplementary Table 4). Furthermore, the variant did not occur in 1467 other osteoarthritis affected subjects (the GARP, RAAK, and PAPRIKA study) nor in 744 random controls (Leiden Longevity study) indicating that the mutation may be private for this family. These data indicate that the A to T nucleotide change (A1205T) at the STOP codon of the *TNFRSF11B* gene encoding osteoprotegerin is likely causal to the FOA phenotype in this family. The read-through mutation results in the generation of a Leucine amino acid (X402L) and subsequent 19 additional amino acids at the C-terminus of OPG (c.1205A=>T; p.Stop402Leu; Supplementary Figure 3).

Effect of read-through mutation on OPG protein function. OPG is a soluble decoy receptor which inhibits osteoclastogenesis by competing with the receptor activator of the nuclear factor-KB (RANK), expressed on the membrane of pre-osteoclasts, for the binding of the nuclear factor-KB ligand (RANKL; Supplementary Figure 4). To investigate the putative effect of the (X402L) read-through mutation on OPG protein function, its activity was investigated in an *in vitro* osteoclastogenesis assay. Mutant OPG gene was generated by site directed mutagenesis and recombinant wild-type and mutant OPG proteins were produced in transfected HEK293T cells. Protein integrity was confirmed by Western blotting (Supplementary Figure 5). Subsequently, recombinant proteins were added to

differentiating poietics human pre-osteoclasts seeded onto OsteoAssay™ Human Bone Plates. After 1 week, efficiency of bone resorption upon osteoclastogenesis of the osteoclast precursors was quantified by determining the amount of CTX-I released in the medium of the differentiating osteoclasts (Figure 2A). A clear difference could be observed between cells cultured in the presence or in the absence of OPG. Figure 2 shows that the addition of RANKL to pre-osteoclasts greatly enhanced osteoclastogenesis as can be concluded from the higher concentrations of CTX-I (Figure 2A, +). Adding medium from HEK293T cells transfected with empty vector did not have an effect (Figure 2A, '+ sup'). However, addition of recombinant wild-type OPG (Figure 2A, '+ wt OPG') or mutant OPG (Figure 2A, '+ mt OPG') significantly reduced bone resorption, respectively, 1.2-fold (p=0.049) and 1.6-fold (p=8.4x10^{-5}) reflecting a statistical significant increased capacity of the mutant-OPG to inhibit osteoclastogenesis as compared to the wild type (p=4.7x10^{-3}). To visually inspect the osteoclastogenesis in this experiment

Generalizability of the OPG/RANK/RANKL pathway in osteoarthritis. To investigate the a tartrate-resistant acid phosphatase (TRAP) staining was performed for the different conditions. Mature osteoclasts are distinguished as large multinucleated TRAP positive cells (Figure 2B, indicated with an arrow). Together, these data demonstrate that the identified read-through mutation in *TNFRSF11B*, causing the FOA phenotype, enhances OPG protein function.response of *TNFRSF11B* and the tumor necrosis factor superfamily member 11 (*TNFSF11*) gene, encoding RANKL, to the general osteoarthritis pathophysiological process in articular cartilage, we explored a microarray expression dataset of osteoarthritis affected and preserved cartilage, taken from the same joints, of 33 independent osteoarthritis patients that underwent joint replacement due to end stage osteoarthritis disease (RAAK study). *TNFRSF11A* encoding RANK appeared to be expressed at a very low level (in the lowest quartile of expression) and was therefore not detected well. In contrast, the expression levels of *TNFRSF11B* in osteoarthritic and preserved cartilage samples appeared in the highest expression quartile with little inter-individual variation. A highly significant up-regulation of both *TNFRSF11B* (2.1-fold change, p=1.9x10^{-8}) and *TNFSF11* (1.2-fold change, p=3.1x10^{-6}) was observed in osteoarthritic cartilage as compared to preserved cartilage of the same joint, indicating that these genes are responsive to the OA process in articular cartilage (Supplementary Table 5).

It has been shown that bone homeostasis merely depends on the ratio between *TNFRSF11B* and *TNFSF11*.[29] Therefore, we compared this ratio in both, osteoarthritic and preserved cartilage. As shown in Figure 3, we found a significant increase of the *TNFRSF11B*/*TNFSF11* ratio (p=2.0x10^{-4}) in osteoarthritic articular cartilage which may contribute to respective mineralization of the cartilage and eventually formation of bone, a major hallmark of the ongoing osteoarthritis disease process.[30]

DISCUSSION

We have identified a heterozygous read-through mutation (c.1205A=>T; p.Stop402

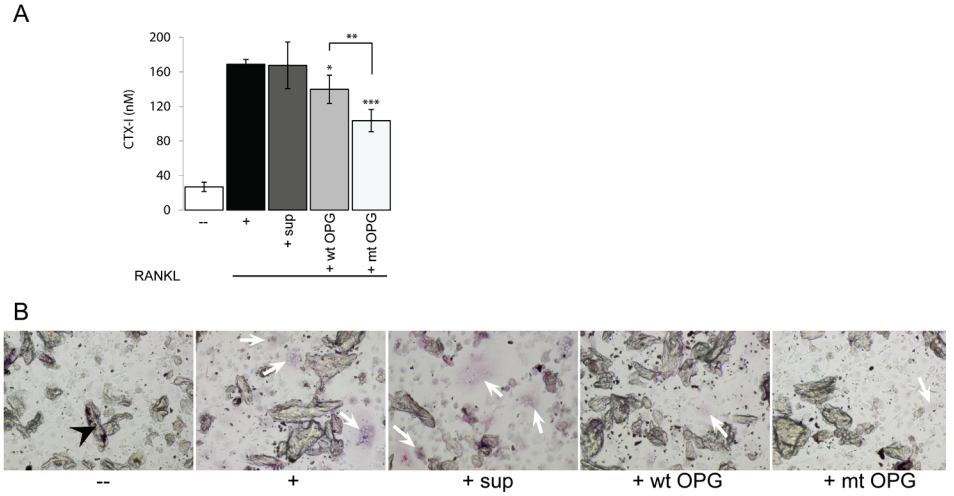


Figure 2. Read-through mutation in the *TNFRSF11B* gene results in gain of OPG function. **A.** Levels of CTX-I were measured in the medium of the assay. Results shown are the average of two independent assays performed in triplicate (*p <0.05 **p <0.005 ***p <0.0001 ANOVA) . Legend: '--': nothing added; '+': RANKL added, '+ sup': RANKL and medium from HEK293T cells transfected with empty vector added; '+ wt OPG': RANKL and recombinant wild-type OPG added; '+ mt OPG': RANKL and recombinant mutant OPG added. Added OPG was present in the tissue culture supernatant at a concentration of 100 ng/ml. **B.** Representative example of histochemical staining for Tartrate-resistant acid phosphatase (TRAP) at day 7 of poietics human pre-osteoclast cultures on the OsteoAssay™ Human Bone Plate. Active and fully differentiated osteoclasts are detected as large multinucleate cells (generally accepted standard for mature osteoclasts is presence of 3 or more nuclei per cell) which are characterized by high TRAP expression (visualized as light purple; white arrows indicate some of the osteoclasts shown in the pictures and black arrowhead points at an example of a bone slices).

Leu) in *TNFRSF11B* that is likely causal to the severe osteoarthritis phenotype in the here studied extended FOA family. Functional analyses showed that this mutation establishes a gain of function of the encoding protein OPG. Given that the mutant OPG decoy receptor more efficiently antagonizes osteoclastogenesis and the FOA phenotype is characterized by calcifications within the joint- and fibro-cartilage, we hypothesize that the detected mutation acts via an unfavorable interplay between (subchondral) bone and cartilage towards enhanced matrix mineralization. Gene expression analyses of preserved and osteoarthritic articular cartilage of independent OA patients showed that enhanced OPG-mediated antagonism is a general phenomenon of the pathophysiological process within osteoarthritic articular cartilage. This is in line with other reports showing increased expression of OPG in arthritic cartilage compared to healthy cartilage.[31, 32] Also during intervertebral disc degeneration, a process highly comparable to the pathological processes occurring in articular cartilage during OA, OPG appeared to be increased in correlation with degeneration grade and coinciding with the presence of microscopic calcifications.[33] Albeit that the role of the molecular pathway of OPG has

been studied in osteoarthritis, a direct causal role towards matrix mineralization has not been reported up until now and it may provide new clues to the development of anti-osteoarthritic drugs.

Retrospectively, we observed in our family a positional linkage signal at the position of the *TNFRSF11B* gene (Supplementary Table 3).[19, 34] Additionally, of note is a report of Baldwin *et al.* that indicated linkage to chromosome 8q in a family with early onset osteoarthritis and chondrocalcinosis at a locus encompassing the *TNFRSF11B* gene.[35] Given the results, it would be interesting to further investigate the status of the *TNFRSF11B* gene of the reported family by next generation sequencing. Although *TNFRSF11B* was not previously detected as OA susceptibility gene by virtue of genome wide association studies, consistent genetic associations of *TNFRSF11B* with knee OA were found in the general population,[32, 36] in particular in association with OA progression as measured by changes in osteophyte grade. More investigations are necessary to explore whether genetic variations in genes of the OPG pathway confer risk to common OA.

Transgenic mice that overexpress *TNFRSF11B* had increased skeletal radiodensity, osteopetrosis, and increased bone density due to decreased numbers of mature osteoclasts,[37] however, no reference or investigation has been made to osteoarthritis as being part of the phenotypes of these mice or, for that matter, radiographic chondrocalcinosis. Since long, epidemiological studies have shown that high bone mineral density is associated with development of OA[38, 39] and may also affect OA progression.[40] Therefore, alternatively to the mentioned enhanced matrix mineralization of the cartilage,

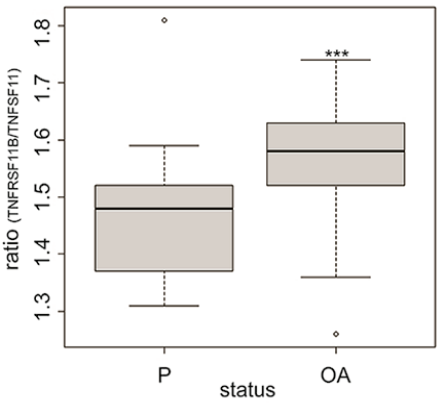


Figure 3. Increased *TNFRSF11B* gene expression in osteoarthritic as compared to preserved cartilage. Boxplots showing the ratio in expression of *TNFRSF11B* and *TNFSF11* in preserved (P) and osteoarthritic (OA) cartilage (ratio of the individual samples is shown as gray dots). Data is from micro-array expression of N=33 patients that underwent joint replacement due to end stage OA. Ratios are determined using the overall relative light intensity signals for both genes as provided by Illumina™ in respectively the preserved and osteoarthritic cartilage (***p <0.0005).

expression of mutant OPG could result in systemic higher bone mass thereby predisposing to OA due to a suboptimal cartilage-bone interplay. Since we have no bone density measurements available of our FOA family members, the role of systemic high bone density in the osteoarthritis onset requires further investigation. In contrast, loss of function mutations in the *TNFRSF11B* gene, that abolish the inhibitory action on osteoclastogenesis, have been found to cause Juvenile Pagets disease characterized by accelerated bone turnover, fractures of the long bone and hyperphosphatasia.[41] Furthermore, SNPs within the *TNFRSF11B* gene have been found to be associated in genome wide association studies to common osteoporosis or low bone mineral density.[42]

Despite the fact that it has been convincingly demonstrated that a strong interaction between subchondral bone and the articular cartilage exists and that changes in mineralization of both bone and cartilage reflect pathophysiology of osteoarthritis several seemingly contradictory observations have been made and the direction of the changes of the subchondral bone density and mineralization in osteoarthritis patients appeared unclear and possibly dependable on the disease status.[43, 44] More recently, osteoarthritis patients were shown to benefit from treatment with strontium ranelate, which is a drug licensed for osteoporosis and acts by increasing bone formation while decreasing bone resorption.[45] These studies, however, have been subject of debate[46, 47] and seem in contrast with our current findings and with the extensive literature indicating that individuals with high systemic bone mass are at increased risk for the incidence of osteoarthritis.[48] It could be speculated that, initially, strontium ranelate enhances stabilization of the dynamically changing OPG/RANKL ratios as it was observed across different grades of OA,[49] resulting in OPG and RANKL levels in the joint balancing towards a net ratio of bone formation. This may temporarily strengthening the joint and decelerate OA progression. Although therapeutic treatment of rats with another bone-forming agent, OPG-Fc, did not directly improve the OA phenotype, it did decrease pain in particular when starting the treatment at early stages.[50] Increased bone thickness as a result of the treatment may prevent enervation of the joint leading to a 'desensitization'. Careful investigation of the long term effects of these bone forming therapies in osteoarthritis is, definitely, necessary[30] thereby taking the existence of different (sub)types of OA characterized by differences in bone mineral density[51] into account.

The molecular pathway of the OPG/RANKL in terms of its manipulation for therapeutic benefit has been studied extensively, especially to enhance the OPG-mediated RANK antagonism in osteoporosis, or, in the case of cancer, growth of bone metastasis.[52] These studies showed that all of the C-terminal domains of OPG are necessary for high affinity association with RANKL.[53] As a result, the design of short recombinant, monomeric OPG molecules that mimic both the receptor binding and dimerization functions of the multi-domain, 100-kDa homodimeric OPG polypeptide has been indicated as challenging.[52] Given that the here detected read-through mutation results in the addition of 19 amino acids at the C-terminus of the OPG protein, we hypothesize that the gain of function may be established through an enhanced stability of OPG dimers thereby increasing the efficiency of RANKL binding. The use of the here identified X402L mutant OPG protein could provide a putative new starting point to further establish such a stable molecule.

In summary, we have identified a read-through, gain of function mutation within the *TNFRSF11B* gene that is likely causal to the FOA phenotype in the here studied family. With this result, we are the first to demonstrate that enhanced OPG mediated RANK antagonism could be directly underlying the onset of osteoarthritis. The exact mechanism on how this is brought about remains to be established. Given that enhanced OPG-mediated antagonism appears to be a more general phenomenon in

the pathophysiological process of osteoarthritic cartilage, we advocate that, in contrast to the use of bone-forming therapies, agents counteracting OPG function could comply with the development of new disease modifying treatments in osteoarthritis.

ACKNOWLEDGEMENTS

The authors want to thank all patients and subjects for their participation. We would like to thank L. Pelascini for providing us with the HEK293T cells and Dr. L. J. Hawinkels for valuable suggestions regarding the ELISAs. We are grateful to Dr. J. K. van der Korst, Dr. P. J. J. M. Rompa, Dr. Kahn, and Dr. J. J. Rasker for providing us with the x-rays and medical histories of their patients.

Licence for Publication: The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in ARD and any other BMJPGL products, and sublicences such use and exploit all subsidiary rights, as set out in the licence.

Author Contributions: YR, ES and IM contributed to the overall study design, data interpretation and writing of the manuscript. YR performed the functional studies. RB contributed to DNA sample preparations and confirmation of exome sequencing findings by *de novo* genotyping. SB contributed the mRNA expression analyses. MK and RN contributed to collection and interpretation of clinical data and cartilage samples. KY and EWL contributed to the bio-informatics of the exome sequencing data. All authors read and approved the manuscript.

Competing Interest: Authors have no conflict of interest to declare.

REFERENCES

1 Woolf AD, Pfleger B. Burden of major musculoskeletal conditions. *Bull World Health Organ* 2003;**81**:646-56.

2 Nuesch E, Dieppe P, Reichenbach S, Williams S, Iff S, Juni P. All cause and disease specific mortality in patients with knee or hip osteoarthritis: population based cohort study. *BMJ* 2011;**342**:d1165.

3 Evangelou E, Chapman K, Meulenbelt I, Karassa FB, Loughlin J, Carr A, *et al.* Large-scale analysis of association between GDF5 and FRZB variants and osteoarthritis of the hip, knee, and hand. *Arthritis Rheum* 2009;**60**:1710-21.

4 Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* 2012;**380**:815-23.

5 Goldstein DB, Allen A, Keebler J, Margulies EH, Petrou S, Petrovski S, *et al.* Sequencing studies in human genetics: design and interpretation. *Nat Rev Genet* 2013;**14**:460-70.

6 Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet* 2013;**14**:681-91.

7 Kannu P, Bateman JF, Randle S, Cowie S, du SD, McGrath S, *et al.* Premature arthritis is a

distinct type II collagen phenotype. *Arthritis Rheum* 2010;**62**:1421-30.

8 Seemann P, Schwappacher R, Kjaer KW, Krakow D, Lehmann K, Dawson K, *et al.* Activating and deactivating mutations in the receptor interaction site of GDF5 cause symphalangism or brachydactyly type A2. *J Clin Invest* 2005;**115**:2373-81.

9 Avcin T, Makitie O, Susic M, Miller S, Thorne C, Tenenbaum J, *et al.* Early-onset osteoarthritis due to otospondylomegaepiphyseal dysplasia in a family with a novel splicing mutation of the COL11A2 gene. *J Rheumatol* 2008;**35**:920-6.

10 Tompson SW, Bacino CA, Safina NP, Bober MB, Proud VK, Funari T, *et al.* Fibrochondrogenesis results from mutations in the COL11A1 type XI collagen gene. *Am J Hum Genet* 2010;**87**:708-12.

11 Mu SC, Lin YJ, Liu HC, Wu JY, Li SC, Michael Lee MT, *et al.* A mutation in cartilage oligomeric matrix protein (COMP) causes early-onset osteoarthritis in a large kindred study. *Ann Hum Genet* 2011;**75**:575-83.

12 van de Laar IM, Oldenburg RA, Pals G, Roos-Hesselink JW, de Graaf BM, Verhagen JM, *et al.* Mutations in SMAD3 cause a syndromic form of aortic aneurysms and dissections with early-onset osteoarthritis. *Nat Genet* 2011;**43**:121-6.

13 Hilhorst-Hofstee Y, Scholte AJ, Rijlaarsdam ME, van HA, Kroft LJ, Reijnierse M, *et al.* An unanticipated copy number variant of chromosome 15 disrupting SMAD3 reveals a three-generation family at serious risk for aortic dissection. *Clin Genet* 2013;**83**:337-44.

14 Chapman K, Takahashi A, Meulenbelt I, Watson C, Rodriguez-Lopez J, Egli R, *et al.* A meta-analysis of European and Asian cohorts reveals a global role of a functional SNP in the 5' UTR of GDF5 with osteoarthritis susceptibility. *Hum Mol Genet* 2008;**17**:1497-504.

15 Valdes AM, Spector TD, Tamm A, Kisand K, Doherty SA, Dennison EM, *et al.* Genetic variation in the SMAD3 gene is associated with hip and knee osteoarthritis. *Arthritis Rheum* 2010;**62**:2347-52.

16 Mio F, Chiba K, Hirose Y, Kawaguchi Y, Mikami Y, Oya T, *et al.* A functional polymorphism in COL11A1, which encodes the alpha 1 chain of type XI collagen, is associated with susceptibility to lumbar disc herniation. *Am J Hum Genet* 2007;**81**:1271-7.

17 Panoutsopoulou K, Southam L, Elliott KS, Wrayner N, Zhai G, Beazley C, *et al.* Insights into the genetic architecture of osteoarthritis from stage 1 of the arcOGEN study. *Ann Rheum Dis* 2011;**70**:864-7.

18 Meulenbelt I, Bijkerk C, Breedveld FC, Slagboom PE. Genetic linkage analysis of 14 candidate gene loci in a family with autosomal dominant osteoarthritis without dysplasia. *J Med Genet* 1997;**34**:1024-7.

19 Min JL, Meulenbelt I, Kloppenburg M, van Duijn CM, Slagboom PE. Mutation analysis of candidate genes within the 2q33.3 linkage area for familial early-onset generalised osteoarthritis. *Eur J Hum Genet* 2007;**15**:791-9.

20 Deelen J, Beekman M, Uh HW, Helmer Q, Kuningas M, Christiansen L, *et al.* Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell* 2011;**10**:686-98.

21 Riyazi N, Meulenbelt I, Kroon HM, Ronday KH, Hellio le Graverand MP, Rosendaal FR, *et al.* Evidence for familial aggregation of hand, hip, and spine but not knee osteoarthritis in siblings with multiple joint involvement: the GARP study. *Ann Rheum Dis* 2005;**64**:438-43.

22 Keurentjes JC, Fiocco M, So-Osman C, Onstenk R, Koopman-Van Gemert AW, Poll RG, *et al.* Patients with severe radiographic osteoarthritis have a better prognosis in physical functioning after hip and knee replacement: a cohort-study. *PLoS One* 2013;**8**:e59500.

23 Boomsma DI, Wijmenga C, Slagboom EP, Swertz MA, Karssen LC, Abdellaoui A, *et al.* The Genome of the Netherlands: design, and project goals. *Eur J Hum Genet* 2014;**22**:221-7.

24 Meulenbelt I, Bos SD, Chapman K, van der Breggen R, Houwing-Duistermaat JJ, Kremer D, *et al.* Meta-analyses of genes modulating intracellular T3 bio-availability reveal a possible role for the DIO3 gene in osteoarthritis susceptibility. *Ann Rheum Dis* 2011;**70**:164-7.

25 Lathrop GM, Lalouel JM, Julier C, Ott J. Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci U S A* 1984;**81**:3443-6.

26 Ramos YF, Stad R, Attema J, Peltenburg LT, van der Eb AJ, Jochemsen AG. Aberrant expression of HDMX proteins in tumor cells correlates with wild-type p53. *Cancer Res* 2001;**61**:1839-42.

27 Goeman JJ, van de Geer SA, de KF, van Houwelingen HC. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics* 2004;**20**:93-9.

28 Kiezun A, Pulit SL, Francioli LC, van DF, Swertz M, Boomsma DI, *et al.* Deleterious alleles in the human genome are on average younger than neutral alleles of the same frequency. *PLoS Genet* 2013;**9**:e1003301.

29 Boyce BF, Xing L. Functions of RANKL/RANK/OPG in bone modeling and remodeling. *Arch Biochem Biophys* 2008;**473**:139-46.

30 Goldring MB, Goldring SR. Articular cartilage and subchondral bone in the pathogenesis of osteoarthritis. *Ann N Y Acad Sci* 2010;**1192**:230-7.

31 Ijiri K, Zerbini LF, Peng H, Otu HH, Tsuchimochi K, Otero M, *et al.* Differential expression of GADD45beta in normal and osteoarthritic cartilage: potential role in homeostasis of articular chondrocytes. *Arthritis Rheum* 2008;**58**:2075-87.

32 Valdes AM, Hart DJ, Jones KA, Surdulescu G, Swarbrick P, Doyle DV, *et al.* Association study of candidate genes for the prevalence and progression of knee osteoarthritis. *Arthritis Rheum* 2004;**50**:2497-507.

33 Rutges JP, Duit RA, Kummer JA, Oner FC, van Rijen MH, Verbout AJ, *et al.* Hypertrophic differentiation and calcification during intervertebral disc degeneration. *Osteoarthritis Cartilage* 2010;**18**:1487-95.

34 Meulenbelt I, Min JL, van Duijn CM, Kloppenburg M, Breedveld FC, Slagboom PE. Strong linkage on 2q33.3 to familial early-onset generalized osteoarthritis and a consideration of two positional candidate genes. *Eur J Hum Genet* 2006;**14**:1280-7.

35 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.

36 Valdes AM, Van OM, Hart DJ, Surdulescu GL, Loughlin J, Doherty M, *et al.* Reproducible genetic associations between candidate genes and clinical knee osteoarthritis in men and women. *Arthritis Rheum* 2006;**54**:533-9.

37 Simonet WS, Lacey DL, Dunstan CR, Kelley M, Chang MS, Luthy R, *et al.* Osteoprotegerin: a novel secreted protein involved in the regulation of bone density. *Cell* 1997;**89**:309-19.

38 Hart DJ, Mootoosamy I, Doyle DV, Spector TD. The relationship between osteoarthritis

and osteoporosis in the general population: the Chingford Study. *Ann Rheum Dis* 1994;**53**:158-62.

39 Sowers M, Lachance L, Jamadar D, Hochberg MC, Hollis B, Crutchfield M, *et al.* The associations of bone mineral density and bone turnover markers with osteoarthritis of the hand and knee in pre- and perimenopausal women. *Arthritis Rheum* 1999;**42**:483-9.

40 Goker B, Sumner DR, Hurwitz DE, Block JA. Bone mineral density varies as a function of the rate of joint space narrowing in the hip. *J Rheumatol* 2000;**27**:735-8.

41 Chong B, Hegde M, Fawkner M, Simonet S, Cassinelli H, Coker M, *et al.* Idiopathic hyperphosphatasia and TNFRSF11B mutations: relationships between phenotype and genotype. *J Bone Miner Res* 2003;**18**:2095-104.

42 Rivadeneira F, Styrkarsdottir U, Estrada K, Halldorsson BV, Hsu YH, Richards JB, *et al.* Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. *Nat Genet* 2009;**41**:1199-206.

43 Bultink IE, Lems WF. Osteoarthritis and osteoporosis: what is the overlap? *Curr Rheumatol Rep* 2013;**15**:328.

44 Tat SK, Pelletier JP, Velasco CR, Padrines M, Martel-Pelletier J. New perspective in osteoarthritis: the OPG and RANKL system as a potential therapeutic target? *Keio J Med* 2009;**58**:29-40.

45 Reginster JY, Badurski J, Bellamy N, Bensen W, Chapurlat R, Chevalier X, *et al.* Efficacy and safety of strontium ranelate in the treatment of knee osteoarthritis: results of a double-blind, randomised placebo-controlled trial. *Ann Rheum Dis* 2013;**72**:179-86.

46 Lafeber FP, van Laar JM. Strontium ranelate: ready for clinical use as disease-modifying osteoarthritis drug? *Ann Rheum Dis* 2013;**72**:157-61.

47 Herrero-Beaumont G, Roman-Blas JA, Largo R, Berenbaum F, Castaneda S. Bone mineral density and joint cartilage: four clinical settings of a complex relationship in osteoarthritis. *Ann Rheum Dis* 2011;**70**:1523-5.

48 Goldring SR. Role of bone in osteoarthritis pathogenesis. *Med Clin North Am* 2009;**93**:25-35, xv.

49 Upton AR, Holding CA, Dharmapatni AA, Haynes DR. The expression of RANKL and OPG in the various grades of osteoarthritic cartilage. *Rheumatol Int* 2012;**32**:535-40.

50 Sagar DR, Ashraf S, Xu L, Burston JJ, Menhinick MR, Poulter CL, *et al.* Osteoprotegerin reduces the development of pain behaviour and joint pathology in a model of osteoarthritis. *Ann Rheum Dis* 2013.

51 Castano-Betancourt MC, Rivadeneira F, Bierma-Zeinstra S, Kerkhof HJ, Hofman A, Uitterlinden AG, *et al.* Bone parameters across different types of hip osteoarthritis and their relationship to osteoporotic fracture risk. *Arthritis Rheum* 2013;**65**:693-700.

52 McGrath EE. OPG/RANKL/RANK pathway as a therapeutic target in cancer. *J Thorac Oncol* 2011;**6**:1468-73.

53 Schneeweis LA, Willard D, Milla ME. Functional dissection of osteoprotegerin and its interaction with receptor activator of NF-kappaB ligand. *J Biol Chem* 2005;**280**:41155-64.

Supplementary methods

STUDIES:

Controls

Leiden Longevity Study (LLS). In the current study, the OPG read-through mutation has been tested for presence among the partners (N=744) of the offspring LLS referred to as random controls. Families were eligible to participate in the LLS study when at least two long-lived siblings were alive and over 89 years old for males and over 91 years for females, representing less than 0.5% of the Dutch population in 2001 (so-called 'nonagenarians').[3]

Cases

Genetics osteoarthritis and Progression (GARP). The GARP study consists of 192 sibling pairs concordant for clinical and radiographically (K/L score) confirmed osteoarthritis at two or more joint sites among hand, spine (cervical or lumbar), knee or hip. Conventional radiographs of the hands (dorso-volar), knees (posterior-anterior (PA) in weight bearing semiflexed and lateral), hips (PA), lumbar (PA and lateral), and cervical spine (anterior-posterior, lateral, and transbuccal) were obtained from all participants. Detailed description of the GARP study has been published previously.[4] Occurrence of the OPG read-through mutation among OA cases was tested by *de novo* genotyping of 177 cases from the GARP study.

Patients Prospectively Recruited in Knee and hip Arthroplasty (PAPRIKA). The Paprika study is performed at the Leiden University Medical Center (Dept. Orthopedics) and consists in a long-term follow-up study of patients that have undergone total joint replacement (TJR) at hip or knee.[5] Patients of Caucasian descent were included when they were diagnosed with primary osteoarthritis based on radiographs and the ACR rheumatology classification criteria. Patients with secondary osteoarthritis or patients requiring a revision were excluded from the PAPRIKA study. For the study described here we genotyped the detected mutations in 1137 patients of the PAPRIKA study.

Research Arthritis and Articular Cartilage (RAAK). The RAAK study is aimed at the biobanking of joint materials (cartilage, bone and where available ligaments) and bone marrow derived mesenchymal stem cells (hip joints only) of OA patients and controls in the Leiden University Medical Center and collaborating outpatient clinics in the Leiden area. In the current study we genotyped the detected mutations in 153 RAAK patients, and determined gene expression in 61 paired samples of osteoarthritic and macroscopically unaffected cartilage.

Gene expression analysis

To explore the generalizability of the discovered pathway towards common forms of osteoarthritis, mRNA expression was analyzed in an available dataset (Illumina HT-12 v3 microarray) of 33 paired preserved and osteoarthritic cartilage samples of the RAAK study.

Cartilage collection. Within the RAAK study we collected paired preserved and OA affected cartilage samples from Caucasian end stage OA patients undergoing joint replacement surgery for primary OA in the Leiden University Medical Centre. At the moment of collection (within

2 hours following surgery) tissue was washed extensively with phosphate buffered saline (PBS) to decrease the risk of contamination by blood, and cartilage was collected of the weight-baring area of the joint. Cartilage was classified macroscopically and collected separately for OA affected and preserved regions (classification was done according to predefined features for OA-related damage based on color/whiteness of the cartilage, based on surface integrity as determined by visible fibrillation/crack formation, and based on depth and hardness of the cartilage upon sampling with a scalpel). During collection with a scalpel, care was taken to avoid contamination with bone or synovium. Collected cartilage was snap frozen in liquid nitrogen and stored at -80°C prior to RNA extraction.

RNA isolation. Cartilage samples were pulverized using a Retsch MM200 under cryogenic conditions. On average 150 mg of pulverized cartilage was dissolved in 1 ml of Trizol reagent, and mixed vigorously. After addition of 200µl of chloroform the sample was mixed and centrifuged for 15 minutes at 16,000g. The clear aqueous layer was transferred to a new vial and 1 volume of 70% ethanol/DEPC-treated water was added to precipitate RNA. The RNA was collected using Qiagen mini columns according to the manufacturers protocol. The RNA quality was assessed using a Bioanalyzer lab-on-a-chip and quantity was assessed using a nanodrop spectrophotometer. RIN values above eight were considered suitable for gene expression analysis.

Microarray expression. 0.5 µg of total RNA was used as template for first- and second-strand reverse transcription steps using an Ambion (Life Technologies Europe BV, Rijswijk, The Netherlands) RNA amplification kit. After in vitro transcription, amplification, and labeling with biotin-labeled nucleotides Illumina HT-12 v3 microarrays were hybridized (Illumina, San Diego). Sample pairs were randomly dispersed over the microarrays, however each pair was measured on a single chip of in total eight chips. The microarrays were read using an Illumina Beadarray 500GX and after basic quality checks using Beadstudio the intensity values were normalized using the "rsn" option in the Lumi R-package. The corresponding signals increase exponentially with relative expression probe-levels. Indicated units are light intensity as provided by Illumina. The subsequently obtained raw probe-level data were exported for analyses using Limma.[6] As implemented in Limma, a paired t-test was applied to all sample pairs. Overall mean normalized probe expression levels of the measured genes in cartilage ranged from 6.58 to 14.91 with a mean of 7.4 and a median of 7.1 with a strong right tailing.

References

1 Meulenbelt I, Bijkerk C, Breedveld FC, Slagboom PE. Genetic linkage analysis of 14 candidate gene loci in a family with autosomal dominant osteoarthritis without dysplasia. *J Med Genet* 1997;**34**:1024-7.

2 Meulenbelt I, Min JL, van Duijn CM, Kloppenburg M, Breedveld FC, Slagboom PE. Strong linkage on 2q33.3 to familial early-onset generalized osteoarthritis and a consideration of two positional candidate genes. *Eur J Hum Genet* 2006;**14**:1280-7.

3 Deelen J, Beekman M, Uh HW, Helmer Q, Kuningas M, Christiansen L, *et al.* Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell* 2011;**10**:686-98.

4 Riyazi N, Meulenbelt I, Kroon HM, Ronday KH, Hellio le Graverand MP, Rosendaal FR, *et al.* Evidence for familial aggregation of hand, hip, and spine but not knee osteoarthritis in siblings with multiple joint involvement: the GARP study. *Ann Rheum Dis* 2005;**64**:438-43.

5 Keurentjes JC, Fiocco M, So-Osman C, Onstenk R, Koopman-Van Gemert AW, Poll RG, *et al.* Patients with severe radiographic osteoarthritis have a better prognosis in physical functioning after hip and knee replacement: a cohort-study. *PLoS One* 2013;**8**:e59500.

6 Goeman JJ, van de Geer SA, de KF, van Houwelingen HC. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics* 2004;**20**:93-9.

Supplementary Table 1. Radiographic and clinical abnormalities in members of a family with autosomal dominant FOA with radiographic signs of chondrocalcinosis. Diagnoses depicted are established at time of ascertainment and as published previously.[1] Legend: age: age of patient at time of taking the radiograph; sex: M = Male, F = Female; OA: Radiographic osteoarthritis (Kellgren Lawrence grade 2 or higher in one or both joints; - = No radiographs available / no joint complaints; N: normal joint bilateral (Kellgren Lawrence grade 0 or 1); H: Heberden nodes; B: Bouchard nodes; C: Radiographic signs of chondrocalcinosis; ? = possible; DIP: distal interphalangeal joints; PIP: proximal interphalangeal joints; MCP: meta-carpophalangeal joints; CMCI: first-carpometacarpal joints; U/L: upper/lower segment ratio; P: Preclinical signs of osteoarthritis (bony enlargements / joint deformity / limited range of movement); BMI = body mass index (kg/m2).

Patient	sex	age	DIP	PIP	MCP	CMC1	Elbow	Shoulder	Hip	Knee	Feet	Ankle	CWK	LWK	Height (m)	BMI (kg/m^2)	U/L
0		79	-	-	-	-	OA	-	N	OA, C	OA	OA	-	-			
1	M	53	OA, H	-	-	OA	OA, C	OA	-	OA, C	OA	OA	OA	OA	1.76	26.8	1.02
2	M	61	OA, H	OA, B	-	OA	P	-	-	OA, C		OA	OA	?	1.64	24.3	0.96
3	M	61	-	-	-	-	-	-	N	-	-	-	OA	OA			
6	F	41	N	N	N	OA	-	-	-	OA	-	N	OA	OA	1.70	26.3	0.83
7	F	31	-	-	-	N	N	-	-	OA, C	-	P	OA	-	1.68	23.4	0.95
8	M	23	-	-	-	-	-	-	-	-	-	-	-	-	1.76		
9	M		N	N	N	N	-	-	N	N, C?	-	-	-	-	1.80		
10	M	18	-	-	-	-	-	-	N	N	-	-	-	-	1.76		
12	F	37	N	N	N	OA	-	-	N	OA	-	-	OA	?			
13	F	36	-	-	-	-	-	-	-	-	-	-	OA	-			
14	M	35	-		-	-	N-	-	-	OA, C	OA	N	OA	N	1.71	27.5	0.91
15	M	26	-		-	-		-	-	?	-	-	-	-	1.83	22.4	0.91

Supplementary Table 2. Variant prioritization scheme after exome sequencing.

	Subject 1	**Subject 4**
Total number of variants identified	57018	60562
exonic (including 5'-utr and 3'-utr) and canonical splice sites	15130	15183
missense, nonsense, readthrough, frame error, splice site	7735	7699
novel	618	636
missense damaging, nonsense, readthrough, frame error, splice site	168	176
not present in in-house whole genome sequencing data (N=694)	110	115
not present in in-house exome sequencing data (N=43)	102	105
shared among distant family members	5	5

Supplementary Table 3. Genotyping of the detected variants in FOA family members.
*variant was not detected by genotyping. ǂDNA not available. ǂTotal number of individuals showing linkage between phenotype and genotype (subjects with diagnostic uncertainty are not taken into account). SS = splice site.

		PRAMEF22	*DNAH10*	*TSPAN8*	*DIS3*	*TNFRSF11B*
FamID	**Status**	**I223T**	**S577Y***	**V209F**	**SS**	***=>L**
1	Affected	**T C**	C C	**C A**	**T C**	**T A**
2	Affected	T T	-	C C	**T C**	**T A**
3	Unaffected[ǂ]	-	-	-	-	-
4	Affected	**T C**	C C	**C A**	**T C**	**T A**
5	Affected	**T C**	C C	**C A**	T T	**T A**
6	Affected	**T C**	C C	C C	**T C**	**T A**
7	Affected	T T	C C	**C A**	**T C**	**T A**
8	Unaffected	T C	C C	C A	T C	**T T**
9	Diagnostic uncertain	T T	C C	C C	T C	T A
10	Unaffected	T C	C C	C A	**T T**	**T T**
11	Unaffected	**T T**	C C	**C C**	**T T**	**T T**
12	Affected	T T	C C	C C	T T	**T A**
13	Diagnostic uncertain	T T	C C	C C	T C	T A
14	Affected	T T	C C	C C	T T	**T A**
15	Diagnostic uncertain	T T	C C	C C	T T	T A
16	Affected	T T	C C	C C	T T	**T C**
	Total linked[ǂ]	5/12	0/12	5/12	7/12	12/12

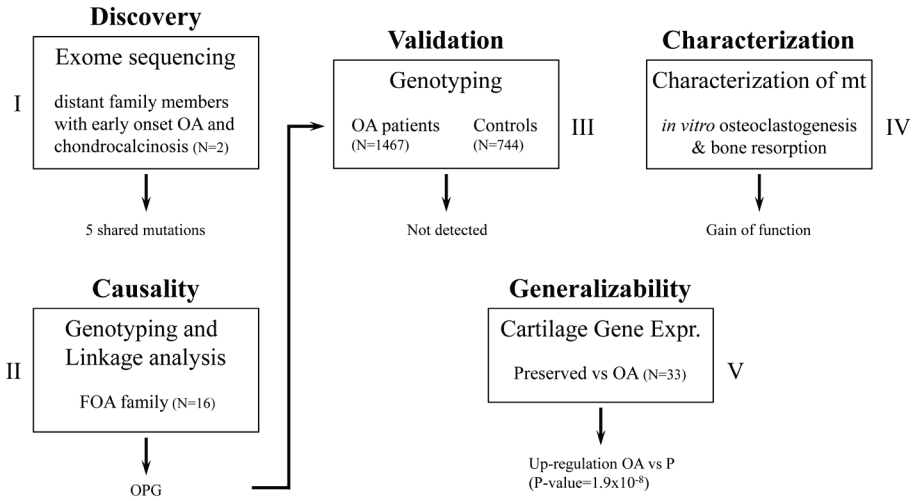
Supplementary Table 4. Two point LOD scores FOA family a. *LOD scores calculated with the original family structure and inheritance pattern[2]
†LOD scores calculated with c.1205A=>T using the updated family structure.

Marker	CHR	POS (start – stop)	LODscore		
			0	0.05	0.1
D8S1132*	8	107328821 - 107329072	1.07	0.96	0.83
TNFRSF11B†	8	119938733 - 119964383	3.48	3.16	3
D8S1179*	8	125907064 - 125907405	0.39	0.33	0.27

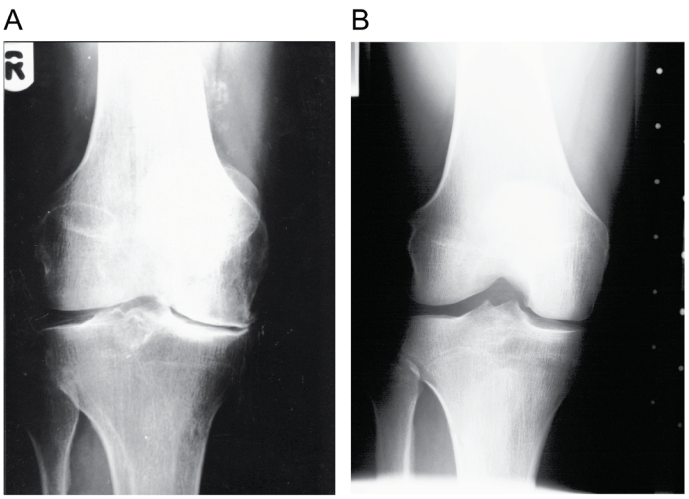
Supplementary Table 5: Expression analyses of *TNFRSF11B* and *TNFSF11* in articular cartilage. Legend: Expr: relative gene expression, assessed by exploring a microarray mRNA expression dataset generated on Illumina HT-12 v3 chips in matched osteoarthritic and preserved cartilage samples of 33 patients that underwent joint replacement due to end stage osteoarthritis (expression levels are expressed as the light intensity values as provided by IlluminaTM. The average level of expression for all probes ranged from 6.58-14.91 with mean 7.4 and median 7.1 with a strong right tailing); Expr QT: Quartile of the expression level where 1 = lowest level of expression and 4 = highest level of expression; FC: fold change between osteoarthritic and preserved cartilage samples (FC > 1 indicated higher expression in osteoarthritic as compared to preserved cartilage).
TNFRSF11B: tumor necrosis factor receptor superfamily member 11b encoding osteoprotegerin (OPG); *TNFSF11*: tumor necrosis factor (ligand) superfamily, member 11 encoding the nuclear factor-KB ligand (RANKL).

Gene	Expr	Expr QT	FC	*p*
TNFRSF11B	10.3	4	2.1	1.9×10^{-8}
TNFSF11	7.1	3	1.2	3.1×10^{-6}

Supplementary Figure 1. Schematic overview of the exome sequencing and follow-up analyses.

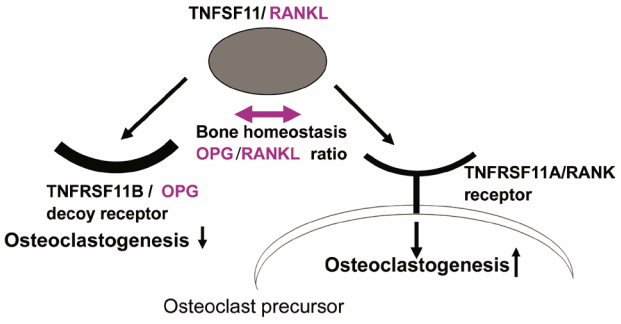


Supplementary Figure 2. Knee radiographs. A Right knee joint of FOA family member one at age 53 years with diagnosed chondrocalcinosis (not clearly visible on this photo) and marked osteoarthritis medially but without evidence of epiphyseal dysplasia.[1] **B** Knee joint of healthy control.

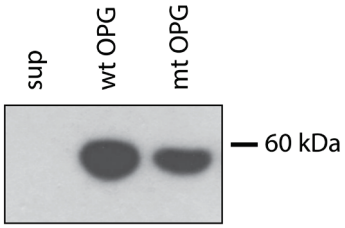


Supplementary Figure 3. Amino acid sequence of wild-type and mutant (X402L) TNFRSF11B or osteoprotegerin. An T>A base pair change at the stop codon in exon 5 (A1205T) results in a STOP to Leucine amino acid change (X402L) in the TNFRSF11B gene. This read through mutation leads to a 19 amino acid elongation of the mutant protein.

Supplementary Figure 4. Bone homeostasis depends on the local OPG/RANK ratio. The RANK ligand (RANKL) competitively binds to the OPG decoy and RANK receptor. Binding of RANKL to the RANK receptor at the osteoclast precursor cell acts as an osteoclast differentiation and activation signal.



Supplementary Figure 5. Recombinant human osteoprotegerin (OPG). To confirm integrity of the recombinant proteins, Western blotting was performed using 10 µl of the concentrated tissue culture supernatant of HEK293T cells 3 days after transfection with either empty vector (sup), wild-type (wt OPG) or mutant OPG (mt OPG), as indicated above the lanes.



Underlying molecular mechanisms of DIO2 susceptibility in symptomatic osteoarthritis

7

Ann Rheum Dis (doi:10.1136/annrheumdis-2013-204739)

Compelling insight into DIO2 function during human in vitro chondrogenesis; towards underlying mechanisms of OA susceptibility

Nils Bomer[1,2†], Wouter den Hollander[1†], Yolande F.M. Ramos[1†], Steffan D. Bos[1,3], Ruud van der Breggen[1], Nico Lakenberg[1], Barry A. Pepers[7], Annelies E. van Eeden[1], Arash Darvishan[8], Elmar W. Tobi[1,2], Bouke J. Duijnisveld[4], Erik B. van den Akker[1,5], Bastiaan T. Heijmans[1,3], Willeke M.C. van Roon-Mom[8], Fons J Verbeek[9], Gerjo J.V.M van Osch[6,7], Rob G.H.H. Nelissen[4], P. Eline Slagboom[1,2,3], Ingrid Meulenbelt[1,3*]

[1]Dept. of Molecular Epidemiology, LUMC, Leiden, The Netherlands. [2]IDEAL The Netherlands. [3]Genomics Initiative, sponsored by the NCHA, Leiden, The Netherlands. [4]Dept. of Orthopaedics, LUMC, Leiden, The Netherlands. [5]The Delft Bioinformatics Lab, Delft University of Technology, Delft, The Netherlands. [6]Dept. of Orthopaedics, Erasmus MC, Rotterdam, The Netherlands. [7]Dept. of Otorhinolaryngology, Erasmus MC, Rotterdam, The Netherlands. [8]Dept. of Human Genetics, LUMC, Leiden, The Netherlands. [9]Dept. of Imaging & BioInformatics, LIACS, Leiden, The Netherlands.

† authors contributed equally

ABSTRACT

Objectives **To investigate how the genetic susceptibility gene *DIO2* confers risk to OA onset in humans and to explore whether counteracting the deleterious effect could contribute to novel therapeutic approaches.**

Methods **Epigenetically regulated expression of *DIO2* was explored by assessing methylation of positional CpG-dinucleotides and the respective *DIO2* expression in OA affected and macroscopically preserved articular cartilage from end stage OA patients. In a human *in vitro* chondrogenesis model, we measured the effects when thyroid signaling during culturing was either enhanced (excess T3 or lentiviral induced *DIO2* over-expression) or decreased (Iopanoic Acid).**

Results **OA-related changes in methylation at a specific CpG dinucleotide upstream of *DIO2* caused significant upregulation of its expression (Beta=4.96; p=0.0016). This effect was enhanced and appeared driven specifically by *DIO2* rs225014 risk-allele carriers (Beta=5.58, p=0.0006). During *in vitro* chondrogenesis *DIO2* overexpression resulted in a significant reduced capacity of chondrocytes to deposit extra cellular matrix (ECM) components, concurrent with significant induction of ECM degrading enzymes (ADAMTS5, MMP13) and markers of mineralization (ALPL, COL1A1). Given their concurrent and significant upregulation of expression, this process is likely mediated via HIF-2α/RUNX2 signalling. In contrast, we showed that inhibiting deiodinases during *in vitro* chondrogenesis contributed to prolonged cartilage homeostasis as reflected by significant increased deposition of ECM components and attenuated upregulation of matrix degrading enzymes.**

Conclusions **Our findings show how genetic variation at *DIO2* could confer risk to OA and raised the possibility that counteracting thyroid signaling may be a novel therapeutic approach.**

Keywords: DIO2; thyroid pathway; chondrogenesis; epigenetics; endochondral ossification

INTRODUCTION

Osteoarthritis (OA) is a prevalent, complex, disabling disease of articular joints, characterized by degradation of articular cartilage and remodeling of the subchondral bone. There is no effective therapy to reverse or slow down the disease except for joint replacement surgery at the end stage. As a result, OA has a large detrimental impact on the quality of life of the elderly and causes a major burden on health and social care.[1] To allow development of new therapies, there is an on-going need for insight into the underlying mechanisms driving OA. Genetic studies provided evidence that genes orchestrating growth plate endochondral ossification play a underlying role in common OA susceptibility,[2] hence functional follow up approaches require focus on both the early developmental and late acting effects of these OA genes. A notable example is the deiodinase iodothyronine type 2 (D2) gene (*DIO2*) with the C-allele of the single nucleotide polymorphism (SNP) rs225014 (frequency ~0.35) located in the coding region that conferred consistent risk to OA.[3, 4] The gene product of *DIO2* is responsible for catalyzing the conversion of intracellular inactive thyroid hormone (T4) to active thyroid hormone (T3). T3 subsequently signals terminal maturation of growth plate chondrocytes leading to cell hypertrophy, cartilage matrix destruction mediated via upregulation of hypoxia inducible factor-2α (*HIF-2a*) and runt-related transcription factor-2 (*RUNX2*),[5, 6] mineralization of the cartilage and eventually formation of bone.[7] There are striking parallels between the chondrocyte signalling events that take place in the growth plate and those of hypertrophic chondrocytes in OA affected articular cartilage.[5] This has led to the hypothesis that with age and environmental stresses the propensity of the highly specialized, maturational arrested articular chondrocytes is affected by loss of epigenic control. Progression of age and disease could result in reactivation of genes involved in endochondral ossification, leading to loss and mineralization of articular cartilage, a process known to contribute to OA.[3, 8-10] Functional genomic studies showed high expression of *DIO2* mRNA and D2 protein levels in osteoarthritic as compared to healthy cartilage.[5, 11, 12] Furthermore, *DIO2* allelic imbalance was assessed and showed that the OA risk allele 'C' was more abundantly present in articular joint tissues than the wild-type allele 'T'.[11] In transgenic rats *DIO2* overexpression conferred risk to articular cartilage destruction.[13] The underlying mechanism how the *DIO2* SNP confers susceptibility to OA in humans remains, however, to be determined[14] but most likely acts via aberrent up-regulation of its expression. The *DIO2* locus in humans contains several putative CCCTC-binding factor (CTCF) binding sites, including one that is overlapping with the rs225014 location.[15] CTCF is considered to facilitate long-range chromatin interactions in order to insulate gene expression and distal transcriptional elements on the genome are brought in close proximity to transcriptional start sites (TSSs) of genes to inhibit expression.[16]

In the current study, we focus on regulatory mechanisms of *DIO2* expression in preserved and osteoarthritic human articular cartilage, thereby taking into account the *DIO2* risk allele. The direct effect of changes in *DIO2* expression on chondrocyte function and

human cartilage extracellular matrix (ECM) homeostasis, is subsequently studied in human *in vitro* chondrogenesis models which should be considered a well defined system for studying changes in the ECM when chondrocytes differentiate, become hypertrophic and start to exhibit cartilage debilitating expression paterns.

MATERIALS AND METHODS

The ongoing RAAK study is aimed at the biobanking of joint materials (cartilage, bone and where available ligaments) and mesenchymal stem cells and primary chondrocytes of patients in the Leiden University Medical Center. In the current study we used paired preserved and OA affected cartilage samples from 52 Caucasian end stage OA patients undergoing joint replacement surgery for primary OA (23 hips, 29 knees). For additional details on the RAAK study, cell cultures, RNA and DNA extraction, quantitative RT-PCR, ChIP and the data analysis, see the Supplemental Methods.

Electrophoretic Mobility Shift Assay. For Electrophoretic mobility shift assays (EMSAs) synthetic oligonucleotides containing the putative CTCF binding site were 5'-end labeled by γ-^{32}P-ATP and subsequently purified by gel filtration on Sephadex G-25 Medium columns. For additional details, see the Supplemental Methods.

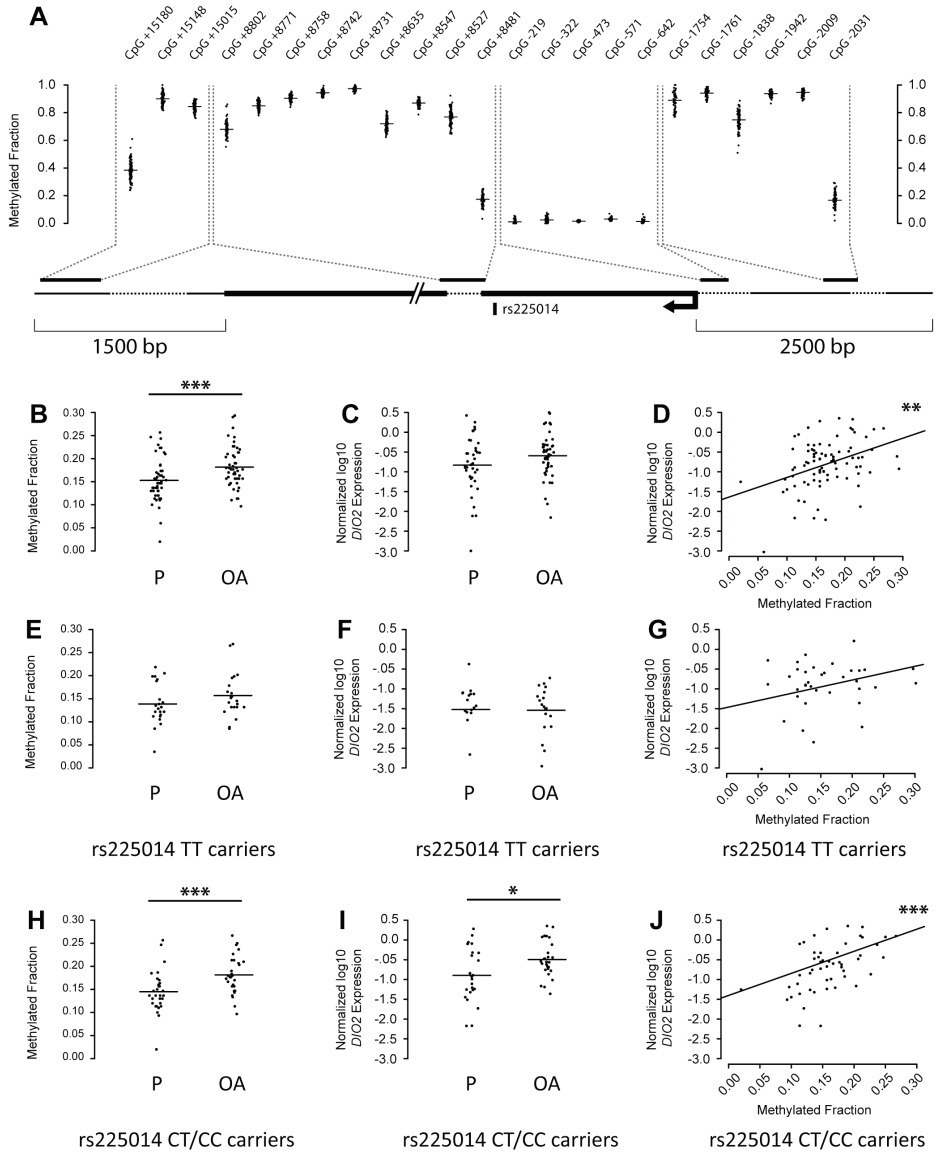
Quantification of methylation. The methylated fraction of CpG dinucleotides was assessed with MALDI-TOF mass spectrometry (Epityper, Sequenom), a commonly applied method to quantify CpG methylation.[17-19] For additional details, see the Supplemental Methods.

Lentiviral constructs and transduction. C-terminal FLAG-tagged cys-D2 (kindly provided by Prof. Dr. Bianco [20]) was digested with *EcoR1* followed by Klenow-treatment and digestion with *Xba1*. Inserts were inserted into the *EcoRV-XbaI* sites of the pLV-CMV-IRES-eGFP Lentiviral backbone. For additional details, see the Supplemental Methods.

***In vitro* chondrogenesis.** 3D pellets were formed using 2.5 x 10^5 hBMSCs. Chondrogenesis was initiated in serum-free chondrogenic differentiation medium. From day 14 onwards cell pellets were maintained either in the standard chondrogenic differentiation medium or in the presence of T_3 (10 nM) or IOP (10 μM). For additional details, see the Supplemental Methods.

Relative pixel intensity. The relative pixel. intensity was computed by loading the

Figure 1. (next page). A functional CpG dinucleotide (CpG -2031) significantly modulates *DIO2* expression. A) Schematic overview of the quantified CpG dinucleotides in pooled preserved and paired OA affected samples across the *DIO2* locus. **B)** Methylation between preserved and OA affected samples for CpG dinucleotide located 2031 base pairs upstream (CpG -2031) of the *DIO2* TSS (GLMM, N=103, Beta: 0.028, *p*=0.001, Bonferroni adjusted; see also **Table S2**). **C)** Real-time RT-qPCR data of *DIO2* expression between preserved and OA affected cartilage (GLMM, N=87, Beta: 0.22, *p*=0.063). **D)** Association between methylation at CpG -2031 and *DIO2* expression (GLMM, N=87, Beta: 4.959, *p*=0.002 Bonferroni adjusted for 23 CpG sites tested). **E)** Methylation between preserved and OA affected cartilage for CpG -2031 among homozygous rs225014 wildtype allele



carriers (GLMM, N=44, Beta: 0.018, p=0.112). **F)** Real-time RT-qPCR data of *DIO2* expression between preserved and OA affected cartilage among homozygous rs225014 wildtype allele carriers (GLMM, N=36, Beta: -0.017, p=0.929). **G)** Association between methylation at CpG-2031 and *DIO2* expression among homozygous rs225014 wildtype allele carriers (GLMM, N=36, Beta: 3.863, p=0.050). **H)** Methylation between preserved and OA affected cartilage for CpG -2031 among the hetero- and homozygous carriers of the rs225014 risk allele (GLMM, N=59, Beta: 0.034, p=2x10^{-5}). **I)** Real-time RT-qPCR data of *DIO2* expression between preserved and OA affected cartilage among rs225014 risk allele carriers (GLMM, N=51, Beta: 0.35, p=0.012). **J)** Association between methylation at CpG -2031 and *DIO2* expression among rs225014 risk allele carriers (GLMM, N=51, Beta: 5.58, p=6x10^{-4}). Legend: *p <0.05, **p <0.01, ***p <0.001. OA: osteoarthritic cartilage P= preserved cartilage.

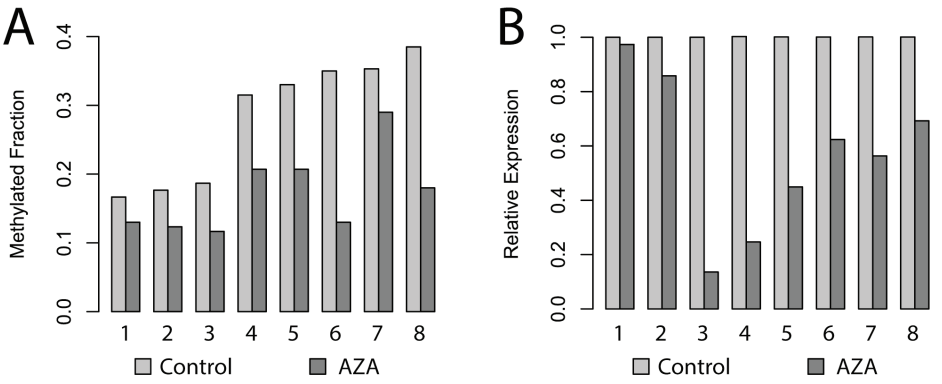


Figure 2. Methylation regulates *DIO2* expression in articular cartilage. Each pair of bars reflects a unique donor cell culture, deriven from total hip replacement patients. **A)** Methylation at dinucleotide CpG -2031 with no treatment and after treatment with 1.5 μM of the demethylating agent 5-aza-2–deoxycytidine (AZA). **B)** Real-time RT-qPCR data of *DIO2* expression with no treatment and after AZA treatment.

photos into ImageJ (v.1.47).[21-23] For additional details, see Supplemental Methods.

RESULTS

Epigenetic regulation of *DIO2* expression by CpG methylation in articular cartilage. Online available ChIP-seq data revealed multiple transcription factors to bind the rs225014 locus of which CTCF was predicted to bind with the highest certainty (Fig. S1). To assess the regulatory properties of the CTCF binding site overlapping rs225014 and to test whether the rs225014 alleles directly affect the binding, we performed an electrophoretic mobility shift assay (EMSA). The putative CTCF sequence overlapping rs225014 (DIO2-CTCF1) was found not functional nor, for that matter, dependent of the rs225014 alleles (Fig. S2, lane 1-6).

As no further TFs were confidently predicted to bind the rs225014 locus, we set out to elucidate putative regulatory mechanisms of *DIO2*, independent of the rs225014 base change. We, therefore, quantified expression of *DIO2* and methylation of 23 CpG dinucleotides across the *DIO2* locus in macroscopically preserved and OA affected cartilage from joints of patients undergoing total arthroplasty of the knee (N=29) or hip (N=23) (RAAK study; Table S1). We found borderline significant upregulation of *DIO2* expression (Beta: 0.22, p=0.063, Fig. 1C) in OA as compared to preserved cartilage. Furthermore, we observed several CpG dinucleotides to be differentially methylated between preserved and OA affected cartilage (Tabe S2). However, only for the CpG site 2031 base pairs upstream of the *DIO2* transcription start site (CpG -2031, Fig. 1A), we observed significant differential methylation between OA and preserved cartilage (Beta: 0.028, p=0.0007, Fig. 1B) and a significant positive association between methylation and *DIO2* expression in all samples (Beta: 4.959; p=0.0016, Fig. 1D; Table S2). To confirm the regulatory properties of CpG -2031 on *DIO2* expression, we applied 5-aza-2'-deoxycytidine (AZA), a demethylating agent, to the culture

medium in 8 primary chondrocyte cultures derived from preserved cartilage of total hip replacement patients. Addition resulted in a decrease in methylation at CpG -2031 (p=0.003, Fig. 2A) corresponding to down regulation of *DIO2* expression (p=0.004, Fig. 2B). Having observed the association between *DIO2* expression and OA associated methylation at CpG -2031, we applied a multivariate model to asses the individual effects of CpG -2031 methylation, joint site, sex, BMI, age and rs225014 alleles on *DIO2* expression in articular cartilage (Table S3). We could hereby ratify the association between increased CpG -2031 methylation and *DIO2* expression (Beta: 4.008, p=0.019). The most notable observation, however, was the significant independent association of rs225014 genotype on *DIO2* expression (Beta: 0.557; p=0.0003), indicating additive effects of both genotype and DNA methylation differences on *DIO2* expression. This effect appeared to be mainly driven by the risk allele. Upon stratification by the rs225014 alleles, no significant differences were observed between preserved and OA affected cartilage among homozygous carriers of the rs225014 wildtype allele T in methylation (Beta: 0.018, p=0.112, Fig. 1E) or expression (Beta: -0.017, p=0.929, Fig. 1F). In carriers of the risk allele, however, an increase was observed in OA cartilage as compared to the preserved tissue, both in the difference of CpG -2031 methylation (Beta: 0.034, p=0.00002, Fig. 1H) and in expression (Beta: 0.35, p=0.012, Fig. 1I), conconmittant with an increased association between methylation and expression (Beta: 6.816, p=0.00001, Fig. 1J). As expected, this association between expression and methylation was much smaller among homozygous wildtype allele carriers (Beta: 3.863, p=0.050, Fig. 1G).

***DIO2* effects on *in vitro* chondrogenesis; *DIO2* overexpression.** To assess the direct effect of *DIO2* upregulation upon cartilage matrix homeostasis, we examined *in vitro* chondrogenesis of human bone marrow derived stem cells (hBMSCs) in a pellet culture for 5 consecutive weeks and generated lentiviral-mediated overexpression of *DIO2* (Fig. 3A and Fig. S3). As a result, we observed a greatly reduced expression of genes encoding the main proteins of articular cartilage ECM, *COL2A1*, *ACAN* and *COL10A1*, by reverse transcriptase qPCR (p<0.0001; Fig. 3A). In parallel, we observed profound and significant upregulated gene expression of the OA markers of hypertrophy and ECM breakdown *ADAMTS5*, *MMP13, RUNX2* and *EPAS1* (encoding HIF-2α) from 3 weeks onwards (p<0.05, Fig. 3A). Histology confirmed a reduced deposition of glycosaminoglycans (GAGs; Fig. 3B, 3C). Although we did not have data to apply proper statistics to produce a significant number, pixel intensity measurements of the pictures reproduced as in Figure 3 (B-C) showed us a 35% difference after 35 days between control and *DIO2* over-expressing pellets, concurrent with visual lower collagen type-II (COL2) and collagen type-X (COL10) protein expression by immunohistochemical staining at consecutive weeks as compared to controls (Fig. 3D-G).

***DIO2* effects on *in vitro* chondrogenesis; thyroid hormone signaling.** To explore whether the effects of *DIO2* overexpression is due to enhanced thyroid signaling, we examined *in vitro* chondrogenesis of hBMSCs from 5 different donors (Table S4) in pellet

cultures for 7 consecutive weeks (Fig. S4) while adding active T3 to the chondrogenic medium from 2 weeks onwards (Fig. 4A-F). Excess T3 resulted in a significant, progressive reduction in pellet sizes as compared to controls, at consecutive weeks of culturing (P<0.05, Fig. 4G). Respective expression analyses showed a reduced expression of the ECM genes *ACAN* and *COL2A1* (Fig. 4J), albeit that the downregulation of *COL2A1* appeared not statistically significant. Assesing the chondrogenic potential by measuring the ratio between *COL2A1* and *COL1A1* expression showed a dramatic effect on the chondrogenic effect of the cell system upon addition of T3 (Fig. 4I). Moreover, the OA markers of hypertrophy and ECM breakdown genes *ADAMTS5*, *RUNX2* and *EPAS1* were consistently up-regulated from 3 weeks of treatment onwards across the donor cultures (P<0.05 Fig. 4J). The differences in mean expression of *MMP13* between controls and T3 treated cells across the donors at consecutive timepoints was not significant. Immunohistochemical studies contributed to the confirmation that excess T3 resulted in a reduced expression of COL2(Fig. 4C-D) and COL10 (Fig. 4E-F) and a pronounced reduction of GAGs (Fig. 4A-B) as reflected by the reduced Alcian blue staining at consecutive time points with an

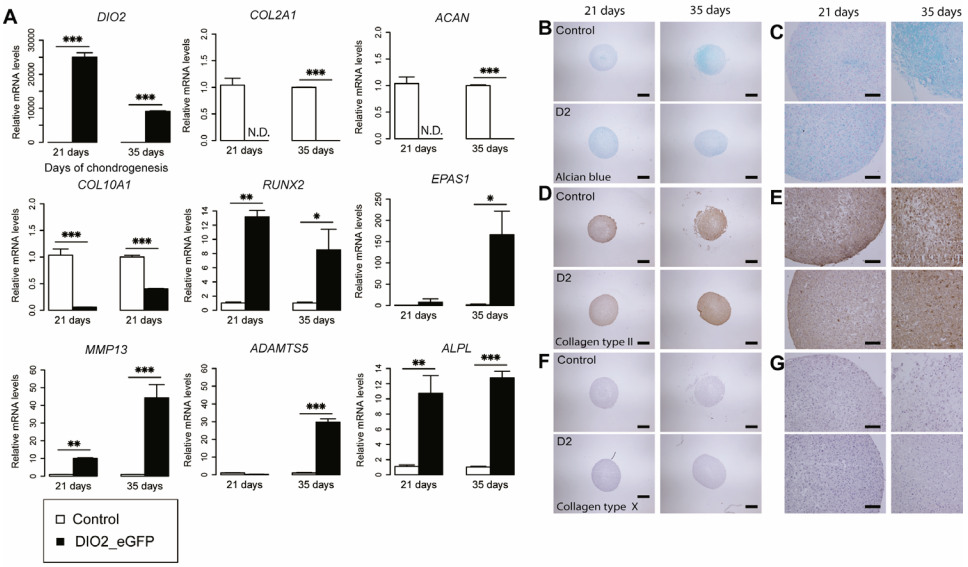


Figure 3. Overexpressing *DIO2* has a detrimental effect on cartilage extracellular matrix homeostasis. A) RT-qPCR analysis of *DIO2, COL2A1, ACAN, COL10A1, RUNX2, EPAS1, MMP13, ADAMTS5* and *ALPL* at 21 and 35 days of treatment in the chondrogenic hBMSC transduced with respectively control vector (eGFP) and DIO2 (DIO2_eGFP; Fig. S3). **(B-G)** Sections comparing control (top) and *DIO2* overexpressing (bottom) chondrogenic hBMSC pellets at 21 and 35 days of treatment. **(B-C)** Alcian Blue staining. (**D-E)** Immunohistochemical staining of collagen type II. (**F-G)** Immunohistochemical staining of collagen type X. **(B, D** and **F)** Scale bar, 400 µm. (**C, E** and **G)** Scale bar, 100 µm. The expression levels were arbitrarily defined as '1' in the pellets grown under control conditions (white), and data from the pellets over-expressing *DIO2* (black) are given as the means ± s.e.m. Legend: *p <0.05, **p <0.01, ***p <0.001. S.e.m. <0.05 are not distinguishable in the figure.

overal mean decrease of 35% (p=0.0002) as measured by the quantitative pixel-intensities (Fig. 4H).

Pharmacological inhibition of D2. To investigate whether thyroid signalling blockade attenuates the detrimental effect on cartilage matrix homeostasis, pellet cultures were treated with the pharmacological deodinase inhibitor iopanoic acid (IOP) during *in vitro* chondrogenesis of hBMSCs of the 5 donors. Quantification of surface areas indicated increasing sizes at consecutive time points in the IOP-treated pellets comparable to non-treated, control pellets (Fig. 5G). Respective expression analyses showed a significantly increased expression of the ECM genes *ACAN, COL2A1* and *COL10A1* after 3 weeks of treatment (p <0.01, Fig. 5J), whereas the OA markers of hypertrophy and ECM breakdown genes *ADAMTS5*, *RUNX2* and *EPAS1* were similar to control cultures (Fig. 5J). Furthermore, we found no significant difference in chondrogenic potential between control and IOP treated pellets, when assesing the ratio between *COL2A1* and *COL1A1* expression (Fig. 5I). Following these expression patterns, histological analysis of matrix components showed slightly higher levels of COL2 and COL10 staining (Fig. 5C-F) and higher levels of GAGs as reflected by a significant 4% increase in Alcian blue staining at week 5 of treatment (p=0.018; Fig. 5A-B and Fig. 5H). Of note is the denser cartilage matrix structure with less cellular lacunae at week 5 of treatment in cells treated with IOP as compared to controls (Fig. 5A). By discriminating "cartilage" form the "lacunae" using the ImageJ data, we could show that administration of IOP resulted to an 10.92% decrease of lacunea (p=4.27x10^{-5}) on day 35 of treatment.

DISCUSSION

In the current study, we provided insights into how genetic variation at the *DIO2* locus confers risk to OA. As a result of OA related changes in articular cartilage, loss of epigenetic silencing results in upregulation of *DIO2* expression among *DIO2* rs225014 risk-allele carriers (Fig. 1, Fig. 2). By applying an *in vitro* chondrogenesis model with genetically modified hBMSC, it was subsequently shown that genetic upregulation of *DIO2* expression resulted in a marked reduction of the capacity of chondrocytes to deposit ECM components, concurrent with induction of OA specific markers of cartilage matrix degeneration (*ADAMTS5* and *MMP13*) and mineralization (*ALPL*) (Fig. 3). Given their concurrent upregulation, this process is likely mediated via *HIF-2a/RUNX2* signalling, a hallmark of the OA disease process.[6, 24-26] Moreover, we show that the detrimental effects of *DIO2* upregulation are a result of increased T3 synthesis as reflected by the identical results when adding T3 to the culture medium (Fig. 4). Given that the effects observed in both treatments are very similar and the fact that the specific downstream effect of D2 action is the conversion of inactive T4 to active T3, we are confident that the reported effects in both experiments are reflecting the same mechanism, albeit that we did not directly asses the levels of the trace-element T3 in our in vitro chondrogenesis model. Together, our data are in line with our previous

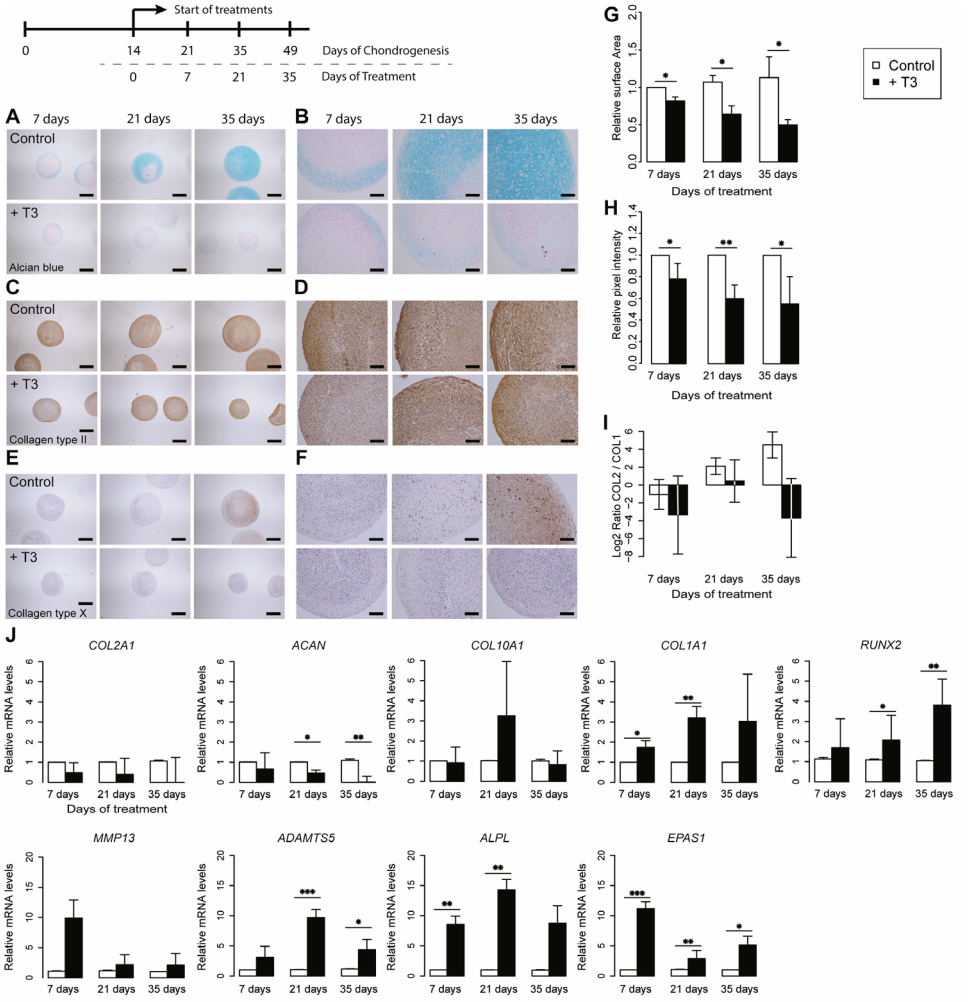


Figure 4. Thyroid hormone is believed to be a key regulator in human cartilage development. (A-F) Sections of a representative donor (See Figure S4) comparing control (top) and T3 treated (bottom) chondrogenic hBMSC pellet at 7, 21 and 35 days of treatment. **(A-B)** Alcian Blue staining. (**C-D)** Immunohistochemical staining of Collagen type II. (**E-F)** Immunohistochemical staining of collagen type X. (**A, C,** and **E)** Scale bar, 400 μm. (**B, D** and **F)** Scale bar, 100 μm. (**G)** Mean surface area measurements at 7, 21 and 35 days of treatment with N = 3 donors in each group. **(H)** Mean quantitative pixel-intensity measurements (Fig. S1) at 7, 21 and 35 days of treatment with N = 5 donors in each group. **(I)** Log2 Ratio of *COL2A1* and *COL1A1* with N = 5 donors in each group. **(J)** Mean real time RT-qPCR analysis of *COL2A1*, *ACAN*, *COL10A1*, *COL1A1, ALPL, ADAMTS5 MMP13*, *RUNX2*, and *EPAS1* at 7, 21 and 35 days of treatment in chondrogenic hBMSCs of N = 5 donors in each group. The surface area measurement, pixel intensities and level of expression were arbitrarily defined as '1' in the pellets grown under control conditions (white) and the data from the pellets grown in the addition of excess T3 (black) are given as the mean ± s.e.m. *p <0.05, **p <0.01, ***p <0.001. S.e.m. < 0.05 are not distinguishable in the figure.

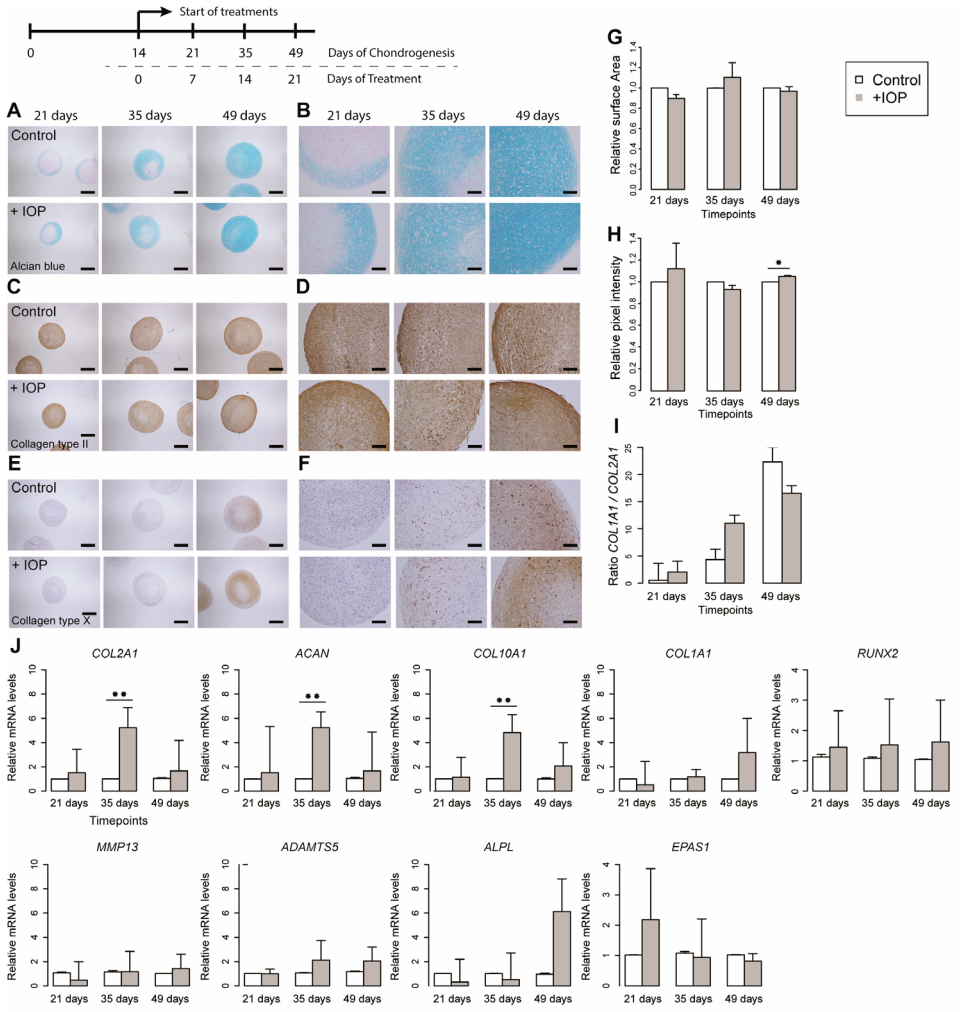


Figure 5. Counteracting thyroid signalling by inhibiting deiodinases in human cartilage development. (A-F) Sections of a representative donor (See Figure S4) comparing control (top) and IOP treated (bottom) chondrogenic hBMSC pellet at 7, 21 and 35 days of treatment. **(A-B)** Alcian Blue staining. (**C-D)** Immunohistochemical staining of Collagen type II. (**E-F)** Immunohistochemical staining of collagen type X. (**A, C,** and **E)** Scale bar, 400 µm. (**B, D** and **F)** Scale bar, 100 µm. (**G)** Mean surface area measurements at 7, 21 and 35 days of treatment with N = 3 donors in each group. **(H)** Mean quantitative pixel-intensity measurements (Fig. S1) at 7, 21 and 35 days of treatment with N = 5 donors in each group. **(I)** Log2 Ratio of *COL2A1* and *COL1A1* with N = 5 donors in each group. **(J)** Mean real time RT-qPCR analysis of *COL2A1*, *ACAN*, *COL10A1*, *COL1A1, ALPL, ADAMTS5 MMP13*, *RUNX2*, and *EPAS1* at 7, 21 and 35 days of treatment in chondrogenic hBMSCs of N = 5 donors in each group. The surface area measurement, pixel intensities and level of expression were arbitrarily defined as '1' in the pellets grown under control conditions (white) and the data from the pellets grown in the addition of IOP (grey) are given as the mean ± s.e.m. *p <0.05, **p <0.01, ***p <0.001. S.e.m. <0.05 are not distinguishable in the figure.

observations that carriers of the *DIO2* risk allele are prone to improper endochondral ossification and respective skeletal morphogenesis that could result in subtle malformations of joints or articular cartilage ECM composition.[27]

In contrast, we showed that inhibiting deiodinases by addition of IOP, contributed to prolonged 'healthy' cartilage homeostasis by virtue of attenuated upregulation of matrix degrading enzymes, a constant *COL2A1*/*COL1A1* ratio, denser cartilage matrix structure with significant less cellular lacunae which indicates a reduced propensity of chondrocytes to enter the terminal maturational process (Fig. 5). In view of these findings, we advocate that attenuation of thyroid signalling by for example inhibiting deiodinases, could contribute to novel therapeutic options of OA or could improve outcomes of cartilage tissue-engineering approaches. Nevertheless, given that T3 has many and various biological functions, both in the circulation as well as in a tissue specific manner, local administration of an thyroid-blocking agent is likely necessary and a challenging aspect. Furthermore, the effect of, for example, IOP on other joint-tissues (e.g. ligament and synovium) requires investigation. It should be noted that IOP is a general inhibitor of deodinases and as such could have aslo inhibited D1 and D3 action. In this respect, expression of *DIO1*, being important mainly in the circulation, is likely absent in cartilage tissues. Furthermore, by inhibiting D3, we prevented the conversion of active T3 to inactive T4 thereby ruling out the effect of T3 depletion due to inactive D2. Despite the beneficial effects of IOP, we observed upregulation of COL10 similar to control cultures reflecting the normal initiation of chondrocyte hypertrophy (Fig. 5J). Vice versa, we showed that upon addition of T3, chondrocytes directly enter the terminal maturational process towards bone, as reflected by the up-regulation of enzymatic breakdown (*ADAMTS5*) and mineralization (*ALPL, COL1A1*), this without significant induction of COL10 deposition (Fig. 4E-F). Together, these data indicate that chondrocyte hypertrophy in our model was not neccesarily detrimental to cartilage homeostasis this in contrast to *DIO2* upregulation. Our data, therefore, indicates that upregulation of *DIO2* does not affect the 'early' hypertrophic expression of *COL10A1* but induces the later "progression stage"-marker *MMP13*, and the "late stage"-markers *ALPL* and *COL1A1*.[28] In our model, we used the expression of *EPAS1* and *RUNX2* to shown downstream effects of up-regulation of *DIO2* and observed a significant positive association between T3 and *EPAS1*. Although such an observation based on association does not imply a direct causal relationship, recent data of Chatonnet et al., (2013)[29] showed, in a ChIP-seq analysis in mouse C17.2 neural progenitor cells, that *EPAS1*, harbours specific thyroid hormone nuclear receptor (THR) binding sites and is directly reactive to thyroid hormone. In view of these data, we advocate that active thyroid hormone, likely by local *DIO2* action, could have an important impact on *EPAS1* up-regulation during the pathophysiology of OA. Additional studies are, however, necessary to elucidate whether T3 is directly affecting expression of *EPAS1*, or *RUNX2* for that matter, by binding to a positional thyroid receptor in humans.

Despite the fact that the direction of gene expression changes of *COL2A1* and *MMP13* upon addition of T3 to the culture media

appeared consistent with our overall results (Fig. 4J), the mean differences in expression were not statistically significant across the donors. Most likey this was the result of the considerable heterogeneity in the differential gene expression patterns of these specific genes across donors, especially with respect to timing the respective down and up-regulation at consecutive time-points (Fig. S4), a phenomenon generally recognized in the *in vitro* chondrogenesis models of primary hBMSCs[30].

We detected several CpG dinucleotides across the *DIO2* locus that were differentially methylated between preserved and OA affected cartilage. Although we have only verified their functionality with respect to *DIO2* expression, we cannot exclude that methylation at these sites regulates expression of more distal genes. We found a consistent positive correlation between methylation at CpG -2031 and *DIO2* expression in articular cartilage among carriers of the rs225014 risk allele, which may not comply to the conventional inverse relation between CpG methylation and gene expression. However, in recent genome wide approaches, it has been recognized that this conventional relation primarily holds among CpG dinucleotides residing in CpG islands and proximal promotors, whereas gene body and distal enhancer methylation, as is the case for CpG -2031, has been shown to correlate in either direction with gene expression.[31-33]

With respect to our observations of *DIO2* expression in preserved and osteoarthritic cartilage, it should be noted that, in contrast to previously reported high upregulation of *DIO2* expression in osteoarthritic when compared to healthy cartilage[5, 11] we showed a moderate upregulation in osteoarthritic when compared to preserved cartilage of the same joint only among carriers of the rs225014 risk allele (Fig. 1I). This difference suggests that upregulation of *DIO2* expression may be an early event in OA pathophysiology and might be set to continue progressively among rs225014 risk allele carriers.

We were unable to validate the functionality of a putative CTCF binding site directly at the rs225014 locus nor for that matter the causality of the previously assessed consistent allelic imbalance.[11] Possibly, the rs225014 SNP affects three dimensional chromatin conformations underlying the relation between the rs225014 tagged allelic imbalance and methylation dependent upregulation of *DIO2* among rs225014 risk allele carriers.

In conclusion, our data provides evidence in humans that genetic predisposition combined with early OA related changes results in loss of epigenetic silencing of *DIO2*, which likely induces *EPAS1* and *RUNX2* mediated up-regulation of cartilage matrix degrading enzymes (*ADAMTS5* and *MMP13*) and mineralization of matrix (*ALPL* and *COL1A1)*, thereby driving the OA process most distinctly among *DIO2* risk allele carriers. Furthermore, our data shows that counteracting the thyroid signalling by inhibiting deiodinases could contribute to needed novel therapeutic approaches of OA.

ACKNOWLEDGEMENTS

We thank all participants of the PAPRIKA and RAAK study. The PAPRIKA and RAAK studies were supported by the Leiden University Medical Centre, the Dutch Arthritis Association (DAA 101-402 and

Reumafonds LRR) and the Centre of Medical System Biology and Netherlands Consortium for Healthy Aging both in the framework of the Netherlands Genomics Initiative (NGI). Furthermore, we acknowledge support by TreatOA and IDEAL which are funded by the European Union's Seventh Framework Program (FP7/2007-2011) under, respective, grant agreement no. 200800 and 259679. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interest: The authors declare no competing financial, personal, or professional interests.

Author contributions:

Conceived and designed the experiments: NB; WdH; YFMR; FJV; PES; IM

Performed the experiments: NB; WdH; YFMR; RvdB; NL; AEvE; AD

Analyzed the data: NB; WdH; YFMR; AEvE; AD

Contributed reagents/materials/analysis tools: SDB; BAP; BJD; RGHHN

Wrote the manuscript: NB; WdH; YFMR; IM

Critically reviewed the manuscript: All authors

REFERENCES

1 Woolf A, Pfleger B. Burden of major musculoskeletal conditions. *Bulletin of the World Health Organization* 2003;**81**:646-56.

2 Reynard LN, Loughlin J. The genetics and functional analysis of primary osteoarthritis susceptibility. *Expert Rev Mol Med* 2013;**15**:e2.

3 Meulenbelt I, Min JL, Bos S, Riyazi N, Houwing-Duistermaat JJ, van der Wijk HJ, *et al.* Identification of DIO2 as a new susceptibility locus for symptomatic osteoarthritis. *Hum Mol Genet* 2008;**17**:1867-75.

4 Meulenbelt I, Bos SD, Chapman K, van der BR, Houwing-Duistermaat JJ, Kremer D, *et al.* Meta-analyses of genes modulating intracellular T3 bio-availability reveal a possible role for the DIO3 gene in osteoarthritis susceptibility. *Ann Rheum Dis* 2011;**70**:164-7.

5 Ijiri K, Zerbini LF, Peng H, Otu HH, Tsuchimochi K, Otero M, *et al.* Differential expression of GADD45beta in normal and osteoarthritic cartilage: potential role in homeostasis of articular chondrocytes. *Arthritis Rheum* 2008;**58**:2075-87.

6 Husa M, Liu-Bryan R, Terkeltaub R. Shifting HIFs in osteoarthritis. *Nat Med* 2010;**16**:641-4.

7 Goldring MB, Goldring SR. Articular cartilage and subchondral bone in the pathogenesis of osteoarthritis. *Ann N Y Acad Sci* 2010;**1192**:230-7.

8 Reynard LN, Bui C, Canty-Laird EG, Young DA, Loughlin J. Expression of the osteoarthritis-associated gene GDF5 is modulated epigenetically by DNA methylation. *Hum Mol Genet* 2011;**20**:3450-60.

9 Young DA, Bui C, Barter MJ. Understanding CpG methylation in the context of osteoarthritis. *Epigenomics* 2012;**4**:593-5.

10 Young DA. More evidence for a role of CpG methylation in the pathogenesis of osteoarthritis. *Arthritis Rheum* 2012.

11 Bos SD, Bovee JV, Duijnisveld BJ, Raine EV, van Dalen WJ, Ramos YF, *et al.* Increased type II deiodinase protein in OA-affected cartilage and allelic imbalance of OA risk polymorphism rs225014 at DIO2 in human

OA joint tissues. *Ann Rheum Dis* 2012;**71**:1254-8.

12 Karlsson C, Dehne T, Lindahl A, Brittberg M, Pruss A, Sittinger M, *et al.* Genome-wide expression profiling reveals new candidate genes associated with osteoarthritis. *Osteoarthritis Cartilage* 2010;**18**:581-92.

13 Nagase H, Nagasawa Y, Tachida Y, Sakakibara S, Okutsu J, Suematsu N, *et al.* Deiodinase 2 upregulation demonstrated in osteoarthritis patients cartilage causes cartilage destruction in tissue-specific transgenic rats. *Osteoarthritis Cartilage* 2013;**21**:514-23.

14 Valdes AM, Spector TD. The clinical relevance of genetic susceptibility to osteoarthritis. *Best Pract Res Clin Rheumatol* 2010;**24**:3-14.

15 Rosenbloom KR, Dreszer TR, Long JC, Malladi VS, Sloan CA, Raney BJ, *et al.* ENCODE whole-genome data in the UCSC Genome Browser: update 2012. *Nucleic Acids Res* 2012;**40**:D912-D917.

16 Holwerda S, de LW. Chromatin loops, gene positioning, and gene expression. *Front Genet* 2012;**3**:217.

17 Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G, *et al.* Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci U S A* 2005;**102**:15785-90.

18 Talens RP, Boomsma DI, Tobi EW, Kremer D, Jukema JW, Willemsen G, *et al.* Variation, patterns, and temporal stability of DNA methylation: considerations for epigenetic epidemiology. *FASEB J* 2010.

19 Tobi EW, Lumey LH, Talens RP, Kremer D, Putter H, Stein AD, *et al.* DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum Mol Genet* 2009;**18**:4046-53.

20 Gereben B, Goncalves C, Harney JW, Larsen PR, Bianco AC. Selective proteolysis of human type 2 deiodinase: a novel ubiquitin-proteasomal mediated mechanism for regulation of hormone activation. *Mol Endocrinol* 2000;**14**:1697-708.

21 Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 2012;**9**:671-5.

22 Schwabe T, Neuert H, Clandinin T. A Network of Cadherin-Mediated Interactions Polarizes Growth Cones to Determine Targeting Specificity. *Cell* 2013;**154**:351-64.

23 Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Meth* 2012;**9**:671-5.

24 Bos SD, Slagboom PE, Meulenbelt I. New insights into osteoarthritis: early developmental features of an ageing-related disease. *Curr Opin Rheumatol* 2008;**20**:553-9.

25 Saito T, Fukai A, Mabuchi A, Ikeda T, Yano F, Ohba S, *et al.* Transcriptional regulation of endochondral ossification by HIF-2[alpha] during skeletal growth and osteoarthritis development. *Nat Med* 2010;**16**:678-86.

26 Yang S, Kim J, Ryu JH, Oh H, Chun CH, Kim BJ, *et al.* Hypoxia-inducible factor-2[alpha] is a catabolic regulator of osteoarthritic cartilage destruction. *Nat Med* 2010;**16**:687-93.

27 Waarsing JH, Kloppenburg M, Slagboom PE, Kroon HM, Houwing-Duistermaat JJ, Weinans H, *et al.* Osteoarthritis susceptibility genes influence the association between hip morphology and osteoarthritis. *Arthritis Rheum* 2011;**63**:1349-54.

28 van der Kraan PM, van den Berg WB. Chondrocyte hypertrophy and osteoarthritis: role in initiation and progression of cartilage degeneration? *Osteoarthritis Cartilage* 2012;**20**:223-32.

29 Chatonnet F, Guyot R, Benoît G, Flamant F. Genome-wide analysis of thyroid hormone receptors shared and specific functions in neural cells. *Proceedings of the National Academy of Sciences* 2013;**110**:E766-E775.

30 Hellingman CA, Davidson EN, Koevoet W, Vitters EL, van den Berg WB, van Osch GJ, *et al.* Smad signaling determines chondrogenic differentiation of bone-marrow-derived mesenchymal stem cells: inhibition of Smad1/5/8P prevents terminal differentiation and calcification. *Tissue Eng Part A* 2011;**17**:1157-67.

31 Bell J, Pai A, Pickrell J, Gaffney D, Pique-Regi R, Degner J, *et al.* DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biology* 2011;**12**:R10.

32 Zhang D, Cheng L, Badner JA, Chen C, Chen Q, Luo W, *et al.* Genetic Control of Individual Differences in Gene-Specific Methylation in Human Brain. *The American Journal of Human Genetics* 2010;**86**:411-9.

33 Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet* 2010;**6**:e1000952.

Supplementary methods

RAAK study

The ongoing RAAK study is approved by the ethical committee of the LUMC (P08.239) and is aimed at the bio-banking of joint materials (cartilage, bone and where available ligaments) and mesenchymal stem cells (hip joints only) and primary chondrocytes of patients and controls in the Leiden University Medical Center and collaborating outpatient clinics in the Leiden area. Informed consent was obtained from each patient. In the current study we used paired preserved and OA affected cartilage samples from 52 Caucasian end stage OA patients undergoing joint replacement surgery for primary OA (23 hips, 29 knees) in the Leiden University Medical Centre (Table S1). At the moment of collection (within 2 hours following surgery) tissue was washed extensively with phosphate buffered saline (PBS) to decrease the risk of contamination by blood, and cartilage was collected of the weight-baring area of the joint. Initial classification of cartilage was performed macroscopically and collected separately as OA affected or preserved regions. Classification was done according to predefined features for OA related damage based on color/whiteness of the cartilage, based on surface integrity as determined by visible fibrillation/crack formation, and based on depth and hardness of the cartilage upon sampling with a scalpel. During collection with a scalpel, care was taken to avoid contamination with bone or synovium. Collected cartilage was snap frozen in liquid nitrogen and stored at -80°C prior to RNA extraction. We histologically assessed cartilage samples with the modified Mankin scoring system.

Nucleic acid isolation and genotyping

Snap frozen cartilage was powderized using a Retsch Mixer Mill 200 with continuous liquid nitrogen cooling. DNA was isolated using the Promega Wizard Genomic DNA Purification kit according to the manufacturer's protocol. RNA was isolated using the Qiagen RNAeasy Mini kits, followed by cDNA synthesis using 1 μg of RNA and random hexamer primers (First Strand cDNA Synthesis Kit, Thermo Scientific). Samples were genotyped for rs225014 using restriction fragment length polymorphism analysis with RsaI (Forward primer (F): 5'-AGTGGCAATGTGTTTAATGTGA-3', Reverse primer (R): 5'-CACACACGTTCAAAGGCTACC-3').) DNA fragment length of wild type alleles were called after gel electrophoresis and were 121, 30 and 389 base pairs. The risk allele C affects the first cut site resulting in two fragments consisting of 151 and 389 base pairs. All primers were ordered at Invitrogen.

Electrophoretic mobility shift assay

For Electrophoretic Mobility Shift Assays (EMSAs) synthetic oligonucleotides containing the putative CTCF binding site were 5'-end labeled by γ-^{32}P-ATP and subsequently purified by gel filtration on Sephadex G-25 Medium columns. The CTCF-11 zinc finger (11ZF) DNA binding domain, full-length CTCF using pIVEX1.4 WG CTCF-11ZF and CTCF-FL constructs were synthesized with the RTS 100 Wheat Germ CECF kit (5 PRIME). For binding reactions, we used buffer containing standard PBS with 5 mM $MgCl_2$, 0.1 mM $ZnSO_4$, 1 mM DTT, 0.05%

NP40, 50 ng/µl poly(dI-dC) and 10% glycerol. The reaction mixtures were incubated for 30 min at RT and analyzed by 5% native PAGE in 0.5x Tris-borate-EDTA buffer. Table S5 contains the sequences of the used probes. Expression of *DIO2* was assessed using TaqMan probe Hs00988260_m1 (Applied Biosystems), normalized for *GAPDH* expression (real-time PCR, F: 5'-TGCCATGTAGACCCCTTGAAG-3', R: 5'-ATGGTACATGACAAGGTGCG G-3') and subsequently log transformed for downstream analysis.

Quantification of methylation

Using the ZymoResearch EZ DNA Methylation kit isolated genomic DNA was treated with sodium bisulphite (BS), thereby reducing unmethylated cytosine residues to uracil, while methylated cytosines remain unchanged. The methylated fraction of CpG dinucleotides was assessed with MALDI-TOF mass spectrometry (Epityper, Sequenom), a commonly applied to quantify CpG methylation. Samples were randomly distributed on PCR plates prior to BS treatment and PCR amplification. PCR amplification and MALDI-TOF measurements were performed in triplicate as technical replicates. Using MethPrimer 9 amplicons (Table S6) were designed upstream, downstream and intragenic of *DIO2*. Amplicons were designed to cover conserved transcription factor binding sites (TFBS) according to Human Genome Assembly, build 19. To avoid interference of SNPs in the MALDI-TOF measurements we avoided SNPs in the amplified regions using dbSNP137. Methylation of several CpG dinucleotides was measured redundantly by separate amplicons, measurements of CpG dinucleotides with the most successful observations per amplicon were used for downstream analyses. Finally, the 9 amplicons constituted 4 independent regions covering 23 unique CpG dinucleotides.

AZA treated cell cultures

Primary articular chondrocytes were isolated from cartilage derived from three OA patients who underwent total joint arthroplasty of the hip (RAAK study). Cartilage tissue was incubated overnight in DMEM (high glucose; Gibco, Bleiswijk, The Netherlands) supplemented with 10% fetal bovine serum (FBS; Gibco), antibiotics (100 U/ml penicillin, 100 µg/ml streptomycin; Gibco) and 2 mg/ml collagenase Type I at 37 ºC in a humidified 5% CO_2 / 95% atmosphere. Subsequently, primary chondrocytes were resuspended and filtered through a 100 µm mesh to remove undigested cartilage fragments and extracellular matrix debris. Cells were expanded at 37 °C in a humidified 5% CO_2 / 95% atmosphere in DMEM supplemented with 10% FBS, penicillin (100 units/mL), streptomycin (100 units/mL) and 0.5 ng/ml FGF-2 (PeproTech, Heerhugowaard, The Netherlands) for 2 passages. 24 Hours into the second passage 1.5 µM of the demethylating agent 5-aza-2 –deoxycytidine (AZA) (Sigma Aldrich; Zwijndrecht, The Netherlands) was added. Cells were harvested for DNA and RNA isolations after being grown to confluence, obtained after three more days.

Analysis of methylation data

Methylation of CpG dinucleotides with fewer than two out of three triplicate measurements or with a SD > 0.1 were discarded prior to analysis. CpG site-containing fragments that had equal or overlapping mass, making them irresolvable by mass spectrometry, and CpG sites

containing fragments whose measurement was confounded by SNPs were removed prior to analysis[1]. Samples with bisulphite conversion rates < 98% were discarded.

All statistical analyses were performed by fitting Generalized Linear Mixed Models (GLMMs). To account for inter-individual differences a random effect for sample donor was added to each model. Homozygous carriers (N = 3) of the rs225014 "C" risk allele were pooled with heterozygous carriers (N = 27). Analyses were carried out using the R programming language with the *lme4* (GLMMs) package.[2] To assess the relation between the cartilage phenotype and methylation of separate features, we fitted the following model: $Methylation_i \sim (1|Donor) + Phenotype$. To identify functional CpG dinucleotides, we fitted the following model: $DIO2\ Expression \sim (1|)\ Donor + Methylation_i$. Where, in both models, $Methylation_i$ represents the methylated fraction of the *i*-th CpG feature. To explore other possible significant covariates, we fitted the following model: $DIO2\ Expression \sim (1|Donor) + Joint\ site + rs225014\ alleles + Methylation_{CpG\text{-}2031}$. Finally, p-values were adjusted for multiple testing using Bonferroni correction. Analyses of differences in methylation and expression in chondrocyte cultures after AZA treatment were done by paired-student T-tests.

Cell culture and *in vitro* chondrogenic differentiation

HEK 293T cells were grown in DMEM (Gibco; high glucose) supplemented with 10% fetal calf serum (FBS; Gibco, Bleiswijk, The Netherlands) and antibiotics (100 U/ml penicillin, 100 µg/ml streptomycin; Gibco) at 37ºC in a humidified 5% CO_2 / 95% atmosphere.

Human bone marrow derived mesenchymal stem cells (hBMSCs) were isolated from hip joints of 5 OA patients who underwent total hip arthroplasty as result of end stage OA as part of the RAAK study at the Leiden University Medical Centre (LUMC). Cells were expanded at 37ºC in a humidified 10% CO_2 / 90% atmosphere in DMEM (high glucose) supplemented with 10% FBS, antibiotics, and 0.5 ng/ml FGF-2 (PeproTech, Heerhugowaard, The Netherlands) for 5 passages. Subsequently, 3D pellets were formed using centrifugal forces (1200 rpm; 5 min) on 2.5 x 10^5 cells in 15 ml polypropylene conical tubes. Chondrogenesis was initiated in 1 ml serum-free chondrogenic differentiation medium (DMEM, supplemented with Ascorbic acid (50 µg/ml; Sigma-Aldrich; Zwijndrecht, The Netherlands), L-Proline (40 µg/ml; Sigma-Aldrich), Sodium Puryvate (100 µg/ml; Sigma-Aldrich), Dexamethasone (0.1 µM; Sigma-Aldrich), ITS+, antibiotics, and TGF-β1 (10 ng/ml; PeproTech)). Medium was changed every 3-4 days. From day 14 onwards cell pellets were maintained either in the standard chondrogenic differentiation medium or in the presence of T_3 (10 nM) or IOP (10 µM). Pellets were collected for immunohistochemistry and RNA isolation at the time points indicated for up to 49 days total. Literature search on the use of T3 in mammalian cell-models showed that in a wide variety of cell-lines and phenotypes, a wide range of concentrations is used (10 nM[3], 100 nM[4] and 1uM[5]). To induce OA-symptoms in chondrogenesis models we tested two concentrations, namely 10 and 100 nM. Using 10 nM T3 was shown sufficient in our model system to induce an OA phenotype comparable to over expressing DIO2 (Data not shown).

Lentiviral constructs and transduction

C-terminal FLAG-tagged cys-D2 (kindly provided by Prof. Dr. Bianco) was digested with

EcoR1 followed by Klenow-treatment and digestion with *Xba1*. Inserts were inserted into the *EcoRV-XbaI* sites of the pLV-CMV-IRES-eGFP Lentiviral backbone (kindly provided by Prof. Dr. R. Hoeben; Dept of Molecular Cell Biology; LUMC). Lentiviral production was performed in HEK 293T cells as described previously.[6] In short, for lentiviral transduction, hBMSCs (6 x 10^5 cells) were seeded onto 20 cm plates. After one day in culture (approximately 80% confluence), cells were incubated for 16 hr with the corresponding Lentiviruses at a multiplicity of infection (MOI) of 1 in the presence of 15 μg/ml Polybrene (Sigma-Aldrich) before the medium was replaced to regular growth medium.

Histochemistry and immunohistochemistry

hBMSC pellets were fixed in 4% formaldehyde upon harvesting and embedded in paraffin. Sections were stained for GAGs with Alcian blue and counterstained with nuclear fast red (Sigma-Aldrich).

For immunohistochemistry, paraffin sections were deparaffinized and rehydrated. Endogenous peroxidase was blocked with MeOH / 0.3% H_2O_2, and antigen retrieval was performed by incubation with proteinase K (5μl/ml; Sigma-Aldrich) for 10 minutes at 37°C and subsequently with Hyaluronidase (5mg/ml; Sigma-Aldrich) for 30 minutes at 37°C. Sections were blocked for 30 minutes at room temperature in TBST (2.5mM Tris-HCl pH 7.6, 7.5mM NaCl, 0.1% Tween-20) with 5% normal goat serum (R&D Systems, Abingdon, Great Britain). Incubation with the primary antibodies was performed overnight at 4°C with commercially available antibodies for collagen type 2 (COL2; MAB1330 Millipore; Amsterdam, The Netherlands; 1:100 in TBST / 10% NGS) and collagen type 10 (COL10; Clone X53, Quartett; Berlin, Germany; 1:100 in TBST/10% NGS). The next day, Powervision (Immunologic; Duiven, The Netherlands) was applied and visualization was carried out in a 3-diaminobenzidine (DAB)-solution (Sigma-Aldrich) for 1 and 10 minutes for COL2 and COL10 respectively. The sections were then counterstained with haematoxylin.

Relative pixel intensity

The relative pixel intensity was computed by loading the photos into ImageJ (v.1.47).[7],[8] Automated thresholding (IsoData) was used to separate the glycosaminoglycans staining and the background (vacuoles and surroundings), after using the concept of rolling ball algorithm to correct for uneven illumination (Fig. S5, panel A-C). The regions depicted as glycosaminoglycans by the automated thresholding were semi-automatically filled by using the 4-connected background elements (Fig. S5, panel D). The glycosaminoglycans-staining was separated from the background (Fig. S5, panel E) by splitting the image into separate channels and only using the red channel. Every pixel of the image was then weighed for color intensity (0 – 255) and the sum of al intensity was averaged by the number of pixels, giving the average pixel intensity per condition.

Surface area meusurements

Surface area measurements were performed using the standard CellSens (Build 9164; version1.5) software from Olympus. Within this microscope software, we used the option to

manually draw a border and let the software calculate the surface area within this border. Measurements are derived from 3 individual pellets per condition, per time point, per donor.

RNA isolation

RNA isolation from the pellets was performed by pooling two pellets for every given condition. The isolation of RNA was performed as described previously.[9] RNA quantity was assessed using a nanodrop spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, USA).

Real time quantitative reverse transcription PCR

RNA was processed with the First Strand cDNA Synthesis Kit according to the manufacturer's protocol (Roche Applied Science, Almere, The Netherlands). cDNA pre-amplification was performed with a DNA Engine Tetrad® 2 Peltier Thermal Cycler (Bio-Rad) for 5 cycles under standard PCR conditions and using the multiplexed primers (Table S8). Primer efficiencies were verified in advance, by performing a concentration curve experiment. Only primers showing a perfect correlation between Ct-value and concentration and with a single peak in the melting-curve calculation (Lightcycler480 SW 1.5 software) were used. The housekeeping gene GAPDH was used as a single reference gene for qPCR.[10, 11] A custom sample-mix was made using the EvaGreen® dye (Biotium, Amsterdam, The Netherlands), ROX reference dye (Life technologies), and FastStart Taq polymerase (Roche Applied Science). Subsequently, qPCR was performed using the custom-designed primers (Table S8) in triplicate with the using Biomark™ 96.96 Dynamic Arrays (Fluidigm). Additional RT-qPCR measurements were performed on the Roche Lightcycler 480 II, using FastStart SYBR Green Master reaction mix according to the manufacturer's protocol (Roche Applied Science).

Gene expression and pixel intensity analyses

Relative gene expressions were corrected for GAPDH by using the $2^{-\Delta\Delta Ct}$ method, in which Ct indicates cycle threshold, the fractional cycle number where the fluorescent signal reaches detection threshold. All values were calculated relative to the first timepoint measured in the assay. For the hBMSC pellet-model, time point t = 7 days of treatment of the control was taken as reference point for expression analysis.

The paired-student T-test was used to calculate the significance of changes in expression. The same method was used to calculate significance of the changes in size of the pellets over time. To calculate the significance of differences in pixel intensities relative to the control conditions, the Wilcoxon–Mann–Whitney test was performed. All nominal P-values < 0.05 were considered statistically significant.

REFERENCES

1 Heijmans BT, Kremer D, Tobi EW, Boomsma DI, Slagboom PE. Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/

H19 locus. *Hum Mol Genet* 2007;**16**:547-54.

2 Douglas Bates, Martin Maechler, Ben Bolker. lme4: Linear mixed-effects models using S4 classes. [R package version 0.999999-2]. 2013.

3 Williams AJ, Robson H, Kester MHA, van Leeuwen JPTM, Shalet SM, Visser TJ, *et al.* Iodothyronine deiodinase enzyme activities in bone. *Bone* 2008;**43**:126-34.

4 Wang L, Shao YY, Ballock RT. Thyroid Hormone Interacts With the Wnt/+¦-Catenin Signaling Pathway in the Terminal Differentiation of Growth Plate Chondrocytes. *J Bone Miner Res* 2007;**22**:1988-95.

5 Smith L, Belanger J, Oberbauer A. Fibroblast growth factor receptor 3 effects on proliferation and telomerase activity in sheep growth plate chondrocytes. *Journal of Animal Science and Biotechnology* 2012;**3**:39.

6 Hamdi M, Kool J, Cornelissen-Steijger P, Carlotti F, Popeijus HE, van der Burgt C, *et al.* DNA damage in transcribed genes induces apoptosis via the JNK pathway and the JNK-phosphatase MKP-1. *Oncogene* 2005;**24**:7135-44.

7 Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Meth* 2012;**9**:671-5.

8 Schwabe T, Neuert H, Clandinin T. A Network of Cadherin-Mediated Interactions Polarizes Growth Cones to Determine Targeting Specificity. *Cell* 2013;**154**:351-64.

9 Hellingman CA, Koevoet W, Kops N, Farrell E, Jahr H, Liu W, *et al.* Fibroblast growth factor receptors in in vitro and in vivo chondrogenesis: relating tissue engineering using adult mesenchymal stem cells to embryonic development. *Tissue Eng Part A* 2010;**16**:545-56.

10 Toegel S, Huang W, Piana C, Unger F, Wirth M, Goldring M, *et al.* Selection of reliable reference genes for qPCR studies on chondroprotective action. *BMC Molecular Biology* 2007;**8**:13.

11 Fujii Y, Kitaura K, Matsutani T, Shirai K, Suzuki S, Takasaki T, *et al.* Immune-Related Gene Expression Profile in Laboratory Common Marmosets Assessed by an Accurate Quantitative Real-Time PCR Using Selected Reference Genes. *PLoS ONE* 2013;**8**:e56296.

Table S1. Characteristics of samples used in methylation and expression assessment. (a) Genotype for the rs225014 SNP

Donor	Gender	Age	Joint	rs225014[a]					
1	Female	78	Hip	TC	27	NA	69	Knee	TT
2	Female	79	Hip	TT	28	Male	64	Hip	TC
3	Male	67	Hip	TT	29	Female	58	Knee	TC
4	Male	69	Knee	TC	30	Female	57	Knee	TC
5	Female	77	Hip	TC	31	Female	62	Hip	TC
6	Female	74	Hip	TT	32	Female	67	Knee	TC
7	Male	72	Hip	TC	33	Male	68	Hip	TC
8	Female	75	Hip	TT	34	Male	67	Hip	TC
9	Male	54	Knee	TT	35	Female	53	Hip	TT
10	Male	64	Knee	TC	36	Male	82	Knee	TT
11	Female	65	Knee	TT	37	Female	70	Knee	TC
12	Female	68	Hip	TT	38	NA	NA	Knee	TT
13	Male	61	Hip	TC	39	Female	60	Knee	TT
14	Female	75	Hip	TT	40	NA	NA	Knee	TC
15	Male	61	Knee	TT	41	Male	69	Knee	TT
16	Female	80	Hip	TT	42	Male	77	Hip	TC
17	Female	78	Knee	TT	43	Female	66	Knee	TT
18	Male	44	Hip	TC	44	NA	NA	Knee	TC
19	Female	79	Hip	TC	45	Female	80	Knee	CC
20	NA	NA	Knee	CC	46	Female	80	Knee	TT
21	NA	NA	Hip	TC	47	NA	NA	Knee	CC
22	NA	NA	Hip	TC	48	NA	NA	Knee	TC
23	Male	63	Hip	TC	49	Male	69	Knee	TT
24	Female	69	Hip	TC	50	Male	NA	Knee	TC
25	Female	46	Knee	TT	51	Female	62	Knee	TC
26	Female	68	Knee	TT	52	Male	NA	Knee	TC

Table S2. Overview of quantified CpG dinucleotides across the DIO2 locus, as depicted in Figure 2. (a) Location of the respective CpG dinucleotides in relation to the DIO2 TSS. **(b)** Difference in methylation between preserved and matched OA affected samples. **(c)** Crude P-value of differential methylation. **(d)** Bonferroni adjusted P-value of differential methylation. **(e)** Number of samples used in the analyses for quantification of methylation **(f)** β-estimate of the methylation variable in the regression analysis with DIO2 expression as outcome. **(g)** Crude P-value of the methylation variable in the regression analysis. **(h)** Bonferroni adjusted P-value of the methylation variable in the regression analysis. **(i)** Number of samples used in the regression analysis. P: preserved

			Differential methylation P vs OA cartilage				Correlation *DIO2* methylation & express			
Location[a]	**Mean P**	**Mean OA**	**Beta**[b]	***p***[c]	p_{adj}[d]	**N**[e]	**Beta**[f]	***p***[g]	p_{adj}[h]	**N**[i]
CpG +15180	0.377	0.391	0.015	0.12293	1	102	-0.581	0.57794	1	85
CpG +15148	0.897	0.902	0.005	0.42454	1	103	-1.960	0.23275	1	86
CpG +15015	0.846	0.842	-0.004	0.39223	1	103	-3.284	0.10764	1	86
CpG +8802	0.678	0.681	0.001	0.92605	1	76	-2.745	0.04668	1	64
CpG +8771	0.852	0.846	-0.007	0.03448	0.793	98	-0.303	0.90671	1	82
CpG +8758	0.906	0.903	-0.003	0.36266	1	101	-1.245	0.69968	1	84
CpG +8742	0.944	0.942	-0.002	0.42346	1	99	-0.126	0.97346	1	82
CpG +8731	0.972	0.974	0.001	0.58860	1	103	-8.680	0.12942	1	86
CpG +8635	0.716	0.724	0.009	0.12923	1	102	0.422	0.80058	1	85
CpG +8547	0.872	0.866	-0.005	0.13632	1	97	2.832	0.35395	1	80
CpG +8527	0.758	0.779	0.017	0.01689	0.389	100	-0.904	0.49324	1	83
CpG +8481	0.167	0.182	0.015	0.00913	0.210	103	-0.201	0.91184	1	86
CpG -219	0.011	0.012	0.000	0.92106	1	94	7.420	0.19666	1	79
CpG -322	0.024	0.026	0.002	0.60637	1	98	-2.373	0.59233	1	83
CpG -473	0.015	0.016	0.001	0.15828	1	102	-7.684	0.64305	1	85
CpG -571	0.031	0.032	0.002	0.28463	1	97	2.355	0.77655	1	80
CpG -642	0.013	0.015	0.003	0.08636	1	98	5.473	0.41953	1	81
CpG -1754	0.874	0.908	0.032	0.00449	0.103	63	1.089	0.42829	1	54
CpG -1761	0.947	0.935	-0.012	0.00011	0.002 **	98	-2.094	0.48780	1	82
CpG -1838	0.763	0.732	-0.031	0.00129	0.030 *	102	-2.468	0.01598	0.368	86
CpG -1942	0.940	0.934	-0.006	0.02555	0.588	103	-1.530	0.67832	1	86
CpG -2009	0.950	0.942	-0.008	0.01004	0.231	103	-2.142	0.54294	1	86
CpG -2031	0.153	0.182	0.028	0.00005	0.001 **	103	4.959	0.00010	0.002 **	87

Table S3 Multivariate analysis the individual effects. Effects of CpG -2031 methylation, joint site (hip or knee) and rs225014 alleles on DIO2 expression in articular cartilage

	Beta	Standard Error	Statistic	*p*
Intercept	-0.905	0.808	-1.120	0.262
CpG -2031	4.008	1.708	2.347	0.019
Joint	-0.247	0.168	-1.471	0.141
rs225014	0.557	0.154	3.629	0.0003
Gender	0.202	0.146	1.385	0.166
BMI	-0.017	0.013	-1.260	0.208
Age	-0.007	0.007	-0.987	0.323

Table S4 Characteristics of hBMSC donors of the RAAK study. (a) Genotype for the rs225014 SNP.

Donor	Sex	Age	rs225014[a]
53	female	66	TC
54	female	72	TT
55	female	81	CC
56	female	59	TC
57	male	80	TC

Supplementary Tables S5-S8 with primer sequences can be downloaded from http://ard.bmj.com/content/early/2014/04/02/annrheumdis-2013-204739/suppl/DC1

Figure S1. TF binding evidence at the rs225014 locus. UCSC browser hg19 screenshot. Darker bars indicate stronger evidence for binding. CTCF shows a black bar, while the other TFs show less strong evidence indicated as grey bars (enlarged figure can be viewed on **http://ard.bmj.com/content/early/2014/04/02/annrheumdis-2013-204739/suppl/DC1**).

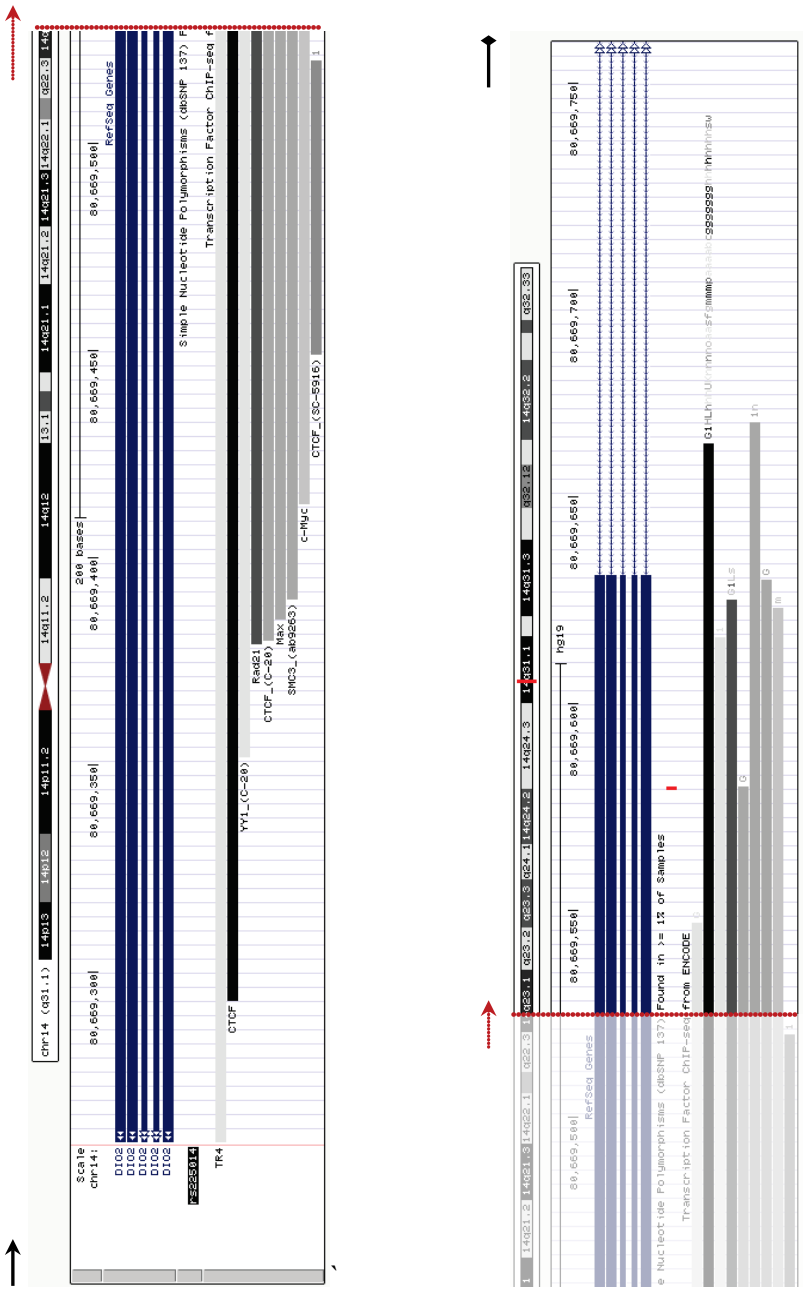


Figure S2. The rs225014 risk allele does not influence local CTCF binding. Electromobility shift assays (EMSAs) using full-length CTCF (CTCF-FL) and a truncated protein, containing just the 11 Zinc Finger binding domain (CTCF-11ZF). We have used three different probes containing the rs225014 common allele (T, Fig. 1, lane 1-2), the minor allele (C, Fig. 1, lane 3-4) and a non-existent allele (G, Fig. 1, lane 5-6). The latter is in highest agreement with the consensus CTCF-binding sequence. As a positive control, a known CTCF site 90 base pairs upstream of rs225014 was used (ENCODE project[1]) (DIO2-CTCF2; **Fig. 1,** lane 7-8). Lanes 1, 3, 5 and 7 depict the truncated CTCF protein, containing just the CTCF DNA binding domain whereas lanes 2, 4, 6 and 8 depict the full length CTCF protein. For lane 1 to 6 no band shifts were observed, indicating that CTCF does not bind the putative CTCF binding site located at the rs225014 SNP. However, we were able to confirm the putative CTCF binding site upstream of rs225014, as for both proteins a band shift was observed (lane 7-8).

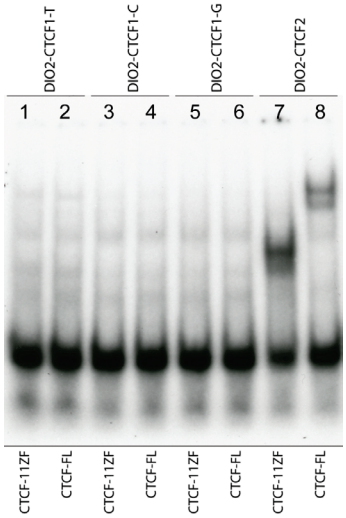


Figure S3. *DIO2* overexpression in hBMSCs, proven by co-transduced eGFP expression. A) GFP expression seen in cells cultured on culture dishes **B)** Normal light microscopy picture of a hBMSC pellet after 3 weeks of culturing. **C)** Fluorescence picture of a non-transduced hBMSC pellet after 3 weeks of culturing. **D)** Fluorescence picture of an hBMSC pellet transduced with the control vector containing only the sequence for eGFP after 3 weeks of culturing. **E)** Fluorescence picture of a hBMSC pellet transduced with the *DIO2* overexpressing vector containing both the sequence for eGFP and *DIO2*. **(B, C, D,and E)** Scale bar, 200 µm.

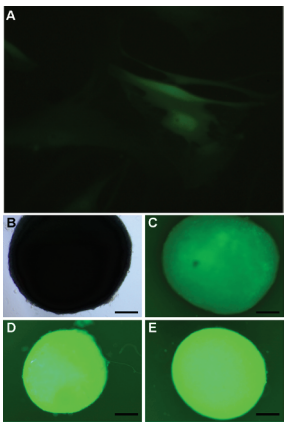


Figure S5. Calculating the relative pixel intensity; an automated histology quantification method. A) Original image algorithm. **B)** IsoData thresholding without rolling ball algorithm. **C)** IsoData thresholding after roling ball algorithm. **D)** masking-area after separation **E)** Glycosaminoglycan histology separated from the background **(A-E)** Scale bar, 400 µm.

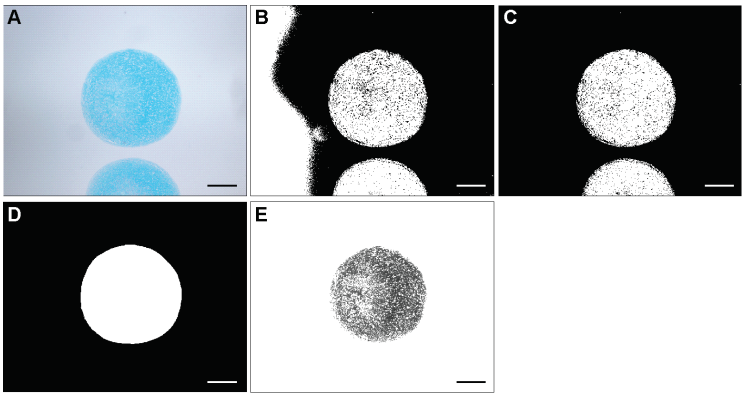
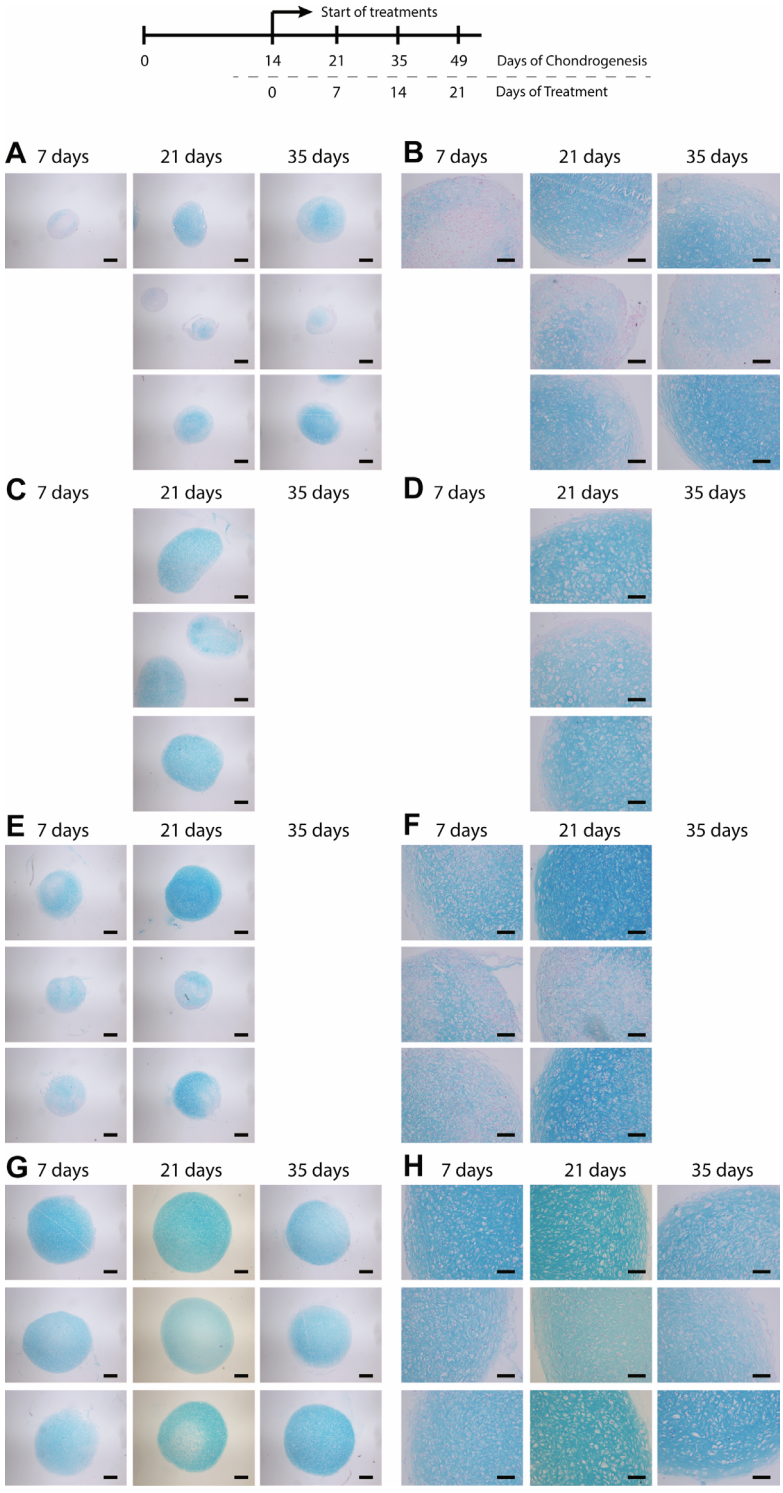


Figure S4 (next page). The effect of excess T3 and IOP in a human BMSC in vitro chondrogenesis model. (A-G) Alcian Blue staining of sections comparing control (top), thyroid treated (middle) and IOP treated (bottem) chondrogenic hBMSC pellets of **(A-B)** donor 54, **(C-D)** donor 55, **(E-F)** donor 56, and **(G-H)** donor 57. **(A, C, E, and G)** Scale bar, 400 µm. **(B, D, F, and H)** Scale bar, 100 µm.


Start of treatments
0
14
21
35
49
Days of Chondrogenesis
0
7
14
21
Days of Treatment
A 7 days
21 days
35 days
B 7 days
21 days
35 days
C 7 days
21 days
35 days
D 7 days
21 days
35 days
E 7 days
21 days
35 days
F 7 days
21 days
35 days
G 7 days
21 days
35 days
H 7 days
21 days
35 days

Discussion

8

Parts of this text will be included in an invited review for Current Rheumatology Reports

GENERAL DISCUSSION

Osteoarthritis (OA) is a complex heterogeneous disease with a considerable heritable component. While the incidence of the disease is steadily increasing, up until now no treatment options are available that could cure or stop disease progression. Main impediments in OA management are the lack of biomarkers that detect OA at an early stage when damage is still reversible, or that allow stratification of subsets of OA patients with different disease modes. This considerably hampers development of effective therapies and requires insight into the heterogeneous etiology of OA.

The aim of this thesis was threefold. Firstly, to address OA biomarker impediments by exploring gene expression profiles as biomarkers in osteoarthritic cartilage and blood (chapter 2 and 3), and by investigating whether we could identify genetic factors that influence levels of promising OA biomarkers such as serum COMP and urinary CTX-II in the common population independent of OA status (chapter 4). We hypothesized that sensitivity of such biomarkers could be improved by taking into account these genetic factors. Secondly, to explore the possibility to identify OA susceptibility genes by either applying a GWA study with sibling pairs (familial OA cases) towards the severe end of common/late onset OA phenotype (symptomatic OA at multiple joint sites) of the GARP study (chapter 5) or by applying whole exome sequencing of family members with early-onset symptomatic OA at multiple joint sites (chapter 6). Thirdly, to initiate the necessary shift from genetic towards functional genomic studies in OA research for the deiodinase iodothyronine type-2 (*DIO2*) gene using *in vitro* chondrogenesis models (chapter 7). Together, results will contribute to the selection of druggable targets and the badly needed development of new treatments as well as the identification of biomarkers to early detect ongoing OA.

BIOMARKERS IN OSTEOARTHRITIS

At present, timely diagnosis of OA before joint damage becomes irreversible is still lacking. This is one of the impediments in OA treatment and thus, biomarkers that can ascertain OA at an early stage are urgently needed. In addition, clinical biochemical markers to classify patients with different OA subtypes and to sensitively monitor OA disease activity in an individual over time or to assess quantitative joint tissue remodeling are required. In 2006, Bauer and colleagues proposed the BIPED biomarker classification to be applied in development and analysis of OA biomarkers.[1] BIPED stands for 'Burden of Disease, Investigative, Prognostic, Efficacy of Intervention and Diagnostic' and serves to facilitate research (both independent and collaborative) and to reduce redundancy while specifying and accelerating the validation of potential biomarkers.

Studies thus far have suggested that serum cartilage oligomeric protein (sCOMP) and urinary C-telopeptide of type II collagen (uCTX-II) are promising candidates hence they were listed among a number of biomarkers recommended to focus on in biomarker research.[2] However, due to the high heterogeneity in OA phenotypes and

relatively large inter-individual variation also among healthy subjects, the use of biomarkers is complicated. In chapter 4, we have investigated whether we could identify genetic variants affecting innate levels of sCOMP and uCTX-II that could possibly explain the variance and improve the use of respective biomarkers in the clinic. We found specific loci in association with biomarker levels. Among others, we found a single nucleotide polymorphism (SNP) in close vicinity to the Mannose receptor C type 1 (*MRC1*) gene in association with levels of sCOMP with genome wide significance which was not in association with hip or knee OA. MRC1 is a membrane receptor highly expressed by macrophages and it has been shown that MRC1 mediates specific recognition of sugar residues to allow endocytosis of glycoproteins.[3] Since *MRC1* is only lowly expressed in articular cartilage and carriers of the risk allele have decreased levels of sCOMP, we hypothesize that carriers of the risk allele may have enhanced phagocytosis of serum glycoproteins such as sCOMP which could negatively affect the sensitivity of sCOMP as biomarker of OA in a clinical setting. Our findings are encouraging and incite additional studies to identify genetic loci associated with other biomarkers. However, most important now is to investigate whether taking the genetic variation into account will improve the use of these biomarkers and may allow for the distinction of OA subtypes.

Unconventional biomarkers. Traditionally, biomarkers are thought of as biochemical substances and studies for development of biomarkers are still mainly focused on biochemical components in body fluids that mark either cartilage, bone, or synovial metabolism, or a combination of these.[4] In addition to these more traditional biomarkers, imaging methods such as quantitative magnetic resonance imaging (MRI) to assess cartilage volume and thickness or shape by statistical shape modeling have shown to be suitable as biomarkers for OA.[5-7] Disadvantage of these methods is, however, their relatively high costs. Currently, molecular profiles of genome wide gene expression or epigenetic variation are increasingly considered as optional targets in biomarker development. The fact that organ-specific pathologies can be detected from easily accessible tissues such as blood which was recently confirmed by integration of multiple studies,[8] encourages research for the identification of novel molecular markers. Differentiation of patients based on gene expression profiles in blood has been successfully performed for example for coronary artery disease,[9, 10] and several types of cancer.[11-13] Also for knee OA patients, gene expression profiles in blood were generated successfully and resulted in the identification of different subsets of patients.[14] In this thesis we have used gene expression profiles in blood to identify diagnostic biomarkers that could distinguish patients from the GARP study from healthy controls. Inclusion of the potential molecular biomarkers in the estimated cross-validated algorithm in receiver operating curve (ROC) analysis yielded a 20% increase of the area under the curve (AUC) to 91-95% (chapter 3). This clinically relevant AUC obtained in our analysis is expected to be accomplished by virtue of the combination of body mass index (BMI), also augmenting the AUC with 20% to approximately 72%, with the expression of a specific set of genes. This was

confirmed in an independent selection of the GARP study, however, controls were overlapping and results await replication with population-based controls. Moreover, because our study has a cross-sectional design, further investigation in a longitudinal study is necessary to assess predictive value and putative benefit for early diagnosis as well as disease progression in a population-based setting.

Gene expression profiles in blood may either be more systemic and thus independent of OA pathophysiological processes in the joint or reflect ongoing OA processes. Therefore we explored the overlap between expression profiles in blood and in cartilage and found that around 25% of the genes with fold-changes of ≥1.5 coincided. For example, a gene overlapping between our blood and cartilage expression profiles was Serine/Arginine-Rich Splicing Factor 5 (*SFRS5*), highly expressed in cartilage, and in blood contributing most to our prediction model of OA patients. *SFRS5* encodes one of the members of the SR family of proteins that binds to exonic splicing enhancer cis-sequences and it was shown to be involved in alternative splicing of the fibronectin gene specifically in chondrocytes.[15] Another gene overlapping between blood and cartilage profiles was *H3F3B*, encoding the H3 histone family 3B. While in cartilage *H3F3B* was only 1.2-fold decreased, in blood the reduction was more than 2-fold making it particularly attractive to further study its putative application as a biomarker. A recent study showed association of a specific mutation in *H3F3B* with chondroblastoma, while mutations in *H3F3A*, a homolog of *H3F3B*, were exclusively found in association with giant cell tumors of the bone.[16] Interestingly, in *H3F3B* knockout mice reduced H3K4me3 (a marker for active gene transcription) was observed preferentially near genes related to extracellular matrix.[17] Together, this could suggests an essential role for *H3F3B* in chondrocyte proliferation and cartilage homeostasis, and maybe in susceptibility to OA development. However, this needs further investigation.

Remarkably, of the genes overlapping between blood and cartilage expression profiles only half appeared to have the same direction of effects. Possibly, this stems from the differences in the study set-up: while the cartilage expression profiles were generated from joints of patients with end-stage OA, patients from the GARP study have symptomatic OA at multiple joints and blood profiles were compared with those of healthy subjects. Expression differences may therefore relate to the differences in timing and mechanism of development of OA. Alternatively, since expression of genes in blood may rather reflect a systemic profile, this can be different from the more direct pathophysiological processes taking place in the joint tissues. For example, a gene that is increased in blood of OA patients but decreased in OA cartilage is *TGFBR3* (transforming growth factor beta receptor III). As known, involvement of members of the TGF-β superfamily in cell proliferation, differentiation, morphogenesis, tissue homeostasis and regeneration is highly cell- and context-dependent.[18, 19] Therefore, it could very well be that OA patients have increased levels of *TGFBR3* in blood cells while the processes taking place during OA development in articular cartilage result in a decreased expression of the gene. For future studies, combining joint tissue databases with blood expression profiles as well as the inclusion of samples (joint tissues and blood)

from healthy individuals may be of help for the selection of biomarkers that are able to sensitively monitor disease activity.

A field in development in biomarker research is the use of non-coding RNAs as biomarkers. Several types of non-coding RNAs can be distinguished that are generally classified as small RNAs such as miRNA, piRNAs, and long non-coding RNAs or lncRNAs.[20] Micro RNAs can be detected in circulating body fluids such as plasma and serum where they are relatively stable because they are protected from RNAse activity by virtue of their association with secreted membrane vesicles or RNA-binding proteins,[21] and multiple studies now suggest that specific circulating miRNAs sensitively reflect ongoing pathophysiological processes[22] thereby qualifying their use as potential biomarkers, also for OA. Studies investigating expression of OA-specific miRNAs however, are still limited. To our knowledge, the largest study thus far was performed by Beyer *et al* including 816 individuals of the Bruneck cohort.[23] In this longitudinal study, miRNA-let-7e was identified as a negative, dose-dependent predictor for severe knee or hip osteoarthritis. Further research is needed with replication and validation in other cohorts, to obtain data on specificity and sensitivity of miRNAs to monitor dynamic changes in the OA pathophysiological processes over time. In addition, it could be investigated whether miRNAs serve to identify patients prone to fast progression. This would allow for more adequate OA disease management and more efficient design of clinical trials.

IDENTIFYING OA SUSCEPTIBILITY GENES

One of the drawbacks scientists encounter in their research to discover genetic variants causing susceptibility to OA is the so-called 'missing heritability'.[24] In part, the problem might be caused by an overestimation of heritability in the general population based on results of twin and family studies. Alternatively, this may be due to the fact that population-based or case-control studies with GWAS are designed specifically to detect common variants which usually have small effect sizes.[25] As a result, variants with larger effect sizes and/or that associate specifically with more rare and severe OA phenotypes, such as the variants identified for *DIO2* and *FRZB*, are not detected at the genome wide significance level in GWA studies.[26-28] Nevertheless, given the functional impact, these variants are likely contributing significantly to the percentage of heritability. To illustrate, the *DIO2* susceptibility SNP rs225014 was shown to have an odds ratio of OR = 1.79 consequently involving a heritability of 5.3 which is in sharp contrast to the average heritability of established loci as identified by GWAS ($H^2_{average}$ = 0.8). In general, family-based studies are prone to detect variants with moderate to large effect sizes. We investigated whether selection based on cases with familial and relatively severe phenotypes could contribute to the identification of variants with larger effects in GWAS (chapter 5) and in whole exome sequencing (chapter 6). Although identified variants will be rare and frequently private to the respective family their value lies in the expectation that more common variants in the same gene or pathway in which the gene acts will be

present in a larger population.[29] In these studies, the availability of a gene expression dataset of (OA affected and unaffected) articular cartilage (the RAAK study) allowed efficient selection of putative candidates and generalizability of the results.

Whole exome sequencing. In chapter 6 we applied whole exome sequencing to 2 distant members of a family suffering from early-onset OA. A suitable prioritization scheme and analysis of gene expression levels in articular cartilage indicated a mutation in the tumor necrosis factor receptor superfamily member 11B (*TNFRSF11B*) gene encoding osteoprotegerin (OPG) to be likely causal to the phenotype in our family. This was in line with the genome wide significant linkage analysis at this locus (LOD-score of 3.48) and complete segregation of the mutation with early-onset OA in the family.

The decoy receptor OPG antagonizes receptor activator of NF-κB (RANK) in the process of osteoclastogenesis. Functional *in vitro* analysis now showed that the mutation in *TNFRSF11B* results in enhanced OPG-mediated RANK antagonism and the significant upregulation of *TNFRSF11B* in articular cartilage suggests that this is a phenomenon in general OA pathology. Our results that the mutation in OPG is causal to the FOA phenotype in our family are highly convincing, nevertheless, further research is required. The phenotype of the family has to be studied in more detail and final proof for the causal role of OPG in OA should be provided by the performance of *in vivo* experiments in transgenic mice bearing the identified mutation to confirm the role of the identified mutation in development of OA. OPG transgenic and knockout mice have been generated previously and are viable,[30, 31] nonetheless, they were never challenged with OA-inducing trials such as mechanical loading through forced running. Interestingly, OPG transgenic mice are osteopetrotic which is caused by the impaired bone remodeling due to failure of the osteoclasts to resorb bone, but their phenotype is different from other mouse strains with osteopetrosis. The phenotype, also characteristic for RANKL knockout mice,[32] is impaired tooth eruption and shortened, 'club-shaped' long bones due to complete lack of mature osteoclasts. It was suggested that, depending on the differences in bone surfaces (e.g. endosteal and periosteal surfaces), the capacity of systemic OPG to inhibit RANKL varies. These circumstances make the application of experimentally induced OA in OPG transgenic mice even more interesting since results are likely to lead to deeper insight into the role of the bone-cartilage interaction in OA. Further *in vitro* analyses which can be performed among others by comparing chondrogenic and osteogenic potential of iPS cells derived from affected and unaffected family members can give more insight into the underlying mechanisms of the mutated OPG in OA.

Increased RANK antagonism leads to decreased osteoclastogenesis and it may thus result in concurrent increased bone mineral density (BMD). The variation in BMD towards higher levels would be in line with standing epidemiological studies showing increased BMD as a potential risk factor for OA.[33-37] On the other hand, 2 clinical studies,[38, 39] although subject of debate,[40-42] suggested that OA patients benefit from treatment with strontium renalate, an anti-osteoporotic drug that increases bone formation while decreasing bone resorption. Also *in vitro*[43] and *in vivo*[44-46] experiments indicated putative beneficial effects of strontium renalate. In this respect it

should be noted that the methods applied in the *in vivo* studies are broadly used in animal studies, nevertheless, it has been questioned whether these methods are the best to investigate pathophysiological processes of primary OA as it occurs in humans (see '*In vivo animal models in OA*'). Possibly, subtypes of OA exist that will benefit from bone forming therapies while others will not. For example, among patients suffering from hip OA those with atrophic OA had systemically lower BMD as compared to those with normotrophic OA and as compared to the controls, while those with hypertrophic (osteophytic) OA had higher BMD.[47] Given that gene expression profiles of human osteoarthritic cartilage indicated enhanced OPG-mediated RANK antagonism as a more general phenomenon in the OA pathophysiological process (chapter 2 of this thesis), we advocate that careful investigation of the long term effects of these bone forming therapies in osteoarthritis are required and that, in contrast to the use of bone-forming therapies such as strontium renalate, (at least for some patients) agents counteracting OPG function could contribute to the development of new disease modifying treatments in osteoarthritis.

GWAS using familial phenotype towards the severe end of the OA spectrum. By performing a genome wide association study in a very specific cohort consisting of familial cases with severe symptomatic and radiographic OA (KL>2) in at least two joints (the GARP study) we aimed at the identification of loci that may be less common but with larger effect sizes.[25] Several loci annotated to interesting genes (e.g. the calcium sensing receptor or *CASR*) were identified with larger effect sizes (mean odds ratio 1.8 with range: 1.5-2.1). However, none of these loci were genome wide significant (smallest p-value: 1.2×10^{-7}) which may indicate a lack of power despite the predictions of power calculation programs. Besides lack of power, a problem when performing GWAS with a specific phenotype is subsequent replication of the results obtained in the discovery stage. The more specific a cohort is, the smaller the chance to find another study with the same phenotype. We have not been able to find other cohorts with genotype data of cases with familial generalized OA comparable with the GARP study. Furthermore, effect sizes of a study with more severe phenotype (deCODE), albeit not selected for familial affiliations, were again much smaller indicating either heterogeneity in the phenotype or false positive findings in the discovery cohort. Since GWAS platforms are generally based on common variants the causal (less common) SNP may not be included or removed by selection for allele frequencies of at least 5%. Replication in cohorts with more common phenotypes did not result in robust associations. This was also demonstrated by a GWA study from Evans *et al* for hip OA.[48] For the discovery analysis with genome wide significant results, cases were carefully examined and strictly selected to acquire a more homogeneous cohort. However, results of the combined analyses of discovery and replication were not genome wide significant ($p=1\times10^{-6}$). In conclusion, genome wide significant results in GWAS derive from a trade-off between large sample sizes and homogeneous cohorts. Confirmation of our results in a cohort similar to the GARP study could ascertain whether selection of samples towards the severe end of the spectrum of OA and with a familial phenotype may indeed be a promising approach to find loci with larger effects (average odds ratio of the

Table 1 Top 25 of gene-based association analysis. Analysis was performed by joining all SNPs within known druggable genes (N= 4,290) and subsequently testing each of the genes for association with OA.

Gene Symbol	# SNPs	*p*	FDR
SLCO2B1	12	7.17E-05	2.79E-01
CXCL11	3	1.62E-04	3.15E-01
KCNQ4	14	2.72E-04	3.53E-01
ABCB11	26	5.82E-04	5.66E-01
SLC12A6	22	6.45E-04	5.02E-01
SELP	24	7.07E-04	4.59E-01
FTL	3	7.44E-04	4.14E-01
SLCO5A1	49	8.46E-04	4.12E-01
KITLG	12	1.02E-03	4.43E-01
STRADA	1	1.36E-03	5.30E-01
DPYD	117	1.50E-03	5.30E-01
RHBDL1	6	1.53E-03	4.96E-01
NCOA3	20	1.86E-03	5.58E-01
BAI3	148	2.08E-03	5.79E-01
GRIN2B	154	2.12E-03	5.50E-01
ADAMTS12	108	2.14E-03	5.20E-01
RPP25	1	2.29E-03	5.25E-01
ASPRV1	1	3.08E-03	6.66E-01
KCNK4	1	3.24E-03	6.64E-01
BNIP3	1	3.33E-03	6.47E-01
CD44	47	3.34E-03	6.20E-01
CTRB2	7	3.45E-03	6.10E-01
IGFBP3	7	3.59E-03	6.07E-01
CRABP2	9	3.92E-03	6.36E-01
PCDHB9	3	4.02E-03	6.26E-01

established OA susceptibility loci detected in large meta-analyses with genome wide significance is 1.2 with range 1.1-2.0 compared to 1.8 in our study).

In the genomic region around rs347842, locus with strongest evidence for association with OA, another strong signal was identified which may be independent but which may also be (partially) shared: rs17236749, located close to *C15orf29*. Therefore, we reasoned that analysis for burden of association per gene could possibly increase the power to identify genetic regions or genes involved in development of OA. Results of our GWAS were used to perform gene-based association analysis by joining all SNPs within known druggable genes (N= 4,290). Subsequently, each of these genes was tested for association with OA. Table 1 shows the 25 most significant genes [YFMR & S Tsonaka, unpublished results]. Unfortunately, none was significant after correction for multiple testing. However, besides *SLC12A6* several other genes previously identified in association with OA were present such as *NCOA3*[28] and *IGFBP3*.[48] This observation may add to the credibility of our GWAS results and validate selection of loci close to druggable targets for further analyses despite the absence of genome wide significant loci.

Identifying strong candidate genes. United efforts of large consortia have ensured the discovery of loci within or close to compelling OA candidate genes by applying GWAS (see Table 1 of the Introduction). However, different from whole exome sequencing where high impact mutations are identified that are damaging for protein function, OA susceptibility loci identified by GWAS generally represent intergenic loci, in regions with multiple genes that have not yet been implicated in OA etiology, or within gene deserts and establishment of the causal gene has been proven difficult. Even when a SNP is localized within a plausible gene, such as the SNP identified within the *CASR* gene that was identified with the GARP GWAS

(chapter 5), after confirmatory genetic replication consecutive verification at multiple levels (molecular, cellular, whole organism) is required before definitive conclusions can be drawn. This was emphasized by the study of Smemo *et al*[49] for a locus associated with obesity in a very large GWAS and localized within the *FTO* (fat mass and obesity associated) gene that recently appeared to mainly target the *IRX3* gene, localized mega bases further away from the SNP. The finding stresses the observation of Freedman and colleagues[50] that 'even if a transcript is associated with a risk allele, it does not necessarily mean that the annotated gene is definitively involved in the trait of interest and that functional follow-up with assays relevant to the trait are needed to actually show that a gene is directly involved with disease development'.

Risk alleles likely act by affecting joint tissue specific gene expression.[51, 52] Therefore, a starting point to come to the gene affected by the genetic variant is analysis of expressed quantitative trait loci (eQTL) which should be performed on a more regular base in GWAS.[53] Challenge for studying eQTLs is the availability and selection of disease relevant tissues beyond the collection of blood. Although online databases exist (http://biogps.org) cartilage is not a tissue commonly included and in this respect comprehensive biobanking in OA research is of enormous importance. In chapter 2, we have shown how comparison of a relatively large number of samples from diseased and unaffected areas of cartilage from the same joint hint at generic pathological processes causing OA. Genome wide eQTL analysis was not performed within the RAAK study, but the generated dataset has been extensively used to prioritize genes in subsequent studies within the scope of this thesis (chapter 3-6) and follow-up research.[28, 54] A typical example for prioritization in regions with multiple genes could be the SNP identified within the *ASTN2* gene, rs4836732, associated with hip OA with genome wide significance in females and thus put forward as novel likely OA susceptibility gene.[27] Although we do not exclude the possibility that *ASTN2* affects OA susceptibility, the gene was not detected well in our microarray expression analysis. Having said this, we advocate that *PAPPA*, a protease for IGF binding proteins localized 100 kb upstream of *ASTN2* which increases osteoblast proliferation[55] and is 2-fold up-regulated in OA affected cartilage, should be considered as the OA susceptibility gene.

Taken together, when the correct datasets are available these can be used to identify, in high throughput manner, strong positional candidates that subsequently can be employed to cell models to address the specific effects of the genetic variation.

MOLECULAR PROCESSES IN OA

During OA, several processes take place in the joint such as cartilage degradation, bone remodeling, and inflammation. At the molecular level, each of these processes is reflected by specific gene expression networks, however, these networks are strongly intertwined among each other. Figure 1 shows a schematic representation of the different processes and some of the genes involved.

A common feature of OA cartilage is that

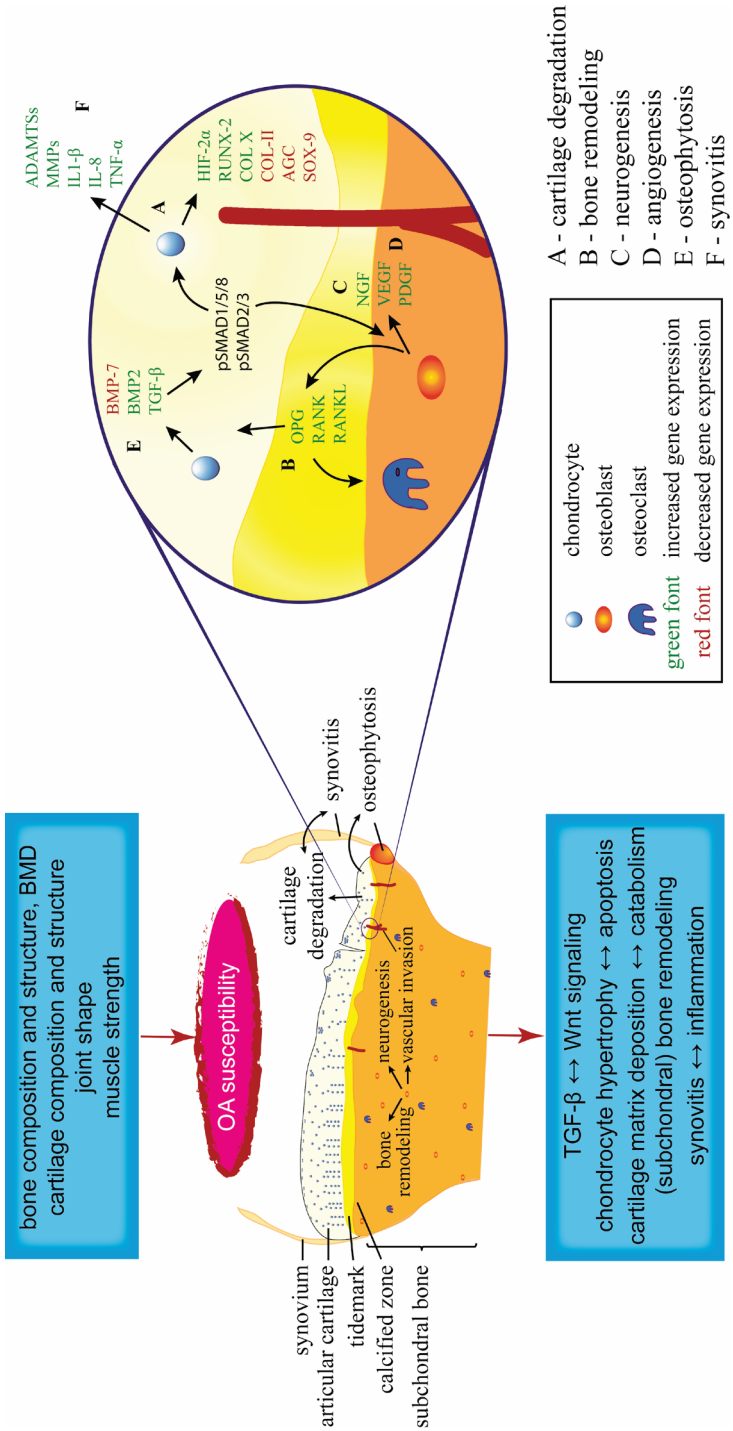


Figure 1 Schematic representation of molecular processes involved in OA.

the chondrocytes have lost their specific maturational arrested phenotype and are characterized by hypertrophy. This results in various degrees from detrimental genetic variation, cellular responses required for matrix-repair processes in damaged cartilage, and deregulated gene expression following loss of epigenetic control due to pathophysiological processes and/or aging. Chondrocyte hypertrophy results in mineralization of the cartilage due to increased expression of *RUNX2* and concurrent increase of *COL10* making the cartilage more prone to damage following mechanical stresses. In addition, hypertrophic chondrocytes express higher levels of secretion factors such as cytokines (*IL1B*, *TNF-a*), catabolic enzymes (*MMPs*, *ADAMTSs*) and members of the transforming growth factor-beta superfamily (reviewed in [56] and [57]). These are responsible for the inflammatory aspects of the arthritic joint.

One of the genetic loci underlying susceptibility to OA is annotated to the *DIO2* gene, and our studies showed that indeed impaired thyroid signaling predisposes articular chondrocytes towards hypertrophy with concurrent increased expression of *COL10*, and catabolic genes such as *MMP13*, resulting in mineralization of the articular cartilage (chapter 7). Possibly, aberrant thyroid signaling involves one of the first episodes on the way to OA. It has been shown that the *EPAS1* promoter encoding Hypoxia-inducible factor-2α (HIF-2α) contains a thyroid binding element both in mice[58] and man.[59] While HIF-1α contributes to cartilage extracellular matrix preservation via induction of *SOX9*, *COL2A1*, and *AGC*,[60] expression of HIF-2α results in cartilage hypertrophy due to up-regulation of *RUNX2* and catabolic factors (*ADAMTS5*, *MMP13*, etc.).[61, 62]

During ongoing OA, expression of *TNFRSF11B* is increased in cartilage (chapter 2) which may contribute to the mineralized phenotype of arthritic cartilage. This increased expression during OA pathophysiology is in line with the identification of a gain of function mutation as probably causal in a family with early onset OA. It has been shown that the promoter of *TNFRSF11B* contains a thyroid binding element.[59] This could be underlying the increased expression of *TNFRSF11B* in arthritic cartilage since also *DIO2* expression is increased and thus the thyroid signaling. In addition, it could be speculated that augmented thyroid signaling not only predisposes to OA via mechanisms that are detrimental to the cartilage, but also via increased bone remodeling. In fact, arthritic bone was also found to express increased levels of *TNFRSF11B*.[63]

The increased bone remodeling in OA triggers formation of channels which extend from the subchondral bone into the articular cartilage thereby breaching the tidemark. These channels allow for increased vascularization which is frequently accompanied by innervation, both of sympathetic and of sensory nerves.[64] Among others, this results in increased nerve growth factor expression within vascular channels and may explain that, besides genes involved in bone development, GWAS and gene expression analyses also have identified multiple genes involved in neurogenesis. Alternatively, it could link to aberrant neurogenesis during development with a concurrent irregular locomotor system. This would predispose to OA during aging as a result from an accumulation of joint damage.

In line with this, in mice targeted deletion of *SLC12A6*, identified in our GWAS as a putative OA susceptibility gene (chapter 5), leads to a locomotor and a sensorimotor gating deficit, suggesting a critical role for *SLC12A6* in development of the nervous system.[65, 66]

Blood expression profiles of OA patients from the GARP study were found to be enriched for genes involved in apoptosis (chapter 3). This is striking given previous observations of increased apoptosis in OA cartilage at early as well as at late stages[67] despite the fact that OA cartilage expression profiles were not enriched for apoptosis-related genes (chapter 2). However, it should be noted that the study had a cross sectional design. Consequently, gene expression differences between cases and controls can be both, the result of OA pathophysiological processes or innate processes. Until further studies have been performed the blood expression profiles should only be used for diagnosis but not for prediction or prognosis. In addition, it is known that multiple interactions exist between apoptotic genes and cytokines (e.g. *IL1B* and *IL8*, also present in the blood expression profiles). Therefore, it is also possible that the apoptotic genes rather reflect an inflammatory state which is also recognized in OA patients.[14]

COMING FROM, GOING TO: FUNCTIONAL GENOMICS

Although the number of published genetic studies that lack experimental data showing the biological relevance of identified genetic variation is still large, in the field of OA the step from genetics to biological function, also named 'functional genomics', has currently started to be implemented.[48, 54, 68, 69] To proceed and improve the approaches taken we propose a combination of those suggested by Freedman *et al*[50] and by Sunyaev[70] (Figure 2), which are similar to those applied in this thesis for OA.

Knowledge of the direction of effects of the susceptibility locus and understanding the functional consequences of the genetic variation of candidate genes is required before translation to the clinic can be considered. Alterations in gene expression most likely originate from changes in epigenetic control mechanisms[51, 71] or from changes in transcription factor binding sites[72] as generally acknowledged and outlined previously.[50, 73] Therefore, a first approach especially interesting for intronic or intergenic SNPs is to investigate whether genetic variation at the identified locus results in differential allelic expression (DAE) of the susceptibility gene as was done for example for *GDF5*,[74] *ALDH1A2*,[54] and *DIO2* by our own group.[52] Alternatively, for high impact protein variants, the effect of the variant needs to be established starting from *in silico* analysis of the predicted impact of the variant (more specifically for protein mutations), to proceed via *in vitro* cell models to the *in vivo* effects.

In vitro cell models in OA. *In vitro* cell models can be broadly applied and are especially suitable to investigate transcriptional regulation of a gene of interest and the consequences of genetic variation[69, 71, 72] which eventually may serve to interfere with its regulation in a clinical setting. Cell models can also be used to determine the gene targeted by the susceptibility locus: Chromosome Confor-

Table 2 Comparison of different cell models

	cell lines	primary cells	iPS cells
lifespan	Unlimited	Limited	Unlimited
transformation	Yes	No	No
donor variation	No	Yes	No
phenotypic stable	Yes	No	Yes
3D pellet model doable	No	Yes	Yes
experiment duration	3 weeks	3 months	3 months

mation Capture (3C), Circularized Chromosome Conformation Capture (4C), or Carbon-Copy Chromosome Conformation Capture (5C) can be applied to explore chromosome folding and regions of the genome interacting with the site of interest. In fact, the use of 4C sequencing has been of valuable help in the identification of *IRX3* as the gene genetically targeted by the obesity associated locus within the *FTO* gene.[49]

Besides the study of gene and protein perturbations to obtain mechanistic insight of OA risk genes, cell models can be used to select drugs or biologicals that *in vitro* counteract the detrimental effect of the mutations and may therefore also be selected for *in vivo* studies. In chapter 7 this was done to investigate the effects of counteracting the thyroid hormone pathway on deposition of cartilage extracellular matrix. Iopanoic acid (IOP), a pharmacological inhibitor of deiodinases, was added to our 3-dimensional (3D) pellet cultures and was found to be beneficial: upregulation of catabolic enzymes was attenuated, while the COL2A1/COL1A1 ratio was more stable and cartilage matrix structure denser leading to prolonged 'healthy' cartilage homeostasis. Although encouraging, results await confirmation of the effects of IOP on articular cartilage *in vivo*. IOP belongs to the so-called 'oral cholecystographic agents', originally used as alternative contrast agent to identify gallbladder pathology.[75] The drug is primarily a potent inhibitor of the conversion of peripheral serum inactive thyroid (T4) to active thyroid (T3) making it an effective medication for the control of hyperthyroidism and thyrotoxicosis. IOP appears to have few side effects within the window of its application, however, long-term effects and consequences for joint health following oral use remain to be established. This should first be tested in mice and can be combined with methods used to induce OA to investigate whether intake of IOP reduces development of the disease. In addition, for the long-term use it could be an advantage to screen for small molecules that act more specifically to inhibit DIO2 but not the other deiodeninases.

Different cell models can also be employed to investigate the effect of strong OA candidates. In these studies, the most important choice is which cells to use. The easy way out is to use established cell lines such as the mouse ATDC5 (chondrogenic) or MC3T3 (osteogenic) cells which, typically, are easy to obtain and maintain promoting that studies can be performed within a relatively short period of time (around 3 weeks). However, within the scope of this thesis, we have used primary human cells since established cell lines are usually transformed and gene expression is likely

modified during the transformation process. Furthermore, in cell lines the association between growth and differentiation may be altered and they have frequently accumulated mutations in the course of the extended proliferation thereby changing the responses such that they do not necessarily reflect the effect of gene variations *in vivo* in humans. Challenges of primary cells such as their phenotypic instability and limited lifespan thereby restricting study possibilities can be circumvented by using induced pluripotent stem cells (iPS) that provides a stable source of cells and which can also be differentiated towards multiple lineages including chondrogenesis and osteogenesis,[76] reflecting more likely the effects of genetic variants in men. Table 2 summarizes properties if the different types of cells.

For models of chondrogenesis it was shown previously that monolayer chondrocyte cultures mainly result in a hypertrophic phenotype,[77] thereby hampering the investigation of the effects of variable expression of the gene of interest. We observed the same for the ATDC5 cells, and attempts to form 3D pellets with ATDC5 cells similar to those formed with human bone marrow derived mesenchymal stem cells (hBMSCs) failed (data not shown). Therefore, we think that to study chondrogenesis the best option is to perform 3D pellet cultures with hBMSCs (chapter 7) or with primary human chondrocytes. To study the effect of the OPG mutation (chapter 6) we also used primary human cells for the generation of mature osteoclasts: commercially available pre-osteoclasts. An alternative for the use of the commercial pre-osteoclasts could be blood mononuclear cells that also can be differentiated towards osteoclasts, but in these cells we observed enormous donor variation and therefore this option was discarded [YFMR & AE Oostlander, unpublished data].

Cell lines are less well fitted to learn about the organic consequences on the joint as a whole since the different tissues actively interact with each other during development and disease. Therefore, likely more suitable although not applied in this thesis is the performance of cocultures. Indirect cocultures can be applied by exchange of conditioned medium as well as by culturing cells in transwell chambers that allow the diffusion of secreted factors to investigate effects on cell function, extracellular matrix deposition, and survival. Direct cocultures can be performed by using different combinations of mesenchymal stem cells, chondrocytes, osteoblasts, explants or synoviocytes. Frequently applied in studies for development of tissue regeneration and replacements suitable for implantation is the combination of mesenchymal stem cells and chondrocytes.[78-81] Of note are some studies in which even multiple-lineage coculture models were employed. Recently, in the lab of R. Tuan a 3-dimensional 'microtissue' was developed suitable to study OA pathogenesis.[82] The model system can also be used to evaluate the effects of different reagents and compounds on osteochondral health.[83] He *et al*[84] established cultures of osteoblasts, mesenchymal stem cells, and fibroblasts on a hybrid silk scaffold to investigate whether BMSCs cocultured in the presence of ligament and bone cells would differentiate into fibrocartilage, and Leyh *et al*[85] showed that mechanical and biochemical properties of extracellular matrix deposited by mesenchymal stem cells and chondrocytes is impaired in the presence of OA cartilage

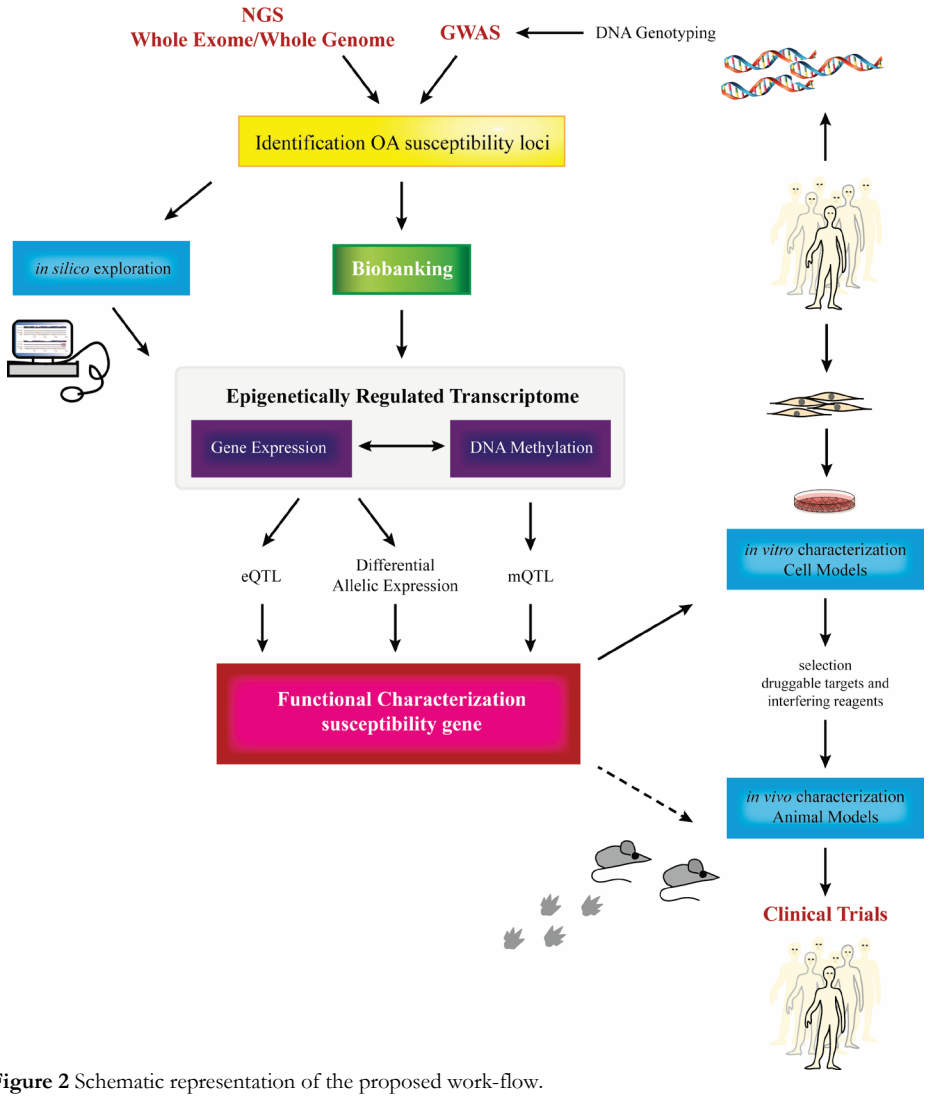


Figure 2 Schematic representation of the proposed work-flow.

explants due to altered fibrillar network. To study the interaction between cartilage and bone, chondrocytes-osteoblasts-osteoclasts, isolation of explants from patients undergoing joint replacement surgery may also be a good alternative.[86] These multi-lineage 3D *in vitro* models are the 'last best' option before getting to animal models.[87]

In vivo animal models in OA. As stated by Sunyaev *et al*[70] genetic manipulation of organisms (mouse, rat, rabbit, horse, zebrafish) provide a possibility to test the phenotypic rather than molecular consequences of human allelic variants, which is by some considered the ultimate proof for the causal effect of a susceptibility locus. Animal models can also be used to investigate new treatment options and to test

compounds selected because of their positive effect in *in vitro* assays as discussed before for IOP. Use of drugs with encouraging outcome, *i.e.* inhibited or reversed disease severity, can subsequently be put forward in clinical trials.

Of the different types of laboratory animals, mice are used more frequently because they are fairly cheap, and relatively easy to handle and to breed. Although multiple different mouse strains have been generated for candidate genes, known to be involved in chondrogenesis or endochondral ossification, this has hardly been done for discovered OA susceptibility genes. An exception has been the generation of *FRZB* knockout mice of which its characteristics were studied using established OA models.[88, 89] *GDF5* knockout mice were generated long before the identification of the *GDF5* susceptibility locus and were post-hoc used to study development of OA following different established methods (reviewed by Cornelis *et al*, 2011[90]). Also *DIO2* knockout mice were generated before discovery of the association with OA to investigate the intracellular effect of thyroid hormone bio-availability with respect to brain function.[91] So far, no studies applying OA models on mice have been published but transgenic rats overexpressing *DIO2* in articular cartilage were shown to be more prone to cartilage damage,[92] thereby confirming that the detected overexpression of *DIO2* in OA cartilage[52] is most likely not only the result of OA pathophysiology but may underlie disease development. Results presented in chapter 7 suggest that inhibition of the deiodinases is protective in cartilage, however, this awaits confirmation in *in vivo* studies to determine whether counteracting the thyroid signaling pathway may be a novel therapeutic approach.

Although not identified as an OA susceptibility gene, in 2005, a study was published reporting for the first time on a single gene deletion, *ADAMTS5*, that resulted in the abrogation of cartilage degeneration in an animal model of osteoarthritis.[93] Many other mouse models for OA have followed since, such as for HIF-2α,[61, 62] and for TGF-β,[94] and most of the recently developed mouse models are inducible and chondrocyte-specific such as the transgenic BMP2 mice[95] or the mTOR knockout mice.[96] This allows controlled expression of the gene of interest: during embryogenesis or post-natal, prior to injury or after the induction of OA. Most transgenic or knockout mice do not spontaneously develop OA but its development is provoked by means of surgery (destabilization of the medial meniscus or DMM; cruciate ligament transection), with enzymes or chemical substances (intra-articular collagenase injections), or by mechanical loading (Figure 2). An exception are STR/Ort mice that upon aging spontaneously develop OA with unknown cause,[97] and the recently generated cartilage-specific double knockout mice Nfatc1/Nfatc2-/- which exhibit early-onset, aggressive OA affecting multiple joints.[98] According to the authors this is highly favorable in OA research since it reduces the time-cost ratio of the model and does not require further surgical or chemical intervention. However, the fact that these mice are extremely prone to develop OA shortly after birth points out that it is a very specific model system, more resembling cases of early-onset OA during development than the slowly progressing 'normal' primary

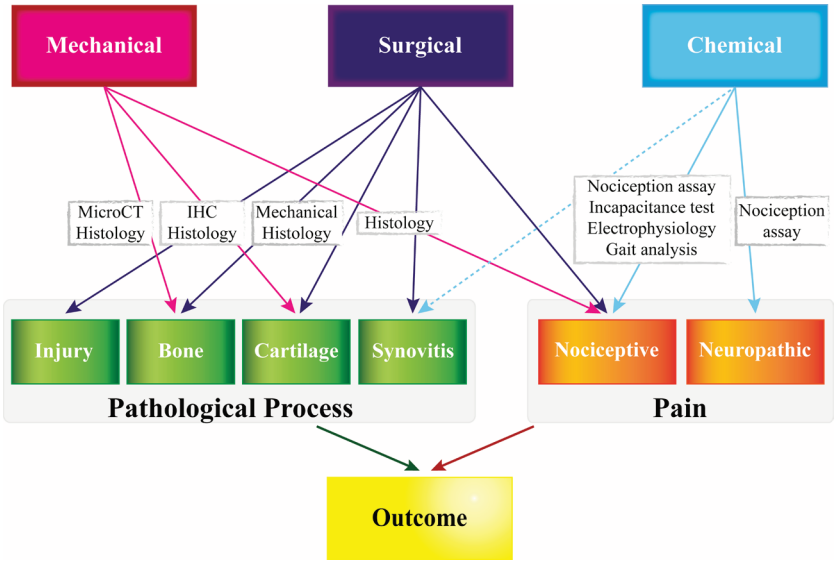


Figure 3. Animal models of OA and methods of assessment (adapted from Vincent et al 2012[89]).

OA at older age. We should be more aware of the susceptibility loci that are identified in genetic studies and generate mice with these variants to functionally characterize the effects of this 'natural' variation as it occurs in humans.

It is of note that, although destabilization of the medial meniscus was called 'minimally traumatic'[99] and here development of OA is less 'aggressive' than when using enzymes, it is yet more similar to the type of osteoarthritis developing in humans after an injurious event better known as secondary or post-traumatic OA.[100, 101] In addition, as reviewed by Aigner *et al*,[102] the pathophysiology and severity of the different methods used to induce OA are diverse and thus far no paradigm has been formulated that encompasses all of its aspects in particular its relation to aging which is an important risk factor in OA. This is not surprising since we know that OA itself is a heterogeneous disease. We should therefore take advantage of the differences between the different methods and, in case economic possibilities permit, try to include aging in the mouse models. Personally, I think mechanical loading (running) is the most appropriate stimulus since it provides a subtle way to induce OA and in particular when performed at different ages (young versus old mice) it may largely resemble the human situation.

FINAL REMARKS AND SUGGESTIONS FOR FUTURE STUDIES

To improve the clinical application of biochemical markers in OA, the genetic factors identified for sCOMP and for uCTX-II should be taken into account. In addition, it should be considered to perform further meta-analyses of GWAS for other promising biomarkers such as serum hyaluronan and serum Procollagen type II N-terminal propeptide (sPIIANP). Additionally, these efforts could identify genetic loci that

influence the OA biomarker levels as well as OA (endo)phenotypes thereby actually reflecting OA status as we have found for the *COMP* gene and sCOMP levels (chapter 4). The use of gene expression profiles which in our hands seemed very promising to distinguish patients from controls could be further analyzed. These profiles should be examined in prospective studies to establish whether the biomarkers are predictive for OA. Alternatively, in case the expression profiles in blood correlate to expression of respective genes in cartilage its change over time could be investigated which would learn whether the profiles are useful as markers to monitor the ongoing disease process.

With respect to genetic studies, given the fact that scientists have been able to collect large numbers of samples by collaborations towards large consortia, we have been able to discover several compelling OA susceptibility loci. To come to clinical applications and druggable targets, the next step in these analyses should be taken. The most recent large-scale analyses were done upon stratification for joint resulting in additional, joint-specific, OA susceptibility loci.[26, 28] We could move forward by performing rigorous phenotyping to identify more homogeneous subtypes for selection of samples as suggested previously by Kerkhof and colleagues.[103] It would be very interesting to investigate whether performing GWAS for the different OA phenotypes as distinguished by Castaño-Betancourt *et al* (atrophic, normotrophic, and hypertrophic[47]) would point towards variation in different genetic loci. Nevertheless, it is likely that the problem will be to replicate the findings since few groups have collected or characterized such particular phenotypes. Depending on the effect sizes of such less heterogenic groups, this may result in lack of power and thus genome wide significant loci may not be found. In fact, this occurred with the GWAS analysis performed within the GARP study (chapter 5). It could be questioned how many of the suggestive loci contain genes relevant for the trait. And instead of only focusing on p-values for 'genome wide significance' we may as well investigate potential susceptibility loci in the neighborhood of compelling genes at multiple biological levels before considering these loci false positives. This approach could very well open doors to interesting new targets in drug research. Functional studies are highly time consuming, however, hence applying balanced selection criteria based on both statistical and biological evidence may be more optimal.

In the course of the studies described in this thesis we have tried to make a step forward in functional genomics. In particular, this was performed by setting up *in vitro* model systems to functionally characterize identified mutations in the *TNFRSF11B* (chapter 6) and the *DIO2* gene (chapter 7), and these can be used in future studies to characterize other genes and variants such as *NCOA3*[28] and *ALDH1A2*.[54] Besides the models used in the course of this thesis, for further investigations I suggest investment in the development of co-culture models in which the interaction between different cell lineages can be studied in more detail since the interactions between the different joint tissues have a high impact on development of OA. This may also be very helpful in drug discovery research, since it can be investigated whether particular drugs have opposite direction of effects on different types of cells which is very important before initializing *in vivo* experiments.

In conclusion, the doors are open towards improvement of genetic studies and their interpretation and towards development of new strategies for the application of functional genomics in OA.

REFERENCES

1 Bauer DC, Hunter DJ, Abramson SB, Attur M, Corr M, Felson D, *et al.* Classification of osteoarthritis biomarkers: a proposed approach. *Osteoarthritis Cartilage* 2006;**14**:723-7.

2 Attur M, Krasnokutsky-Samuels S, Samuels J, Abramson SB. Prognostic biomarkers in osteoarthritis. *Curr Opin Rheumatol* 2013;**25**:136-44.

3 Largent BL, Walton KM, Hoppe CA, Lee YC, Schnaar RL. Carbohydrate-specific adhesion of alveolar macrophages to mannose-derivatized surfaces. *J Biol Chem* 1984;**259**:1764-9.

4 van Spil WE, Jansen NW, Bijlsma JW, Reijman M, DeGroot J, Welsing PM, *et al.* Clusters within a wide spectrum of biochemical markers for osteoarthritis: data from CHECK, a large cohort of individuals with very early symptomatic osteoarthritis. *Osteoarthritis Cartilage* 2012;**20**:745-54.

5 Betancourt MC, Linden JC, Rivadeneira F, Rozendaal RM, Zeinstra SM, Weinans H, *et al.* Dual energy x-ray absorptiometry analysis contributes to the prediction of hip osteoarthritis progression. *Arthritis Res Ther* 2009;**11**:R162.

6 Agricola R, Reijman M, Bierma-Zeinstra SM, Verhaar JA, Weinans H, Waarsing JH. Total hip replacement but not clinical osteoarthritis can be predicted by the shape of the hip: a prospective cohort study (CHECK). *Osteoarthritis Cartilage* 2013;**21**:559-64.

7 Castano-Betancourt MC, van Meurs JB, Bierma-Zeinstra S, Rivadeneira F, Hofman A, Weinans H, *et al.* The contribution of hip geometry to the prediction of hip osteoarthritis. *Osteoarthritis Cartilage* 2013;**21**:1530-6.

8 Sung J, Kim PJ, Ma S, Funk CC, Magis AT, Wang Y, *et al.* Multi-study integration of brain cancer transcriptomes reveals organ-level molecular signatures. *PLoS Comput Biol* 2013;**9**:e1003148.

9 Rosenberg S, Elashoff MR, Beineke P, Daniels SE, Wingrove JA, Tingley WG, *et al.* Multicenter validation of the diagnostic accuracy of a blood-based gene expression test for assessing obstructive coronary artery disease in nondiabetic patients. *Ann Intern Med* 2010;**153**:425-34.

10 Elashoff MR, Wingrove JA, Beineke P, Daniels SE, Tingley WG, Rosenberg S, *et al.* Development of a blood-based gene expression algorithm for assessment of obstructive coronary artery disease in non-diabetic patients. *BMC Med Genomics* 2011;**4**:26.

11 Habermann JK, Bader FG, Franke C, Zimmermann K, Gemoll T, Fritzsche B, *et al.* From the genome to the proteome--biomarkers in colorectal cancer. *Langenbecks Arch Surg* 2008;**393**:93-104.

12 Cui J, Chen Y, Chou WC, Sun L, Chen L, Suo J, *et al.* An integrated transcriptomic and computational analysis for biomarker identification in gastric cancer. *Nucleic Acids Res* 2011;**39**:1197-207.

13 Tanase CP, Neagu M, Albulescu R, Hinescu ME. Advances in pancreatic cancer detection. *Adv Clin Chem* 2010;**51**:145-80.

14 Attur M, Belitskaya-Levy I, Oh C, Krasnokutsky S, Greenberg J, Samuels J, *et al.* Increased interleukin-1beta gene expression in peripheral blood leukocytes is associated with increased pain and predicts risk for progression of symptomatic knee osteoarthritis. *Arthritis Rheum* 2011;**63**:1908-17.

15 Kuo BA, Uporova TM, Liang H, Bennett VD, Tuan RS, Norton PA. Alternative splicing during chondrogenesis: modulation of fibronectin exon EIIIA splicing by SR proteins. *J Cell Biochem* 2002;**86**:45-55.

16 Behjati S, Tarpey PS, Presneau N, Scheipl S, Pillay N, Van LP, *et al.* Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. *Nat Genet* 2013;**45**:1479-82.

17 Bush KM, Yuen BT, Barrilleaux BL, Riggs JW, O'Geen H, Cotterman RF, *et al.* Endogenous mammalian histone H3.3 exhibits chromatin-related functions during development. *Epigenetics Chromatin* 2013;**6**:7.

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

19 Massague J. TGFbeta signalling in context. *Nat Rev Mol Cell Biol* 2012;**13**:616-30.

20 Morris KV, Mattick JS. The rise of regulatory RNA. *Nat Rev Genet* 2014;**15**:423-37.

21 Kosaka N, Yoshioka Y, Hagiwara K, Tominaga N, Katsuda T, Ochiya T. Trash or Treasure: extracellular microRNAs and cell-to-cell communication. *Front Genet* 2013;**4**:173.

22 Reid G, Kirschner MB, van ZN. Circulating microRNAs: Association with disease and potential use as biomarkers. *Crit Rev Oncol Hematol* 2011;**80**:193-208.

23 Beyer C, Zampetaki A, Lin NY, Kleyer A, Perricone C, Iagnocco A, *et al.* Signature of circulating microRNAs in osteoarthritis. *Ann Rheum Dis* 2014.

24 Maher B. Personal genomes: The case of the missing heritability. *Nature* 2008;**456**:18-21.

25 Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, *et al.* Finding the missing heritability of complex diseases. *Nature* 2009;**461**:747-53.

26 Evangelou E, Valdes AM, Kerkhof HJ, Styrkarsdottir U, Zhu Y, Meulenbelt I, *et al.* Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. *Ann Rheum Dis* 2011;**70**:349-55.

27 Zeggini E, Panoutsopoulou K, Southam L, Rayner NW, Day-Williams AG, Lopes MC, *et al.* Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* 2012;**380**:815-23.

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

29 Goldstein DB, Allen A, Keebler J, Margulies EH, Petrou S, Petrovski S, *et al.* Sequencing studies in human genetics: design and interpretation. *Nat Rev Genet* 2013;**14**:460-70.

30 Simonet WS, Lacey DL, Dunstan CR, Kelley M, Chang MS, Luthy R, *et al.* Osteoprotegerin: a novel secreted protein involved in the regulation of bone density. *Cell* 1997;**89**:309-19.

31 Min H, Morony S, Sarosi I, Dunstan CR, Capparelli C, Scully S, *et al.* Osteoprotegerin reverses osteoporosis by inhibiting endosteal osteoclasts and prevents vascular calcification by blocking a process resembling osteoclastogenesis. *J Exp Med* 2000;**192**:463-74.

32 Kong YY, Yoshida H, Sarosi I, Tan HL, Timms E, Capparelli C, *et al.* OPGL is a key regulator of osteoclastogenesis, lymphocyte

development and lymph-node organogenesis. *Nature* 1999;**397**:315-23.

33 Hannan MT, Anderson JJ, Zhang Y, Levy D, Felson DT. Bone mineral density and knee osteoarthritis in elderly men and women. The Framingham Study. *Arthritis Rheum* 1993;**36**:1671-80.

34 Hart DJ, Mootoosamy I, Doyle DV, Spector TD. The relationship between osteoarthritis and osteoporosis in the general population: the Chingford Study. *Ann Rheum Dis* 1994;**53**:158-62.

35 Burger H, van Daele PL, Odding E, Valkenburg HA, Hofman A, Grobbee DE, *et al.* Association of radiographically evident osteoarthritis with higher bone mineral density and increased bone loss with age. The Rotterdam Study. *Arthritis Rheum* 1996;**39**:81-6.

36 Nevitt MC, Zhang Y, Javaid MK, Neogi T, Curtis JR, Niu J, *et al.* High systemic bone mineral density increases the risk of incident knee OA and joint space narrowing, but not radiographic progression of existing knee OA: the MOST study. *Ann Rheum Dis* 2010;**69**:163-8.

37 Hardcastle SA, Dieppe P, Gregson CL, Hunter D, Thomas GE, Arden NK, *et al.* Prevalence of radiographic hip osteoarthritis is increased in high bone mass. *Osteoarthritis Cartilage* 2014.

38 Reginster JY, Badurski J, Bellamy N, Bensen W, Chapurlat R, Chevalier X, *et al.* Efficacy and safety of strontium ranelate in the treatment of knee osteoarthritis: results of a double-blind, randomised placebo-controlled trial. *Ann Rheum Dis* 2013;**72**:179-86.

39 Bruyere O, Reginster JY, Bellamy N, Chapurlat R, Richette P, Cooper C. Clinically meaningful effect of strontium ranelate on symptoms in knee osteoarthritis: a responder analysis. *Rheumatology (Oxford)* 2014;**53**:1457-64.

40 Herrero-Beaumont G, Roman-Blas JA, Largo R, Berenbaum F, Castaneda S. Bone mineral density and joint cartilage: four clinical settings of a complex relationship in osteoarthritis. *Ann Rheum Dis* 2011;**70**:1523-5.

41 Lafeber FP, van Laar JM. Strontium ranelate: ready for clinical use as disease-modifying osteoarthritis drug? *Ann Rheum Dis* 2013;**72**:157-61.

42 Tenti S, Cheleschi S, Guidelli GM, Galeazzi M, Fioravanti A. What about strontium ranelate in osteoarthritis? Doubts and securities. *Mod Rheumatol* 2014.

43 Henrotin Y, Labasse A, Zheng SX, Galais P, Tsouderos Y, Crielaard JM, *et al.* Strontium ranelate increases cartilage matrix formation. *J Bone Miner Res* 2001;**16**:299-308.

44 Pelletier JP, Kapoor M, Fahmi H, Lajeunesse D, Blesius A, Maillet J, *et al.* Strontium ranelate reduces the progression of experimental dog osteoarthritis by inhibiting the expression of key proteases in cartilage and of IL-1beta in the synovium. *Ann Rheum Dis* 2013;**72**:250-7.

45 Yu DG, Ding HF, Mao YQ, Liu M, Yu B, Zhao X, *et al.* Strontium ranelate reduces cartilage degeneration and subchondral bone remodeling in rat osteoarthritis model. *Acta Pharmacol Sin* 2013;**34**:393-402.

46 Sagar DR, Ashraf S, Xu L, Burston JJ, Menhinick MR, Poulter CL, *et al.* Osteoprotegerin reduces the development of pain behaviour and joint pathology in a model of osteoarthritis. *Ann Rheum Dis* 2013.

47 Castano-Betancourt MC, Rivadeneira F, Bierma-Zeinstra S, Kerkhof HJ, Hofman A, Uitterlinden AG, *et al.* Bone parameters across different types of hip osteoarthritis and their relationship to osteoporotic fracture risk. *Arthritis Rheum* 2013;**65**:693-700.

48 Evans DS, Cailotto F, Parimi N, Valdes AM, Castano-Betancourt MC, Liu Y, *et al.* Genome-wide association and functional

studies identify a role for IGFBP3 in hip osteoarthritis. *Ann Rheum Dis* 2014.

49 Smemo S, Tena JJ, Kim KH, Gamazon ER, Sakabe NJ, Gomez-Marin C, *et al.* Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* 2014;**507**:371-5.

50 Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, *et al.* Principles for the post-GWAS functional characterization of cancer risk loci. *Nat Genet* 2011;**43**:513-8.

51 Egli RJ, Southam L, Wilkins JM, Lorenzen I, Pombo-Suarez M, Gonzalez A, *et al.* Functional analysis of the osteoarthritis susceptibility-associated GDF5 regulatory polymorphism. *Arthritis Rheum* 2009;**60**:2055-64.

52 Bos SD, Bovee JV, Duijnisveld BJ, Raine EV, van Dalen WJ, Ramos YF, *et al.* Increased type II deiodinase protein in OA-affected cartilage and allelic imbalance of OA risk polymorphism rs225014 at DIO2 in human OA joint tissues. *Ann Rheum Dis* 2012;**71**:1254-8.

53 Westra HJ, Franke L. From genome to function by studying eQTLs. *Biochim Biophys Acta* 2014.

54 Styrkarsdottir U, Thorleifsson G, Helgadottir HT, Bomer N, Metrustry S, Bierma-Zeinstra S, *et al.* Severe osteoarthritis of the hand associates with common variants within the ALDH1A2 gene and with rare variants at 1p31. *Nat Genet* 2014;**46**:498-502.

55 Qin X, Wergedal JE, Rehage M, Tran K, Newton J, Lam P, *et al.* Pregnancy-associated plasma protein-A increases osteoblast proliferation in vitro and bone formation in vivo. *Endocrinology* 2006;**147**:5653-61.

56 Sun MM, Beier F. Chondrocyte hypertrophy in skeletal development, growth, and disease. *Birth Defects Res C Embryo Today* 2014;**102**:74-82.

57 Yuan XL, Meng HY, Wang YC, Peng J, Guo QY, Wang AY, *et al.* Bone-cartilage interface crosstalk in osteoarthritis: potential pathways and future therapeutic strategies. *Osteoarthritis Cartilage* 2014;**22**:1077-89.

58 Chatonnet F, Guyot R, Benoit G, Flamant F. Genome-wide analysis of thyroid hormone receptors shared and specific functions in neural cells. *Proc Natl Acad Sci U S A* 2013;**110**:E766-E775.

59 Ayers S, Switnicki MP, Angajala A, Lammel J, Arumanayagam AS, Webb P. Genome-wide binding patterns of thyroid hormone receptor beta. *PLoS One* 2014;**9**:e81186.

60 Duval E, Leclercq S, Elissalde JM, Demoor M, Galera P, Boumediene K. Hypoxia-inducible factor 1alpha inhibits the fibroblast-like markers type I and type III collagen during hypoxia-induced chondrocyte redifferentiation: hypoxia not only induces type II collagen and aggrecan, but it also inhibits type I and type III collagen in the hypoxia-inducible factor 1alpha-dependent redifferentiation of chondrocytes. *Arthritis Rheum* 2009;**60**:3038-48.

61 Saito T, Fukai A, Mabuchi A, Ikeda T, Yano F, Ohba S, *et al.* Transcriptional regulation of endochondral ossification by HIF-2alpha during skeletal growth and osteoarthritis development. *Nat Med* 2010;**16**:678-86.

62 Yang S, Kim J, Ryu JH, Oh H, Chun CH, Kim BJ, *et al.* Hypoxia-inducible factor-2alpha is a catabolic regulator of osteoarthritic cartilage destruction. *Nat Med* 2010;**16**:687-93.

63 Chou CH, Wu CC, Song IW, Chuang HP, Lu LS, Chang JH, *et al.* Genome-wide expression profiles of subchondral bone in osteoarthritis. *Arthritis Res Ther* 2013;**15**:R190.

64 Suri S, Gill SE, Massena de CS, Wilson D,

McWilliams DF, Walsh DA. Neurovascular invasion at the osteochondral junction and in osteophytes in osteoarthritis. *Ann Rheum Dis* 2007;**66**:1423-8.

65 Howard HC, Mount DB, Rochefort D, Byun N, Dupre N, Lu J, *et al.* The K-Cl cotransporter KCC3 is mutant in a severe peripheral neuropathy associated with agenesis of the corpus callosum. *Nat Genet* 2002;**32**:384-92.

66 Sun YT, Lin TS, Tzeng SF, Delpire E, Shen MR. Deficiency of electroneutral K+-Cl-cotransporter 3 causes a disruption in impulse propagation along peripheral nerves. *Glia* 2010;**58**:1544-52.

67 Grogan SP, D'Lima DD. Joint aging and chondrocyte cell death. *Int J Clin Rheumtol* 2010;**5**:199-214.

68 Castano Betancourt MC, Cailotto F, Kerkhof HJ, Cornelis FM, Doherty SA, Hart DJ, *et al.* Genome-wide association and functional studies identify the DOT1L gene to be involved in cartilage thickness and hip osteoarthritis. *Proc Natl Acad Sci U S A* 2012;**109**:8218-23.

69 Reynard LN, Bui C, Syddall CM, Loughlin J. CpG methylation regulates allelic expression of GDF5 by modulating binding of SP1 and SP3 repressor proteins to the osteoarthritis susceptibility SNP rs143383. *Hum Genet* 2014.

70 Sunyaev SR. Inferring causality and functional significance of human coding DNA variants. *Hum Mol Genet* 2012;**21**:R10-R17.

71 Bomer N, den HW, Ramos YF, Bos SD, van der Breggen R, Lakenberg N, *et al.* Underlying molecular mechanisms of DIO2 susceptibility in symptomatic osteoarthritis. *Ann Rheum Dis* 2014.

72 Syddall CM, Reynard LN, Young DA, Loughlin J. The identification of trans-acting factors that regulate the expression of GDF5 via the osteoarthritis susceptibility SNP rs143383. *PLoS Genet* 2013;**9**:e1003557.

73 Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011;**12**:745-55.

74 Dodd AW, Syddall CM, Loughlin J. A rare variant in the osteoarthritis-associated locus GDF5 is functional and reveals a site that can be manipulated to modulate GDF5 expression. *Eur J Hum Genet* 2013;**21**:517-21.

75 Tyer NM, Kim TY, Martinez DS. Review on Oral Cholecystographic Agents for the Management of Hyperthyroidism. *Endocr Pract* 2014;1-30.

76 Diederichs S, Tuan RS. Functional comparison of human-induced pluripotent stem cell-derived mesenchymal cells and bone marrow-derived mesenchymal stromal cells from the same donor. *Stem Cells Dev* 2014;**23**:1594-610.

77 Caron MM, Emans PJ, Coolsen MM, Voss L, Surtel DA, Cremers A, *et al.* Redifferentiation of dedifferentiated human articular chondrocytes: comparison of 2D and 3D cultures. *Osteoarthritis Cartilage* 2012;**20**:1170-8.

78 Fischer J, Dickhut A, Rickert M, Richter W. Human articular chondrocytes secrete parathyroid hormone-related protein and inhibit hypertrophy of mesenchymal stem cells in coculture during chondrogenesis. *Arthritis Rheum* 2010;**62**:2696-706.

79 Yang YH, Lee AJ, Barabino GA. Coculture-driven mesenchymal stem cell-differentiated articular chondrocyte-like cells support neocartilage development. *Stem Cells Transl Med* 2012;**1**:843-54.

80 Wu L, Prins HJ, Helder MN, van Blitterswijk CA, Karperien M. Trophic effects of mesenchymal stem cells in chondrocyte co-cultures are independent of culture conditions

and cell sources. *Tissue Eng Part A* 2012;**18**:1542-51.

81 Maumus M, Manferdini C, Toupet K, Peyrafitte JA, Ferreira R, Facchini A, *et al.* Adipose mesenchymal stem cells protect chondrocytes from degeneration associated with osteoarthritis. *Stem Cell Res* 2013;**11**:834-44.

82 Lozito TP, Alexander PG, Lin H, Gottardi R, Cheng AW, Tuan RS. Three-dimensional osteochondral microtissue to model pathogenesis of osteoarthritis. *Stem Cell Res Ther* 2013;**4 Suppl 1**:S6.

83 Lin H, Lozito TP, Alexander PG, Gottardi R, Tuan RS. Stem cell-based microphysiological osteochondral system to model tissue response to interleukin-1beta. *Mol Pharm* 2014;**11**:2203-12.

84 He P, Ng KS, Toh SL, Goh JC. In vitro ligament-bone interface regeneration using a trilineage coculture system on a hybrid silk scaffold. *Biomacromolecules* 2012;**13**:2692-703.

85 Leyh M, Seitz A, Durselen L, Springorum HR, Angele P, Ignatius A, *et al.* Osteoarthritic cartilage explants affect extracellular matrix production and composition in cocultured bone marrow-derived mesenchymal stem cells and articular chondrocytes. *Stem Cell Res Ther* 2014;**5**:77.

86 Wuelling M, Vortkamp A. Cartilage explant cultures. *Methods Mol Biol* 2014;**1130**:89-97.

87 Mobasheri A, Kelly S, Clutterbuck AL, Buhrmann C, Shakibaei M. Three-dimensional, high-density and tissue engineered culture models of articular cartilage. In: Davies J, ed. *Replacing animal models: a practical guide to creating and using culture-based biomimetic alternatives*. 1 ed. John Wiley & Sons, Ltd. 2012:167-92.

88 Lories RJ, Peeters J, Bakker A, Tylzanowski P, Derese I, Schrooten J, *et al.* Articular cartilage and biomechanical properties of the long bones in Frzb-knockout mice. *Arthritis Rheum* 2007;**56**:4095-103.

89 Lories RJ, Peeters J, Szlufcik K, Hespel P, Luyten FP. Deletion of frizzled-related protein reduces voluntary running exercise performance in mice. *Osteoarthritis Cartilage* 2009;**17**:390-6.

90 Cornelis FM, Luyten FP, Lories RJ. Functional effects of susceptibility genes in osteoarthritis. *Discov Med* 2011;**12**:129-39.

91 Schneider MJ, Fiering SN, Pallud SE, Parlow AF, St Germain DL, Galton VA. Targeted disruption of the type 2 selenodeiodinase gene (DIO2) results in a phenotype of pituitary resistance to T4. *Mol Endocrinol* 2001;**15**:2137-48.

92 Nagase H, Nagasawa Y, Tachida Y, Sakakibara S, Okutsu J, Suematsu N, *et al.* Deiodinase 2 upregulation demonstrated in osteoarthritis patients cartilage causes cartilage destruction in tissue-specific transgenic rats. *Osteoarthritis Cartilage* 2013;**21**:514-23.

93 Glasson SS, Askew R, Sheppard B, Carito B, Blanchet T, Ma HL, *et al.* Deletion of active ADAMTS5 prevents cartilage degradation in a murine model of osteoarthritis. *Nature* 2005;**434**:644-8.

94 Zhen G, Wen C, Jia X, Li Y, Crane JL, Mears SC, *et al.* Inhibition of TGF-beta signaling in mesenchymal stem cells of subchondral bone attenuates osteoarthritis. *Nat Med* 2013;**19**:704-12.

95 Blaney Davidson EN, Vitters EL, Bennink MB, van Lent PL, van Caam AP, Blom AB, *et al.* Inducible chondrocyte-specific overexpression of BMP2 in young mice results in severe aggravation of osteophyte formation in experimental OA without altering cartilage damage. *Ann Rheum Dis* 2014.

96 Zhang Y, Vasheghani F, Li YH, Blati M, Simeone K, Fahmi H, *et al.* Cartilage-specific deletion of mTOR upregulates autophagy and

protects mice from osteoarthritis. *Ann Rheum Dis* 2014.

97 Altman FP. A metabolic dysfunction in early murine osteoarthritis. *Ann Rheum Dis* 1981;**40**:303-6.

98 Greenblatt MB, Ritter SY, Wright J, Tsang K, Hu D, Glimcher LH, *et al.* NFATc1 and NFATc2 repress spontaneous osteoarthritis. *Proc Natl Acad Sci U S A* 2013;**110**:19914-9.

99 Vincent TL, Williams RO, Maciewicz R, Silman A, Garside P. Mapping pathogenesis of arthritis through small animal models. *Rheumatology (Oxford)* 2012;**51**:1931-41.

100 Little CB, Hunter DJ. Post-traumatic osteoarthritis: from mouse models to clinical trials. *Nat Rev Rheumatol* 2013;**9**:485-97.

101 Fang H, Beier F. Mouse models of osteoarthritis: modelling risk factors and assessing outcomes. *Nat Rev Rheumatol* 2014;**10**:413-21.

102 Aigner T, Cook JL, Gerwin N, Glasson SS, Laverty S, Little CB, *et al.* Histopathology atlas of animal model systems - overview of guiding principles. *Osteoarthritis Cartilage* 2010;**18 Suppl 3**:S2-S6.

103 Kerkhof HJ, Meulenbelt I, Akune T, Arden NK, Aromaa A, Bierma-Zeinstra SM, *et al.* Recommendations for standardization and phenotype definitions in genetic studies of osteoarthritis: the TREAT-OA consortium. *Osteoarthritis Cartilage* 2011;**19**:254-64.

Appendix

A

SUMMARY

Introduction

Osteoarthritis, also referred to in short as OA, is a disorder of the joints characterized by pain, crepitus, and stiffness resulting in decreased mobility. Cause is degeneration of the cartilage and/or the development of protrusions of the bone in medical terms named osteophytes. Ongoing disease can also be accompanied by inflammation. Prevalence of OA rapidly increases with increasing age. Due to a growing elderly population currently OA is ranking 11th of 291 diseases causing lack of mobility among elderly.

OA can occur in any joint, however, it is most common in hands, hips, and knees. Development can be triggered by several factors such as joint injury, suboptimal metabolism (e.g. obesity or a particular dietary pattern), and sex. Also genetic predisposition plays an important role in development of OA. It can determine which joint becomes affected, what are the roots for its development (defective cartilage or bone), and how is the shape of the joint.

At present, treatment options for OA are still limited. Reason is the relatively limited knowledge of its etiology and that it is not diagnosed until at late stages, in addition to the high heterogeneity among OA patients. Medication is mainly directed at pain relief until joint arthroplasty at end stage is required.

Most important questions of this thesis

In the course of this PhD we aimed to improve research related to OA at 3 points. Firstly, with respect to the application of biomarkers. To gain insight into OA pathophysiological processes, markers in cartilage and blood were studied. Also, it was investigated whether genetic variants exist that influence levels of known biochemical markers possibly resulting in the perturbation of their application in the clinic. Secondly, it was investigated which variants in the human genome are associated with susceptibility to OA. Therefore, a genome wide association study was performed with a group of patients suffering from familiar and severe OA at multiple joint sides and the genetic make-up (exome sequencing) of 2 distant family members with severe, early-onset, OA was determined. Thirdly, by combining genetic and functional studies we tried to come to deeper insight into mechanisms underlying susceptibility to OA. Filling this gap will, in due time, facilitate translation of scientific studies to the application in the clinic.

In order to perform their research, scientists depend on patients that want to participate in a study. Research described in this thesis was mainly based on 2 groups of patients. In collaboration with the Department of Rheumatology we previously included patients for the 'Genetics osteoARthritis and Progression' or GARP study. This group exists in 191 sib pairs suffering from symptomatic and radiographic OA in 2 or more joints. The other group is the ongoing RAAK study (Research Artrose en Articulair Kraakbeen') consisting in patients that underwent joint arthroplasty. Within the RAAK study a wealth of cells and tissues, such as cartilage, bone, and blood, have been collected. Besides the GARP and the RAAK study, within Europe but also outside of Europe scientists collaborate in large consortia such as

TREAT~OA and arcOGEN, that have the availability of information on thousands of OA patients which can be analyzed together. Such collaborations are required for epidemiological research.

Biomarkers in OA

Biomarkers are biological variables which can be quantified and used to indicate existing biological circumstances or to predict occurrence of pathological conditions. Good biomarkers have to be sensitive as well as specific for a certain disorder. With respect to OA biomarkers are required for early diagnosis, ideally even before damage becomes visible with x-ray, to monitor disease progression, and for prognosis. Available are traditional biochemical markers that reflect joint damage. For example, cartilage matrix molecules and fragments thereof released and measurable in blood serum (e.g. serum COMP or sCOMP) or urine (e.g. uCTX2). Unfortunately, sensitivity and specificity of these markers are relatively low hampering their application in the clinic. This urges further research. Recently it has been studied whether joint shape can serve as a marker to predict development of OA. In addition, currently molecular biomarkers such as DNA and RNA in tissues and body fluids received a lot of attention.

By comparing gene expression in OA affected cartilage with that in macroscopically 'normal' cartilage of the same joint in the RAAK study, it was investigated in **chapter 2** which genes are changed during ongoing OA. Although these changes can be cause as well as consequence of the arthritic processes, it gives insight into the different pathways taking place in chondrocytes during development of OA. Results showed that especially genes involved in skeletal development were changed. This suggests that processes normally taking place during growth to become inactive in adulthood, become activate again in OA. In addition, also several genes involved in neurogenesis were identified. Thus, creating a potential link to nociception. Development of treatment options incurring at these mechanisms could result in more efficient therapy of OA.

In **chapter 3** a gene expression profile was generated in blood of OA patients of the GARP study. It was found that in particular genes involved in cell death (apoptosis) were different in blood of patients when compared to healthy controls. A link between OA and apoptosis has been described before in chondrocytes, however, this finding in blood is remarkable. Besides the specific arthritic profile it was also found that, based on expression of not more than 4 genes, OA patients could be distinguished very well from healthy subjects (~90% certainty). It is important now to investigate whether the genes identified here can also be applied to other groups of OA patients, and whether the genes differ already at an early stage and may thus contribute to early diagnosis.

Besides interest in the (relatively new) molecular markers such as gene expression that was studied in **chapter 2** and **3**, there is still interest in the more traditional biochemical markers. Perturbation of their application in the clinic could be the existence of particular genetic factors that, independent of OA, affect innate levels. This was investigated in **chapter 4** for sCOMP and uCTX2. Meta-analysis of many patients of the TREAT~OA consortium identified genetic variants influencing variation of sCOMP and uCTX2 in serum and urine, respectively. Taking advantage of this knowledge in the clinic may improve the use of

biochemical markers in OA. However, is cannot be excluded that also other factors interfere, such as differential clearance of degradation products of cartilage and bone in kidney.

Genetic variation underlying OA

It has been known for many years now that OA has a considerable genetic component. Based on family and twin studies, heritability was estimated between 40 and 60 percent. Genetic studies identified a number of the susceptibility genes such as *DIO2*, *GDF5*, and *NCOA3*, however, these seem to contribute very little to the total percentage of heritability. In part, this could be caused by overestimation in the family studies. Also, heterogeneity among OA patients may hamper the identification of genes underlying susceptibility to OA and result in discovery of variants with small effect sizes. To find genes with moderate effects, a genome wide association study was performed with the GARP study. Unfortunately, the study appeared too small to come to genome wide significant signals, and too specific to replicate findings in other cohorts. However, the analysis provided with several interesting variants. For example, *CDKAL1* was identified, previously found in association with diabetes (a risk factor for progression of OA), and expression of *CDKAL1* is different in blood of OA patients when compared to healthy controls.

Another way to determine which genetic factors contribute to development of OA is by applying exome sequencing to highly affected family members. This is described in **chapter 6**. Although the variant that is detected will most likely be particular to the family involved, it is established that results point towards common genetic pathways underlying OA in the general population. Our analysis resulted in the identification of a specific genetic variant highly likely to be causal in the family. It was also observed that expression of the gene, *TNFRSF11B*, was highly abundant in OA affected cartilage of the RAAK study. The fact that the identified variant of *TNFRSF11B* results in increased bone formation is very important with respect to the current discussion whether OA patients may benefit from strontium ranelate, a medicine that enhances bone formation. Our study suggests that for certain patients this treatment could be highly disadvantageous, and that before establishing treatment options it is very important to classify the patient (occurrence of increased bone formation or not).

A bridge from laboratory to clinic

In the course of the studies described here we have tried to go beyond the results of molecular epidemiological studies to increase insights into underlying mechanisms. This will facilitate translation of knowledge of genetic variants to the needs of OA patients and thus to application in the clinic. Therefore, in **chapter 6** and particularly in **chapter 7**, model systems were developed to investigate the effect of genetic aberrations on extracellular matrix deposition of cartilage or bone. Results of the studies performed here underline the causal role of the genetic variants in OA and show the importance of combining genetic and functional studies. Such an approach can result in better selection of medical treatments of OA: following the identification of the *DIO2* gene as a risk factor for OA, the effect on cartilage formation was investigated. It appeared that DIO2 negatively affects genes involved in cartilage matrix deposition, while increasing the expression of genes involved in degradation of the matrix.

DIO2 is an enzyme that converts inactive thyroid hormone to active thyroid hormone. Inhibition of this conversion by addition of the drug iopanoic acid or IOP to our model system showed a structural improvement of the cartilage matrix.

Conclusion and recommendation

Studies described in this thesis suggest that the application of known biochemical markers in OA may possibly improve by taking genetic variation of the patients into account. Moreover, additional research for the application of molecular markers to diagnose, monitor, and possibly also for the prognosis of OA seem worthwhile. For the first time, it was shown that, using an easily accessible tissue such as blood, it is possible to distinguish OA patients from healthy controls. This is encouraging and stimulates subsequent research for the application of molecular markers for diagnosis, monitoring, and prognosis of OA.

Our studies show the importance of patient classification before starting a specific treatment. Wrong combinations of respectively the needs of the patient and the underlying mechanism may result in acceleration of the pathophysiological process leading to more total joint replacement surgeries.

Studies described in this thesis show how functional genomics can be achieved by combining genomic and functional studies. Model systems were developed for the investigation of the mechanisms underlying susceptibility to OA via genetic variants. In addition, systems serve to determine which drugs potentially inhibit these mechanisms. Taking into account that OA is a disease of the whole joint, I suggest to invest in development of complex, multi-lineage, co-culture systems that can also address the interaction between the different tissues in a joint. In addition, this could also be of importance to determine putative controversial effects of drugs on different types of tissues.

SUMMARY IN DUTCH

Inleiding

Osteoartrose, kortweg ook wel artrose genoemd, is een aandoening van de gewrichten die gekenmerkt wordt door pijn, kraken van het aangedane gewricht (crepitus), en stijfheid, met als gevolg verminderde mobiliteit. Oorzaak is de afbraak van kraakbeen en/of het ontstaan van botuitstulpingen bij de gewrichten aangeduid met de medische term 'osteofieten'. Bovendien gaat het ziekteverloop gepaard met ontstekingsverschijnselen. De kans op artrose neemt snel toe bij het ouder worden. Door de toenemende vergrijzing staat artrose nu al op de 11de plaats van 291 aandoeningen die onderzocht zijn als oorzaak voor immobiliteit bij ouderen.

Artrose kan zich voordoen in vrijwel elk gewricht, maar komt voornamelijk voor in hand-, knie- en heupgewrichten. Hiervoor zijn verschillende oorzaken te benoemen zoals (sport)blessures, een suboptimale stofwisseling (bijvoorbeeld door obesitas of een bepaald voedingspatroon) en geslacht. Ook erfelijke aanleg beïnvloedt het risico op de ontwikkeling van artrose. Deze kan beïnvloeden welke gewrichten gevoeliger zijn, waar de basis ligt voor het ontstaan (zwakker kraakbeen of juist zwakker bot), en welke vorm het gewricht heeft.

Tot op heden zijn de behandelmogelijkheden voor artrose zeer beperkt. Dit komt doordat er nog relatief weinig bekend is over de oorzaken, diagnose pas in een gevorderd stadium gesteld wordt, en er onder patiënten grote verscheidenheid voorkomt in het ziektebeeld. Medicatie is voornamelijk gericht op pijnvermindering tot tenslotte een gewrichtsvervangende operatie moet plaatsvinden.

Belangrijkste vragen die in dit proefschrift gesteld zijn

In het kader van dit promotieonderzoek is getracht het onderzoek naar artrose op drie punten vooruit te brengen. Ten eerste betreft dit de toepassing van biologische markers (biomarkers) voor artrose. Er is gezocht naar bepaalde markers in kraakbeen en bloed om een beter inzicht te krijgen in het ziekteproces en naar mogelijke genetische factoren die de toepassing van bekende biomarkers negatief beïnvloeden. Ten tweede is onderzocht welke variatie in het menselijke genoom, al het DNA in een cel, de kans op artrose verhoogt. Hiertoe is een genoomwijde associatieanalyse gedaan bij een groep van patiënten met familiaire en symptomatische artrose in meerdere gewrichten. Daarnaast is de erfelijke code bepaald van twee verre familieleden met ernstige artrose beginnend op jonge leeftijd door middel van de zogeheten exome sequencing techniek. Ten derde is er gewerkt aan het slaan van een brug tussen genetische en functionele studies om te komen tot kennis van de mechanismen die ten grondslag liggen aan het ontstaan van artrose. Dit vergemakkelijkt de translatie van bevindingen in het laboratorium naar toepassingen in de kliniek.

Voor onderzoek naar ziekten zijn wetenschappers altijd afhankelijk van patiënten die bereid zijn deel te nemen aan hun studies. Onderzoeken, beschreven in dit proefschrift, zijn gebaseerd op twee groepen patiënten die hieraan hun deelneming verleenden. Gegevens van de 'Genetics osteoARthritis and Progression' of GARP studie waren voorheen bijeengebracht op onze afdeling in samenwerking met de afdeling Reumatologie van het Leids Universitair

Medisch Centrum. Deze groep bestaat uit 191 broer-zus (zus-zus) paren die lijden aan relatief ernstige symptomatische en radiologische artrose in meerdere gewrichten. De andere groep is de lopende RAAK studie ('Research Artrose en Articulair Kraakbeen') welke bestaat uit patiënten die een gewrichtsvervangende operatie hebben ondergaan. In het kader van deze studie is een grote verscheidenheid aan cellen en weefsels verzameld, zoals kraakbeen, bot, en bloed. Behalve de GARP en de RAAK studies wordt voor het genetische onderzoek naar artrose binnen Europa, maar ook daarbuiten, samengewerkt in grote verbanden zoals de TREAT~OA en arcOGEN consortia, waarbij gegevens van duizenden patiënten beschikbaar zijn en gezamenlijk worden geanalyseerd. Voor epidemiologisch onderzoek zijn zulke samenwerkingen een noodzaak.

Biomarkers in artrose

Biomarkers zijn meetbare biologische variabelen die gebruikt kunnen worden als indicator van een bepaalde bestaande biologische situatie of die het optreden van een aandoening kunnen voorspellen. Een goede biomarker moet zowel gevoelig als specifiek zijn voor een bepaalde aandoening. Met betrekking tot artrose zijn biomarkers nodig voor het stellen van een vroegtijdige diagnose, idealiter nog voordat de schade zichtbaar wordt op een röntgenfoto, voor het monitoren van het ziekteproces, en voor prognose. Beschikbaar zijn de meer traditionele biochemische markers die gewrichtsschade reflecteren. Voorbeelden zijn restproducten van afgebroken kraakbeen welke in bloedserum (bv. serum COMP of sCOMP) of in urine (bv. uCTX2) gemeten kunnen worden. Helaas zijn de sensitiviteit en de specificiteit hiervan relatief laag wat de klinische toepassing ervan belemmert en uitnodigt tot het verrichten van verder onderzoek. Zo is recentelijk bijvoorbeeld bekeken in hoeverre informatie over de vorm van een gewricht kan dienen als biomarker om te voorspellen of iemand binnen bepaalde termijn artrose zal ontwikkelen. Daarnaast is er op dit moment veel belangstelling voor nieuwe moleculaire biomarkers, zoals DNA en RNA in weefsels en lichaamsvloeistoffen.

Door de expressie van genen in artrotisch kraakbeen te vergelijken met die in macroscopisch 'normaal' kraakbeen van hetzelfde gewricht in de RAAK studie is in **hoofdstuk 2** onderzocht welke veranderingen plaatsvinden gedurende het artroseproces. Hoewel deze veranderingen zowel oorzaak als gevolg kunnen zijn, geeft het inzicht in de verschillende wegen die een kraakbeencel doorloopt tijdens het ziekteproces. De resultaten lieten zien dat met name genen die betrokken zijn bij skeletontwikkeling veranderd zijn. Dit suggereert dat processen die normaal gesproken alleen plaatsvinden tijdens de vroege ontwikkeling om daarna tot stilstand te komen, bij artrose weer actief zijn. Tevens vonden we een aantal genen die betrokken zijn bij neurogenese en dus mogelijk bij pijnperceptie. Het ontwikkelen van behandelingen die op deze mechanismen aangrijpen zouden een nieuw aanknopingspunt kunnen zijn om te komen tot een efficiëntere behandeling van artrose.

In **hoofdstuk 3** is een profiel gemaakt van genexpressie in bloed van artrosepatiënten van de GARP studie. Uit onze studie bleek dat met name genen die betrokken zijn bij celdood (apoptose), verschilden in bloed van patiënten en gezonde mensen. De link tussen artrose en apoptose was al eerder gevonden in kraakbeencellen, maar het is opmerkelijk dat dit ook in bloed van patiënten naar voren is gekomen. Naast het specifieke artroseprofiel werd ontdekt dat op basis van de expressie van slechts vier verschillende genen, artrosepatiënten zeer goed

van gezonde mensen konden worden onderscheiden (met ~90% zekerheid). Het is nu zeer belangrijk te onderzoeken of deze bevindingen ook in andere groepen artrosepatiënten bevestigd kunnen worden om te zien of het genexpressie profiel met vergelijkbare efficiëntie deze patiënten van gezonde personen kan onderscheiden. Vervolgens moet onderzocht worden of deze genen al in een vroeg stadium anders zijn en dus zouden kunnen bijdragen aan een vroegtijdige diagnose van de ziekte.

Naast interesse in (relatief nieuwe) moleculaire markers zoals genexpressie welke in **hoofdstuk 2** en **3** werden onderzocht, is er nog steeds belangstelling voor de meer traditionele biochemische markers. Een mogelijke beperking bij de klinische toepassing hiervan is het feit dat erfelijke, aangeboren verschillen tussen mensen de meetwaardes beïnvloeden, onafhankelijk van artrose. Deze optie is onderzocht in **hoofdstuk 4** voor sCOMP en uCTX2. De gezamenlijke analyse van een groot aantal mensen van TREAT~OA heeft aangetoond dat bepaalde genetische factoren invloed hebben op de variatie van sCOMP en uCTX2 in respectievelijk serum en urine. Hoewel de effecten van erfelijkheid op de meetwaardes klein lijken te zijn zou het benutten van deze wetenschap toepassing van genoemde biomarkers in de kliniek mogelijk kunnen verbeteren. Het mag echter niet uitgesloten worden dat ook andere factoren van invloed kunnen zijn zoals bijvoorbeeld de snelheid waarmee in de nieren afbraakproducten van kraakbeen en bot uit het bloed worden gefilterd.

Erfelijke factoren betrokken bij artrose

Reeds vele jaren geleden is er op basis van tweeling- en familiestudies vastgesteld dat 40-60 procent van het risico om artrose te ontwikkelen door erfelijkheid wordt bepaald. Genetische studies hebben al een aantal van de risicogenen geïdentificeerd zoals *DIO2*, *GDF5*, en *NCOA3*, maar deze lijken slechts in geringe mate bij te dragen aan het totale erfelijkheidspercentage. Voor een deel zou er sprake kunnen zijn van een overschatting in de eerdere studies, en voor een ander deel wordt het veroorzaakt door heterogeniteit onder de patiënten. Dit laatste bemoeilijkt namelijk de identificatie van risicogenen en resulteert vooral in factoren die een klein effect hebben in de populatie. Om genen te vinden die in een wat grotere mate bijdragen aan artrose is in **hoofdstuk 5** een genoomwijde associatiestudie uitgevoerd met de GARP studie. De studie is echter te klein gebleken om te komen tot genoomwijd significante resultaten en te specifiek om de resultaten in andere studies te kunnen repliceren. Het heeft daarentegen wel een aantal mogelijke risicogenen aangeduid zoals bijvoorbeeld het *CDKAL1* gen welke eerder ontdekt was in associatie met diabetes (een risicofactor voor progressie van artrose) en waarvan de expressie in bloed van artrosepatiënten anders blijkt te zijn dan in bloed van gezonde mensen.

Een andere manier om genetische factoren te vinden die betrokken zijn bij de aanleg tot het ontwikkelen van artrose is met behulp van exome sequencing van familieleden met zeer ernstige artrose. Dit is beschreven in **hoofdstuk 6**. Hoewel het waarschijnlijk is dat op deze wijze afwijkingen worden gevonden die specifiek zijn voor de betrokken familie wees eerder gedaan onderzoek erop dat het wel bepaalde routes kan aanduiden die ook in de doorsnee-populatie tot artrose leiden. De analyse heeft geresulteerd in de identificatie van een specifieke afwijking in het erfelijke materiaal van de familie, die met grote waarschijnlijkheid de

familieleden gevoelig maakt voor artrose. Het betrokken gen, *TNFRSF11B*, bleek ook een sterk verhoogde expressie te hebben in artrotisch kraakbeen in de RAAK studie. Het feit dat het betreffende gen een rol speelt in de botbalans en de afwijking resulteert in verhoogde botaanmaak is zeer belangrijk voor de huidige discussie betreffende behandeling van artrosepatiënten met medicijnen die de botaanmaak nog meer zouden stimuleren. Ons onderzoek laat zien dat zulk een behandeling nadelig kan zijn voor bepaalde patiënten en dat het vaststellen van het type artrose (al dan niet met verhoogde botaanmaak) vóór het bepalen van de behandelwijze erg belangrijk is.

Een brug van laboratorium naar kliniek

In de loop van het beschreven onderzoek is geprobeerd niet te blijven staan bij de moleculaire epidemiologie maar om te komen tot meer inzicht in onderliggende mechanismen. Dit vergemakkelijkt de 'vertaling' van kennis van een erfelijke variant naar de behoefte van de artrosepatiënt en dus naar het vinden van praktische toepassingen en behandelingen in de kliniek. In **hoofdstuk 6** en met name in **hoofdstuk 7** zijn hiertoe modelsystemen opgezet om de effecten van bepaalde genetische afwijkingen op kraakbeen- en botvorming te onderzoeken. Resultaten van deze studies onderstrepen dat gevonden genvarianten betrokken zijn bij de ontwikkeling van artrose en laten het belang zien van de combinatie van genetische en functionele studies. Een dergelijke benadering kan selectie van geschikte medicijnen voor de behandeling van artrose verbeteren: nadat een erfelijke variatie in het gen *DIO2* gevonden was als risicofactor voor het ontwikkelen van artrose, is bestudeerd wat het effect hiervan was op de vorming van kraakbeen. De capaciteit om matrix te maken bleek verlaagd, terwijl stoffen die het kraakbeen afbreken juist verhoogd waren. DIO2 is een enzym dat schildklierhormoon activeert. Remming van deze activatie door het toevoegen van de drug IOP (iopanoic acid) in het modelsysteem lijkt de structuur van het kraakbeen dat is geproduceerd te verbeteren.

Conclusie en aanbeveling

De toepassing van bekende biochemische markers in artrose kan mogelijk verbeterd worden door rekening te houden met bepaalde genetische variatie in patiënten. Verder zijn de resultaten met moleculaire markers zeer veelbelovend. Voor het eerst is hiermee aangetoond dat een makkelijk af te nemen weefsel zoals bloed gebruikt kan worden om artrosepatiënten te onderscheiden van gezonde personen. Dit stimuleert tot verder onderzoek naar het gebruik van moleculaire markers voor diagnose, en mogelijk ook voor het volgen van pathologische processen in de tijd en prognose van artrose.

Onze studies tonen aan dat het van groot belang is om patiënten beter te karakteriseren vóórdat de behandeling wordt vastgesteld. Verkeerde combinatie van respectievelijk de medische behoefte van de patiënt en werkingsmechanisme van de medicatie kan leiden tot een versneld ziekteproces met als gevolg méér gewrichtsvervangende operaties.

Studies beschreven in dit proefschrift laten zien hoe geprobeerd is te komen tot 'functional genomics': de combinatie van genomische en functionele studies. Er zijn modelsystemen opgezet om te onderzoeken wat de onderliggende mechanismen zijn waarop erfelijke varianten resulteren in aanleg voor artrose en welke medicijnen dit eventueel zouden kunnen remmen. De modellen kunnen ook in de toekomst gebruikt worden om bevindingen van moleculair

epidemiologische studies te verifiëren. Wetende dat artrose een ziekte is van het hele gewricht, is het sterk aanbevelenswaardig om, ter verbetering van de modelsystemen, complexe, meercellige systemen te ontwikkelen, die de interactie tussen de verschillende weefsels van het gewricht in het onderzoek kunnen betrekken. Dit is ook voor het selecteren van medicatie van belang, ter voorkoming van tegengestelde effecten die deze kan hebben op verschillende weefsels van het gewricht.

Curriculum Vitae

Yolande Francisca Maria Ramos was born in 1970 on February 8 in Eindhoven, the Netherlands. She obtained her Bachelors degree at the University of Utrecht in 1991 and her Master's degree with specialization tumor biology and neuropharmacology at the Free University of Amsterdam in 1995. After working for a year at the Dept. of Pathology of the Harvard Medical School, Boston, USA, in the laboratory of Prof. dr. Tsai, she started her first PhD in 1996 at the Dept. of Molecular Cell Biology, section Molecular Carcinogenesis with Prof. dr. van der Eb at the University of Leiden, in the laboratory of dr. Jochemsen. Research performed in this period resulted in 2001 in the thesis entitled 'Analysis of MDMX: functions and expression pattern'.

After several years of research as a post doctoral collaborator of dr. Zantema and dr. van Dam at the Dept. of Molecular Cell Biology while specializing in transcription regulation, the author of this thesis decided to perform a second PhD at the Dept. of Medical Statistics, section Molecular Epidemiology with Prof. dr. Slagboom at the Leiden University Medical Center, in the laboratory of dr. Meulenbelt.

Currently, dr. Ramos continues her research as a post doctoral scientist in the laboratory of dr. Meulenbelt. She has a strong interest in translating genetic and molecular biologic studies towards applications in the clinic.

LIST OF PUBLICATIONS

Y. F. Ramos, S. D. Bos, R. van der Breggen, M. Kloppenburg, K. Ye, E. W. Lameijer, R. G. Nelissen, P. E. Slagboom, I. Meulenbelt, *Ann. Rheum. Dis.* **2014**.

Y. F. Ramos*, W. den Hollander*, J. V. Bovee, N. Bomer, R. van der Breggen, N. Lakenberg, J. C. Keurentjes, J. J. Goeman, P. E. Slagboom, R. G. Nelissen, S. D. Bos, I. Meulenbelt, *PLoS. One.* **2014**, *9*, e103056.

Y. F. Ramos*, S. Metrustry*, N. Arden, A. C. Bay-Jensen, M. Beekman, A. J. de Craen, L. A. Cupples, T. Esko, E. Evangelou, D. T. Felson, D. J. Hart, J. P. Ioannidis, M. Karsdal, M. Kloppenburg, F. Lafeber, A. Metspalu, K. Panoutsopoulou, P. E. Slagboom, T. D. Spector, E. W. van Spil, A. G. Uitterlinden, Y. Zhu, A. M. Valdes, J. B. van Meurs, I. Meulenbelt, *J. Med. Genet.* **2014**, *51*, 596.

N. Bomer*, W. den Hollander*, **Y. F. Ramos***, S. D. Bos, R. van der Breggen, N. Lakenberg, B. A. Pepers, A. E. van Eeden, A. Darvishan, E. W. Tobi, B. J. Duijnisveld, E. B. van den Akker, B. T. Heijmans, W. M. van Roon-Mom, F. J. Verbeek, G. J. van Osch, R. G. Nelissen, P. E. Slagboom, I. Meulenbelt, *Ann. Rheum. Dis.* **2014**.

N. Bomer, F. M. Cornelis, **Y. F. Ramos**, W. den Hollander, L. Storms, R. van der Breggen, N. Lakenberg, P. E. Slagboom, I. Meulenbelt, R. J. Lories, *Ann. Rheum. Dis.* **2014**.

W. den Hollander, **Y. F. Ramos**, S. D. Bos, N. Bomer, R. van der Breggen, N. Lakenberg, W. J. de Dijcker, B. J. Duijnisveld, P. E. Slagboom, R. G. Nelissen, I. Meulenbelt, *Ann. Rheum. Dis.* **2014**, *73*, 2208.

Y. F. Ramos, S. D. Bos, N. Lakenberg, S. Bohringer, W. J. den Hollander, M. Kloppenburg, P. E. Slagboom, I. Meulenbelt, *Ann. Rheum. Dis.* **2014**, *73*, 1844.

E. Evangelou, H. J. Kerkhof, U. Styrkarsdottir, E. E. Ntzani, S. D. Bos, T. Esko, D. S. Evans, S. Metrustry, K. Panoutsopoulou, **Y. F. Ramos**, G. Thorleifsson, K. K. Tsilidis, N. Arden, N. Aslam, N. Bellamy, F. Birrell, F. J. Blanco, A. Carr, K. Chapman, A. G. Day-Williams, P. Deloukas, M. Doherty, G. Engstrom, H. T. Helgadottir, A. Hofman, T. Ingvarsson, H. Jonsson, A. Keis, J. C. Keurentjes, M. Kloppenburg, P. A. Lind, A. McCaskie, N. G. Martin, L. Milani, G. W. Montgomery, R. G. Nelissen, M. C. Nevitt, P. M. Nilsson, W. E. Ollier, N. Parimi, A. Rai, S. H. Ralston, M. R. Reed, J. A. Riancho, F. Rivadeneira, C. Rodriguez-Fontenla, L. Southam, U. Thorsteinsdottir, A. Tsezou, G. A. Wallis, J. M. Wilkinson, A. Gonzalez, N. E. Lane, L. S. Lohmander, J. Loughlin, A. Metspalu, A. G. Uitterlinden, I. Jonsdottir, K. Stefansson, P. E. Slagboom, E. Zeggini, I. Meulenbelt, J. P. Ioannidis, T. D. Spector, J. B. van Meurs, A. M. Valdes, *Ann. Rheum. Dis.* **2014**, *73*, 2130.

C. Rodriguez-Fontenla, M. Calaza, E. Evangelou, A. M. Valdes, N. Arden, F. J. Blanco, A. Carr, K. Chapman, P. Deloukas, M. Doherty, T. Esko, C. M. Garces Aleta, J. J. Gomez-Reino Carnota, H. Helgadottir, A. Hofman, I. Jonsdottir, H. J. Kerkhof, M. Kloppenburg, A. McCaskie, E. E. Ntzani, W. E. Ollier, N. Oreiro, K. Panoutsopoulou, S. H. Ralston, **Y. F. Ramos**, J. A. Riancho, F. Rivadeneira, P. E. Slagboom, U. Styrkarsdottir, U.

Thorsteinsdottir, G. Thorleifsson, A. Tsezou, A. G. Uitterlinden, G. A. Wallis, J. M. Wilkinson, G. Zhai, Y. Zhu, D. T. Felson, J. P. Ioannidis, J. Loughlin, A. Metspalu, I. Meulenbelt, K. Stefansson, J. B. van Meurs, E. Zeggini, T. D. Spector, A. Gonzalez, *Arthritis Rheumatol.* **2014**, *66*, 940.

D. Bos, J. V. Bovee, B. J. Duijnisveld, E. V. Raine, W. J. van Dalen, **Y. F. Ramos**, R. van der Breggen, R. G. Nelissen, P. E. Slagboom, J. Loughlin, I. Meulenbelt, *Ann. Rheum. Dis.* **2012**, *71*, 1254.

N. A. Laurie, S. L. Donovan, C. S. Shih, J. Zhang, N. Mills, C. Fuller, A. Teunisse, S. Lam, **Y. F. Ramos**, A. Mohan, D. Johnson, M. Wilson, C. Rodriguez-Galindo, M. Quarto, S. Francoz, S. M. Mendrysa, R. K. Guy, J. C. Marine, A. G. Jochemsen, M. A. Dyer, *Nature* **2006**, *444*, 61.

P. de Graaf, N. A. Little, **Y. F. Ramos**, E. Meulmeester, S. J. Letteboer, A. G. Jochemsen, *J. Biol. Chem.* **2003**, *278*, 38315.

Y. F. Ramos, R. Stad, J. Attema, L. T. Peltenburg, A. J. van der Eb, A. G. Jochemsen, *Cancer Res.* **2001**, *61*, 1839.

R. Stad, **Y. F. Ramos**, N. Little, S. Grivell, J. Attema, A. J. van der Eb, A. G. Jochemsen, *J. Biol. Chem.* **2000**, *275*, 28039.

V. Bottger, A. Bottger, C. Garcia-Echeverria, **Y. F. Ramos**, A. J. van der Eb, A. G. Jochemsen, D. P. Lane, *Oncogene* **1999**, *18*, 189.

A. Shvarts, M. Bazuine, P. Dekker, **Y. F. Ramos**, W. T. Steegenga, G. Merckx, R. C. van Ham, van der Houven van Oordt, A. J. van der Eb, A. G. Jochemsen, *Genomics* **1997**, *43*, 34.

M. Nikolic, H. Dudek, Y. T. Kwon, **Y. F. Ramos**, L. H. Tsai, *Genes Dev.* **1996**, *10*, 816.

G. Hateboer, A. Gennissen, **Y. F. Ramos**, R. M. Kerkhoven, V. Sonntag-Buck, H. G. Stunnenberg, R. Bernards, *EMBO J.* **1995**, *14*, 3159.

* Shared first author

DANKWOORD

Aan het einde gekomen van deze 2de promotie was het een aangename verrassing dat, in tegenstelling tot zo'n 10 jaar geleden, volgens het huidige promotiereglement iedereen die op de één of andere wijze betrokken is geweest bij de totstandkoming van dit boekje officieel bedankt mag worden! Je zult begrijpen, zo'n boekje komt er niet vanzelf. Zeker niet door het werk van één persoon alléén, hoe hard deze ook kan werken. Dus, het is erg fijn dat ik nu mijn waardering kan uitspreken. Daar komt'ie:

Geachte Professor Slagboom, beste Eline, bedankt voor je inzet. Hoewel je uurtjes voor mij geteld waren, ze waren altijd vruchtbaar en leerzaam. Nooit ben ik zonder nieuwe moed en ideeën je kamer weer uitgelopen. Verder waardeer ik erg dat je als zeer actieve hoogleraar naast de wetenschap ook belangstelling hebt voor allerlei andere zaken waaronder muziek. Geachte dr. Meulenbelt, beste Ingrid, ontzettend bedankt voor je goede begeleiding, je hulp, en je interesse. Heel fijn dat je, intussen alweer 6 jaar geleden, met mij in zee wilde gaan. Zelf was ik na ons eerste gesprek erg enthousiast over je onderzoek en de moleculaire epidemiologie waar ik me eerder nooit in had verdiept. Gelukkig was het enthousiasme wederzijds. Ik heb in alle opzichten heel veel van je mogen leren de afgelopen jaren, en bewonder je gedrevenheid. Hoe druk je het ook hebt, je maakt altijd tijd voor overleg tot in de kleine uurtjes. Fantastisch! En fijn dat we de komende tijd nog samen verder kunnen werken.

Geachte leden van mijn promotiecommissie: Professor Bovée, Professor Karperien, en dr. Van Meurs. Hartelijk dank dat jullie gelegenheid hebben gevonden het manuscript te lezen. Ik waardeer de suggesties die ieder van jullie heeft gedaan ter verbetering van dit proefschrift. Beste Judith, tijdens het scoren van de verschillende gewrichten heb ik veel van je mogen leren, en ik ben erg dankbaar dat je altijd bereid bent een professionele blik te werpen op de verschillende coupes waar ik telkens mee aankom.

Rotsen in de branding: goede collega's met wie je alles van werk samen kunt doen. Nils: bedankt voor de fijne samenwerking en dat je m'n paranimf wilde zijn, Wouter: wat had ik zondeR gedaan?, Evelyn: bedankt voor je enthousiasme en inzet. Dennis, die ons intussen alweer een tijd geleden heeft verlaten maar wiens bijdrage aan mijn eerste Sequenome ik niet ben vergeten, Eka, Nico: heel fijn zoals we de microarrays samen hebben gedaan, Ruud, Wesley.

Speciale dank ook aan mijn kamergenootjes. De eerste 4 jaren Anika: ik vond het erg leuk dat ik je paranimf mocht zijn; Bianca met wie ik niet alleen over wetenschap maar regelmatig ook over de 'Grote Vragen des Levens' kan praten; Erik die net zelf deze mijlpaal achter de rug heeft en weet wat de steun van anderen hiervoor kan betekenen, Joris die altijd zo goed is mijn vragen over GWAS te beantwoorden of mee te kijken; last but not least: Matthijs, 'chef-kabels-en-méér'! Hartelijk dank dat je altijd bereid bent even mee te kijken, een sequentie uit de database te trekken, en heel veel méér.

Natuurlijk wil ik ook alle andere (ex) Malle Eppies bedanken: Bas, Caroline, Elmar, Eric-Wubbo, Jennifer, Jenny, Joyce, Kai, Koen, Leen, Maarten, Marian, Mark, Niels, René, Rita, Roderick, Steffan; en onze stagiaires. Ontzettend bedankt voor jullie kritische vragen tijdens

werkbesprekingen, jullie suggesties, het meedenken, etc. Dan, zeker niet op de laatste plaats: Inge, waar zouden we zijn zonder jou? Ik wil je bedanken voor alles wat je altijd voor ons doet. Maar vooral wil ik je bedanken voor je voorbeeld: dat je altijd zo opgewekt voor iedereen klaarstaat, ook als je zelf heel druk met van alles bezig bent!

Verder zo velen van andere afdelingen die op de één of andere manier hebben bij gedragen door onze samenwerkingen: Professor Jeanine Houwing-Duistermaat, Quinta, Ramin, Roula, Stefan; Inge Briaire; Professor Margreet Kloppenburg, Andreea; Professor Rob Nelissen, Angela, Karien, Kim, Monique, Nienke, Paul; Daan, Karoly; Henk, Yavuz; Professor Christine Mummery, Christian, en Valeria; Professor Peter ten Dijke, Amaya, Luuk; en Ietje, Kit, Shosh en Twan die ons op weg hebben geholpen in de wondere wereld van de stamcel.

Ook wil ik mijn oud collega's hier bedanken met wie ik zo'n goede tijd heb gehad en bij wie ik nog altijd terecht kan. Met name AG, Alfred, Alt, Amina, Hans, Kirsten, Mariam, Matty, Suzanne; Anabel, Hester, Leatitia. En ik wil niet nalaten Peter-Bram hier te noemen: uiteindelijk heeft onze samenwerking een paar jaar geleden er toe geleid dat ik deze kant van de wetenschap op ben gegaan. En natuurlijk alle mensen buiten het LUMC waar we zo fijn mee samenwerken; dank ook aan iedereen die ik niet persoonlijk hier heb genoemd of die ik onbedoeld vergeten ben!

Ten slotte, na deze bijna oneindige lijst van genoemde en niet-genoemde collega's die ik heel dankbaar ben en van wie ik veel heb geleerd, wil ik ook zeker niet vergeten alle artrose patiënten te bedanken die hun medewerking verlenen aan ons onderzoek, en natuurlijk ook mijn familie en vrienden. Marjolijn, samen uit, samen thuis: je was erbij toen ik begon, super dat je er bij kunt zijn nu ik afrond: bedankt dat je mijn paranimf wil zijn.

Lieve (oud)huisgenootjes en familie die de 'onderzoeks-perikelen' van dag tot dag mee hebben moeten beleven: het wel en het wee van een onderzoeker. Bedankt, Bernadet, Elisa, Inma, Loles, Maaike, Marinke, Pascale, Tiny, Yvonne; Helena, Bregje, Miriam, Marta, Nelly, Bego, en Paula. Jullie zijn de èchte Toppers! Bedankt, lieve papa&mama, lieve Marjolein&Johan, Bart&Suzan, Eddy&Sharon, lieve familie, ik weet: met of zonder tijd, we zullen er altijd voor elkaar zijn! Van jullie heb ik het belangrijkste geleerd, dat wat we door ons werk soms een beetje vergeten: jullie leren me te Leven! En ook al zien jullie me vaak op het nippertje nog binnen rennen bij een verjaardag of ander feest, en veel te vroeg weer opstappen, op weg naar de volgende afspraak of activiteit, bij jullie kan ik altijd terecht. Voor veel plezier, maar ook voor de nodige steun, goede raad, of gewoon een luisterend oor.

Voor allen: **mijn dank is heel groot!**