OSTEOARTHRITIS:

GENETICS AND PHENOTYPES IN ALL THEIR COMPLEXITY

Cindy Germaine Boer

The works described in this thesis was conducted at the department of Internal Medicine at the Erasmus MC University Medical Center Rotterdam, the Netherlands.

The Rotterdam study is funded by Erasmus MC University Medical Center and Erasmus University Rotterdam; the Netherlands Organization for Scientific Research (NWO); the Netherlands Organization for Health Research and Development (ZonMW); the Research institute for Diseases in the elderly (RIDE); the Dutch Ministry of Education, Culture and Science; the Dutch Ministry of Health, Welfare and sports; the European Commission (DG XII); and the municipality of Rotterdam.

ISBN: 978-94-6416-105-2

Cover Design: Cindy G. Boer
Design: Cindy G. Boer and Ridderprint
Illustrations: Cindy G. Boer
Layout and printing: Ridderprint

Osteoarthritis:
genetics and phenotypes in all their complexity

Artrose:
genetica en fenotypen in al hun complexiteit

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam

op gezag van de rector magnificus

Prof. Dr. R.C.M.E. Engels

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
dinsdag 15 december 2020 om 13:30 uur

Door

Cindy Germaine Boer
geboren te Sliedrecht

Erasmus University Rotterdam

Promotiecommissie

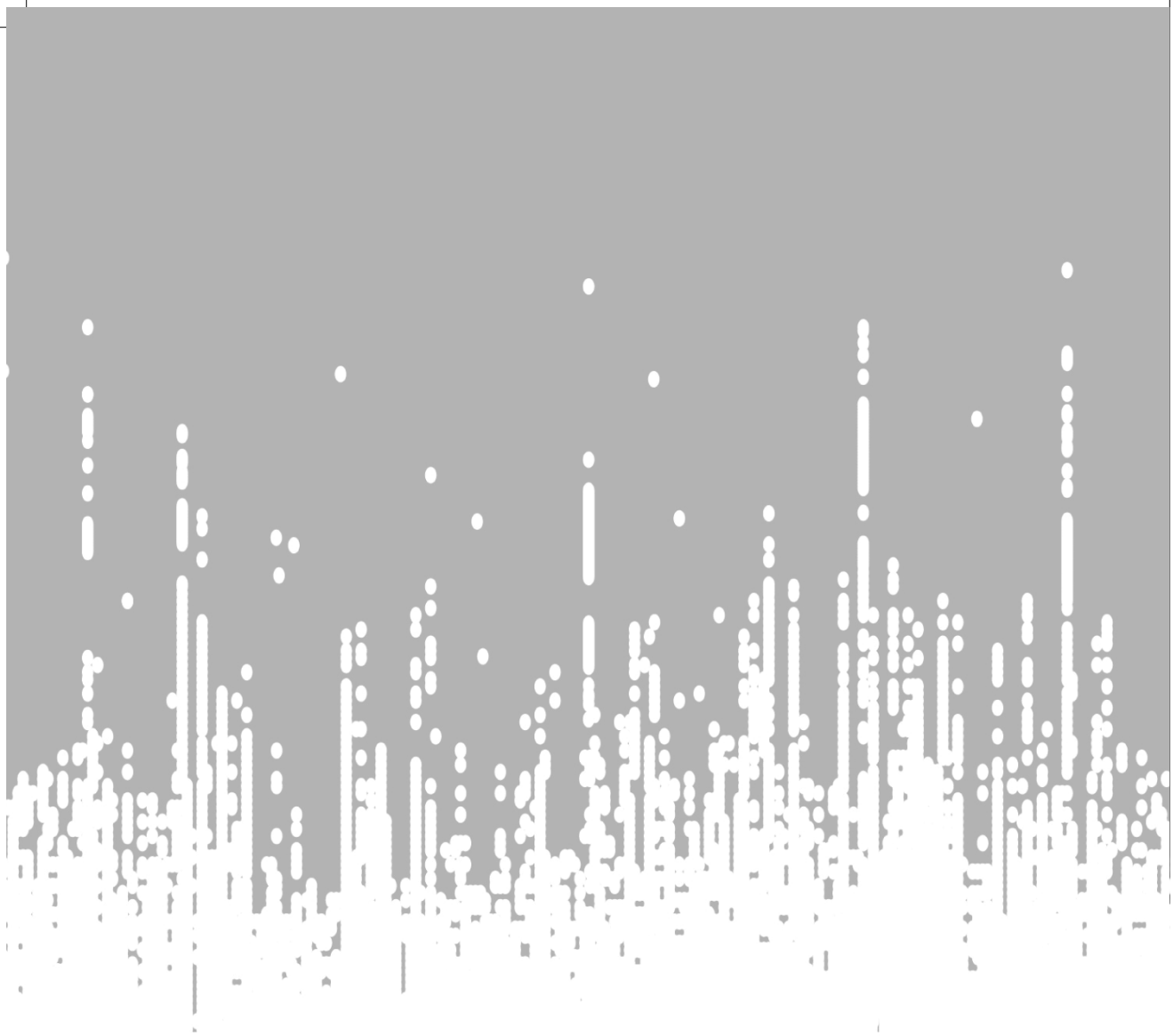
Promotor:	Prof. Dr. A.G. Uitterlinden
Overige Leden:	Prof. Dr. G. van Osch
	Prof. Dr. M.A. Ikram
	Prof. Dr. E. Zeggini
Co-promotor:	Dr. J.B.J. van Meurs

Paranimfen:	Justin D. Boer, BSc.
	Joost A.M. Verlouw, BSc.

Contents

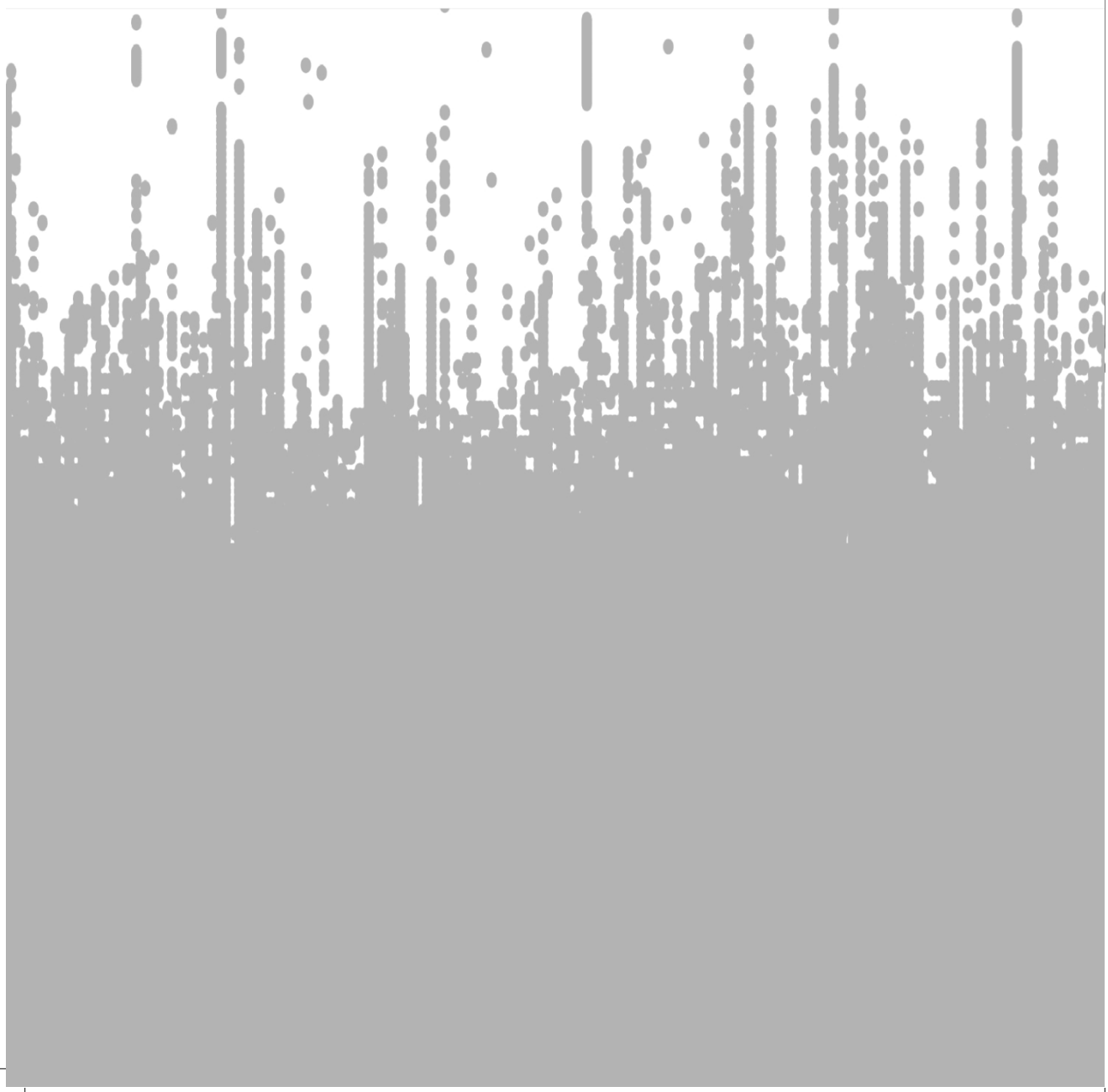
A PhD is not a Journey, it's a Quest.

Alistair McCulloch



Chapter 1

General Introduction



Chapter 1.1

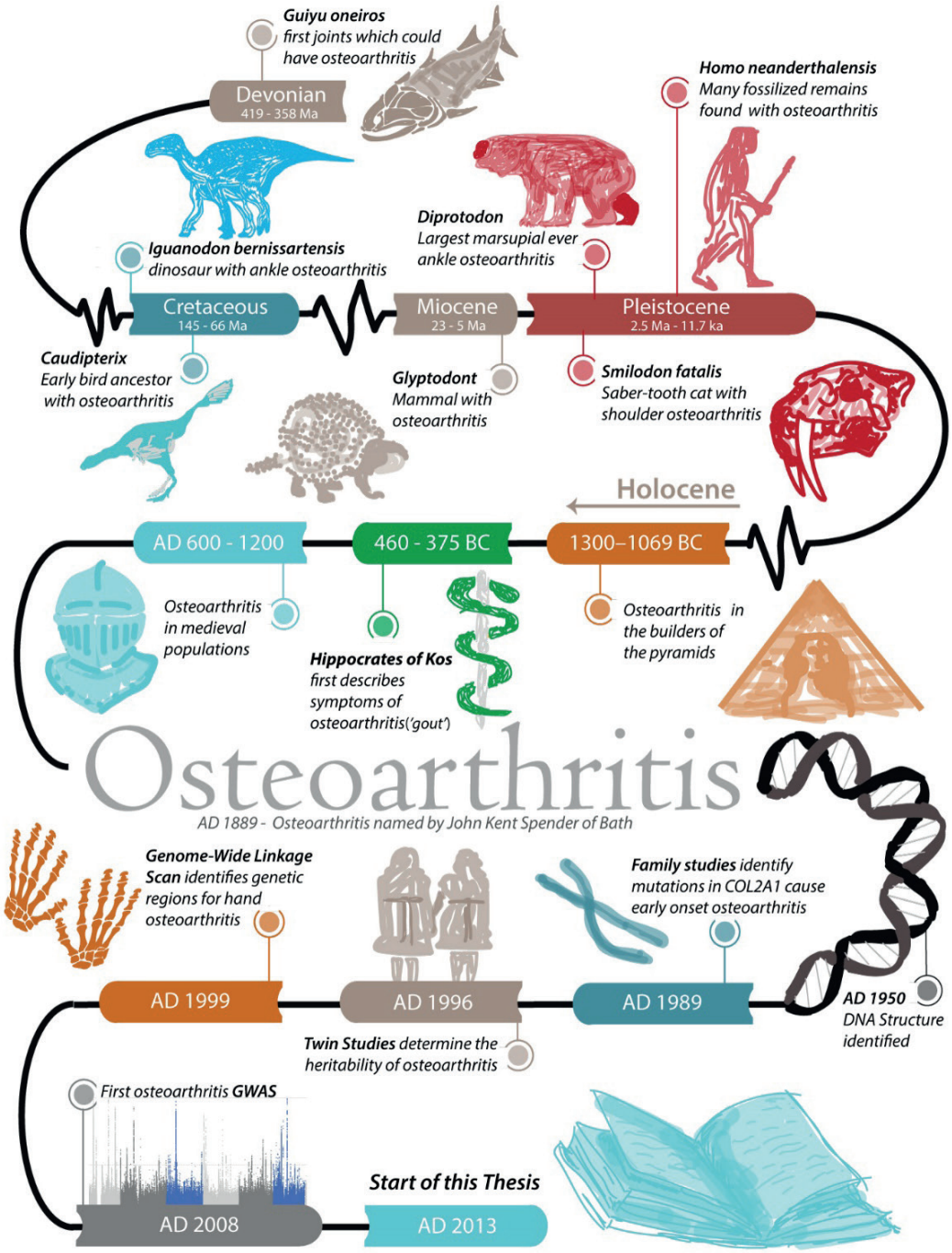
General Introduction

The problem

Osteoarthritis is one of the world's oldest diseases, and yet no curative treatments are available. Osteoarthritis is the most common chronic degenerative joint disease worldwide and has been present throughout history, evidenced by the presence of osteoarthritis in the fossilized ankle bones of *Iguanodon bernissartensis* (**Figure 1**), the first ever recognized dinosaur species[1, 2]**.** Not only dinosaur fossils have been found with osteoarthritis, also early bird ancestor *(Caudipteryx*)[3], marsupial (*Diprotodon*) [4], and mammal (*Glyptodont* and *Canis dirus*)[5, 6] fossils are known. In fact, probably the first common ancestor to all bony vertebrates (*Eutelestomi*) could have been affected by os-teoarthritis[7]. Early man was also plagued by osteoarthritis. The disease is one of the most common features seen in archeo-paleontology[8], with evidence from Neander-thals[9], early inhabitants of America[10], Egyptian mummies[11, 12], medieval citi-zens and knights[13] (**Figure 1**)**,** up to the modern day where 4-5% of the entire world population has at least one joint affected by OA[2].

Symptoms of the disease are stiffness, loss of mobility and loss of function in the affected joint, with pain described as the most dominant and disabling symptom of the disease[14]. Not surprisingly, osteoarthritis is the single biggest cause for joint pain and disability worldwide[2]. Currently osteoarthritis is the third most rapidly growing chronic disease worldwide[2]. Combined with the global increase in risk factors for osteoarthritis (aging, joint injury and obesity), the impact and burden of osteoarthritis will only continue to rise. Yet, no curative treatments for osteoarthritis are known, and much of the pathology of osteoarthritis remains a mystery. Thus, it is high time to unravel the mysteries surrounding osteoarthritis, a disease plaguing its hosts for over hundreds of millions of years.

►**Figure 1: A brief Timeline of Osteoarthritis**. Osteoarthritis is one of the oldest known diseases in the world; the first known bony vertebrate (*Guiyu onerios*) might already have had osteoarthritis. Fossil evidence dates back to the cretaceous (~125 Ma), and ample evidence of the disease exists in the Pleistocene, where even the *H. neanderthalensis* was affected by osteoarthritis. Osteoarthritis is seen throughout human history, from ancient Egypt, Greece (where Hippocrates first described signs of the disease) and medieval Europe to the modern day, however it was not until 1889 that osteoarthritis got its name. With the discovery of the helix structure of DNA, genetic research into osteoarthritis started in full; from family studies, twins studies, linkage scans and the first Genome-wide association study to the start of this thesis on "*Osteoarthritis; Genotypes and phenotypes in all their complexity*". Ma: Mega *annum*, ka: kilo *annum, annum*: one year. BC: Before Christ, AD: *Anno Domini*


Guiyu oneiros
first joints which could have osteoarthritis
Devonian
419 - 358 Ma
Homo neanderthalensis
Many fossilized remains found with osteoarthritis
Diprotodon
Largest marsupial ever ankle osteoarthritis
Iguanodon bernissartensis
dinosaur with ankle osteoarthritis
Cretaceous
145 - 66 Ma
Miocene
23 - 5 Ma
Pleistocene
2.5 Ma - 11.7 ka
Caudipterix
Early bird ancestor with osteoarthritis
Glyptodont
Mammal with osteoarthritis
Smilodon fatalis
Saber-tooth cat with shoulder osteoarthritis
Holocene
AD 600 - 1200
460 - 375 BC
1300–1069 BC
Osteoarthritis in medieval populations
Hippocrates of Kos
first describes symptoms of osteoarthritis('gout')
Osteoarthritis in the builders of the pyramids
Osteoarthritis
AD 1889 - Osteoarthritis named by John Kent Spender of Bath
Genome-Wide Linkage Scan identifies genetic regions for hand osteoarthritis
Family studies identify mutations in COL2A1 cause early onset osteoarthritis
AD 1999
AD 1996
AD 1989
AD 1950
DNA Structure identified
Twin Studies determine the heritability of osteoarthritis
First osteoarthritis GWAS
Start of this Thesis
AD 2008
AD 2013

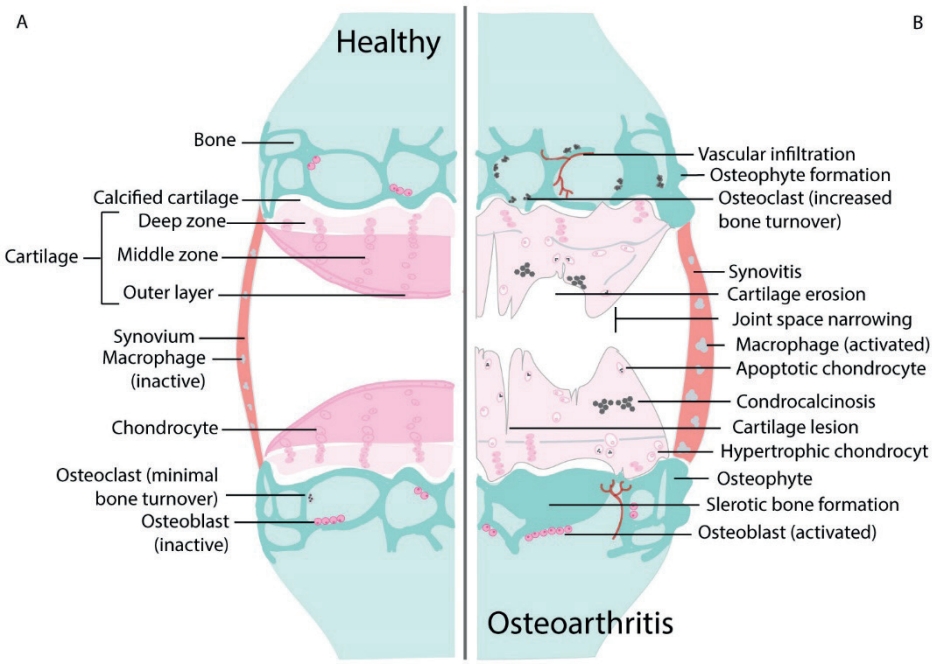
What is osteoarthritis?

The first descriptions of osteoarthritis can be traced back to writings of Hippocrates (460-375 B.C.E.)[15](**Figure 1**). However, from these first writings until the beginnings of the 19th century, all forms of rheumatic joint diseases, including osteoarthritis, were called "*gout*". In 1802, William Herben the Elder recognized osteoarthritis as a separate disease for the first time[16]. However, it took until 1889 that John Kent Spender of Bath England coined the name osteoarthritis[17]. The name referring back to the ancient Greek of Hippocrates, where osteo- comes from the Ancient Greek ὀστέον (ostéon) meaning "bone" and arthritis from the combination of arthr- ἄρθρον (árthron) meaning "joint, limb" and -itis from ῖτις(îtis) meaning "pertaining to", and together forming osteoarthritis: "bone pertaining to the joint/limb".

As implied by the name, osteoarthritis pathology commonly involves the bone, however, we now know that osteoarthritis is a disease of the whole joint (**Figure 2**). Affecting multiple tissues within the joint, most notably the cartilage and bone, but also the synovium, and possibly the muscles and tendons of the joint[18]. Osteoarthritis is a degenerative disease involving an active and complex process of multiple mechanical, metabolic and inflammatory pathways leading to the destruction and failure of the affected joint[19]. the most common and characteristic features of the disease pathology are the degradation of the articular cartilage, together with the formation of new bone, osteophytes, at the margins of the joint. Other known pathological features of osteoarthritis are bone sclerosis, cartilage lesions, cysts, chondrocalcinosis, and synovial inflammation.

Long has it been thought that osteoarthritis was an inevitable part of aging, that it was a passive "wear-and-tear" disease. However, we now know that osteoarthritis is an active and dynamic disease, an imbalance between the repair and destruction of joint tissue[19]. In addition, osteoarthritis is associated with increased comorbidity and even increased mortality[20]. More and better understanding of the pathological processes and pathways involved is imperative, if we want to develop treatment strategies for osteoarthritis.

The definition of osteoarthritis is based on the pathological effects of the disease on any combination of the tissues involved. Therefore multiple definitions of osteoarthritis exist. However, the most used definitions are radiographic or clinical osteoarthritis. The radiological definition usually consists of the Kellgren-Lawrence grading scale, based on the presence and severity of joint space narrowing (JSN) and osteophytosis (OST)[22], while the clinical definition is focused on the clinical symptoms; pain, swelling and stiffness of the joint[23]. Although there is some standardization in osteoarthritis definitions [24], definitions can still vary per study[25]. Although there is some



▲ **Figure 2: Characteristics of an osteoarthritic joint**. **A)** Schematic overview of a healthy articular joint. **B)** Schematic overview of characteristics and structural changes in a joint affected by osteoarthritis. Any combination of these changes can be seen in an osteoarthritic joint. Figure adapted from[21].

standardization in osteoarthritis definitions [24], definitions can still vary per study[25]. Also, as osteoarthritis is a heterogenous disease, involving many pathological pathways all leading to the same outcome, subgroups with distinct pathology are possibly unrecognized. All contributing to heterogeneity and complexity of osteoarthritis research.

The complexity of osteoarthritis

Osteoarthritis is a complex multifactorial disease. Not just because it has had a turbulent and long history or that the pathology affects multiple tissues of the whole joint. Osteoarthritis is a complex disease in the scientific sense, where complex means: both genetic and environmental factors play an important role in the development of the disease. Multiple environmental risk factors have been recognized for osteoarthritis, with age, physical activity and obesity as the most well-established risk factors[21].

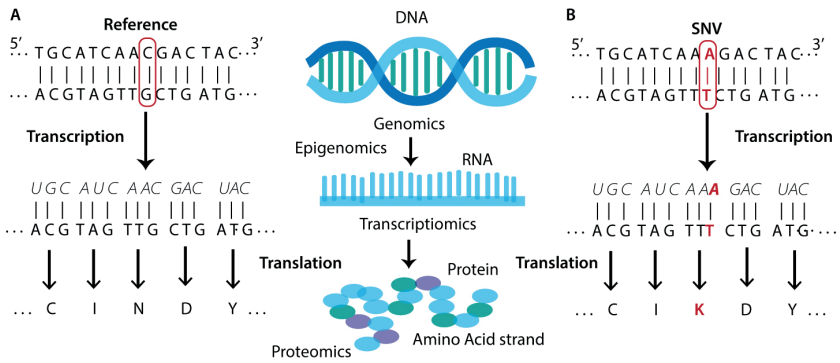
The complex interplay between the different environmental risk factors themselves and with the underlying genetic risk factors, determines an individual's risk for developing osteoarthritis. Mostly from twin studies we have learned that the predicted

risk for osteoarthritis that can be contributed to genetics is estimated to be ~39%-65% depending on the joint affected[20]. The heritability component of knee osteoarthritis is ~40%, hip osteoarthritis ~60% and for osteoarthritis of the hand ~39%-65% depending on the hand joint[26-28]. Thus, a large part of the risk for osteoarthritis is determined by genetic variation. Understanding how and which genetic variations contribute to OA risk is important, as this will lead to more knowledge of disease pathogenesis, which in turn could lead to the development of new treatment or preventive strategies.

Genetics: a primer

Genetics is the study of heredity, genes and genetic variation. DNA, or deoxyribonucleic acid, is a complex molecule that stores all of the genetic information, the instructions for development, function and maintenance of an organism. The DNA consists of two strands that wind around each other in a double helix, like a twisted ladder[29]. Each strand consists of a sugar-phosphate 'backbone' to which one of four nucleotides (bases) are bound; Adenine (A), Thymine (T), Cytosine (C) or Guanine (G). Each "rung" of the ladder consists of a base pair (bp), which are two nucleotides held together by hydrogen bonds; A always pairs with T, and C always pairs with G. The human DNA contains the genetic code for ~20.000-25.000 genes (average number of parts in a car is ~30.000). A gene is a part of the DNA that gets copied (transcribed) into RNA, which is translated into proteins. Proteins are the essential building blocks and machinery of a cell, performing all manner of functions. The transcription of DNA to RNA and the translation of RNA to protein is called the "central dogma of molecular biology" [30] (**Figure 3**).

Although the DNA sequence between any two humans is for ~99.5% similar, each individual is unique, as is their DNA sequence. Variations in the genetic code can explain differences between individuals, such as hair colour, height, shape of the face, or disease risk. On average, an individuals' genome differs on 4.000.000 to 5.000.000 sites from the human reference genome[31]. Many different types of genetic variation exist, of which the most common in the human genome is the Single Nucleotide Variation (SNV or SNP)[32] (**Figure 3**). Each SNV, named with its reference SNV cluster ID (rsID), represents a single nucleotide difference on the DNA from the human reference genome. In total there are currently ~335.000.000 SNVs known based on ~1 million human genomes sequenced [33]**,** and they can occur anywhere on the DNA, both in coding and in non-coding regions. In the coding regions a SNV could affect the translation of a gene into a protein, and in the non-coding regions it could affect the regulation of gene expression (gene transcription).



▲**Figure 3: The central Dogma of Molecular Biology** DNA can be copied into DNA or translated into RNA, RNA can be copied into RNA or translated into protein. **A)** Schematic representation of the molecular dogma, depected as nucelotide letters or as an schematic image. **B)** Representation of the possible effect of a single nucelotide variant on protein translation, due to the SNV the protein triplet code does not code for the amino acid Asparagine (N) but for Lysine(K), a nonsynonymous SNV, as the SNV changes the protein code. A synonymous SNV does effect DNA and RNA sequence but not the protein amino acid sequence.

However, SNVs that have a large effect on the functioning of a gene can impact reproductive success and will not propagate into a population, and thus will occur infrequent in a population (<1%). Common SNVs, those that occur frequently in a population(>1%), have usually relatively small effects, and, thus, are under less selective pressure[34]**.** Certain SNVs in relative proximity tend to always co-occur together, they are said to be in linkage disequilibrium (LD). Such a group of SNVs is called a haplotype, and, as with individual SNVs, different haplotypes occur at different frequencies in different human populations, and they will differ more in frequency if the populations are geographically and ancestrally more different[35]

Complex genetics

Since genetic variation can explain differences in disease risk between individuals, investigating which and how specific genetic variations can increase or decrease disease risk could provide insight into the disease pathology. In order to identify which genetic variation is related to a disease, several approaches have been tried, including linkage analysis in families or sib-pairs and, genome wide association studies (GWAS).

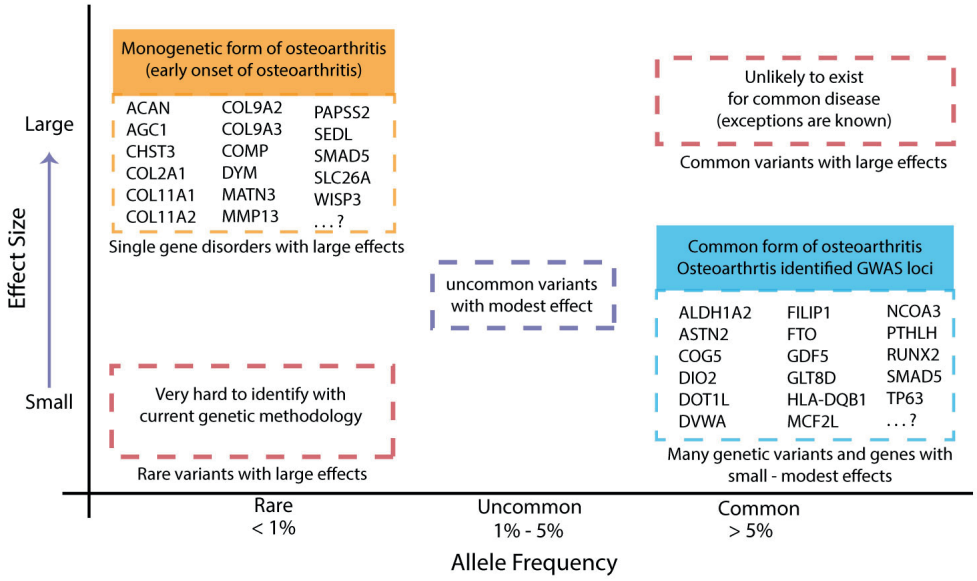
Gregor Johan Mendel was the first to demonstrate that traits can be inherited from parent to offspring in his plea plant (*Pisum sativum*) experiments in 1865[36]. With his experiments he laid the foundation for the field of genetics. However, it was not until the 20th century that his work was rediscovered and, it took until 1941 that Stecher *et al.*, first postulated the possible involvement of genetics in osteoarthritis. He first described that the Herben's nodes (bony nodes on the joints) seen in hand osteo-

arthritis were inherited[37, 38]. it would not be It would not be until 1989, after the discovery of the DNA structure by Watson and Crick[29], that through linkage studies in families the first gene associated with osteoarthritis would be found. In genetic linkage studies inheritance of DNA regions, marked via genetic markers, is linked with the occurrence of the disease in a family. Using this method, a genomic region containing mutations (rare SNVs) in the *COL2A1* gene was found to co-occur with osteoarthritis, in two families with high occurrence of premature osteoarthritis of several joints[39]. However, these mutations in the *COL2A1* gene are very rare, while osteoarthritis is very common in the population, nor could the *COL2A1* mutations explain the occurrence of other familial forms of osteoarthritis[40]. All indicating that osteoarthritis is also complex at the genetic level. Indicating that not one gene is involved in osteoarthritis risk but many genes are[40].

Since linkage studies are primarily useful to identify large effects of rare genetic variants for monogenetic (single gene) diseases, they are not suited for the study of complex diseases. In complex diseases, multiple genetic variations in multiple genes and their interaction with the environment determine the risk of disease. Since hundreds, if not thousands of variants can be involved, each individual variant will have a small effect on disease risk[41]. Osteoarthritis has both rare monogenetic forms, such as the *COL2A1* mutations, and the much more common complex forms of the disease for which we now know that several hundreds of common variants with more subtle effects are involved in the genetic architecture of the disease (**Figure 4**). Thus other methods are needed for the study of complex diseases, methods than can identify subtle effects of many hundreds of variants across the genome. Genome-wide association studies (GWASs) are such a method.

Genome-Wide Association Studies

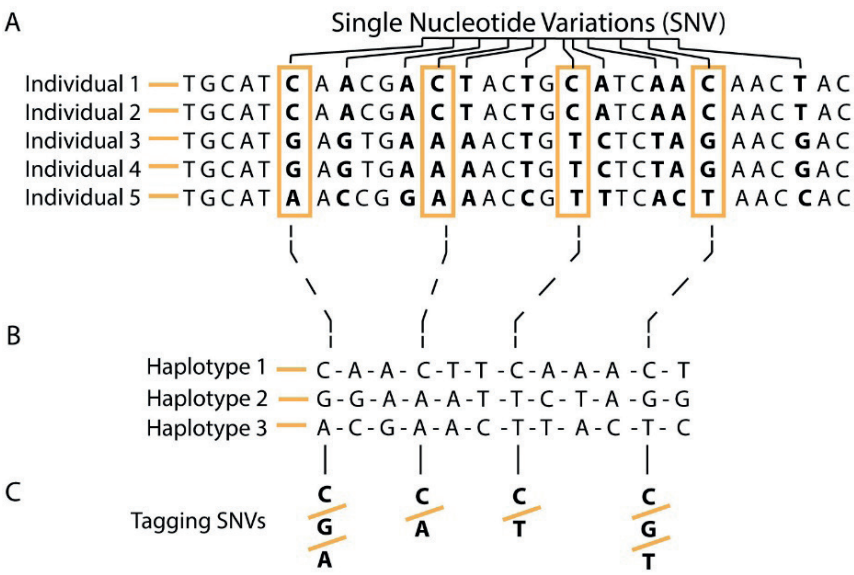
In a genome-wide association study (GWAS) many millions of SNVs are examined whether a certain SNV occurs more frequently in individuals with the disease than in individuals without the disease[41]. The GWAS study design was made possible with the advent of cheap genotyping methods, where hundreds of thousands sites of genetic variation (SNVs) are measured on a single array. Through these measured SNVs the genotype of other SNVs within the same linkage(LD) block can be inferred by imputation (**Figure 5)**. Early arrays measured the genotype of hundreds of thousands of SNV uniformly distributed across the genome. However, this did not accurately or efficiently captured all LD blocks within the human genome, thus resulted in missing information on genetic variation in certain sections of the genome. Current arrays more efficiently



▲**Figure 4: Allele frequency and osteoarthritis disease risk.** Early onset osteoarthritis or Mendelian/ monogenetic forms of osteoarthritis, are caused by single gene mutations. Such genetic variants have large effects, but are rare in the population. Known genes with rare mutations that give rise to early onset osteoarthritis are given in the box. For late onset or the common form of osteoarthritis, many variants and genes confer risk. Meaning that almost all of these variants must have small to moderate effect sizes. Thus common variants with large effect sizes are unlikely to exist for common diseases, although it is possible. In the box known GWAS identified loci associated with osteoarthritis are given. Figure adapted from [34, 42]

utilizes the LD between SNVs by only measuring the genotype of "tagging" SNVs, SNVs which "tag" the genotype of their LD block (**Figure 5**), resulting that only several hundreds of thousands SNV are needed to efficiently "tag" the full variation across the entire genome.

Although imputation methods thus provide a cheap method for identifying the full genomic variation, the quality and accuracy of the imputation is dependent on the quality of the reference information. The SNV linkage blocks (haplotypes) needed for imputation are based on reference haplotypes generated via whole genome sequencing in populations. The first efforts to create the human "haplotype-map" was done by the HapMap consortium, in 2005, with the last of the data published in 2009[35, 43]. This first haplotype map was based on the whole genome sequencing data of 692 individuals from 11 global populations[43]. Different haplotypes occur with different frequencies in different populations, some might even be population specific. Thus, more whole genome sequences across multiple population are needed to improve the human haplotype map, and thereby the quality and accuracy of imputation. This was done in the following 1000 Genomes project, which contained 1,092 whole genome sequences from



▲**Figure 5: Single Nucleotide Variants, Haplotypes , Imputation and Tagging SNVs. A**) 5 short stands of DNA from four versions of the same genetic location in 5 different individuals. Most of the DNA sequence is identical, except for 12 single Nucleotide Variations (SNV) nucleotide variations (SNVs). Each SNV has several possible alleles (A/T/G/C). **B**) From the 5 individuals DNA sequence, 3 haplotypes can be distinguished. A haplotype is made up of a particular combination of alleles that are inherited together (in linkage). Only the SNVs are shown and 4 of them are marked. **C**) Tagging SNVs. By just genotyping these 4 tagging SNVs the genotype of the other 8 SNVs in linkage could be determined a.k.a. imputed, and identify these 3 haplotypes. Thus if an individual has the genotype G-A-T-G at these 4 SNVs, their haplotype would be determined as number 2. Note that haplotypes can occur at different frequencies in different populations, haplotype 3 in this example occurs at a lower frequency than haplotype 1 and 2. More reference sequences are needed to be able to detect other possible haplotypes, thereby determining the best tagging SNVs and thus to improve the accuracy and quality of imputation.

11 global populations[44], and this number has rapidly increased over the last years with the Haplotype Reference Consortium (HRC, n=64,976)[31] and Trans-Omics for Precision Medicine (TOPMed) imputation (n= 4,800,000)[45]. Resulting in more and more accurate imputations and more genetic variants for GWAS to use.

About 1 million of haplotypes (in the Caucasian genome) are examined at the same time in a GWAS, thus strict multiple testing corrections are needed to prevent false positive results. For this reason SNVs are genome wide significant in a GWAS if the SNV if the threshold of p-value≤$5x10^{-08}$ is reached. Since this genome wide significance still is nominally a p-value≤0.05 (but now corrected for the 1 million LD blocks across the human genome), such genome wide discoveries have to be replicated in an independent study to distinguish true findings from coincidence. From the many GWASs performed since the first GWAS in 2005, we know that the effect sizes of the majority of associated SNVs (GWAS hits) are relatively small. Thus, due to the small effect of the

SNVs associated with complex diseases, GWASs need very large sample sizes to have the statistical power to robustly identify associated genetic variations.

In sum, a GWAS uses imputed genotype information of millions of SNVs to identify genetic associations in large sample sizes. SNVs are included into the GWAS, without any a-priori selection, all possible SNVs that can be imputed are included. This is also termed "hypothesis-free". However, contrary to popular belief, GWASs are not truly hypothesis-free, as GWASs assume that the disease or investigated trait has a genetic component[41]. This is why GWAS has proven to be a very successful method for finding genetic variation associated with complex disease, such as osteoarthritis [41] (**Figure 4**) and, not so much for being a *"dog-person"*[46].

Epigenetics: beyond genetic complexity

A GWAS identifies SNVs or a genomic region (locus) to be associated with a disease or trait, it only very rarely directly identifies genes. Only through many subsequent steps in downstream bio-informatics and functional studies to test the candidate genes in the region associated genes can be identified. However, it is not always very easy to determine how a SNV could affect a gene or which gene it would affect, especially within the context of a disease or trait. SNVs can be located anywhere in the genome, and only 2% of the genome consists of coding DNA. However, the 98% non-coding DNA is not without function and is thought to have several important roles including in gene regulation.: the regulation of the 2% coding DNA, via gene regulatory sequences. These are stretches of DNA sequences that are involved in the regulation of gene expression. Usually DNA binding proteins, such as transcription factors (TF), can bind to these DNA sequences and regulate gene expression. Such gene regulatory sequences can be identified in the DNA by examining where these TF bind the DNA, or by examining epigenetic modifications.

Epigenetic modifications are modifications to the DNA base pairs and DNA binding proteins that do not alter the DNA sequence, but do affect gene activity and expression. Many such epigenetic modifications are known, these include DNA methylation and histone modifications. How these epigenetic modifications can be used for GWASs is detailed in **Chapter 1.2.** Epigenetics is also a mechanism by which the environment can affect gene activity and regulation, and, thus, influence disease risk. What the role is of epigenetics in skeletal disorders such as osteoarthritis is also detailed in **Chapter 1.2.** Throughout this thesis, epigenetics is used in GWAS, in order to gain more insight into osteoarthritis pathology, and ultimately provide novel treatment options. **Chapters 2-4** use epigenetics for GWAS interpretation and in **Chapter 3** these GWAS interpretations even lead to possible clinical implications for osteoarthritis patients.

Genome-wide osteoarthritis association studies

The research field of osteoarthritis was an early adaptor of GWASs and has proven to be relatively successful in finding underlying genetic variations, results of which (for 2013) are summarized in **Figure 4.** The first GWAS performed for osteoarthritis was done in 2008, and identified a single region/locus of the DNA [47]. As stated before, several hundreds of genes are probably involved in osteoarthritis. Each common SNV, as identified in GWAS, will only pose a small risk for osteoarthritis, meaning that probably hundreds if not thousands of SNVs will be associated. Due to these small effect sizes of individual SNVs, very large numbers of osteoarthritis cases and controls are needed for successful and robust SNV discovery through GWAS [41]. For this reason, GWAS researchers are forced to collaborate in order to collect such large sample sizes.

The first osteoarthritis collaboration, the TREAT-OA (Translational Research in Europe Applied Technologies for OsteoArthritis) consortium, included 6,709 osteoarthritis cases, and 44,439 controls [48]. This consortium only confirmed the previously found single genetic locus in 2008, a far cry from the hundreds thought to exist. Although this was a massive effort in 2008, the sample size was still too small for robust GWAS findings. Thus, even larger collaborative efforts were needed. **Chapter 4** presents the results from the Genetics of Osteoarthritis (GO) consortium. This is the most recent and largest collaborative effort on osteoarthritis genetics, including ~180.000 osteoarthritis cases and 600.000 controls, a ~26 fold increase in sample size from the first collaborative effort.

Phenotypes for osteoarthritis

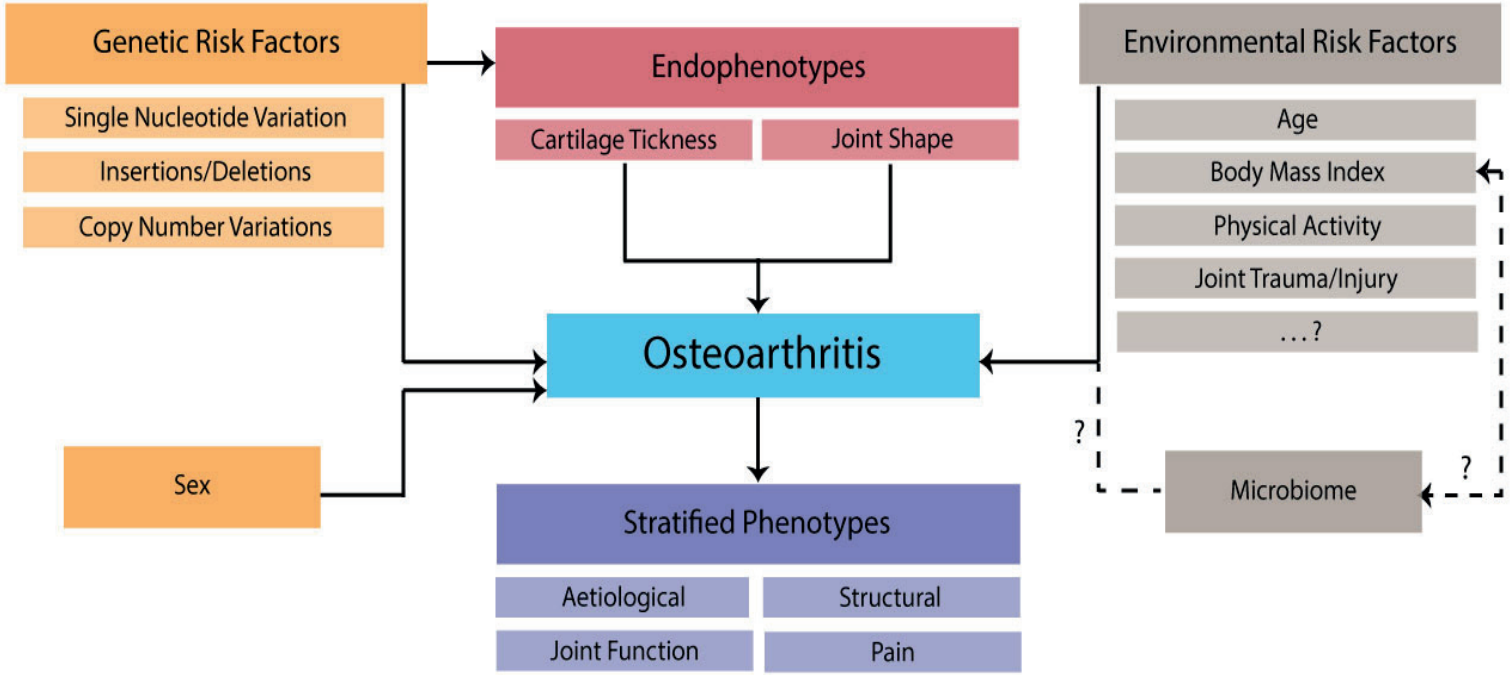
Increasing the number of cases and controls involved in a study is not the only way to increase the power to robustly detect associated genetic variants. To reduce the heterogeneity seen in complex diseases(such as in osteoarthritis), the use of endophenotypes or stratified phenotypes of the disease is possible (**Figure 6**). The use of such phenotypes to reduce osteoarthritis heterogeneity will be presented in **Chapter 2**. Endophenotypes are phenotypes (characteristics, traits) of the disease that are more closely related to the underlying genetics, than to the disease itself [49]. For example, the thickness of cartilage in the joint is considered an endophenotype for osteoarthritis[50]. This is because the amount of articular cartilage is fixed, there is limited replacement of the collagen matrix after skeletal maturity, even with the occurrence of disease. Meaning that the total cartilage thickness in a joint is largely determined by underlying genetics[51]. **Chapter 2.1** presents the results of using cartilage thickness in the hip joint as an endophenotype for hip osteoarthritis. This and other osteoarthritis endophenotypes are depicted in **Figure 6**.

Other than using endophenotypes, stratified osteoarthritis phenotypes can also be used to reduce heterogeneity and increase GWAS power. Stratified phenotypes aim to differentiate subgroups within the diagnosis of osteoarthritis[25]. These can be based on a single pathological phenotype of osteoarthritis (e.g., amount of osteophytes or bone marrow lesions present) or based on multiple characteristics (clinical, radiographic osteoarthritis, joint function and disability)(**Figure 6**). Indeed, multiple broad term categories could be recognized for osteoarthritis stratified phenotypes based on either structural (pathological structure formations) aetiological (based on the underlying cause of osteoarthritis) characteristics or pain (clinical osteoarthritis, presence of inflammation), joint function/disability (range of motion, gait), and molecular (which pathological molecular pathway is dominant) (**Figure 6**). Many more stratified osteoarthritis phenotypes might be possible and will be discovered as continued work on osteoarthritis and its phenotypes yields more genetic loci and etiological insight into the biological mechanism underlying this disease. **Chapter 2.2** and **3.1** will demonstrate the results of using structural osteoarthritis phenotypes for the identification of genetic variants associated with hand osteoarthritis.

Osteoarthritis pain, inflammation & the microbiome

As described, the -itis suffix in osteoarthritis comes from the Ancient Greek ῖτις (îtis) meaning "pertaining to". However, in recent years the –itis suffix in a disease name has come to be associated with inflammation. In recent history there has been much debate, as osteoarthritis was thought to be a disease without any inflammatory involvement[52, 53]. Currently, however, the involvement of synovial inflammation in osteoarthritis pathology has been well established, and thought to be one of the causes for the pain seen in osteoarthritis affected joints[54]. Several causes for this inflammation have been postulated, such as traumatic joint injury (sports, accidents, falls etc.), leading to joint damage and inflammation, triggering more damage and inflammation. A novel factor postulated to be involved in osteoarthritis related inflammation is the gastrointestinal(gut) microbiome[55, 56] (**Figure 6**).

The gastrointestinal microbiome consists of trillions of microorganisms, mainly bacteria, living in our intestinal tract. These microorganisms are important in the digestion of our food, and recent research has shown that there is also a very important role for the microbiome in health and disease. This has become clear by observing the composition of the microbiome, or rather the change in composition of the microbiome (dysbiosis) which has been linked to several disorders and diseases [57, 58]. Most notably the effect of the microbiome on obesity and the effect of obesity on the microbiome

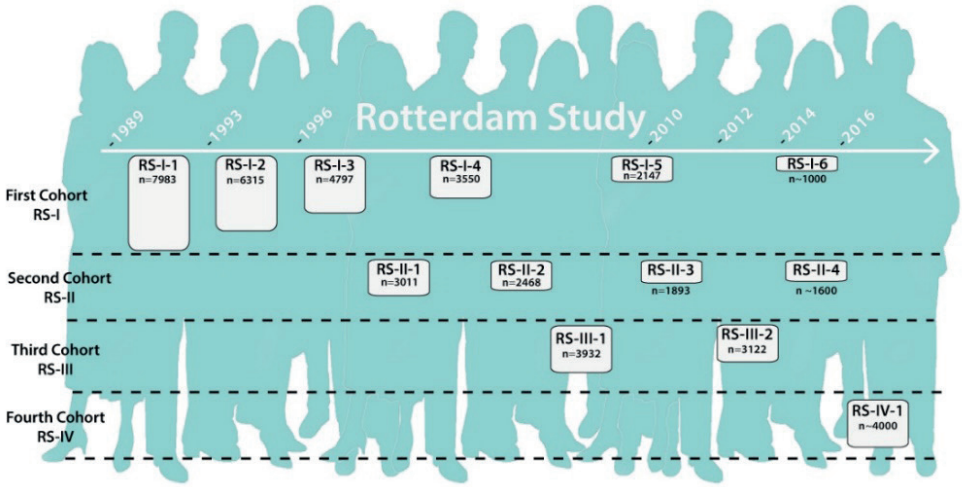


▲ **Figure 6: Risk factors and phenotypes for osteoarthritis**. Osteoarthritis is a complex disease determined by genetic risk factors and environmental risk factors. Several known types of genetic risk factors and environmental risk factors are given in the figure. Endophenotypes are intermediary phenotypes that are more closely associated to the underlying genetic risk factors than to the disease. osteoarthritis. Examples are the shape of the joint and the amount (thickness) of the cartilage present in the joint. In contrast to endophenotypes, Stratified phenotypes aim to differentiate subgroups within the diagnosis of osteoarthritis to reduce the heterogeneity of the osteoarthritis diagnosis, for example by grouping osteoarthritis patients based on the underlying different aetiologies, different structural pathologies (osteophytes, joint space narrowing and lesions), amount of pain experienced or by the amount of function of the joint (mobility, range of motion).

has been well-documented[59, 60]. Individuals with high BMI (body mass Index) have a significantly different microbiome than those who are not obese.Obesity is one of the most well-known and characterized risk factor for osteoarthritis [61, 62]. The increased risk for osteoarthritis is thought to be due to increased loading of the joint. However, this would not explain the observed increased risk for osteoarthritis in non-weight bearing joints, such as the hands[63]. Another explanation is the increased systemic inflammation seen in obese individuals[64], which is thought to arise due to the differences/dysbiosis in the microbiome of obese individuals[65]. Postulated is that changes in the gut microbiome due to the environment (obesity) can lead to an increased systemic inflammatory state (low grade inflammation)[55]. This is something that will be explored in **Chapter 5**.

Osteoarthritis In the Rotterdam Study

All of the studies in this thesis were performed with the use of the "Rotterdam Study" cohorts, also known in Dutch as the "Erasmus Rotterdam Gezondheid Onderzoek"(ERGO) [66]**.** This is a population-based prospective cohort, designed to study the determinants of aging. The Rotterdam Study consists of several sub-cohorts (**Figure 7**)**,** all consisting of individuals (>45 years of age) living in the well-defined suburb of Ommoord in the city of Rotterdam, the Netherlands. Participants were examined in detail at baseline in 1990 via home interviews and an extensive set of examinations at the specifically built research enterer in Ommoord. In addition, follow-up examinations were repeated every 3-4 years. The Rotterdam Study focusses on 14 different research lines in the context of "healthy aging" which includes all major organ systems and related diseases and , important for the studies described in this thesis, also diseases of the musculoskeletal system such as osteoporosis and osteoarthritis, In addition, many important determinants of disease are documented and studied including, e.g., medication use, dietary patterns, and genetics and genomics. Also extensive bio-banking is part of this long running cohort study, in particular blood is collected at each follow-up measurement from which DNA was extracted. The DNA has been subjected to several genetic analyses, most notably a large SNP array genotyping effort, to allow GWAS studies taking place with all the data in RS, which started in 2008 and has since been extended to include almost the complete cohort. Relevant for OA research, radiographs were made of multiple joints, such as knees, hips and hands at baseline and during follow-up examinations. All radiographs have been scored on the Kellgren-Lawrence osteoarthritis severity scoring by expert clinicians.



▲**Figure 7: Overview of the Rotterdam Study.** The four Rotterdam Study (RS) cohorts are indicated, and their start and follow-up visits.The Rotterdam Study started in 1990 and is currently still ongoing. Boxes indicate the number of participants for each cohort and visit. RS-I-1 is the largest cohort, and it's box size functions as reference for the other box sizes. Data used in this thesis: X-ray data was available for RS-I-1, RS-1-2, RS-1-4, RS-II-1, RS-II-2, RS-III-1 and RS-III-2. Microbiome and pain data was available for RS-III-2. Pharmacological data was avalible for RS-I, and RS-II.

Outline and aim of this thesis

The aim of this thesis is to identify novel genes involved in the pathogenesis of osteoarthritis, in order to gain greater insight into the pathology of osteoarthritis and bring possible preventive or curative treatments closer for one of the world's oldest diseases.

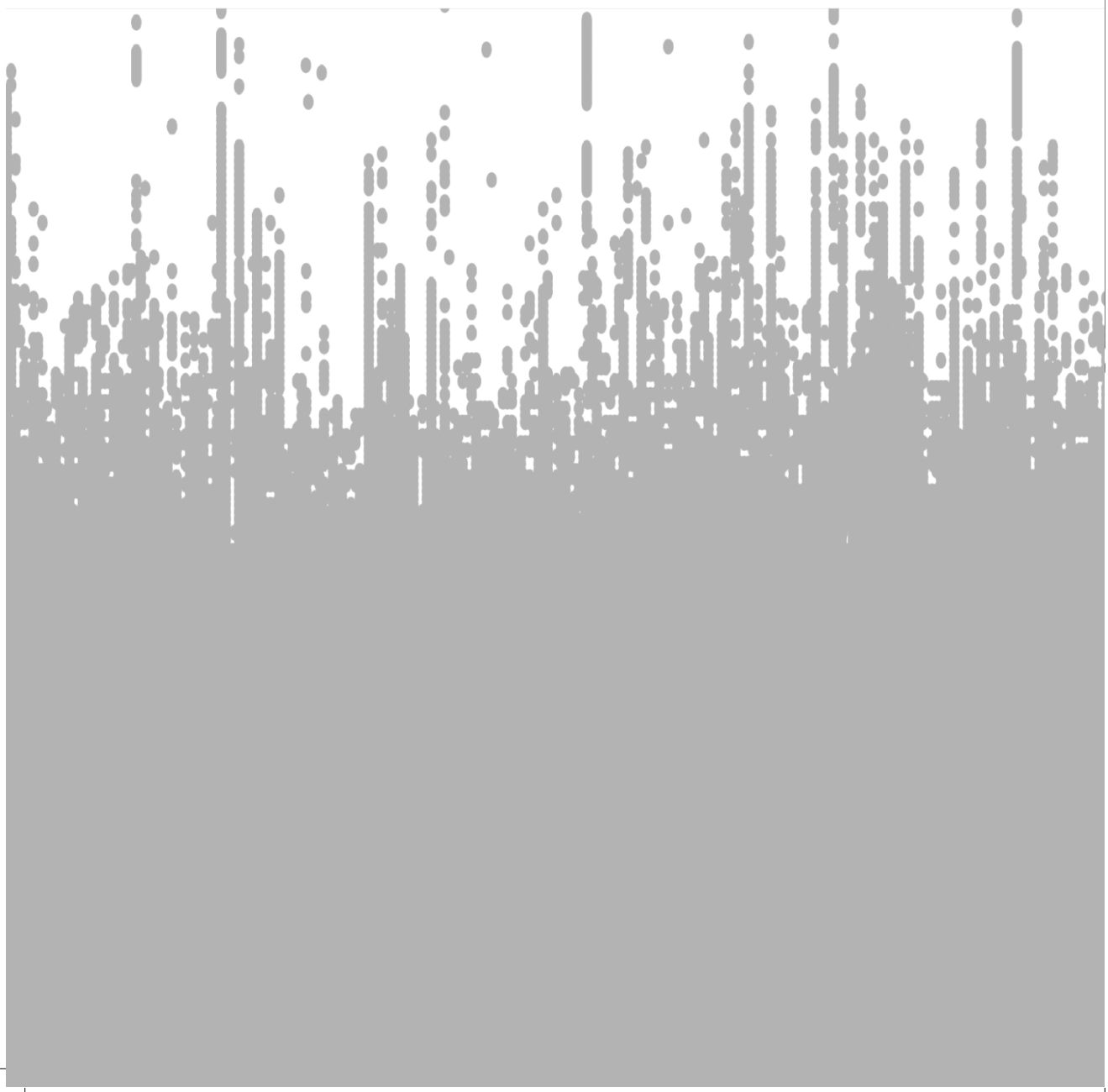
In **Chapter 2**, osteoarthritis endophenotypes and structural phenotypes are used to increase GWAS power and identify novel genes associated with osteoarthritis. Next, , in **Chapter 3** also the interaction with the environment is investigated, in particular the use of vitamin K antagonists, to identify novel, and perhaps modifiable, genetic influences associated with osteoarthritis. In **Chapter 4**, the results of the largest GWAS to multiple osteoarthritis phenotypes is presented, including functional follow-up to identity possible causal genes. Lastly, in **Chapter 5**, the possible influence of the gastrointestinal microbiome composition on osteoarthritis related knee pain and inflammation is investigated.

References

1. Rothschild, B.M., Radiologic assessment of osteoarthritis in dinosaurs Annals of Carnegie Museum 1990. 59(4): p. 295-301.
2. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet, 2018. 392(10159): p. 1789-1858.
3. Rothschild, B.M.X., Z.; Martina, L. D. and S. more, Osteoarthritis in the early avian radiation: Earliest recognition of the disease in birds. Cretaceous Research, 2012. 35: p. 178-180.
4. Rothschild, B.M.M., R. E. , Osteoarthritis in fossil marsupial populations in Australia Annals of Carnegie Museum, 1988. 57: p. 155-158.
5. Barbosa, F.H., et al., Arthritis in a glyptodont (Mammalia, Xenarthra, Cingulata). PLoS One, 2014. 9(2): p. e88646.
6. Brown, C., et al., Skeletal trauma reflects hunting behaviour in extinct sabre-tooth cats and dire wolves. Nat Ecol Evol, 2017. 1(5): p. 131.
7. Askary, A., et al., Ancient origin of lubricated joints in bony vertebrates. Elife, 2016. 5.
8. Woolf, A.D., Bone and Joint Features. 2002, London: BMJ Books.
9. Haeusler, M., et al., Morphology, pathology, and the vertebral posture of the La Chapelle-aux-Saints Neandertal. Proc Natl Acad Sci U S A, 2019. 116(11): p. 4923-4927.
10. Lawrence Angel, J., Early skeletons from Tranquillity, California. Smithsonian Contributions to Anthropology., 1966. 1.
11. Fritsch, K.O., et al., The orthopedic diseases of ancient Egypt. Anat Rec (Hoboken), 2015. 298(6): p. 1036-46.
12. Austin, A.E., The Cost of a Commute: A Multidisciplinary Approach to Osteoarthritis in New Kingdom Egypt. International Journal of Osteoarcheology 2016. 27(4): p. 537-550.
13. Plomp, K.A. and A. Boylston, Frequency and patterns of costovertebral osteoarthritis in two Medieval English populations. Int J Paleopathol, 2016. 14: p. 64-68.
14. Neogi, T., The epidemiology and impact of pain in osteoarthritis. Osteoarthritis Cartilage, 2013. 21(9): p. 1145-53.
15. Adams, F., The Genuine Works of Hippocrates. 1849, Sydenham Society: London.
16. Herben, W., Commentaries on the History and Cure of Diseases, ed. T. Payne. Vol. (Reprinted 1982 by The Classics of Medicine Library Division of Gryphon Editions Ltd. Birmingham, AL). 1802, London Mews-gate.
17. Spender , J.K., The Early Symptoms and the Early Treatment of Osteo-arthritis (commonly called rheumatoidarthritis. 1889, London: H.K. Lewis.
18. Brandt, K.D., et al., Yet more evidence that osteoarthritis is not a cartilage disease, in Ann Rheum Dis. 2006: England. p. 1261-4.
19. Hunter, D.P.-A., D.; Arden, N., Osteoarthrtis: the facts. 2nd ed. ed. 2014, Oxford: Oxford Univercity Press.
20. Swain, S., et al., Comorbidities in Osteoarthritis: A systematic review and meta-analysis of observational studies. Arthritis Care Res (Hoboken), 2019.
21. Hunter, D.J. and S. Bierma-Zeinstra, Osteoarthritis. Lancet, 2019. 393(10182): p. 1745-1759.
22. Kellgren, J.H. and J.S. Lawrence, Radiological Assessment of Osteo-Arthrosis. Ann Rheum Dis, 1957. 16(4): p. 494-502.
23. Altman, R.D., Criteria for the classification of osteoarthritis of the knee and hip. Scand J Rheumatol Suppl, 1987. 65: p. 31-9.
24. Kerkhof, H.J., et al., Recommendations for standardization and phenotype definitions in genetic studies of osteoarthritis: the TREAT-OA consortium. Osteoarthritis Cartilage, 2011. 19(3): p. 254-64.
25. Bierma-Zeinstra, S.M. and M. van Middelkoop, Osteoarthritis: In search of phenotypes. Nat Rev Rheumatol, 2017. 13(12): p. 705-706.
26. Spector, T.D., et al., Genetic influences on osteoarthritis in women: a twin study. Bmj, 1996. 312(7036): p. 940-3.
27. MacGregor, A.J., et al., The genetic contribution to radiographic hip osteoarthritis in women: results of a classic twin study. Arthritis Rheum, 2000. 43(11): p. 2410-6.
28. Ishimori, M.L., et al., Heritability patterns in hand osteoarthritis: the role of osteophytes. Arthritis Res Ther, 2010. 12(5): p. R180.

29. Watson, J.D. and F.H. Crick, Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature, 1953. 171(4356): p. 737-8.
30. Crick, F., Central dogma of molecular biology. Nature, 1970. 227(5258): p. 561-3.
31. Auton, A., et al., A global reference for human genetic variation. Nature, 2015. 526(7571): p. 68-74.
32. Lander, E.S., et al., Initial sequencing and analysis of the human genome. Nature, 2001. 409(6822): p. 860-921.
33. Sherry, S.T., M. Ward, and K. Sirotkin, dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. Genome Res, 1999. 9(8): p. 677-9.
34. McCarthy, M.I., et al., Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet, 2008. 9(5): p. 356-69.
35. The International HapMap Project. Nature, 2003. 426(6968): p. 789-96.
36. Mendel, G., Versuche über Plflanzen-hybriden. Verhandlungen des naturforschenden Ver-eines in Brünn. Abhand-lungen, 1866: p. 3*47.
37. Stecher, R.M. and A.H. Hersh, Herben's Nodes: the mechanism of inheritance in hypertrophic arthritis of the fingers. J Clin Invest, 1944. 23(5): p. 699-704.
38. Stecher, R.M., Herben's nodes. Heredity in hypertropic arthritis of the finger joints. am J Med Sci, 1941: p. 201-801.
39. Palotie, A., et al., Predisposition to familial osteoarthrosis linked to type II collagen gene. Lancet, 1989. 1(8644): p. 924-7.
40. Cicuttini, F.M. and T.D. Spector, The genetics of osteoarthritis. J Clin Pathol, 1996. 49(8): p. 617-8.
41. Visscher, P.M., et al., 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet, 2017. 101(1): p. 5-22.
42. van Meurs, J.B., Osteoarthritis year in review 2016: genetics, genomics and epigenetics. Osteoarthritis Cartilage, 2017. 25(2): p. 181-189.
43. Altshuler, D.M., et al., Integrating common and rare genetic variation in diverse human populations. Nature, 2010. 467(7311): p. 52-8.
44. Abecasis, G.R., et al., A map of human genome variation from population-scale sequencing. Nature, 2010. 467(7319): p. 1061-73.
45. Taliun, D., et al., Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. 2019: p. 563866.
46. Fall, T., et al., Evidence of large genetic influences on dog ownership in the Swedish Twin Registry has implications for understanding domestication and health associations. Sci Rep, 2019. 9(1): p. 7554.
47. Valdes, A.M., et al., Genome-wide association scan identifies a prostaglandin-endoperoxide synthase 2 variant involved in risk of knee osteoarthritis. Am J Hum Genet, 2008. 82(6): p. 1231-40.
48. Evangelou, E., et al., Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. Ann Rheum Dis, 2011. 70(2): p. 349-55.
49. Gottesman, II and T.D. Gould, The endophenotype concept in psychiatry: etymology and strategic intentions. Am J Psychiatry, 2003. 160(4): p. 636-45.
50. Castano Betancourt, M.C., 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(21): p. 8218-23.
51. Heinemeier, K.M., et al., Radiocarbon dating reveals minimal collagen turnover in both healthy and osteoarthritic human cartilage. Sci Transl Med, 2016. 8(346): p. 346ra90.
52. Berenbaum, F., Osteoarthritis as an inflammatory disease (osteoarthritis is not osteoarthrosis!). Osteoarthritis Cartilage, 2013. 21(1): p. 16-21.
53. Attur, M.G., et al., Osteoarthritis or osteoarthrosis: the definition of inflammation becomes a semantic issue in the genomic era of molecular medicine, in Osteoarthritis Cartilage. 2002: England. p. 1-4.
54. Scanzello, C.R. and S.R. Goldring, The role of synovitis in osteoarthritis pathogenesis. Bone, 2012. 51(2): p. 249-57.
55. Huang, Z. and V.B. Kraus, Does lipopolysaccharide-mediated inflammation have a role in OA? Nat Rev Rheumatol, 2016. 12(2): p. 123-9.
56. Huang, Z.Y., et al., Both systemic and local lipopolysaccharide (LPS) burden are associated with knee OA severity and inflammation. Osteoarthritis Cartilage, 2016. 24(10): p. 1769-1775.
57. Jackson, M.A., et al., Gut microbiota associations with common diseases and prescription medications in a population-based cohort. Nat Commun, 2018. 9(1): p. 2655.

58. Falony, G., et al., Population-level analysis of gut microbiome variation. Science, 2016. 352(6285): p. 560-4.
59. Turnbaugh, P.J., et al., An obesity-associated gut microbiome with increased capacity for energy harvest. Nature, 2006. 444(7122): p. 1027-31.
60. Backhed, F., et al., The gut microbiota as an environmental factor that regulates fat storage. Proc Natl Acad Sci U S A, 2004. 101(44): p. 15718-23.
61. Silverwood, V., et al., Current evidence on risk factors for knee osteoarthritis in older adults: a systematic review and meta-analysis. Osteoarthritis Cartilage, 2015. 23(4): p. 507-15.
62. Saberi Hosnijeh, F., et al., Development of a prediction model for future risk of radiographic hip osteoarthritis. Osteoarthritis Cartilage, 2018. 26(4): p. 540-546.
63. Jiang, L., et al., Body mass index and hand osteoarthritis susceptibility: an updated meta-analysis. Int J Rheum Dis, 2016. 19(12): p. 1244-1254.
64. Wellen, K.E. and G.S. Hotamisligil, Obesity-induced inflammatory changes in adipose tissue. J Clin Invest, 2003. 112(12): p. 1785-8.
65. Tilg, H. and A. Kaser, Gut microbiome, obesity, and metabolic dysfunction, in J Clin Invest. 2011. p. 2126-32.
66. Ikram, M.A., et al., The Rotterdam Study: 2018 update on objectives, design and main results. Eur J Epidemiol, 2017. 32(9): p. 807-850.



Chapter 1.2

Role of Epigenetics in Bone and Cartilage

van Meurs J.B.J., **Boer C.G.**, Lopez-Delgado L., Riancho J.A.

Published in: J Bone Miner Res. 2019 Feb;34(2):215-230.

Abstract

Phenotypic variation in skeletal traits and diseases is the product of genetic and environmental factors. Epigenetic mechanisms include information-containing factors, other than DNA sequence, that cause stable changes in gene expression and are maintained during cell divisions. They represent a link between environmental influences, genome features, and the resulting phenotype. The main epigenetic factors are DNA methylation, posttranslational changes of histones, and higher-order chromatin structure. Sometimes non-coding RNAs, such as microRNAs (miRNAs) and long non-coding RNAs (lncRNAs), are also included in the broad term of epigenetic factors. There is rapidly expanding experimental evidence for a role of epigenetic factors in the differentiation of bone cells and the pathogenesis of skeletal disorders, such as osteoporosis and osteoarthritis. However, different from genetic factors, epigenetic signatures are cell- and tissue-specific and can change with time. Thus, elucidating their role has particular difficulties, especially in human studies. Nevertheless, epigenome-wide association studies are beginning to disclose some disease-specific patterns that help to understand skeletal cell biology and may lead to development of new epigenetic-based biomarkers, as well as new drug targets useful for treating diffuse and localized disorders. Here we provide an overview and update of recent advances on the role of epigenomics in bone and cartilage diseases.

Epigenomics as a Link Between Environment, Genotype, and Phenotype

Phenotypic variation in skeletal traits and diseases is the product of genetic and environmental factors. The extent to which genetics shapes the phenotype is different across the different skeletal conditions. Nevertheless, the genetic component of skeletal traits is large, varying from 30% (knee osteoarthritis [OA]) to 80% (bone mineral density [BMD]). This does not mean there is no effect of the environment. In fact, these DNA-sequence variants form the template upon which environmental factors can influence the phenotype, by a number of mechanisms, including epigenetic marks. Waddington coined the term epigenetics to describe the interactions between the environment and the genes leading to the development of phenotype[1]. The modern definition of epigenetic mechanisms includes information-containing factors, other than DNA sequence, that cause stable changes in gene expression and are maintained during cell divisions[2].

The main epigenetic factors are DNA methylation, posttranslational changes of histones, and higher-order chromatin structure. Sometimes non-coding RNAs, such as microRNAs (miRNAs) and long non-coding RNAs (lncRNAs), are also included in the broad term of epigenetic factors. However, the exact definition of epigenetics and its components is still a matter of controversy[3]. Together these different epigenetic mechanisms are key factors behind the regulation, function, and cell fate of all tissues and cells. Not surprisingly, there is a large body of evidence supporting the role of epigenetics in skeletal development, the maintenance of bone mass, and skeletal disorders. Our purpose here is to provide an overview and update of recent advances on the role of epigenomics in bone and cartilage diseases.

Epigenomic marks

DNA methylation

DNA methylation refers to the covalent addition of a methyl group to cytosines in DNA, particularly when they are part of CpG dinucleotides. In somatic cells, more than 80% CpGs are methylated, especially in repetitive sequences in intergenic regions and introns, whereas CpGs in gene promoters may be methylated or not. In general, the methylation of CpGs in gene promoters is associated with repression of gene expression, whereas the methylation of gene bodies and other regulatory regions (such as enhancers)[4] has a less predictable effect. Up to 20% of the variation in DNA methylation is influ-

enced by genetic variation, but the majority of the variation in methylation is caused by other factors, including environmental and stochastic variation. Indeed, variation in DNA methylation increases with age[5] and is thought to have a role in the relation between environmental risk factors and disease risk. As all epigenetic marks, DNA methylation is dynamic; methyl groups can be added and removed from the DNA by specialized proteins, DNA-methyl-transferases (DNMTs) and ten-eleven translocation (TET) proteins, respectively (see **Box 1** for an explanation of epigenomic terms).

Chromatin structure

The spatial organization of the DNA itself in the cell nucleus, the chromatin structure, is also important for the functional read-out of the genome. Histones are critical components of the chromatin (**Figure 1**). In fact, DNA-bound histones play major roles in the regulation of gene transcription. Posttranslational modifications (PTM) of specific amino acids in the N-terminal tail of histones, such as methylation, phosphorylation, acetylation, and ubiquitylation, remodel the shape of the chromatin. This, in turn, alters the DNA's accessibility for proteins involved in the transcription machinery, thereby regulating gene expression[6]. Histone PTMs are dynamic and a number of enzymes are able to add or remove histone marks. For example, acetyl groups can be added by histone acetylases (HATs) and removed by histone deacetylases (HDACs).

Next to histone PTMs, also the spatial organization of the chromatin itself in the nucleus can modulate gene expression. Chromosome-conformation capture techniques have shown that the genome is divided into so-called topological associated domains (TADs), which are large (megabase scale) compartments of the genome. These regions interact more frequently with themselves than the rest of the genome and enhancers usually contact genes located within these TADs but not outside[7]. Distant enhancers and their target gene promoters are brought into contact with each other using the formation of so-called "DNA-loops," mediated by, for example, CTCF and cohesins (**Figure 1**).

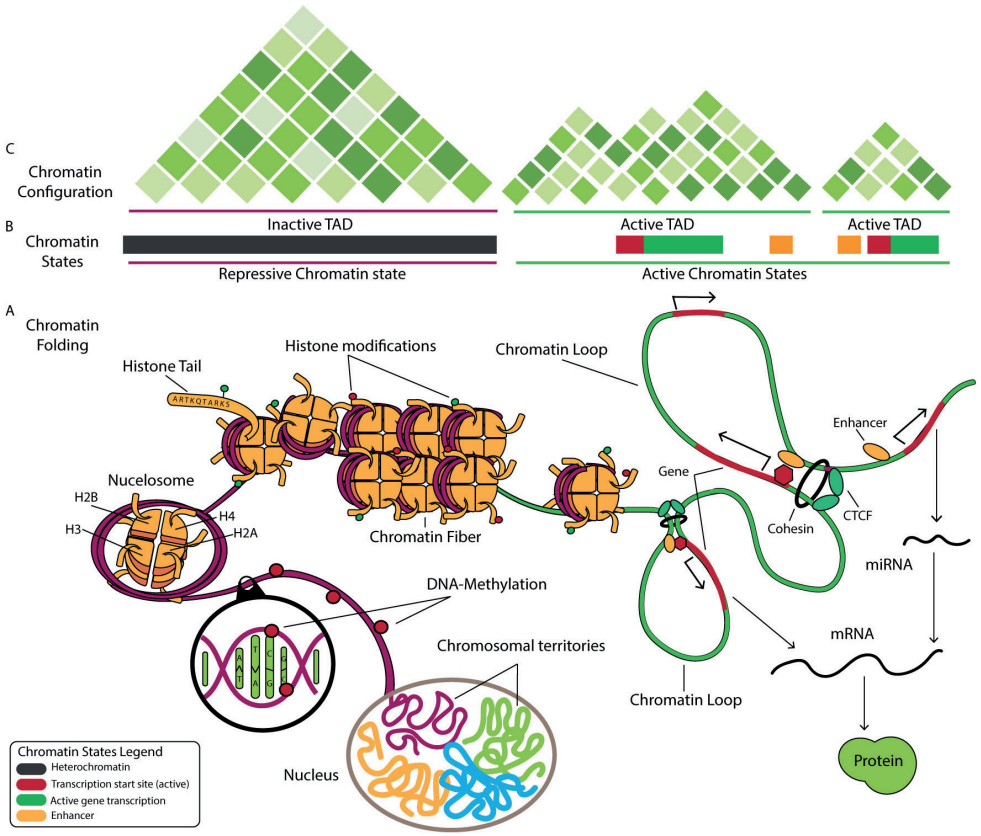
Non-coding RNAs

Besides chromatin-related marks and structure, non-coding RNAs (ncRNAs) are also frequently included among the mechanisms of epigenetic control[8]. They are classified as small RNAs (<200 nucleotides) and long RNAs (>200 nucleotides). The best-known subset of small RNAs are microRNAs (miRNAs, 18 to 25 nucleotides), which inhibit protein synthesis by binding to the 3′-untranslated region of target mRNAs. Long non-coding RNAs (lncRNAs) modulate the activity of both nearby genes and distant genes by a variety of mechanisms. For instance, they often serve as scaffolds for transcription factors and other molecules involved in initiation of transcription, including repressive

chromatin modifiers such as polycomb repressive complex proteins (PRC1 and PRC2) or activating chromatin modifiers[9]. Some lncRNAS are mainly located in the cytosol, where they target mRNAs and downregulate protein translation. Interestingly, they may also act as decoys for miRNAs, thus preventing the inhibitory effect of the binding of miRNAs to their target mRNAs[10].

Box 1: Epigenomics-Related Terms

- **Chromatin:** The complex of DNA and its packaging molecules. The core of the chromatin is the nucleosome, which consists of an octamer of 4 histones around which 147 bp of DNA is wrapped around.
- **Chromosome conformation capture and Hi-C:** Techniques used to map the spatial (3D) organization of the chromatin in the nucleus. Chromosome conformation capture (3C) quantifies the number of interactions between a given loci and the rest of the genome. In Hi-C, all genomic interaction between all genomic regions are quantified.
- **CTCF**: CCCTC-binding factor, highly conserved zinc finger protein involved in diverse genomic regulatory functions, including transcriptional activation/repression, insulation, imprinting, and X-chromosome inactivation, through mediating the formation of chromatin loops.
- **DNA methyl-transferases (DNMTs)**: Family of enzymes responsible for the methylation of DNA. DNMT1 recognizes hemimethylated CpG sites on newly replicated DNA and thus it maintains the methylation pattern through cell divisions. On the other hand, DNMT3A/3B are the novo methylases, capable of converting unmethylated CpGs into methylated CpGs in double-strand DNA, which is particularly important during embryogenesis and cell differentiation.
- **Epigenetics:** Mechanisms causing changes in gene expression that are heritable through cell divisions and do not include modifications of DNA sequence.
- **Epigenome-wide association study (EWAS):** Studies of the relationship between many epigenetic marks distributed throughout the genome and phenotypic characteristics. So far, most studies aimed to analyze DNA methylation.
- **Epigenomics:** Usually refers to the epigenetic changes in many genes or even through the whole genome.
- **Genome-wide association study (GWAS):** Studies of the relationship between many genetic variants (usually hundred thousands or millions) distributed throughout the genome and phenotypic characteristics.
- **Histone code**: Hundreds of different posttranslational modifications (PTMs) and their combinatorial patterns form a code, the histone code. This code can give rise to a prescribed transcriptional or other genomic regulatory response, interpreted by specialized proteins that can read, write, and erase histone PTMs.
- **Histone deacetylases (HDACs):** Family of enzymes removing acetylation marks from histone tails. These epigenetic "erasers" are very important for modulating gene expression because histone acetylation is usually associated with active chromatin.
- **Long non-coding RNAs (lncRNAs):** ncRNAs with more than 200 nucleotides that regulate gene expression and interact with other epigenetic mechanisms.
- **Methylation quantitative trait loci (meQTL):** A DNA locus (usually a single-nucleotide polymorphism [SNP]) that is associated with DNA-methylation levels from a certain CpG.
- **MicroRNAs (miRNAs):** Small ncRNAs, 18 to 25 nucleotides long. One known function (of the many that are known) of miRNA's target mRNAs, which interferes with protein translation.
- **Non-coding RNAs (ncRNAs):** RNAs that do not code proteins but have regulatory roles on chromatin structure, gene expression, or translation. There are multiple types, with different sizes and functions.
- **Quantitative trait loci (QTL):** A DNA locus that is associated with a particular quantitative phenotypic trait.
- **Ten–eleven translocation (TET):** Family of proteins involved in the demethylation of CpGs.
- **Topologically associated domain (TAD):** Chromatin conformation capture techniques, such as Hi-C, showed that the genome was divided in compartments that interact more frequently with themselves than the rest of the genome. Regulatory regions, such as enhancers, usually contact genes located within the same TAD as the regulatory region but not outside of their TAD.



▲ **Figure 1: Schematic overview of epigenetics**: **chromatin modifications and structure.** (A) The DNA in the nucleus is present in the form of chromosomes. These inhabit distinct territories within the nucleus, allowing for the formation of distinct intra- or interchromosomal contacts. DNA methylation is the chemical addition of a methyl group to cytosines in DNA. The chromatin is organized in nucleosomes made of DNA wrapped around an octamer of 4 histones (H2A, H2B, H3, and H4). Posttranslational modifications (PTM) of specific amino acids in the N-terminal tail of histones, such as methylation, phosphorylation, acetylation, and ubiquitylation, remodel the shape and subsequently the function of the chromatin into repressive and active chromatin. Chromatin loops enable distant enhancers to come into close contact with their target gene promoters or create regions of gene silencing. The proteins CTCF and cohesin are known to be involved in the mediation of such chromatin loops. (B) Histone tails can contain many different or even multiple PTMs. This combination of PTMs, the histone code, confers meaning on the function or the state of that particular part of the chromatin. For several histone PTMs, their chromatin state is known, such as H3K4me1 for active enhancers and H3K4me3 for active promoters. (C) Chromatin conformation capture techniques, such as Hi-C, have revealed the compartmentalization of the genome into topologically associated domains (TADs), which are regions of preferential chromatin interactions. Depicted here is a stylistic interpretation of Hi-C data, where the darker colored the bloc between two genomic regions, the more genomic interactions are quantified.

Epigenomic interplay

It is worth emphasizing that epigenetic mechanisms often act in concert by interacting with each other. For example, MeCP2, a protein recognizing methylated CpGs, promotes the activity of HDACs. On the other hand, some histone marks modulate the binding of

DNMTs and subsequently DNA methylation. The methylation of promoters regulates the transcriptional activity not only of protein-coding genes but also of miRNAs and other non-coding RNAs. In turn, miRNAs contribute to modulating the synthesis of DNMTs and histone-modifying enzymes. lncRNAs also influence the activity of genes encoding chromatin-modifying enzymes and miRNAs[10]. Although the sequence of molecular steps is still unclear, there is evidence for the notion that DNA, RNA, and histone proteins, along with their modifications, act in a concerted fashion to bring about chromatin states that are important for dictating genomic functions[8] (**Figure 1**).

Epigenetic programming during development

From the moment of conception to adulthood, the environment shapes the phenotypic output. It is thought that there are certain "high sensitive" windows especially during development that have major influence on the epigenome[11]. The developmental origins of health and disease concept suggests that poor developmental experience can increase the risk of non-communicable diseases in later life, including cardiovascular, metabolic, neurological, and skeletal disorders[12]. A variety of mechanisms, including DNA methylation and other long-lasting epigenetic marks, may mediate the influence of the environment on the developing organism[13, 14]. A few studies have explored the role of developmental factors in skeletal disorders. In a systematic review of the literature, a positive association between birth weight and bone mass was clear among children, unclear among adolescents, and weak among adults. The effect was stronger on bone mineral content (BMC) than on BMD regardless of age[15]. This suggests that intrauterine growth is more closely related to bone size than to bone density and that the effect tends to be mitigated by postnatal influences. It seems that early life exposures are important for determining peak bone mass, which may be a reflection of the combined influence of intrauterine and early postnatal environmental exposures.

Maternal nutrition and specifically the maternal vitamin D status may be a critical factor for an adequate intrauterine growth rate[16], but studies have shown conflicting results.[17]. Rather surprisingly, in the Rotterdam cohort, severe maternal 25(OH)D deficiency (<25 nmol/L) during mid-pregnancy was associated with higher offspring BMC and bone area at 6 years of age, while no associations were found between maternal vitamin D status and offspring BMD[18]. In experimental animals, vitamin D status has a transgenerational effect on the methylation of multiple genes[19]. However, human studies about the relationship between maternal vitamin D levels and DNA methylation in offspring have given controversial results[16]. Therefore, the actual relevance of maternal vitamin D on DNA methylation and the bone mass of the offspring is still unclear.

Whether related to parental influences or not, a few studies reported an association of the methylation of some genes (such as *NOS, RXRA*, and *CDKN2A*) in cord blood and childhood bone mass[20-22]. However, those results have not been confirmed in other cohorts yet. Although less studied than the relationship between early life experiences and osteoporosis, some data support a developmental component in OA. For example, exposure to Chinese famine during childhood has been associated with arthritis (including both OA and inflammatory arthritis) in later life[23]. Similarly, in a British study, lower weight at birth and year 1 was associated with higher rates of OA[24]. Weight and body length differences, which have a clear developmental component, may explain, at least in part, those associations. Another example is finger length pattern, which is thought to be an indicator of prenatal androgen exposure. Type 3 finger length pattern (longer fourth digit than second digit) has been associated with having symptomatic knee OA and chronic pain[25]. which might be explained by an influence of embryogenic sex hormone exposure on brain development[26]. It is thought that epigenetic programming plays a role in all of these associations, but the exact role of epigenetic factors in the relation between prenatal exposures and skeletal diseases has not been elucidated yet.

Epigenomic plasticity in adult life

In addition to the developmental epigenetic programming, there is an enormous epigenomic plasticity in adult life. The variance in epigenetic marks increases with age, which is thought to reflect the response to environmental exposures in such a way that they modulate the expression of genes. However, studies in highly inbred rodent lines highlighted that a part of the phenotypic variation could not be attributed to environmental exposures[27]. Also, monozygotic twin studies examining discordances have shown that part of the phenotypic variation is attributed to so-called “stochastic” variation, possibly caused by “molecular noise” due to imperfect control of the molecular interactions in the cell[28]. Stochasticity, or random variation, is thought to have a large impact on disease susceptibility[29, 30].

DNA Methylation and Skeletal Disorders

DNA methylation and the differentiation of skeletal cells

Osteoclast precursors derive from hematopoietic stem cells, whereas the bone- and cartilage-forming cells, osteoblasts and chondrocytes, derive from mesenchymal stem

cells (MSCs). Pluripotent MSCs can also differentiate into other cell types, such as adipocytes and myocytes. The differentiation of MSCs toward the osteoblastic lineage is induced by the master transcription factors *RUNX2* and *osterix* and is stimulated by ligands of the *Wnt* and *BMP* pathways[31]. As it happens in other tissues, epigenetic factors play critical roles in determining the fate of MSCs. Specifically, genes that are characteristic of the osteoblast-osteocyte lineage (such as *alkaline phosphatase, sclerostin, RANKL, osteoprotegerin*, etc.) tend to undergo demethylation and de-repression during the differentiation of MSCs[32-34]. In line with this concept, the demethylating agents 5-azacytidine and 5-deoxy-azacytidine improve the osteogenic differentiation of MSCs[35, 36]. The ability of MSCs to proliferate and differentiate may decrease with aging, a phenomenon that may be explained, at least partially, by changes in the methylation and hydroxymethylation of DNA[37-39].

The differentiation of osteoclast precursors is associated with marked changes in their DNA methylation signature, with hypermethylation and hypomethylation occurring in a variety of gene categories. In this process, PU.1 may play an important role, by recruiting DNMT3B to hypermethylated promoters, and TET2, which converts 5-methylcytosine to 5-hydroxymethylcytosine, to genes that become demethylated[40]. Another DNA methyltransferase, DNMT3A, is essential to methylate and repress anti-osteoclastogenic genes, thus allowing osteoclast differentiation to continue[41]. On the other hand, by influencing the expression of *RANKL* and *OPG* in cells of the osteoblast-osteocyte lineage, DNA methylation indirectly contributes to regulating osteoclastogenesis[33]. Although cartilage does not undergo a remodeling process as bone does, epigenetic mechanisms also contribute to regulation of chondrogenesis and cartilage maintenance. DNA methylation modulates the expression of genes involved in cartilage homeostasis, such as *GDF5*, *SOX9*, and *MMP13*[42].

A number of studies have explored the relation between genome-wide methylation patterns and skeletal disease status in humans. These so-called epigenome-wide association studies (EWAS) have a hypothesis-free approach that have the potential to find novel genes and/or pathways involved in skeletal disease. Because epigenetic marks are tissue specific, it makes sense to perform these studies in the tissue of interest. For osteoporosis, this would be bone, but for osteoarthritis, the tissue of interest is less straightforward because there are many tissues involved, including cartilage, bone, and synovial tissue. In addition, epigenetic marks in the circulation (blood) could be indicative of systemic mechanisms playing a role in disease and could also be easily accessible biomarkers for the disease. An overview of published studies in the field of osteoarthritis and osteoporosis is given in **Table 1.**

Table 1: Summary of Epigenomewide Studies of DNA Methylation in Osteoporosis and Osteoarthritis

Joint trait	# and type of samples	Tissue/Cell type	Adjustments	Technique	Results	Pathway enrichment	Ref.
Knee/hip OA	24 intact vs damaged	Cartilage	No	450K array	550 DMS	SMAD3, angiogenesis, inflammation	[46]
Knee/hip OA	31 intact vs damaged	Cartilage	Sex, age, BMI, disease status	450K array	6272 DMS, 357 DMR	Skeletal development	[45]
Knee/hip OA	Intact vs damaged 12 discovery, 26 replication	Cartilage	No	450K array	9896 DMS, 271 DMR	Skeletal system development	[43]
Knee and hip OA and neocartilage	12 cartilage vs 8 neocartilage from MSCs	Cartilage	No	450K array	5706 DMS	Developmental and transcription regulation processes	[51]
Knee OA	5 intact vs damaged	Cartilage	No	Agilent Promoter array	1214 promoters	Development, TGFbeta-, MAPK, hedgehog	[47]
Knee OA	10 intact vs damaged	Cartilage	No	RRBS	>1000 DMRs	Embryogenesis and skeletal development	[44]
Knee OA progression	12 early vs interm; early vs late	Cartilage Subchondral bone	No	450K array	0;519 DMS 72;397 DMS	Skeletal system development Morphogenesis, skeletal development, HOX	[142]
Knee OA and controls	18 controls vs 23 OA	Cartilage	No	27K array	91 DMS	Inflammation, transcription activity, phosphorylation, MAPK	[50]
Knee OA and controls	11 controls vs 12 OA	Cartilage	Sex, BMI (age)	450K array	929 DMS (0 after age adjustment)	Integrin, Wnt, FGF	[143]
Knee OA and controls	4 TKR vs 4 controls	Chondrocytes	No	MeDIP seq	70,591 DhMR	Many, including Wnt, bone remodeling, inflammation	[144]

Table 1 (continued): Summary of Epigenomewide Studies of DNA Methylation in Osteoporosis and Osteoarthritis

Joint trait	# and type of samples	Tissue/Cell type	Adjustments	Technique	Results	Pathway enrichment	Ref.
Hip OA	9 osteophytes vs cartilage	Chondrocytes	No	450K array	3161 DMR	Extracellular matrix, skeletal system development, endodermal cell differentiation	[145]
Hip fracture and OA	22 fracture vs 17 OA	Bone marrow MSCs	Age	450K array	9038 DMS	Only enhancer CpGs: stem-cell development, osteoblast differentiation, Wnt pathway	[58]
Hip fracture and OA	7 hip fracture vs 11 OA	Cartilage	No	450K array	103 DMS	Embryonal skeletal development, HOX	[48]
Hip fracture and OA	27 fracture vs 26 OA	Bone	Age	27K array	241 DMS	Neuron differentiation, glycoprotein, skeleton, Homeobox	[57]
OP	40 normal BMD vs 26 OP	Bone Blood	No Age, sex, cell counts, batch	450K array	63 13	Not reported	[59]
OP	22 normal BMD vs 12 early OP/10 advanced OP	Blood	No	450K array	1233	Not reported	[61]
BMD	4614 discovery; 901 replication	Blood	Age, weight, sex, smoking, family structure, cell counts, batch	450K array	0	NA	[146]

OA = osteoarthritis; DMS = differentially methylated site; BMI = body mass index; DMR = differentially methylated region; DhMR = differentially hydroxymethylated region; RRBS = reduced representative bisulphite sequencing; OP = osteoporosis; BMD = bone mineral density.

Epigenomewide studies in osteoarthritis

A relatively large number of EWAS have been done focusing on the differences in the methylome of diseased versus preserved cartilage[43-47], cartilage from hip fracture[48, 49], controls[48, 50], or neocartilage engineered from MSCs[51] (**Table 1**). Although the exact genes that are identified in the studies can differ, the identified pathways that emerge are robust. They involve skeletal development and morphogenesis and known signalling pathways (such as the *TGFβ, Wnt, HOX*) involved in skeletal development. Also, inflammation seems to be a pathway that is identified by multiple studies and two studies have also shown the possibility of clustering a subset of patients characterized by methylation differences in (or near) inflammatory genes[50, 52]. It is interesting to note that genome-wide methylation studies consistently found enrichment of differentially methylated CpGs in enhancer regions[53]. This suggests that modifications of DNA methylation marks may be more important in distant regulatory regions than in proximal promoters. This observation is consistent with other EWAS of complex diseases, where most of the associations found have been highly enriched in enhancer regions. Good annotation and interpretation of these enhancers are therefore of great importance. Methylation patterns can also be used to calculate a so-called "epigenetic age," which is thought to reflect biological aging and has been linked to a whole range of age-related diseases and time of death[54, 55]. One study examined the relation between this "epigenetic clock" and osteoarthritis and observed accelerated aging in OA cartilage[56].

Epigenomewide studies in osteoporosis

The number of EWAS studies that examine osteoporosis as an endpoint is much lower than the cartilage studies described above. The first study examined femoral head bone tissue from hip fracture and OA patients using a limited set of methylation sites, and showed among a number of pathways, differentially methylated regions in the family of HOX-genes[57]. A second, more recent study, examined MSCs in fracture versus OA patients and identified a number of differentially methylated genes in stem cell and osteoblast differentiation pathways[58]. In another study using bone biopsies of women with low (osteoporotic) or normal BMD[63]. differentially methylated CpGs were found at a lenient false discovery rate (FDR<0.1)[59]. Because bone is a multicellular tissue, methylation differences could also reflect the changes in cell-type proportions, which was not accounted for in the published studies.

Methylation signatures in the circulation can potentially be powerful biomarkers for disease. A large EWAS examining methylation patterns of circulating leukocytes was

performed in a collaborative study examining the relationship with BMD in a total of 5,515 participants[60]. However, this did not result in robustly associated CpGs, suggesting that blood might not be the correct tissue to study the epigenomic features in relation to bone phenotypes. In contrast, another small study examining 22 controls versus 22 osteoporotic women (defined by low BMD) identified a large number of differentially methylated sites[61]. Importantly, this study lacked replication and adjustment of several important possible confounders, such as cell counts and BMI.

Non-coding RNAs in Skeletal Disorders

Among ncRNAs, miRNAs have been most extensively studied in relation to bone and cartilage diseases. More than 35 miRNAs modulate the differentiation of osteoblast precursors in vitro, through various mechanisms, including targeting master regulators such as RUNX2 (see recent reviews)[62-64]. A few of them have demonstrated effects on bone in animal models in vivo and may be involved in osteoclast-osteoblast communication. For example, osteoclast-derived exosomal miR-214-3p inhibits osteoblast activity in vitro and reduces bone formation in vivo[65]. The miR-34 family and miR-214 also tend to have a negative influence on bone formation[66]. On the other hand, miR-2861 and miR-29a stimulate bone formation. Interestingly, they target HDACs, thus illustrating the interactions between different layers of epigenetic marks[67]. As with the osteoblastic lineage, several miRNAs influence osteoclast differentiation in vitro. Some of them have been validated in vivo. For instance, miR-503, which targets RANKL, and miR-34a inhibit bone resorption in animal models, whereas miR-148a tends to stimulate resorption[62, 64, 68]. Some miRNAs are abundant in cartilage and appear to be important for the regulation of metalloproteases, toll-like receptor signalling, and other genes involved in catabolic pathways[69]. A recent review of 57 studies about miRNA expression in cartilage revealed 46 differentially expressed miRNAs in OA, which were involved in autophagy, chondrocyte homeostasis, and degradation of the extracellular matrix[70]. For instance, miR-140 is involved in the pathogenesis of OA by regulating, at least in part, MMP13 and ADAMTS5. miR-140 is downregulated in OA cartilage and the intra-articular injection of miRNA-140 alleviates OA progression in rats[71]. The potential role of other miRNAs in OA pathogenesis has been recently reviewed[42, 72]. In a few cases, their effects have been validated by gain-of-function or loss-of-function experiments in vivo (**Table 2**).

Table 2: Selected miRNAs With Effects on Bone or Osteoarthritis Shown in Loss-of-Function or Gain-of-Function Experiments In Vivo

Positive effects on bone formation or bone mass [ref.]	Negative effects on bone formation or bone mass [ref.]	Amelioration of osteoarthritis [ref.]	Aggravation of osteoarthritis [ref.]
miR-135 [147]	miR-103 [148]	miR-140-5p [149]	miR-101 [150]
miR-141 [151]	miR-125b [152]	miR-142-3p [153]	miR-181a-5p [138]
miR-145[a] [154]	miR-133 [155]	miR-210-5p [156]	miR-221 [157]
miR-148 [158]	miR-138-5p [159]	miR-370 [160]	
miR-199a-5p [161]	miR-140 [76]	miR-373 [160]	
miR-21[a] [74]	miR-145a [162]	miR-98 [163]	
miR-216a [164]	miR-146 [165]		
miR-26a[a] [166]	miR-148 [167]		
miR-29b-3p [168]	miR-208 [169]		
miR-335 [77]	miR-21[a] [170]		
miR-34a-5p [171]	miR-214-3p [66]		
miR-503 [172]	miR-222 [75]		
miR-286 [167]	miR-23b [173]		
miR-375 [174]	miR-26a[a] [175]		
	miR-31-5p [78]		
	miR-341 [76]		
	miR-3831 [77]		

[A] Some miRNAs have shown contradictory effects in different models.

Use of miRNAs to diagnose and treat disease

The translational potential of miRNAs is large, since they can potentially be used to directly treat disease. A number of miRNAs enhance the osteogenic differentiation of MSCs; hence, they have been pointed out as potential therapies to promote bone regeneration. For example, scaffolds and particles loaded with analogs of miR-21 or miR-148 potentiate bone regeneration in experimental models, at least in part, by interacting with the RUNX2 pathway[73, 74]. On the other hand, anti-sense oligonucleotides and "sponges" inhibiting miR-22, miR-29, miR-31, miR-133, miR-138, or miR-214 have a stimulatory effect on osteogenesis and may improve fracture healing[62, 75, 76]. Also, miRNA-engineered MSCs may find a role in bone regeneration[77, 78]. miRNA-based therapies have also been explored in joint disorders. In this view, the intra-articular injection of lentiviruses expressing miR-140 and miR-210 ameliorated joint disease in experimental models of OA[79].

Many cells can secrete miRNAs (mainly via exosomes) that are detectable in the synovial fluid and the circulation. Hence, several miRNAs have been suggested as potential biomarkers for the diagnosis or follow-up of skeletal disorders (see recent reviews)

[72, 80-82]. However, in general, the identity of regulated miRNAs varies widely across the studies and sometimes conflicting evidence is found. In addition, given the absence of replication across study outcomes and the small sample sizes, these studies should be considered as exploratory and further replication and validation is warranted. It is also to mention that circulating levels of miRNAs reflect processes in the whole body, most notably those of blood cells, and therefore miRNA signatures could be influenced by co-morbid conditions and other circumstances (such as aging in general, low-grade inflammation, etc.)[80]. Thus, published data are promising, but more work is needed before miRNAs can serve as robust diagnostic and prognostic tools in the clinic.

Histones and Chromatin Structure in Skeletal Disorders

Evidence is slowly being accumulated for a role of histone PTMs in chondrocyte and osteoblast differentiation, skeletal development, and the pathogenesis of OA and osteoporosis[83-88]. There is a well-established role for HATs and HDACs in chondrogenesis, involving the regulation and function of *Sox9* and its downstream targets. *Sox9* is an essential regulator of chondrocyte differentiation and homeostasis[89]. The HDAC KDM4B mediates *Sox9* activation by removing a repressive histone PTM (namely, methylation of lysines 9 in histone 3; H3K9me3) from the *Sox9* promoter region, which in turn triggers TGF-β-mediated chondrogenesis.[90] Regulation of downstream targets of *Sox9* is also dependent of histone remodelers. The HDAC Sirt1 binds to the enhancer and promoter region of *COL2A1*, which in turn recruits the HATs p300/CBP[91]. These form a complex with *Sox9* at the *COL2A1* promoter. By relaxing the chromatin structure near the *COL2A1* promoter by P300/CBP through their HAT activity, Sox9 is able to initiate *COL2A1* expression[92, 93]. Recently, the HDAC KDM6B has also been suggested to regulate *COL2A1* expression, possibly also by direct interaction with the promoter of *COL2A1*[94]. Conversely, the HDAC ELF3 suppresses *COL2A1* transcription by inhibiting the HAT activity of the Sox9/CBP complex[95, 96].

Many other histone remodelers are being implicated in osteoblast and chondrocyte development[83, 97]. Recently, KDM6B, a HDAC, has been shown to be significantly increased in expression during cartilage development[94]. Also, identified through genome-wide association studies (GWAS), *DOT1L*, a histone lysine methyltransferase, might be involved in chondrocyte differentiation and homeostasis[98, 99]. The proteins that "read" the histone PTMs are also important for cell differentiation. Proteins from the bromodomain and extra-terminal domain (BET) family recognize and bind to acetylated lysine on histones, forming a scaffold for protein complexes involved in gene transcription[100, 101]. The inhibition of BET proteins leads to a suppression of os-

teoclast differentiation and activity in vitro[100]. Because of these seemingly key roles, HDACs, HAT, and BET have been suggested as novel therapeutic targets in a range of skeletal diseases[97, 102]. Thus, JQ1 and other inhibitors of BET proteins ameliorate bone loss in ovariectomized mice and several preclinical models of inflammatory disorders, such as arthritis and periodontitis[102, 103].

However, it is important to keep in mind that most histone remodelers and readers also fulfil essential roles in other tissues and cell types[104]. For example, DOT1L is not only essential for chondrogenesis and cartilage homeostasis[99] but, also for telomere silencing, meiotic checkpoint control, and DNA damage response[105, 106]. Murine knockout models of *DOT1L* show multiple developmental abnormalities, not only restricted to skeletal abnormalities[106]. Similar observations can be made for most histone remodelers associated with skeletal development and disease, such as the sirtuins (SirT1)[107], KDM6B[108], and BET proteins[109]. Because histone remodelers and readers are involved in a plethora of cellular processes in diverse cell types[110], globally administered therapeutic targeting may produce off-target and side effects, illustrating the need to examine such histone remodelers and readers in careful detail.

The chromatin structure itself can also modulate gene regulation and expression. For example, disruption of the TAD structure near certain genes can cause congenital skeletal disorders. Depending on the type and size of the of TAD disruption, brachydactyly, syndactyly, and polydactyly may be caused by changes in enhancer-promoter regulation in the *WNT6/IHH/EPHA4/PAX3* locus[111]. Further, the deletion of a TAD boundary as a disease mechanism has also been proposed for Liebenberg syndrome, a rare disorder where the arms of the patient acquire morphological characteristics similar to those of the legs[112]. For a comprehensive review regarding the disruption of TADs and skeletal disorders, see Lupiáñez and colleagues[7].

Up to now, only small-scale data are available on histone modifications and 3D chromatin structure of chondrocytes and bone. This may reflect the novelty of these data, the costs, material amounts, and skills needed to perform such experiments and analysis. However, recent large-scale efforts from the ROADMAP consortium[113] and the encyclopedia of DNA elements (ENCODE)[114] have built genome-wide maps of several histone modifications and chromatin conformations in multiple human cells and tissues, including bone and cartilage[4]. Those data are freely accessible (**Table 3**) and can be used as an epigenomic reference map for the locations of regulatory elements in osteoblasts and chondrogenic cells. To our knowledge, no chromatin conformation capture data of chondrocyte or osteoblast is available. However, TADs have been shown to be stable across cells, tissues, and even species, which highlights biological relevance and suggests that they function as a general 3D framework to determine domains of

Table 3: Publicly Available Epigenetics Data on Bone and Cartilage Cells

Project	Cell	Type	Source	Experiment	Data	Combined data	Ref.
ENCODE	Chondrocyte	Primary cells	Knee articular chondrocyte	RNA-seq	Small RNA expression Total RNA expression		[114]
	Osteoblast	Primary cells		RNA-seq RNA-microarray CAGE RRBS DNase-seq ChIP-Seq	Small RNA expression Total RNA expression Transcriptional start sites Methylation DNA accessibility Histone marks: H2AFZ, H3K27ac, H3K27me3, H3K36me3, H3K4me1, H3K4me2, H3K4me3, H3K79me2, H3K9me3, H4K20me1 DNA binding proteins: CTCF, EP300	Full reference epigenome by ROADMAP: E129	[4, 113]
ROADMAP	Chondrocyte	Cultured cells	Mesenchy mal stem cell derived	ChIP-Seq CAGE	Transcriptional start sites Histone marks: H3K27ac, H#K27me3, H3K36me3, H3K4me1, H3K4me3, H3K-9ac, H3K9me3	Full reference epigenome by ROADMAP: E049 FANTOM5: active enhancers	[4, 113, 182]

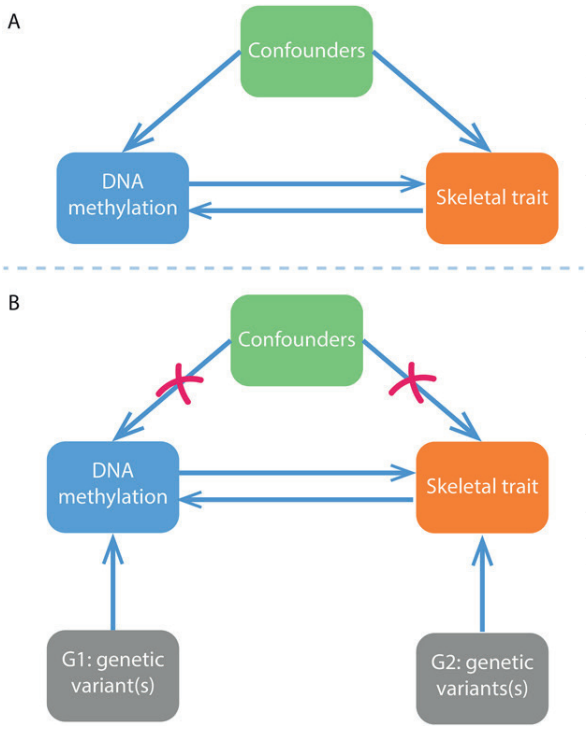
ENCODE=encyclopedia of DNA elements; ROADMAP=Roadmap Epigenomics Mapping Consortium; RNA-seq=RNA sequencing; CAGE=cap analysis gene expression; RRBS=reduced representation bisulfite sequencing; DNase-seq=DNase-I hypersensitive sites sequencing; ChIP-seq=chromatin immunoprecipitation sequencing.

possible interaction partners. Thus, chromatin conformation maps are a valuable resource to identify potential causal variants and possible causal genes in GWAS. TAD boundaries can be used to limit possible causal genes, and reference epigenome maps can help to identify active genes and variants in cell-specific active regulatory elements.

Interpretation of Epigenetic Studies: Problems and Tools

Cause or effect: confounding and reverse causation

Interpretation of epigenetic studies is not trivial because many factors can influence the association between the phenotype and epigenetic features. In contrast to DNA-sequence variants, epigenomic features are dynamic, meaning that they can be highly tissue specific (especially enhancers), dependent on environmental stimuli and developmental stage. This makes epigenetic analysis vulnerable for classical epidemiological pitfalls such as confounding and reverse causation. Besides the previously mentioned cellular heterogeneity as a potential confounder in case of multicellular tissues, such as bone or synovial tissue, a major issue is reverse causation. In cross-sectional studies, one can never be sure whether the epigenomic features cause the phenotype or vice versa. With respect to the epigenetic studies performed in cartilage and bone, this problem is also realistic. The massive epigenomic deregulation apparent in degraded cartilage can be a consequence of a process initiated by an entirely different cause of the disease. It is possible to use known genetic association and the concept of mendelian randomization to investigate the potential causal relationships between DNA methylation and the phenotype of interest.[115] In these kinds of studies, single nucleotide polymorphisms (SNPs) known to influence the CpG site are used as genetic instruments to test for causation. Similarly, SNPs associated with the trait/disease of interest can then be used to test whether they are associated with methylation levels at the same CpG site. In this way, the direction of cause can be disentangled (**Figure. 2**). This approach has recently been used to show that DNA methylation differences associated with BMI are predominantly a consequence of adiposity, rather than a cause[116]. In addition, this methodology has also been used to suggest that several known risk factors do not show a causal effect on bone fracture, except for BMD[117]. Similarly, high BMI was shown to cause OA, but other known clinical risk factors did not have a causal relationship with OA[118]. Typically, these mendelian randomization studies need genetic data with large sample sizes, which is increasingly available for both osteoporosis (GEFOS consortium) and osteoarthritis (Genetics of OA-consortium). However, sample sizes for the methylation studies in target tissue are typically small (**Table 1**), and therefore collaboration and meta-analysis across the different data sets are needed. This is increasingly recognized in the field[119].



◀Figure 2: Mendelian randomization to study causal relationships in epigenomic epidemiology. (A) The relationship between DNA methylation and the studied trait is difficult to interpret because it can be subject to classical epidemiological pitfalls, such as confounders and/or reverse causation. (B) To establish causal pathways for observed associations between (molecular) markers and clinical outcome, genetic variation can be used as a causal anchor. When examining the causal relation from the DNA methylation marker to the skeletal trait, a genetic variant (or combination of genetic variants) (G1) can be used that are robustly associated with the methylation marker. To study reverse causation, one can examine the causal relationship from skeletal trait to methylation by using a second set of genetic variants (G2) that are robustly associated with the skeletal trait of interest. Statistical methods to perform this analysis have been reviewed elsewhere.[115]

Integrating different omic levels

Interpretation of epigenomic data is complicated by the fact that the function of the genome is not fully known. It has become apparent that gene expression can be regulated by genomic features (enhancers/insulators) far away from the gene[120]. In fact, recent studies in cancer show that variation in methylation in distal enhancers account for a larger part in the regulation of expression, then promoter-methylation variation[121]. Similarly, den Hollander and colleagues observed that only 10% of the differentially methylated regions in cartilage is associated with differential expression of the nearest gene[45]. Integrating epigenomic data with data from other molecular layers, such as RNA expression and/or protein expression, can help to identify the function of the epigenomic features.

Epigenetics as mechanism for genetic etiology of disease

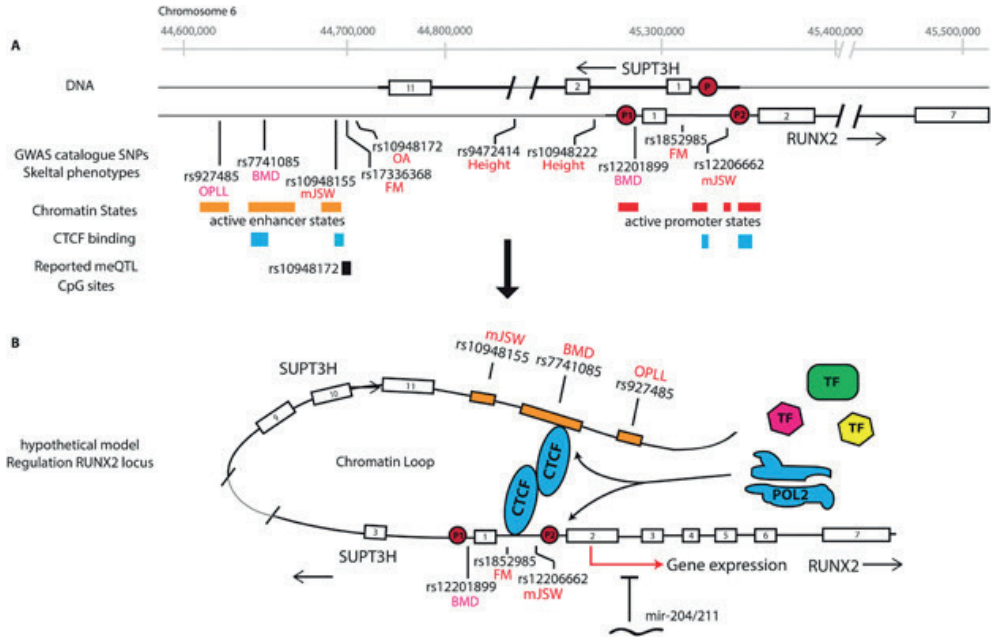
Of all the genetic loci identified by GWAS for complex diseases (such as OA and osteoporosis), the majority does not affect the protein coding but is instead thought to affect gene expression regulation. Epigenetics may mediate genetic risk, meaning that the methylation state of a specific locus is driven by a nearby genetic variant(s), also called a methylation quantitative trait locus (meQTL). In this way, integration of epigenomic

and genetic data can be used to elucidate the function and biological pathway underly-ing the genetic association. Large-scale studies in blood have shown that meQTLs are widespread (>30% of all CpGs have a meQTL), but that only 10% of the meQTLs also associate with expression levels of nearby genes[122]. A number of studies have ex-amined the relation between previously identified genetic loci for BMD and/or OA and methylation (and/or gene expression) in the target tissue. These studies identified a number of meQTLs specific for cartilage[123-125], suggesting that methylation medi-ates the association between the genetic locus and disease. Notably, absence of a meQTL in these studies does not mean that methylation does not play a role, since methylation is dynamic and context-dependent and can be dependent on a specific stimulus or de-velopmental stage.

Annotating epigenomic features: available information regarding skeletal tissues

Recent large-scale efforts have provided new understanding in the function of epigen-etic modifications, and efforts are ongoing to map histone modifications, transcription factor binding, and 3D chromatin structure of multiple cell types and tissues[5, 43, 44]. The data generated by these large-scale efforts are publicly available and contain DNA annotation on multiple levels: histone modification, binding of transcription factors and other DNA binding proteins, methylation, gene expression, DNA accessibility, and chro-matin conformation. Some also contain epigenetic information on cartilage and bone tissues; most notably are the "reference" epigenomes for osteoblasts and chondrocytes generated by the ROADMAP consortium (**Table 3**). The reference epigenomes construct an annotation of functional elements, ie, enhancers, promoters, etc., of the DNA per cell type,[4] which has already been proven to be a valuable resource for skeletal GWAS and EWAS, to finemap causal SNPs, CpGs, and genes, and also to provide for a hypothesis on the mechanism underlying GWAS/EWAS findings for skeletal outcomes[126, 127].

The interesting *SUPT3H-RUNX2* locus provides a good example on how different levels of molecular data can be combined to build a model for the regulatory mecha-nisms involved in the *RUNX2* locus, a master regulator of osteoblast and chondrocyte regulation (**Figure 3**). This locus encompasses a region of roughly ~700 kb, where multiple genetic association signals have been identified for OA[127]. BMD[128], OPLL (ossification of the posterior longitudinal ligament of the spine), cartilage thickness (measured as the minimum joint space width)[127]. height[129]. and facial morpholo-gy[130]. Interestingly, these GWAS signals are independent from each other[127]. Data from the ROADMAP consortium shows that several of the GWAS signals are located in different (potential) enhancers in cartilage and osteoblast cells. This suggests that the genetic association with various skeletal phenotypes might



▲**Figure 3: Hypothetical model for RUNX2 regulation and expression**. Several GWAS findings are linked to skeletal phenotypes located in the SUPT3H-RUNX2 locus. OA=osteoarthritis; BMD=bone mineral density; OPLL=ossification of the posterior longitudinal ligament of the spine; FM=facial morphology. (A) The locations of the GWAS findings coincide with possible active enhancers and promoter elements in chondrogenic cells and osteoblast. In addition, near some of these enhancer and promoter elements, CTCF binding was found in Chip-seq experiments in osteoblasts. For rs10948172, also meQTL effects on 4 CpG sites have been found in chondrogenic cells. (B) We hypothesize that the SNPs are located in long-range enhancers, which regulate RUNX2 gene expression during osteoblast and chondrogenic differentiation.

arise due to a difference in control of *RUNX2* expression. Recent studies have shown that regulation of *RUNX2* is tightly regulated by multiple epigenetic mechanisms, such as miR-204/211 and other miRNAs[131, 132], HDAC[133], and DNA methylation[123, 134]. Also, some of the genetic associa-tion signals near *RUNX2* exert effects on epigenetics. For the osteoarthritis-associated SNP, rs10948172, it has been shown to operate as a methylation quantitative trait loci (mQTL) for 4 CpG sites in the *RUNX2* locus[123, 134]. Nevertheless, the GWAS signals are positioned at a relative large distance from *RUNX2,* and chromatin interaction from regulatory region to gene promoter is needed. Using TAD localizations from the 3D genome browser, we notice that all skeletal genetic RUNX2 signals are within one TAD. This makes physical interaction of enhancers—promoters through chromatin "looping" likely to occur within this region. Such loops are mostly regulated by several DNA binding proteins such as cohesin and/or CTCF[135]. Using data from the ENCODE database, *RUNX2* P1 and P2 promoter regions (**Figure 3**). One can imagine that depending on the

developmental stage, the different enhancer regions are actively involved in regulation of expression, and different chromatin loops are established with the P1 and/or P2 promoter of RUNX2. The *RUNX2* regulatory region carries all elements of a so-called super-enhancer region, which is a genomic region comprising multiple enhancers. These super-enhancers are typically identified near genes important for cell-identity/master transcription regulation genes, which require detailed regulation of expression, such as is the case for *RUNX2.*

What Will the Future Hold?

Advances in understanding the epigenomic mechanisms have greatly expanded our knowledge about the molecular mechanisms involved in the differentiation and activity of cells responsible for skeletal homeostasis, which are central players in the pathogenesis of disorders such as osteoporosis and OA. Thus, we can anticipate that epigenetic mechanisms must also play an important role in these disorders and may become the foundation for new therapies. However, much more research is needed in order to use epigenetic marks as biomarkers of disease risk or progression and to introduce epigenetic-based therapies into the clinic.

To date, association studies aimed to elucidate the association of epigenetic signals with bone phenotypes have had limited reproducibility, particularly regarding DNA methylation marks. Similarly, some miRNA signatures have been produced in a few small-size studies, but replication in larger studies is still pending. Of course, this is critical before introducing the analysis of epigenetic markers as tools for diagnosis or prognosis of disease. It will likely require large collaborative, epigenome-wide studies.

A major issue regarding therapy aimed at modulating epigenetic mechanisms is related to lack of specificity and undesired effects. Drugs interfering with DNA methylation are already being used to treat some bone marrow disorders, but their widespread effects make them rather unsuitable for non-neoplastic disorders. Thus, CRISPR-based methods and other procedures to induce targeted DNA methylation changes to specific loci in somatic cells may greatly help to first delineate the effects of methylation marks and then to epigenetically modify the activity of specific genes[136]. SAHA-PIPs are a novel class of histone modifiers made by conjugating selective DNA binding pyrrole-imidazole polyamides (PIPs) with the histone deacetylase inhibitor SAHA. They show some selectivity and modulate the transcription of certain clusters of genes[137].

Localized skeletal disorders, such as some forms of OA and delayed-union fractures may be amenable to local therapies, such as agonists and antagonists of specific

miRNAs or genetically/epigenetically engineered MSCs. The local nature of the disease and the therapy may help to avoid generalized undesired effects. In this line, antisense oligonucleotides of miRNA 181a-5p attenuate cartilage destruction when injected into the knees of rats and mice with experimentally induced OA[138].

The task of revealing the role of epigenomic variants and pathways is even more difficult because, unlike the genome, the epigenome is cell- and tissue-specific and may change over time. Thus, studies may need to be done using skeletal samples, which poses obvious difficulties for human studies and particularly for those requiring sampling at multiple time points. For studies on developmental processes for bone and cartilage in humans, model systems such as human stem cells and IPSCs might prove a good alternative. In addition, organ on a chip technology is promising for studying early disease processes. In this regard, recent findings that suggest that some molecules circulating in blood may be used as biomarkers of the status of solid tissues, if confirmed, may facilitate using epigenetic elements as biomarkers. In the skeletal field, miRNAs have been mostly studied from this perspective. However, methylation marks in circulating cell-free DNA are also being actively explored in cancer and other disorders.

Technical advances to facilitate high-throughput analysis of epigenomic marks at decreasing costs are emerging and will facilitate larger-scale applications. Sequencing costs are continuously going down, making it possible to generate increasing amounts of data. The bottleneck for these "big data" studies is the data analysis, which in general requires the same costs as generation of the raw data itself. An emerging technique is single-molecule sequencing, with Pac Bio's single-molecule real-time sequencing (SMRT) and Oxford Nanopore's (ON) nanopore sequencing as the most prominent players in the field. SMRT and ON are able to directly sequence native DNA or RNA, making it possible to directly measure chemical groups attached to the nucleic acid sequence (such as methylation), avoiding the bias of other techniques due to the necessary amplification and/or bisulphite treatment. Although single-molecule sequencing is still in a developmental stage, more and more applications are being developed, indicating that the technique is almost ready for wider-scale applications.

Another important technical advance is single-cell epigenomics, which is developing rapidly. Single-cell "omic" technologies have emerged as powerful tools to explore cellular heterogeneity at individual cell resolution. Previous scientific knowledge in cell biology is largely limited to data generated by bulk profiling methods, which only provide averaged read-outs that generally mask cellular heterogeneity. Single-cell technology has recently led to the identification of a self-renewing skeletal stem cell that generates progenitors of bone, cartilage, and stroma but not fat[139]. In addition, the first study applying single-cell RNAseq on OA cartilage identified seven chondrocyte

subpopulations in cartilage[140].

Because of the high sensitivity of single-cell genomics, attention must be put into experimental setup and execution of this technology. Careful handling and processing of cells is critical to preserve the native epigenetic and expression profile; only then meaningful analysis and conclusions can be drawn. This is especially important for the study of skeletal cells because their natural habitat is normally deeply embedded in the extracellular matrix and getting a single-cell solution from cartilage or bone is not trivial.

A new field of epigenetic regulation is epi-transcriptomics[141]. These are RNA modifications that represent a novel layer of regulation of gene expression. Early reports show the existence of many different dynamic RNA modifications. The role of these RNA modifications in regulating gene expression is largely unexplored, but recent evidence suggests that these RNA modifications are key switches for RNA metabolism.

It took more than 1,500 years to evolve from Ptolomeo's map to good world maps. The map of the human epigenome is probably much more complex. Fortunately, technical advances allow us to be optimistic and not expect such a long time for the final drawing. However, building the maps of the human epigenome and specifically of human skeletal cells, as well as the tracks to be travelled to get into the core mechanisms driving the move from healthy bones and joints toward disease, is still a formidable work. Intensive collaboration of multiple research groups working together is a must to accomplish such a daunting task. Previous experience in the genetic field shows that it may be difficult to do, but it is feasible. A new world of hope for millions of patients is ahead.

Disclosures

All authors state that they have no conflicts of interest.

Acknowledgments

Authors' studies were supported by grants from Instituto de Salud Carlos III to JAR (PI16/915; which may be co-funded by EU funds) and Reuma Nederland (16-1-404).

Authors' roles: JBJM, CGB, LLD and JAR contributed to conception, writing and revision of the manuscript.

References

1. Tronick E, Hunter RG. Waddington, dynamic systems, and epigenetics. Front Behav Neurosci. 2016; 10: 107.
2. Feinberg AP. The key role of epigenetics in human disease prevention and mitigation. N Engl J Med. 2018; 378: 1323– 34.
3. Greally JM. A user's guide to the ambiguous word “epigenetics.” Nat Rev Mol Cell Biol. 2018; 19(4): 207– 8.
4. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015; 518(7539): 317– 30.
5. van Dongen J, Nivard MG, Willemsen G, et al. Genetic and environmental influences interact with age and sex in shaping the human methylome. Nat Commun. 2016; 7: 11115.
6. Rothbart SB, Strahl BD. Interpreting the language of histone and DNA modifications. Biochim Biophys Acta. 2014; 1839: 627– 43.
7. Lupiáñez DG, Spielmann M, Mundlos S. Breaking TADs: how alterations of chromatin domains result in disease. Trends Genet. 2016; 32(4): 225– 37.
8. Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. Nat Rev Genet. 2016; 17(8): 487– 500.
9. Achour C, Aguilo F. Long non-coding RNA and polycomb: an intricate partnership in cancer biology. Front Biosci. 2018; 23: 2106– 32.
10. Schmitz SU, Grote P, Herrmann BG. Mechanisms of long noncoding RNA function in development and disease. Cell Mol Life Sci. 2016; 73(13): 2491– 509.
11. Fleming TP, Watkins AJ, Velazquez MA, et al. Origins of lifetime health around the time of conception: causes and consequences. Lancet. 2018; 391(10132): 1842– 52.
12. Baird J, Jacob C, Barker M, et al. Developmental origins of health and disease: a lifecourse approach to the prevention of non-communicable diseases. Healthcare. 2017; 5(1): 14.
13. Tobi EW, Slieker RC, Luijk R, et al. DNA methylation as a mediator of the association between prenatal adversity and risk factors for metabolic disease in adulthood. Sci Adv. 2018; 4(1): eaao4364.
14. Tobi EW, Goeman JJ, Monajemi R, et al. DNA methylation signatures link prenatal famine exposure to growth and metabolism. Nat Commun. 2014; 5: 5592.
15. Martínez-Mesa J, Restrepo-Méndez MC, González DA, et al. Life-course evidence of birth weight effects on bone mass: systematic review and meta-analysis. Osteoporos Int. 2013; 24(1): 7– 18.
16. von Websky K, Hasan AA, Reichetzeder C, Tsuprykov O, Hocher B. Impact of vitamin D on pregnancy-related disorders and on offspring outcome. J Steroid Biochem Mol Biol. 2018; 180(August 2017): 51– 64.
17. Mahon P, Harvey N, Crozier S, et al. Low maternal vitamin D status and fetal bone development: cohort study. J Bone Miner Res. 2010; 25: 14– 9.
18. Garcia AH, Erler NS, Jaddoe VWV, et al. 25-hydroxyvitamin D concentrations during fetal life and bone health in children aged 6 years: a population-based prospective cohort study. Lancet Diabetes Endocrinol. 2017; 5(5): 367– 76.
19. Xue J, Schoenrock SA, Valdar W, Tarantino LM, Ideraabdullah FY. Maternal vitamin D depletion alters DNA methylation at imprinted loci in multiple generations. Clin Epigenetics. 2016; 8: 107.
20. Curtis EM, Murray R, Titcombe P, et al. Perinatal DNA methylation at CDKN2A is associated with offspring bone mass: findings from the Southampton Women's Survey. J Bone Miner Res. 2017; 32: 2030– 40.
21. Harvey NC, Lillycrop KA, Garratt E, et al. Evaluation of methylation status of the eNOS promoter at birth in relation to childhood bone mineral content. Calcif Tissue Int. 2012; 90(0171–967; 2): 120– 7.
22. Harvey NC, Sheppard A, Godfrey KM, et al. Childhood bone mineral content is associated with methylation status of the RXRA promoter at birth. J Bone Miner Res. 2014; 29(3): 600– 7.

23. Xu X, Liu L, Xie W, et al. Increase in the prevalence of arthritis in adulthood among adults exposed to Chinese famine of 1959 to 1961 during childhood: a cross-sectional survey. Medicine (Baltimore). 2017; 96(13): e6496.
24. Clynes MA, Parsons C, Edwards MH, et al. Further evidence of the developmental origins of osteoarthritis: results from the Hertfordshire Cohort Study. J Dev Orig Health Dis. 2014; 5(06): 453– 8.
25. de Kruijf M, Kerkhof HJM, Peters MJ, et al. Finger length pattern as a biomarker for osteoarthritis and chronic joint pain: a population-based study and meta-analysis after systematic review. Arthritis Care Res. 2014; 66(9): 1337– 43.
26. Berenbaum SA, Beltz AM. Sexual differentiation of human behavior: effects of prenatal and pubertal organizational hormones. Front Neuroendocrinol. 2011; 32(2): 183– 200.
27. Gärtner K. A third component causing random variability beside environment and genotype. A reason for the limited success of a 30 year long effort to standardize laboratory animals? int J Epidemiol. 2012; 41(2): 335– 41.
28. Czyz W, Morahan JM, Ebers GC, Ramagopalan SV. Genetic, environmental and stochastic factors in monozygotic twin discordance with a focus on epigenetic differences. BMC Med. 2012; 10(1): 93.
29. Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: from theories to phenotypes. Nat Rev Genet. 2005; 6(6): 451– 64.
30. Vogt G. Stochastic developmental variation, an epigenetic source of phenotypic diversity with far-reaching biological consequences. J Biosci. 2015; 40(1): 159– 204.
31. Lin GL, Hankenson KD. Integration of BMP, Wnt, and notch signaling pathways in osteoblast differentiation. J Cell Biochem. 2011; 112(12): 3491– 501.
32. Delgado-Calle J, Sanudo C, Sanchez-Verde L, Garcia-Renedo RJ, Arozamena J, Riancho JA. Epigenetic regulation of alkaline phosphatase in human cells of the osteoblastic lineage. Bone. 2011; 49(4): 830– 8.
33. Delgado-Calle J, Sanudo C, Fernandez AF, Garcia-Renedo R, Fraga MF, Riancho JA. Role of DNA methylation in the regulation of the RANKL-OPG system in human bone. Epigenetics. 2012; 7(1): 83– 91.
34. del Real Á, Riancho JA, Delgado-Calle J. Epigenetic regulation of Sost/sclerostin expression. Curr Mol Biol Rep. 2017; 3(2): 85– 93.
35. El Serafi AT, Oreffo RO, Roach HI. Epigenetic modifiers influence lineage commitment of human bone marrow stromal cells: differential effects of 5-aza-deoxycytidine and trichostatin A. Differentiation. 2011; 81(1): 35– 41.
36. Yan X, Ehnert S, Culmes M, et al. 5-azacytidine improves the osteogenic differentiation potential of aged human adipose-derived mesenchymal stem cells by DNA demethylation. PLoS One. 2014; 9(3): 1– 10.
37. Fernández AF, Bayón GF, Urdinguio RG, et al. H3K4me1 marks DNA regions hypomethylated during aging in human stem and differentiated cells. Genome Res. 2015; 25(1).
38. Toraño EG, Bayón GF, del Real Á, et al. Age-associated hydroxymethylation in human bone-marrow mesenchymal stem cells. J Transl Med. 2016; 14(1): 207.
39. Roforth MM, Farr JN, Fujita K, et al. Global transcriptional profiling using RNA sequencing and DNA methylation patterns in highly enriched mesenchymal cells from young versus elderly women. Bone. 2015; 76: 49– 57.
40. de la Rica L, Rodríguez-Ubreva J, García M, et al. PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. Genome Biol. 2013; 14(9): R99.
41. Nishikawa K, Iwamoto Y, Kobayashi Y, et al. DNA methyltransferase 3a regulates osteoclast differentiation by coupling to an S-adenosylmethionine–producing metabolic pathway. Nat Med. 2015; 21(3): 281– 7.
42. Ramos YFM, Meulenbelt I. The role of epigenetics in osteoarthritis. Curr Opin Rheumatol. 2017; 29(1): 119– 29.

43. Steinberg J, Ritchie GRS, Roumeliotis TI, et al. Integrative epigenomics, transcriptomics and proteomics of patient chondrocytes reveal genes and pathways involved in osteoarthritis. Sci Rep. 2017;7(April): 8935.
44. Bonin CA, Lewallen EA, Baheti S, et al. Identification of differentially methylated regions in new genes associated with knee osteoarthritis. Gene. 2016; 576(1): 312– 8.
45. den Hollander W, Ramos YFM, Bos SD, et al. Knee and hip articular cartilage have distinct epigenomic landscapes: implications for future cartilage regeneration approaches. Ann Rheum Dis. 2014; 73(12): 2208– 12.
46. Jeffries MA, Donica M, Baker LW, et al. Genome-wide DNA methylation study identifies significant epigenomic changes in osteoarthritic cartilage. Arthritis Rheumatol. 2014; 66(10): 2804– 15.
47. Moazedi-Fuerst FC, Hofner M, Gruber G, et al. Epigenetic differences in human cartilage between mild and severe OA. J Orthop Res. 2014; 32(12): 1636– 45.
48. Aref-Eshghi E, Zhang Y, Liu M, et al. Genome-wide DNA methylation study of hip and knee cartilage reveals embryonic organ and skeletal system morphogenesis as major pathways involved in osteoarthritis. BMC Musculoskelet Disord. 2015; 16(1): 287.
49. Rushton MD, Reynard LN, Barter MJ, et al. Characterization of the cartilage DNA methylome in knee and hip osteoarthritis. Arthritis Rheumatol (Hoboken, NJ). 2014; 66(9): 2450– 60.
50. Fernández-Tajes J, Soto-Hermida A, Vázquez-Mosquera ME, et al. Genome-wide DNA methylation analysis of articular chondrocytes reveals a cluster of osteoarthritic patients. Ann Rheum Dis. 2014; 73(4): 668– 77.
51. Bomer N, den Hollander W, Suchiman H, et al. Neo-cartilage engineered from primary chondrocytes is epigenetically similar to autologous cartilage, in contrast to using mesenchymal stem cells. Osteoarthritis Cartilage. 2016; 24(8): 1423– 30.
52. Rushton MD, Young DA, Loughlin J, Reynard LN. Differential DNA methylation and expression of inflammatory and zinc transporter genes defines subgroups of osteoarthritic hip patients. Ann Rheum Dis. 2015; 74(9): 1778– 82.
53. van Meurs JBJ. Osteoarthritis year in review 2016: genetics, genomics and epigenetics. Osteoarthritis Cartilage. 2017; 25(2): 181– 9.
54. Chen BH, Marioni RE, Colicino E, et al. DNA methylation-based measures of biological age: meta-analysis predicting time to death. Aging (Albany NY). 2016; 8(9): 1844– 65.
55. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. Nat Rev Genet. 2018; 19(6): 371– 84.
56. Vidal-Bralo L, Lopez-Golan Y, Mera-Varela A, et al. Specific premature epigenetic aging of cartilage in osteoarthritis. Aging (Albany NY). 2016; 8(9).
57. Delgado-Calle J, Fernández AF, Sainz J, et al. Genome-wide profiling of bone reveals differentially methylated regions in osteoporosis and osteoarthritis. Arthritis Rheum. 2013; 65(1): 197– 205.
58. del Real A, Pérez-Campo FM, Fernández AF, et al. Differential analysis of genome-wide methylation and gene expression in mesenchymal stem cells of patients with fractures and osteoarthritis. Epigenetics. 2017; 12(2): 113– 22.
59. Reppe S, Lien TG, Hsu Y-H, et al. Distinct DNA methylation profiles in bone and blood of osteoporotic and healthy postmenopausal women. Epigenetics. 2017; 12: 674– 87.
60. Morris JA, Tsai P-C, Joehanes R, et al. Epigenome-wide association of DNA methylation in whole blood with bone mineral density. J Bone Miner Res. 2017; 32(8): 1644– 50.
61. Cheishvili D, Parashar S, Mahmood N, et al. Identification of an epigenetic signature of osteoporosis in blood DNA of postmenopausal women. J Bone Miner Res. 2018; 33(11): 1980– 9.
62. Yang Y, Fang S. Small non-coding RNAs-based bone regulation and targeting therapeutic strategies. Mol Cell Endocrinol. 2017; 456: 16– 35.
63. Sera SR, zur Nieden NI. microRNA regulation of skeletal development. Curr Osteoporos Rep. 2017; 15(4): 353– 66.

64. Taipaleenmäki H. Regulation of bone metabolism by microRNAs. Curr Osteoporos Rep. 2018; 16(1): 1– 12.
65. Li D, Liu J, Guo B, et al. Osteoclast-derived exosomal miR-214-3p inhibits osteoblastic bone formation. Nat Commun. 2016; 7: 10872.
66. Wang X, Guo B, Li Q, et al. miR-214 targets ATF4 to inhibit bone formation. Nat Med. 2013; 19(1546–170; 1): 93– 100.
67. Li H, Xie H, Liu W, et al. A novel microRNA targeting HDAC5 regulates osteoblast differentiation in mice and contributes to primary osteoporosis in humans. J Clin Invest. 2009; 119(12): 3666– 77.
68. Seeliger C, Er B, van Griensven M. miRNAs related to skeletal diseases. Stem Cells Dev. 2016; 25: 1261– 8.
69. Kolhe R, Hunter M, Liu S, et al. Gender-specific differential expression of exosomal miRNA in synovial fluid of patients with osteoarthritis. Sci Rep. 2017; 7(1): 2029.
70. Cong L, Zhu Y, Tu G. A bioinformatic analysis of microRNAs role in osteoarthritis. Osteoarthritis Cartilage. 2017; 25(8): 1362– 71.
71. Si HB, Zeng Y, Liu SY, et al. Intra-articular injection of microRNA-140 (miRNA-140) alleviates osteoarthritis (OA) progression by modulating extracellular matrix (ECM) homeostasis in rats. Osteoarthritis Cartilage. 2017; 25(10): 1698– 707.
72. Trachana V, Ntoumou E, Anastasopoulou L, Tsezou A. Studying microRNAs in osteoarthritis: critical overview of different analytical approaches. Mech Ageing Dev. 2018; 171: 15– 23.
73. Li X, Guo L, Liu Y, et al. MicroRNA-21 promotes osteogenesis of bone marrow mesenchymal stem cells via the Smad7-Smad1/5/8-Runx2 pathway. Biochem Biophys Res Commun. 2017; 493(2): 928– 33.
74. Zhao Y, Wang Z, Wu G, et al. Improving the osteogenesis of human bone marrow mesenchymal stem cell sheets by microRNA-21-loaded chitosan/hyaluronic acid nanoparticles via reverse transfection. Int J Nanomedicine. 2016; 11: 2091.
75. Yoshizuka M, Nakasa T, Kawanishi Y, et al. Inhibition of microRNA-222 expression accelerates bone healing with enhancement of osteogenesis, chondrogenesis, and angiogenesis in a rat refractory fracture model. J Orthop Sci. 2016; 21(6): 852– 8.
76. Li K-C, Chang Y-H, Yeh C-L, Hu Y-C. Healing of osteoporotic bone defects by baculovirus-engineered bone marrow-derived MSCs expressing MicroRNA sponges. Biomaterials. 2016; 74: 155– 66.
77. Zhang L, Tang Y, Zhu X, et al. Overexpression of MiR-335-5p promotes bone formation and regeneration in mice. J Bone Miner Res. 2017; 32(12): 2466– 75.
78. Deng Y, Zhou H, Zou D, et al. The role of miR-31-modified adipose tissue-derived stem cells in repairing rat critical-sized calvarial defects. Biomaterials. 2013; 34(28): 6717– 28.
79. Grol MW, Lee BH. Gene therapy for repair and regeneration of bone and cartilage. Curr Opin Pharmacol. 2018; 40: 59– 66.
80. Hackl M, Heilmeier U, Weilner S, Grillari J. Circulating microRNAs as novel biomarkers for bone diseases—complex signatures for multifactorial diseases? Mol Cell Endocrinol. 2016; 432: 83– 95.
81. Materozzi M, Merlotti D, Gennari L, Bianciardi S. The Potential role of miRNAs as new biomarkers for osteoporosis. Int J Endocrinol. 2018; 2018: 1– 10.
82. Gennari L, Bianciardi S, Merlotti D. MicroRNAs in bone diseases. Osteoporos Int. 2017; 28(4): 1191– 213.
83. Cantley MD, Zannettino ACW, Bartold PM, Fairlie DP, Haynes DR. Histone deacetylases (HDAC) in physiological and pathological bone remodelling. Bone. 2017; 95: 162– 74.
84. Bradley EW, Carpio LR, van Wijnen AJ, McGee-Lawrence ME, Westendorf JJ. Histone deacetylases in bone development and skeletal disorders. Physiol Rev. 2015; 95(4): 1359– 81.
85. Pike JW, Meyer MB, St John HC, Benkusky NA. Epigenetic histone modifications and master regulators as determinants of context dependent nuclear receptor activity in bone cells. Bone. 2015; 81: 757– 64.
86. LR, Westendorf JJ. Histone deacetylases in cartilage homeostasis and osteoarthritis. Curr Rheumatol Rep. 2016; 18(8): 52.

87. Zhang Y-X, Sun H-L, Liang H, Li K, Fan Q-M, Zhao Q-H. Dynamic and distinct histone modifications of osteogenic genes during osteogenic differentiation. J Biochem. 2015; 158(6): 445– 57.

88. Batlle-López A, Cortiguera MG, Rosa-Garrido M, et al. Novel CTCF binding at a site in exon1A of BCL6 is associated with active histone marks and a transcriptionally active locus. Oncogene. 2015; 34: 246– 56.

89. Bi W, Deng JM, Zhang Z, Behringer RR, de Crombrugghe B. Sox9 is required for cartilage formation. Nat Genet. 1999; 22(1): 85– 9.

90. Lee H-L, Yu B, Deng P, Wang C-Y, Hong C. Transforming growth factor-β-induced KDM4B promotes chondrogenic differentiation of human mesenchymal stem cells. Stem Cells. 2016; 34(3): 711– 9.

91. Dvir-Ginzberg M, Gagarina V, Lee E-J, Hall DJ. Regulation of cartilage-specific gene expression in human chondrocytes by SirT1 and nicotinamide phosphoribosyltransferase. J Biol Chem. 2008; 283(52): 36300– 10.

92. Tsuda M, Takahashi S, Takahashi Y, Asahara H. Transcriptional co-activators CREB-binding protein and p300 regulate chondrocyte-specific gene expression via association with Sox9. J Biol Chem. 2003; 278(29): 27224– 9.

93. Furumatsu T, Tsuda M, Yoshida K, et al. Sox9 and p300 cooperatively regulate chromatin-mediated transcription. J Biol Chem. 2005; 280(42): 35203– 8.

94. Dai J, Yu D, Wang Y, et al. Kdm6b regulates cartilage development and homeostasis through anabolic metabolism. Ann Rheum Dis. 2017; 76(7): 1295– 303.

95. Otero M, Peng H, Hachem K El, et al. ELF3 modulates type II collagen gene (COL2A1) transcription in chondrocytes by inhibiting SOX9-CBP/p300-driven histone acetyltransferase activity. Connect Tissue Res. 2017; 58(1): 15– 26.

96. Wondimu EB, Culley KL, Quinn J, et al. Elf3 contributes to cartilage degradation in vivo in a surgical model of post-traumatic osteoarthritis. Sci Rep. 2018; 8(1): 6438.

97. Khan NM, Haqqi TM. Epigenetics in osteoarthritis: potential of HDAC inhibitors as therapeutics. Pharmacol Res. 2018; 128: 73– 9.

98. Evangelou E, Valdes AM, Castano-Betancourt MC, 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(7): 1264– 5.

99. Monteagudo S, Cornelis FMF, Aznar-Lopez C, et al. DOT1L safeguards cartilage homeostasis and protects against osteoarthritis. Nat Commun. 2017; 8: 15889.

100. Niu N, Shao R, Yan G, Zou W. Bromodomain and extra-terminal (BET) protein inhibitors suppress chondrocyte differentiation and restrain bone growth. J Biol Chem. 2016; 291(52): 26647– 57.

101. Jiang Y, Zhu L, Zhang T, et al. BRD4 has dual effects on the HMGB1 and NF-κB signalling pathways and is a potential therapeutic target for osteoarthritis. Biochim Biophys Acta Mol Basis Dis. 2017; 1863(12): 3001– 15.

102. Baud'huin M, Lamoureux F, Jacques C, et al. Inhibition of BET proteins and epigenetic signaling as a potential treatment for osteoporosis. Bone. 2017; 94: 10– 21.

103. Park-Min K-H, Lim E, Lee MJ, et al. Inhibition of osteoclastogenesis and inflammatory bone resorption by targeting BET proteins and epigenetic regulation. Nat Commun. 2014; 5: 5418.

104. Dimitrova E, Turberfield AH, Klose RJ. Histone demethylases in chromatin biology and beyond. EMBO Rep. 2015; 16(12): 1620– 39.

105. Wood K, Tellier M, Murphy S. DOT1L and H3K79 methylation in transcription and genomic stability. Biomolecules. 2018; 8(1).

106. Jones B, Su H, Bhat A, et al. The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. PLoS Genet. 2008; 4(9): e1000190.

107. Matsushima S, Sadoshima J. The role of sirtuins in cardiac disease. Am J Physiol Circ Physiol. 2015; 309(9): H1375– 89.

108. Burchfield JS, Li Q, Wang HY, Wang R-F. JMJD3 as an epigenetic regulator in development and disease. Int J Biochem Cell Biol. 2015; 67: 148– 57.

109. Stathis A, Bertoni F. BET proteins as targets for anticancer treatment. Cancer Discov. 2018; 8(1): 24– 36.
110. Wang Z, Zang C, Rosenfeld JA, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet. 2008; 40(7): 897– 903.
111. Lupiáñez DG, Kraft K, Heinrich V, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell. 2015; 161(5): 1012– 25.
112. Spielmann M, Brancati F, Krawitz PM, et al. Homeotic arm-to-leg transformation associated with genomic rearrangements at the PITX1 locus. Am J Hum Genet. 2012; 91(4): 629– 35.
113. Bernstein BE, Stamatoyannopoulos JA, Costello JF, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 2010; 28(10): 1045– 8.
114. Rosenbloom KR, Sloan CA, Malladi VS, et al. ENCODE data in the UCSC Genome Browser: year 5 update. Nucleic Acids Res. 2013; 41(Database issue): D56– 63.
115. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. Int J Epidemiol. 2012; 41(1): 161– 76.
116. Wahl S, Drong A, Lehne B, et al. Epigenome-wide association study of body mass index and the adverse outcomes of adiposity. Nature. 2017; 541(7635): 81– 6.
117. Trajanoska K, Morris JA, Oei L, et al. Assessment of the genetic and clinical determinants of fracture risk: genome wide association and mendelian randomisation study. BMJ. 2018; 362: k3225.
118. Zengini E, Hatzikotoulas K, Tachmazidou I, et al. Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. Nat Genet. 2018; 50(4): 549– 58.
119. Reynard LN. Analysis of genetics and DNA methylation in osteoarthritis: what have we learnt about the disease? Semin Cell Dev Biol. 2017; 62: 57– 66.
120. Whalen S, Truty RM, Pollard KS. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. Nat Genet. 2016; 48(5): 488– 96.
121. Aran D, Sabato S, Hellman A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. Genome Biol. 2013; 14(3): R21.
122. Bonder MJ, Luijk R, Zhernakova DV, et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nat Genet. 2017; 49(1): 131– 8.
123. Rushton MD, Reynard LN, Young DA, et al. Methylation quantitative trait locus analysis of osteoarthritis links epigenetics with genetic risk. Hum Mol Genet. 2015; 24(25): 7432– 44.
124. den Hollander W, Ramos YFM, Bomer N, et al. Transcriptional associations of osteoarthritis-mediated loss of epigenetic control in articular cartilage. Arthritis Rheumatol. 2015; 67(8): 2108– 16.
125. Bomer N, den Hollander W, Ramos YFM, et al. Underlying molecular mechanisms of DIO2 susceptibility in symptomatic osteoarthritis. Ann Rheum Dis. 2015; 74(8): 1571– 9.
126. Medina-Gomez C, Kemp JP, Dimou NL, et al. Bivariate genome-wide association meta-analysis of pediatric musculoskeletal traits reveals pleiotropic effects at the SREBF1/TOM1L2 locus. Nat Commun. 2017; 8(1): 121.
127. Castaño-Betancourt MC, Evans DS, Ramos YFM, et al. Novel genetic variants for cartilage thickness and hip osteoarthritis. PLoS Genet. 2016; 12(10): e1006260.
128. Medina-Gomez C, Kemp JP, Trajanoska K, et al. Life-course genome-wide association study meta-analysis of total body BMD and assessment of age-specific effects. Am J Hum Genet. 2018; 102(1): 88– 102.
129. Wood AR, Esko T, Yang J, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet. 2014; 46(11): 1173– 86.
130. Adhikari K, Fuentes-Guajardo M, Quinto-Sánchez M, et al. A genome-wide association scan implicates DCHS2, RUNX2, GLI3, PAX1 and EDAR in human facial variation. Nat Commun. 2016;7(May): 1– 11.
131. Huang J, Zhao L, Xing L, Chen D. MicroRNA-204 regulates Runx2 protein expression and mesenchymal progenitor cell differentiation. Stem Cells. 2010; 28(2): 357– 64.

132. Zhao W, Zhang S, Wang B, Huang J, Lu WW, Chen D. Runx2 and microRNA regulation in bone and cartilage diseases. Ann N Y Acad Sci. 2016; 1383(1): 80– 7.
133. Vishal M, Ajeetha R, Keerthana R, Selvamurugan N. Regulation of Runx2 by histone deacetylases in bone. Curr Protein Pept Sci. 2016; 17(4): 343– 51.
134. Rice SJ, Aubourg G, Sorial AK, et al. Identification of a novel, methylation-dependent, RUNX2 regulatory region associated with osteoarthritis risk. Hum Mol Genet. 2018; 27(19): 3464– 74.
135. Marsman J, O'Neill AC, Kao BR-Y, et al. Cohesin and CTCF differentially regulate spatiotemporal runx1 expression during zebrafish development. Biochim Biophys Acta. 2014; 1839: 50– 61.
136. Liu XS, Wu H, Ji X, et al. Editing DNA methylation in the mammalian genome. Cell. 2016; 167(1): 233– 247.e17.
137. Prachayasittikul V, Prathipati P, Pratiwi R, et al. Exploring the epigenetic drug discovery landscape. Expert Opin Drug Discov. 2017; 12(4): 345– 62.
138. Nakamura A, Rampersaud YR, Nakamura S, et al. microRNA-181a-5p antisense oligonucleotides attenuate osteoarthritis in facet and knee joints. Ann Rheum Dis. 2019; 78(1): 111– 21.
139. Chan CKF, Gulati GS, Sinha R, et al. Identification of the human skeletal stem cell. cell. 2018; 175(1): 43– 56.
140. Ji Q, Zheng Y, Zhang G, et al. Single-cell RNA-seq analysis reveals the progression of human osteoarthritis. Ann Rheum Dis. 2019; 78(1): 100– 10.
141. Kadumuri RV, Janga SC. Epitranscriptomic code and its alterations in human disease. Trends Mol Med. 2018; 24: 886– 903.
142. Zhang Y, Fukui N, Yahata M, et al. Identification of DNA methylation changes associated with disease progression in subchondral bone with site-matched cartilage in knee osteoarthritis. Sci Rep. 2016; 6(1): 34460.
143. Alvarez-Garcia O, Fisch KM, Wineinger NE, et al. Increased DNA methylation and reduced expression of transcription factors in human osteoarthritis cartilage. Arthritis Rheumatol. 2016; 68(8): 1876– 86.
144. Taylor SEB, Li YH, Wong WH, Bhutani N. Genome-wide mapping of DNA hydroxymethylation in osteoarthritic chondrocytes. Arthritis Rheumatol. 2015; 67(8): 2129– 40.
145. Steinberg J, Brooks RA, Southam L, et al. Widespread epigenomic, transcriptomic and proteomic differences between hip osteophytic and articular chondrocytes in osteoarthritis. Rheumatology. 2018; 57(8): 1481– 9.
146. Morris JA, Tsai P-C., Joehanes R, et al. Epigenome-wide association of DNA methylation in whole blood with bone mineral density. J Bone Miner Res. 2017; 32(8): 1644– 50.
147. Xie Q, Wang Z, Zhou H, et al. The role of miR-135-modified adipose-derived mesenchymal stem cells in bone regeneration. Biomaterials. 2016; 75: 279– 94.
148. Zuo B, Zhu J, Li J, et al. microRNA-103a functions as a mechanosensitive microRNA to inhibit bone formation through targeting Runx2. J Bone Miner Res. 2015; 30(2): 330– 45.
149. Tao S-C, Yuan T, Zhang Y-L, Yin W-J, Guo S-C, Zhang C-Q. Exosomes derived from miR-140-5p-overexpressing human synovial mesenchymal stem cells enhance cartilage tissue regeneration and prevent osteoarthritis of the knee in a rat model. Theranostics. 2017; 7(1): 180– 95.
150. Dai L, Zhang X, Hu X, et al. Silencing of miR-101 prevents cartilage degradation by regulating extracellular matrix–related genes in a rat model of osteoarthritis. Mol Ther. 2015; 23(8): 1331– 40.
151. Yang S, Zhang W, Cai M, et al. Suppression of bone resorption by miR-141 in aged rhesus monkeys. J Bone Miner Res. 2018; 33(10): 1799– 812.
152. Wang H, Xie Z, Hou T, et al. MiR-125b regulates the osteogenic differentiation of human mesenchymal stem cells by targeting BMPR1b. Cell Physiol Biochem. 2017; 41(2): 530– 42.
153. Wang X, Guo Y, Wang C, Yu H, Yu X, Yu H. MicroRNA-142-3p inhibits chondrocyte apoptosis and inflammation in osteoarthritis by targeting HMGB1. Inflammation. 2016; 39(5): 1718– 28.
154. Yu F, Xie C, Sun J, Peng W, Huang X. Overexpressed miR-145 inhibits osteoclastogenesis in RANKL-induced bone marrow-derived macrophages and ovariectomized mice by regulation of Smad3. Life Sci. 2018; 202: 11– 20.

155. Peng H, Lu S-L, Bai Y, Fang X, Huang H, Zhuang X-Q. MiR-133a inhibits fracture healing via targeting RUNX2/BMP2. Eur Rev Med Pharmacol Sci. 2018; 22(9): 2519– 26.

156. Zhang D, Cao X, Li J, Zhao G. MiR-210 inhibits NF-κB signaling pathway by targeting DR6 in osteoarthritis. Sci Rep. 2015; 5(1): 12775.

157. Lolli A, Narcisi R, Lambertini E, et al. Silencing of antichondrogenic microRNA-221 in human mesenchymal stem cells promotes cartilage repair in vivo. Stem Cells. 2016; 34(7): 1801– 11.

158. Qureshi AT, Doyle A, Chen C, et al. Photoactivated miR-148b-nanoparticle conjugates improve closure of critical size mouse calvarial defects. Acta Biomater. 2015; 12: 166– 73.

159. Eskildsen T, Taipaleenmaki H, Stenvang J, et al. MicroRNA-138 regulates osteogenic differentiation of human stromal (mesenchymal) stem cells in vivo. Proc Natl Acad Sci. 2011; 108(15): 6139– 44.

160. Song J, Kim D, Chun C-H, Jin E-J. miR-370 and miR-373 regulate the pathogenesis of osteoarthritis by modulating one-carbon metabolism via SHMT-2 and MECP-2, respectively. Aging Cell. 2015; 14(5): 826– 37.

161. Chen X, Gu S, Chen B-F, et al. Nanoparticle delivery of stable miR-199a-5p agomir improves the osteogenesis of human mesenchymal stem cells via the HIF1a pathway. Biomaterials. 2015; 53: 239– 50.

162. Fukuda T, Ochi H, Sunamura S, et al. MicroRNA-145 regulates osteoblastic differentiation by targeting the transcription factor Cbfb. FEBS Lett. 2015; 589(21): 3302– 8.

163. Wang G-L, Wu Y-B, Liu J-T, Li C-Y. Upregulation of miR-98 inhibits apoptosis in cartilage cells in osteoarthritis. Genet Test Mol Biomarkers. 2016; 20(11): 645– 53.

164. Li H, Li T, Fan J, et al. miR-216a rescues dexamethasone suppression of osteogenesis, promotes osteoblast differentiation and enhances bone formation, by regulating c-Cbl-mediated PI3K/AKT pathway. Cell Death Differ. 2015; 22(12): 1935– 45.

165. Xie Q, Wei W, Ruan J, et al. Effects of miR-146a on the osteogenesis of adipose-derived mesenchymal stem cells and bone regeneration. Sci Rep. 2017; 7(1): 42840.

166. Zhang X, Li Y, Chen YE, Chen J, Ma PX. Cell-free 3D scaffold with two-stage delivery of miRNA-26a to regenerate critical-sized bone defects. Nat Commun. 2016; 7: 10376.

167. Cheng P, Chen C, He H-B, et al. miR-148a regulates osteoclastogenesis by targeting V-maf musculoaponeurotic fibrosarcoma oncogene homolog B. J Bone Miner Res. 2013; 28(5): 1180– 90.

168. Lee WY, Li N, Lin S, Wang B, Lan HY, Li G. miRNA-29b improves bone healing in mouse fracture model. Mol Cell Endocrinol. 2016; 430: 97– 107.

169. Arfat Y, Basra MAR, Shahzad M, Majeed K, Mahmood N, Munir H. miR-208a-3p suppresses osteoblast differentiation and inhibits bone formation by targeting ACVR1. Mol Ther Nucleic Acids. 2018; 11: 323– 36.

170. Hu C-H, Sui B-D, Du F-Y, et al. miR-21 deficiency inhibits osteoclast function and prevents bone loss in mice. Sci Rep. 2017; 7(1): 43191.

171. Krzeszinski JY, Wei W, Huynh H, et al. miR-34a blocks osteoporosis and bone metastasis by inhibiting osteoclastogenesis and Tgif2. Nature. 2014; 512: 431– 5.

172. Sun Y, Xu J, Xu L, et al. MiR-503 promotes bone formation in distraction osteogenesis through suppressing Smurf1 expression. Sci Rep. 2017; 7(1): 409.

173. Deng L, Hu G, Jin L, Wang C, Niu H. Involvement of microRNA-23b in TNF-α-reduced BMSC osteogenic differentiation via targeting runx2. J Bone Miner Metab. 2018; 36(6): 648– 60.

174. Chen S, Zheng Y, Zhang S, Jia L, Zhou Y. Promotion effects of miR-375 on the osteogenic differentiation of human adipose-derived mesenchymal stem cells. Stem Cell Rep. 2017; 8(3): 773– 86.

175. Su X, Liao L, Shuai Y, et al. MiR-26a functions oppositely in osteogenic differentiation of BMSCs and ADSCs depending on distinct activation and roles of Wnt and BMP signaling pathway. Cell Death Dis. 2015; 6(8): e1851.

176. Wei J, Shi Y, Zheng L, et al. miR-34s inhibit osteoblast proliferation and differentiation in the mouse by targeting SATB2. J Cell Biol. 2012; 197(4): 509– 21.

177. Tang J, Zhang Z, Jin X, Shi H. miR-383 negatively regulates osteoblastic differentiation of bone marrow mesenchymal stem cells in rats by targeting Satb2. Bone. 2018; 114: 137– 43.

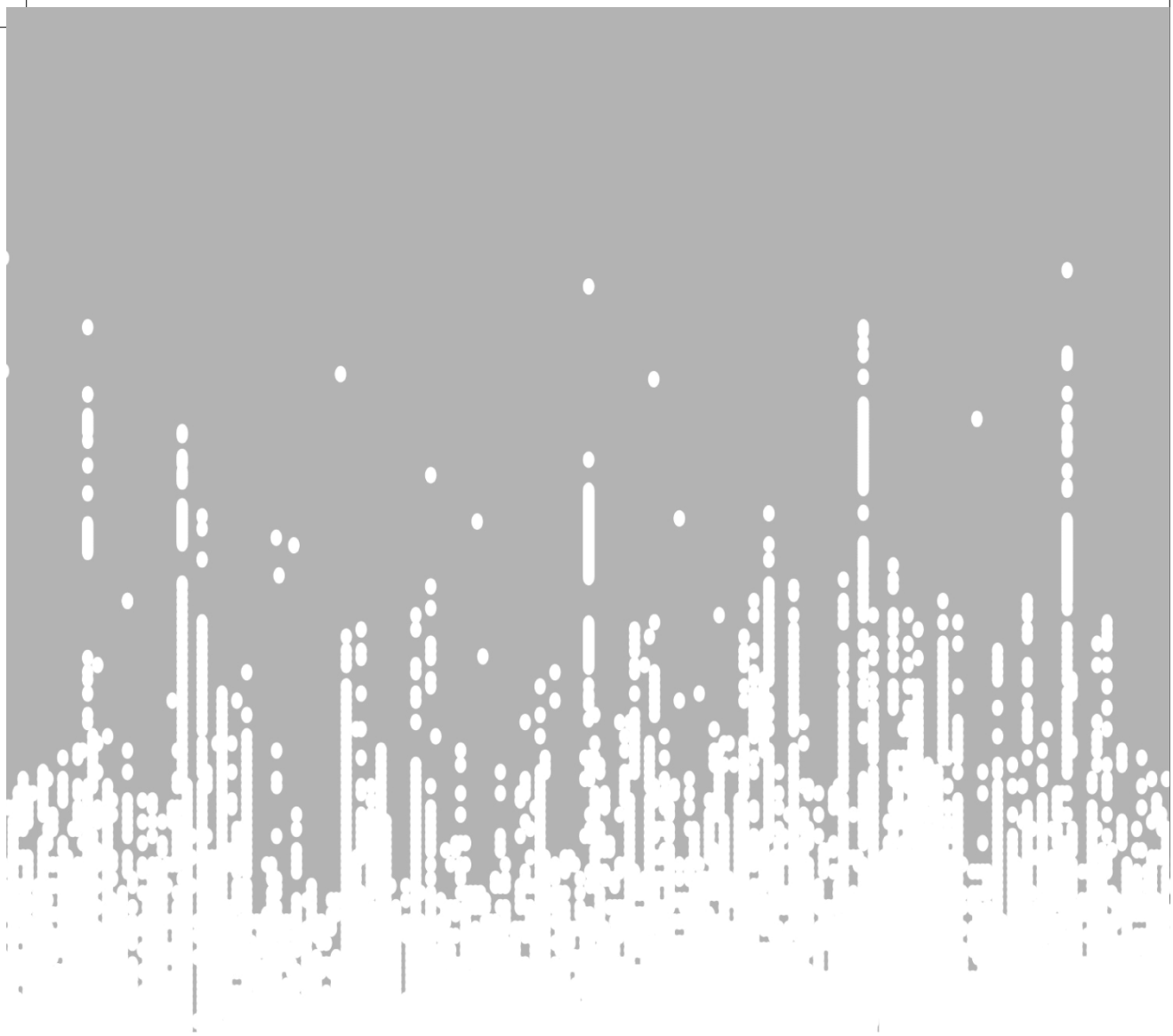
178. Tian Z, Zhou H, Xu Y, Bai J. MicroRNA-495 inhibits new bone regeneration via targeting high mobility group AT-Hook 2 (HMGA2). Med Sci Monit. 2017; 23: 4689– 98.

179. Liu L, Liu M, Li R, et al. MicroRNA-503-5p inhibits stretch-induced osteogenic differentiation and bone formation. Cell Biol Int. 2017; 41(2): 112– 23.

180. Kureel J, Dixit M, Tyagi AM, et al. miR-542-3p suppresses osteoblast cell proliferation and differentiation, targets BMP-7 signaling and inhibits bone formation. Cell Death Dis. 2014; 5(2): e 1050.

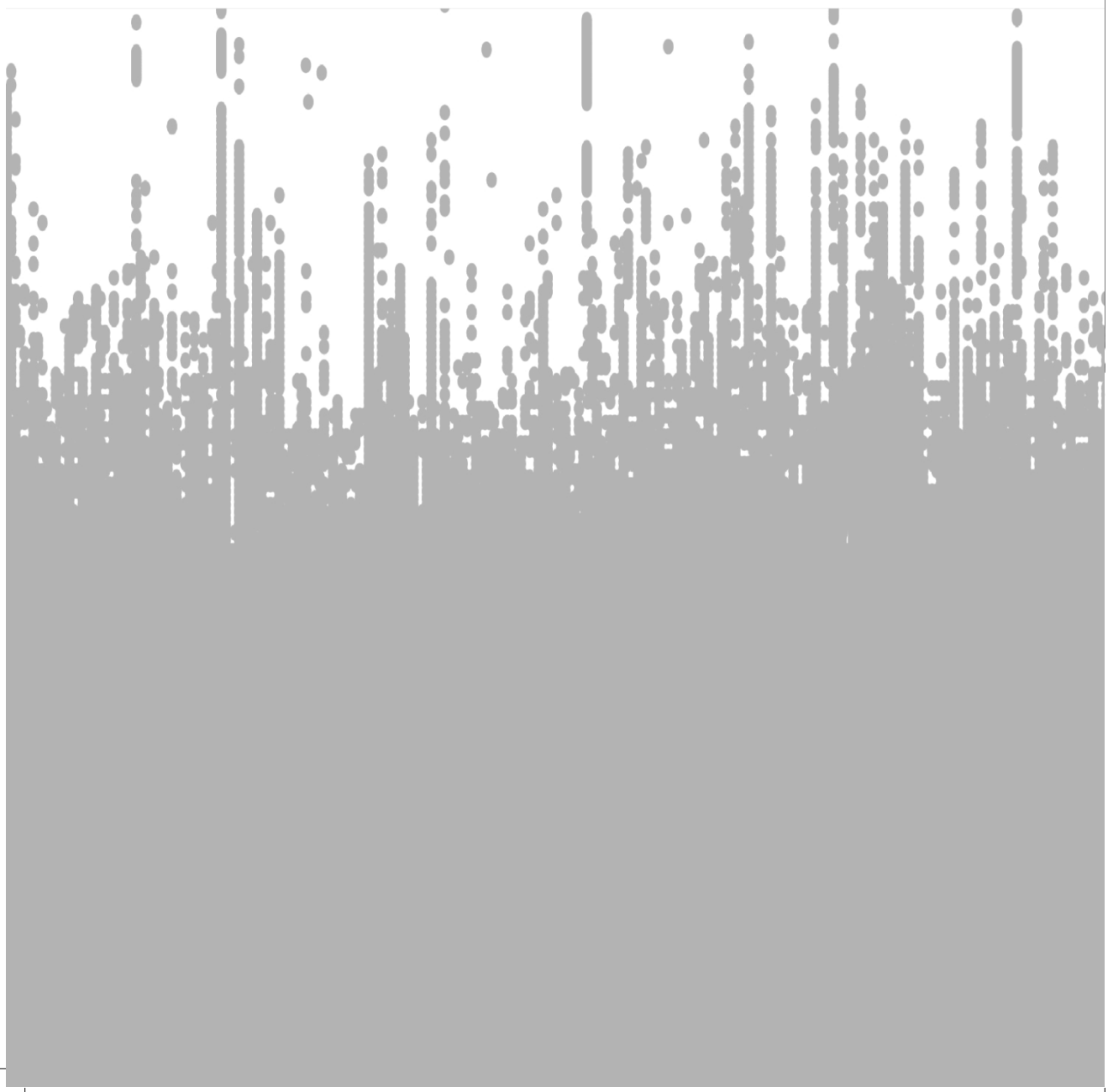
181. Murata K, Ito H, Yoshitomi H, et al. Inhibition of miR-92a enhances fracture healing via promoting angiogenesis in a model of stabilized fracture in young mice. J Bone Miner Res. 2014; 29(2): 316–26.

182. Andersson R, Gebhard C, Miguel-Escalada I, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014; 507(7493): 455– 61.



Chapter 2

PHENOTYPES OF OSTEOARTHRITIS



Chapter 2.1

Novel Genetic Variants for Cartilage Thickness and Hip Osteoarthritis

Castaño-Betancourt, M.C.*, Evans, D.S.*, Ramos, Y.F.M.*, **Boer C.G**.,*, Metrustry, S., Liu Y., den Hollander, W., van Rooij, J., Kraus V.B., Yau, M.S., Mitchell, B.D., Muir , K., Hofman, A., Doherty, M., Doherty, S., Zhang, W., Kraaij, R., Rivadeneira, F., Barrett-Connor, E., Maciewicz R.A., Arden, N., Nelissen, R.G.H.H., Kloppenburg, M.,Jordan, J.M., Nevitt, M.C, Slagboom, E.P., Hart, D.J., Lafeber, F., Styrkarsdottir, U., Zeggini, E., Evangelou,E., Spector, T.D., Uitterlinden,A.G., Lane, N.E.†, Meulenbelt, I.†, Valdes, A.M.† & van Meurs, J.B.J.†

* MCC-B, DSE, YFMR and CGB contributed equally as first authors

† NEL, IM, AMV and JBJvM contributed equally as senior authors

Published in: PLoS Genet. 2016 Oct; 12(10): e1006260.

Abstract

Objective Osteoarthritis is one of the most frequent and disabling diseases of the elderly. Only few genetic variants have been identified for osteoarthritis, which is partly due to large phenotype heterogeneity. To reduce heterogeneity, we here examined cartilage thickness, one of the structural components of joint health.

Methods We conducted a genome-wide association study of minimal joint space width (mJSW), a proxy for cartilage thickness, in a discovery set of 13,013 participants from five different cohorts and replication in 8,227 individuals from seven independent cohorts.

Results We identified five genome-wide significant (GWS, p-value≤5·0*10^{-08}) SNPs annotated to four distinct loci. In addition, we found two additional loci that were significantly replicated, but results of combined meta-analysis fell just below the genome wide significance threshold. The four novel associated genetic loci were located in/near *TGFA* (rs2862851), *PIK3R1* (rs10471753), *SLBP/FGFR3* (rs2236995), and *TREH/DDX6* (rs496547), while the other two (*DOT1L* and *SUPT3H/RUNX2*) were previously identified. A systematic prioritization for underlying causal genes was performed using diverse lines of evidence. Exome sequencing data (n = 2,050 individuals) indicated that there were no rare exonic variants that could explain the identified associations. In addition, *TGFA, FGFR3* and *PIK3R1* were differentially expressed in OA cartilage lesions versus non-lesioned cartilage in the same individuals.

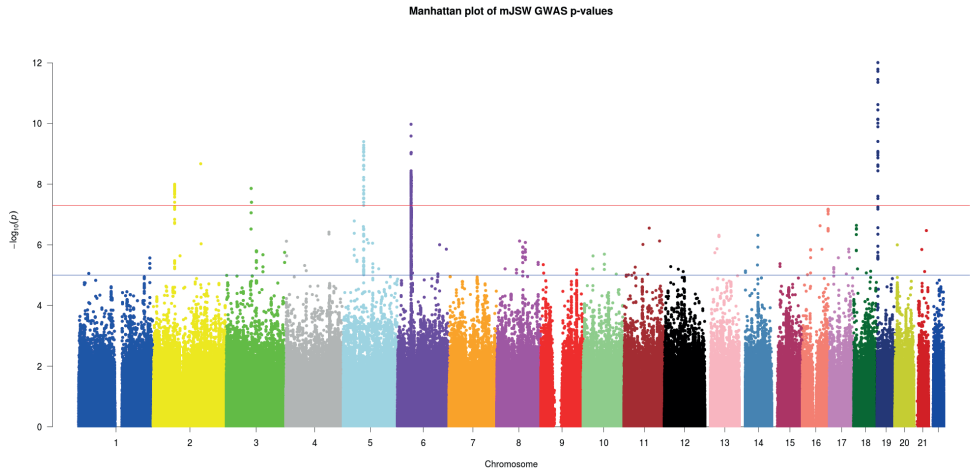
Conclusion We identified four novel loci (*TGFA, PIK3R1, FGFR3* and *TREH*) and confirmed two loci known to be associated with cartilage thickness. The identified associations were not caused by rare exonic variants. This is the first report linking *TGFA* to human OA, which may serve as a new target for future therapies.

Introduction

In spite of advances in the understanding of OA, the absence of effective therapeutic targets demonstrates that a better comprehension of its causes and pathophysiological mechanisms is needed. Since genome-wide genetic studies are hypothesis-free and do not suffer from the bias of previous knowledge, they have the potential to identify novel biological pathways involved in OA. The discovery of novel genes has the potential to identify novel treatment options. In addition, more personalized medicine approaches for OA can be explored through prediction of risk for disease as well as classification of disease subtypes.

Heritability of hip OA has been estimated to be around 40–60%. However, to date only few genetic variants have been successfully identified[1,2]. The reasons for finding only a modest number of genetic loci associated with hip OA can be attributed partially to relatively modest samples sizes in comparison to other complex diseases, such as myocardial infarction[3]. In addition, phenotype heterogeneity is an important issue in OA genetics as this is well known to reduce power to robustly detect signals. The problem of heterogeneity in genetic association studies of OA has been highlighted before[4,5] and is exemplified by the fact that the definition of the phenotype is a combination of bone and/or cartilage features as well as clinical complaints. Moreover, there is growing consensus that OA can be divided into multiple sub-phenotypes each with their own aetiology and risk factors. For example, it has been demonstrated that individuals with hip OA, where only cartilage degradation is involved (atrophic OA form), are linked to a different systemic bone phenotype compared to individuals with OA where bone formation is also present[6]. As a way to overcome this, we examined a quantitative trait, which is one of the structural components of joint health, cartilage thickness, as a distinct phenotype.

Joint Space Width (JSW) is considered to be a proxy for cartilage thickness measured on hip radiographs. Minimal JSW (mJSW) has been shown to be a more reliable measure for hip joint health compared to the classical Kellgren & Lawrence score[7]. Previously, we have demonstrated that using only a modest discovery sample size (n = 6,000), we were able to successfully identify a genome-wide significant association of the *DOT1L* locus with mJSW as well as hip OA[1,8]. We now aimed to perform a more powerful analysis by combining data from five studies in the discovery phase, and subsequent replication in seven additional studies, amounting to a total sample size of 21,240 to identify new genes implicated in joint health using mJSW as a proxy for cartilage thickness. Using whole exome sequence data from 2,050 individuals we screened the discovered genes for potential functional variants. Subsequently we used multiple



▲**Figure 1: Manhattan plot for association of mJSW in the discovery phase**. The -$\log_{10}$ p-values, for each of the 2.5 million tests performed as part of the genome wide association of minimal joint space width of the hip (mJSW), plotted against their position per chromosome. Full results of the discovery GWAS are accessible through https://www.glimdna.org. The grey solid horizontal line corresponds to the genome-wide significant threshold (p-value=$5x10^{-08}$). The dotted grey line corresponds to the selection for replication threshold (p-value=$1x10^{-05}$).

approaches that leverage different levels of information to enforce evidence of candidate genes annotated close to the associated signals.

Results

Identification of novel genetic loci associated with cartilage thickness

Genome-wide Association analysis of mJSW of the hip with genetic variants was performed in a discovery set that included 13,013 individuals (see **Supplementary Table 1** and **Supplementary Text 1** for cohort specifics) with data on ±2,5 million genotyped or HapMap Phase II imputed SNPs. We applied extensive quality control measures (see **Supplementary Table 2** and **Supplementary Table 3** for details on quality control and exclusion criteria) leaving a total of 2,385,183 SNPs available for association analyses. Genomic control inflation factors for the p-values of the RS, TwinsUK, MrOS, and SOF GWAS were low (λ=1.02, 1.01, 1.02 and 0.99 respectively), and the inter quantile-quantile plot (**Supplementary Figure 1**) also indicated no residual population stratification due to cryptic relatedness, population substructure or other biases.

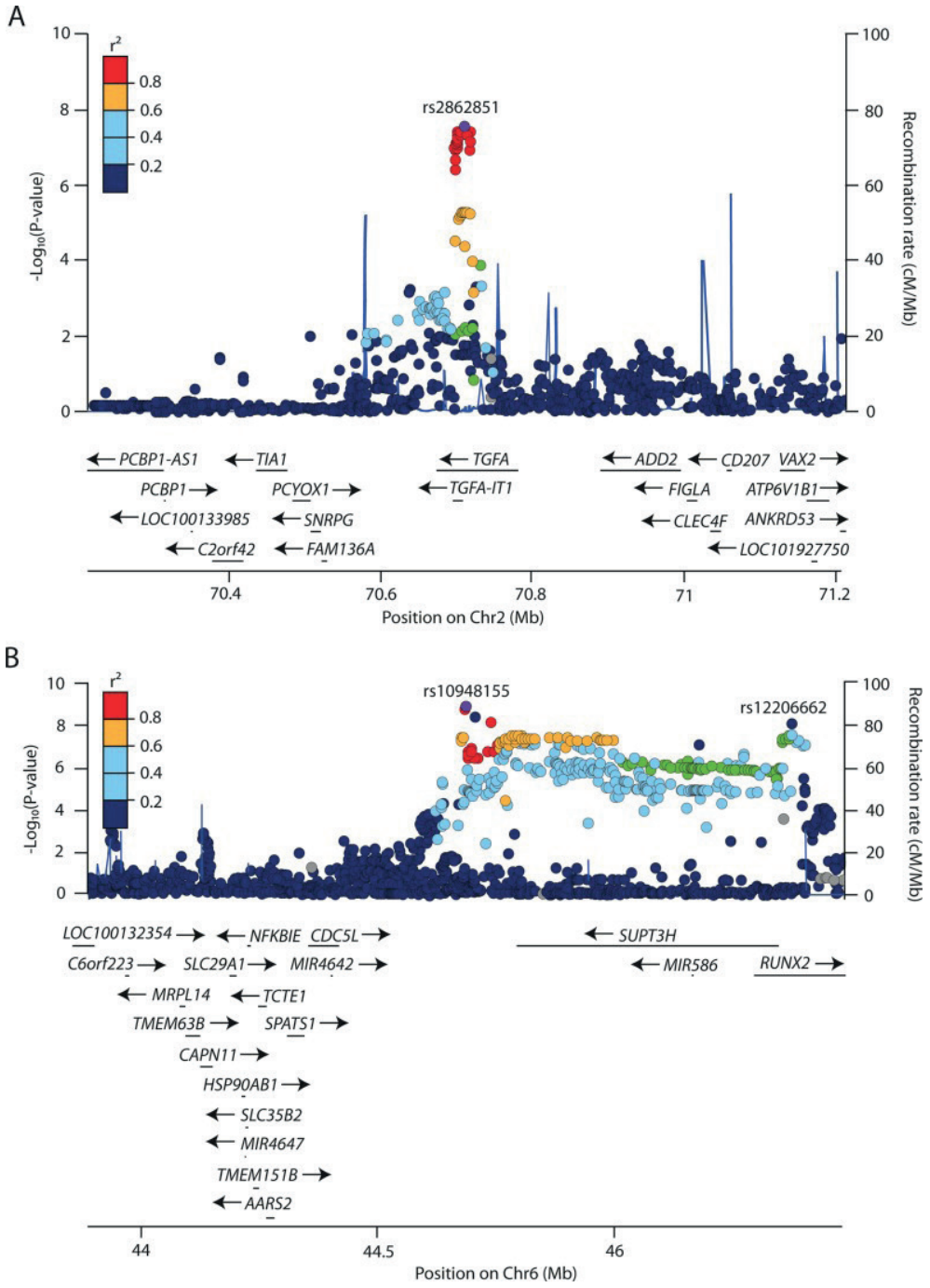
The discovery analysis yielded eighteen independent SNPs with suggestive evidence for association (p-value<$1x10^{-05}$) with mJSW, of which five (four genetic loci) met the genome-wide significance threshold of p-value≤$5*10^{-08}$ (**Figure 1**). The top SNPs

from these eighteen loci were selected for replication in an additional 8,227 individuals from seven different cohorts. We observed that six of the eighteen SNPs significantly replicated (p-value<0.05) with the same direction of effect (**Table 1**). When we combined discovery and replication results in a meta-analysis, the five SNPs that met genome-wide significance in the discovery analysis became more significant and another two SNPs that replicated in independent cohorts reached suggestive evidence (p-value≤$1*10^{-06)}$ for association in the combined meta-analysis.

The top signal in the combined meta-analysis, rs1180992 (**Table 1**, p-value$_{combined}$ = $3.2x10^{-16}$), is located in the intronic region of the previously OA associated *DOT1L* gene. This variant is very close to and in linkage disequilibrium with rs12982744 (D'=1, r^2=1), which was previously found in association with mJSW and hip OA [1,8]. The *DOT1L* signal was followed in strength of association by rs2862851 (p-value$_{combined}$ = $5.2x10^{-11}$), which is annotated to the intronic region of *TGFA* (**Figure 2a**). Two variants near *RUNX2*, rs10948155 and rs12206662, also reached genome-wide significance for association with mJSW (**Figure 2b**). The two variants in the *RUNX2* locus were weakly correlated (r^2<0.2). Conditional analysis, using GCTA, showed that both SNPs represented different signals (**Supplementary Table 4**). Finally, the last signal that reached genome-wide significance was rs10471753, an intergenic variant closer to *PIK3R1* (~450 Kb) than to *SLC30A5* (~750Kb) (**Table 1**, p-value$_{combined}$ = $3.8x10^{-9}$). Other suggestive signals for association with mJSW at a p-value≤ $1x10^{-6}$ including signals with significant replica-tion were rs496547 (p-value=$1.5*10^{-07}$), a downstream gene variant located in the 3' of *TREH* and, an intron variant annotated near *SLBP* (rs2236995; p-value=$9x10^{-07}$). All other additional signals selected in the discovery stage did not replicate.

Association of identified loci with hip OA and other musculoskeletal phenotypes

We examined whether the five GWS and two suggestive mJSW loci were also associated with hip OA in a total of 8,649 cases and >57,000 controls. Detailed description of the cohorts and OA definitions is given in **supplementary Table 1. Table 2** shows the asso-ciations found with hip OA. We observed that five of the seven identified mJSW loci were also associated with hip OA (p-value<0.05). Apart from the known *DOT1L* locus, the vari-ant near *TGFA* was significantly associated with hip OA (**Table 2**, p-value=$4.3x10^{-05}$). In addition, the SNP near *SLBP* and the two SNPs near *RUNX2* were associated with hip OA. One of the latter SNPs, rs10948155, is in high LD with a variant (rs10948172, D'=0.95 and r^2=0.90) previously found in association with hip OA in males at borderline GWS level[2]. However, in our study, rs10948155 was just marginally associated with hip OA in the overall analysis (**Table 2**, p-value=0.021).



▲ **Figure 2: Regional association plot for the locus of rs2862851** rs12206662/rs10948155. SNPs are plotted by position in a 800-kb window versus association with mJSW (–log10 P) for A. rs2862851 (TGFA locus) and B. rs12206662/rs10948155 (SUPT3H/RUNX2 locus). The purple dot highlights the most sig-nificant SNP in discovery analysis. Blue peaks indicate recombination rates. The SNPs surrounding the most significant SNP are color coded to reflect their LD with this SNP (from pairwise r2 values from the HapMap CEU). Genes, exons and the direction of transcription from the University of California at Santa Cruz genome browser are depicted. Plots were generated using Locus Zoom [9].

Table 1: Results from the minimal Joint Space Width genome-wide association study; discovery, replication and joint meta-analysis

					Discovery			Replication			Joint Meta-analysis				
SNP	Chr	AE	NEA	EAF	Beta	SE	p-value	Beta	SE	p-value	Beta	SE	p-value	N	Locus
rs11880992	19	A	G	0.40	0.10	0.01	$9.1x10^{-14}$	0.09	0.03	$1.5x10^{-03}$	0.10	0.01	$1.1x10^{-14}$	15,234	DOT1L
rs2862851	2	T	C	0.49	-0.07	0.01	$2.9x10^{-08}$	-0.06	0.02	$4.8x10^{-04}$	-0.07	0.01	$1.6x10^{-10}$	18,366	TGFA
rs10948155	6	T	C	0.67	0.08	0.01	$1.2x10^{-09}$	0.05	0.02	$1.5x10^{-02}$	0.07	0.01	$4.1x10^{-10}$	18,366	SUPT3H-RUNX2
rs12206662	6	A	G	0.07	-0.19	0.03	$7.9x10^{-09}$	-0.09	0.04	$7.3x10^{-03}$	-0.15	0.02	$6.0x10^{-10}$	16,856	SUPT3H-RUNX2
rs10471753	5	c	G	0.61	0.08	0.01	$1.5x10^{-10}$	0.02	0.02	$3.3x10^{-01}$	0.06	0.01	$4.9x10^{-09}$	17,777	PIK3R1
rs2236995	4	T	G	0.49	0.06	0.01	$6.4x10^{-06}$	0.04	0.02	$3.2x10^{-02}$	0.05	0.01	$1.5x10^{-07}$	18,366	SLBP
rs496547	11	A	T	0.34	-0.06	0.01	$1.5x10^{-06}$	-0.05	0.02	$3.8x10^{-02}$	-0.06	0.01	$4.7x10^{-07}$	16,879	TREH-DDX6
rs717433	20	T	C	0.76	0.07	0.02	$8.0x10^{-06}$	0.06	0.03	$8.3x10^{-02}$	0.07	0.01	$3.9x10^{-06}$	15,234	HAO1
rs6437120	2	T	C	0.50	-0.06	0.01	$4.0x10^{-07}$	0.006	0.02	$7.5x10^{-01}$	-0.04	0.01	$5.3x10^{-06}$	18,365	ACVR1-UPP2
rs6592847	11	A	G	0.28	0.07	0.02	$3.7x10^{-06}$	-0.01	0.06	$8.1x10^{-01}$	0.06	0.02	$9.5x10^{-06}$	13,412	NARS2
rs10495106	1	T	C	0.94	0.12	0.03	$6.2x10^{-06}$	0.05	0.04	$1.3x10^{-01}$	0.10	0.02	$1.2x10^{-05}$	18,362	TGFB2
rs6429001	1	A	G	0.83	0.08	0.02	$3.1x10^{-06}$	-0.006	0.04	$8.8x10^{-01}$	0.07	0.02	$2.4x10^{-05}$	15,233	RIR2
rs13148031	4	A	G	0.60	0.06	0.01	$9.8x10^{-06}$	0.016	0.02	$3.6x10^{-01}$	0.04	0.01	$3.7x10^{-05}$	18,366	HHIP-GYPA
rs4837613	9	c	G	0.50	-0.06	0.01	$2.0x10^{-06}$	0.002	0.02	$9.1x10^{-01}$	-0.04	0.01	$3.7x10^{-05}$	16,850	ASTN2
rs2703529	17	A	G	0.07	-0.11	0.02	$2.9x10^{-06}$	0.006	0.03	$8.6x10^{-01}$	-0.08	0.02	$4.3x10^{-05}$	18,366	HRN3P3
rs11045356	12	A	G	0.17	0.08	0.02	$4.4x10^{-06}$	-0.05	0.04	$1.5x10^{-01}$	0.06	0.02	$1.0x10^{-04}$	15,225	PPE3A
rs7739938	6	A	G	0.03	-0.31	0.07	$1.1x10^{-06}$	-0.03	0.03	$2.5x10^{-01}$	-0.08	0.02	$1.3x10^{-04}$	17,164	CITED2

Table 2: Association of mJSW loci with Hip osteoarthritis

			Hip osteoarthritis		
SNP	**Locus**	**EA**	**Beta**	**SE**	**p-value**
rs2862851	DOT1L	A	0.06	0.02	$6.9x10^{-05}$
rs2236995	TGFA	T	-0.05	0.02	$2.5x10^{-03}$
	SUPT3H-RUNX2	T	0.01	0.02	$7.6x10^{-01}$
rs10948155	SUPT3H-RUNX2	A	-0.04	0.02	$2.3x10^{-02}$
	PIK3R1	C	0.14	0.00	$1.3x10^{-04}$
rs496547	TREH-DOX6	A	0.01	0.02	$5.8x10^{-01}$
rs11880992	SLBP	T	-0.06	0.02	$9.7x10^{-05}$

SNP: single nucleotide polymorphism, EA: effect Allele, SE: standard Error.

We observed the second variant in this genomic region, an intronic variant in *RUNX2*, rs12206662, to have a larger effect size (β=0.14, p-value=$1.1*10^{-04}$, r^2=0.09 with rs10948172).

We further examined whether the identified loci were found associated with other phenotypes in earlier reports (**Table 3**). Five of the seven identified mJSW SNPs mapped to loci that have previously been associated with other bone-related phenotypes, primarily height. However, many of the identified height loci were not highly correlated with the mJSW signal (**Table 3**). Additional adjustment for height did not have an effect on the described association with mJSW; they showed an independent, possibly pleiotropic effect, on both traits. A particularly dense number of associations with different bone related phenotypes were present in the *RUNX2* 5'region, where variants have been associated to BMD[10], height[11], osteoarthritis[2] and ossification of the spine[12]. Given the low LD between the variants underlying the different GWAS signals, it is likely that these represent independent associations.

Prioritization of genes underlying the genetic loci

We used multiple approaches that leverage different levels of information (e.g., gene expression, regulatory regions, published literature, mouse phenotypes) to prioritize candidate genes at each mJSW locus. **Table 4** shows the summarized results from these analyses. In addition to the seven loci identified in the current study, we also analyzed five previously published loci for hip OA[2]. First, we focused on two gene prioritization methods: (i) DEPICT, a novel tool designed to identify the most likely causal gene in a given locus, and identify gene sets that are enriched in the genetic associations[21], and (ii) GRAIL which uses existing literature to identify connections between genes in the associated loci[22]. The DEPICT analysis yielded seventeen significantly prioritized

Table 3: Association of mJSW loci with other skeletal phenotypes

			Hip osteoarthritis	
SNP	**Locus**	**EA**	**Beta**	**SE**
rs2862851	DOT1L	A	OA[1]	rs12982744 (r^2=1.0)
			Height[13]	rs2523178(r^2=0.1)
rs2236995	TGFA	T	-	-
Rs1220662	SUPT3H-RUNX2	T	OA[2],	rs10948172(r^2=0.09)
			Height[13]	rs10948222(r^2=0.0)
			BMD[10],	rs117755164(r^2=0.04)
			OPLL[12]	927485(r^2=0.0)
rs10948155	SUPT3H-RUNX2	A	OA[2],	rs10948172(r^2=0.84)
			Height[13]	rs10948222(r^2=0.0)
			BMD[10],	rs117755164(r^2=0.21)
			OPLL[12]	927485(r^2=0.0)
Rs10471753	PIK3R1	C	-	-
rs496547	TREH-DOX6	A	Height[14]	rs494459 (r2=0.36)
			Metabolite levels[15]	rs507080(r2=21)
			Vitiligo[16]	rs638893 ((r2=0.02)
			Celiac Disease/ RS[17]	rs10892279(r2=0.36)
			SLE[18]	rs4639966(r2=0.36)
rs11880992	SLBP	T	Height[14]	rs2247341(r^2=0.6)

SNP: single nucleotide polymorphism, EA: effect Allele, SE: standard Error.

genes (FDR >0.05), from which three genes were also significantly prioritized in the GRAIL analysis (**Supplementary Table 5** and **Supplementary Table 6**). Next, using the Online Mendelian Inheritance in Man (OMIM) database (http://omim.org), we identified genes with mutations implicated in abnormal skeletal growth in humans; for 50% of the loci, a skeletal syndrome gene was present (**Supplementary Table 7**). Similarly, we investigated if any of the genes had a known bone and cartilage development phenotype in mice. Very similar to the human phenotypes, the mice knockouts of the same genes resulted in bone and cartilage phenotypes (http://mousemutant.jax.org/) (**Supplementary Table 7**). Other supporting biological evidence that we gathered consisted of known expression quantitative loci (eQTL) and nonsynonymous variants in LD (r^2>0.6) with the lead SNP of a locus (**Supplementary Table 8** and **Supplementary**

Table 9), as well as expression in bone and cartilage tissue in mice using data from the Jackson lab database (**Supplementary Table 7**).

To further explore which genes are possibly underlying the genetic associations identified in this study, we analyzed gene expression in a paired set of non-lesioned and OA-lesioned cartilage samples of the RAAK study acquired from 33 donors at the time of joint replacement surgery for primary OA[19]. We first examined which genes are expressed in a set of seven human healthy cartilage samples (**Supplementary Table 7**). Additionally, we tested which of the genes located in 1MB region surrounding the lead SNP were differentially expressed in OA-lesioned cartilage versus non-lesioned cartilage of the same hip. Of the 152 genes that were selected, 129 genes were represented on the array. Of those, 64 genes were significantly expressed in the cartilage samples. For eight of the twelve loci, we found genes that were differentially expressed in OA lesioned cartilage versus non-lesioned cartilage (**Table 4**, **Supplementary Table 10**). Differential expression in cartilage healthy vs OA affected cartilage was performed likewise (**Supplementary Table 10**), while additionally adjusting for sex and age. Given the relatively small number of healthy samples (n=7) with large age range these data are less robust and we did not use these data in gene prioritization.

For each gene a prioritization score was computed, based on equally weighting of the ten lines of evidence (**Table 4**). Following this approach, *RUNX2* is highly likely to be the causal gene associated with rs12206662 and rs10948155. Similar strong evidence is found for rs788748 (*IGFBP3*) and rs10492367 (*PTHLH*). In addition, suggestive evidence for a causal gene is found for the following: rs10471753 (*PIK3R1*), rs835487 (*CHST11*), rs2862851 (*TGFA*), rs6094710 (*SULF2*), rs9350591 (*COL12A1*) and rs11177 (*GNL3*). However for some loci the current evidence is ambiguous, suggesting more than one gene as the potentially causal one; rs2236995 (*FGFR3* or *SLBP*), rs11880992 (*GADD45B* or *DOT1L*) and rs496547(*KMT2A* or *UPK2*) (**Table 4**)

Exome sequencing of prioritised genes

In 2,628 individuals from the Rotterdam Study, exome sequencing was performed at a mean depth of 55x. Of those, 2,050 individuals also had mJSW and hip OA phenotype data. Baseline characteristics of those individuals were similar to the source population, mean age was 67.3 years, 57% of the individuals were female and mean of mJSW was 3.81 mm (SD 0.82). Details of the experimental procedure and variant calling are given in the supplementary material (**Supplementary Text 2**). Only the variants with a minimal allele count of three in the total population were selected for analysis. Within the sixteen prioritized genes, a total number of 158 variants were identified in the protein-coding region, of which 85 were non-synonymous and one was a stop-gain mutation (**Table 5**, **Supplementary Table 11**). We first performed a single variant test,

Table 4: Prioritized genes for Osteoarthritis loci. Biological and database evidence was collected to identify the causal gene in each investigated osteoarthritis associated locus. Locus gene sets were constructed by taking a region of 500 Kb upstream and 500Kb downstream of the lead SNP of that locus (1MB locus region), in total we analysed 289 genes. The tables gives a summary of the biological evidence for each osteoarthritis associated locus: 1) Nearest located genes 2) DEPICT gene prioritization results, only genes that were significantly prioritized (FDR <0.05) are listed 3) GRAIL results, only nominal significant results are listed (P<0.05) 4) Mouse cartilage or bone gene expression 5) bone or cartilage development phenotype in mouse. Mouse expression and phenotype data was obtained from the Jackson lab database (http://www.informatics.jax.org/). 6) Human cartilage tissue gene expression and 7) differential gene expression between OA cartilage tissue and Preserved cartilage tissue (Ramos et al., 2014,GEO series accession number: GSE57218) 8)OMIM phenotype 9) eQTL evidence (blood eQTL browser, and Westra et al) and 10) Nonsynonymous variants in LD >0.6 with lead SNP. All evidence for each gene is summarized, genes with the highest number of evidence are reported.

		rs10471753	rs10492367	rs10948155	rs11177	rs11880992	rs12206662	rs2236995	rs2862851	rs496547	rs6094710	rs788748	rs835487	Rr9350591
	GWAS association	mJSW	HipOA	mJSW	TJR	mJSW	mJSW	mJSW	mJSW	mJSW	Hip OA	HipOA	THR	Hip OA
	Nearest Genes	PIK3R1	KLHL42 PTHLH	CDC5LSUPT3H	GNL3	DOT1L	RUNX2	SLBP	TGFα	TRH DDX6	ZMYND8 NCOA3	IGFBP3	CHST11	FILIP1 SENP6
GRAIL		Not Significant		Not Significant	Not Significant	Not Significant	Not Significant	Not Significant	Not Significant	Not Significant	SULF2	IGFBP3	Not Significant	Not Significant
	GRAIL p-value		$6.4x10^{-03}$								$2.4x10^{-02}$	$4.4x10^{-03}$		
DEPICT	Gene	PIK3R1	KLHL42	RUNX2	No prioritization	Multiple genes prioritized	RUNX2	No prioritization	VAX2	UPK2	Multiple genes prioritized	IGFBP3	CHST11	No prioritization
	DEPICT p-value	$7.4x10^{-05}$	$5.0x10^{-03}$	$1.5x10^{-03}$			$1.5x10^{-03}$		$8.9x10^{-04}$	$3.81E^{-03}$		$2.37E^{-03}$	$1.04E^{-03}$	
Mouse	Gene	-	PTHLH	RUNX2	BAP1	GADD45B	RUNX2	FGFR3	TGFα	KMT2A	SULF2	IGFBP3	CHST11	COL12A1
	Bone & Cartilage Expr.	-	+	+	+	+	+	+	+	-	+	+	+	+
	Bone & Cartilage Phenotype	-	+	+	+	-	+	+	-	+	-	+	+	+
Human	Gene	PIK3R1	PTHLH	RUNX2	GNL3	GADD45B	RUNX2	FGFR3	TGFa	-	SULF2			
	NCOA3	IGFBP3	-	COL12A1										
	Cartilage tissue expr.	+	+	+	+	+	+	+	+	-	+	+	-	+
	Diff. expr. OA cartilage tissue	PIK3R1	PTHLH	-	GNL3	GADD45B	-	FGFR3	TGFα	Multiple genes	NCOA3	IGFBP3	TXNRD1	COL12A1

Table 4 (continued):

		rs10471753	rs10492367	rs10948155	rs11177	rs11880992	rs12206662	rs2236995	rs2862851	rs496547	rs6094710	rs788748	rs835487	Rr9350591
OMIM	Gene	PIK3R1	PTHLH	RUNX2	-	-	RUNX2	FGFR3	-	KMT2A	-	-	-	-
	OMIM skeletal disorder	+	+	+	-	-	+	+	-	+	-	-	-	-
	OMIM disorder	SHORT syndrome	Brachydactyly	Cleidocranial dysplasia	-	-	Cleido - cranial dysplasia	skeletal dysplasia	-	Wiedemann-Steiner syndrome	-	-	-	-
eQTL	Gene	-	-	SUPT3H	NT5DC2	AP3D1	-	SLBP	-	-	-	-	-	-
	Cis-eQTL p-value	-	-	4.2x10-22	1.1x10-10	1.1x10-15	-	6.8x10-10	-	-	-	-	-	-
	eQTL tissue	-	-	Whole Blood	Lympho-blastoid	Whole Blood	-	Whole Blood	-	-	-	-	-	-
SNV	Nr. Non-synonymous SNVs	-	-	-	8	1	-	-	-	-	1	-	-	-
	Gene	-	-	-	GNL3;NEK4; STAB1; SPCS1; ITIH1	DOT1L	-	-	-	-	NCOA3	-	-	-
Gene	Prioritized gene	PIK3R1	PTHLH	RUNX2	GNL3	DOT1L	RUNX2	FGFR3	TGFα	KMT2A	SULF2	IGFBP3	CHST11	COL12A1
	Lines of evidence	5	6	5	4	3	6	5	4	2	4	7	4	4
	Alternate gene	-	-	-	-	GADD45B	-	SLBP	-	UPK2	NCOA3			
	Lines of evidence	-				3		5		2	4			

where we tested each of the 86 variants changing the amino-acid sequence for association with the mJSW trait (**Supplementary Table 11**). We observed four nominal significant associations, with rare variants in *SULF2*, *TGFA*, *RUNX2* and *FGFR3*. None of these rare exonic variants explained the original association between the GWAS hit and mJSW or hipOA when tested in a multivariate model (**Supplementary Table 12**). Next, we performed a burden test (SKAT)[23], to investigate whether the cumulative effects of the variants present in the sixteen selected genes were associated to mJSW, while adjusting for age and gender (**Table 5**). We observed a nominal significance burden test (p-value<0.05) for *TGFA*, *SULF2*, *CHST11* and *RUNX2* for mJSW. However, none of these findings reached significance after correction for multiple testing.

mJSW associated SNPs in regulatory regions

For most of the loci, no obvious protein-coding variants were found that could explain the associations. In previous studies it was shown that disease-associated variants are enriched in regulatory DNA regions[24,25]. We therefore examined whether the identified DNA variants (or SNPs in high LD) resided in chondrocyte and/or osteoblast specific enhancer regions, using data from ENCODE and ROADMAP[26–28]. To this end, we compared CHIP-seq signals from five different chromatin state markers (H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me3) in chondroblasts and osteoblasts to four cell lines from another origin. Together, these chromatin state markers identify promoter and enhancer activity in each of the cell lines. With the exception of rs2862851, we observed that for all mJSW genetic loci, SNPs in high LD were located in cell regulatory regions in chondroblast and/or osteoblast cells (see **Figure 3** for an overview and **Supplementary Tables 13-18**. for each locus).

Discussion

Only a modest number of genetic variants has been successfully identified through genome-wide association studies for OA This can in part be explained by the phenotypic heterogeneity of OA. Therefore, we used mJSW, a proxy for cartilage thickness in the hip joint, as one of the structural components of joint health. An additional advantage of this phenotype is its continuous nature, which increases power compared to a dichotomous trait, such as OA-status. We identified six independent loci associated with cartilage thickness in the hip joint, of which four surpassed genome-wide significance (*TGFA*, *PIK3R1, SUPT3H-RUNX2, DOT1L*) and two were suggestive for association with mJSW (*SLBP/FGFR3, TREH-DDX6*). Four of these loci (*TGFA, SUPT3H-RUNX2, DOT1L* and *FGFR3*) were also associated with hip OA.

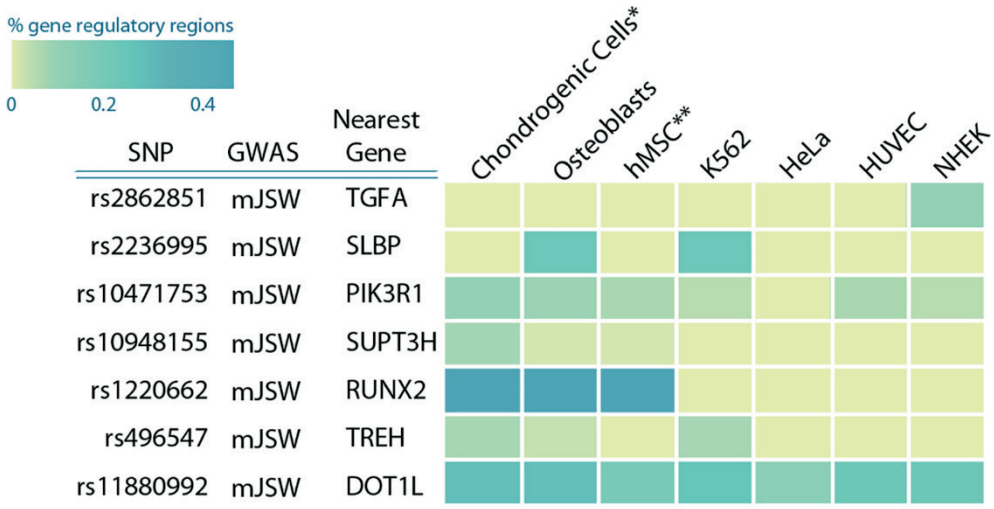
Table 5: Association between protein coding variants identified by exome-sequencing and mJSW.

	Genetic variants tested		
Gene	Non-synonymous	STOP gained	mJSW P-value*
CHST11	3	-	0.05
COL12A1	23	1	1.00
DOT1L	13	-	1.00
FGFR3	12	-	0.64
GADD45B	1	-	0.91
GNL3	5	-	0.64
IGFBP3	2	-	0.28
KMT2A	8	-	0.40
NCOA3	7	-	0.24
PIK3R1	3	-	0.80
PTHLH	0	-	-
RUNX2	2	-	0.05
SLBP	1	-	0.16
SULF3	3	-	0.05
TGFA	1	-	0.04
UPK2	1	-	0.74

*Burden Test P-value: burden test was performed between the protein-coding variants in the candidate genes and mJSW.

The fact that we were able to identify six loci with the current sample size (13K individuals in the discovery) indicates that cartilage thickness is a phenotype providing a better yield in number of discoveries than the efforts ran with traditional composite radiographic scores. As a comparison, the largest GWAS study up to now, arcOGEN with 7,4K cases and 11K controls as discovery, yielded one locus in the overall analysis, and seven additionally in a number of stratified analyses. Interestingly, in the current manuscript we report on rs10948155, which is in high LD (r^2>0.8) with a locus from arcOGEN

which was only marginally associated (p-value below genome-wide significance threshold) with OA in males only[2]. By using a cartilage specific endophenotype, evidence for this locus is elevated here to genome-wide significance in the total population, underscoring the increased power when more specific endophenotypes are used. Endophenotypes are quantifiable biological traits intermediate in the causal chain between genes and disease manifestation (in this case osteoarthritis). JSW can be precisely measured throughout the life of individuals[7] and also displays variation in normal subjects. Therefore, mJSW may be more tractable for the genetic dissection of OA. Across the cohorts in this manuscript, mJSW has been measured in different ways, using both hand measured JSW on radiographs as well as (semi) automatic software which could have added some noise to the overall meta-analysis. Future cross-calibration of JSW measurements might aid in a more precise measurement and additional power to pick up genetic loci.



* Human mesenchymal stem cell derived chondrocyte cultured cells
** Bone marrow derived cultured human mesenchymal stem cells

▲Figure 3: mJSW and OA associated variants are co-localized with potential gene regulatory markers. We examined the epigenetic histone marks in Chondrogenic cells, osteoblasts, hMSC, K562, HUVEC, HeLA and NHEK cells. This heatmap of the percentage of variants in gene regulatory regions (enhancer/promoter associated regions) in high LD (r2 >0.8) with lead GWAS SNP. Enrichment was calculated according [25].

To the best of our knowledge, we are the first to scrutinize exome variants in relation to OA identified by large scale re-sequencing. We did not find low frequency exonic variants in any of the prioritized genes that could explain the observed associations with mJSW. We do have to keep in mind that the power of the exome sequencing effort is smaller than the original discovery analysis. We were unable to examine variants with allele frequencies below 0,07%. In addition, for rare or low allele frequencies, we only had power to detect relatively large effect sizes. For example, we had 80% power to detect a beta of 0,7 mm (almost 1SD) difference for a variant with 1% allele frequency. However, we tested all of the discovered exome variants in a multivariate analysis, and found that the novel identified rare exome variants did not affect the association between the GWAS-identified variants and mJSW in the same sample. This suggests that the associations between mJSW and the identified SNPs are not explained by rare exonic variants and likely exert their effects through regulation of expression. Indeed, supporting this hypothesis, we found that many these variants (or SNPs in LD) were annotated in regions that were annotated as regulatory active in chondroblastic and/or osteoblastic cells. However, more work is needed to examine the exact biological mechanism underlying the identified genetic loci.

TGFA (Transforming Growth Factor Alpha, rs2862851) was the strongest novel locus associated with cartilage thickness and hip OA. *TGFA* has been suggested to be

involved in endochondral bone formation in mice, specifically the transition from hypertrophic cartilage to bone[29]. Recent, *TGFA* has also been implicated in the degeneration of articular cartilage during OA in rats[30]. Our results now imply a relationship between *TGFA* and human OA. In addition to the genetic association, we also show that *TGFA* expression is higher in human OA affected versus non-lesioned cartilage, possibly indicating that *TGFA* has a role in cartilage remodelling.

Functional characterization of the *TGFA*-associated locus by an examination of the histone methylation marks representing promoter or enhancer activity, did not reveal an obvious explanation for the functional impact of the SNP. However, the examined histone mark data represent unstimulated cells, and it is anticipated that the promoter and enhancer activity change upon stimulation of the cells. It is becoming more clear that effects of SNPs can be stimulus and context dependent, as has recently been shown for human monocytes, where many regulatory variants display functionality only after pathophysiological relevant immune stimuli[31].

The identified *SUPT3H-RUNX2* locus contains two variants, rs12206662 and rs10948155, which are partially independent of each other. Where rs12206662 is located in the first intron of the *RUNX2* gene near the second transcription start site (the so-called P2 promoter), rs10948155 is located more than 500kb away from *RUNX2* between *CDC5L* and *SUPT3H*. However, rs10948155 is in high linkage disequilibrium with SNPs near in the P2 promoter and SNPs located in chondroblast specific enhancer regions. Possibly, these enhancer regions regulate *RUNX2* gene expression during endochondral differentiation. *RUNX2*(Runt-related transcription factor 2) is a master transcription factor for controlling chondrocyte hypertrophy and osteoblast differentiation[32]. Previous genome-wide association studies have identified variants in the *SUPT3H-RUNX2* locus associated with other bone and cartilage related phenotypes including height[14], bone mineral density[10] and ossification of the posterior longitudinal ligament of the spine[12]. All these previously published loci are independent of the two mJSW SNPs identified in the current study. We hypothesize that the SNPs are located in long-range enhancers, which regulate *RUNX2* gene expression during endochondral differentiation via a chromatin-loop mediating protein.

We have also identified rs10471753, with *PIK3R1*(Phosphoinositide-3-Kinase, Regulatory subunit 1 alpha) as the closest and strongest prioritized gene, related to rs10471753 associated with mJSW. Mutations in this gene are known to cause the SHORT syndrome, which is a rare multisystem disease with several manifestations including short stature, hernias, hyper extensibility and delayed dentition [33]. Taken together with the fact that *PIK3R1* is differentially expressed in OA affected cartilage, these results identify *PIK3R1* as the most likely causal gene. Another possibility is that

not *PIK3R1* but rather a long-non-coding RNA (lncRNA), *lnc-PIK3R1-4:1*, is causal, since a variant in LD with the lead SNP is located in the predicted transcription start site of this lncRNA potentially affecting its expression. Although conserved in mice and zebrafish, thus far no function has been ascribed to this lncRNA [34].

We confirmed a locus previously associated with cartilage thickness, the *DOT1L* locus. Our identified SNP, rs11880992 is in high LD with the previously reported SNP rs129827744, and both are associated with cartilage thickness and hip OA[1]. Despite the previously presented suggestive evidence for involvement of *DOT1L* in chondrogenic differentiation, *DOT1L* did not receive a high score in our systematic prioritization study; the gene *GADD45B,* located in the region 500Kb downstream of the lead SNP, received a similar score. *GADD45B* is a transcriptional co-factor for *C/EBP-β*, a master regulator of chondrocyte differentiation[35]. Thus, it remains unclear which gene or genes in this locus contribute to the cartilage phenotype. Further research is needed to determine whether *DOT1L* is the true causal gene in this locus.

Our analyses suggest that the majority of prioritized genes in hip OA associated loci are involved in cartilage and bone developmental pathways; including *TGFA*, *RUNX2*, *FGFR3*, *PTHLH*, *COL12A1* and others that seem to affect bone and/or cartilage development such as *PIK3R1* and *KMT2A* We hypothesize that the mJSW and OA associated variants influence gene expression regulation. The dysregulation of these genes and mechanisms during development may, later in life, result in an increased risk for OA.

The identified mJSW SNPs are associated with hip OA, but not with knee OA. We have analysed the identified SNPs also for association with knee OA in the TREAT-OA meta-analysis dataset[36], but found no association. This observation fits in the overall finding that many of the identified genetic loci for OA seem to be site-specific [37], and support the hypothesis that the aetiology of OA is different in each joint. Nevertheless, this observation can still be a result of low power in the GWAS studies that have been done for OA till now[38], and final conclusions on this aspect cannot be drawn at this point.

This is the first report linking *TGFA* to human OA most likely by affecting mJSW. It may serve as a new target for future therapies. We have identified multiple mJSW associated loci which have previously been associated with other bone and cartilage related phenotypes such as bone mineral density and height, displaying a possible pleiotropic effect for the analysed traits. It will be important to understand how mJSW and OA associated variants can affect the developmental processes that regulate morphometry of the hip joint, including the formation of articular cartilage. Therefore further expression and functional studies are warranted of genes identified to be associated with hip OA phenotypes.

Materials and Methods

Ethics statement

The participating studies were approved by the medical ethics committees of all participating centres, and all participants gave their written informed consent before entering the study

Discovery GWAS, replication and meta-analysis

We conducted genome-wide association studies of mJSW for each cohort of the discovery stage: Rotterdam Study I (RS-I), Rotterdam Study II (RS-II), TwinsUK, SOF and MrOS using standardized age-, gender and population stratification (four principal components) adjusted residuals from linear regression. Cohort description and details of the single GWAS studies are given in **Supplementary Text 1** and **Supplementary Table 1.** The 6 cohorts used in the discovery stage were combined in a joined meta-analysis using inverse variance weighting with METAL[39]. Genomic control correction was applied to the standard errors and P-values before meta-analysis. SNPs with a p-value<=5×10^{-6} were selected for replication. The top SNPs for each independent locus were taken for replication in seven studies: the Genetics of Osteoarthritis and Lifestyle (GOAL) study, the Chingford study, CHECK (Cohort Hip & Cohort Knee), Genetics osteoARthritis and progression (GARP) study, the Genetics of Generalized Osteoarthritis (GOGO), the Johnston County Osteoarthritis Project (JoCo) and additionally the Nottingham OA case-control study for association with Hip OA (see Supplemental material for detailed information of the cohorts). Association of the SNPs with mJSW was additionally adjusted for height to test its independence. Secondary analyses included: association of the top SNPs with hip OA using logistic regression analysis (age and gender adjusted and by study centres an/or relatedness when it was pertinent). We used conditional analyses to investigate whether there are any independent signals in the identified associated loci, which were implemented using GCTA-COJO analysis[40].

Phenotype description of minimal Joint Space width (mJSW)

The mJSW was assessed at pelvic radiographs in anterior-posterior position. The mJSW was measured in mm, along a radius from the center of the femoral head, and defined as the shortest distance found from the femoral head to the acetabulum. Within the Rotterdam Study, we used a 0.5 mm graduated magnifying glass laid directly over the radiograph to measure the minimal joint space width of the hip joints[41]. Within SOF and MrOS, a handheld caliper and reticule was used to measured mJSW to the nearest 0.1mm between the acetabular rim and proximal head of the femur[42]. For CHECK, mJSW was measured semi-automatic with the Software tool HOLY[43].

Phenotype description of radiographic hip OA

Radiographic hip OA was defined in the RS-I, RS-II, RSIII, Twins-UK, Chingford, and JoCo studies using Kellgren and Lawrence (KL) grade. Hip OA cases were defined as a KL grade≥2 on either side of the hip or THR due to OA. Hip OA controls were defined as no THR for OA and KL grade≤1 and JSN≤1. In MrOS and SOF cohorts, radiographic hip OA case-control was defined by a modified Croft grade, as previously described[44], where cases were defined as a Croft score ≥ 2 on either side of the hip or THR due to OA and controls were defined as a Croft score ≤ 1 on both sides of the hip and no THR. Hip OA cases in the GOAL and Nottingham OA studies were defined by having THR, and controls were radiographically free of hip OA, as previously described [45]. In GARP, hip osteoarthritis was defined as pain or stiffness in the groin and hip region on most days of the preceding month in addition to femoral or acetabular osteophytes or axial joint space narrowing on radiography or prosthesis due to osteoarthritis. In GOGO, hip OA was defined as KL grade>=2, or minimal joint space width<=2.5 mm, or the combination of joint space narrowing grade>=2 and any osteophyte of grade>=1, or history of joint replacement for OA. In JoCo, hip OA cases were defined as KL grade>=2 or THR in at least one hip. Hip OA controls were defined as KL grade<= 1 in both hips.

Gene prioritization analysis

We have used several available tools and publicly available databases to prioritize genes in known and newly discovered osteoarthritis associated regions. Locus gene sets were constructed by taking a region of 500Kb upstream and 500Kb downstream of the lead SNP of that locus. We analysed 152 genes in 13 independent loci associated with minimal joint space width in the hip joint (mJSW) for 7 loci, hip OA for 4 loci, total joint replacement (TJR) for 1 locus and total hip replacement (THR) for 1 locus[2]. We analysed the following biological evidence for each gene at all loci; Nearest located genes: Taken from the UCSC genome browser, GRCh37/hg19[46]. DEPICT gene prioritization: Data-driven Expression-Prioritized Integration for Complex Traits, a novel tool designed to identify the most likely causal gene in a given locus and to gene sets that are enriched in the genetic associations[21]. DEPICT was used to prioritize genes in a 1MB region around the found SNPs that were significant associated with the osteoarthritis phenotype, taking a region of 500Kb upstream and 500Kb downstream of the lead SNP of that locus. Gene prioritization analysis was performed to directly investigate functional similarities among genes from different associated regions, significance was defined by false discovery rate (FDR ≤ 5%). GRAIL gene prioritization: Gene Relationships Across Implicated Loci (GRAIL), was used to determine connectively between genes across OA implicated loci based on literature associations [22]. A GRAIL analysis was performed

on 10 independent OA associated loci, based on existing literature in PubMed till August 2014. Mouse gene expression and phenotype: For each investigated gene, expression in mouse bone and/or cartilage tissue during several developmental stages as well as for adult tissue was determined using data from the Jackson lab database (http://www.informatics.jax.org/). In addition mouse phenotype data was also obtained for each gene. OMIM phenotype: Using the Online Mendelian Inheritance in Man (OMIM) database we examined which genes were involved in abnormal skeletal growth syndromes when mutated (http://omim.org). Expression quantitative trait loci: eQTL information was taken from the Blood eQTL browser (http://genenetwork.nl/bloodeqtlbrowser/) and the eQTL browser (http://www.ncbi.nlm.nih.gov/projects/gap/eqtl/index.cgi) using the lead SNP in each locus [20]. Non-synonymous variants: Last we determined if there were any nonsynonymous variants in LD (r^2>0.6) with the lead SNP of a locus, using HaploReg V2 and the SNP Annotation and Proxy Search (SNAP) tools[47,48]. For each gene we assigned a score based on equally weighted lines of evidence.

Human cartilage gene expression

We have used cartilage samples from the RAAK study to study gene expression in preserved and affected cartilage from individuals undergoing joint replacement[19]. The ongoing Research Arthritis and Articular Cartilage (RAAK) study is aimed at the biobanking of blood, joint materials (cartilage, bone and where available ligaments of knees and hips) and bone marrow stem cells (hip joints only) of patients and controls in the Leiden University Medical Center and collaborating outpatient clinics in the Leiden area. 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-bearing area of the joint. Cartilage was classified macroscopically and collected separately for macroscopically OA affected and preserved regions. Classification was done according to predefined features for OA related damage based on colour/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. Tissues have been stored tailored to apply staining and immunohistochemistry(IHC). Furthermore, DNA and RNA have been isolated from the preserved and affected areas of the respective tissues in order to apply genetic, transcriptomic and epigenomic profiling with respect to the OA pathophysiological process.

After in vitro transcription, amplification, and labeling with biotin-labeled nucleotides (Illumina TotalPrep RNA Amplification Kit) Illumina HumanHT-12 v3 mi-

croarrays 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. After removal of probes that were not optimally measured (detection p-value>0.05 in more than 50% of the samples) a paired t-test was performed 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.

Exome sequencing

Exome sequencing was performed in 2628 individuals from the Rotterdam Study the average mean coverage was 55x, corresponding to approximately 80% of the targeted regions covered by at least 20 reads. The exome sequencing was performed in house (HuGe-F, www.glimDNA.org). Details of the technical procedure and variant calling are given in S2 Text. We tested the exome variants for association with mJSW and/or hip OA in two ways. Each individual variant was tested for association with mJSW using the single variant option within RV-test, while adjusting for age and sex. In addition, we did a burden test for each of the selected genes by using SNP-set kernel association test (SKAT-O). SKAT aggregates individual score test statistics of SNPs in a SNP set and computes SNP-set level p-values for a gene[23].

Visualization of the regulatory landscape of mJSW associated loci

For each of the top mJSW GWAS associated SNPs the LD region was determined using the 1000G Phase 1 population using the Haploreg tool [47]. The LD threshold was set at $r^2 \geq 0.8$. For each of these SNPs it was determined if the variant was located in a potential enhancer region using the Roadmap consortium reference epigenomes data set[27]. Heatmaps were constructed by calculating the percentage of variants in LD with the top mJSW GWAS found SNP located in enhancer regions as defined by the Roadmap epigenome chromatin states. The reference epigenomes were downloaded from the official data portal accompanying[27]. Reference epigenome data was used from mesenchymal stem cell derived chondrocyte cultured cells, Osteoblast, Bone marrow derived cultured mesenchymal stem cells, K562, HUVEC, HeLA and NHEK cells. Reference epigenomes were chromatin state models based on ChIPseq data of 5 core histone marks

(H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me3) and an additional H3K27ac histone mark, the Roadmap expanded 18-state model.

ChIPseq data of mesenchymal stem cell derived chondrocyte cultured cells, and bone marrow derived cultured mesenchymal stem cells were generated by the NHI roadmap epigenomics project [28]. ChIPseq data of, Osteoblast, K562, HUVEC, HeLA and NHEK cells were generated by the ENCODE consortium[26]. All data and annotation tracks were downloaded through the UCSC genome browser table tool. Visualization of all ChIPseq annotation and roadmap full epigenomes tracks was done through the UCSC genome browser on GRCh37/hg19. Heatmaps were plotted in R using the CRAN software packages gplots and RcolorBrewer. Enrichment was calculated according to methods described in Trynka et al[25].

Acknowledgments

The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists.

We thank Mr. Pascal Arp, Ms. Mila Jhamai, and Mr. Marijn Verkerk for their help in creating the RS-Exome Sequencing database. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters, MSc, and Carolina Medina-Gomez, MSc, for their help in creating the GWAS database, and Karol Estrada, PhD, Yurii Aulchenko, PhD, and Carolina Medina-Gomez, MSc, for the creation and analysis of imputed data.

Supplementary Material

Additional material, as well as a funding statement, is published online only, to view please visit the paper online at the journal website:

https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006260

References

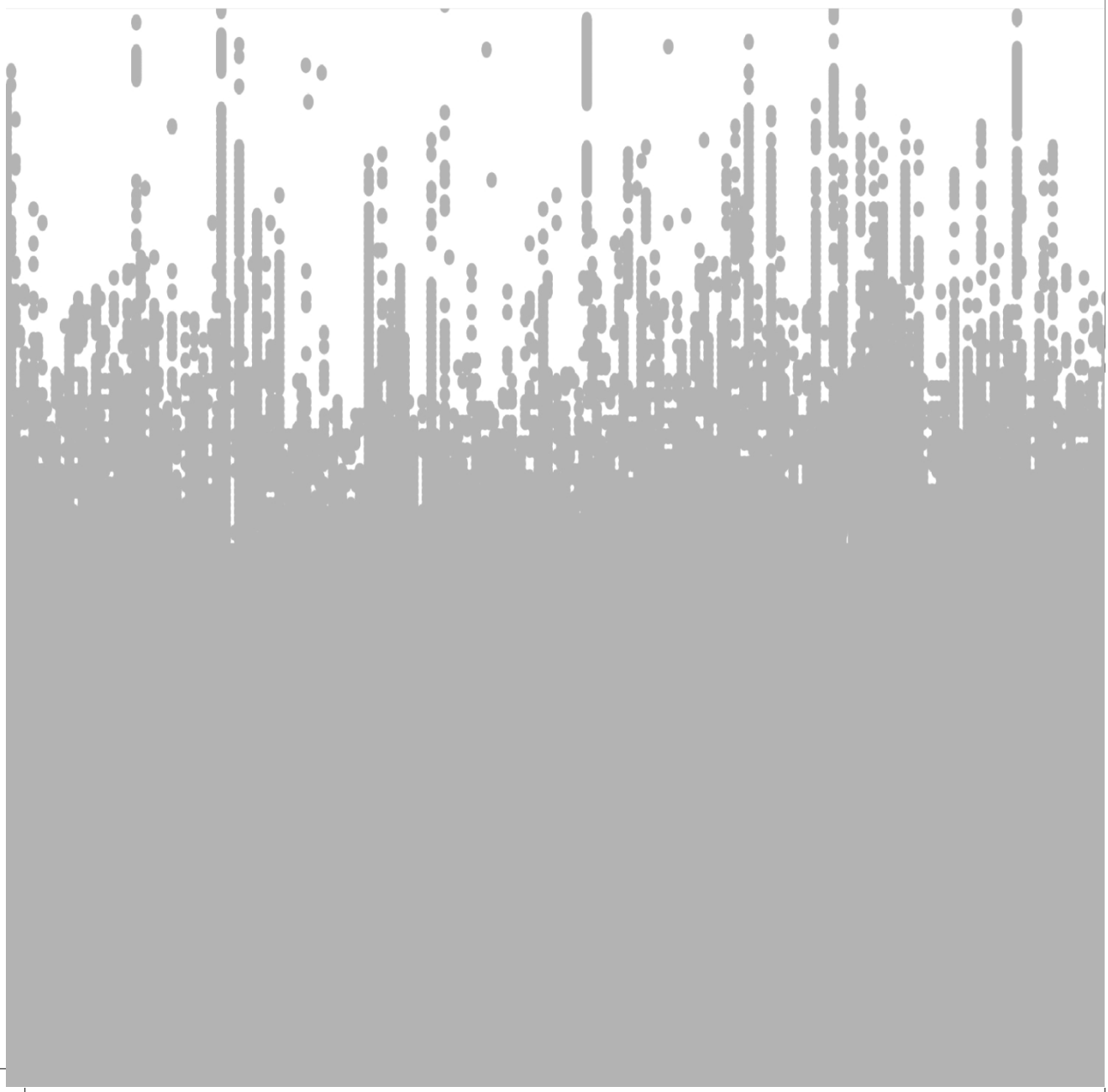
1. Castaño Betancourt MC, Cailotto F, Kerkhof HJ, Cornelis FMF, Doherty SA, Hart DJ, et al. Genome-wide association and functional studies identify the DOT1L gene to be involved in cartilage thickness and hip osteoarthritis. Proc Natl Acad Sci U S A. 2012;109: 8218–23. doi: 10.1073/pnas.1119899109
2. Consortium A, Collaborators A. Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. Lancet. Elsevier Ltd; 2012;380: 815–23. doi: 10.1016/S0140-6736(12)60681-3
3. Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, Mannucci PM, et al. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. Nat Genet. 2009;41: 334–341. doi: 10.1038/ng.327
4. Kerkhof HJM, Meulenbelt I, Akune T, Arden NK, Aromaa A, Bierma-Zeinstra SMA, et al. Recommendations for standardization and phenotype definitions in genetic studies of osteoarthritis: the TREAT-OA consortium. Osteoarthritis Cartilage. 2011;19: 254–64. doi: 10.1016/j.joca.2010.10.027
5. 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: 864–867. doi: 10.1136/ard.2010.141473
6. Castaño-Betancourt MC, Rivadeneira F, Bierma-Zeinstra S, Kerkhof HJM, 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. doi: 10.1002/art.37792 [PubMed]
7. Ingvarsson T, Hägglund G, Lindberg H, Lohmander LS. Assessment of primary hip osteoarthritis: comparison of radiographic methods using colon radiographs. Ann Rheum Dis. 2000;59: 650–3.
8. 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. doi: 10.1136/annrheumdis-2012-203182
9. 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: 2336–2337. doi: 10.1093/bioinformatics/btq419
10. Estrada K, Styrkarsdottir U, Evangelou E, Hsu Y-H, Duncan EL, Ntzani EE, et al. Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. Nat Genet. 2012;44: 491–501. doi: 10.1038/ng.2249
11. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet. 2014;46: 1173–86. doi: 10.1038/ng.3097
12. Nakajima M, Takahashi A, Tsuji T, Karasugi T, Baba H, Uchida K, et al. A genome-wide association study identifies susceptibility loci for ossification of the posterior longitudinal ligament of the spine. Nat Genet. Nature Publishing Group; 2014;46: 1012–6. doi: 10.1038/ng.3045
13. Berndt SI, Gustafsson S, Mägi 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. doi: 10.1038/ng.2606
14. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. Nature Publishing Group; 2010;467: 832–8. doi: 10.1038/nature09410
15. Yu B, Zheng Y, Alexander D, Morrison AC, Coresh J, Boerwinkle E. Genetic determinants influencing human serum metabolome among African Americans. PLoS Genet. 2014;10: e1004212 doi: 10.1371/journal.pgen.1004212
16. Tang X-F, Zhang Z, Hu D-Y, Xu A-E, Zhou H-S, Sun L-D, et al. Association analyses identify three susceptibility Loci for vitiligo in the Chinese Han population. J Invest Dermatol. 2013;133: 403–10. doi: 10.1038/jid.2012.320
17. Zhernakova A, Stahl EA, Trynka G, Raychaudhuri S, Festen EA, Franke L, et al. Meta-analysis of ge-

nome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. PLoS Genet. 2011;7: e1002004 doi: 10.1371/journal.pgen.1002004

18. Han J-W, Zheng H-F, Cui Y, Sun L-D, Ye D-Q, Hu Z, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat Genet. 2009;41: 1234–7. doi: 10.1038/ng.472
19. Ramos YFM, den Hollander W, Bovée JVMG, 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 doi: 10.1371/journal.pone.0103056
20. Westra H-J, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2013;45: 1238–1243. doi: 10.1038/ng.2756
21. Pers TH, Karjalainen JM, Chan Y, Westra J, Wood AR, Yang J, et al. Biological interpretation of genome-wide association studies using predicted gene functions. Nat Commun. Nature Publishing Group; 2015;6: 5890 doi: 10.1038/ncomms6890
22. Raychaudhuri S. VIZ-GRAIL: Visualizing functional connections across disease loci. Bioinformatics. 2011;27: 1589–1590. doi: 10.1093/bioinformatics/btr185
23. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011;89: 82–93. doi: 10.1016/j.ajhg.2011.05.029
24. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337: 1190–5. doi: 10.1126/science.1222794
25. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat Genet. 2013;45: 124–30. doi: 10.1038/ng.2504
26. Rosenbloom KR, Sloan C a, Malladi VS, Dreszer TR, Learned K, Kirkup VM, et al. ENCODE data in the UCSC Genome Browser: year 5 update. Nucleic Acids Res. 2013;41: D56–63. doi: 10.1093/nar/gks1172
27. Consortium RE, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. doi: 10.1038/nature14248 [PMC free article] [PubMed]
28. Bernstein BE, Stamatoyannopoulos J a, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. Nature Publishing Group; 2010;28: 1045–8. doi: 10.1038/nbt1010-1045
29. Appleton CTG, Usmani SE, Bernier SM, Aigner T, Beier F. Transforming growth factor alpha suppression of articular chondrocyte phenotype and Sox9 expression in a rat model of osteoarthritis. Arthritis Rheum. 2007;56: 3693–705. doi: 10.1002/art.22968
30. Usmani SE, Pest M a, Kim G, Ohora SN, Qin L, Beier F. Transforming growth factor alpha controls the transition from hypertrophic cartilage to bone during endochondral bone growth. Bone. Elsevier Inc.; 2012;51: 131–41. doi: 10.1016/j.bone.2012.04.012
31. Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, Lau E, et al. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science. 2014;343: 1246949 doi: 10.1126/science.1246949
32. Enomoto H, Enomoto-Iwamoto M, Iwamoto M, Nomura S, Himeno M, Kitamura Y, et al. Cbfa1 is a positive regulatory factor in chondrocyte maturation. J Biol Chem. 2000;275: 8695–702. Available: http://www.ncbi.nlm.nih.gov/pubmed/10722711
33. Dyment D a, Smith AC, Alcantara D, Schwartzentruber J a, Basel-Vanagaite L, Curry CJ, et al. Mutations in PIK3R1 Cause SHORT Syndrome. Am J Hum Genet. 2013; 158–166. doi: 10.1016/j.ajhg.2013.06.005
34. Volders P-J, Verheggen K, Menschaert G, Vandepoele K, Martens L, Vandesompele J, et al. An update on LNCipedia: a database for annotated human lncRNA sequences. Nucleic Acids Res.

2014;43: D174–D180. doi: 10.1093/nar/gku1060

35. Tsuchimochi K, Otero M, Dragomir CL, Plumb DA, Zerbini LF, Libermann TA, et al. GADD45beta enhances Col10a1 transcription via the MTK1/MKK3/6/p38 axis and activation of C/EBPbeta-TAD4 in terminally differentiating chondrocytes. J Biol Chem. 2010;285: 8395–407. doi: 10.1074/jbc.M109.038638
36. Evangelou E, Valdes a M, Kerkhof HJM, Styrkarsdottir U, Zhu YY, Meulenbelt I, et al. Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. Ann Rheum Dis. 2013;70: 349–355. doi: 10.1136/ard.2010.132787.Meta-analysis
37. Reynard LN. Analysis of genetics and DNA methylation in osteoarthritis: What have we learnt about the disease? Semin Cell Dev Biol. 2016; doi: 10.1016/j.semcdb.2016.04.017
38. van Meurs JBJ, Uitterlinden a G. Osteoarthritis year 2012 in review: genetics and genomics. Osteoarthritis Cartilage. Elsevier Ltd; 2012;20: 1470–6. doi: 10.1016/j.joca.2012.08.007
39. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26: 2190–2191. doi: 10.1093/bioinformatics/btq340
40. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet. The American Society of Human Genetics; 2011;88: 76–82. doi: 10.1016/j.ajhg.2010.11.011
41. Reijman M, Hazes JMW, Pols HAP, Bernsen RMD, Koes BW, Bierma-Zeinstra SMA. Validity and reliability of three definitions of hip osteoarthritis: cross sectional and longitudinal approach. Ann Rheum Dis. 2004;63: 1427–33. doi: 10.1136/ard.2003.016477 [PMC free article] [PubMed]
42. Lane NE, Nevitt MC, Hochberg MC, Hung Y-Y, Palermo L. Progression of radiographic hip osteoarthritis over eight years in a community sample of elderly white women. Arthritis Rheum. 2004;50: 1477–86. doi: 10.1002/art.20213
43. Kinds MB, Vincken KL, Vignon EP, ten Wolde S, Bijlsma JWJ, Welsing PMJ, et al. Radiographic features of knee and hip osteoarthritis represent characteristics of an individual, in addition to severity of osteoarthritis. Scand J Rheumatol. 2012;41: 141–9. doi: 10.3109/03009742.2011.617311
44. Nevitt MC, Lane NE, Scott JC, Hochberg MC, Pressman a. R, Genant HK, et al. Radiographic osteoarthritis of the hip and bone mineral density. The Study of Osteoporotic Fractures Research Group. Arthritis Rheum. 1995;38: 907–16.
45. Valdes AM, McWilliams D, Arden NK, Doherty SA, Wheeler M, Muir KR, et al. Involvement of different risk factors in clinically severe large joint osteoarthritis according to the presence of hand interphalangeal nodes. Arthritis Rheum. 2010;62: 2688–95. doi: 10.1002/art.27574
46. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler a. M, et al. The Human Genome Browser at UCSC. Genome Res. 2002;12: 996–1006. doi: 10.1101/gr.229102
47. Ward LD, Kellis M. HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012;40: 930–934. doi: 10.1093/nar/gkr917
48. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, De Bakker PIW. SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 2008;24: 2938–2939. doi: 10.1093/bioinformatics/btn564



Chapter 2.2

Genome-wide association of phenotypes based on clustering patterns of hand osteoarthritis identify WNT9A as novel osteoarthritis gene

Boer, C.G., Yau, M.S., Rice, S.J., Coutinho de Almeida, R., Cheung, K., Styrkarsdottir, U., Southam, L., Broer, L., Wilkinson, J.M., Uitterlinden, A.G., Zeggini, E., Felson, D., Loughlin, J., Young, M., Capellini, T., Meulenbelt, I., and van Meurs, J.B.J.

Published in Ann Rheum Dis. 2020 Oct 14: annrheumdis-2020-217834

Abstract

Background: Despite recent advances in the understanding of the genetic architecture of osteoarthritis (OA), only two genetic loci have been identified for OA of the hand, in part explained by the complexity of the different hand joints and heterogeneity of OA pathology..

Methods: We used data from the Rotterdam Study (RSI, RSII and RSIII) to create three hand OA phenotypes based on clustering patterns of radiographic OA severity to increase power in our modest discovery GWAS in the Rotterdam Study (RS) (n=8,700), and sought replication in an independent cohort, the Framingham Heart Study (n=1,203). We used multiple approaches that leverage different levels of information and functional data to further investigate the underlying biological mechanisms and candidate genes for replicated loci. We also attempted to replicate known OA loci at other joint sites, including the hips and knees.

Results: We found two novel genome-wide significant loci for OA in the thumb joints. We identified WNT9A as a possible novel causal gene involved in OA pathogenesis. Furthermore, several previously identified genetic loci for OA seem to confer risk for OA across multiple joints: TGFa, RUNX2, COL27A1, ASTN2, IL11 and GDF5-loci.

Conclusions: We identified a robust novel genetic locus for hand OA on chromosome 1, of which WNT9A is the most likely causal gene. In addition, multiple genetic loci were identified to be associated with OA across multiple joints. Our study confirms the potential for novel insight into the genetic architecture of OA by using biologically meaningful stratified phenotypes.

Introduction

Osteoarthritis (OA) is a serious destructive joint disorder and the third most rapidly rising condition associated with disability[1]. Despite this, no effective treatments that target OA are available. Current treatments only manage pain, not the underlying mechanisms of disease etiology. An estimated 5% of the world population is affected by OA, with hand OA as one of its most prevalent forms. Given the high prevalence with age and high estimated lifetime risk rates for symptomatic hand OA (39.8%), the number of individuals affected will only continue to increase[2, 3]. Hand OA has a high clinical burden, involving considerable pain, deformity, and impaired function[4-6]. In addition, hand joints are non-weight bearing, and therefore may reflect the systemic aspects of the disease more than knee and/or hip OA, where mechanical loading is a dominant risk factor[7]. A better understanding of hand OA, its causes and pathophysiological mechanisms is therefore urgently needed.

Hand OA is a complex multifactorial disorder. It shares risk factors, such as repetitive movements and obesity, with OA at other joint sites[8, 9]. Hand OA also has a strong genetic component, with heritability estimates ranging from 39%–84% depending on the hand joint affected[10, 11]. Recently, major advances were made in elucidating the genetic background of hip and knee OA, using large (n>400,000 individuals) genome-wide association studies (GWAS)[12-14]. Yet, for hand OA only two common genetic loci have been found near the MGP and ALDH1A2 genes[15, 16]. The lack of findings for hand OA may be attributed to the modest samples sizes (n<9,000 individuals) in previous GWAS and disease heterogeneity[17], which is known to reduce power to robustly detect genetic loci[18].

Osteoarthritis in the hand may be present in any joint, but is most prevalent in the distal interphalangeal joints (DIP), first carpometacarpal (CMC1) and trapezioscaphoid (TS) joints, followed by the proximal interphalangeal joints (PIP). Osteoarthritis is least prevalent in metacarpophalangeal joints (MCP)(19, 20). Diverse definitions of hand OA have been used including nodal hand OA (interphalangeal (IP) joints), thumb base OA (CMC1/TS), and generalized hand OA (DIP/PIP/CMC1)[19, 20] with varying results, suggesting that disease etiology may differ between the joints. Moreover, OA affects multiple tissues within the joint including cartilage and bone. This further contributes to disease heterogeneity and warrants the assessment of hand OA by endophenotypes such as joint space narrowing (JSN) and osteophytes (OST) that may capture separate biological processes underlying OA pathology. The use of endophenotypes, quantifiable biological phenotypes intermediate to the genes and the disease, has been successfully used in OA for the detection of novel genetic loci[16, 21, 22] in knee and hip OA, and

may provide new insights into hand OA. For heterogeneous diseases such as OA, stratification of OA phenotypes into different dimensions of disease is one way of increasing power in GWAS[18]. There are few GWAS of hand OA and none have examined hierarchically defined clusters of OA joint presentation in the hands or hand OA endophenotypes that may provide new insights into hand OA pathogenesis. We therefore set out to examine the occurrence of OA across hand joints and conduct a GWAS stratified by hand OA patterns [23] to identify novel genetic loci for hand OA.

Methods

GWAS, discovery, replication and meta-analysis

We conducted genome-wide association studies (GWASs) on radiographic structural phenotypes for OA of the hand using data from the Rotterdam Study (RS) (n~8,700) [24]. For a detailed description of the RS, sub-cohorts and GWAS methods, see **online Supplementary Text** and **Table 1**. Briefly, genotypes were imputed to the Haplotype Reference Consortium reference panel (v1.0) using the Michigan Imputation Server[25]. We assessed genetic associations in each RS sub-cohort using linear regression models adjusted for age, sex and the first four genetic principal components. RVtests[26] was used for the GWAS analyses and results were quality controlled using EasyQC[27]. Variants with an imputation quality<0.3, minor allele frequency<0.05 or effective allele count<5 were excluded and genomic control correction was applied to all standard error and p-values. Meta-analysis between the discovery cohorts was performed using fixed-effects inverse variance weighting with METAL[28]. Manhattan plots and QQ-plots were generated using R and R package qqman[29]. Independent variants with a p-value<$1x10^{-06}$ and a chi-squared statistic test of heterogeneity p-value>$1x10^{-06}$ were selected for replication in the Framingham Heart Study (FHS) (n=1,203)[30]. For a detailed description of the FHS, see the **online Supplementary Text**. Summary level data from the discovery and replication stage were combined in a joint meta-analysis (METAL)[28]. Variants met criteria for replication if the association reached a p-value<0.05, had the same direction of effect as the discovery sample, and reached a joint meta-analysis p-value<$5x10^{-08}$. Replicated variants were also examined for association with clinical OA (i.e., hospital diagnosed OA) based on GWAS summary statistics from a large-scale OA meta-analysis of data from the UK Biobank and Icelandic deCODE populations[14, 15]. For a detailed description of the UK Biobank and deCODE, see the **online Supplementary Text.** Associations that reached a p-value<0.01 were considered statistically significant.

Table 1: General Characteristics of the study population

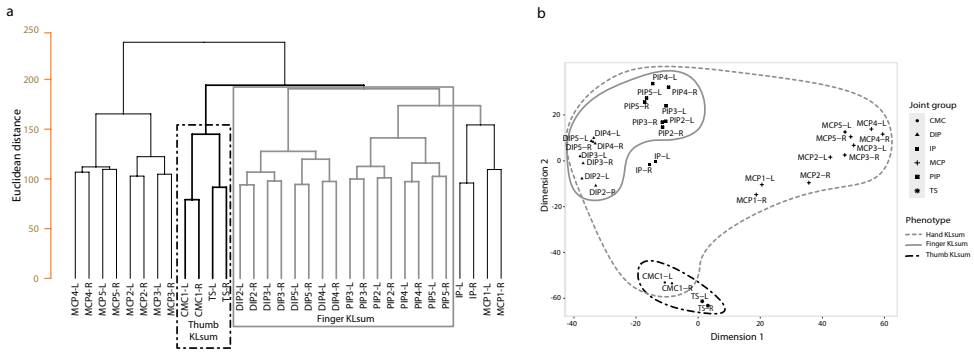
Cohort	RS-I	RS-II	RS-III
Hand			
n (OA cases)	4,829 (1,830)	1,791 (688)	2,071 (526)
Female (%)	2773 (0.57)	964 (0.54)	1180 (0.57)
Age (sd)	67.6(7.9)	64.6(7.9)	57.1(7.0)
Klsum (aver.(sd))	8.4 (9.9)	6.9 (4.0)	4.7 (6.5)
Osteophytes (aver.(sd))	7.1 (7.8)	6.8 (8.2)	4.6 (6.3)
JSN (aver.(sd))	0.84 (2.4)	0.34 (1.3)	0.2 (0.8)
Finger			
n (OA cases)	4,839 (1,244)	1,803 (474)	2,072 (298)
Female (%)	2779 (0.57)	972 (0.54)	1181 (0.57)
Age (sd)	67.6(7.9)	64.6(7.9)	57.1(7.0)
Klsum (aver.(sd))	5.8 (7.4)	4.5 (6.1)	3.0 (4.7)
Osteophytes (aver.(sd))	4.7 (5.5)	4.3 (5.8)	2.9 (4.5)
JSN (aver.(sd))	0.6 (1.9)	0.25 (1.1)	0.1 (0.7)
Thumb			
n (OA cases)	4,882 (916)	1,813 (255)	2,083 (166)
Female (%)	2785 (0.57)	972 (0.54)	1184 (0.57)
Age (sd)	67.6(7.9)	64.6(7.9)	57.1(7.0)
Klsum (aver.(sd))	2.1 (3.3)	1.2 (2.3)	0.8 (1.7)
Osteophytes (aver.(sd))	1.3 (2.1)	0.93 (1.7)	0.6 (1.3)
JSN (aver.(sd))	0.4 (1.0)	0.18 (0.63)	0.1 (0.4)

OA: osteoarthritis, KL: Kellgren-Lawerence, JSN: joint space narrowing, aver.: average, sd: standard deviation, RS: Rotterdam Study

Detailed phenotype descriptions

For each participant, all hand joints (16 joints per hand, 32 joints per individual) were scored for Kellgren and Lawrence (KL) grade[31] based on hand radiographs. KL grade is a semi-quantitative score ranging from 0-4, where higher scores indicate more severe disease. Radiographic OA was defined as KL-grade ≥2 (definite joint space narrowing and definite osteophytes). Each joint was also scored for individual radiographic features including joint space narrowing (JSN) and osteophytosis (OST)[31]. The JSN and OST scores are semi-quantitative scores ranging from 0-3, where 0=none, 1=possible, 2=definite and 3=marked.

We conducted hierarchical clustering of KL grade across all hand joints to identify patterns of disease occurrence defined by location and disease severity (see online supplementary text). This yielded three semi-quantitative hand OA phenotypes for analysis: 1) hand KLsum=sum of KL grades across all DIP, PIP, MCP, IP, and CMC1 joints in both hands (15 joints per hand, 30 joints per individual, hand KLsum score range: 0 - 120), 2) finger KLsum=sum of KL grades across all DIP and PIP joints in both hands



▲Figure 1: Tree-dendrogram and multidimensional scaling (MDS) plot of KL scores in the joints of the hand. a. tree-dendogram of complete hierarchal clustering of Euclidean distance matrix of KL scores of groups of joints. Joints were grouped per 'type' and per hand, left or right. **b.** MDS plot of KL scores for all hand joints left (L) and right (R), not grouped per "type". Data consisted of all available data from RSI, RSII and RSIII (n=8,691). Selected clusters are depicted by the different dashed lines. OA: osteoarthritis, KL: Kellgren-Lawrence OA severity score, TS: trapezioscaphoid joints, CMC: carpometacarpal joints, PIP: proximal interphalangeal joints, DIP: distal interphalangeal joints, MCP: metacarpophalangeal joints and IP: interphalangeal joints. L/R: left or right joint, number denotes which joint, i.e., PIP2L: the second PIP joint at the left hand. See supplementary figures 2-3 for dendrogram and MDS plots for joint space narrowing and osteophytosis.

(8 joints per hand, 16 joints per individual, KLsum score range: 0 - 64), and 3) Thumb KLsum=sum of KL grades across the CMC1 and TS joint (2 joints per hand, 4 joints per individual, KLsum score range: 0 - 16). Individuals with a missing KL grade in one or more hand joints were excluded from the analysis of phenotypes that required scoring of the missing joint(s). Also, individuals with missing age, sex or genetic principal components were excluded from all analyses (**Table 1**).

Post GWAS analysis

Post GWAS analysis consisted of multiple bioinformatic and functional analyses (online **Supplementary Text**). Briefly, all GWAS variants were annotated using HaploReg (V4.1) and FUMA[32, 33]. Intersection of variants with epigenetic markers, proteins, tran-scription factor (TF) motifs and binding, and chromatin interactions was done using data from ROADMAP, ENCODE, HaploReg (V4.1) and the 3D genome browser[34-37]. Functional studies included expression quantitative trait loci (eQTL) analysis, methyl-ation expression quantitative trait loci (meQTL) analysis, ATAC-seq analysis and dif-ferential gene expression analysis. All functional analyses were performed in human articular cartilage. Details are provided in the online **Supplementary Text**.

Results

Patterns of osteoarthritis severity in joints of the hand

We used hierarchal cluster analysis on the KL-grades of all 32 hand joints of the left hand and right hand to identify clusters (**Figure 1**). Tight symmetric clustering between the left and right joints was seen, in addition to clustering based on joint group, which is in line with previous findings [19, 38, 39] (**Figure 1a – b**). Clustering based on individual radiographic features, joint space narrowing (JSN) and number of osteophytes (OST), produced similar symmetric and joint group clusters (online **Supplementary Figures S1 and S2**). We observed consistent clustering of the PIP with the DIP joints and the TS with the CMC joints (**Figure 1,** online **Supplementary Figure S1 and S2**). Based on these analyses, we created three semi-quantitative hand OA phenotypes: hand KLsum, thumb KLsum and finger KLsum (**Table 1**).

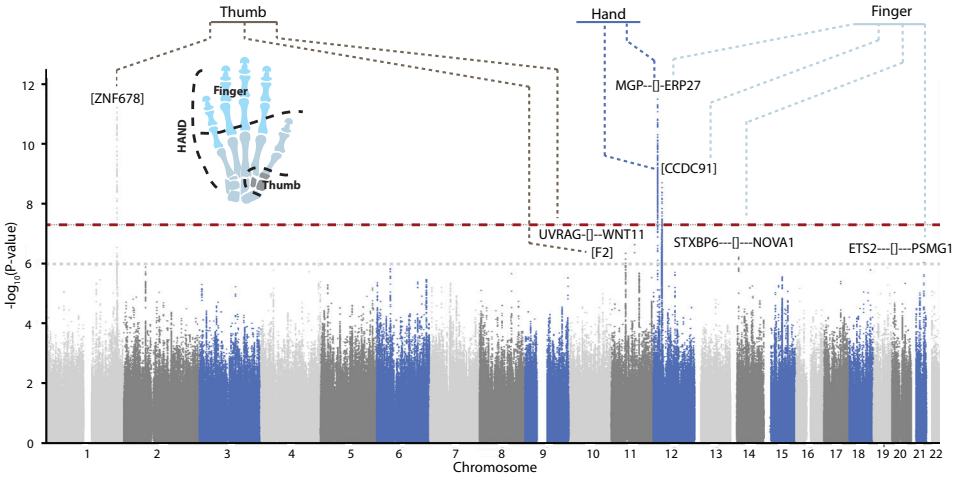
Identification of genetic hand osteoarthritis loci

We conducted genome-wide association studies (GWAS) on each of the three identified hand OA phenotypes in a discovery sample that included RS-I, RS-II and RS-III cohorts (n~ 8,700) (**Table S1**). In total, we identified seven independent signals with genome-wide suggestive association (p-value<$1x10^{-06}$) which were taken forward for replication in the Framingham Heart Study (FHS, n=1,203) (**Figure 2, Table 2**). In total, four independent signals were genome-wide significant in the meta-analysis (p-value≤$5x10^{-08}$), of which three were significantly replicated (p-value<0.05) (**Table 2**). Two of these signals were novel osteoarthritis associated loci. The first and most significant novel locus was located on chromosome 1 near the *ZNF678*, *WNT3A* and *WNT9A* genes, with rs10916199 as the lead single nucleotide variant (SNV). This signal is replicated (p-value<0.05) and genome-wide significantly associated with thumb KLsum (beta=-0.31, p-value=$2.36x10^{-13}$). The second replicated novel locus is located on chromosome 11 containing the *F2, LRP4* and *CREB3L1* genes, and is associated with thumb KLsum (beta=-0.19, p-value=$4.7x10^{-08}$). We also identified two known OA associated loci. The first locus was located near the *MGP* gene, with rs4767133 as the lead SNV. This locus was previously found to be associated with hand KLsum [16]. The second known OA locus, also located on chromosome 12, is the *CCDC91*-locus, which was also previously found to be associated with hand KLsum, though it did not reach genome-wide significance[16]. Here, the lead variant rs12049916, was genome-wide significantly associated with hand KLsum (beta=0.78, p-value=$1.5x10^{-08}$) and finger KLsum (beta=0.58, p-value=$2.0x10^{-08}$), but did not reach nominal significance in the replication cohort, though the direction of effect was the same between discovery and replication cohorts (**Table 2**).

Table 2: Summary of radiographic hand OA structural phenotypes GWAS results

				*Discovery**				*Replication†*					*Meta-Analysis‡*	
rsID	*Chr*	*EA*	*NEA*	*EAF*	*Beta*	*SE*	*P-value*	*Beta*	*SE*	*P-value*	*Beta*	*SE*	*P-value*	*Locus§*
Thumb KLsum														
rs10916199	1	A	G	0.81	-0.31	0.04	$2.1x10^{-12}$	-0.27	0.13	$3.8x10^{-02}$	-0.31	0.04	$2.4x10^{-13}$	[ZNF678]
rs2070852	11	C	G	0.69	-0.19	0.04	$4.5x10^{-07}$	-0.22	0.11	$3.4x10^{-02}$	-0.19	0.04	$4.7x10^{-08}$	[F2]
rs621457	11	A	G	0.52	-0.18	0.03	$2.4x10^{-07}$	-0.07	0.11	$5.2x10^{-01}$	-0.17	0.03	$3.3x10^{-07}$	UVRAG-[]--WNT11
Finger KLsum														
rs4764133	12	T	C	0.38	0.61	0.09	$5.7x10^{-12}$	1.38	0.38	$2.7x10^{-04}$	0.65	0.09	4.8x10-14	MGP—[]-ERP27
rs12049916	12	G	A	0.22	0.57	0.11	$7.5x10^{-08}$	0.73	0.44	$9.9x10^{-02}$	0.58	0.10	$2.0x10^{-08}$	[CCDC91]
rs1950427	14	T	C	0.12	0.66	0.13	$6.2x10^{-07}$	0.63	0.53	$2.4x10^{-01}$	0.66	0.13	$3.0x10^{-07}$	STXBP6---[]---NOVA1
rs1029003	21	A	G	0.46	0.43	0.09	$9.3x10^{-07}$	0.35	0.37	$3.4x10^{-01}$	0.42	0.09	$5.9x10^{-07}$	ETS2---[]---PSMG1
Hand Klsum														
rs4764133	12	T	C	0.38	0.75	0.12	$3.0x10^{-10}$	1.76	0.49	$2.9x10^{-04}$	0.81	0.12	$2.9x10^{-12}$	MGP—[]-ERP27
rs12049916	12	A	G	0.77	0.77	0.14	$5.3x10^{-08}$	0.90	0.56	0.11	0.78	0.14	$1.5x10^{-08}$	[CCDC91]

EA: Effect Allele, NEA: Non Effect Allele, EAF: Effec
t Allele Frequency, SE: Standard Error
* Discovery consists of Rotterdam Study RS-I, RSII, RSIII, samples sizes per phenotype are; Thumb KLsum n=8,778; Finger KLsum n=8,714; Hand KLsum n=8629.
† Replication cohorts consist of Framingham Heart Study n=1203
‡ Meta-analysis is a meta-analysis of discovery and replication cohorts using inverse variance weighted meta-analysis using METAL.
§SNP location represented by [], if the SNP is localized intergenic the dashes denotes the distance,-≤10 kb,--≤100kb,---≤1Mkbkb, ----->1Mkb.



▲Figure 2: Combined Manhattan-plot of GWAS discovery results of all radiographic hand OA structural phenotypes. thumb= thumb KLsum score, finger= finger KLsum score and hand: handklsum score. GWAS discovery consists of RSI, RSII and RSIII, and was adjusted for age, sex, and the first four genetic principle components. The –log10 p-values, for each of the ~11 million SNPs analyzed (remaining after EASYQC quality control) from the association studies is plotted against their position per chromosome. The dotted red horizontal line corresponds to the genome-wide significant threshold (p-vslur=5*10–08). The solid grey line corresponds to the selection for replication threshold (p-value=5*10–06). Lead SNP location is represented by [] (intronic), if the SNP is localized intergenic the dashes denotes the distance, -≤10 kb, --≤100 kb, ----≤1 Mkb, ----≥1 Mkb. GWAS, genome-wide association studies. For plots of the individual GWAS see Supplementary Figures 3-5.

To examine if replicated loci were also associated with clinically defined OA, we looked up findings in GWAS summary statistics from a recent large-scale OA meta-analysis that included the UK Biobank and deCODE populations[14, 15](**Table 3**). Of the novel signals, only rs10916199 (thumb KLsum) was significantly associated with its matching clinical OA phenotype (thumb OA: OR=0.9, 95%CI=0.86-0.95, p-value=5.7x10^{-05}). Of the known signals, rs4764133 was significantly associated with multiple clinical OA phenotypes: thumb OA (OR=1.07, 95%CI= 1.03-1.12, p-value=6.3x10^{-04}), finger OA (OR=1.12, 95%CI=1.07-1.17, p-value=5.7-10^{-07}), hand OA (OR=1.09, 95%CI=1.04-1.13, p-value=6.7x10^{-05}), and nominal significantly with knee OA (OR=0.99, 95%CI=0.96-1.00, p-value 1.8x10^{-02})(**Table 3**).

WNT9A as potential causal gene for thumb OA

We examined the rs10916199 locus in more detail given the strong and consistent association with thumb OA. Different levels of information were leveraged for all genes within 1Mb surrounding rs10916199 in order to prioritize a putative causal gene (**Figure 3, online supplementary text**). First, we meta-analyzed two human osteoarthritic cartilage expression quantitative trait loci (eQTL) datasets (n=116, hip/knee joints, **online supplementary Table S1**)[40, 41] and found no significant effect of rs10916199

Table 3: Association with xosteoarthritis phenotypes

			Thumb OA (7,280 cases; 605,132 controls)			Finger OA (7,037 cases; 222,772 controls)		
rsID	**EA**	**EAF**	**OR**	**95%CI**	**P-value**	**OR**	**95%CI**	**P-value**
rs10916199	A	0.81	0.91	0.86-0.95	$5.7x10^{-05}$	0.98	0.93-1.04	0.57
rs2070852	C	0.69	0.97	0.89-1.05	0.47	0.99	0.94-103	0.58
rs4764133	T	0.39	1.07	1.03-1.12	$6.3x10^{-04}$	1.12	1.07-1.17	$5.7x10^{-07}$
rs12049916	A	0.77	1.00	0.96-1.05	0.87	0.99	0.94-1.04	0.67
			Hand OA *(8,591 cases; 224,326 controls)*			**Hip OA** *(17,151 cases, 613,790)*		
rsID	EA	EAF	OR	95%CI	P-value	OR	95%CI	P-value
rs10916199	A	0.81	1.03	0.98-1.08	0.32	1.00	0.98-1.03	0.95
rs2070852	C	0.69	0.99	0.96-103	0.80	0.98	0.96-1.01	0.22
rs4764133	T	0.39	1.09	1.04-1.13	$6.7x10^{-04}$	0.99	0.96-1.01	0.30
rs12049916	A	0.77	0.99	0.94-1.04	0.68	0.96	0.93-0.99	$2.9x10^{-03}$
			KneeOA *(24,919 cases, 613,702 controls)*					
rsID	**EA**	**EAF**	**OR**	**95%CI**	**P-value**			
rs10916199	A	0.81	0.99	0.97-1.02	0.59			
rs2070852	C	0.69	0.99	0.97-1.01	0.40			
rs4764133	T	0.39	0.98	0.96-1.00	0.018			
rs12049916	A	0.77	0.99	0.97-1.01	0.27			

Thumb OA, Finger OA, HandOA GWAS summary statistics are from the DECODE cohort, Hip OA and Knee OA summary statistics are from a meta-analysis of DECODE and UKbiobank cohorts. OA: osteoarthritis, EA: Effect Allele, NEA: Non Effect Allele, EAF: Effect Allele Frequency, SE: Standard Error, OR: odds ratio, CI: Confidence Interval

on gene expression in these datasets. Next, there were significant methylation quantitative trait loci (meQTL) associated with rs10916109 and two CpG sites in human osteoar-thritic cartilage (knee/hip joints): CpG09796739 (beta=0.39, FDR p-value=8.3x10-07) and CpG11520395 (beta=0.21, FDR p-value=$5.5x10^{-03}$) (**Figure 3e, zone 2**). These methylation sites are located in a region flanking an active transcriptional start site in primary osteoblasts and chondrogenic cells (**Figure 3b**).

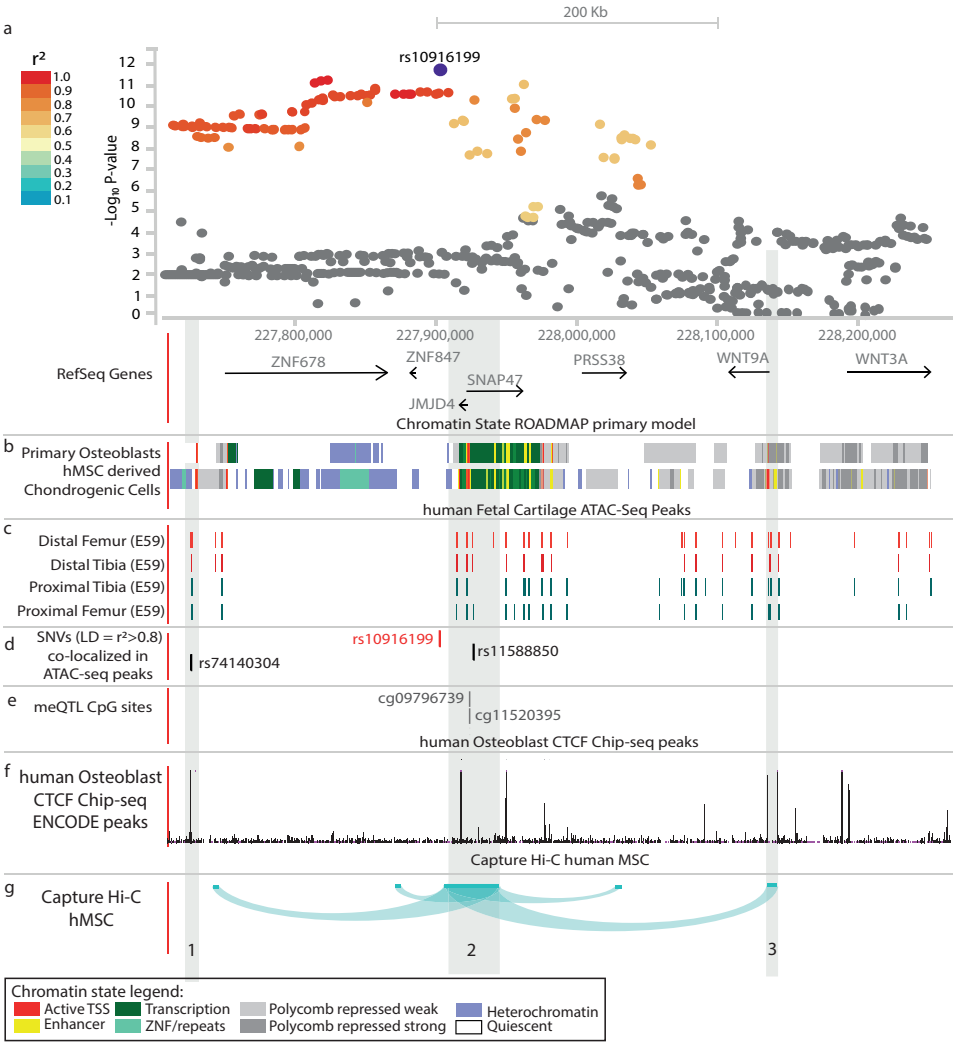
To further assess co-localization of the identified genetic loci with regulatory function during cartilage differentiation, we intersected the GWAS signals in the lo-cus with accessible chromatin regions (ATAC-seq peaks) in rare human fetal cartilage acquired from proximal and distal long bones from gestational day(E) 59 of development[42]. Two of the SNVs in high LD (r2≥0.8) with rs10916199 (rs74140304 and rs11588850) intersected with accessible chromatin regions across multiple human long bone cartilage (**Figure 3c**). In addition, rs74140304 also intersected with an active transcription start site in osteoblast and chondrogenic cells (**Figure 3b**)[28]. Next, we examined 3D chromatin conformation in the locus.

Since no chromatin conformation capture data were available for chondrogenic or bone cells, we examined data from hu-man mesenchymal stem cells (hMSC), which are stem cell progenitors for chondrocytes and osteoblasts (**Figure 3g**)[28]. The genomic location of rs1158850 (**Figure 3g, zone 2**) appears to come into close proximity with the promoter region of WNT9A (**Figure 3g, zone 3**). In addition, CTCF binding peaks in osteoblasts also intersect with the WN-T9A promoter region (**Figure 3f, zone 3**), and are located near rs11588850 (**Figure 3f, zone 2**).

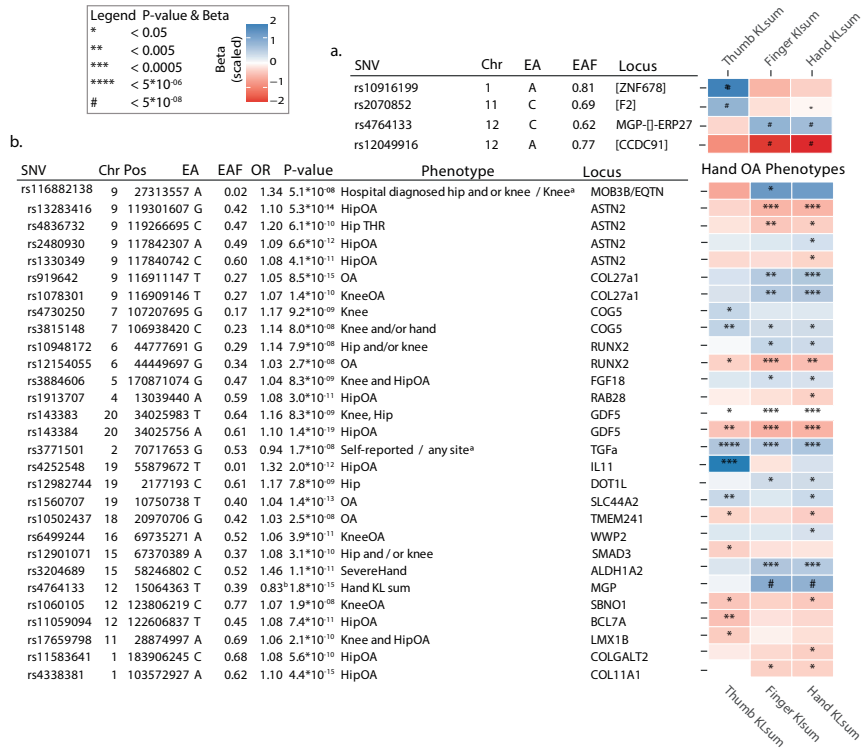
Next, differential expression analysis between OA lesioned and preserved cartilage (hip/knee joints) identified *WNT9A* as the most significant result, with increased expression in OA lesioned cartilage (n=21, fold change=2.42, p-value=$9.4x10^{-08}$, online **Supplementary Table S2**). Lastly, we examined whether rs11588850 may significantly affect (p-value<$4.0x10^{-08}$) the regulatory transcription factor (TF) binding motifs (37) located within the *WNT9A* promoter[35]. The minor allele (G) of rs11588850 is in high linkage disequilibrium (r^2>0.8) with the OA risk increasing allele (G) of rs10916199, which significantly increases the binding affinity of the TF binding motif for RAD21 (G-allele LOD=11.2, A-allele LOD=9.8). The RAD21 protein has been previously shown to bind to the *WNT9A* promoter region (online **Supplementary Table S3).** Thus, our results indicate *WNT9A* as novel OA associated gene, where rs1158850 is a potential regulatory variant for *WNT9A* (**Figure 3, zone 2-3**).

Additional hand osteoarthritis associated loci

Osteoarthritis is highly heritable and co-occurrence of OA in multiple joint sites is well recognized[43]. As the hand joints are non-weight bearing, causes of OA in these joints may reflect effects of systemic risk factors, unlike the hip and knee joint where mechanical loading is a dominant risk factor[7]. Thus, we examined whether other known OA loci may also confer risk for hand OA (**Figure 4**). For 29 of the previously reported OA associated SNVs[13, 14], nominally significant associations (p-value<0.05) were observed for one or more hand OA phenotypes. Strong associations were seen for known-hand OA loci: *ALDH1A2*-locus (rs3204689), *MGP*-locus (rs4767133)(**Figure 4a** and **4b**) and *COG5* (rs3815148)[44], an SNV identified from a candidate gene study for hand OA. Interestingly, the *MGP*- and *ALDH1A2*-loci, were only associated with finger and/or hand OA phenotypes, but not with thumb OA. In contrast, the *BCL7A*-locus (rs11059094), was only associated with thumb OA. Strikingly, several reported knee and hip OA loci were also associated with hand OA phenotypes in our study (Bonferroni, p-value<5.8x10-04): *RUNX2*-locus (rs12154055), *COL27A1*-locus (rs919642), *ASTN2*-locus (rs13253416), *IL11*-locus (rs4252548), *TGFa*-locus (rs3771501) and *GDF5*-locus (rs143384)(**Figure 4b**). Since some of these known loci



▲Figure 3: Schematic overview of part of the rs10916199 locus. a. LocusZoom plot of rs10916199 locus, the Y-axis depicts the-Log10 p-value of the SNV from the thumb KLsum GWAS. Colours depict the LD (r2) between the variant and the LD SNV rs10916199. The X-axis depicts the relative genomic location, depict are the protein coding genes at those genomic locations. For this genomic region depicted in figure **b – g** are several epigenetic annotations are plotted. **b**. Chromatin state, as predicted by the ROADMAP 15-state model on histone modifications, for primary osteoblasts and human mesenchymal stem cell derived chondrocytes. Colours depict chromatin states, legend at bottom of full figure. **c.** ATAC-seq peaks from human embryonic cartilage at different bone development sites at gestation day (E)59. **d.** Location of the lead SNV, rs10916199 and two putative causal SNV's which co-locative with ATAC-seq peaks. **e.** Genomic location of rs10916199 meQTL CpG sites. **f.** Chip-seq CTCF protein binding peaks from human primary osteoblasts from ENCODE. **g.** Capture Hi-C chromatin interactions from 3D genome browser for human mesenchymal Stem Cells (hMSC) and mesoderm. Depicted are the chromatin interactions from the promoters of the queried gene to the genomic location of interaction, this was done for the JMJD4/SNAP47 transcription start site (TSS), WNT9A TSS and the WNT3A TSS. For details on methods and underlying data see supplementary methods.



▲Figure 4: Heatmap depicting the effect of OA associated SNVs in in each hand OA phenotype. a. The found associated lead SNVs of the hand OA phenotypes and their beta and p-value in the other phenotypes. **b.** Depicted are all known OA SNVs which had a nominal significant effect in at least one stratified hand OA phenotype. All beta's were calculated for the reported effect allele and scaled. Colours represent the scaled beta of the effect allele, which is here the minor allele. P-values are represented by * in the box. Gene: reported gene from the GWAs study. Chr, chromosome; Pos: base pair position on the chromosome Hg19; EA, effect allele; EAF, effect allele frequency; OR, odds ratio; KL, kellgren-Lawrence score.

were previously found to be associated with knee and/or hip OA, they may reflect common mechanisms across all joints in OA.

Discussion

We identified four genome-wide significant loci associated with hand OA phenotypes, of which two were novel and specific for thumb OA. Integration of multiple lines of data provided cumulative evidence that *WNT9A* may be a causal gene for thumb OA. We first conducted a cluster analysis of hand joints to identify less heterogeneous clusters of joints that served as the basis of the hand OA phenotypes assessed in this GWAS. With this approach of using radiographically defined biologically relevant OA phenotypes to reduce phenotype heterogeneity and increase statistical power, we were able to robust-ly identify known and novel OA loci despite our modest sample

size (n~9,900)[18, 22]. This indicates that assessment of stratified phenotypes in OA may be warranted to im-prove GWAS statistical power and provide novel insight into OA biology.

Using bioinformatic analysis and functional genomics datasets, we were able to identify rs1158850 as potential causal variant. This SNV is nearby meQTL CpGs and the G allele of this variant is predicted to increase RAD21 (RAD21 Cohesin Complex Compo-nent) binding affinity in a region that has chromatin interactions with the WNT9A pro-moter. RAD21 is a part of the cohesion complex, involved in the formation of chromatin loops with CTCF[45]. Both RAD21 and CTCF bind to the *WNT9A* promoter region. In line with these findings, *WNT9A* expression was significantly increased in OA lesioned cartilage compared to preserved OA cartilage. Combining all results, we postulate that rs1158850 is located in a gene regulatory element, increases RAD21 binding, and me-diated by CTCF, interacts with the WNT9A promoter to influence *WNT9A* expression.

WNT9A (Wingless-type MMTV integration site family, member 9A) previously known as *WNT14*, is a member of the WNT gene family, and has been shown to play a central role in synovial joint formation[46, 47]. Knockout *WNT9A* mice have severe skel-etal developmental defects, and are neonatal lethal[48]. Expression of WNT members by chondrocytes leads to the destruction of the cartilage matrix by the upregulation of Wnt/β-catenin signaling. Inhibition of WNT members has been suggested as a plausible OA therapeutic strategy, with recent success in a murine model of OA[49, 50]. However, this is the first evidence for WNT9A, a non-canonical Wnt ligand, in human OA.

In addition, to identifying novel OA loci, we also provide evidence for a sub-set of generalized OA genetic risk loci: *GDF5 (rs143384), TGFα (rs2862851/rs3771501), RUNX2 (rs12154055), ASTN2 (rs2480930, rs13823416), COL27A1 (rs919642), and IL11 (rs4252548)*. These loci should be given priority as potential therapeutic targets since genetically supported drug targets have been shown to double the success rate of ther-apeutics in clinical development and intervention at these target loci may be beneficial regardless of which joint site is affected by OA [51].

Although collectively our findings implicate *WNT9A* in thumb OA, there are sever-al limitations. First, age is the most predominant risk factor for OA, yet the genetic back-ground may determine the age of onset, rather than the lifetime risk for OA. Therefore, future genetic studies may benefit from examining the age of onset rather than adjust for age[52]. Second, our functional findings are based on chondrogenic data sourced from several different tissues that did not include tissues from the hand joints. How-ever, consistent results were found across the

available chondrogenic source material, indicating a more general role for the WNT9A locus in chondrocyte functional pathways. Given the complex nature of OA susceptibility and the fact that pathophysiologic causes are not uniform across skeletal sites, alterations in *WNT9A* expression may be seen in other joints, but may have a more marked detrimental effect on the thumb joints.

In summary, by examining the distribution of radiographic OA features in the hand joints, we identified three distinct hand OA phenotypes that provided the basis for identification of a novel locus for thumb OA despite our modest sample size. We iden-tified *WNT9A* as a plausible causal gene for thumb OA, providing new insights into the genetic architecture of hand OA and a new candidate for OA therapeutic development.

Supplementary matarial

All supplementary materials: Methods, Figures and Tables, Can be viewed and down-loaded online at this location: https://drive.google.com/drive/folders/1iqiLVJm1MbeNhefsbLNzQhLNmvHhj5OB?usp=sharing

Data Availability Statement

All relevant data supporting the key findings of this study are available within the article and its supplementary information files. Other data are available from the corresponding author upon reasonable request. Due to ethical and legal restrictions, indi-vidual-level data of the Rotterdam Study (RS) cannot be made publicly available. Data are available upon request to the data manager of the Rotterdam Study Frank van Rooij (f.vanrooij@erasmusmc.nl) and subject to local rules and regulations. This includes submitting a proposal to the management team of RS, where upon approval, analysis needs to be done on a local server with protected access, complying with GDPR regula-tions.

Acknowledgments

The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. The generation and man-agement of GWAS genotype data for the Rotterdam Study (RS I, RS II, RS III) was exe-cuted by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters, MSc, and Carolina Medina-Gomez, MSc, for their help in creating the GWAS database, and Linda Broer PhD, for the creation of the imputed data.

Funding

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 Rotterdam Study GWAS datasets are supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project nr. 050-060-810. The Framingham Heart Study of the National Heart, Lung, and Blood Institute of the National Institutes of Health and Boston University School of Medicine was supported by the National Institutes of Health (contract no. HHSN268201500001I, N01-HC-25195, AG18393, AR47785) and its contract with Affymetrix, Inc. for genotyping services (N02-HL-6-4278). Analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. MSY was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) and the National Institute on Aging (NIA) (R01AR075356). Rice, Cheung and Loughlin were supported by Versus Arthritis (grant 20771), by the Medical Research Council and Arthritis Research UK as part of the MRC-Arthritis Research UK Centre for Integrated Research into Musculoskeletal Ageing (CIMA, grant references JXR 10641, MR/P020941/1 and MR/R502182/1), and by the European Union's Seventh Framework Program for research, technological development and demonstration under grant agreement number No. 305815 (D-BOARD). The research leading to the RAAK biobank and the current 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 no. 259679. ATAC-seq datasets generated by MY and TDC were funded by a Harvard University Dean's Competitive Award.

Author Contributions

CGB designed the study, performed the analyses, made the figures and tables, and wrote the manuscript. MSY performed the replication analysis. SJR, RCdA, KC and LS performed functional analysis and look-ups(EQTL, meQTL and differential expression), MY and TDC performed ATAC-seq analysis, US provided GWAS data of deCODE and UK-

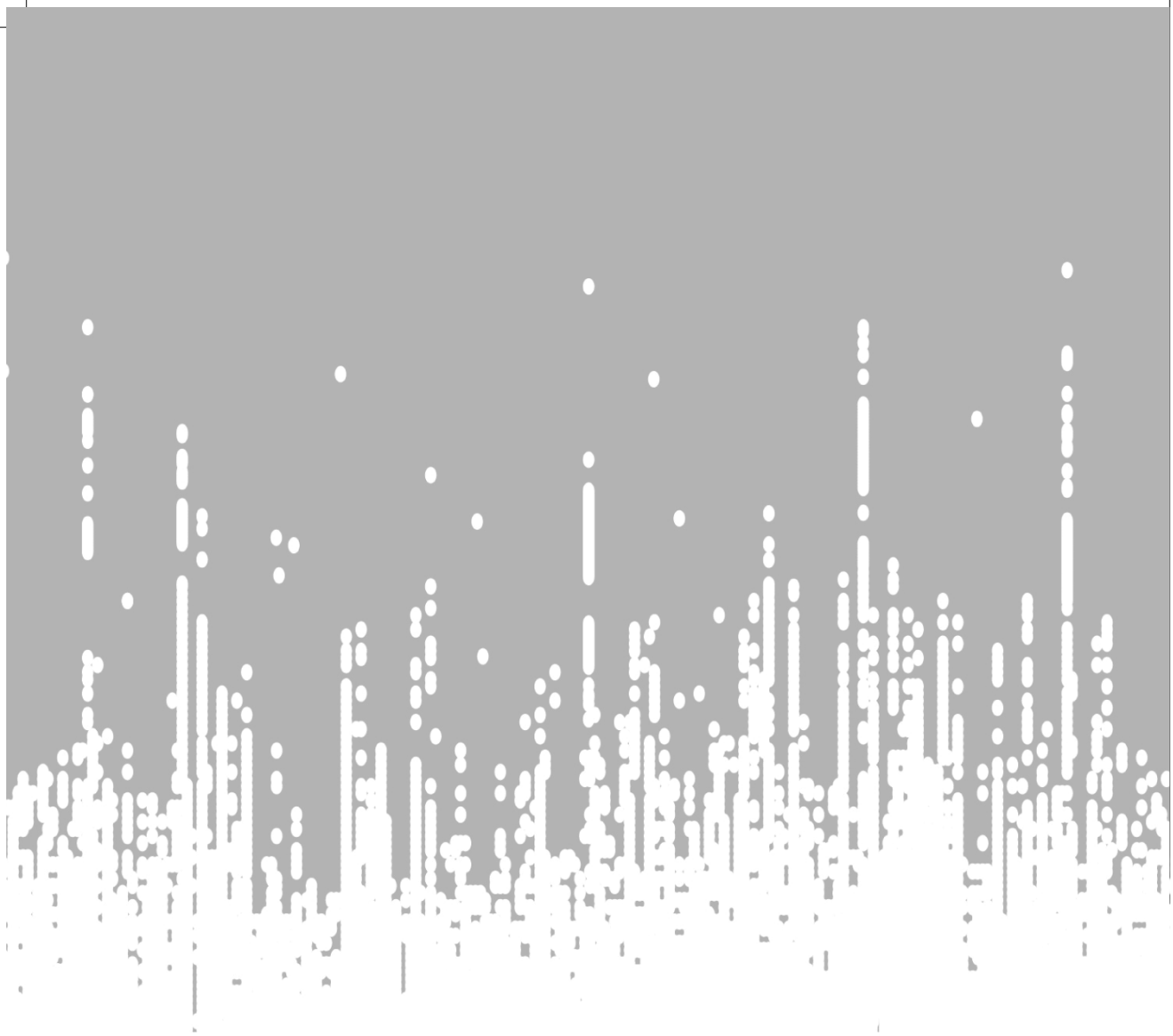
biobank, LB provided analysis help. AGU provided access to the Rotterdam study data set, DF provided replication date, EZ, JL, TDC, and IM provided functional data. J.B.J.v.M. designed the study and supervised this work. All authors critically assessed the manuscript.

References

1. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392(10159):1789-858.
2. Barbour KE, Lui LY, Nevitt MC, Murphy LB, Helmick CG, Theis KA, et al. Hip Osteoarthritis and the Risk of All-Cause and Disease-Specific Mortality in Older Women: A Population-Based Cohort Study. Arthritis Rheumatol. 2015;67(7):1798-805.
3. Qin J, Barbour KE, Murphy LB, Nelson AE, Schwartz TA, Helmick CG, et al. Lifetime Risk of Symptomatic Hand Osteoarthritis: The Johnston County Osteoarthritis Project. Arthritis Rheumatol. 2017;69(6):1204-12.
4. Bijsterbosch J, Watt I, Meulenbelt I, Rosendaal FR, Huizinga TW, Kloppenburg M. Clinical burden of erosive hand osteoarthritis and its relationship to nodes. Ann Rheum Dis. 2010;69(10):1784-8.
5. Kwok WY, Vliet Vlieland TP, Rosendaal FR, Huizinga TW, Kloppenburg M. Limitations in daily activities are the major determinant of reduced health-related quality of life in patients with hand osteoarthritis. Ann Rheum Dis. 2011;70(2):334-6.
6. Kjeken I, Dagfinrud H, Slatkowsky-Christ B, Mowinckel P, Uhlig T, Kvien T, et al. Activity limitations and participation restrictions in women with hand osteoarthritis: patients' descriptions and associations between dimensions of functioning. Ann Rheum Dis. 2005;64(11):1633-8.
7. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. Lancet. 2019;393(10182):1745-59.
8. Jiang L, Xie X, Wang Y, Lu Y, Tian T, Chu M, et al. Body mass index and hand osteoarthritis susceptibility: an updated meta-analysis. Int J Rheum Dis. 2016;19(12):1244-54.
9. Jensen V, Boggild H, Johansen JP. Occupational use of precision grip and forceful gripping, and arthrosis of finger joints: a literature review. Occup Med (Lond). 1999;49(6):383-8.
10. Spector TD, Cicuttini F, Baker J, Loughlin J, Hart D. Genetic influences on osteoarthritis in women: a twin study. Bmj. 1996;312(7036):940-3.
11. Ishimori ML, Altman RD, Cohen MJ, Cui J, Guo X, Rotter JI, et al. Heritability patterns in hand osteoarthritis: the role of osteophytes. Arthritis Res Ther. 2010;12(5):R180.
12. Zengini E, Hatzikotoulas K, Tachmazidou I, Steinberg J, Hartwig FP, Southam L, et al. Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. Nat Genet. 2018;50(4):549-58.
13. Tachmazidou I, Hatzikotoulas K, Southam L, Esparza-Gordillo J, Haberland V, Zheng J, et al. Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. Nat Genet. 2019;51(2):230-6.
14. Styrkarsdottir U, Lund SH, Thorleifsson G, Zink F, Stefansson OA, Sigurdsson JK, et al. Meta-analysis of Icelandic and UK data sets identifies missense variants in SMO, IL11, COL11A1 and 13 more new loci associated with osteoarthritis. Nat Genet. 50. United States2018. p. 1681-7.
15. 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(5):498-502.
16. den Hollander W, Boer CG, Hart DJ, Yau MS, Ramos YFM, Metrustry S, et al. Genome-wide association and functional studies identify a role for matrix Gla protein in osteoarthritis of the hand. Ann Rheum Dis. 2017;76(12):2046-53.
17. Kloppenburg M, Kwok WY. Hand osteoarthritis--a heterogeneous disorder. Nat Rev Rheumatol. 2011;8(1):22-31.
18. Manchia M, Cullis J, Turecki G, Rouleau GA, Uher R, Alda M. The Impact of Phenotypic and Genetic Heterogeneity on Results of Genome Wide Association Studies of Complex Diseases. PLoS One. 82013.
19. Egger P, Cooper C, Hart DJ, Doyle DV, Coggon D, Spector TD. Patterns of joint involvement in osteoarthritis of the hand: the Chingford Study. J Rheumatol. 1995;22(8):1509-13.

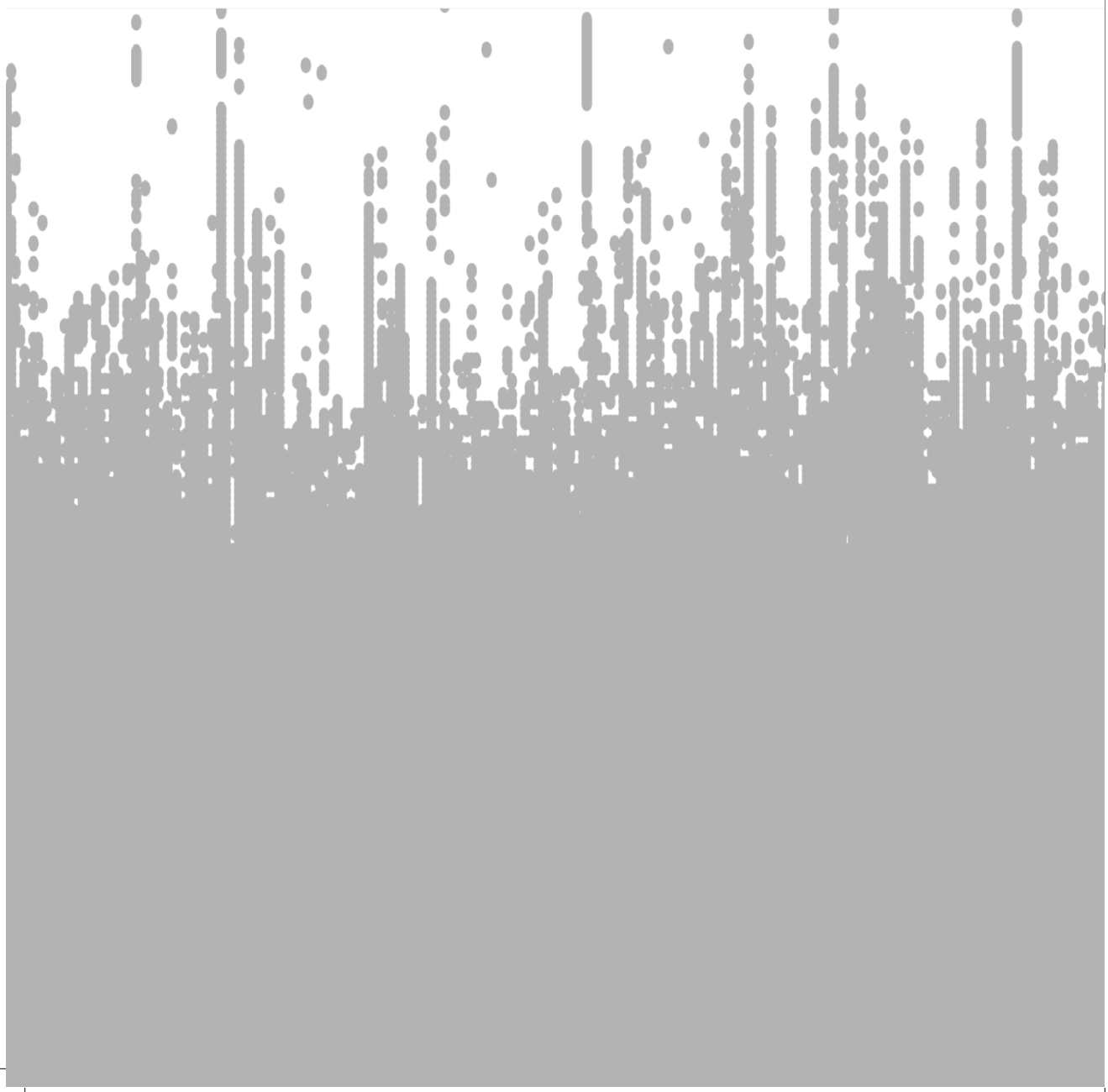
20. Dahaghin S, Bierma-Zeinstra SM, Ginai AZ, Pols HA, Hazes JM, Koes BW. Prevalence and pattern of radiographic hand osteoarthritis and association with pain and disability (the Rotterdam study). Ann Rheum Dis. 2005;64(5):682-7.
21. Castano-Betancourt MC, Evans DS, Ramos YF, Boer CG, Metrustry S, Liu Y, et al. Novel Genetic Variants for Cartilage Thickness and Hip Osteoarthritis. PLoS Genet. 2016;12(10):e1006260.
22. Panoutsopoulou K, Thiagarajah S, Zengini E, Day-Williams AG, Ramos YF, Meessen JM, et al. Radiographic endophenotyping in hip osteoarthritis improves the precision of genetic association analysis. Ann Rheum Dis. 2017;76(7):1199-206.
23. Bierma-Zeinstra SM, van Middelkoop M. Osteoarthritis: In search of phenotypes. Nat Rev Rheumatol. 2017;13(12):705-6.
24. Ikram MA, Brusselle GGO, Murad SD, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2018 update on objectives, design and main results. Eur J Epidemiol. 2017;32(9):807-50.
25. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016;48(10):1284-7.
26. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics. 2016;32(9):1423-6.
27. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Magi R, et al. Quality control and conduct of genome-wide association meta-analyses. Nat Protoc. 2014;9(5):1192-212.
28. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1.
29. Team RC. R: A language and environment for statistical computing.: R Foundation for Statistical Computing, Vienna, Austria.; 2014 [Available from: http://www.R-project.org/.
30. Haugen IK, Englund M, Aliabadi P, Niu J, Clancy M, Kvien TK, et al. Prevalence, incidence and progression of hand osteoarthritis in the general population: the Framingham Osteoarthritis Study. Ann Rheum Dis. 2011;70(9):1581-6.
31. Kellgren JH, Lawrence JS. Radiological Assessment of Osteo-Arthrosis. Ann Rheum Dis. 1957;16(4):494-502.
32. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res. 2016;44(D1):D877-81.
33. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8(1):1826.
34. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317-30.
35. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.
36. Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, et al. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol. 2018;19(1):151.
37. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 402012. p. D930-4.
38. Kalichman L, Cohen Z, Kobyliansky E, Livshits G. Patterns of joint distribution in hand osteoarthritis: contribution of age, sex, and handedness. Am J Hum Biol. 2004;16(2):125-34.
39. Marshall M, van der Windt D, Nicholls E, Myers H, Hay E, Dziedzic K. Radiographic hand osteoarthritis: patterns and associations with hand pain and function in a community-dwelling sample. Osteoarthritis Cartilage. 2009;17(11):1440-7.
40. den Hollander W, Pulyakhina I, Boer C, Bomer N, van der Breggen R, Arindrarto W, et al. Annotating Transcriptional Effects of Genetic Variants in Disease-Relevant Tissue: Transcriptome-Wide Allelic Imbalance in Osteoarthritic Cartilage. Arthritis Rheumatol. 712019. p. 561-70.

41. Steinberg J, Ritchie GRS, Roumeliotis TI, Jayasuriya RL, Clark MJ, Brooks RA, et al. Integrative epigenomics, transcriptomics and proteomics of patient chondrocytes reveal genes and pathways involved in osteoarthritis. Sci Rep. 2017;7(1):8935.
42. Richard D, Liu Z, Cao J, Kiapour AM, Willen J, Yarlagadda S, et al. Evolutionary Selection and Constraint on Human Knee Chondrocyte Regulation Impacts Osteoarthritis Risk. Cell. 2020.
43. Vaes RB, Rivadeneira F, Kerkhof JM, Hofman A, Pols HA, Uitterlinden AG, et al. Genetic variation in the GDF5 region is associated with osteoarthritis, height, hip axis length and fracture risk: the Rotterdam study. Ann Rheum Dis. 2009;68(11):1754-60.
44. Hamalainen S, Solovieva S, Vehmas T, Luoma K, Leino-Arjas P, Hirvonen A. Genetic influences on hand osteoarthritis in Finnish women--a replication study of candidate genes. PLoS One. 2014;9(5):e97417.
45. Phillips JE, Corces VG. CTCF: master weaver of the genome. Cell. 2009;137(7):1194-211.
46. Spater D, Hill TP, O'Sullivan R J, Gruber M, Conner DA, Hartmann C. Wnt9a signaling is required for joint integrity and regulation of Ihh during chondrogenesis. Development. 2006;133(15):3039-49.
47. Hartmann C, Tabin CJ. Wnt-14 plays a pivotal role in inducing synovial joint formation in the developing appendicular skeleton. Cell. 2001;104(3):341-51.
48. Maupin KA, Droscha CJ, Williams BO. A Comprehensive Overview of Skeletal Phenotypes Associated with Alterations in Wnt/β-catenin Signaling in Humans and Mice. Bone Res. 12013. p. 27-71.
49. Zhou Y, Wang T, Hamilton JL, Chen D. Wnt/beta-catenin Signaling in Osteoarthritis and in Other Forms of Arthritis. Curr Rheumatol Rep. 2017;19(9):53.
50. Lietman C, Wu B, Lechner S, Shinar A, Sehgal M, Rossomacha E, et al. Inhibition of Wnt/beta-catenin signaling ameliorates osteoarthritis in a murine model of experimental osteoarthritis. JCI Insight. 2018;3(3).
51. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. Nat Genet. 2015;47(8):856-60.
52. Mars N, Koskela JT, Ripatti P, Kiiskinen TTJ, Havulinna AS, Lindbohm JV, et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. Nat Med. 2020;26(4):549-57.



Chapter 3

HAND OSTEOARTHRITIS GENETICS



Chapter 3.1

Genome-wide association and functional studies identify a role for Matrix-Gla protein in osteoarthritis of the hand

den Hollander, W.,* **Boer, C.G.,*** Hart, D.J., Yau, M.S., Ramos, Y.F.M., Metrustry, S., Broer, L., Deelen, J., Cupples, L.A., Rivadeneira, F., Kloppenburg, M., Peters, M., Spector, T.D., Hofman, A., Slagboom, P.E., Nelissen, R.G.H.H., Uitterlinden, A.G., Felson, D.T., Valdes, A.M., Meulenbelt, I.,† & van Meurs, J.B.J.†

* WH and CGB contributed equally as first authors
† IM and JBJvM contributed equally as senior authors

Published in: Ann Rheum Dis. 2017 Dec;76(12):2046-2053

Abstract

Objective Osteoarthritis (OA) is the most common form of arthritis and the leading cause of disability in the elderly. Of all the joints, genetic predisposition is strongest for OA of the hand; however, only few genetic risk loci for hand OA have been identified. Our aim was to identify novel genes associated with hand OA and examine the underlying mechanism.

Methods We performed a genome-wide association study of a quantitative measure of hand OA in 12,784 individuals (discovery: 8,743, replication: 4,011). Genome-wide significant signals were followed up by analysing gene and allele-specific expression in a RNA sequencing dataset (n=96) of human articular cartilage.

Results We found two significantly associated loci in the discovery set: at chr12 (p-value=$3.5x10^{-10}$) near the matrix Gla protein (MGP) gene and at chr12 (p-value=$6.1x10^{-9}$) near the CCDC91 gene. The DNA variant near the MGP gene was validated in three additional studies, which resulted in a highly significant association between the MGP variant and hand OA (rs4764133, $Beta_{meta}$=0.83, $p\text{-value}_{meta}=1.8x10^{-15}$). This variant is high linkage disequilibrium with a coding variant in *MGP*, a vitamin K-dependent inhibitor of cartilage calcification. Using RNA sequencing data from human primary cartilage tissue (n=96), we observed that the MGP RNA expression of the hand OA risk allele was significantly lower compared with the MGP RNA expression of the reference allele (40.7%, p-value<$5x10^{-16}$).

Conclusions Our results indicate that the association between the MGP variant and increased risk for hand OA is caused by a lower expression of *MGP*, which may increase the burden of hand OA by decreased inhibition of cartilage calcification.

Introduction

Osteoarthritis (OA) is the most frequent joint disorder worldwide. An estimated 22% of the adult population has a joint affected by OA and this incidence increases to 49% in individuals over 65 years of age.[1] All synovial joints can be affected by OA, with hand OA as one of the most common forms of OA. Hand OA is characterised by osteophyte formation, bony enlargements of finger joints and cartilage degradation in the joints. One of the factors contributing to cartilage degradation is the increase of calcified cartilage in the joint.[2,3] In addition, hand OA is related to the occurrence of OA at other sites, most notably with knee OA.[4,5] Patients affected by hand OA suffer from pain and disability, impacting their quality of life. OA is a leading cause of chronic disability,[6] yet currently no effective therapeutic treatments against OA are known. It is therefore imperative to dissect the underlying mechanism of disease aetiology as this may enhance effective and targeted drug development.

OA has a strong genetic component. Depending on the joint affected, the heritability of OA is estimated in the range of 40%–60%,[7,8] with hand OA having the largest heritability, that is, ~60%.[9,10] Therefore, in recent years, several large-scale genetic studies have been performed to identify the underlying genes and pathways leading to OA. Multiple significantly associated loci for OA of the hip and knee have been identified through genome-wide association studies (GWAS).[11–18] However, thus far, only one report has described a robust association with OA of the hand.[19] In this previous report, common variants in the *ALDH1A2* and rare variants in chromosome 1p31 were genome-wide significantly associated with hand OA using a discovery cohort of 837 cases and 77,325 controls.

In this study, we aimed to identify novel genes and pathways involved in the aetiology of OA of the hand by performing a large-scale genome-wide association study (GWAS). We used a semiquantitative measure for OA of the hand in order to increase statistical power. We gathered a large sample size of 12,754 individuals for analysis, by combining data from three studies in the discovery phase and an additional three cohorts for replication. Next, we conducted functional follow-up of our top finding to investigate the underlying mechanism.

Table 1: GWA Table 1. Results GWAS quantitative bilateral phenotype of osteoarthritis of the hand (KL sum score), discovery, replication and meta-analysis

					Discovery*			Replication**			Combined			
SNP	Chr	EA	NEA	EAF	Beta	SE	P-value	Beta	SE	P-value	Beta	SE	p-value	Locus†
rs1494593	5	T	C	0.88	-0.83	0.18	$4.5x10^{-06}$	-0.15	0.32	0.63	-0.67	0.16	$2.3x10^{-05}$	PRDM9--[]---C5orf17
rs114370021	5	A	G	0.27	-0.84	0.18	$2.1x10^{-06}$	0.48	0.32	0.13	-0.53	0.16	$7.0x10^{-04}$	[TENM2]
rs7770034	6	A	G	0.48	-0.54	0.12	$3.4x10^{-06}$	-0.36	0.21	$8.7x10^{-02}$	-0.50	0.10	$1.0x10^{-06}$	CDC5L--[]---SUPT3H
6:132063842:D‡	6	D	I	0.27	0.64	0.13	$1.3x10^{-06}$	0.41	0.23	$7.8x10^{-02}$	0.58	0.11	$3.8x10^{-07}$	[ENPP3]
11:90657297:D	11	D	I	0.11	0.93	0.19	$1.4x10^{-06}$	0.33	0.33	0.31	0.78	0.17	$2.9x10^{-06}$	DISC1FP1-[]----FAT3
rs4764133	12	T	C	0.39	0.75	0.12	$3.5x10^{-10}$	1.11	0.22	$3.3x10^{-07}$	0.83	0.10	$1.8x10^{-15}$	MGP--[]- ERP27
rs7139060	12	A	G	0.67	-0.73	0.12	$6.1x10^{-09}$	0.11	0.22	0.62	-0.52	0.11	$1.5x10^{-06}$	[CCDC91]
rs1950427	14	T	C	0.12	0.86	0.18	$1.1x10^{-06}$	-0.28	0.30	0.36	0.57	0.15	$1.8x10^{-04}$	STXBP6---[]---NOVA1
rs6108226	20	T	C	0.77	0.70	0.15	$3.9x10^{-06}$	0.20	0.27	0.44	0.58	0.13	$1.1x10^{-05}$	PLCB1--[]--PLCB4

* Discovery: RS-I, RS-II, RS-III, n= 8,743
** Replication : GARP, LSS, TwinsUK & FHS, n= 4,011
†SNP location represented by [], if the SNP is localized intergenic the dashes denotes the distance, -≤10 kb, --≤100kb, ---≤1000kb, ----≤1Mkb, -----≥1Mkb
‡For TwinsUK a proxy SNP was used: rs3850251 r2=1 D'=1 (as calculated in the RSI, RSII and RSIII cohorts)
EA: effect allele, NEA: non effect allele, SE: standard Error

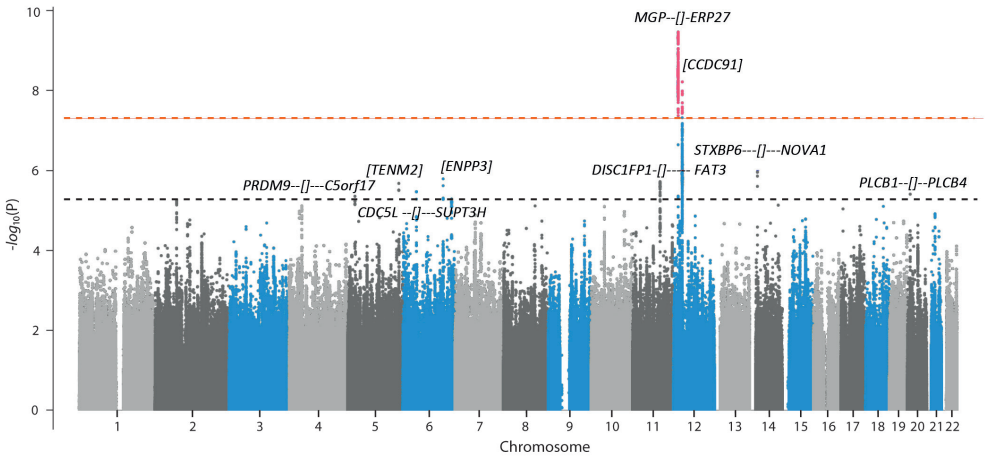
Results

GWAS of KL sum score

We conducted a GWAS of a semiquantitative measure of hand OA, a bilateral summed score of KL scores,[20] which grades radiographic OA severity, across all hand joints (KL sum score, range of 0–120). The discovery set consisted of three Rotterdam Study cohorts (RSI, RSII and RSIII) and included 8,743 participants with KL sum scores. Replication was done in another 4011 individuals from three different cohorts; LS, Framingham Heart Study and Twins UK (TUK). General characteristics of the discovery cohorts and replication cohorts can be found in **Supplementary Table 3** and **Supplementary Text 1**.

The discovery analysis yielded two novel independent genome-wide significant loci (p-value≤$5*10^{-8}$) on chromosome 12, an intergenic region between *matrix Gla protein (MGP)* and *ERP27* and an intronic region in *CCDC91*. We also identified seven other novel loci with suggestive significance (p-value<$5x10^{-06}$) (**Figure 1**). In total, nine loci were selected for replication in 4,011 individuals from three different cohorts (LS, FHS and TUK). Using a Bonferroni corrected p-value <$5.56x10^{-03}$, we significantly replicated one of nine loci, rs4764133 ($Beta_{meta}$=0.83, SE_{meta}=0.10, $p\text{-value}_{replication}=3.4x10^{-07}$, $p\text{-value}_{meta}=1.8x10^{-15}$) with the same direction of effect as identified in the discovery analysis (**Table 1** and **Supplementary Figure 1**). This locus maintained genome-wide significance and another locus near *ENPP3* reached near genome-wide significance (chr6: 132063842:D, $Beta_{meta}$=0.58, SE_{meta}=0.11, $p\text{-value}_{meta}=3.8x10^{-07}$) in the combined discovery and replication joint meta-analysis (**Table 1**). Since the KL sum score has a skewed distribution, the top hit was also reanalysed in the discovery set using a Poisson regression (rs4764133, $Beta_{poisson}$=0.12, $SE_{poisson}$=0.02, $p\text{-value}_{poisson}=1.98*10^{-11}$).

Our top replicated and genome-wide significant finding, rs4764133 [T] ($p\text{-value}_{meta}=1.80*10^{-15}$, Beta=0.83, MAF=0.39) is located in a non-coding intergenic region between *MGP* and *endoplasmic reticulum protein 27* (*ERP27*). However, variants in high LD with rs4764133 (r^2≥0.8) span a ~80 Kb region encompassing multiple genes, including *MGP* and an open-reading frame *C12orf60* (**Figure 2a**). Moreover, several of these variants are located in an mRNA transcript, including a non-synonymous variant in *MGP*, and variants in 3′ and 5′UTR of *MGP* and *C12orf60* (**Table 2, Figure 2b**). The non-synonymous variant in *MGP*, rs4236, is predicted to be non-damaging (STIFT=1, tolerated; polyPhen=0, benign) causing a threonine to alanine amino acid substitution. Two variants are located in predicted active promoter region of *MGP* (rs1800801) and *C12orf60* (rs9668569) in chondrogenic cells and primary osteoblasts (**Table 2**).

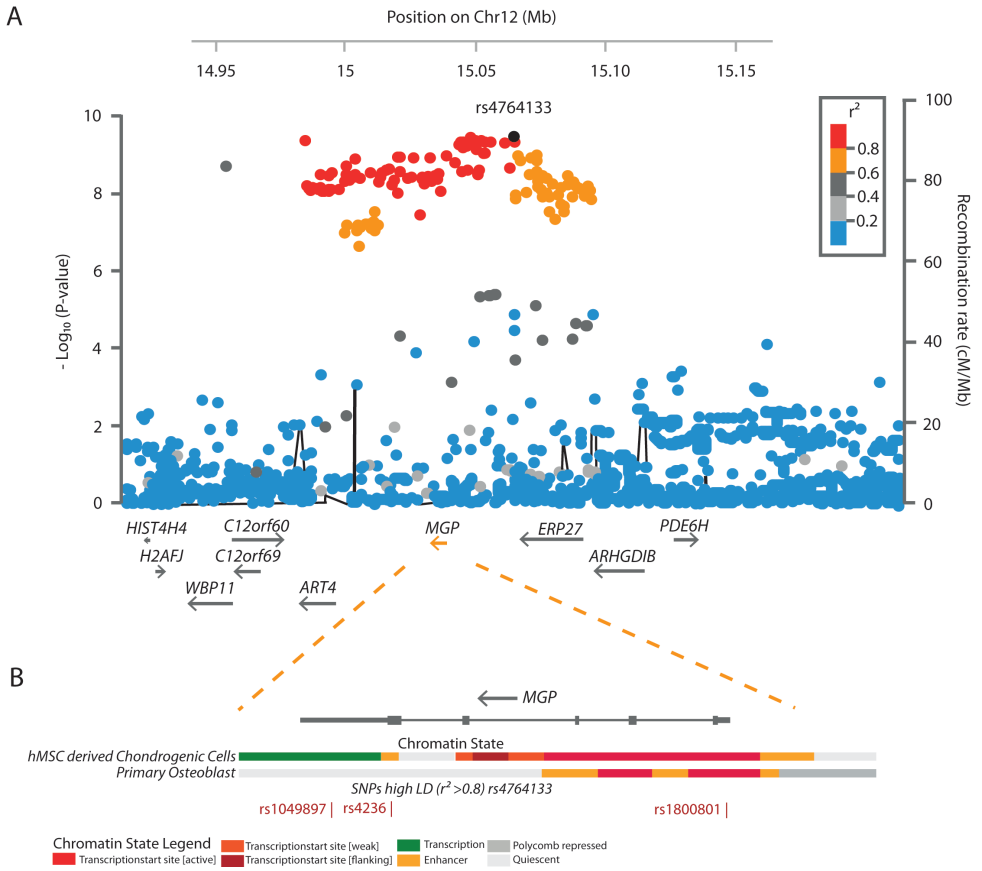


▲ **Figure 1: GWAS results for association with the KL sum score in the discovery phase.** Manhattan plot for association with the KL sum score, adjusted for age and sex, in the discovery cohorts of RSI, RSII and RSIII. The −log10 p-values, for each of the ~11 million SNPs analysed (remaining after EASYQC quality control) as part of the genome-wide association with the KL sum score, plotted against their position per chromosome. The red dotted horizontal line corresponds to the genome-wide significant threshold (p-value=5×10−08). The dotted grey line corresponds to the selection for replication threshold (p-value=5×10−06). SNP location represented by [], if the SNP is localised intergenic the dashes denotes the distance, -≤10 kb, --≤100 kb, ---≤1000 kb, ----≤1 Mkb, -----≥1 Mkb. GWAS, genome-wide association studies; KL, Kellgren and Lawrence score

Next, we investigated the association of rs4764133 with bilateral severe hand OA and bilateral finger OA using the discovery set (RSI, RSII and RSIII). We found a strong association with finger OA (p-value=3.09x10^{-08}, OR=1.25) and nominal significant association with severe hand OA (p-value=2.80x10^{-02}, OR=1.36), which has a low frequency in the population (**Supplementary Table 4.**). To see if rs4764133 also confers risk for other forms of OA, that is, OA of the hip and knee, we used the GWAS summary data of the treat OA consortium [27] and the recently published minimal joint space width of the hip (mJSW) meta-analysis.[18] No association was found between rs4764133 and hip or knee OA (**Supplementary Table 4.**). However, we did find a nominal significant association between rs1049897 and cartilage thickness in the hip joint(mJSW) (r^2=0.98 with rs4764133) (p value=1.28x10^{-02}, Beta=−0.398).

Gene expression analyses

In order to identify potential causal genes located in the LD block surrounding rs4764133, we assessed gene expression of *MGP*, *ERP27*, *ART4*, *SMOC3 (C12orf69)* and *C12orf60* in articular cartilage, the primary OA affected tissue. RNA sequencing was obtained on articular cartilage from patients with primary OA who had total joint replacement surgeries of either the knee (n=25) or hip (n=22) joint. Expression levels of *ERP27*, *C12orf60*, *ART4* and *SMOC3* were substantially lower than *MGP* expression levels



▲Figure 2: Locus zoom plot for rs4764133. 150kb upstream and downstream of rs4764133 has been taken as plotted region (A). Zoom in on MGP and three SNPs in high LD with top SNP that are located in the MGP mRNA transcript (B). Also represented is ROADMAP chromatin 18-state data of two tissue types: human mesenchymal stem cell (hMSC)-derived cultured chondrogenic cells and primary osteoblasts. In both these cell types, the chromatin contains active marks surrounding the MGP promoter. LD:linkage disequilibrium; MGP: matrix Gla protein.

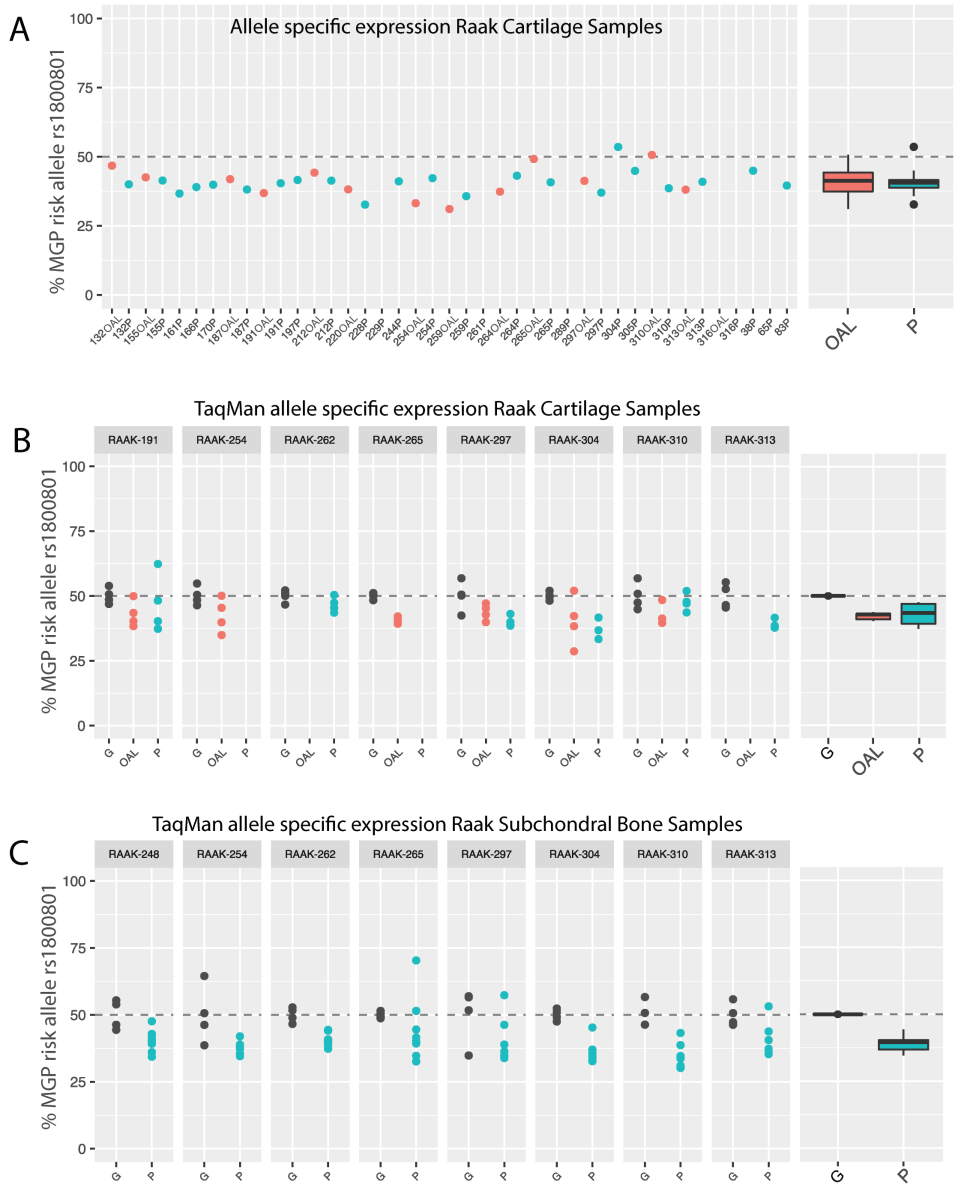
in articular cartilage (**Supplementary Figure 2a**). Nonetheless, neither *MGP*, *ERP27*, *ART4*,*SMOC3* nor *C12orf60* showed significant difference in gene expression between paired Preserved (P) and OA lesioned (OAL) articular cartilage. However, while these genes are not differentially expressed in OA affected cartilage, it is possible that the identified GWAS SNPs affect gene transcription. When we analysed the relationship between the top SNP and expression analysis in a classical expression quantitative trait loci (eQTL) analysis, we did not to detect significant correlations between rs1049897, rs4236 or rs1800801 and absolute *MGP*, *ERP27*, *ART4*, *SMOC3* or *C12orf60* expression levels (**Supplementary Figure 2b**). However, we did observe several variants in high LD located in the mRNA transcript of *MGP* and *C12orf60*, allowing us to assess allele specific expression (ASE) for these genes. We were unable to study ASE for *ART4*, *SMOC3*

and *ERP27*, since no SNP in high LD with rs4764133 is present in the coding region. In ASE, the influence of exonic alleles on gene expression *in -cis* is measured within heterozygote subjects, circumventing strong effects from environmental or trans-acting influences. This property results in ASE analysis to be a more statistically powerful approach, when compared with classical eQTL analysis.[28] Subsequently, we found that the OA risk alleles for three coding variants in high LD with the lead variant, rs4236 (**Supplementary Figure 3a**, 39.6% C allele, p-value<$5x10^{-16}$), rs1049897 (**Supplementary Figure 3b**, 44.4% A allele, p-value<$5x10^{-10}$) and rs1800801 (**Figure 3a**, 40.7% T allele, p-value<$5x10^{-16}$), were significantly correlated with lower expression of *MGP*, marking imbalanced expression among heterozygotes, independent of the disease status of the articular cartilage. No ASE was observed between SNPs rs11276, rs3088189 and rs1861698 (residing in *C12orf60* and in high LD with the lead SNP, r^2>0.8, **Table 2**). Technical and biological replication was performed using a custom allele-specific TaqMan assay for rs1800801 in eight additional heterozygous individuals for which we isolated RNA from P cartilage (n=2), OAL (n=2) or both (n=4) from patients with primary knee OA and confirmed the observed imbalance in preserved articular cartilage (**Figure 3b**, relative allelic difference=0.92, p-value<$1x10^{-06}$), as well as in eight knee subchondral bone samples (**Figure 3c**, relative allelic difference=0.78, p-value<$1x10^{-04}$).

Discussion

Here, we show for the first time, that there is a robust genome-wide significant association between rs4764133, located near *MGP*, and hand OA. Furthermore, we performed functional validation showing that *MGP* coding variants in LD with rs4764133 are associated with ASE of *MGP*, which may increase risk of hand OA by lowering inhibition of articular cartilage calcification, since MGP is an essential inhibitor of cartilage calcification. [29,30] These findings suggest that MGP could be considered a prioritised drug target for hand OA, since genetically supported drug targets double the success rate of therapeutics in clinical development.[31]

MGP is an essential inhibitor of cartilage calcification, and genetic deficiencies of *MGP* in humans and mice have been linked to abnormal mineralisation of soft tissues, including cartilaginous tissue.[29,32] Furthermore, *MGP* has been previously implicated in relation to OA. A small candidate study reported marginally significant association between hand OA and genetic variants in *MGP* (rs1800802 and rs4236).[33] This is consistent with our findings that the minor allele for rs4764133 and related coding variants in high LD (r2>0.8), rs1800802 and rs4236, increase the risk of hand OA and that we found high expression of *MGP* in both P and OAL articular cartilage. In contrast,



▲Figure 3: Allelic imbalanced expression of MGP marked by the alleles among heterozygotes of rs1800801 (a), in the assessed cartilage RNA sequencing dataset. Validation of selected rs1800801 using a custom TaqMan assay confirmed the imbalance (b). Allelic imbalance was also assessed with a custom TaqMan assay in subchondral bone samples (C). Preserved (P) and OA lesioned (OAL) samples are shown respectively in blue and red, and genomic DNA (TaqMan control) in black (G). For ASE results for rs4236 and rs1049879, see Supplementary Figure 3 and for information on the samples, see Supplementary Table 2 ASE, allele-specific expression; MGP, matrix Gla protein; OA, osteoarthritis.

another study showed that an MGP protein complex is excreted by healthy articular chondrocytes, but not by OA-affected chondrocytes,[34] although we only assessed *MGP* expression and not MGP protein complex excretion.

Although the loci with ASE are known to be enriched for eQTLs, [35] we were unable to detect an association between the MGP genotype and *MGP* RNA expression levels in cartilage. This could have been due to our modest sample size (knee joint, n=25 and or hip joint, n=22) in combination with large heterogeneity of the tissue. Notably, the available cartilage samples originated from different joint sites (knee and hip) and different disease stage (preserved versus affected) and had large age range of the individuals. Also, it is known that ASE is a more powerful technique than classical eQTL analysis to identify functional SNPs influencing expression of genes.[28] While the extent of imbalance could be considered relatively modest, an increasing number of OA associated SNP alleles appear to mark ASE by comparable amount.[19,36–38] From a more biological perspective, one could consider a prolonged, although slight, deviation from homeostasis due to modest ASE of cartilage relevant genes to be of substantial influence over time. This latter hypothesis could contain the molecular basis for increased risk towards developing OA among the ageing population. Additionally, we observed that the rs1800801 alleles also affected expression of MGP in subchondral bone samples. This could imply that, in parallel to an effect in cartilage, the presumed disturbed cartilage homeostasis is further affected by the underlying bone, further enabling the view that OA is a pathology of the entire joint.

Our findings may give an explanation for the known vitamin K association with OA: MGP-mediated calcification inhibition is dependent on γ-carboxylation by vitamin K.[39] It has been shown that low vitamin K intake is correlated with OA.[40] Thus, vitamin K intake may be a potential therapeutic treatment in OA. Recently, a first randomised control trial testing the effects of vitamin K on OA was published, which reported no overall effect of vitamin K on hand OA.[41] Despite the low power of the trial, there was a significant beneficial effect on joint space narrowing (cartilage degradation) among those individuals that were vitamin K deficient at the start of the trial.[41] Thus, an adequately powered study of vitamin K may be justified based on the found *MGP* association. Furthermore, genetic predisposition for hand OA was not taken into account in the trial. Perhaps, genetic predisposition for hand OA (*MGP*-risk variants) in combination with insufficient vitamin K intake might potentiate cartilage calcification and subsequent risk for developing hand OA. Therefore, future OA trails, therapeutic and preventive treatments might benefit from taking a personalised medicine approach since genetically supported drug targets double the success rate of therapeutics in clinical development.[31]

Styrkarsdottir *et al [*19] reported on common genetic variants that associate with severe hand OA, among the replication cohorts were the Leiden and Rotterdam cohorts. [19] Although we observe suggestive signals at the reported locus (*ALDH1A2* gene, 1p31), the respective variants did not meet the genome-wide significance threshold in our analyses (**Supplementary Table 5**). This difference is likely caused by the markedly different phenotypes that were used for either analyses. Where Styrkarsdottir *et al* studied a dichotomous severe hand OA phenotype, our phenotype was semi-quantitatively phenotype.

To conclude, we here present coding variants in *MGP* that are associated with radiographic hand OA, and the hand OA risk allele marks lower expression of *MGP* in articular cartilage. Our findings suggest that *MGP* might play an important role in hand OA pathogenesis through pathways related to articular cartilage calcification and vitamin K. Better understanding of MGP gene and protein regulation and its relation to vitamin K intake and OA may reveal novel therapeutic drug targets for hand OA.

Materials & Methods

Discovery GWAS, replication and meta-analysis

For a detailed description on the GWAS methods, participating studies, quality control procedures for genotyping and imputation, see online **Supplementary Text 1** and **Supplementary Table 1**.

Detailed phenotype description of Kellgren and Lawrence sum score

We have used a semi-quantitative bilateral measure of OA of the hand based on the radiographic Kellgren and Lawrence (KL) score.[20] Using radiographs of both hands, the KL score was determined for each joint in the hand. Using these KL scores we defined the KL sum score: the total KL score, the sum, of the following hand joints for both hands (left and right): all distal interphalangeal joints, all proximial interphalangeal joints, all metacarpophalangeal (MCP) joints, the interphalangeal joint and the first carpometacarpal joint, which gives the sum of 15 joints on each hand, and in total 30 joints for both hands together, resulting in a minimum score of 0 and a maximum score of 120. In the Leiden Studies (LS) cohort, no KL scoring was done of the MCP joints, resulting in a KL sum score of maximum 88. Individuals lacking KL grading for both hands or one hand and individuals with missing age or gender information were excluded from all analyses in all cohorts. As the KL sum score has a skewed distribution, the top finding of the meta-analysis was repeated in the discovery cohorts using a Poisson regression.

Visualisation of the associated loci and the regulatory landscape

For the top GWAS-associated single nucleotide polymorphism (SNP), the linkage disequilibrium (LD) region (r^2 >0.8) was determined using the 1000G Phase-1 population using the HaploReg V3 tool.[21] Using the ROADMAP-generated reference epigenomes, we determined if any of the variants in high LD were located in potential gene regulatory regions in primary osteoblasts (generated by ENCODE) and bone marrow-derived chondrocytes (ROADMAP).[22,23] The 18-state chromatin reference epigenomes were downloaded from the ROADMAP epigenomes data portal.[23] SNPs and regulatory annotations were visualised using the UCSC Genome Browser on GRCh37/hg19.[24] For each variant, it was also determined if the alternative allele would disrupt a protein binding motif; this was done using the HaploReg V3 tool.[21]

RNA sequencing data

Post-RNA isolation (Qiagen RNeasy Mini Kit, RIN >7) of 40 knee (15 paired preserved (P) and OA lesioned (OAL), 7 P only and 3 OAL only) and 28 hip (six paired P and OAL, 14 P only and 2 OAL only) cartilage samples (**online supplementary table S2**), paired-end 2×100 bp RNA library sequencing (Illumina TruSeq RNA-Library Prep Kit, Illumina HiSeq2000) resulted in an average of 10 million fragments per sample. Reads were aligned using GSNAP against GRCh37/hg19, in which SNPs from the Genome of the Netherlands consortium with a minor allele frequency (MAF) >1% were masked to prevent alignment bias. Number of fragments per gene were used to assess quantile-adjusted conditional maximum likelihood (edgeR, R-package). Subsequently, differential gene expression analysis was performed pairwise between P and OAL samples for which we had RNA of both (n=21). Allele-specific expression (ASE) was assessed using SNVMix2 [25] with default settings(min coverage=25, 10 reads per allele). The extent of ASE was defined as the fraction of risk allele among all counts at the respective location. Meta-analysis was done only across P samples or OAL when no P counterpart sample was present. p Values were calculated using canonical binominal test (metagen R-package).

TaqMan assay

Conventional TaqMan genotyping was performed on both genomic DNA (gDNA), articular cartilage and subchondral bone cDNA. An allele-specific custom TaqMan assay for rs1800801 (Thermo Fisher Scientific) was used to quantify the allele ratio in cDNA samples and were normalised against the gDNA ratio, which was used as an 1:1 allele ratio reference. Each sample has been measured in four (cartilage) or eight (subchon-

dral bone) times, while calculations and statistics were performed as described previously.[19,26] Cartilage samples that yielded fewer than four measurements (n=2) were discarded prior to further analyses. All subchondral bone samples were assessed by eight technical replicates.

Acknowledgments

The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters, MSc, and Carolina Medina-Gomez, MSc, for their help in creating the GWAS database, and Karol Estrada, PhD, Yurii Aulchenko, PhD, and Carolina Medina-Gomez, MSc, for the creation and analysis of imputed data. We thank Nico Lakenberg, Ruud van der Breggen and Eka Suchiman for their help in preparing DNA and RNA samples.

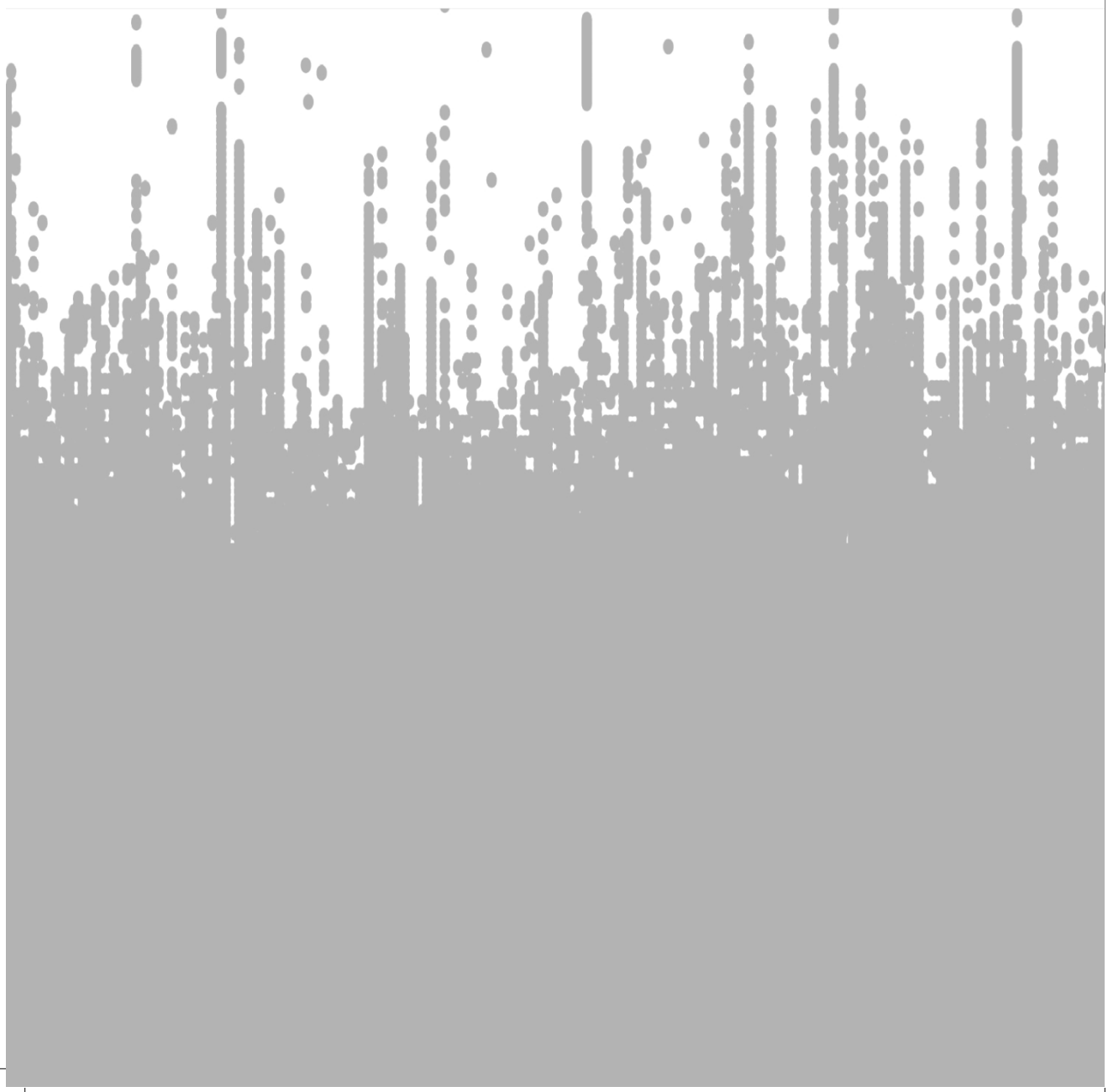
Supplementary Material

Additional material is published online only, to view please visit the paper online at the journal website (https://ard.bmj.com/content/76/12/2046.long)

References

1. CDC. Prevalence of Doctor-Diagnosed Arthritis and Arthritis-Attributable Activity Limitation — United States, 2010–2012. MMWR Morb Mortal Wkly Rep 2013;62:869–73. doi:10.1001/jama.296.22.2671
2. Patel N, Buckland-Wright C. Advancement in the zone of calcified cartilage in osteoarthritic hands of patients detected by high definition macroradiography. Osteoarthr Cartil 1999;7:520–5.http://www.idealibrary.com (accessed 2 Nov 2016).
3. Fuerst M, Bertrand J, Lammers L, et al. Calcification of articular cartilage in human osteoarthritis. Arthritis Rheum 2009;60:2694–703. doi:10.1002/art.24774
4. Nelson AE, DeVellis RF, Renner JB, et al. Quantification of the whole-body burden of radiographic osteoarthritis using factor analysis. Arthritis Res Ther 2011;13:R176. doi:10.1186/ar3501
5. Jonsson H, Helgadottir GP, Aspelund T, et al. Hand Osteoarthritis Severity is Associated with Total Knee Joint Replacements Independently of BMI. The Ages-Reykjavik Study. Open Rheumatol J 2011;5:7–12. doi:10.2174/1874312901105010007
6. CDC. Prevalence of Disabilities and Associated Health Conditions Among Adults --- United States, 1999. MMWR Morb Mortal Wkly Rep 2009;58:421–6.
7. Spector TD, Cicuttini F, Baker J, et al. Genetic influences on osteoarthritis in women: a twin study. BMJ 1996;312:940–3. doi:10.1136/bmj.312.7036.940
8. Felson DT. Risk factors for osteoarthritis: understanding joint vulnerability. Clin Orthop Relat Res 2004;:S16–21. doi:10.1097/01.blo.0000144971.12731.a2
9. Bijkerk C, Houwing-Duistermaat JJ, Valkenburg H a., et al. Heritabilities of radiologic osteoarthritis in peripheral joints and of disc degeneration of the spine. Arthritis Rheum 1999;42:1729–35. doi:10.1002/1529-0131(199908)42:8<1729::AID-ANR23>3.0.CO;2-H
10. Ishimori ML, Altman RD, Cohen MJ, et al. Heritability patterns in hand osteoarthritis: the role of osteophytes. Arthritis Res Ther 2010;12:R180. doi:10.1186/ar3144
11. Kerkhof HJM, Lories RJ, Meulenbelt I, et al. A genome-wide association study identifies an osteoarthritis susceptibility locus on chromosome 7q22. Arthritis Rheum 2010;62:499–510. doi:10.1002/art.27184
12. Nakajima M, Takahashi A, Kou I, 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. doi:10.1371/journal.pone.0009723
13. Castaño Betancourt MC, Cailotto F, Kerkhof HJ, 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. doi:10.1073/pnas.1119899109
14. Consortium A, Collaborators A. Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. Lancet 2012;380:815–23. doi:10.1016/S0140-6736(12)60681-3
15. Day-Williams AG, Southam L, Panoutsopoulou K, et al. A variant in MCF2L is associated with osteoarthritis. Am J Hum Genet 2011;89:446–50. doi:10.1016/j.ajhg.2011.08.001
16. Evangelou E, Kerkhof HJ, Styrkarsdottir U, et al. A meta-analysis of genome-wide association studies identifies novel variants associated with osteoarthritis of the hip. Ann Rheum Dis 2013;:1–7. doi:10.1136/annrheumdis-2012-203114
17. Miyamoto Y, Shi D, Nakajima M, et al. Common variants in DVWA on chromosome 3p24.3 are associated with susceptibility to knee osteoarthritis. Nat Genet 2008;40:994–8. doi:10.1038/ng.176
18. Castaño-Betancourt MC, Evans DS, Ramos YFM, et al. Novel Genetic Variants for Cartilage Thickness and Hip Osteoarthritis. PLOS Genet 2016;12:e1006260. doi:10.1371/journal.pgen.1006260
19. Styrkarsdottir U, Thorleifsson G, Helgadottir HT, 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. doi:10.1038/ng.2957
20. Kellgren JH, Lawrence JS. Radiological Assessment of Osteo-Arthrosis. Ann Rheum Dis 1957;16:494–502. doi:10.1136/ard.16.4.494
21. Ward LD, Kellis M. HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 2012;40:930–4. doi:10.1093/nar/gkr917

22. Chadwick LH. The NIH Roadmap Epigenomics Program data resource. Epigenomics 2012;4:317–24. doi:10.2217/epi.12.18
23. Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518:317–30. doi:10.1038/nature14248
24. Kent WJ, Sugnet CW, Furey TS, et al. The Human Genome Browser at UCSC. Genome Res 2002;12:996–1006. doi:10.1101/gr.229102
25. Goya R, Sun MGF, Morin RD, et al. SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. Bioinformatics 2010;26:730–6. doi:10.1093/bioinformatics/btq040
26. Bos SD, Bovée JVMG, Duijnisveld BJ, 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. doi:10.1136/annrheumdis-2011-200981
27. Kerkhof HJM, Doherty M, Arden NK, et al. Large-scale meta-analysis of interleukin-1 beta and interleukin-1 receptor antagonist polymorphisms on risk of radiographic hip and knee osteoarthritis and severity of knee osteoarthritis. Osteoarthr Cartil 2011;19:265–71. doi:10.1016/j.joca.2010.12.003
28. Pastinen T. Genome-wide allele-specific analysis: insights into regulatory variation. Nat Rev Genet 2010;11:533–8. doi:10.1038/nrg2815
29. Luo G, Ducy P, McKee MD, et al. Spontaneous calcification of arteries and cartilage in mice lacking matrix GLA protein. Nature 1997;386:78–81. doi:10.1038/386078a0
30. Yagami K, Suh JY, Enomoto-Iwamoto M, et al. Matrix GLA protein is a developmental regulator of chondrocyte mineralization and, when constitutively expressed, blocks endochondral and intramembranous ossification in the limb. J Cell Biol 1999;147:1097–108.http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2169349&tool=pmcentrez&rendertype=abstract
31. Nelson MR, Tipney H, Painter JL, et al. The support of human genetic evidence for approved drug indications. Nat Genet 2015;47:856–60. doi:10.1038/ng.3314
32. Munroe PB, Olgunturk RO, Fryns JP, et al. Mutations in the gene encoding the human matrix Gla protein cause Keutel syndrome. Nat Genet 1999;21:142–4. doi:10.1038/5102
33. Misra D, Booth SL, Crosier MD, et al. Matrix Gla protein polymorphism, but not concentrations, is associated with radiographic hand osteoarthritis. J Rheumatol 2011;38:1960–5. doi:10.3899/jrheum.100985
34. Wallin R, Schurgers LJ, Loeser RF. Biosynthesis of the vitamin K-dependent matrix Gla protein (MGP) in chondrocytes: a fetuin-MGP protein complex is assembled in vesicles shed from normal but not from osteoarthritic chondrocytes. Osteoarthritis Cartilage 2010;18:1096–103. doi:10.1016/j.joca.2010.05.013
35. Hangoc G, Williams DE, Falkenburg JH, et al. Influence of IL-1 alpha and -1 beta on the survival of human bone marrow cells responding to hematopoietic colony-stimulating factors. J Immunol 1989;142:4329–34.
36. Gee F, Clubbs CF, Raine EVA, et al. 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. doi:10.1186/1471-2350-15-53
37. Raine EVA, Dodd AW, Reynard LN, et al. Allelic expression analysis of the osteoarthritis susceptibility gene COL11A1 in human joint tissues. BMC Musculoskelet Disord 2013;14:85. doi:10.1186/1471-2474-14-85
38. Bos SD, Bovée JVMG, Duijnisveld BJ, 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. doi:10.1136/annrheumdis-2011-200981
39. Schurgers LJ, Uitto J, Reutelingsperger CP. Vitamin K-dependent carboxylation of matrix Gla-protein: a crucial switch to control ectopic mineralization. Trends Mol Med 2013;19:217–26. doi:10.1016/j.molmed.2012.12.008
40. Shea MK, Kritchevsky SB, Hsu F-C, et al. The association between vitamin K status and knee osteoarthritis features in older adults: the Health, Aging and Body Composition Study. Osteoarthritis Cartilage 2015;23:370–8. doi:10.1016/j.joca.2014.12.008
41. Neogi T, Felson DT, Sarno R, et al. Vitamin K in hand osteoarthritis: results from a randomised clinical trial. Ann Rheum Dis 2008;67:1570–3. doi:10.1136/ard.2008.094



Chapter 3.2

Vitamin K antagonist anticoagulant usage is associated with increased incidence and progression of osteoarthritis

Boer, C.G., Szilagyi, I., Nguyen, N.L., Neogi, T., Meulenbelt, I., Ikram, M.A., Uitterlinden, A.G., Bierma-Zeinstra, S., Stricker, B. & van Meurs, J.B.J.

Manuscript submitted to Annals of the Rheumatic Diseases (2020)

Abstract

Objectives: Vitamin K is hypothesized to play a role in osteoarthritis pathogenesis through effects on vitamin K-dependent bone and cartilage proteins, and therefore may represent a modifiable risk factor. A genetic variant in a vitamin K-dependent protein that is an essential inhibitor for cartilage calcification, Matrix Gla Protein (MGP), is associated with an increased risk for osteoarthritis. Vitamin K antagonists anticoagulants (VKA), such as warfarin and acenocoumarol, act as anticoagulants through inhibition of vitamin K-dependent blood coagulation proteins. VKA likely also affect the functioning of other vitamin K-dependent proteins such as MGP.

Methods: We investigated the effect of acenocoumarol usage on progression and incidence of radiographic osteoarthritis in 4,492 participants of the Rotterdam Study cohort. We also examined the effect of MGP and VKORC1 single nucleotide variants on this association.

Results: Acenocoumarol usage was associated with an increased risk of osteoarthritis incidence and progression (OR=2.50, 95%CI=1.94-3.20), both for knee (OR=2.34, 95%CI=1.67-3.22) and hip osteoarthritis (OR=2.74, 95%CI=1.82-4.11). Among acenocoumarol users, carriers of the high *VKORC1(BB)* expression haplotype together with the *MGP* osteoarthritis risk allele (rs1800801-T) had an increased risk of osteoarthritis incidence and progression (OR=4.18, 95%CI=2.69-6.50), while this relationship was not present in non-users of that group (OR=1.01, 95%CI=0.78-1.33).

Conclusions: These findings support the importance of vitamin K and vitamin K-dependent proteins, as MGP, in the pathogenesis of osteoarthritis. Additionally, these results may have direct implications for the clinical prevention of osteoarthritis, supporting the consideration of direct oral anticoagulants in favour of VKAs.

Introduction

Osteoarthritis (OA) is a chronic disabling joint disease that also increases in prevalence with age. It is the most common form of arthritis, one of the fastest growing chronic diseases worldwide[1], and is the fourth leading cause of years lived with disability globally[2]. To date, there are no known therapies that can alter its progression or prevent its occurrence. Apart from obesity and knee injury, very few other modifiable risk factors have been identified. Vitamin K has been hypothesized to play a role in OA pathogenesis through its effects on several vitamin K-dependent bone and cartilage proteins, and therefore may represent a modifiable risk factor. A number of observational studies reported an association between vitamin K status and prevalence and incidence of OA[3-5]. There has been one modestly sized clinical trial studying the effect of Vitamin K supplementation on OA progression. This ancillary study, originally designed to study vascular calcification, reported no overall beneficial effects of vitamin K supplementation. However, in individuals with insufficient Vitamin K levels at baseline a beneficial effect was observed [6]. No studies to date have evaluated the relation of Vitamin K antagonists anticoagulants (VKAs), which can be expected to result in low vitamin K functioning, which may lead to OA incidence or progression[3-5].

Vitamin K is an essential cofactor in the posttranslational γ-carboxylation of glutamic acid to form γ-carboxyglutamic acid (Gla) residues, which confer functionality to Gla-proteins. VKAs deplete the active form of vitamin K by inhibiting the enzyme vitamin K epoxide reductase complex 1 (VKORC1). Genetic variants of *VKORC1* account for approximately 25% of the variance in VKA dose[7]. Matrix Gla protein (MGP) is a vitamin K-dependent Gla- protein that is an essential inhibitor of cartilage and vascular mineralization[8, 9]. Recent, genome-wide association study (GWAS) and functional studies identified *MGP* to be causally involved in osteoarthritis[10].

Vitamin K antagonists anticoagulants (VKAs) such as warfarin and acenocoumarol are primarily prescribed for the prevention of thromboembolic events in patients with atrial fibrillation (AF)[11]. With aging-related increases in prevalence of AF, the projected number of individuals with AF needing anti-coagulation is predicted to rise to 17.9 million by 2060 in the European Union[12]. While a new class of anti-coagulants are available, the non-vitamin K oral anti-coagulants (NOACs), VKAs are still widely prescribed, particularly to older adults[13]. Whether long-term VKA use with resultant impairment of vitamin K-dependent proteins such as MGP increases risk of OA incidence or progression is not known. Given the high prevalence of VKA users in addition to the high prevalence of OA globally, clarifying this relationship would have substantial public health impact by identifying a potentially modifiable risk factor for OA.

We therefore examined the relation of VKA use to progression and incidence of hip and knee OA in two subcohorts of the large prospective population-based cohort of the Rotterdam Study. We additionally examined how the impact of VKA use varies by presence of the *MGP* risk allele that influences MGP expression and SNVs affecting VKORC1 gene expression, which impact VKA dosage.

Methods

Study population and clinical data

The Rotterdam study (RS) is a large prospective population-based cohort study ongoing since 1990 to study determinants of chronic disabling diseases in the elderly[14]. It consists of separate subcohorts (RS-I, RS-II, RS-III). All RS cohort participants live in the Ommoord district of the city of Rotterdam, the Netherlands. Residents of 55 years and older were first recruited in 1990. In 2000, a second cohort, RS-II, was started with individuals who had become 55 years of age or moved into the study district since the start of the study. Follow-up data were collected as follow-up visits every ~5-6 years. Details of the design and rationale of the Rotterdam Study has been published elsewhere[14]. Participant measurements at baseline and follow-up were obtained during visits to the research center for physical examinations, computerized pharmacy records, and from home interviews. Our study included participants of RS-I and RS-II for whom radiographs of knee and hip joints at baseline and follow-up visit were present, obtained and scored (**Supplementary Figure S1**). Additional information included: sex, age at baseline visit, body mass index (BMI, kg-1m2), physical activity (metabolic equivalent of task/hours per week), smoking (never, former and current smoker), locomotor disability, education level (UNESCO education classification), Diabetes mellitus, hypertension, femoral neck bone mineral density (FN-BMD), HDL/total cholesterol ratio and the Stanford Health Assessment Questionnaire (**see Supplementary Text for details**).

The Rotterdam Study has been approved by the institutional review board (Medical Ethics Committee) of the Erasmus Medical Center and by the review board of The Netherlands Ministry of Health, Welfare and Sports. The approval has been renewed every 5 years (MEC 02.1015). All participants provided written informed consent for participation in the Rotterdam Study.

Incidence and Progression of OA

Our study included participants of RS-I and RS-II for whom radiographs of knee and hip joints at baseline and follow-up visit were obtained and scored by trained medical

professionals for OA severity using Kellgren and Lawrence Grade (KLG)[15, 16] (**Supplementary Figure S1**). Individuals which had at baseline locomotor disability were excluded from out study population[17] (**Supplementary Figure 1**). We evaluated any OA progression, defined as an increase of KLG between baseline and follow-up of ≥1 and/or joint replacement; if baseline KLG was 0, progression was defined as increase of KLG ≥2 (incidence) [18]. Joints with a baseline KLG of 4 or baseline joint replacement were excluded from analysis (**Supplementary Figure S1**). OA progression was defined in a joint- and side-specific manner (knee, hip; left and right). Joints with no progression of OA comprised the referent group. Joints with missing data were excluded (**Supplementary Figure S1**), with the exception of joints with missing baseline data and a KLG of ≤1 at follow-up, which were included in the referent group (**Supplementary Figure S1**).

Vitamin K antagonists

For each participant, we extracted the usage of VKAs (acenocoumarol) for the period between baseline visit (RS-I-1, RS-II-1) and follow-up visits (RS-I-3, RS-II-2), from computerized pharmacy data (**supplementary text**). Acenocoumarol is the main prescribed VKA in the Netherlands as warfarin is not registered for use as a drug. All participants taking VKAs attended an anticoagulation clinic, which is standard practice in the Netherlands[19]. We excluded participants who were taking VKA (n=148) during the baseline visit to avoid prevalent user bias. We defined VKA usage as any acenocoumarol or usage during the period between the baseline (RS-I, RS-II) and follow-up visit (RS-I 3, RS-II 2), regardless of duration or dosage. To examine the effects of increasing duration of use, we defined duration of use by tertiles: ≤180 days, >180 days and ≤556 days, and, >556 days of use.

Genetic data and haplotype analysis

Methods for DNA-isolation, genotyping, quality control and data processing have been described elsewhere[10]. Data from 11 SNVs were extracted from the GWAS dataset: the *MGP* SNV previously found associated with OA[10] and ten SNVs needed for the *VKORC1* H-haplotypes as described on the PharmGKB database[20] (Supplementary Table S1). Haplotypes were inferred from all available genotypes(N=8,448), using genotype dosage data (HRC panel v.1.1[21]) and R-package haplo.stats[22]. Haplotypes were grouped based on VKA maintenance dose/*VKORC1* expression association; A: low dose VKA requirement/low *VKORC1* expression and B: high dose VKA requirement/high *VKORC1* expression[23]. Study participants were further stratified into: low expression/dose (AA), intermediate expression/dose (AB) and high expression/dose (BB) groups (**Table 1, Supplementary Table S1**).

Statistical analysis

We evaluated the relation of VKA/antiplatelet to the risk of overall progression of OA of either the knee or hip using logistic regression with generalized estimating equations (GEE) to account for correlations between joints within an individual[18]. We repeated the analyses stratified by *VKORC1* and *MGP* genotype/haplotype. The following covariates were included in all analyses: sex, age, BMI, physical activity, smoking, education level, Diabetes mellitus, hypertension, femoral neck BMD, HDL/total cholesterol ratio, baseline OA severity, time between follow-up visits, joint modeled and RS cohort (**Supplementary Text**).

Results

Relation of Acenocoumarol use to OA progression

A total of 4,018 participants of two large prospective older-age population-based cohorts, the Rotterdam Study (RS), with RS-I, contributing 2,748 individuals, while RSII contributed 1,002 participants[14]. See **Table 1** for the general characteristics of the study population. At baseline there were 358 individuals with OA (KLG ≥2), 75 with hip OA and 290 with knee OA. We identified 268 new users of acenocoumarol (VKA) (RS-I n=230, RS-II n=38) in our study population in our follow-up period. When we examined the incidence and progression of OA in acenocoumarol users and non-users, there was a >2-fold higher risk for overall OA incidence/progression (i.e., OA of the knee or hip) in acenocoumarol users compared with non-users (OR=2.50, 95% CI=1.94-3.20, **Table 2**). This association was also observed in each subcohort (RS-I and RS-II) separately (**Supplementary Table S2**). Overall OA incidence/progression risk estimates remained similar for longer duration of acenocoumarol use (tertiles): ≤180 days (OR=2.82, 95% CI=1.90-4.20, >180 days and ≤556 days (OR=2.94, 95% CI=2.00-4.32), with only a reduction of risk with long term use, >556 days of acenocoumarol use (OR=1.74, 95% CI = 1.10-2.76) (**Table 2**).

Increased risk of overall OA incidence/progression in acenocoumarol users was also observed in the knee and hip joints separately (**Table 2**) [knee OA (OR=2.34, 95% CI=1.69-3.22) and hip OA(OR=2.74, 95% CI=1.82-4.11)], as well as in each subcohort separately (**Supplementary Table S2**). Interestingly, longer duration of acenocoumarol use do seem to have slight different effects om knee osteoarthritis incidence and progression than on hip. Knee OA risk seems to increase for longer duration of acenocoumarol use, whereas risk for hip OA seems to decrease for longer durations of aceno-

Table 1: Study population characteristics

	Study population (RSI & RSII)	
	Non-users (N = 3,750)	Acenocoumarol users (N = 268)
General characteristics	Mean (SD) or N (%)	
RSI	2,748 (73.3%)	230 (85.8%)
RSII	1,002 (26.7%)	38 (14.2%)
Age, years	64.2 (6.4)	66.6 (6.8)
Females (%)	2,068 (55.1%)	130 (48.5%)
Body Mass Index (BMI), kg/m2	26.3 (3.5)	26.9 (3.5)
Follow-up period, months, median, [IQR]	76.1 [55.2, 78.5]	76.7 [75.1, 79.4]
Osteoarthritis(OA) prevalence at baseline visit†		
Hip OA (%)	65 (1.7%)	10 (3.7%)
Knee OA (%)	264 (7.4%)	26 (9.7%)
MGP allele occurrence (rs1800801)		
Non-risk allele (A/A)	1,345 (38.6%)	90 (37.2%)
Risk allele carrier (T/*)	2,137 (61.4%)	152 (62.8%)
VKORC1 group occurrence§		
Low - AA	522 (15.0%)	45 (18.7%)
Intermediate - AB	1,303 (37.5%)	94 (39.0%)
High - BB	1,646 (47.4%)	102 (42.3%)

* P-values calculated either by the Students T-test for normal distributed data, for non-normal distributions the Kruskal-Wallis test and, categorical variables were tested with Chi-squared test.
†Osteoarthritis at baseline was defined as radiographic OA, Kellgren-Lawrence score ≥2 in either the left or right or both investigated joints.
§VKORC1 groups based on VKORC1 H-Haplotypes and their association with VKORC1 expression/VKAs maintenance dosage; low: low VKORC1 expression and associated with lower required dosage, High: high VKORC1 expression and associated with higher required dosage.

coumarol use. These differences may represent true biological differences between the joints. However, this is more likely the effect of low statistical power (tertiles analysis) or the statistical power difference between the joints. As there were more cases of knee OA incidence and progression (n=385 joints) in our study cohorts than for hip OA (n=216 joints).

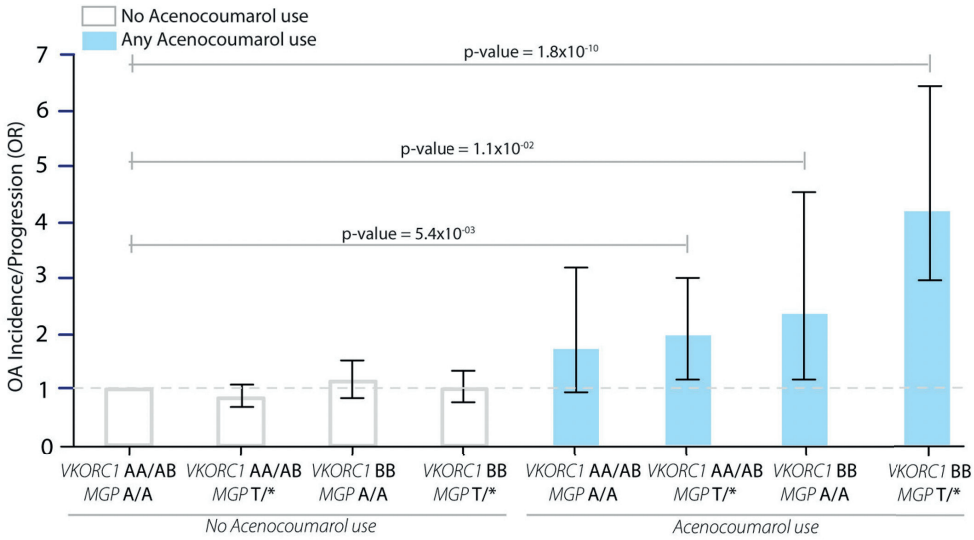
MGP, VKORC1 genetics and acenocoumarol affect OA progression

Maintenance dosages of VKAs are dependent on genetic variants (single nucleotide variations, SNVs) affecting *VKORC1* expression activity[19, 24]. This altered *VKORC1* expression can also affect *MGP* γ-carboxylation, as *VKORC1* is needed for γ-carboxylation by vitamin K, process that is needed to activate *MGP*. We therefore examined the extent to which acenocoumarol users with a genetic predisposition for decreased *MGP* and/or

Table 2: Association between acenocoumarol use and risk of OA progression in RSI and RSII

	Overall Osteoarthritis Progression					Overall progression of Knee Osteoarthritis					Overall progression of Hip Osteoarthritis				
	Joints N*	Progression N (%)	OR	95% CI	p-value	Joints N*	Progression N (%)	OR	95% CI	p-value	Joints N*	Progression N (%)	OR	95% CI	p-value
Non-users	12,594	506 (4.0%)	1	-	-	6,162	329 (5.3%)	1			6,432	177 (2.8%)	1	-	-
Users	863	94 (10.9%)	2.50	1.94- 3.20	$9.0x10^{-13}$	426	55 (12.9%)	2.34	1.69- 3.22	$2.4x10^{-07}$	437	39 (8.9%)	2.74	1.82- 4.11	$1.2x10^{-06}$
Duration of Acenocoumarol usage															
Non users	12,594	506 (4.0%)	1	-	-	6,162	329 (5.3%)	1	-	-	6,432	177 (2.8%)	1	-	-
≤ 180 days	279	35 (12.5%)	2.82	1.90- 4.20	$3.3x10^{-07}$	144	15 (10.4%)	1.78	1.01- 3.18	$4.8x10^{-02}$	135	20 (14.8%)	4.84	2.73- 8.56	$5.9x10^{-08}$
≤ 556 days	285	36 (12.6%)	2.94	2.00- 4.32	$3.8x10^{-08}$	135	20 (14.8%)	2.69	1.62- 4.46	$1.2x10^{-04}$	150	16 (10.7%)	3.33	1.79- 6.18	$1.4x10^{-04}$
>556 days	299	23 (7.7%)	1.74	1.10- 2.76	$1.9x10^{-02}$	147	20 (13.6%)	2.65	1.57- 4.46	$2.5x10^{-04}$	152	3 (2.0%)	0.27	0.18- 1.85	0.35

Overall progression of Osteoarthritis(OA) in RS-I and RS-II within the follow-up time associated with acenocoumarol use. Model used is a GEE (Generalized Estimated Equations) multivariate logistic regression model including acenocoumarol use and adjusted for age, sex, BMI, smoking, time between baseline and follow-up visit, baseline OA severity in Kellgren-Lawrence score, joint modeled, lower limb disability, femoral neck BMD, HDL/total cholesterol ratio, physical activity, education level, hypertension, diabetes mellites and Rotterdam Study cohort. OR: odds ratio, CI: confidence interval, Progression: number of joints showing overall progression of either hip or knee joints or both. OR: odds ratio, CI: confidence interval, Progression: number of joints showing overall progression of either hip or knee joints or both. *Number of individual knee and/or hip joints studied from RSI and RSII, excluding joints at baseline with a KL-score =4, a total joint replacement or missing radiograph (n=1,262 joints).



▲ Figure 1: Acenocoumarol use interacts with MGP OA risk single nucleoti de variants (SNVs) and VKORC1 haplotype groups, leading to increased risk of overall progression of osteoarthriti s. VKORC1 haplotype groups are based on the VKORC1 H-haplotypes, which can be divided into three groups based on their VKA dosage/VKORC1 expression associati on: AA = low VKA dose/VKORC1 expression, AB=intermediate VKA dose/VKORC1 expression and BB = high VKA dose/VKORC1 expression. c) Aceno-coumarol use in VKORC1 BB haplotype carriers and MGP risk allele carriers and OA risk. OR, CI and p-value based on GEE multi variate logisti c regression adjusted for age, sex, BMI, smoking, ti me between baseline and follow-up visit, baseline OA severity in Kellgren-Lawrence score, joint modeled, femoral neck BMD, HDL/total cholesterol rati o, physical acti vity, educati on level, hypertension and diabetes mel-lites. OA: osteoarthriti s, A/A: non carrier of MGP risk allele, T/*: carrier of MGP risk allele. VKORC1 AA: homozygous VKORC1 A haplotype carriers, VKORC1 BB: homozygous VKORC1 BB haplotype carriers OR: odds rati o, error bars indicate 95% Confi dence Interval (CI) for the OR. P-values belong to depicted OR. See Supplementary Table 3 for the exact values depicted in this graph.

altered *VKORC1* expression had an altered risk for overall incidence and progression of OA. The study population was stratified in *MGP* risk allele carriers (T/*) and non-carriers (A/A)(**Table 1**).

Within acenocoumarol users among both *MGP* genotype groups had a significantly higher risk of overall OA progression than non-users (**Table 3**). Using the *VKORC1*-Haplotypes, the study population could be also stratified into low (AA), intermediate (AB), and high (BB) *VKORC1* expression/VKA dose groups[23](**Table 1** and **Supplementary Table S1**). Similarly to the *MGP* genotypes, we observed no effect of the *VCORC1* genotypes on the risk for overall OA incidence/progression (**Table 3**), although the *VKORC1*-BB group had the highest risk of overall OA progression (OR=3.35, 95% CI=2.22-5.05, **Table 3**).

As, individuals can be both carriers of *MGP* risk alleles and *VKROC1* haplotypes, we examined the combined effects of the *VKORC1*-BB haplotype and MGP -risk allele carriers (T/*). We stratified our study population into VKORC1 BB-haplotype or AA/AB carriers, which was then further stratified into carriers (T/*) and non-carriers (A/A) of the *MGP* risk allele, rs1800801 (**Figure 1** and **Supplementary Table 3**). Acenocoumarol users whom either were carriers of the *MGP* risk alleles (T/*) or carriers of the *VKORC1* BB-haplotype has a significant increased risk of osteoarthritis incidence and progression. Individuals whom were carriers of both the *VKORC1* BB-haplotype and the *MGP* risk allele had a fourfold increased risk of osteoarthritis incidence and progression (OR=4.18, 95%CI=2.69-6.50, **Figure 1**). Interestingly, *VKORC1* AA/AB-haplotype carriers who were not carriers of the *MGP*-risk allele (A/A) did not have a significant increased risk of overall OA progression/incidence when using acenocoumarol (OR=1.72, 95%CI=0.93-3.19, **Figure 1**).

Discussion

We demonstrated that use of the acenocoumarol was associated with a higher risk of overall OA incidence/progression in than non-users. We observed that the increased OA risk in acenocoumarol users varied based on genetic variants affecting the vitamin K cycle with VKA use. Acenocoumarol users with the MGP risk allele and VKORC1 BB-haplotype, had a fourfold higher risk for overall OA incidence/progression.

The *VKORC1* BB-haplotype is associated with a higher expression of *VKORC1*, associated with greater vitamin K activity; however, individuals who are *VKORC1* BB-haplotype carriers also require and receive higher dosages of VKA for the desired anticoagulation effect[23]. Intuitively, the anticoagulation, amount of *VKORC1* inhibition, should be similar in all users regardless of *VKORC1* haplotype since the amount of anticoagulation is the dosing measurement, not *VKORC1* expression. However, vitamin K availability levels differ significantly between tissues[25, 26] and warfarin affects vitamin K inhibition differently in liver compared to bone[27-29]. In liver, another enzyme in addition to *VKORC1*, is available for the recycling of vitamin K into its active form, NQO1(NAD(P)H quinone oxidoreductase 1). This enzyme is not present in bone tissue, causing bone tissue to be more susceptible to VKA dosages than liver tissue[28-30]. Thus, higher VKA dosages in *VKORC1* BB-haplotype carriers needed for desired inhibition of vitamin K-dependent blood coagulation proteins in the liver might be too high of a dosage for *VKORC1* functioning in the joint.

Oral VKAs, which includes acenocoumarol and warfarin, were the only oral anticoagulants available for decades[31]. New oral anticoagulant drugs developed over the past decade target thrombin (IIa) or factor X (Xa) instead of vitamin K, known as non-vitamin K inhibiting anticoagulants (NOACs) or direct oral anticoagulants (DOACs). Recent years have seen a rise in the use of NOACs[13], which have an improved efficacy-to-safety ratio over VKAs. Additionally, they do not need routine coagulation monitoring and have fewer food and drug interactions compared to VKAs; however NOACs are more costly and difficult to reverse[32-34]. Nonetheless, VKAs continue to be commonly prescribed, and are the only indicated anticoagulant class for certain indications (e.g., antiphospholipid antibody syndrome, mechanical heart valves). With aging of the population, the number of people with OA and requiring anti-coagulation medication will continue to rise. Given our findings of increased risk of OA incidence and progression with VKA and the lack of effective treatment options for OA, it may be reasonable to consider NOACs over VKAs for medical indications in which NOACs can be used. This may be particularly the case for those who are carriers of the *MGP* risk allele and/or the *VKORC1* BB-haplotype, an estimated ~30% of individuals of European ancestry.

The strengths of our study include the robust underlying biological hypothesis, large sample size and high quality prospectively collected data. However, some limitations of our study should be acknowledged. First, while we found similar results in RS-I and RS-II, analyses in other independent cohorts are warranted, specifically since RSII has a much smaller sample size and number of acenocoumarol users compared to RSI. Also replication in non-Central European (CEU) ancestry based cohorts is warranted. Second, the association we noted in this study may be due to a shared disease pathology between OA and VKA indications[35, 36]. We attempted to address this issue by adjusting for cardiovascular disease risk factors. However, this possible confounding by indication bias, need to be addressed more directly, by examining direct (new) oral anticoagulants (DOAC/NOAC) users as a comparator group, as these oral anticoagulants do not inhibit the vitamin K-cycle. Unfortunately, our study population contains too few DOAC/NOAC users for such an analysis (n=9). Last, as with all observational studies, we cannot rule out residual confounding.

In summary, we found an increased risk of OA progression in users of the VKA acenocoumarol, which was further increased in *VKORC1* BB-haplotype and *MGP* risk allele carriers. These findings are consistent with the known biology of MGP and vitamin K, and are in keeping with prior studies of vitamin K in OA. Taken together, these studies, including the current one, highlight the importance of vitamin K and vitamin K dependent Gla-proteins, including MGP, in the pathogenesis of OA. Given that there are

as yet no treatment options for preventing OA onset or progression, vitamin K may represent a modifiable risk factor. These data provide strong rationale for a properly powered randomized clinical trial of vitamin K in an appropriate patient population, such as those with insufficient vitamin K levels and/or MGP risk allele carriers. Additionally, these data lend support to the consideration of DOAC/NOACs in favour of VKAs when appropriate for a supported indication, and highlight the future possibility of genetic screening to identify individuals at high risk of OA incidence and progression.

Data Availability Statement

All relevant data supporting the key findings of this study are available within the article and its supplementary data. Due to ethical and legal restrictions, individual-level data of the Rotterdam Study (RS) cannot be made publicly available. Data are available upon request to the data manager of the Rotterdam Study Frank van Rooij (f.vanrooij@erasmusmc.nl) and subject to local rules and regulations. This includes submitting a proposal to the management team of RS, where upon approval, analysis needs to be done on a local server with protected access, complying with GDPR regulations

Acknowledgments

The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. The generation and management of GWAS genotype data for the Rotterdam Study (RS-I, RS-II, RS-III) was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters, MSc, and Carolina Medina-Gomez, MSc, for their help in creating the GWAS database, and Linda Broer PhD, for the creation of the imputed data.

Funding

This research was funded by the Dutch Arthritis Foundation (ReumaNederland), (project nr DAA: 16-1-404). 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 Rotterdam

Study GWAS datasets are supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project nr. 050-060-810. TN was supported by NIH K24 AR070892.

Author Contributions

CGB designed the study, performed the analyses, made the figures and tables, and wrote the manuscript. IS contributed to study design, NLN performed analysis, TN and IM contributed to study design, AGU, MAI provided access to the Rotterdam study data set, SB-Z contributed to study design. BS provided pharmacological data, contributed to study design and analysis, J.B.J.v.M. designed the study and supervised this work. All authors critically assessed the manuscript.

References

1. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet, 2018. 392(10159): p. 1789-1858.
2. Hunter, D.J. and S. Bierma-Zeinstra, Osteoarthritis. Lancet, 2019. 393(10182): p. 1745-1759.
3. Neogi, T., et al., Low vitamin K status is associated with osteoarthritis in the hand and knee. Arthritis Rheum, 2006. 54(4): p. 1255-61.
4. Misra, D., et al., Vitamin K deficiency is associated with incident knee osteoarthritis. Am J Med, 2013. 126(3): p. 243-8.
5. Shea, M.K., et al., The association between vitamin K status and knee osteoarthritis features in older adults: the Health, Aging and Body Composition Study. Osteoarthritis Cartilage, 2015. 23(3): p. 370-8.
6. Neogi, T., et al., Vitamin K in hand osteoarthritis: results from a randomised clinical trial. Ann Rheum Dis, 2008. 67(11): p. 1570-3.
7. Yin, T. and T. Miyata, Warfarin dose and the pharmacogenomics of CYP2C9 and VKORC1- rationale and perspectives. Thromb Res, 2007. 120(1): p. 1-10.
8. Luo, G., et al., Spontaneous calcification of arteries and cartilage in mice lacking matrix GLA protein. Nature, 1997. 386(6620): p. 78-81.
9. Newman, B., et al., Coordinated expression of matrix Gla protein is required during endochondral ossification for chondrocyte survival. J Cell Biol, 2001. 154(3): p. 659-66.
10. den Hollander, W., et al., Genome-wide association and functional studies identify a role for matrix Gla protein in osteoarthritis of the hand. Ann Rheum Dis, 2017. 76(12): p. 2046-2053.
11. Holbrook, A., et al., Evidence-based management of anticoagulant therapy: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest, 2012. 141(2 Suppl): p. e152S-e184S.
12. Krijthe, B.P., et al., Projections on the number of individuals with atrial fibrillation in the European Union, from 2000 to 2060. Eur Heart J, 2013. 34(35): p. 2746-51.
13. Zhu, J., et al., Trends and Variation in Oral Anticoagulant Choice in Patients with Atrial Fibrillation, 2010-2017. Pharmacotherapy, 2018. 38(9): p. 907-20.
14. Ikram, M.A., et al., Objectives, design and main findings until 2020 from the Rotterdam Study. Eur J Epidemiol, 2020. 35(5): p. 483-517.
15. Kellgren, J.H. and J.S. Lawrence, Radiological Assessment of Osteo-Arthrosis. Ann Rheum Dis, 1957. 16(4): p. 494-502.
16. Kerkhof, H.J., et al., Recommendations for standardization and phenotype definitions in genetic studies of osteoarthritis: the TREAT-OA consortium. Osteoarthritis Cartilage, 2011. 19(3): p. 254-64.
17. Odding, E., et al., Association of locomotor complaints and disability in the Rotterdam study. Ann Rheum Dis, 1995. 54(9): p. 721-5.
18. Clockaerts, S., et al., Statin use is associated with reduced incidence and progression of knee osteoarthritis in the Rotterdam study. Ann Rheum Dis, 2012. 71(5): p. 642-7.
19. Teichert, M., et al., A genome-wide association study of acenocoumarol maintenance dosage. Hum Mol Genet, 2009. 18(19): p. 3758-68.
20. Whirl-Carrillo, M., et al., Pharmacogenomics knowledge for personalized medicine. Clin Pharmacol Ther, 2012. 92(4): p. 414-7.
21. McCarthy, S., et al., A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet, 2016. 48(10): p. 1279-83.
22. Sinnwell JP, S.D., haplo.stats: Statistical Analysis of Haplotypes with Traits and Covariates when linkag pahs is Ambigous. 2018.

23. Rieder, M.J., et al., Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. N Engl J Med, 2005. 352(22): p. 2285-93.
24. Teichert, M., et al., Dependency of phenprocoumon dosage on polymorphisms in the VKORC1, CYP2C9, and CYP4F2 genes. Pharmacogenet Genomics, 2011. 21(1): p. 26-34.
25. Roncaglioni, M.C., et al., Warfarin-induced accumulation of vitamin K-dependent proteins. Comparison between hepatic and non-hepatic tissues. Biochem Biophys Res Commun, 1983. 114(3): p. 991-7.
26. Thijssen, H.H. and M.J. Drittij-Reijnders, Vitamin K status in human tissues: tissue-specific accumulation of phylloquinone and menaquinone-4. Br J Nutr, 1996. 75(1): p. 121-7.
27. Price, P.A. and Y. Kaneda, Vitamin K counteracts the effect of warfarin in liver but not in bone. Thromb Res, 1987. 46(1): p. 121-31.
28. Sato, T., et al., Difference in the metabolism of vitamin K between liver and bone in vitamin K-deficient rats. Br J Nutr, 2002. 87(4): p. 307-14.
29. Hara, K., M. Kobayashi, and Y. Akiyama, Comparison of inhibitory effects of warfarin on gamma-carboxylation between bone and liver in rats. J Bone Miner Metab, 2005. 23(5): p. 366-72.
30. Ulrich, M.M., et al., Vitamin K is no antagonist for the action of warfarin in rat osteosarcoma UMR 106. Thromb Res, 1988. 50(1): p. 27-32.
31. Gomez-Outes, A., et al., Discovery of anticoagulant drugs: a historical perspective. Curr Drug Discov Technol, 2012. 9(2): p. 83-104.
32. Mekaj, Y.H., et al., New oral anticoagulants: their advantages and disadvantages compared with vitamin K antagonists in the prevention and treatment of patients with thromboembolic events. Ther Clin Risk Manag, 2015. 11: p. 967-77.
33. Kim, I.S., et al., Non-vitamin K antagonist oral anticoagulants have better efficacy and equivalent safety compared to warfarin in elderly patients with atrial fibrillation: A systematic review and meta-analysis. J Cardiol, 2018. 72(2): p. 105-112.
34. Haas, S., et al., Predictors of NOAC versus VKA use for stroke prevention in patients with newly diagnosed atrial fibrillation: Results from GARFIELD-AF. Am Heart J, 2019. 213: p. 35-46.
35. Veronese, N., et al., Association of Osteoarthritis With Increased Risk of Cardiovascular Diseases in the Elderly: Findings From the Progetto Veneto Anziano Study Cohort. Arthritis Rheumatol, 2016. 68(5): p. 1136-44.
36. Wang, H., et al., Osteoarthritis and the risk of cardiovascular disease: a meta-analysis of observational studies. Sci Rep, 2016. 6: p. 39672.

Supplementary Materials

Study population and clinical data

The Rotterdam study (RS) is a large prospective population-based cohort study ongoing since 1990 to study determinants of chronic disabling diseases in the elderly[1]. All cohort participants live in the Ommoord district of the city of Rotterdam, the Netherlands. Residents 55 years and older were first recruited in 1990. In 2000, a second cohort, RS-II, was started with individuals who had become 55 years of age or moved into the study district since the start of the study. Follow-up data were collected as follow-up visits every ~4-5 years. Details of the design and rationale of the Rotterdam Study has been published elsewhere1. Participant measurements at baseline and follow-up were obtained during visits to the research center for physical examinations, computerized pharmacy records, and from home interviews. Our study included participants of RS-I and RS-II for whom radiographs of knee and hip joints at baseline and follow-up visit were present obtained and scored. Additional information collected included: sex, age at baseline visit, body mass index (BMI, kg-1m2), physical activity (metabolic equivalent of task/hours per week), smoking (never, former and current smoker), education level (UNESCO education classification), Diabetes mellitus, hypertension, femoral neck bone mineral density (FN-BMD), HDL/total cholesterol ratio and Stanford Health Assessment Questionnaire. The Medical Ethics Committee of the Erasmus MC, University Medical Center, has approved of the Rotterdam Study (MEC 02.1015). All participants provided written consent prior to participation in the Rotterdam Study.

Computerized pharmacy data of VKA and antiplatelet use

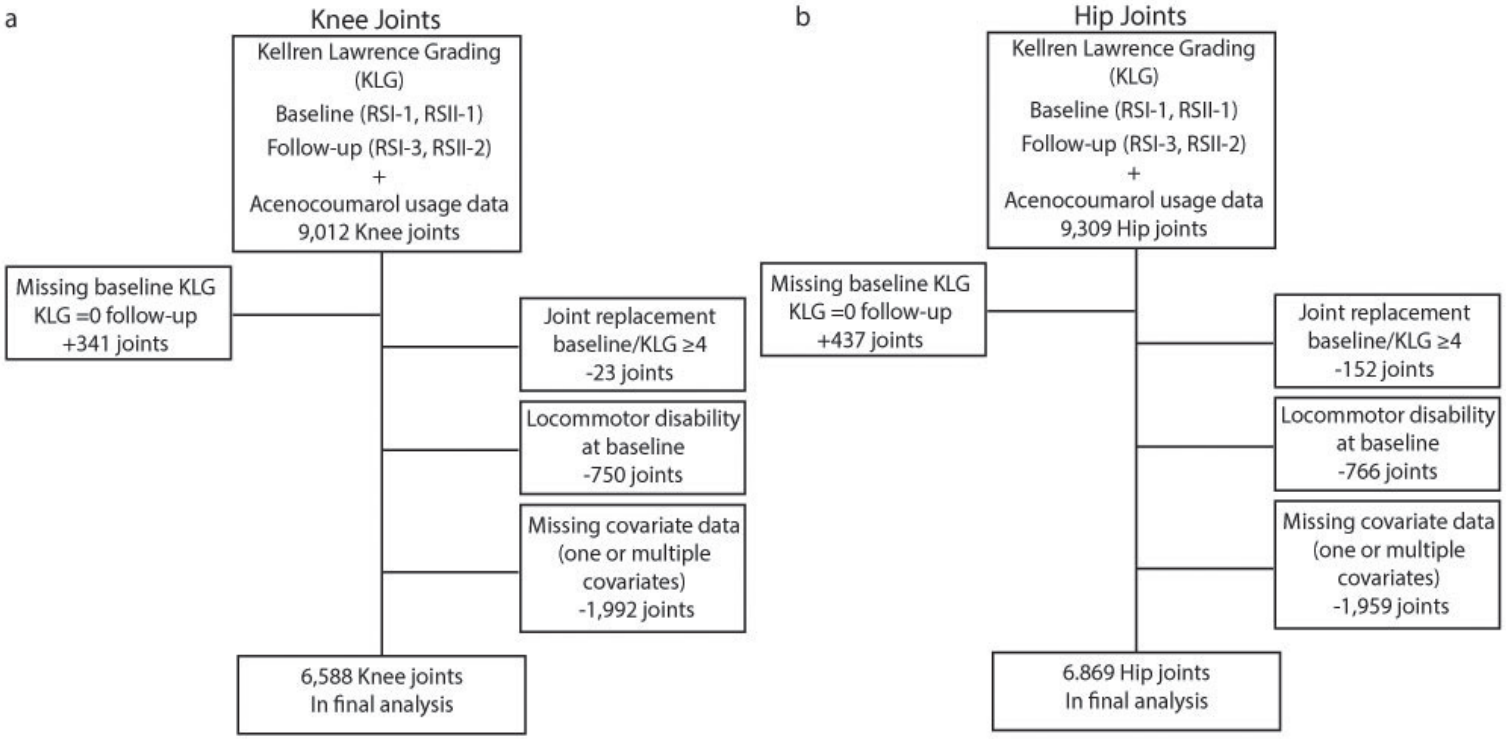
For 98% of the participants of the RS-I and RS-II, medication usage was available through fully computerized pharmacies within the Ommoord district. All data on dispensed drugs were available in computerized form on a daily basis. Information was available on the date of prescribing, the total amount of drug units for each prescription, the prescribed daily number of units, the product name and the therapeutic chemical code1. For each participant, we extracted the usage of VKAs(acenocoumarol) and antiplatelet agents (acetylsalicylic acid, carbasalate calcium and Clopidogrel), for the period between baseline and follow-up. We extracted information on the prescription of the VKA acenocoumarol. Warfarin is not prescribed, as it is not registered for use as a drug in the Netherlands. All patients taking VKAs in this study attended an anticoagulation clinic, which is standard practice in the Netherlands [2].

Detailed description of Potential confounders and risk factors

Potential confounders and risk factors included age, sex, body Mass Index(BMI) smoking status, physical activity, total cholesterol/high-density lipoprotein (HDL) ratio, education level, baseline OA severity, lower-limb disability, femoral neck Bone Mass Density (BMD), cohort(RS-I/RS-II), and comorbidities (hypertension and diabetes mellitus) as potential confounders/risk factors.

Age (at baseline visit in years), sex, BMI(kg-1m2), smoking status (never, former and current smoker), physical activity(metabolic equivalent of task hours/week), education level (0-3) and lower-limb disability(yes/no) were assessed through home interviews taken by trained interviewers1. Education levels were assessed according to UNESCO classification to define education level (0=primary education, 1 = lower/ intermediate general education or lower vocational education, 2= intermediate vocational education or higher general education, 3= higher vocational education or university).Lower-limb disability was determined for each participant based on the Stanford Health Assessment Questionnaire3. Locomotor disability of the lower-limb was defined as the mean of the scores (with zero indicating no impairment and three indicating unable to perform) on the six questions related to lower limb functions[3]. Disability was defined as a lower limb disability index of 0.5 or over.

Baseline OA severity was determined scoring of radiographs taken at baseline visit. Radiographs were scored by trained medical professional according to the Kellgren-Lawrence osteoarthrtis severity grading scale4. Femoral neck BMD was measured by DXA-scans as described previously5. Arterial hypertension was defined as a systolic blood pressure of 160 mm Hg or higher, diastolic blood pressure of 95 mm Hg or higher, or the use of antihypertensive drugs for hypertension. Diabetes mellitus was defined as use of glucose-lowering medication or non-fasting random or postload glucose levels exceeding 11.0 mmol/liter. Total cholesterol/HDL ration was determined via serum total cholesterol determined by an enzymatic procedure, HDL measured after precipitation of non-HDL cholesterol.



▲ **Figure S1: Flow-chart of included knee and hip joints of the Rotterdam Study** a) Flow-chart for the number of knee joints included in the analysis from the Rotterdam study. b) Flow-chart for the number of hip joints included in the analysis from the Rotterdam study. RS: Rotterdam Study, RS-I: Rotterdam study subcohort I, RS-II: Rotterdam Study subcohort II, KLG: Kellgren-Lawrence Grading.

Table S1: Distribution of VKORC1 H-haplotype in the Rotterdam Study Cohorts

VKORC1 Haplotype	Haplotype Sequence*	RS-I (*n = 12,582*)†	RS-II (*n = 4,314*)†	Total Frequency
Haplotype distribution		*Proportion (number of haplotypes)*		
H1	GGCTAGAGAC	0.13 (1,655)	0.14 (597)	0.13
H2	GGCTCGAGAC	0.24 (3,012)	0.23 (980)	0.24
H3	GGCCAGGGGC	<0.01 (66)	<0.01 (17)	$4.9*10^{-3}$
H4	GGCCAGGCAC	<0.01 (8)	<0.01 (5)	$7.1*10^{-4}$
H5	AGCTCGAGAC	0.02 (221)	0.02 (66)	0.02
H6	AGCCAGGCGC	<0.01 (8)	<0.01 (2)	$5.9*10^{-4}$
H7	AGCCAGGCGT	0.26 (3,238)	0.25 (1,097)	0.26
H8	ATCCAGGCGT	0.12 (1,498)	0.11 (490)	0.12
H9	ATGCAAGCGC	0.22 (2,818)	0.24 (1,022)	0.23
H10‡	GGGCAAGCGC	<0.01 (47)	<0.01 (22)	$4.1*10^{-3}$
Other Haplotypes§	-	<0.01 (11)	<0.01 (17)	$1.7*10^{-3}$
Group distribution				
(Low VKORC1) Group A	H1, H2	0.37 (4,667)	0.37 (1,577)	0.37
(High VKORC1) Group B	H7, H8, H9	0.60 (7,554)	0.60 (2,609)	0.60

* for each haplotype the ten investigated single nucleotide variants (SNVs) are listed in sequential order along the ***VKORC1*** gene, alleles are coded to the plus-strand, and match the Haplotypes reported on PHARMGKB: rs7196161, rs17880887, rs17881535, rs9923231, rs2884737, rs17708472, rs9934438, rs8050894, s2359612 and rs7294.

† Haplotypes have been calculated on all individuals of whom genotype information was available, not only individuals included in our acenocoumarol study population. The total number of haplotypes is twice the number of individuals examined.

‡ The H10 ***VKORC1*** haplotype has not been described previously[6].

§ Other haplotypes consist of another 16 haplotypes, occurring on average only once in either RS-I or RS-II.

RS: Rotterdam Study, ***VKORC1***: Vitamin K Epoxide Reductase Complex 1, H: haplotype

Table S2: Association between acenocoumarol use and risk of overall OA progression in RSI and RSII separately

	RS-I: Overall progression of Osteoarthritis					RS-II: Overall progression of Osteoarthritis				
Acenocoumarol usage	**Joints N***	**Progression N (%)**	**OR**	**95% CI**	**p-value**	**Joints N***	**Progression N (%)**	**OR**	**95% CI**	**p-value**
Non-users	10,196	500 (4.9%)	1	-	-	3,668	117 (3.2%)	1	-	-
Users	938	119 (12.7%)	2.33	1.85 - 294	$1.1*10^{-12}$	171	23 (13.5%)	3.42	1.96 - 6.02	$1.6*10^{-05}$
≤ *180 days*	305	43 (14.1%)	2.64	1.84 - 3.78	$1.3*10^{-07}$	52	7 (13.5%)	4.54	1.88 - 10.99	$7.9*10^{-04}$
≤ *556 days*	305	47 (15.4%)	2.56	1.80 - 3.65	$1.9*10^{-07}$	64	11 (17.2%)	5.01	2.42 - 10.73	$1.8*10^{-05}$
>556 days	328	29 (8.8%)	1.74	1.14 - 2.66	$1.1*10^{-02}$	55	5 (9.1%)	1.26	0.42 - 3.72	$6.8*10^{-01}$
	RS-I: Overall progression of Knee Osteoarthritis					**RS-II: Overall progression of Knee Osteoarthritis**				
Non-users	4.953	319 (6.4%)	1	-	-	1,749	82 (4.7%)	1	-	-
Users	474	65 (13.7%)	1.99	1.46 - 2.72	$1.3*10^{-05}$	80	9 (11.3%)	1.95	0.87 - 4.37	$1.1*10^{-01}$
≤ *180 days*	159	21 (13.2%)	1.87	1.15 - 3.04	$1.1*10^{-02}$	25	2 (8.0%)	1.89	0.42 - 8.54	$4.1*10^{-01}$
≤ *556 days*	151	23 (15.2%)	2.00	1.22 - 3.30	$5.9*10^{-03}$	30	3 (10.0%)	1.96	0.56 - 6.88	$3.0*10^{-01}$
>556 days	164	21 (12.8%)	2.09	1.23 - 3.57	$6.6*10^{-03}$	25	4 (16.0%)	1.72	0.53 - 5.62	$3.7*10^{-01}$
	RS-I: Overall progression of Hip Osteoarthritis					**RS-II: Overall progression of Hip Osteoarthritis**				
Non-users	5,243	181 (3.4%)	1	-	-	1,837	35 (1.9%)	1	-	-
Users	464	54 (11.6%)	2.81	1.95 - 4.06	$3.3*10^{-08}$	91	14 (15.4%)	7.25	2.91 - 18.05	$2.1*10^{-05}$
≤ *180 days*	146	22 (15.1%)	4.25	2.47 - 7.30	$1.6*10^{-07}$	27	5 (18.5%)	12.00	3.68 - 39.15	$3.8*10^{-05}$
≤ *556 days*	154	24 (15.6%)	3.33	1.91 - 5.81	$2.2*10^{-05}$	34	8 (23.5^)	13.09	3.61 - 47.42	$9.0*10^{-05}$
>556 days	164	8 (4.9%)	1.15	0.54 - 2.44	$7.2*10^{-01}$	30	1 (3.3%)	1.23	0.14 - 11.07	$8.5*10^{-01}$

Overall progression of Osteoarthritis(OA) in RS-I and RS-II within the follow-up time associated with acenocoumarol use. Model used is a GEE (Generalized Estimated Equations) multivariate logistic regression model including acenocoumarol use and adjusted for age, sex, BMI, smoking, time between baseline and follow-up visit, baseline OA severity in Kellgren-Lawrence score, joint modeled, lower limb disability, femoral neck BMD, HDL/total cholesterol ratio, physical activity, education level, hypertension and diabetes mellites. OR: odds ratio, CI: confidence interval, Progression: number of joints showing overall progression of either hip or knee joints or both.
*Number of individual knee and/or hip joints studied from RS-I and RS-II.

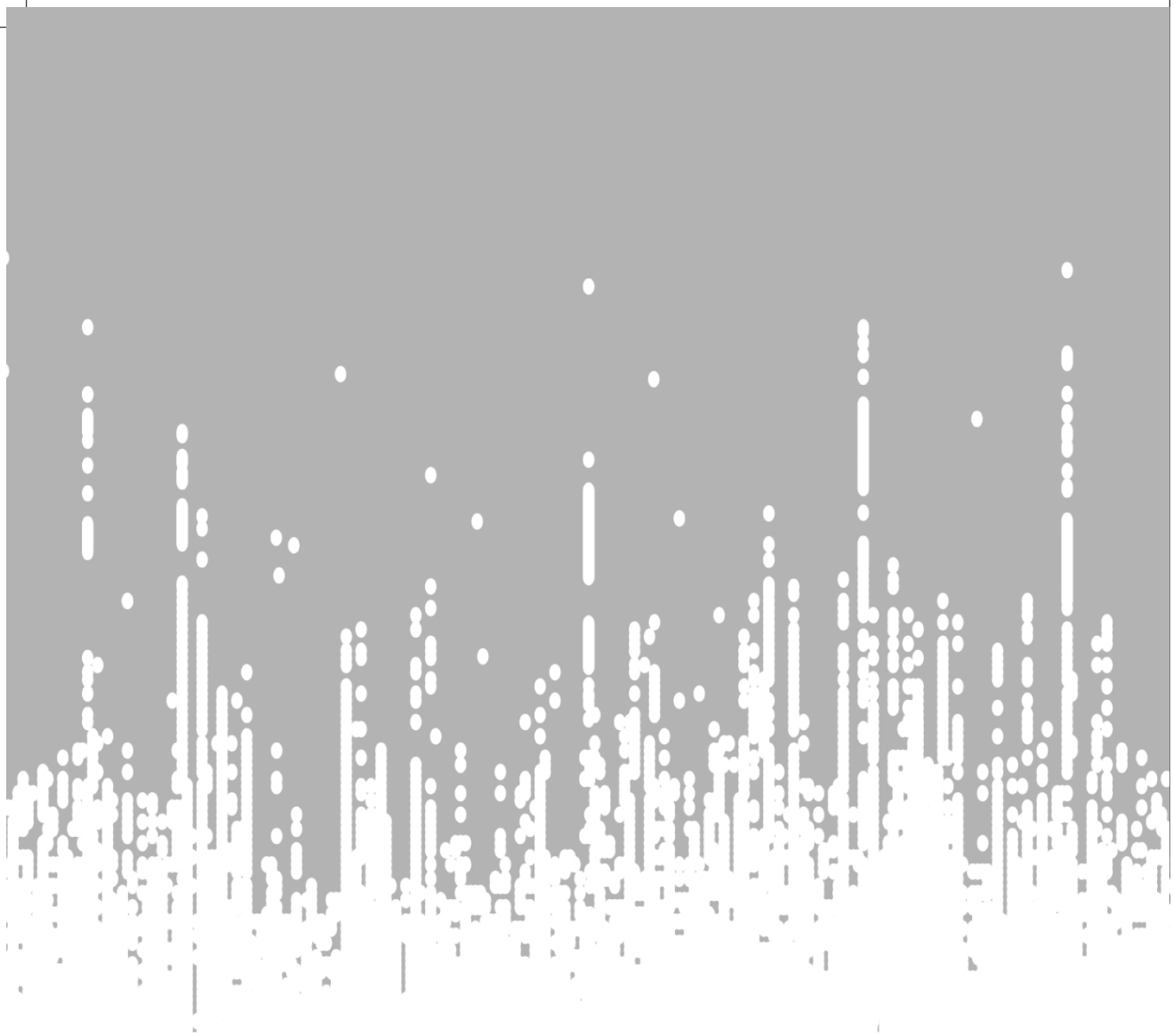
Supplementary Table S3: Association between acenocoumarol and VKORC1 haplotypes and MGP genotypes

Acenocoumarol use	Joints*	Progression	OR	95% CI	P-value
High VKORC1 haplotype and MGP risk allele status					
Non users VKORC1 AA/AB and MGP A/A	2,745	118 (4.3%)	1	-	-
Non users VKORC1 AA/AB and MGP T/*	4,515	160 (3.5%)	0.85	0.67-1.09	0.19
Non users VKORC1 BB and MGP A/A	1,780	87 (4.9%)	1.15	0.86-1.54	0.38
Non users VKORC1 BB and MGP T/*	2,668	114 (4.3%)	1.01	0.78-1.33	0.89
Users VKORC1 AA/AB and MGP A/A	165	13 (7.8%)	1.72	0.93-3.19	$8.5x10^{-02}$
*Users VKORC1 AA/AB and MGP T/**	304	27 (8.9%)	1.96	1.21-3.00	$5.4x10^{-03}$
Users VKORC1 BB and MGP A/A	117	11 (9.4%)	2.35	1.21-4.56	$1.1x10^{-02}$
Users VKORC1 BB and MGP T/*	188	31 (16.5%)	4.18	2.69-6.50	$1.8x10^{-10}$

Overall progression of Osteoarthritis(OA) in RS-I and RS-II within the follow-up time associated with acenocoumarol use. Model used is a GEE (Generalized Estimated Equations) multivariate logistic regression model including acenocoumarol use and adjusted for age, sex, BMI, smoking, time between baseline and follow-up visit, baseline OA severity in Kellgren-Lawrence score, joint modeled, femoral neck BMD, HDL/total cholesterol ratio, physical activity, education level, hypertension, diabetes mellites and Rotterdam Study Cohort. OR: odds ratio, CI: confidence interval, Progression: number of joints showing overall OA progression. *Number of individual Knee and Hip Joints studied from RSI and RSII, excluding all joints of individuals with lower limb disability (n = 514).

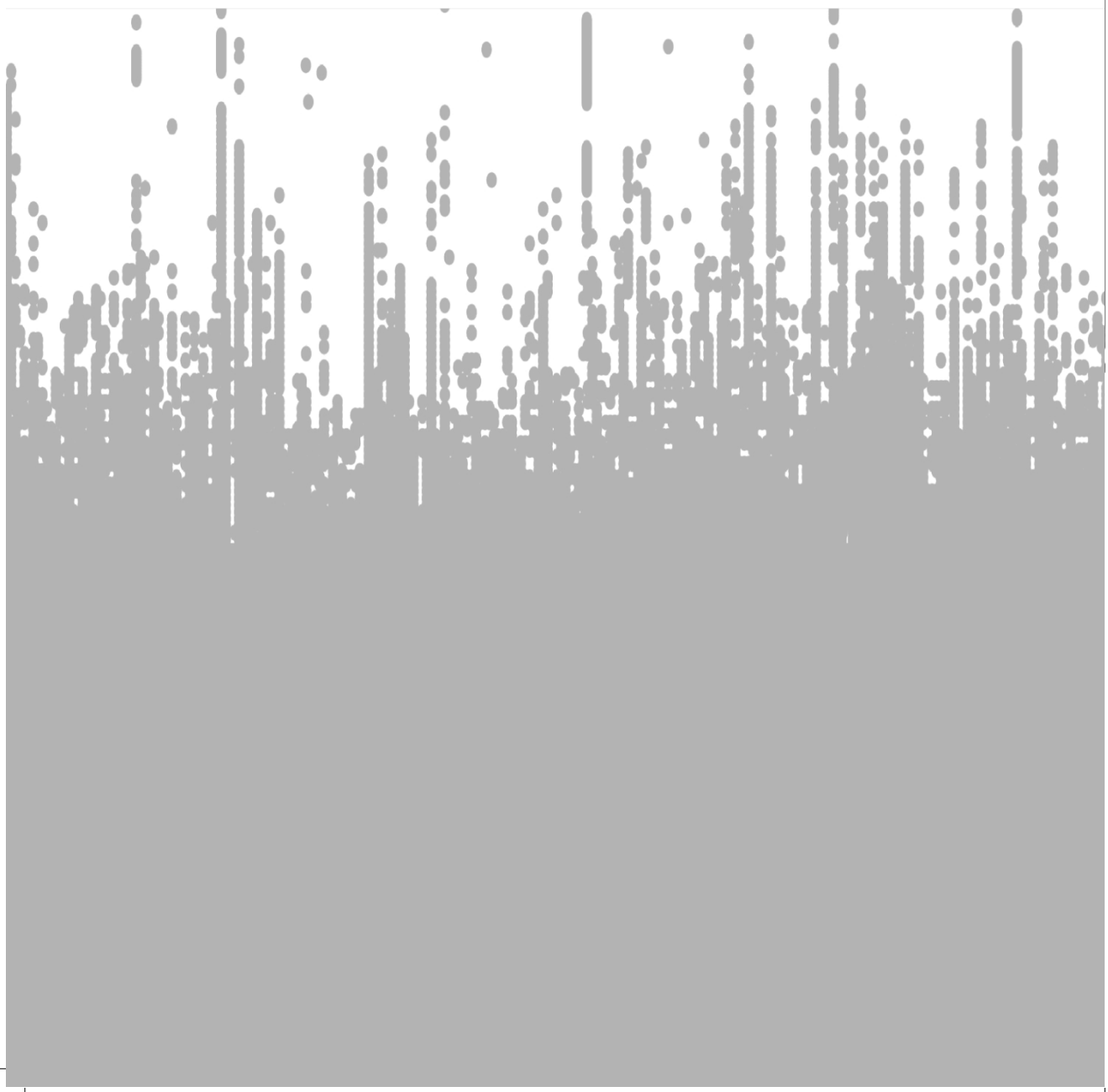
Supplementary References

1. Ikram, M.A., et al., Objectives, design and main findings until 2020 from the Rotterdam Study. Eur J Epidemiol, 2020. 35(5): p. 483-517.
2. Teichert, M., et al., A genome-wide association study of acenocoumarol maintenance dosage. Hum Mol Genet, 2009. 18(19): p. 3758-68.
3. Fries, J.F., P.W. Spitz, and D.Y. Young, The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. J Rheumatol, 1982. 9(5): p. 789-93.
4. Odding, E., et al., Association of locomotor complaints and disability in the Rotterdam study. Ann Rheum Dis, 1995. 54(9): p. 721-5.
5. Burger, H., et al., The association between age and bone mineral density in men and women aged 55 years and over: the Rotterdam Study. Bone Miner, 1994. 25(1): p. 1-13.
6. Rieder, M.J., et al., Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. N Engl J Med, 2005. 352(22): p. 2285-93.



Chapter 4

LARGE SCALE OSTEOARTHRITIS GENETICS



Chapter 4.1

Deciphering Osteoarthritis Genetics Across 826,690 Individuals from 9 Populations

Boer, C.G., Hatzikotoulas K., Southam L.,, Stefánsdóttir L., Zhang Y., Coutinho de Almeida R., Zheng, J., Barysenka A., [Genetics of Osteoarthritis Consortium], Meulenbelt I., Styrkársdóttir, U, Lee M.T.M., van Meurs J.B.J. & Zeggini E.

Manuscript under consideration in Cell

Abstract

Osteoarthritis is the most prevalent musculoskeletal disease with over 300 million people affected worldwide. Here, we conduct a genome-wide association study meta-analysis for osteoarthritis across 826,690 individuals, and identify 55 newly associated risk variants, including 3 female-specific signals, bringing the total number to 148. We identify the first thumb and spine osteoarthritis risk variants, sex-specific risk loci, and differences between weight-bearing and non-weight bearing joints. By integrating functional genomics data from primary patient tissues, we identify likely causal genes for osteoarthritis loci. We find strong evidence for genetic correlation with phenotypes related to pain, the main disease symptom, and identify circulating protein levels that are on the causal path for osteoarthritis. Our results provide new insights into disease aetiopathogenesis and highlight attractive drug targets to accelerate translation.

Introduction

Osteoarthritis is one of the leading causes of disability and pain worldwide, with over 300 million people affected[1]. As one of the most rapidly-rising conditions globally, the societal burden of osteoarthritis is enormous[1, 2], and is accompanied by substantial multimorbidity[3]. Currently no curative treatments are available, and management strategies focus on symptom alleviation through pain relief and joint replacement surgery. A detailed understanding of disease aetiopathology and novel drug targets are therefore urgently needed and eagerly anticipated.

Osteoarthritis is a complex degenerative disease of the whole joint, characterised by cartilage degeneration, subchondral bone thickening, osteophyte formation, synovial inflammation and structural alterations the joint capsule, ligaments and associated muscles[4]. Risk for developing osteoarthritis consists of a complex interplay between environmental factors and underlying genetic variation[5]. Recently, major advances were made in elucidating the genetic background of osteoarthritis, using genome-wide association studies (GWAS)[6-8], with 96 risk single nucleotide variants (SNVs) established to date. However, these SNVs only explain a small proportion of the phenotypic variance[8] and are mainly associated with osteoarthritis affecting the knee and hip joints.

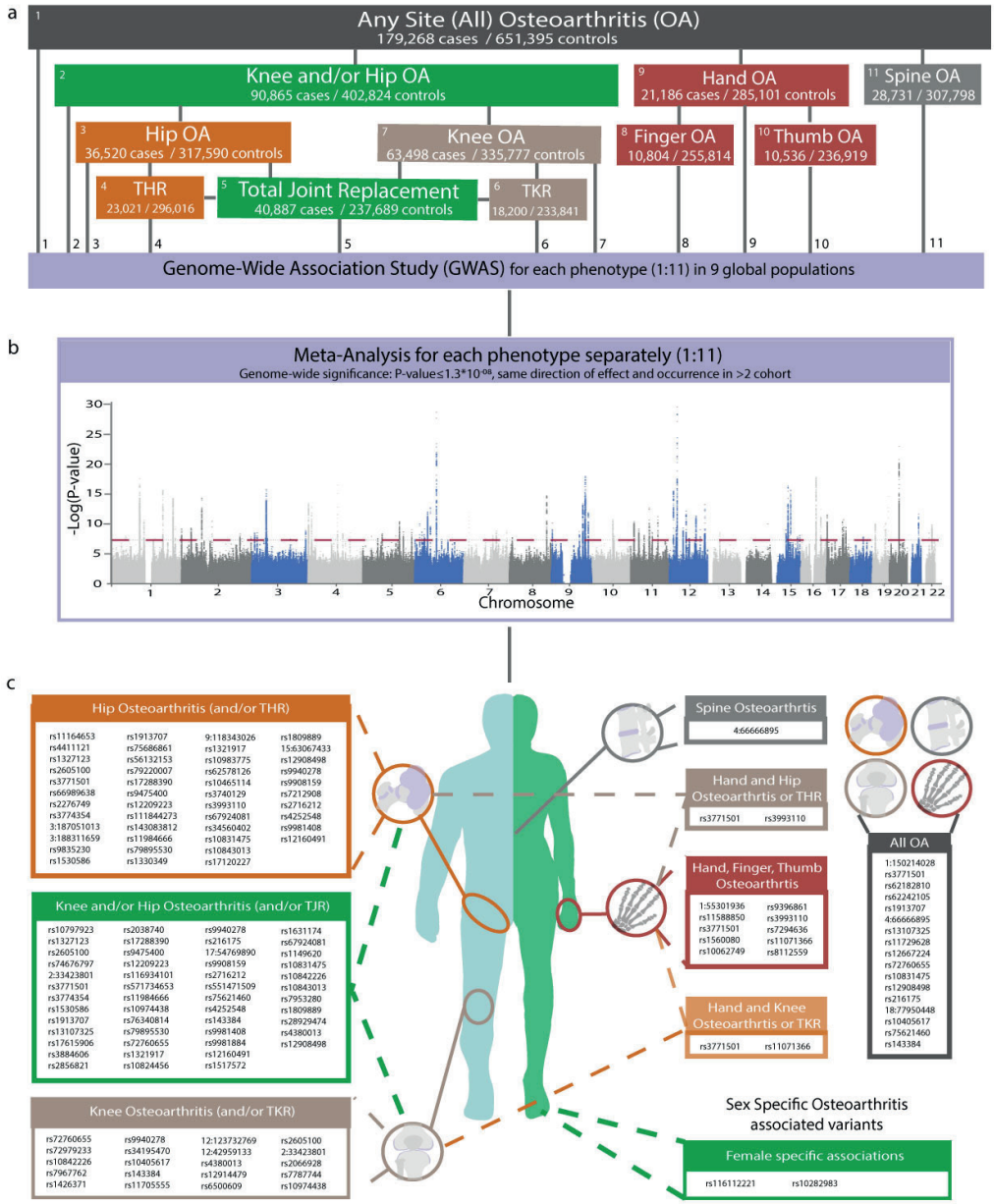
Osteoarthritis can affect every synovial joint and few general osteoarthritis SNVs have been identified to date. An increase in body mass index (BMI) is associated with risk of disease. A better understanding of the genetic differences between weight-bearing (knee, hip and spine) and non-weight-bearing joints (hand, finger, thumb) is needed to help disentangle the metabolic and biomechanical effects in osteoarthritis. Here, we conducted a GWAS meta-analysis across knee, hip, finger, thumb and spine osteoarthritis phenotypes in 826,690 individuals of European and East Asian descent. We integrated functional genomics analyses from disease-relevant tissue, including gene expression and protein abundance, coupled to mouse knockout phenotyping data and complementary computational fine-mapping, colocalization and causal inference approaches, to identify likely effector genes and facilitate much-needed translation into therapies by enhancing our understanding of disease aetiopathology.

Results

Genetic Architecture - Identification of osteoarthritis risk variants

We performed genome-wide meta-analyses for osteoarthritis across 13 international cohorts stemming from 9 populations (**Supplementary Table 1**), in up to 826,690 individuals (177,517 osteoarthritis patients). Two of the cohorts are of East Asian, and 11 of European descent. We defined 11 phenotypes encompassing all major sites for osteoarthritis at both weight-bearing and non weight-bearing joints (**Supplementary Table 1 and Supplementary Table 2**). We identified 223 genome-wide significantly associated SNVs across osteoarthritis phenotypes using a threshold of p-value<$1.3x10^{-08}$, to account for the effective number of independent tests (**Figure 1**, **Table 1**). Eighty-four of these variants have not been associated with osteoarthritis before. Using conditional analyses, we identified 100 unique and independent signals across osteoarthritis phenotypes, 52 of which have not been associated with an osteoarthritis phenotype before (**Table 2 and Supplementary Table 3**). For each of the 100 signals we define the lead SNV as the risk variant with the strongest statistical evidence for association. There was no evidence for multiple independently-associated SNVs at any signal. All lead SNVs have a minimum minor allele frequency (MAF) ≥0.01% in each cohort, and demonstrate the same direction of effect in two or more contributing studies. Six lead SNVs are coding (all missense), 59 reside within a gene transcript and 35 are intergenic.

We report here the first signals for spine (n=1) and thumb (n=2) osteoarthritis, and increase the number of SNVs for hand (five new, three known) and finger (three new, two known) osteoarthritis, as this is the first large-scale GWAS effort for these phenotypes. Of the 100 independent genome-wide significant SNVs, 90 are common (MAF ≥5%) and four are low-frequency (MAF <5% and ≥0.5%). We detected six rare variant associations (MAF 0.03% to 0.11%) with large effect sizes (odds ratio range 3.03 to 9.52) (**Table 2** and **Supplementary Table 2**); one previously reported and five new findings. The new rare variant associations all arise in a large extended family in Iceland, with some branches heavily burdened with osteoarthritis (**Supplementary Information**). The previously-reported hip osteoarthritis rs143083812 locus is a mis-sense variant located in the smoothened, frizzled family receptor (SMO) gene, for which we found association with total hip replacement (p-value=$1.11x10^{-11}$, OR=3.30, 95% CI=2.34-4.66)[7]. All of the new rare variant associations reside in non-coding sequence and are driven by the Icelandic cohort, in which imputation accuracy is excellent, with support from either the UK Biobank or Estonian cohorts.



▲ **Figure 1**: **Graphical summary of the Genetics of Osteoarthritis Consortium workflow and results.** a) Overview of the 11 defined osteoarthritis phenotypes, their relationship with each other and their sample sizes (cases/controls). TKR: total Knee Replacement, THR: total Hip Replacement. b) merged Manhattan-plot of all individual meta-analysis results of all 11 examined osteoarthritis phenotypes. Individual Manhattan plots and QQ-plots can be found in the Supplementary Figure 1. The dashed line represents the genome-wide significance threshold p-value≤1.3x10-8. c) Graphical overview of all genome-significant independent osteoarthritis associated single nucleotide variants (SNVs) and the osteoarthritis phenotypes with which they are associated.

Table 1: Summary of the Go consortium GWAS results

Genome-wide association study	Cases / Controls	Signals†	Novel Signals†	Known Signals†
All osteoarthritis*	177,517/649,173	21	8	13
Knee and/or hip osteoarthritis	89,741/400,604	31	9	22
Hip osteoarthritis	36,445/316,943	45	17	28
Knee osteoarthritis	62,497/333,557	24	11	13
Total hip replacement (THR)	23,021/296,016	38	12	26
Total Knee replacement (TKR)	18,200/233,841	10	4	6
Total joint replacement (TJR)	40,887/327,689	37	12	25
Hand osteoarthritis	20,901/282,881	7	5	2
Finger osteoarthritis	10,804/255,814	5	3	2
	10,536/236,919	4	2	2
Thumb osteoarthritis	28,3721/3057,578	1	1	0
Spine osteoarthritis	177,517/649,173	21	8	13
Independent signals§	100			
Independent novel OA signals§	52			
Independent known OA signals§	48			

* Cases are defined as any-site Osteoarthritis: hip, knee, hand, finger, thumb and spine.
† Signals numbers presented are based per defined OA phenotype, novel/known are thus also based per phenotype, Number represent only genome-wide significant signals p-value≤$1.3x10^{-08}$
‡ Known OA loci defined as known OA signal (regardless of discovery OA phenotype) present within 1Mb upstream or downstream of the identified GWAS signal.
§ Independence calculated within and across OA phenotypes.

Among the 96 previously-reported osteoarthritis-associated SNVs, 91 showed association at nominal significance (p-value≤0.05), and 57 reached genome-wide significance (p-value≤$5x10^{-08}$) in at least one osteoarthritis phenotype (**Supplementary Table 2**). As a sensitivity analysis, we excluded the UK Biobank dataset (up to 68,621 osteoarthritis cases and 247,846 controls) from the meta-analyses, and found that 74 of the lead SNVs continue to demonstrate association at nominal significance (18 at p-value≤$5x10^{-08}$) (**Supplementary Table 3**). Signals from 4 osteoarthritis phenotypes (spine, knee, knee and/or hip, and osteoarthritis at any site) included individuals of non-European ancestry (between 0.9 to 2.8% of cases were of East Asian descent). Even though sample sizes in the East Asian cohorts are small, we observed that 62% of the signals have supportive evidence in East Asian ancestry-only meta-analysis, with the same direction of effect, and that 20% of these signals are also nominally significant (p-value≤0.05) (Signals across European and non-European ancestry, **Supplementary Information**).

Table 2: Summary statistics of all Genome-Wide Significant Signals in the GO-Consortium

Phenotype	Other Phenotypes	SNV	Chr:pos	EA	NEA	EAF	OR	95% CI	P-value
New Osteoarthrit SNVs									
FingerOA		rs11588154	1:55301936	T	G	0.17	0.8	0.79 - 0.88	$6.08x10^{-10}$
HipOA	THR	rs4411121	1:118757034	T	C	0.31	1.1	1.05 - 1.09	$2.16x10^{-11}$
THR	ipOA, TJR	rs1327123	1:184014593	C	G	0.35	0.9	0.89 - 0.93	$2.44x10^{-16}$
ThumbOA		rs11588850	1:227927242	A	G	0.82	0.9	0.84 - 0.91	$3.53x10^{-10}$
KneeHipOA	KneeOA	rs74676797	2:633063	A	G	0.82	1.1	1.03 - 1.07	$6.39x10^{-10}$
THR	HipOA	rs66989638	2:106689736	A	G	0.13	1.1	1.08 - 1.15	$3.31x10^{-11}$
THR		rs2276749	3:11643465	T	C	0.05	0.9	0.82 - 0.90	$3.34x10^{-09}$
AllOA		rs62242105	3:20630395	A	G	0.33	1	0.96 - 0.98	$2.93x10^{-09}$
HipOA		rs781661531	3:187051013	T	C	$7x10^{-4}$	0.1	0.05 - 0.21	$8.36x10^{-11}$
HipOA		rs747952496	3:188311659	A	G	$4x10^{-4}$	7	3.93 - 12.55	$4.91x10^{-11}$
HipOA		rs9835230	3:189735461	A	G	0.24	1.1	1.04 - 1.09	$1.34x10^{-09}$
AllOA	SpineOA	rs201194999	4:66666895	T	C	0.3	0.9	0.85 - 0.92	$3.05x10^{-09}$
AllOA		rs11729628	4:121584282	T	G	0.24	1	0.96 - 0.98	$4.74x10^{-09}$
THR		rs75686861	4:145621328	A	G	0.09	1.1	1.08 - 1.16	$3.04x10^{-09}$
KneeOA		rs2066928	5:30843787	A	G	0.48	1	0.95 - 0.97	$1.20x10^{-08}$
THR	HipOA	rs56132153	5:67825133	A	C	0.61	1.1	1.05 - 1.09	$3.80x10^{-09}$
HandOA		rs1560080	5:115338732	A	G	0.83	0.9	0.88 - 0.94	$9.61x10^{-09}$
KneeHipOA	TJR, AllOA, HipOA, THR	rs17615906	5:128018413	T	C	0.84	1	0.93 - 0.96	$3.76x10^{-11}$
HandOA	ThumbOA, KneeOA	rs10062749	5:141805088	T	G	0.27	1.1	1.06 - 1.11	$2.04x10^{-09}$
FingerOA	HandOA	rs9396861	6:18404133	A	C	0.61	1.1	1.09 - 1.17	$9.35x10^{-11}$
TJR		rs2038740	6:35114542	T	C	0.72	0.9	0.93 - 0.96	$6.20x10^{-10}$
TJR		rs116934101	7:101775597	A	G	0.27	1.1	1.04 - 1.08	$7.12x10^{-09}$
AllOA		rs12667224	7:114024316	A	G	0.52	1	0.96 - 0.98	$1.66x10^{-09}$
KneeHipOA		rs571734653	7:137143697	A	C	$3x10^{-4}$	6	3.30 - 11.03	$5.56x10^{-09}$
TKR		rs7787744	7:150521096	A	G	0.67	1.1	1.05 - 1.11	$1.29x10^{-09}$
TJR		rs76340814	9:98321412	A	G	0.05	0.9	0.86 - 0.92	$1.87x10^{-09}$
THR	HipOA, TJR, KneeHipOA	rs79895530	9:110416422	T	C	0.13	0.9	0.85 - 0.91	$3.86x10^{-14}$
HipOA		rs7862601	9:118343026	A	G	0.62	0.9	0.92 - 0.96	$6.19x10^{-09}$
HipOA		rs10983775	9:120521100	T	C	0.54	1	0.93 - 0.97	$4.65x10^{-09}$
HipOA		rs10465114	9:129917824	A	G	0.22	1.1	1.04 - 1.09	$9.04x10^{-09}$
THR	HipOA	rs3740129	10:73767859	A	G	0.46	1.1	1.05 - 1.10	$1.70x10^{-11}$
TJR		rs10824456	10:78615458	C	G	0.58	1	0.94 - 0.97	$1.16x10^{-08}$
HandOA	THR	rs3993110	11:12794530	A	C	0.61	1.1	1.06 - 1.11	$3.75x10^{-11}$
KneeHipOA		rs1631174	11:47974373	A	C	0.34	1	1.03 - 1.05	$7.28x10^{-09}$
TKR	KneeOA	rs72979233	11:74355523	A	G	0.75	0.9	0.89 - 0.95	$2.52x10^{-09}$

Table 2 (continued): Summary statistics of all Genome-Wide Significant Signals in the GO-Consortium

Phenotype	Other Phenotypes	SNV	Chr:pos	EA	NEA	EAF	OR	95% CI	P-value
TJR	AllOA, KneeHipOA, HipOA, THR	rs10831475	11:95796907	A	G	0.81	1.1	1.05 - 1.10	$5.89x10^{-12}$
KneeHipOA	KneeOA, TKR	rs10842226	12:23959589	A	G	0.42	1	1.03 - 1.06	$4.68x10^{-10}$
TKR	KneeOA	rs7967762	12:48420214	T	C	0.16	1.1	1.07 - 1.15	$4.41x10^{-10}$
KneeOA		rs1426371	12:108629780	A	G	0.27	1	0.93 - 0.97	$8.86x10^{-10}$
KneeOA		rs58973023	13:42959133	A	T	0.49	1.1	1.04 - 1.08	$4.72x10^{-10}$
TJR		rs28929474	14:94844947	T	C	0.02	0.8	0.76 - 0.86	$1.06x10^{-10}$
THR	HipOA	rs746239049	15:63067433	D	I	0.21	0.9	0.87 - 0.93	$8.19x10^{-12}$
KneeOA		rs12914479	15:99174828	C	G	0.66	1	1.03 - 1.06	$7.12x10^{-09}$
KneeOA		rs6500609	16:4515334	C	G	0.11	0.9	0.91 - 0.96	$5.16x10^{-09}$
TJR		rs227732	17:54769890	T	C	0.3	1.1	1.04 - 1.09	$1.61x10^{-09}$
KneeHipOA	HipOA, AllOA	rs9908159	17:54841961	T	C	0.51	1	1.03 - 1.05	$4.44x10^{-11}$
AllOA		18:77950448	18:77950448	T	C	$6x10^{-4}$	3.6	2.35 - 5.60	$6.56x10^{-09}$
KneeHipOA		rs551471509	19:9943264	T	C	$6x10^{-4}$	0.2	0.10 - 0.32	$1.15x10^{-08}$
HandOA	FingerOA	rs8112559	19:46390455	C	G	0.89	1.1	1.09 - 1.18	$7.32x10^{-11}$
TJR		rs9981884	21:40585633	A	G	0.49	1	0.94 - 0.97	$7.93x10^{-09}$
KneeOA		rs11705555	22:28206912	A	C	0.76	1.1	1.03 - 1.07	$3.00x10^{-09}$
THR	TJR, HipOA	rs12160491	22:38195796	A	G	0.71	0.9	0.90 - 0.95	$1.28x10^{-10}$
Previously Reported									
HipOA	THR, TJR, AllOA, KneeHipOA	rs11164653	1:103464210	T	C	0.41	0.9	0.91 - 0.94	$2.77x10^{-18}$
AllOA		1:150214028	1:150214028	D	I	0.38	1	1.02 - 1.05	$8.58x10^{-10}$
TJR		rs10797923	1:183901966	T	C	0.69	1.1	1.04 - 1.07	$6.20x10^{-09}$
TJR	KneeHipOA, KneeOA, HipOA, THR	rs2605100	1:219644224	A	G	0.32	1.1	1.05 - 1.09	$4.49x10^{-15}$
KneeHipOA	KneeOA	rs7581446	2:33423801	T	C	0.48	1	0.94 - 0.97	$4.87x10^{-11}$
AllOA	HipOA, TJR, THR, ThumbOA, KneeHipOA, HandOA	rs3771501	2:70717653	A	G	0.47	1	1.03 - 1.05	$4.05x10^{-15}$
AllOA		rs62182810	2:204387482	A	G	0.54	1	1.02 - 1.04	$3.82x10^{-9}$
THR	KneeHipOA, TJR, HipOA	rs3774354	3:52817675	A	G	0.37	1.1	1.07 - 1.12	$1.40x10^{-16}$
TJR	TKR, HipOA, AllOA, KneeOA, THR, KneeHipOA	rs1530586	4:1760927	T	C	0.8	1.1	1.06 - 1.11	$3.34x10^{-14}$
THR	TJR, HipOA, KneeHipOA, AllOA	rs1913707	4:13039440	A	G	0.6	1.1	1.06 - 1.11	$1.23x10^{-13}$
AllOA	HipOA, KneeHipOA	rs13107325	4:103188709	T	C	0.07	1.1	1.06 - 1.10	$3.25x10^{-17}$
KneeHipOA	HipOA	rs3884606	5:170871074	A	G	0.52	1	0.95 - 0.97	$8.96x10^{-10}$

Table 2 (continued): Summary statistics of all Genome-Wide Significant Signals in the GO-Consortium

Phenotype	Other Phenotypes	SNV	Chr:pos	EA	NEA	EAF	OR	95% CI	P-value
HipOA		rs79220007	6:26098474	T	C	0.93	0.9	0.87 - 0.93	$2.22x10^{-09}$
KneeHipOA		rs2856821	6:33046742	T	C	0.79	1.1	1.03 - 1.06	$5.71x10^{-09}$
THR	KneeHipOA, HipOA, TJR	rs17288390	6:45384018	T	C	0.65	0.9	0.90 - 0.94	$9.16x10^{-13}$
THR	HipOA, TJR	rs9475400	6:55638258	T	C	0.1	1.2	1.10 - 1.19	$1.73x10^{-13}$
THR	HipOA, TJR	rs12209223	6:76164589	A	C	0.11	1.2	1.18 - 1.26	$1.92x10^{-29}$
HipOA	THR	rs111844273	7:18436337	A	G	0.02	1.3	1.18 - 1.34	$1.05x10^{-12}$
THR	HipOA	rs143083812	7:128843410	T	C	$1.1x10^{-3}$	3.3	2.34 - 4.66	$1.11x10^{-11}$
THR	HipOA, TJR	rs11984666	8:130730280	A	C	0.2	0.9	0.87 - 0.92	$1.69x10^{-15}$
KneeHipOA	KneeOA	rs10974438	9:4291928	A	C	0.65	1	1.03 - 1.06	$7.39x10^{-11}$
KneeHipOA	TKR, KneeOA, TJR, AllOA	rs72760655	9:116916214	A	C	0.33	1.1	1.03 - 1.06	$5.97x10^{-13}$
THR	HipOA	rs1330349	9:117840742	C	G	0.59	1.1	1.07 - 1.12	$6.47x10^{-17}$
THR	HipOA, TJR	rs1321917	9:119324929	C	G	0.41	1.1	1.08 - 1.13	$9.87x10^{-19}$
THR	HipOA	rs62578126	9:129375338	T	C	0.37	0.9	0.90 - 0.94	$1.39x10^{-12}$
KneeHipOA	TJR	rs1517572	11:28829882	A	C	0.41	1	1.03 - 1.05	$6.79x10^{-10}$
THR	HipOA, TJR	rs67924081	11:65342981	A	G	0.74	1.1	1.07 - 1.12	$2.14x10^{-13}$
THR	HipOA	rs34560402	11:66872320	T	C	0.06	0.9	0.82 - 0.90	$2.64x10^{-10}$
KneeHipOA		rs1149620	11:76506572	A	T	0.44	1	0.95 - 0.97	$2.87x10^{-9}$
FingerOA		rs7294636	12:15054016	A	G	0.37	1.2	1.12 - 1.20	$2.99x10^{-16}$
THR	TJR, KneeHipOA, HipOA	rs10843013	12:28025196	A	C	0.78	0.9	0.84 - 0.88	$2.53x10^{-30}$
THR	HipOA	rs17120227	12:59289349	T	C	0.07	1.2	1.12 - 1.22	$7.21x10^{-13}$
KneeHipOA	TJR	rs7953280	12:94136009	C	G	0.5	1	1.03 - 1.06	$4.84x10^{-12}$
KneeOA		rs753350451	12:123732769	D	I	0.2	0.9	0.91 - 0.95	$3.36x10^{-10}$
TJR	HipOA, THR	rs1809889	12:124801226	T	C	0.28	1.1	1.05 - 1.09	$5.70x10^{-14}$
KneeOA	KneeHipOA	rs4380013	15:50759428	A	G	0.19	1.1	1.04 - 1.08	$8.73x10^{-10}$
HandOA	KneeOA, TKR, FingerOA, ThumbOA	rs11071366	15:58334244	A	T	0.61	0.9	0.88 - 0.92	$4.88x10^{-17}$
HipOA	TJR,THR, KneeHipOA, AllOA	rs12908498	15:67366488	C	G	0.54	1.1	1.06 - 1.10	$1.85x10^{-16}$
KneeHipOA	TJR, HipOA, KneeOA	rs9940278	16:53800200	T	C	0.43	1.1	1.04 - 1.07	$1.45x10^{-18}$
KneeOA	TKR	rs34195470	16:69955690	A	G	0.45	1	0.94 - 0.96	$3.13x10^{-13}$
AllOA	TKR, KneeHipOA, KneeOA	rs216175	17:2167690	A	C	0.83	1	1.03 - 1.06	$2.74x10^{-12}$
THR	HipOA	rs7212908	17:59654593	A	G	0.8	0.9	0.89 - 0.94	$1.95x10^{-11}$
THR	TJR, HipOA	rs2716212	17:67503653	A	G	0.62	0.9	0.91 - 0.95	$3.56x10^{-10}$
AllOA	KneeOA	rs10405617	19:10752968	A	G	0.32	1	1.02 - 1.04	$9.33x10^{-11}$
TJR	AllOA	rs75621460	19:41833784	A	G	0.03	1.2	1.14 - 1.28	$2.72x10^{-10}$
THR	HipOA, TJR	rs4252548	19:55879672	T	C	0.02	1.4	1.29 - 1.49	$2.49x10^{-19}$

Table 2 (continued): Summary statistics of all Genome-Wide Significant Signals in the GO-Consortium

Phenotype	Other Phenotypes	SNV	Chr:pos	EA	NEA	EAF	OR	95% CI	P-value
KneeOA	AllOA, TJR, KneeHipOA, TKR	rs143384	20:34025756	A	G	0.59	1.1	1.06 - 1.09	$1.01x10^{-23}$
THR	TJR	rs9981408	21:40017446	T	G	0.23	1.1	1.07 - 1.12	$2.21x10^{-12}$
Female specific									
THR		rs116112221	2:59439973	T	C	$6.1x10^{-3}$	1.95	1.58 - 2.41	$4.61x10^{-10}$
THR		rs10282983	8:69590554	T	C	0.22	1.15	1.11 - 1.19	$2.21x10^{-14}$
AllOA		rs10453201	9:34050345	T	C	0.22	1.05	1.02 - 1.06	$1.05x10^{-08}$

Phenotype: Osteoarthritis phenotype with which the SNV has a genome-wide significant association (p-values 1.3x10-08).Other Phenotypes: Other osteoarthritis phenotypes that are associated with this SNV at GWS level but are less significant than OA Phenotype; AllOA, osteoarthritis (OA) at any of one or more of the investigated joint sites: hip, knee, hand, finger, thumb and spine; KneeOA OA or joint replacement at the left and/or right knee joint; HipOA, OA or joint replacement at the left and/or right hip joint; KneeHipOA, KneeOA and/or HipOA; TJR, Undergone total knee and/or hip replacement due to OA in left, right or both joints; TKR, Undergone total knee replacement due to OA in left right or both knee joints; THR, Undergone total hip replacement due to OA in left, right or both hip joints; SpineOA, Lumbar disk degeneration: narrowing, osteophytes or both present on two or more vertebral levels [REF: PMID: 20147869]; FingerOA, OA at one of the distal interphalangeal (DIP) joints 1-4, the proximial interphalangeal (PIP) joints 1-4 in the left finger and right fingers. If only data from one side (left/right) was available cases were defined on that side only. ; ThumbOA, OA at the first carpometacarpal (CMC1) joint and /or trapezioscaphoid (TS) joint in the left thumb and right thumb. If only data from one side (left/right) was available cases were defined on that side only. HandOA, FingerOA and/or ThumbOA. EA: effect allele, NEA: Non-Effect allele, EAF: effect allele frequency: aoverage frequency in the tested cohorts, OR: odds ratio. 95%CI: 95% confidence interval of the OR

Female-specific osteoarthritis risk variants

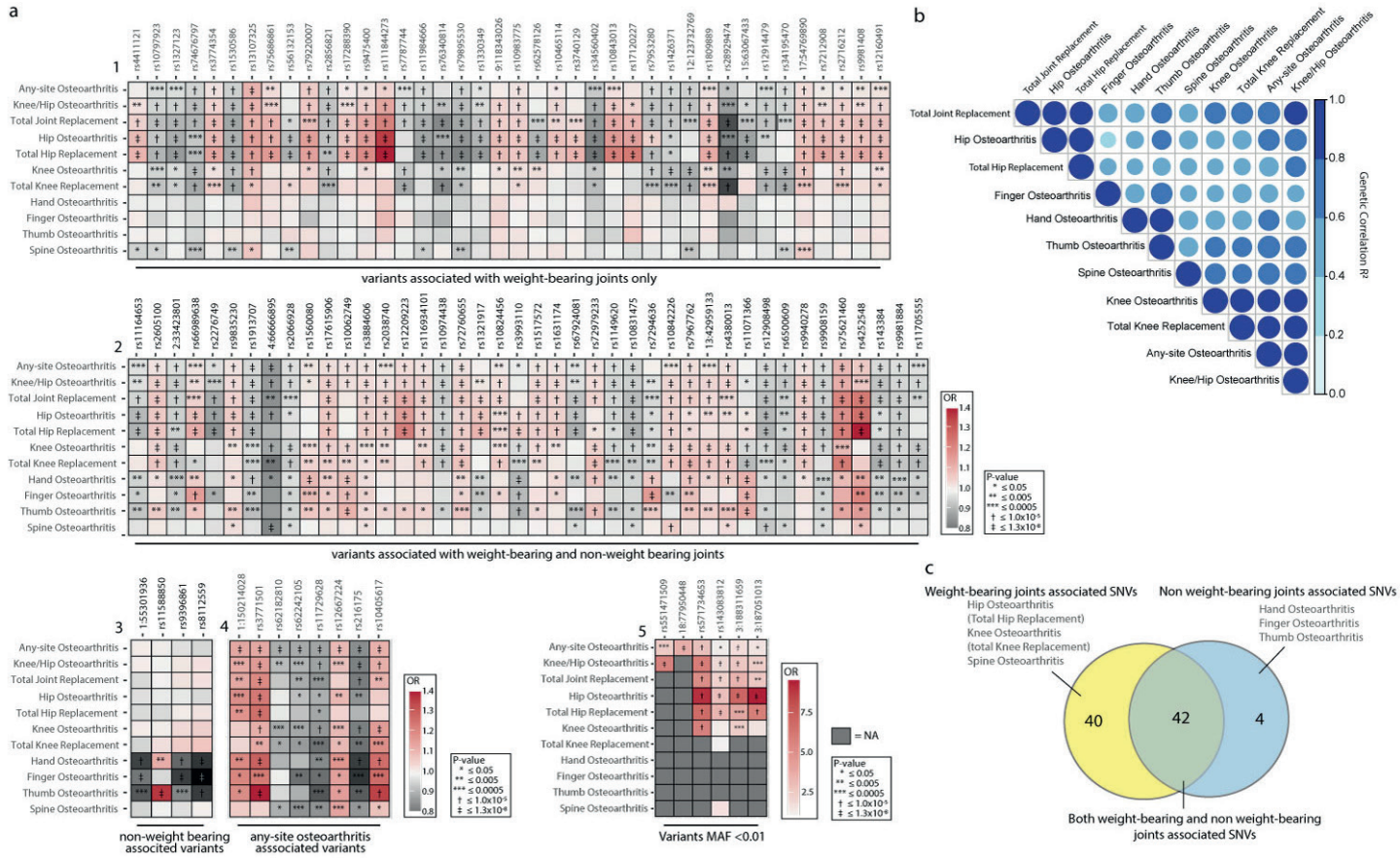
To investigate if the 100 SNVs confer sex-specific effects, we combined association summary statistics from male-specific meta-analysis, consisting of up to 56,462 cases and 153,808 controls, and female-specific meta-analysis, consisting of up to 90,838 cases and 192,697 controls. None of the 100 SNVs showed significant differences in effect size between men and women ($p\text{-value}_{het\text{-}diff} < 0.0005$) (**Supplementary Table 4**). To investigate the presence of any additional osteoarthritis signals specific to males only, females only, or with effects of opposite direction in men and women, we performed for each of the 11 osteoarthritis phenotypes separately, a sex-differentiated test of association and a test of heterogeneity in allelic effects[9, 10]. We identified three new female-specific independent SNVs, two of which showed significant ($p\text{-value}_{het\text{-}diff} < 0.016$) differences in effect size between sexes (**Supplementary Table 4**). rs116112221 ($p\text{-value}_{sex}$-diff=$3.20x10^{-09}$,$p\text{-value}_{het\text{-}diff}$=$4.09x10^{-04}$; female OR=1.95, 95% CI= 1.58-2.41, $p\text{-value}_{female}$=$4.61x10^{-10}$; male OR=1.06, 95% CI=0.82-1.38, $p\text{-value}_{male}$=0.64) is significant in the female-only total hip replacement phenotype and is located in a region containing long intergenic non coding RNAs with the closest protein coding gene being *FANCL* . We

further identified a signal associated with total hip replacement with opposite direction of effects between men and women, rs10282983 (p-value$_{sex\text{-}diff}$=4.93x10^{-16}, p-value$_{het\text{-}diff}$=7.66x10^{-14}; female OR=1.15, 95% CI=1.11-1.19, p-value$_{female}$=2.21x10^{-14}; male OR=0.92, 95% CI=0.88-0.96, p-value$_{male}$=5.16x10^{-04}). rs10282983 resides in an intron of *C8orf34* and has also been associated with waist-to-hip ratio[11] and heel bone mineral density[12].

Similarities and differences of signals across phenotypes

We observed that some variants demonstrate a joint-specific effect (within the power bracket afforded by the study). We found that the majority of SNVs, 60 out of the 100, were genome-wide significantly associated with more than one osteoarthritis phenotype (**Figure 2, Supplementary Information**). 40 of the identified SNVs show genome-wide significant associations with weight-bearing joints only (spine, knee and/or hip), and four display associations with non weight-bearing joints only (hand, finger and thumb) (**Figure 2**). Although several core pathways are known to underpin osteoarthritis pathology, regardless of joint site affected, no common genetic osteoarthritis SNVs have been found previously, with the exception of the *GDF5* locus[13, 14]. Here, we have identified 42 SNVs with strong association across both weight-bearing and non weight-bearing joints. Several of these SNVs, rs3771501 (*TGFA*), rs3993110 (*TEAD1/DKK3*), rs72979233 (*CHRDL2*), rs7967762 (*FPKM/WNT10B*) (**Figure 2, a panel 2,4**) are strongly associated with multiple osteoarthritis joint sites. These SNVs likely represent a common underlying mechanism in osteoarthritis pathology and could therefore be prime candidates for drug development.

Novel insights may be gleaned from the comparison of association signals across osteoarthritis phenotypes. For example, most of the SNVs associated with knee, hip, and knee and/or hip osteoarthritis have a larger effect size on the respective joint replacement-defined phenotypes, all of which are notably of smaller sample size. This could be driven by homogeneity of phenotype definition (**Supplementary Table 1**), or can rep-resent an actual biological and functional relevance of the association, indicating that these loci might play more important roles in receiving a joint replacement (i.e., pain and inflammation) than in osteoarthritis pathology itself. For example, rs76340814 (*PTCH1*) and rs28929474 (missense variant in *SERPINA1*) have stronger associations and larger effect sizes with THR, TKR and TJR, than with hip or knee osteoarthritis (**Figure 2, panel 2**). Indeed, *PTCH1* is thought to function in neurogenetic/brain de-velopment (PMID 20583177, 16405370) and *SERPINA1* in inflammation. Studies in rat osteoarthritis models have shown that early treatment with sivelestat and serpinA1 blocked the proteolytic activity of neutrophil elastase and caused lasting improvements in joint inflammation, pain and saphenous nerve damage[15].



▲**Figure 2**:**Correlation and overlap between osteoarthritis genetics**. a) Heatmap-plot of all GO consortium identified osteoarthritis associated SNVs (single nu-cleotide variants). Effect sizes (OR: odds ratio) and p-values were examined foreach lead SNV in each performed osteoarthritis phenotype GWAS results. OR are plotted as colour, and p-values are represented as symbols in the box. Heatmap-plots are ordered according to SNV associations: 1: weight-bearing joints only (hip, knee and spine), 2: both weight and non-weight bearing joints (hip, knee, spine, hand, finger and thumb), 3: non-weight bearing joints (hand, finger and thumb), 4: any-site osteoarthritis SNVs, and 5: rare variants MAF ≤0.01. b) Heatmap-plot of the genetic correlation (R2) between the examined osteoarthritis phenotypes. c) Venn diagram depicting the number and overlap of SNVs associated with weight-bearing (Knee, hip, spine) joints and non-weight bearing joints (hand, thumb and finger).

Although the majority of SNVs confer risk or protection for disease consistently across phenotypes, two lead SNVs for hand osteoarthritis demonstrated opposite di-rection of effects with other osteoarthritis phenotypes (**Figure 2**): rs3993110 in the TEA domain family member 1 (*TEAD1*) gene with hip osteoarthritis and THR, and rs11071366 in the aldehyde dehydrogenase 1 family member A2 (ALDH1A2) gene with knee osteoarthritis and TKR. They are both common variants with MAF of 0.393 and 0.387, respectively (**Supplementary Information**).

Genetic links between phenotypes

We examined the extent to which subtypes of osteoarthritis phenotypes share a genetic component by cross trait LD Score regression using equally sized, non-overlapping, subgroups of the sample-sets from the meta-analysis[16, 17]. We found osteoarthritis subtypes to share substantial genetic components, albeit with a wide range. The lowest genetic correlation among anatomically very distinct joints (hip, knee, spine, finger and thumb) was observed between finger and hip osteoarthritis (rg=0.38), and the highest was observed between osteoarthritis of finger and thumb (rg=0.72)(**Figure 2 b**, **supplementary Table 5**).

We also investigated if osteoarthritis genetic components are shared with other traits using cross trait LD score regression[16, 17]. Using LD-Hub, we found that 274 of the 763 traits in this database were significantly correlated with osteoarthritis (Bonferroni p-value<$6.0x10^{-06}$): anthropometric traits (BMI, obesity, weight and fat mass), type 2 diabetes, education, depressive symptoms, smoking behaviour, bone mineral density, reproductive phenotypes and intelligence as previously reported (PMID: 29559693, PMID: 30664745), and, for the first time, several pain phenotypes (**Supplementary Table 5 and supplementary Table 6, Supplementary Information**).

We also investigated cross-trait LD score regression of these phenotypes and others not included in the LD-Hub database using the UK Biobank and the Icelandic deCODE genetic datasets (**Supplementary Table 5**). In these analyses, BMI is also genetically correlated with all osteoarthritis phenotypes, with highest correlation with knee osteoarthritis, rg= 0.42. We note that although fat mass measured by DXA is highly correlated using LD-Hub, neither fat mass nor lean mass are very significant in our cross-trait analysis where we correct the lean mass and fat mass measures for sex, age, and importantly, BMI (**Supplementary Table 5**), whereas weight is. This indicates that increased weight, rather than increased fat tissue that results in higher BMI, is causal in osteoarthritis pathogenesis.

Of note is the high correlation between osteoarthritis and sciatica, fibromyalgia, headaches and other pain phenotypes, where the highest correlation is with spine osteoarthritis (rg = 0.61, 0.87, 0.39, 0.79, respectively). *SOX5*, one of the new signals identified in this study, has been previously reported to be associated with back pain and with imaging-detected lumbar intervertebral disc degeneration, attributing the pain to possible structural problems of the spine[18]. These findings are supported by animal model data where inactivation of *SOX5* leads to defects in cartilage and skeletogenesis in mice[19]. We also observed strong correlation between osteoarthritis and pain phenotypes in the LD-Hub database (all derived from the UK Biobank dataset), in particular between spine osteoarthritis and dorsalgia (rg=0.87), leg pain on walking (rg=0.82), knee pain (rg=0.63), hip pain (rg=0.76) back pain (rg=0.75) and neck/shoulder pain (rg=0.74) (**Supplementary Table 6**).

Pain is the most disabling symptom experienced by osteoarthritis patients, one of the main reasons to proceed to total joint replacement[20] and the main reason for physician consultation. The aetiology of pain in osteoarthritis is multifactorial including significant soft tissue inflammation, the sensitisation of pain pathways involving the joint nociceptors, the nociceptive processing in the central nervous system and neuropathic pain components in osteoarthritis models[21-24]. Although a main symptom, no genetic determinants of osteoarthritis pain have been discovered before. Our data suggest that part of the identified signals are also associated with osteoarthritis pain.

Resolution of GWAS Signals

Identification of putative causal variants and involved tissues

We employed complementary computational approaches (**Methods**) to fine-map the GWAS signals to a small set of likely causal variants, identify relevant tissues based on signal enrichment, and provide mechanistic insights based on expression quantitative trait locus colocalization. 12 signals were fine-mapped to variant sets contained entirely within the transcript of a single gene with >95% posterior probability (PP), although we note that this does not provide conclusive evidence for the effector gene. However, of note, *ALDH1A2*, which fine maps to 6 intronic variants with 99% PP, is currently the target of approved drugs in use for other indications, providing a potential opportunity for drug repositioning (**Table 3**, **Supplementary Table 8**).

▲ Figure 3: Heatmap depicting tissue-specific gene-regulatory region enrichment significance(-log10 p-value) for all osteoarthritis(OA) GWAS phenotypes. Tissue/cell type (full name, E-identifier, group name) and P value (−log10) of most significant enrichments (p-value≤1.3*10-08) are shown. Enrichment was calculated using all OA associated lead SNVs and the fine mapped variants, per OA phenotype and all together. Only rows and columns containing a significant enrichment(p-value≤1.3x10-8) for all OA phenotypes (Total) are shown, see supplementary table xx for a full matrix of all 127 examined cell and tissue types. OA: osteoarthritis.

For six SNVs (three new and three known), a single variant could be postulated as causal with >95% PP (**Supplementary Table 7**). The known signals were successfully fine mapped to the same variants as previously reported (rs4252548, a missense variant in *IL11*; rs13107325, a missense variant in *SLC398A*; and rs75621460, an intronic variant in*TGFB1*). The new fine-mapped variants were rs72979233 (total knee replacement) residing in an intron of the polymerase (DNA-directed), delta 3, accessory subunit (*POLD3*) gene; rs74852393 (total hip replacement) situated in the intron of talin 2 (*TLN2*); and rs201194999 (osteoarthritis at any site) which is located between *RP11-807H7.2* and *RNU2-40P*. We identified four lead missense SNVs directly affecting protein coding sequence that have the highest prior of being causal than the other variants in the credible set (rs28929474 in *SERPINA1*, rs3740129 in *CHST3*, rs143083812 in *SMO* and rs2276749 in *VGLL4* (**Supplementary Table 7**). Additionally, 5 lead SNVs were in high LD with coding variants with moderate to severe consequences ($R^2 \geq 0.8$) in the 95% PP set (**Supplementary Table 7**) (**Table 3**).

The majority of complex disease SNVs is thought to exert their risk by affecting regulation of a target gene in a tissue and cell-specific context[25, 26]. We examined if variants in the fine mapped 95% credible sets (**Supplementary Table 7**) were enriched for active gene regulatory histone marks in 127 cell tissues[26]. Osteoarthritis affects multiple tissues within the joint, most notably the cartilage and bone, but there is also evidence for involvement of the synovium, and possibly the muscles and tendons of the joint [27](PMID. We found significant enrichment (p-value<1.3×10^{-08}) of osteoarthritis-associated SNVs in active gene regulatory histone marks in known osteoarthritis tissues, such as cartilage (mesenchymal stem cell derived chondrocytes), bone (primary osteoblast) and muscle (muscle, heart) (**Figure 3**). In line with this we also find enrichment with tissues from the same developmental lineage (mesoderm, mesenchymal stem cells, adipose-derived mesenchymal stem cells). Interestingly, we also find enrichment in epithelial tissues (i.e., fibroblasts). The epithelial cell type enrichment may be due to the extracellular matrix structure of skin, which like cartilage, muscle and tendons, is also for a large part made out of glycans and collagens [28, 29]. Enrichment in gene regulatory elements was also seen with neurosphere, endocrine (ovary, placenta and adrenal gland) and non-endocrine organs (lung, stomach and spleen).

Functional genomics in primary disease tissue

In order to identify putative effector genes for each of the lead SNVs, we examined if any osteoarthritis signals colocalized with the expression of genes residing within ~1Mb of the osteoarthritis-associated lead variant, using quantitative trait loci (eQTLs) across the 48 tissues in the Genotype-Tissue Expression (GTEx) project [30]. We found colocal-

ization for 72 genes which are associated with 42 signals in at least one GTEx tissue. The GTEx dataset does not contain cartilage or bone expression data; however, 33 signals colocalized in tissues and cell types that were found to have enrichment of active gene regulatory elements at the osteoarthritis-associated loci (**Figure 3, Supplementary Table 7**).

We also assessed if any of the genes residing within 1Mb of the osteoarthritis-associated lead variants, showed differential gene expression and protein abundance in primary osteoarthritis-affected tissue, in chondrocytes extracted from osteoarthritis patients undergoing joint replacement surgery. From each patient, two paired cartilage samples were taken; a sample with a low OARSI grade which signifies healthy or macroscopically intact cartilage tissue (intact) and a sample with a high OARSI grade, which denotes highly degenerated (degraded) cartilage tissue. We combined data from two different sources: 35 paired samples from the RAAK study and 38 paired samples from a UK cohort (**Methods**). We only considered genes that were significantly (FDR <0.05) differentially expressed in both studies with the same direction of effect. Forty-four genes showed significant differential expression across both studies, and 34 proteins showed differential abundance (FDR <0.05) (**Supplementary Table 8**). Similarly, we compared gene expression of subchondral bone tissue underneath the intact cartilage and degraded cartilage from two different sources: 11 paired samples from the Taiwan OA study and 24 paired samples from the RAAK study (**Methods**). Thirteen genes were significantly differential expressed across both studies, of which five were also differentially expressed in cartilage tissue (**Supplementary Table 8**). We have used these functional genomics lines of evidence (eQTL, gene expression and protein abundance differences) for the identification of likely effector genes (See Amassing evidence to identify effector genes section below, **Table 3**). For example, the tenascin C (*TNC*) gene demonstrated significant evidence of both increased gene expression and increased protein abundance in degraded compared to intact cartilage. *TNC* resides in the vicinity (<1Mb away) of three independent signals: rs7862601 (hip osteoarthritis); rs72760655 (osteoarthritis of the hip and/or knee); and rs1330349 (total hip replacement), which fine mapped to 3 variants all within the transcript of *TNC*. TNC is a component of the extracellular matrix and is involved with inflammation and cardiovascular disease[31, 32].

Causal inference analysis

We used two-sample Mendelian randomisation (MR) to investigate potential causal relationships between 1,640 plasma proteins measured in up to 6,000 individuals and osteoarthritis[33-37]. Following validation using orthogonal information on DNA

methylation and molecular QTLs from eQTLGen and GTEx (**Methods, Supplementary Information**), three osteoarthritis phenotype associations for 2 proteins were considered robust (**Supplementary Table 9**).

Matrix gla protein (MGP) levels showed strong MR and colocalization evidence for hand and finger osteoarthritis. The lead SNV at this signal (rs7294636) is highly correlated with UTR and missense variants in *MGP* (rs1800801 and rs4236; r^2>0.8), which have previously been found to be associated with hand osteoarthritis severity and with allele-specific expression of *MGP*[38]. Whilst it is possible that the pQTL proxies the effect of the missense variant on probe binding, previous functional studies have identified correlation between the expression of *MGP* and osteoarthritis in cartilage and other joint tissues[38, 39]. MGP is a vitamin K-dependent protein secreted by chondrocytes and plays a role in the inhibition of ectopic tissue calcification.

Spondin 2 (SPON2) protein levels demonstrated MR and colocalization evidence for knee osteoarthritis. The pQTL of *SPON2* (rs878323) is in strong LD with a missense variant (rs11247975) therefore epitope-binding artefacts cannot be ruled out. SPON2 is expressed in human osteoarthritis synovial fluid[40], and is a secreted extracellular matrix protein. It is essential for the initiation of immune responses and a unique pattern recognition molecule for microbial lipopolysaccharide[41, 42], which has been linked to knee osteoarthritis severity[43].

Amassing evidence to identify disease pathways

By combining the results from the complementary functional genomics and computational approaches outlined above, we identified 376 genes with at least one line of evidence pointing to a putative effector gene (**Table 3**, **Supplementary Table 8**) and assessed pathway and gene set signal enrichment in osteoarthritis pathology (**Supplementary Table 10**). We found significant enrichment of signal in biological processes known to be involved in osteoarthritis, specifically in pathways relevant to skeletal, bone and cartilage development. Bone and cartilage development pathways were also found to be enriched in signals traversing weight-bearing and non-weight-bearing joints (**Supplementary Table 10**), indicating joint development as a common mechanisms for any form of osteoarthritis[13]. Interestingly, when we only examined genes from SNVs associated with weight-bearing joints only, the most significant pathways were related to cellular response to endogenous stimuli, neurogenesis and Alzheimer's (**Supplementary Table 10**). Skeletal developmental pathways remained in the top 20 most significantly enriched pathways. We did not have sufficient power to detect significant enrichment for the non-weight-bearing joint signals.

Amassing evidence to identify effector genes

From the 376 genes (mentioned above, **Supplementary Table 8**) we combined supportive information from the fine-mapping and eQTL colocalization analyses, animal model data, human musculoskeletal phenotype, functional genomics and plasma pQTL causal inference analysis and identify 38 genes that have at least three different lines of evidence pointing to them as a putative effector gene (**Table 3**). Of these 38 genes, 23 provide strong additional supportive evidence for known osteoarthritis-associated genes (**Table 3**) and 15 reside in newly-associated signals: *C2orf40* (lead SNV rs66989638), *TSEN15* (rs1327123), *VGLL4* (rs2276749), *RNF144B* (rs9396861), *CHST3* (rs3740129), *TRIOBP* (rs12160491), *CUX1* (rs116934101), *FBN2* (rs17615906), *IGF1R* (rs12914479), COL2A1, *PFKM* and *WNT10B* (rs7967762), *PTCH1* (rs76340814), *TNFSF11* (rs58973023) and SNAP47 (rs11588850). We discuss the 15 genes associated with the new signals here.

Four of the putative osteoarthritis causal genes are involved in developmental pathways: *PTCH1*, *WNT10B*, *CHST3* and *VGLL4.* Developmental pathways, particularly for skeletal, bone and cartilage development, are known to be involved in osteoarthritis pathology[13]. Mutations and deficiency of carbohydrate sulfotransferase 3 protein (*CHST3*), an enzyme involved in the metabolism of the major proteoglycan present in cartilage chondroitin sulphate, have been previously associated with short stature, congenital joint dislocations, clubfoot and elbow joint dysplasia[44]. *CHST3* has been shown to be associated with lumbar disc degeneration[45]. In addition, mutations in *CHST3* have been observed in patients diagnosed with autosomal recessive Larsen syndrome, which affects bone development[46].

PTCH1 (Patched 1) is a member of the hedgehog (Hh) signaling pathway, important in embryonic development and tissue maintenance and regulation. PTCH1 is a receptor for Hh ligands and regulates the activity of SMO. When bound, PTCH1 relinquishes its inhibitory effect on SMO and activates the Hh signalling cascade. The Hh pathway plays an important role in the longitudinal growth of long bones during development[47] and there are currently drug repurposing opportunities for this pathway (see SMO below).

WNT10B (Wnt family member 10B) is involved in the Wnt signalling pathway, which has an established role in osteoarthritis pathogenesis[48]. Mutations in *WNT10B* have been linked to limb defects and dental abnormalities[49-51]. Downregulation of *VGLL4* is linked to the upregulation of Wnt/β-catenin pathway target genes[52]. VGLL4 functions via interacting with TEA domain (TEAD) transcription factors[52, 53].

Table3: Possible Causal Osteoarthritis Genes

Lead OA SNV	Coding Variant & Fine-map	FineMapped Gene	eQTL colocalization (Gtex/OA tissue)	Cartilage		Subchon-dral Bone	Blood pQTL MR CoLoc	musculoskeletal phenotype		Score	Gene	Druggable Genome
				Diff. Expresion	Diff. Abundance	Diff. Expresion		Mouse	Human			
rs10405617			ILF3(2/1)					ILF3		3	ILF3	.
rs11071366		ALDH1A2			ALDH1A2 (-)			ALDH1A2		3	ALDH1A2	Tier 1
rs11164653		COL11A1						COL11A1	COL11A1	3	COL11A1	Tier 3A
rs1149620		TSKU	TSKU(5/4)							3	TSKU	.
rs11588850			SNAP47 (3/3)					SNAP47		3	SNAP47	.
rs116934101		CUX1		CUX1 (+)				CUX1		3	CUX1	.
rs12160491			TRIOBP(22/13)	TRIOBP (-)						3	TRIOBP	.
rs12908498		SMAD3	SMAD3(1/1)					SMAD3	SMAD3	5	SMAD3	Tier 2
rs12914479			IGF1R(1/1)					IGF1R		3	IGF1R	Tier 1
rs1327123	TSEN15		TSEN15 (8/5)							3	TSEN15	.
rs1330349		TNC		TNC (+)	TNC (+)					3	TNC	Tier 1
rs1330349				COL27A1 (-)				COL27A1	COL27A1	3	COL27A1	Tier 3A
rs143083812	SMO	SMO						SMO		3	SMO	Tier 1
rs143384		GDF5	GDF5(4/1)					GDF5	GDF5	5	GDF5	Tier 3A
rs1530586			TACC3(9/6)					TACC3		3	TACC3	.
rs1530586			FGFR3(6/4)					FGFR3	FGFR3	4	FGFR3	Tier 1
rs17615906			FBN2(2/2)					FBN2	FBN2	4	FBN2	Tier 3A
rs2276749	VGLL4	VGLL4	VGLL4(1/0)							3	VGLL4	.
rs2716212		MAP2K6	MAP2K6(1/1)							3	MAP2K6	Tier 1
rs2716212					PRKAR1A (-)			PRKAR1A	PRKAR1A	3	PRKAR1A	.
rs34195470		WWP2		WWP2 (-)				WWP2		3	WWP2	.
rs3740129	CHST3		CHST3(6/1)					CHST3	CHST3	5	CHST3	Tier 1
rs3771501		TGFA	TGFA(2/2)					TGFA		4	TGFA	Tier 1
rs3884606		FGF18		FGF18 (+)				FGF18		3	FGF18	Tier 1
rs4252548	IL11	IL11		IL11 (+)		IL11 (+)				4	IL11	Tier 3A
rs4380013		USP8	USP8(6/2)							3	USP8	.
rs58973023				TNFSF11 (+)		TNFSF11 (+)		TNFSF11	TNFSF11	4	TNFSF11	Tier 1
rs62578126			LMX1B(3/2)					LMX1B	LMX1B	4	LMX1B	.
rs66989638	C2orf40	C2orf40	C2orf40 (40/21)							4	C2orf40	Tier 3B

Table 3 (continued): Possible Causal Osteoarthritis Genes

Lead OA SNV	Coding Variant & Fine-map	FineMapped Gene	eQTL colocalization (Gtex/OA tissue)	Cartilage		Subchon-dral Bone	Blood pQTL MR CoLoc	musculoskeletal phenotype		Score	Gene	Druggable Genome
				Diff. Expresion	Diff. Abundance	Diff. Expresion		Mouse	Human			
rs67924081			LTBP3(1/1)		LTBP3 (+)			LTBP3	LTBP3	5	LTBP3	Tier 3A
rs7212908				TBX4 (-)				TBX4	TBX4	3	TBX4	.
rs72760655				COL27A1 (-)				COL27A1	COL27A1	3	COL27A1	Tier 3A
rs7294636			MGP(4/2)				MGP	MGP	MGP	5	MGP	Tier 3B
rs75621460		TGFB1						TGFB1	TGFB1	3	TGFB1	Tier 1
rs7581446		LTBP1	LTBP1(1/0)					LTBP1		3	LTBP1	Tier 3A
rs76340814			PTCH1(5/2)					PTCH1		3	PTCH1	.
rs7967762			COL2A1(1/0)					COL2A1	COL2A1	3	COL2A1	Tier 3A
rs7967762			PFKM(2/2)		PFKM (-)			PFKM		4	PFKM	.
rs7967762				WNT10B (+)		WNT10B (+)		WNT10B	WNT10B	4	WNT10B	Tier 3B
rs9396861		RNF144B	RNF144B(2/1)							3	RNF144B	.
rs9981408		ERG	ERG(1/1)					ERG		4	ERG	Tier 2

Table 3: Gene s are identified according to the Ensembl GeneName for the gene. Abbreviations are: Lead OA SNV, rsid of the lead variant; EA, effect allele: EAF effect allele frequency; OA, if the signal is new (N) or previously reported (K); Coding variant & FineMap, gene in which the lead SNV has a moderate to high severity or gene in which a moderate to high severity variant* is in high LD (≥r20.8) with the lead SNV and is present in the 95% credible set; Fine mapped gene, all variants in the 95% credible set reside within the transcript of this gene. eQTL colocalisation, gene was colocalised in at least 1 GTEx tissue with the number of GTEx tissues in parentheses followed by the number of these tissues that were also enriched in tissue enrichment analysis , which is suggestive of a role in osteoarthritis pathology. Cartilage Differential Expr, Gene was differentially expressed in high-grade compared to low-grade osteoarthritic cartilage; Cartilage Differential Abund, gene that codes for a protein that was differentially expressed in high-grade compared to low-grade osteoarthritic cartilage. Bone Differential Expr, Gene was differentially expressed in high-grade compared to low-grade in sub-chondral bone; Blood pQTL MR Coloc, gene that was identified as on the causal path in the mendelian randomisation for cis-eQTLs and also colocalised; Score, cumulative score for each gene based on the supporting Finemap and functional analysis;Human musculoskeletal phenotype, indicates if a gene has been linked to a musculoskeletal phenotype according to the Nosology and classification of genetic skeletal disorders (PMID: 31633310); Mouse musculoskeletal phenotype, indicates if a musculoskeletal phenotype is observed in any mouse knockout from www.boneandcartilage.com and the MGI (Mouse Genome Informatics Database) Batch Query was used to extract all mouse knockout phenotypes from the GDX (Mouse Gene Expression Database)
we used the Nosology and classification of genetic skeletal disorders (PMID: 31633310) to identify genes within 1Mb (upstream or downstream) of the 100 SNVs that had links to human musculosketal phenotypes:
*Moderate to high consequences for severity: transcript_ablation, splice_acceptor, splice_donor, stop_gained, frameshift, stop_lost, start_lost, transcript_amplification, inframe_insertion, inframe_deletion, missense, protein_altering

Notably, we identified another new total hip replacement- and hand osteoarthritis-associated signal located in such a transcription factor, the *TEAD1* gene, indicating a common molecular pathway underlying both signals. TEAD1 functions in the Hippo signalling pathway and is transcriptionally regulated by the YAP1 and TAZ protooncogene proteins, which are involved in mechanosensing and mechanotransduction[54, 55]. Mechanoadaptation of articular cartilage is an important factor in osteoarthritis [56, 57].

Two of the likely effector genes have bone remodelling, immune response and inflammatory links. Activation of T-cells can lead to osteoclastogenesis and bone resorption by influencing the expression of *TNFSF11* (TNF Superfamily Member 11)[58]. *TNFSF11* encodes receptor activator of nuclear factor kappa-β ligand (RANKL), a cytokine that has been linked to inflammatory bone remodelling in rheumatoid arthritis, with increased TNFSF11 levels associated with worsening arthritis severity[59, 60] and a well-established role in osteoclastogenesis[61] . *FBN2* (fibrillin 2) encodes fibrillin2, a glycoprotein that forms microfibrils in the extracellular matrix and has a major role during early morphogenesis. Fibrillins potently regulate pathways of the immune response, inflammation and tissue homeostasis[62], are important in bone remodelling and regulate local availability of BMP and TGFα[62]. Mutations in FBN2 cause contractual arachnodactyly[63].

Three genes have a neurological connection. Both *C2orf40* and *TSEN15* are involved in central nervous system development. Augurin, the protein encoded by *C2orf40* (also called *ECRG4*), is involved in central nervous system development in animal models[64] and shows association with neuropathologic features of Alzheimer's disease and related dementias in humans[65]. *TSEN15* belongs to a family of proteins (t-RNA splicing endonucleases) that have been found to be mutated in patients with pontocerebellar hypoplasia and microcephaly, indicating a role of the gene in brain development[66].The last gene *CUX1* (Cut like homeobox 1) is a transcription factor involved with brain neuronal differentiation and synaptogenesis[67]. Increased CUX1 expression may stimulate cell motility and proliferation[68]. Cux1 expression during developing limb was observed at chondrogenic interzones suggesting a regulatory role in joint formation[69].

The five remaining genes have varied biological functions. *COL2A1* (Collagen Type II Alpha 1 Chain) encodes for an essential structural component of cartilage and is important for joint formation and bone growth. A wide spectrum of diseases are associated with *COL2A1* including cartilage and bone abnormalities, such as spondyloepimetaphyseal dysplasia and Kniest dysplasia[70, 71]. *SNAP47* (Synaptosome Associat-

ed Protein 47) has widespread expression in nervous tissue[72] and demonstrates an altered bone phenotype in mutant mouse lines. *RNF144B* (Ring Finger Protein 144B) is involved in the protein ubiquitination pathway. *IGF1R* (Insulin like growth factor 1 receptor) has tyrosine kinase activity and mediates the action of insulin-like growth factor. *PFKM* (phosphofructokinase) encodes a muscle isozyme that catalyses the phosphorylation of fructose-6-phosphate during glycolysis. Finally, we identified TRIO And F-Actin Binding Protein (*TRIOBP*) as the likely causal gene for rs12160491, which is associated with total joint replacement, total hip replacement and hip osteoarthritis. This gene has been linked to cardiovascular function[73] and hearing loss[74].

Drug target identification

We examined the druggability status[75] of the likely causal genes (**Table 3)** using the druggable genome database[8](**Supplementary Table 8**). From these osteoarthritis likely causal genes 11 reside in tier 1, which incorporates the targets of approved drugs (licensed drugs) and drugs in clinical development (**Supplementary Information, Methods**, **Table 3**), and 2 genes reside in tier 2, which includes genes encoding targets with known bioactive drug-like small molecule binding partners and those with ≥50% identity with approved drug targets.

Within tier 1, two candidates are already in clinical trials of efficacy for osteoarthritis. *FGF18* encodes fibroblast growth factor 18 that signals through the FGF receptor 3 to promote chondrogenesis. The recombinant FGF18 Sprifermin has recently been shown to increase cartilage thickness in knee osteoarthritis in a phase 2 clinical trial (PMID: 32098758). *TNFSF11* encodes receptor activator of nuclear factor kappa-β ligand (RANKL), a pivotal protein in bone remodelling[76], consistent with the concept that bone remodelling plays a role in osteoarthritis pathogenesis[77]. The human monoclonal antibody to RANKL, denosumab, is the most potent anti-resorptive agent approved for osteoporosis treatment and is also currently in a phase 2 clinical trial for knee osteoarthritis progression (ISRCTN96920058).

The other tier 1 druggable targets of which we have strong supportive evidence as putative effector genes (**Table 3**): *ALDH1A2, IGF1R, TNC, SMO, FGFR3, MAP2K6, CHST3, TGFA, TGFB,* all have market authorisation or are in clinical development for other indications. The functional and epidemiological evidence of their role in clinical osteoarthritis presented here provides support for early repurposing investigation.

Discussion

Osteoarthritis is the leading cause of joint disability and pain worldwide, but to date no curative treatment options exist. By conducting a large-scale GWAS meta-analysis, we have identified 52 new signals implicated in osteoarthritis risk, bringing the total to 148. We have identified multiple risk variants that traverse osteoarthritis phenotypes, indicating partly shared underpinning mechanisms and pathways relating to bone and cartilage development regardless of joint affected. We have further identified clear differences between osteoarthritis affecting the weight-bearing and non weight-bearing joints. For the first time, we have been able to establish genetic links between the disease and its main symptom, pain. By integrating fine-mapping, colocalization and causal inference analyses with transcriptomics and proteomics studies in primary disease tissue, we have identified likely effector genes that represent high-value targets for therapeutics. These findings enhance our understanding of the genetic aetiology of disease, shed novel biological insights, and provide a stepping stone for translating genetic associations into osteoarthritis drug development.

Methods

GWAS cohorts, phenotype definition and meta-analysis

Genome-wide association analysis for osteoarthritis was performed across 21 cohorts (**Supplementary Table 1**), for a total of 826,690 individuals (177,517 osteoarthritis patients). We defined 11 stratified osteoarthritis phenotypes: osteoarthritis at any site, osteoarthritis of the hip and/or knee, knee osteoarthritis, hip osteoarthritis, total joint replacement, total knee replacement, total hip replacement, hand osteoarthritis, finger osteoarthritis, thumb osteoarthritis and spine osteoarthritis (**Figure 1**, **Supplementary Table 1**, **Supplementary Information**). Osteoarthritis was defined by either a) self-reported osteoarthritis, b) clinical diagnosed, c) ICD10 codes or b) radiographic as defined by the TREAT-OA consortium[78], depending on the data available in the cohort (**Supplementary Table 1**). GWAS analysis were performed by each cohort, and adjusted for cohort specific covariates (**Supplementary Table 1**). GWAS summary statistics from all cohorts were collected and checked to contain all the data needed for the meta-analysis. The quality control (QC) was performed centrally by using EasyQC[79]. Briefly, missing data, mono-allelic SNVs, nonsensical values (P-value>1, infinite beta's etc.) and duplicates were removed from the data. We excluded variants with poor imputation quality (R2<0.3), if the effective sample size was <20 and if the minor allele count was <6 . Allele coding was harmonized across cohorts (A/T/C/G or I/D). Allele frequen-

cy was checked against the imputation reference (HRC or 1000G; URLs) to identify possible allele coding errors. p-values values were checked to match the corresponding beta values. Cleaned data was used as input for the meta-analysis. Meta-analysis was performed using inverse variance weighting in METAL[80]. Genomic control was performed on all datasets, except those which had already carried out genomic-control adjustments prior to centralized QC and meta-analysis. Genome-wide significance threshold was set at p-value<$1.3x10^{-08}$, corrected for multiple testing. For each phenotype we only considered variants reported in at least two cohorts with the same direction of effect with a minimum MAF>=0.0001 in any contributing cohort.

Sex-differentiated meta-analysis

The meta-analyses and QC steps described above were repeated for males and females separately in a subset of cohorts (**Supplementary Table 4**). We then combined the resulting association summary statistics to conduct a sex-differentiated test of association and a test of heterogeneity in allelic effects, as implemented in GWAMA[9, 79]. This method allows for heterogeneity of allelic effects in magnitude and/or direction between males and females and offers substantial gains in power to detect new SNV associations. The genome-wide significance threshold was set at p-value<$1.3x10^{-08}$, corrected for multiple testing. Heterogeneity in allelic effect sizes was assessed with Cochran's Q statistic and the significance threshold was set at p-value<0.016, corrected for the three independent new signals identified across the 11 osteoarthritis phenotypes.

Significance threshold

The testing of M=11 osteoarthritis phenotypes in this study needed to be taken into account in the interpretation of genome-wide statistical significance. Applying a Bonferroni correction would be inherently conservative as this method assumes independence among the tests considered. Therefore, we first used LD Score regression method[16, 17]with genome-wide meta-analysis summary statistics to estimate the genetic correlation matrix between the 11 osteoarthritis traits and then calculated the effective number of independent traits (M_{eff}) from the eigenvalues λ_i of the correlation matrix[81]:

$$M_{eff}=M-\sum_{i}=1M[I(\lambda_i>1)(\lambda i-1)]$$

For the M = 11 osteoarthritis phenotypes in this study, M_{eff} = 4.6565. The threshold corrected for the effective number of traits to report genome-wide significance is p-value<$1.3x10^{-08}$.

Statistical independence

To define independent signals for each osteoarthritis phenotype, we used clumping function in PLINK 1.9[82] with the following parameters: (a) significance threshold for index variants: p-value>$1x10^{-07}$, (b) LD threshold for clumping: 0.10, and (c) physical distance threshold for clumping: 1Mb (2Mb window around the index variant). LD calculations were based on the full UK Biobank imputed set. To test that the index variants defined by clumping were statistically independent, we performed an approximate stepwise model-selection procedure, as implemented by COJO in GCTA[83, 84]. A signal in a region was defined as independent if its P-value of association in the stepwise regression was less than the adjusted genome-wide significant threshold (p-value<$1.3x10^{-08}$). To define independent signals across the 11 osteoarthritis phenotypes, we performed reciprocal approximate conditional analyses, as implemented by COJO in GCTA[83, 84], of each independent variant of one osteoarthritis phenotype conditioned on each independent variant of the other osteoarthritis phenotypes within 1-Mb region. Two signals were considered dependent if the p-value for either signal conditioned on the other was either ≥$1x10^{-07}$, or attenuated by at least two orders of magnitude. Among dependent variants, the one with the lowest P-value was classified as independent. Using an approximate conditional and joint multiple-SNP analysis, as implemented by COJO in GCTA, we investigated the statistical independence between index signals per osteoarthritis phenotype and previously reported osteoarthritis variants within a 1-Mb region. The index variant was classified as a new association if it had a conditional p-value≤$1x10^{-07}$ or the P-value difference between conditional and unconditional analysis increased by more than two orders of magnitude. Index variants were classified as known if they have previously been reported or the association signal disappeared after conditioning on the variant of a previously reported locus.

Genetic signals across osteoarthritis phenotypes

Results from all independent lead SNVs (n=100) across all osteoarthritis phenotypes were extracted from the full meta-analysis results. All OR were calculated on the minor allele (allele frequency < 50%), and SNVs with a MAF<1% were excluded (n=6). For all the remaining SNVs (n=94) the OR for each osteoarthritis phenotype GWAS was plotted in a heatmap, together with the corresponding association *p-value* (**Figure 2**). SNVs with MAF<1% were plotted in a separate heatmap (**Figure 2**). All figures were plotted using R and adjusted for publication quality using Adobe illustrator. We also created three classification groups: 0=Weight-bearing joints only (hip and/or knee, knee, hip, total joint replacement, total knee replacement, total hip replacement and spine), 1=Both, 2=Non-weight bearing joints only (hand, finger, thumb). Osteoarthritis at any

site wasn't included in this analysis as it wasn't clear which osteoarthritis subphenotype was leading the signal. Each of the 100 independent genome wide significant SNVs was assigned to the above groups only if it wasn't nominally significant (p-value>0.05) for any of the other phenotypes in the other classification groups, resulting in 86 SNVs to be further analysed.

Fine mapping

For each independent signal we included all variants within 1Mb of the index variant. In situations where there was more than one independent signal in the region we used the conditional summary statistics of the meta-analysis conditioned on all other signals. We calculated Wakefield's asymptotic Bayes factors [85] and we determined the posterior probability of each variant being causal. To produce a 95% credible set of variants we ranked according to posterior probability and included those variants with the highest probability of being causal until the shared probability was at least 95%. Some regions were large therefore we considered only variants in the 95% credible with a posterior probability of causality > 3% (**Supplementary Table 7)**.

RNA sequencing analysis of the RAAK cohort

Preserved and lesioned cartilage and subchondral bone samples from the same donor were obtained from the Research in Articular osteoArthritis Cartilage (RAAK) study consisting of patients with osteoarthritis who underwent joint replacement surgery due to an end-stage disease [86-88]. Total RNA from articular cartilage and subchondral bone was isolated using Qiagen RNeasy Mini Kit (Qiagen, GmbH, Hilden, Germany). Paired-end 2×100 bp RNA-sequencing (Illumina TruSeq RNA Library Prep Kit, Illumina HiSeq2000 and Illumina HiSeq4000) was performed. Strand specific RNA-Seq libraries were generated which yielded a mean of 20 million reads per sample. More details on mapping and quality control (QC) from cartilage are previously described (PMID: 30504444, 30298554). Methods of subchondral bones RNA sequencing have been previously described[89]. After QC, 35 paired cartilage samples (N = 70) and 24 paired subchondral bone samples (18 paired knee and 6 paired hip samples) remained for further differential expression analysis. Normalization and statistical framework was performed using the DESeq2 v1.20 R package. A general linear model (GLM) assuming a negative binomial distribution was applied followed by a paired Wald-test between preserved and lesioned osteoarthritis cartilage and subchondral bone samples. Benjamini-Hochberg multiple testing corrected p-values with significance cut-off of 0.05 are reported as False Discovery Rate (FDR).

RNA sequencing and proteomic analysis of the UK cohort

Full methods have been described previously[90, 91]. Briefly: DNA, RNA, and protein was extracted from matched intact and degraded cartilage samples from 38 patients undergoing total joint replacement surgery: 29 knee and 9 hip osteoarthritis patients. All patients provided full written informed consent before participation. The human biological samples were sourced ethically and their research use was in accord with the terms of the informed consents under an Institutional Review Board (IRB)- or Ethics Committee (EC)-approved protocol. Proteomics analysis was performed on intact and degraded cartilage samples from 24 individuals and gene expression analysis on samples from all 38 patients. For the proteomics we performed LC-MS analysis on the Dionex Ultimate 3000 UHPLC system coupled with the Orbitrap Fusion Tribrid Mass Spectrometer. Abundance values were normalised by the sum of all protein abundances in a given sample, then log2-transformed and quantile normalised. We restricted the analysis to 3,917 proteins that were quantified in all samples. Differential abundance was performed using a within-individual paired sample design in limma in R[92]. RNA sequencing was performed on the Illumina HiSeq 2000 (75bp paired-end read length). We determined the gene-level counts from transcript-level quantification using salmon 0.8.219 with GRCh38 cDNA assembly release 87 downloaded from Ensembl. Limma-voom[93] was used to remove heteroscedasticity from the estimated expression data. We tested genes for differential expression using limma in R (with lmFit and eBayes), based on a within-individual paired sample design. For the proteomics and RNA sequencing significance was defined at 5% Benjamini-Hochberg FDR to correct for multiple testing (**Supplementary Table 7**).

Transcriptomics and analysis in subchondral bone in the Taiwan OA cohort

Knee joints samples were collected from 11 Han Chinese patients underwent total knee replacement surgery (TKR). The subchondral bone tissues underneath the intact and eroded cartilage were obtained as previously described[94]. RNA was extracted as described[95]. RNA (400ng) per sample were used for cRNA synthesis and amplification. Cyanine 3-labeled cRNA was then purified and hybridized to Agilent whole human genome 44 k microarray chips (Agilent Technologies) according to the manufacturer's instructions (Agilent Technologies, Santa Clara, CA). The array signal intensities were analysed by the Agilent GeneSpring GX software (version 11.5; Agilent Technologies). Gene expression values were normalized using quantile normalization; probes with low signal intensities were excluded by setting the filter above 32. The normalized values were log transformed and compared using t test. Differentially expressed genes were defined at ≥2 fold-change with Benjamini-Hochberg corrected P-value ≤0.05.

Mouse and human musculoskeletal phenotype

We investigated if any of the genes within 1Mb (upstream and downstream) of the 100 SNVs had a musculoskeletal phenotype in mouse knockouts using information from the Mouse Gene expression database (GDX) of the Mouse Genome Informatics (MGI) database[96]. Mouse orthologues of the all the genes within 1 MB (upstream and downstream) of the 100 osteoarthritis associated SNVs, were extracted from Ensemble Genomes, using biomarkt (GRCh37, Version 69)[97]. The MGI Batch Query was used to extract all mouse knockout phenotypes from the GDX. We also investigated if any genes had a musculokeletal phenotype in mouse knockouts using information from www.boneandcartilage.com. Genes which had a mouse musculoskeletal phenotype were reported (**Supplementary Table 8**). We also investigated human skeletal genetic disorders. We used the Nosology and classification of genetic skeletal disorders[98] to identify genes within 1Mb (upstream or downstream) of the 100 SNVs that had links to human musculosketal phenotypes.

Tissue specificity

We selected all independent genome-wide significant SNVs across osteoarthritis GWAS (n=100). For each signal we investigated the lead SNV and all fine mapped SNVs (95% posterior probability) (n=542) to see if they colocalized with active enhancer histone marks as defined by the ROADMAP epigenomics project[26]. All tissue and cell types available in the ROADMAP epigenomics and ENCODE project were used (n=127)[26]. For each lead SNV and their Fine-Mapped SNVs, the percentage of SNVs located in active enhancer marks was calculated. For the enrichment analysis a background value was calculated (1000 permutations) using 100 random SNVs selected from 1000 Genomes Project[99](MAF>0.05). For these 100 random SNVs and all SNV in high LD (r^2=0.8, LD based on 1000 Genomes Project) their percentage of colocalization with enhancer histone marks was calculated. Once all background permutations were done, an average of all results was taken as the final background values. Enrichment was calculated for each osteoarthritis phenotype and investigated cell type separately, by using the two-proportions Z-test to test if. The significance for enrichment was set to genome-wide significance (p-value<$1.3x10^{-08}$). As the analysis is highly dependent on the amount of variants included (power) significance was only based on enrichment analysis including all 100 independent SNVs across osteoarthritis phenotypes. For the eQTL colocalization using GTEx tissues, we considered the following GTEx tissues as possible osteoarthritis target tissues, based on the tissue specificity analysis: Adipose, Brain, Heart, Lung, Muscle, Nerve, Ovary, Placenta, Skin, Stomach, Cultured fibroblasts, Adrenal gland, and Breast tissue.

eQTL Colocalization

For cis-eQTL colocalization we used summary statistics of SNPs from 48 human tissues from the GTEx v7[100]. For each signal and each tissue we included genes that contained at least 1 eQTL (using a threshold of <5% false discovery rate) in GTEx and that overlapped 100kb either side of our signal. For the colocalization analysis we included all variants in common between the GO meta-analysis and the GTEx cis-eQTL analysis with the exception of indels. We used the Bayesian statistical methodology which implements the method of Giambartolomei[101]. This method evaluates whether the GWAS and molecular QTL associations best fit a model in which the associations are due to a single shared variant, summarized by the posterior probability (PP). Evidence for colocalization was assessed using the PP4 indicating that there is an association for both traits and they are driven by the same causal variant. A PP4 > 0.8 was considered evidence for colocalization (**Supplementary Table 7**).

Causal inference analysis

pQTL Mendelian randomization and colocalization

Two-sample Mendelian randomization (MR) was applied to understand the association between plasma proteins on osteoarthritis. In the MR analysis, 1640 proteins were treated as the exposure and the 11 osteoarthritis phenotypes as the outcomes. The genetic instruments of the plasma proteins were obtained from Zheng et al., where the conditional independent pQTLs were pooled from 5 recent GWAS of plasma proteins[33-37]. The genetic instruments were further split into two groups: 1) cis-acting pQTLs within a 500Kb window from each side of the leading pQTL of the protein were used for the MR analysis; 2) trans-acting pQTLs outside the 500Kb window of the leading pQTL were designated as trans instruments. For the MR analysis, the meta-analysis summary statistics of osteoarthritis including UK Biobank participants were used as outcomes. We selected a p-value threshold of 0.05, corrected for 11 osteoarthritis phenotypes and the number of independent tests, as our threshold for prioritising MR results for follow up colocalization analyses (number of tests= 18.030; P-value< 2.77×10^{-6}) (**Supplementary Table 9**).

For 28 protein-osteoarthritis associations that survived the multiple testing threshold in the MR analysis, we further conducted colocalization[101] analysis to distinguish causal effects from genomic confounding due to linkage disequilibrium. A colocalization probability more than 80% in this analysis would suggest that the two association signals are likely to colocalize within the test region. Colocalization analysis was applied to both cis and trans pQTLs. For protein and phenotype GWAS lacking suf-

ficient SNP coverage or missing key information (e.g. allele frequency or effect size) in the test region, we conducted a LD check for the sentinel variant for each pQTL against the 30 strongest SNPs in the region associated with the phenotype as an approximate colocalization analysis. r^2 of 0.8 between the sentinel pQTL variant and any of the 30 strongest SNPs associated with the phenotype was used as evidence for approximate colocalization.

Nine protein-osteoarthritis associations showed reliable MR and colocalization evidence (Bonferroni corrected, p-value<$2.77x10^{-06}$ and colocalization probability >70%) for a total of six proteins on seven osteoarthritis phenotypes (**Supplementary Table 9, Methods**). Of the eight protein quantitative trait loci (pQTLs) used as genetic predictors of these six proteins, five were in strong LD with missense variants (r^2>0.8). As missense variants may cause epitope-binding artefacts, we also evaluated the effect of these eight pQTLs on other molecular traits: DNA methylation (meQTL)[102] and gene expression (whole blood eQTLs from eQTLGen and all tissues eQTLs from GTEx; Six of the eight pQTLs were also cis meQTLs and cis eQTLs in the same region, in which four pQTLs are in LD (r2>0.3) with the top meQTL and eQTL in the region (**Supplementary Table 9**).

Pathway analyses

Pathway and gene set enrichment analysis were performed using the Gene2Func function of FUMA (Functional Mapping and Annotation of Genome-Wide Association Studies)[103]. Analysis was performed as described in [103] using the following settings: all known genes and transcripts were included in the background gene set, we included the MHC-region, a minimum of 5 genes needed to overlap with the examined gene sets and the significance threshold was set at FDR<0.05. For the total pathway analysis we used all 205 genes included in **Table 3**, Two genes were not recognized (*C1orf40* and *ICAL2*) and thus were therefore not included in this analysis. We also performed 3 additional stratified pathway analysis on (I) all SNVs associated with only weight bearing joints(85 genes), (II) all SNVs associated with only non-weight-bearing joints(8 genes), and (III) all SNVs associated with both weight bearing and non-weight bearing joints(85 genes).

Genetic correlation

We estimated the genetic correlation between osteoarthritis traits and secondary traits using the cross-trait LD Score regression method as implemented in LDHub[16, 17] in our meta-analyses and summary statistics from traits in the deCODE and UKBB datasets. We used results for about 1.1 million well imputed variants, and for LD informa-

tion we used pre-computed LD scores for European populations. LD scores for the East Asian populations couldn't be calculated as LD Score method requires a sample size of >4000 individuals. To avoid bias due to overlapping samples, we calculated the genetic correlation between meta-analysis of Icelandic, Norwegian and USA samples and UK, Dutch, Estonian and Greek samples (**Supplementary Table 5**). The results of the two analyses were subsequently meta-analysed. For genetic correlations with other traits, we calculated the genetic correlation between a meta-analysis of UK, Dutch, Estonian and Greek samples and the Icelandic GWAS summary statistics for each secondary trait, and also between a meta-analysis of Icelandic, Norwegian and USA samples and UKBB GWAS summary statistics for each secondary trait. The results of the two analyses were subsequently meta-analysed. For the analysis of genetic correlation between the osteoarthritis subtypes, we split the sample-sets of the meta-analysis in two equally sized groups and performed LD score regression between the two groups for each subtype.

Annotation of protein coding variants

For coding SNVs we considered only the following moderate to high impact annotations when weighting genes for prioritisation: transcript_ablation, splice_acceptor_variant, splice_donor_variant, stop_gained, frameshift_variant, stop_lost, start_lost, transcript_amplification, inframe_insertion, inframe_deletion, missense_variant, protein_altering_variant.

Investigation of GWAS meta-analysis signals in the Druggable Genome

We examined the druggability status for the 376 prioritized genes as mentioned in **Table 3**, using the druggable gene set as defined by Finan et al., 2017 Science Translational Med. This druggable genome contained 4,479 genes and it was divided in three tiers of druggable gene sets based on their drug development. Tier 1 included efficacy targets of approved small molecules and biotherapeutic drugs as well as clinical-phase drug candidates. Tier 2 contained targets with known bioactive drug-like small molecule binding partners and those with significant sequence similarity to approved drug targets. Tier 3 contained proteins with more distant similarity to approved drug targets, and members of key druggable gene families not already included in tier 1 or 2. This tier was further subdivided to prioritize those genes that were in proximity (±50 kbp) to a GWAS SNP from GWAS catalogue and had an extracellular location (Tier 3A). The remainder of the genes were assigned to Tier 3B.

Supplementary Data

Supplementary information and supplementary tables can be examined at:

https://drive.google.com/drive/folders/1iqiLVJm1MbeNhefsbLNzQhLNmvHhj5OB?usp=sharing

Acknowledgements

This research was conducted using the UK Biobank Resource under application numbers 9979 and 23359.

The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. GWAS data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. DXA scanning was specifically funded by Wellcome Trust and MRC (grant refs 076467/Z/05/Z and 084632/Z/08/Z).

Regeneron Genetics Center provided funding for the collection of MyCode samples, the generation of genotype data and genotype imputation. Geisinger provided funding for clinical data extraction and genetic association analysis. TRG receives funding from the UK Medical Research Council (MC_UU_00011/4)

TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union, Chronic Disease Research Foundation (CDRF), Zoe Global Ltd and the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. AMV is funded by the NIHR Nottingham BRC.

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 Rotterdam Study GWAS datasets are supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project nr. 050-060-810.

The deCODE genetics osteoarthritis study: We thank the study subjects for their valuable participation.

HUNT study: The genotyping was financed by the National Institute of health (NIH), University of Michigan, The Norwegian Research council, and Central Norway Regional Health Authority and the Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU). The genotype quality control and imputation has been conducted by the K.G. Jebsen center for genetic epidemiology, Department of public health and nursing, Faculty of medicine and health sciences, Norwegian University of Science and Technology (NTNU).

GARP: We thank all study participants of the GARP study. Leiden University Medical Centre, Pfizer Groton, Connecticut, USA and the Dutch Arthritis Society have supported the GARP study. Furthermore, the research leading to these results has received funding from the Biobanking and BioMolecular resources Research Infrastructure the Netherlands (BBMRI-NL) (complementation project CP2013-84), and the Dutch Arthritis Society (DAF_10_1-402). We are indebted to drs. N. Riyazi, J. Bijsterbosch, H.M. Kroon and I. Watt for collection of data.

RAAK: We thank all study participants of the RAAK study. The Leiden University Medical Centre have and are supporting the RAAK. We thank Evelyn Houtman, Enrike van der Linden, Robert van der Wal, Peter van Schie, Shaho Hasan, Maartje Meijer, Daisy Latijnhouwers, Anika Rabeling-Hoogenstraaten, and Geert Spierenburg for their contribution to the collection of the joint tissue. Furthermore, the research leading to these results has received funding from the: TreatOA which is funded by the European Commission framework 7 programme grant 200800. BBMRI Metabolomics Consortium funded by BBMRI-NL, a research infrastructure financed by the Dutch government (NWO, grant nr 184.021.007 and 184033111). Dutch Scientific Research council NWO /ZonMW VICI scheme (91816631/528)

The Leiden Longevity Study has received funding from the European Union's Seventh Framework Programme (FP7/2007-2011) under grant agreement number 259679. This study was financially supported by the Innovation-Oriented Research Program on Genomics (SenterNovem IGE05007), the Centre for Medical Systems Biology and the Netherlands Consortium for Healthy Ageing (grant 050-060-810), all in the framework of the Netherlands Genomics Initiative, Netherlands Organization for Scientific Research (NWO), and by BBMRI-NL, a Research Infrastructure financed by the Dutch government (NWO 184.021.007 and 184033111). This work is supported by the research programme VOILA with project number 457001001, which is (partly) financed by ZonMw (The Netherlands Organization for Health Research and Development).

Competing interests

TRG receives research funding from GlaxoSmithKline and Biogen

The authors U.S., K.S., L.S., and G.T. are employed by deCODE genetics / Amgen Inc. AMV is a consultant for Zoe Global Ltd

References

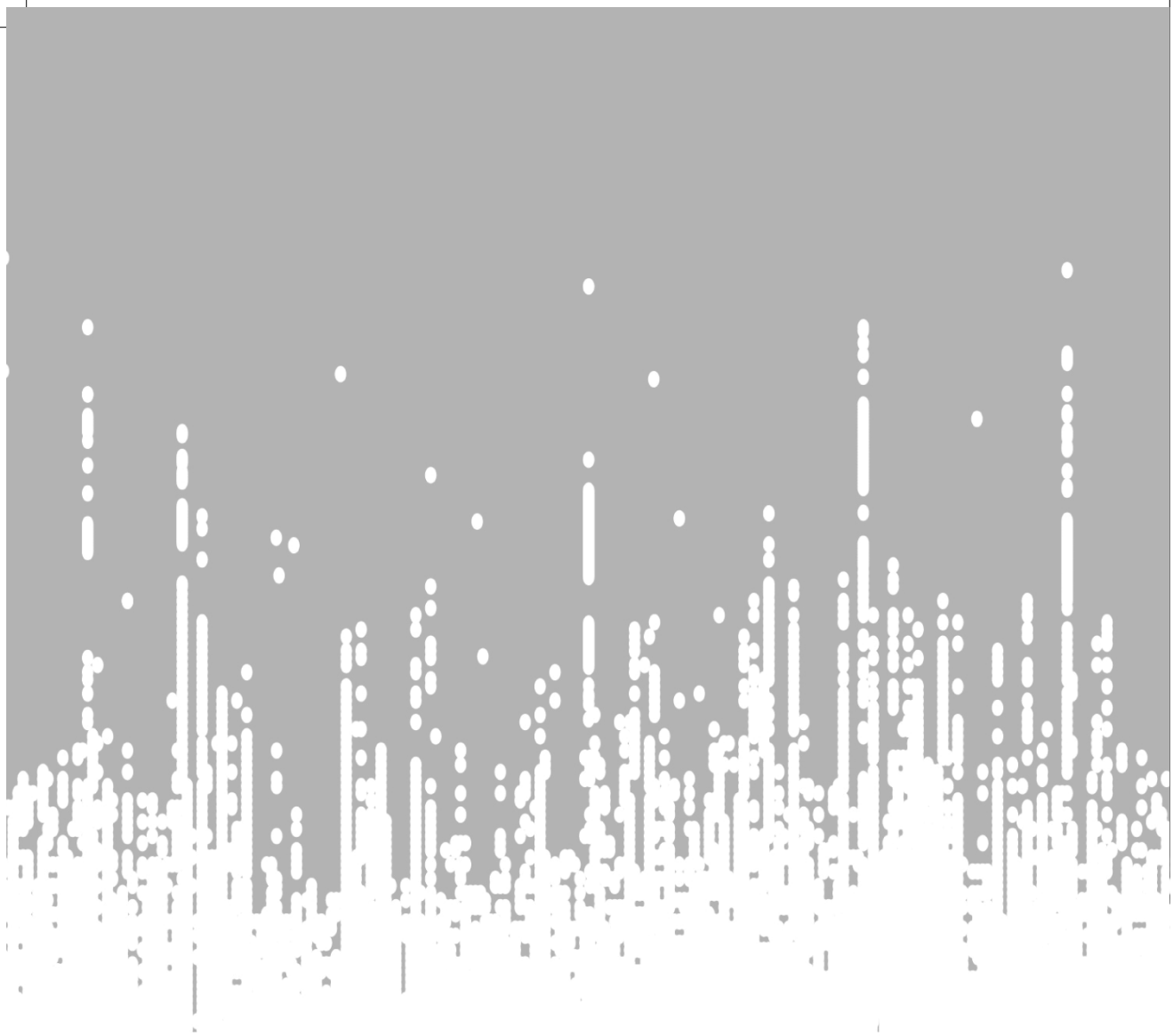
1. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet, 2018. 392(10159): p. 1789-1858.
2. Hunter, D.J., D. Schofield, and E. Callander, The individual and socioeconomic impact of osteoarthritis. Nat Rev Rheumatol, 2014. 10(7): p. 437-41.
3. Lowe, D.B., M.J. Taylor, and S.J. Hill, Associations between multimorbidity and additional burden for working-age adults with specific forms of musculoskeletal conditions: a cross-sectional study. BMC Musculoskelet Disord, 2017. 18(1): p. 135.
4. Hunter, D.J. and S. Bierma-Zeinstra, Osteoarthritis. Lancet, 2019. 393(10182): p. 1745-1759.
5. Spector, T.D. and A.J. MacGregor, Risk factors for osteoarthritis: genetics. Osteoarthritis Cartilage, 2004. 12 Suppl A: p. S39-44.
6. Zengini, E., et al., Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. Nat Genet, 2018. 50(4): p. 549-558.
7. Styrkarsdottir, U., et al., Meta-analysis of Icelandic and UK data sets identifies missense variants in SMO, IL11, COL11A1 and 13 more new loci associated with osteoarthritis, in Nat Genet. 2018: United States. p. 1681-1687.
8. Tachmazidou, I., et al., Whole-Genome Sequencing Coupled to Imputation Discovers Genetic Signals for Anthropometric Traits. Am J Hum Genet, 2017. 100(6): p. 865-884.
9. Mägi, R. and A.P. Morris, GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics, 2010. 11: p. 288.
10. Magi, R., C.M. Lindgren, and A.P. Morris, Meta-analysis of sex-specific genome-wide association studies. Genet Epidemiol, 2010. 34(8): p. 846-53.
11. Pulit, S.L., et al., Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. Hum Mol Genet, 2019. 28(1): p. 166-174.
12. Kichaev, G., et al., Leveraging Polygenic Functional Enrichment to Improve GWAS Power. Am J Hum Genet, 2019. 104(1): p. 65-75.
13. Reynard, L.N. and J. Loughlin, Insights from human genetic studies into the pathways involved in osteoarthritis. Nat Rev Rheumatol, 2013. 9(10): p. 573-83.
14. Sandell, L.J., Etiology of osteoarthritis: genetics and synovial joint development. Nat Rev Rheumatol, 2012. 8(2): p. 77-89.
15. Muley, M.M., et al., Prophylactic inhibition of neutrophil elastase prevents the development of chronic neuropathic pain in osteoarthritic mice. J Neuroinflammation, 2017. 14(1): p. 168.
16. Bulik-Sullivan, B.K., et al., LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet, 2015. 47(3): p. 291-5.
17. Bulik-Sullivan, B., et al., An atlas of genetic correlations across human diseases and traits. Nat Genet, 2015. 47(11): p. 1236-41.
18. Suri, P., et al., Genome-wide meta-analysis of 158,000 individuals of European ancestry identifies three loci associated with chronic back pain. PLoS Genet, 2018. 14(9): p. e1007601.
19. Smits, P., et al., The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. Dev Cell, 2001. 1(2): p. 277-90.
20. Schaible, H.G., Osteoarthritis pain. Recent advances and controversies. Curr Opin Support Palliat Care, 2018. 12(2): p. 148-153.
21. Dimitroulas, T., et al., Neuropathic pain in osteoarthritis: a review of pathophysiological mechanisms and implications for treatment. Semin Arthritis Rheum, 2014. 44(2): p. 145-54.
22. Fu, K., S.R. Robbins, and J.J. McDougall, Osteoarthritis: the genesis of pain. Rheumatology (Oxford), 2018. 57(suppl_4): p. iv43-iv50.
23. Kidd, B., Mechanisms of pain in osteoarthritis. Hss j, 2012. 8(1): p. 26-8.

24. Hsia, A.W., et al., Osteophytes and fracture calluses share developmental milestones and are diminished by unloading. J Orthop Res, 2018. 36(2): p. 699-710.
25. Maurano, M.T., et al., Systematic localization of common disease-associated variation in regulatory DNA. Science, 2012. 337(6099): p. 1190-5.
26. Kundaje, A., et al., Integrative analysis of 111 reference human epigenomes. Nature, 2015. 518(7539): p. 317-30.
27. Brandt, K.D., et al., Yet more evidence that osteoarthritis is not a cartilage disease, in Ann Rheum Dis. 2006: England. p. 1261-4.
28. McKee, T.J., et al., Extracellular matrix composition of connective tissues: a systematic review and meta-analysis. Sci Rep, 2019. 9(1): p. 10542.
29. Uitto, J., D.R. Olsen, and M.J. Fazio, Extracellular matrix of the skin: 50 years of progress. J Invest Dermatol, 1989. 92(4 Suppl): p. 61s-77s.
30. Genetic effects on gene expression across human tissues. Nature, 2017. 550(7675): p. 204-13.
31. Golledge, J., et al., The role of tenascin C in cardiovascular disease. Cardiovasc Res, 2011. 92(1): p. 19-28.
32. Yasuda, M., et al., Characterization of tenascin-C as a novel biomarker for asthma: utility of tenascin-C in combination with periostin or immunoglobulin E. Allergy Asthma Clin Immunol, 2018. 14: p. 72.
33. Sun, B.B., et al., Genomic atlas of the human plasma proteome. Nature, 2018. 558(7708): p. 73-79.
34. Suhre, K., et al., Connecting genetic risk to disease end points through the human blood plasma proteome. Nat Commun, 2017. 8: p. 14357.
35. Folkersen, L., et al., Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. PLoS Genet, 2017. 13(4): p. e1006706.
36. Yao, C., et al., Genome-wide mapping of plasma protein QTLs identifies putatively causal genes and pathways for cardiovascular disease. Nat Commun, 2018. 9(1): p. 3268.
37. Emilsson, V., et al., Co-regulatory networks of human serum proteins link genetics to disease. Science, 2018. 361(6404): p. 769-773.
38. den Hollander, W., et al., Genome-wide association and functional studies identify a role for matrix Gla protein in osteoarthritis of the hand. Ann Rheum Dis, 2017. 76(12): p. 2046-2053.
39. Shepherd, C., et al., Expression analysis of the osteoarthritis genetic susceptibility mapping to the matrix Gla protein gene MGP. Arthritis Res Ther, 2019. 21(1): p. 149.
40. Balakrishnan, L., et al., Proteomic analysis of human osteoarthritis synovial fluid. Clin Proteomics, 2014. 11(1): p. 6.
41. Jia, W., H. Li, and Y.W. He, The extracellular matrix protein mindin serves as an integrin ligand and is critical for inflammatory cell recruitment. Blood, 2005. 106(12): p. 3854-9.
42. He, Y.W., et al., The extracellular matrix protein mindin is a pattern-recognition molecule for microbial pathogens. Nat Immunol, 2004. 5(1): p. 88-97.
43. Huang, Z.Y., et al., Both systemic and local lipopolysaccharide (LPS) burden are associated with knee OA severity and inflammation. Osteoarthritis Cartilage, 2016. 24(10): p. 1769-1775.
44. Unger, S., et al., Phenotypic features of carbohydrate sulfotransferase 3 (CHST3) deficiency in 24 patients: congenital dislocations and vertebral changes as principal diagnostic features. Am J Med Genet A, 2010. 152a(10): p. 2543-9.
45. Song, Y.Q., et al., Lumbar disc degeneration is linked to a carbohydrate sulfotransferase 3 variant. J Clin Invest, 2013. 123(11): p. 4909-17.
46. Superti-Furga, A. and S. Unger, CHST3-Related Skeletal Dysplasia, in GeneReviews(®), M.P. Adam, et al., Editors. 1993, University of Washington, Seattle Copyright © 1993-2020, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.: Seattle (WA).
47. Alman, B.A., The role of hedgehog signalling in skeletal health and disease. Nat Rev Rheumatol, 2015. 11(9): p. 552-60.

48. Zhou, Y., et al., Wnt/beta-catenin Signaling in Osteoarthritis and in Other Forms of Arthritis. Curr Rheumatol Rep, 2017. 19(9): p. 53.
49. Ullah, A., et al., Homozygous sequence variants in the WNT10B gene underlie split hand/foot malformation. Genet Mol Biol, 2018. 41(1): p. 1-8.
50. Yu, P., et al., Mutations in WNT10B Are Identified in Individuals with Oligodontia. Am J Hum Genet, 2016. 99(1): p. 195-201.
51. Kantaputra, P.N., et al., WNT10B mutations associated with isolated dental anomalies. Clin Genet, 2018. 93(5): p. 992-999.
52. Jiao, S., et al., VGLL4 targets a TCF4-TEAD4 complex to coregulate Wnt and Hippo signalling in colorectal cancer. Nat Commun, 2017. 8: p. 14058.
53. Lin, Z., et al., Acetylation of VGLL4 Regulates Hippo-YAP Signaling and Postnatal Cardiac Growth. Dev Cell, 2016. 39(4): p. 466-79.
54. Dupont, S., et al., Role of YAP/TAZ in mechanotransduction. Nature, 2011. 474(7350): p. 179-83.
55. Low, B.C., et al., YAP/TAZ as mechanosensors and mechanotransducers in regulating organ size and tumor growth. FEBS Lett, 2014. 588(16): p. 2663-70.
56. Vincent, T.L. and A.K.T. Wann, Mechanoadaptation: articular cartilage through thick and thin. J Physiol, 2019. 597(5): p. 1271-1281.
57. Zhao, Z., et al., Mechanotransduction pathways in the regulation of cartilage chondrocyte homoeostasis. J Cell Mol Med, 2020. 24(10): p. 5408-5419.
58. Kong, Y.Y., et al., Activated T cells regulate bone loss and joint destruction in adjuvant arthritis through osteoprotegerin ligand. Nature, 1999. 402(6759): p. 304-9.
59. Remuzgo-Martínez, S., et al., Expression of osteoprotegerin and its ligands, RANKL and TRAIL, in rheumatoid arthritis. Sci Rep, 2016. 6: p. 29713.
60. Papadaki, M., et al., New Insights for RANKL as a Proinflammatory Modulator in Modeled Inflammatory Arthritis. Front Immunol, 2019. 10: p. 97.
61. Kohli, S.S. and V.S. Kohli, Role of RANKL-RANK/osteoprotegerin molecular complex in bone remodeling and its immunopathologic implications. Indian J Endocrinol Metab, 2011. 15(3): p. 175-81.
62. Zeyer, K.A. and D.P. Reinhardt, Fibrillin-containing microfibrils are key signal relay stations for cell function. J Cell Commun Signal, 2015. 9(4): p. 309-25.
63. Putnam, E.A., et al., Fibrillin-2 (FBN2) mutations result in the Marfan-like disorder, congenital contractural arachnodactyly. Nat Genet, 1995. 11(4): p. 456-8.
64. Gonzalez, A.M., et al., Ecrg4 expression and its product augurin in the choroid plexus: impact on fetal brain development, cerebrospinal fluid homeostasis and neuroprogenitor cell response to CNS injury. Fluids Barriers CNS, 2011. 8(1): p. 6.
65. Beecham, G.W., et al., Genome-wide association meta-analysis of neuropathologic features of Alzheimer's disease and related dementias. PLoS Genet, 2014. 10(9): p. e1004606.
66. Ferrari, N., et al., Dickkopf-3 links HSF1 and YAP/TAZ signalling to control aggressive behaviours in cancer-associated fibroblasts. Nat Commun, 2019. 10(1): p. 130.
67. Cubelos, B., et al., Cux1 and Cux2 regulate dendritic branching, spine morphology, and synapses of the upper layer neurons of the cortex. Neuron, 2010. 66(4): p. 523-35.
68. Kedinger, V. and A. Nepveu, The roles of CUX1 homeodomain proteins in the establishment of a transcriptional program required for cell migration and invasion. Cell Adh Migr, 2010. 4(3): p. 348-52.
69. Lizarraga, G., et al., Studies on the role of Cux1 in regulation of the onset of joint formation in the developing limb. Dev Biol, 2002. 243(1): p. 44-54.
70. Xiong, Q., et al., A novel de novo mutation in COL2A1 leading to spondyloepiphyseal dysplasia congenita in a Chinese family. Hum Genome Var, 2018. 5: p. 17059.
71. Wilkin, D.J., et al., Small deletions in the type II collagen triple helix produce kniest dysplasia. Am J Med Genet, 1999. 85(2): p. 105-12.

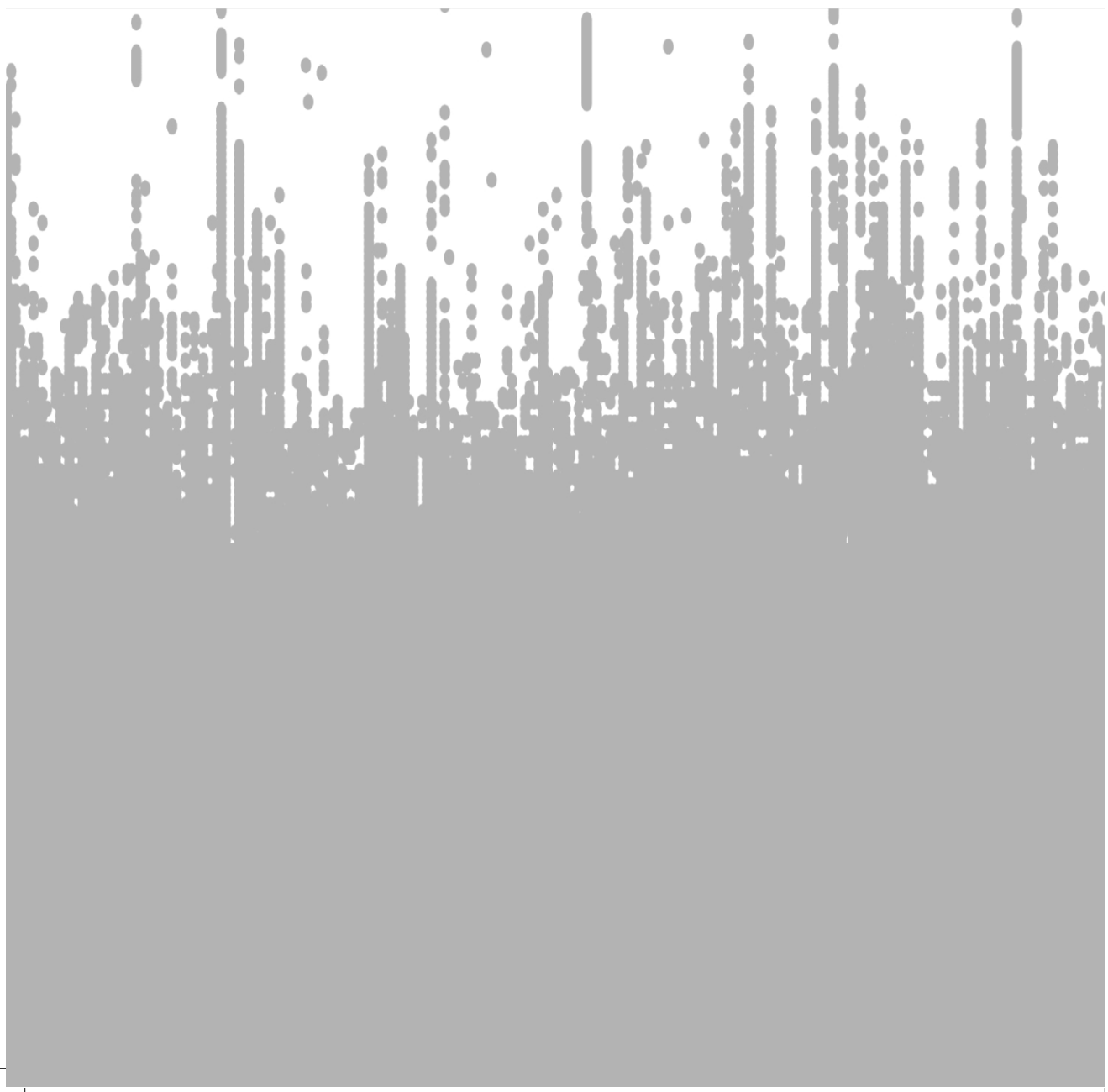
72. Holt, M., et al., Identification of SNAP-47, a novel Qbc-SNARE with ubiquitous expression. J Biol Chem, 2006. 281(25): p. 17076-83.
73. Jones, D.K., et al., Localization and functional consequences of a direct interaction between TRIOBP-1 and hERG proteins in the heart. J Cell Sci, 2018. 131(6).
74. Park, S., et al., Emerging roles of TRIO and F-actin-binding protein in human diseases, in Cell Commun Signal. 2018.
75. Finan, C., et al., The druggable genome and support for target identification and validation in drug development. Sci Transl Med, 2017. 9(383).
76. Anderson, D.M., et al., A homologue of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function. Nature, 1997. 390(6656): p. 175-9.
77. Roh, Y.S., J. Dequeker, and J.C. Mulier, Cortical bone remodeling and bone mass in primary osteoarthrosis of the hip. Invest Radiol, 1973. 8(4): p. 351-4.
78. Kerkhof, H.J., et al., Recommendations for standardization and phenotype definitions in genetic studies of osteoarthritis: the TREAT-OA consortium. Osteoarthritis Cartilage, 2011. 19(3): p. 254-64.
79. Winkler, T.W., et al., Quality control and conduct of genome-wide association meta-analyses. Nat Protoc, 2014. 9(5): p. 1192-212.
80. Willer, C.J., Y. Li, and G.R. Abecasis, METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics, 2010. 26(17): p. 2190-1.
81. Li, M.X., et al., Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. Hum Genet, 2012. 131(5): p. 747-56.
82. Purcell, S., et al., PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet, 2007. 81(3): p. 559-75.
83. Yang, J., et al., GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet, 2011. 88(1): p. 76-82.
84. Yang, J., et al., Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet, 2012. 44(4): p. 369-75, s1-3.
85. Wakefield, J., Bayes factors for genome-wide association studies: comparison with P-values. Genet Epidemiol, 2009. 33(1): p. 79-86.
86. Ramos, Y.F., et al., Genes involved in the osteoarthritis process identified through genome wide expression analysis in articular cartilage; the RAAK study. PLoS One, 2014. 9(7): p. e103056.
87. Coutinho de Almeida, R., et al., RNA sequencing data integration reveals an miRNA interactome of osteoarthritis cartilage. Ann Rheum Dis, 2019. 78(2): p. 270-277.
88. den Hollander, W., et al., Annotating Transcriptional Effects of Genetic Variants in Disease-Relevant Tissue: Transcriptome-Wide Allelic Imbalance in Osteoarthritic Cartilage. Arthritis Rheumatol, 2019. 71(4): p. 561-570.
89. Tuerlings, M., et al., RNA sequencing reveals interacting key determinants of osteoarthritis acting in subchondral bone and articular cartilage. 2020: p. 2020.03.13.969386.
90. Tachmazidou, I., et al., Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. Nat Genet, 2019. 51(2): p. 230-236.
91. Steinberg, J., et al., Integrative epigenomics, transcriptomics and proteomics of patient chondrocytes reveal genes and pathways involved in osteoarthritis. Sci Rep, 2017. 7(1): p. 8935.
92. Ritchie, M.E., et al., limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res, 2015. 43(7): p. e47.
93. Law, C.W., et al., voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol, 2014. 15(2): p. R29.
94. Chou, C.H., et al., Genome-wide expression profiles of subchondral bone in osteoarthritis. Arthritis Res Ther, 2013. 15(6): p. R190.

95. Chou, C.H., et al., Direct assessment of articular cartilage and underlying subchondral bone reveals a progressive gene expression change in human osteoarthritic knees. Osteoarthritis Cartilage, 2013. 21(3): p. 450-61.
96. Finger, J.H., et al., The mouse gene expression database: New features and how to use them effectively. Genesis, 2015. 53(8): p. 510-22.
97. Yates, A.D., et al., Ensembl 2020. Nucleic Acids Res, 2020. 48(D1): p. D682-d688.
98. Mortier, G.R., et al., Nosology and classification of genetic skeletal disorders: 2019 revision. Am J Med Genet A, 2019. 179(12): p. 2393-2419.
99. Auton, A., et al., A global reference for human genetic variation. Nature, 2015. 526(7571): p. 68-74.
100. The Genotype-Tissue Expression (GTEx) project. Nat Genet, 2013. 45(6): p. 580-5.
101. Giambartolomei, C., et al., Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet, 2014. 10(5): p. e1004383.
102. Gaunt, T.R., et al., Systematic identification of genetic influences on methylation across the human life course. Genome Biol, 2016. 17: p. 61.
103. Watanabe, K., et al., Functional mapping and annotation of genetic associations with FUMA. Nat Commun, 2017. 8(1): p. 1826.



Chapter 5

OSTEOARTHRITIS AND THE MICROBIOME



Chapter 5.1

Diversity, compositional and functional differences between gut microbiota of children and adults

Radjabzadeh, D., **Boer, C.G.**, Beth, S.A., van der Wal, P., Kiefte-De Jong, J.C., Jansen, M.A.E., Konstantinov, S.R., Peppelenbosch, M.P., Hays, J.P., Jaddoe, V.W.V., Ikram, M.A., Rivadeneira, F., van Meurs, J.B.J., Uitterlinden, A.G., Medina-Gomez, C., Moll, H.A., & Kraaij, R.

Published in: Sci Rep. 2020;10(1):1040. Published 2020 Jan 23.

Abstract

The gut microbiota has been shown to play diverse roles in human health and disease although the underlying mechanisms have not yet been fully elucidated. Large cohort studies can provide further understanding into inter-individual differences, with more precise characterization of the pathways by which the gut microbiota influences human physiology and disease processes. Here, we aimed to profile the stool microbiome of children and adults from two population-based cohort studies, comprising 2,111 children in the age-range of 9 to 12 years (the Generation R Study) and 1,427 adult individuals in the range of 46 to 88 years of age (the Rotterdam Study). For the two cohorts, 16S rRNA gene profile datasets derived from the Dutch population were generated. The comparison of the two cohorts showed that children had significantly lower gut microbiome diversity. Furthermore, we observed higher relative abundances of genus Bacteroides in children and higher relative abundances of genus Blautia in adults. Predicted functional metagenome analysis showed an overrepresentation of the glycan degradation pathways, riboflavin (vitamin B2), pyridoxine (vitamin B6) and folate (vitamin B9) biosynthesis pathways in children. In contrast, the gut microbiome of adults showed higher abundances of carbohydrate metabolism pathways, beta-lactam resistance, thiamine (vitamin B1) and pantothenic (vitamin B5) biosynthesis pathways. A predominance of catabolic pathways in children (valine, leucine and isoleucine degradation) as compared to biosynthetic pathways in adults (valine, leucine and isoleucine biosynthesis) suggests a functional microbiome switch to the latter in adult individuals. Overall, we identified compositional and functional differences in gut microbiome between children and adults in a population-based setting. These microbiome profiles can serve as reference for future studies on specific human disease susceptibility in childhood, adulthood and specific diseased populations.

Introduction

The human gut microbiome is dynamic, shaped by multiple factors and has been shown to play an important role in human health. Several studies have reported an association between alterations in the composition of the gut microbiome and various gastrointestinal (GI)[1-9] and non-GI[10-15] disease conditions in both children and adults. These changes in composition of the microbiome (known as dysbiosis) and their associations with health and disease have led to an increased interest in the complex microbial community of the gut[16,17]. However, our understanding of the relationship between the gut microbiome and health and disease still remains superficial at best. A fundamental issue hampering progress in this respect is that our lack of knowledge on the composition and variation of the gut microbiome across the human lifespan remains limited. So far, most microbiome studies have been relatively small in terms of sample size and have usually focused on specific diseases and phenotypes, with most publications having used a case-control study design, what makes the analyses more vulnerable to methodological challenges. A better understanding of the variation in microbiome composition (from stool or other sources) and its role in the aetiology of chronic diseases can be achieved through the study of large and well-characterized population-based cohorts, which generate rich information on many different physiological parameters, disease status, medication use, dietary intake, and other layers of "omics" data from individual participants.

Up to now, microbiome studies have focused mainly on adult populations, whereas it has been shown that the intestinal microbiome undergoes dynamic changes in diversity and composition during the human lifespan and particularly during development; with the most substantial changes believed to occur throughout childhood[18-20]. Interestingly, emerging data suggest that early alterations in the gut microbiome are associated with an increased risk of developing diseases later in childhood and adulthood e.g., asthma[21,22] and Crohn's disease[23]. These studies were, however, limited by low sample sizes. If the field of microbiota dynamics is to move forward, it requires studies in well-characterized general population cohorts of different ages, including studies investigating paediatric cohorts. So far, a complete description of diversity, compositional and functional differences between children and adults in a large, homogeneous and population-based cohorts have not been reported.

The above-mentioned considerations prompted us to profile the stool microbiome composition of children and adults within two large, independent and extensively characterized population-based cohorts: the Generation R Study (GenR, visit at 9 years) and the Rotterdam Study (RS, sub-cohort RSIII-2). In this publication, we report our

findings on the differences in gut microbiota between 2,111 children aged 9-12 years and 1,427 adults >40 years of age living in the same city with similar urban surroundings.

Methods

Study populations and sample collection

The Generation R Study (GenR) is a population-based prospective multi-ethnic pregnancy cohort study from foetal life until young adulthood conducted in the city of Rotterdam[24]. The study was designed to identify early environmental and genetic factors and causal pathways underlying normal and abnormal growth during development during childhood. GenR recruited 9,749 children undergoing several rounds of follow-up after birth. Stool sample collection started in 2012 at a mean age of 9.8 years (SD: 0.32). Ethics approval was obtained from the Medical Ethical Committee of Erasmus MC (MEC-2012-165) and written informed consent was obtained from all participants' parents. All methods were performed in accordance with the Declaration of Helsinki.

The Rotterdam Study (RS) is a prospective population-based cohort study established in 1990 to study determinants of disease and disability in Dutch adult individuals. The original design and updates of this study have been described in detail[25]. RS consists of four sub-cohorts and comprises approximately 18,000 inhabitants of the Ommoord suburb in Rotterdam (which is predominantly populated by individuals of European ethnicity (about 96%)), aged ≥40 years. The collection of fecal samples started in 2012 among the RS-III sub-cohort comprising 3,932 participants. This study was approved by the Medical Ethical Committee of Erasmus MC (MEC-02-1015) and by the Ministry of Health, Welfare and Sport of the Netherlands. All subjects provided written consent prior to participation in the study. All methods were performed in accordance with the Declaration of Helsinki.

Stool samples were collected at home by the participants using a Commode Specimen Collection System (Covidien, Mansfield, MA). An aliquot of approximately 1 g was transferred to a 25 × 76 mm feces collection tube (Minigrip Nederland, Lelystad, The Netherlands) without preserving agent included and sent through regular mail to the Erasmus MC. A short questionnaire addressing date and time of defecation, current or recent antibiotics use (past year), recent probiotics use (past 3 days), and recent travel activities (past month), was filled out by the participants and included in the package.

Upon arrival at Erasmus MC, samples were recorded and stored at −20 °C. The only modification in the collection protocol for the GenR cohort (as compared to RS) was that, in case of delay, the samples were stored by participants at 4 °C (home fridge) before mailing to Erasmus MC. This modification allowed to better preserve samples that were produced in the evening or during the weekend.

DNA isolation

Stool samples from the two cohorts were randomly taken from the −20 °C freezer and allowed to thaw for 10 minutes at room temperature prior to DNA isolation. Samples with inconsistent or lack of information on sample production and samples in which mold growth was observed were excluded (**Supplementary Figure 1**). An aliquot of approximately 300 mg was homogenized in stool stabilizing buffer according to the manufacturer's protocol (Arrow Stool DNA; Isogen Life Science, De Meern, The Netherlands). Homogenized samples were bead beated in Lysing Matrix B tubes containing 0.1 mm silica beads (MP Biomedicals, LLC, Bio Connect Life Sciences BV, Huissen, The Netherlands) using the MagNA Lyser instrument (Roche Diagnostics, Almere, The Netherlands) at 7,000 rpm for 45 seconds. Samples were then centrifuged at 6,000 × g for 5 min and 0.5 ml of supernatant was subjected to automated DNA isolation (Arrow; DiaSorin S.P.A., Saluggia, Italy) according to the manufacturer's protocol using setting 'Stool DNA 2.0' in batches of 12 samples per run. DNA concentration was measured using Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific, Waltham, MA) and DNA was stored at −20 °C.

16S rRNA gene sequencing

The V3 and V4 variable regions of the 16 S rRNA gene were amplified using the 309F-806R primer pair and dual indexing (12 base pairs (bp) each on the forward and reverse primers) as previously described[26]. Amplicons were normalized using the SequalPrep Normalization Plate kit (Thermo Fischer Scientific) and pooled. The pools were purified prior to sequencing using Agencourt AMPure XP (Beckman Coulter Life Science, Indianapolis, IN) and the amplicon size and quantity of the pools were assessed on the LabChip GX (PerkinElmer Inc., Groningen, The Netherlands). PhiX Control v3 library (Illumina Inc., San Diego, CA) was spiked into (~10%) the pooled amplicon libraries and each pool was sequenced on an Illumina MiSeq sequencer (MiSeq Reagent Kit v3, 2 × 300 bp) at an average depth of 50,000 read-pairs per sample.

Data pre-processing, OTU picking and quality control

Phylogenetic multi-sample profiling was performed using an in-house developed analysis pipeline (microRapTor) based on QIIME (version 1.9.0)[27] and UPARSE (version 8.1)[28] software packages. Briefly, index sequences (12 bp) were removed from each read and concatenated to generate a unique index of 24 bp for each read-pair. Spacer and primer sequences were removed using TAGCleaner (version 0.16)[29]. Paired reads were merged using PEAR (version 0.9.6)[30] with the following settings: minimum overlap of 10 bp (default) and an average read quality phred-score of 20 over a 30 bp sliding window. Merged reads shorter than 200 bp were discarded. Reads were de-multiplexed using QIIME including extra quality filtering steps: merged reads were truncated before three consecutive low-quality bases, ambiguous bases were not allowed. Chimeric reads were removed using UCHIME (version 8.1)[28]. Duplicate samples, samples with less than 10,000 reads, and samples from participants that have used antibiotics (self-reported) in one year prior to sample production were excluded (**Supplementary Figure 1**). The 16 S sequence reads of the remaining samples (2,214 for GenR and 1,544 for RS) were randomly subsampled at 10,000 reads per sample (after rarefaction analysis). Combined reads of all samples, in each cohort separately, were clustered into operational taxonomic units (OTUs) using UPARSE at a minimum cluster identity of 97%. The representative read from each OTU was then mapped to the SILVA rRNA database version 128[31] using RDP Naïve Bayesian Classifier version 2.12[32]. OTUs containing less than 40 reads were removed as described by Benson et al[33]. This threshold was established based on the correlation analysis of OTU tables of 5 pairs of technical replicates, of which DNA was amplified, sequenced and profiled twice (**Supplementary Figure 2**). The sequence data was then analyzed for α-diversity metrics (Shannon diversity Index, species richness and Inverse Simpson Index). Final OTU filtering was performed by removing OTUs with a total read count less than 0.005% of all reads and OTUs observed in less than 1% of the total number of samples of each cohort as described previously[34]. The final OTU table was divided into 5 sub-tables at different taxonomic levels (in QIIME environment): phylum, class, order, family, and genus.

Statistical analyses

All statistical analyses were performed in R[35] using vegan[36], phyloseq[37] and MaAsLin[38] packages. As MaAsLin performs paralleled multiple analyses, q-value < 0.05 (false discovery rate (FDR) multiple testing corrected) was used as significance threshold. All MaAsLin models were adjusted for technical covariates and other confounders as described in the sections below.

Technical covariates in the initial stool 16S datasets of the two cohorts

After generating the initial 16S datasets of both the RS and GenR cohorts, the effects of collection time (time in mail; TIM) and the yield of DNA isolation runs (Batch; **Supplementary Figure 3**) on the 16S profiles were assessed. To identify the impact of these technical covariates on the α-diversity and overall profiles, Shannon α-diversity and Bray-Curtis dissimilarity metrics were calculated in each cohort using 'diversity' and 'vegdist' functions in vegan, respectively. The effects of the covariates on α-diversity were assessed by linear regression 'lm' function in package stats. The basic model was adjusted for age and sex. Each of the two technical covariates were stepwise added in order to identify its contribution to the model. Then, likelihood ratio test was used to compare these nested models. Additionally, in GenR self-reported ethnicity was added to the model to inspect and compare the effect of this covariate. Subsequently, permutation analysis of variance (PERMANOVA) was performed to inspect the global effects of these technical covariates on the overall profiles using the 'adonis' function in vegan. Furthermore, for TIM, additive general linear regression analyses of single genus-level OTUs were performed in MaAsLin. For this, we used sub-sampled datasets that included only samples up to a certain TIM (up to 7 days). Regression analyses in MaAsLin were performed after arcsine square root transformation of the relative abundance data and were adjusted for age, sex and BMI to assess the effect of TIM exclusively.

Validation of the final stool 16S datasets of GenR and RS

Final datasets were constructed by excluding samples that had been in the mail for 3 (in RS) or 5 (for GenR) days (see Results section). Average relative abundances of the six major phyla were compared to those of other large cohorts. GenR profiles were compared to those reported by the Copenhagen Prospective Study on Asthma in Childhood (COPSAC) cohort (Copenhagen, Denmark; n = 156; age range = 4–6 years[39]), and RS profiles were compared to those reported by the Dutch LifeLines-DEEP (LLD) cohort (Groningen, the Netherlands; n = 1,135; age range = 20–90 years[40,41]) and the Belgian Flemish Gut Flora Project (FGFP) cohort (Leuven, Belgium; n = 1,106; age range = 20–80 years[42]). In addition, we performed association analyses of BMI with Shannon α-diversity and single OTUs at genus level to confirm this well-established association11,13,[40] in our cohorts, and by comparing associated OTUs observed in our cohorts to those reported by the above-mentioned cohorts.

The LifeLines DEEP (LLD) study was approved by the ethics committee of the University Medical Centre Groningen, document number METC UMCG LLD: M12.113965 and all methods were performed in accordance with the Declaration of Helsinki. All participants signed an informed consent form prior to study enrolment.

Flemish Gut Flora Project (FGFP) procedures were approved by the medical ethics committee of the University of Brussels/Brussels University Hospital (approval 143201215505, 5/12/2012) and all methods were performed in accordance with the Declaration of Helsinki. A declaration concerning the FGFP privacy policy was submitted to the Belgian Commission for the Protection of Privacy. Participants were recruited through repeated announcements in print, audiovisual, and social media as well as through the FGFP website, where volunteers could enroll from January 2013 onwards

The Copenhagen Prospective Study on Asthma in Childhood (COPSAC) was approved by all relevant authorities including the Danish Ethical Committee (H-B-2008-093), and the Danish Data Protection Agency (2015-41-3696) and all methods were performed in accordance with the Declaration of Helsinki. Both parents provided informed consent prior to participation.

Comparing the gut microbiome profiles and functions between children and adults

To compare gut microbiome profiles between children and adults, we selected samples from European participants. Ethnic background of GenR participants was assessed based on self-reported country of birth of four grandparents as described elsewhere[43]. Ethnic background of RS participants was assessed based on self-reported ethnic backgrounds of the four grandparents. We selected European participants by combining all European countries together with Americans, Oceanics and North Africans as evaluated elsewhere[44]. In short, the participant was deemed to be of non-Dutch origin if one parent was born abroad. If both parents were born abroad, the country of birth of the participant's mother defined the ethnic background. Ethnicities of the parents were derived from the grandparents using the same protocol. The different ethnicities based on the parents country of birth or ethnic background were narrowed to three main ancestry groups: Europeans, including all European countries, together with Americans, Oceanics, and North Africans (with parents from Algeria, Egypt, Libya, Morocco, Sudan, Tunisia, and Western Sahara); Africans, including Sub-Saharan Africans, Dutch Antilleans, and Surinamese Creoles; and Asians, including all Asian countries and Surinamese Hindustanis. Reads were subsampled at 10,000 reads per sample and pooled. Classification was performed (as described above) in one combined run. Shannon α-diversities were calculated and tested for significant differences between the two cohorts using Wilcoxon signed-rank test (1000 permutations). Between-sample Bray-Curtis dissimilarities were calculated and principal coordinate analysis (PCoA) was performed and tested for significant differences between the two cohorts using PERMANOVA. During the beta diversity analysis we adjusted for DNA isolation batch and time in mail since these were

the technical covariates. Linear regression models intending to determine significantly different genera between both cohorts were performed in MaAsLin. During analysis we adjusted for BMI, sex, technical covariates (TIM and Batch), and multiple testing by FDR (q-value < 0.05).

To compare the functional metagenome of gut bacteria between children and adults, we used the PICRUSt (v.1.1.0) tool[45] to obtain predicted bacterial functions. HUMAnN2 (v0.99) was used to identify the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways[46]. For the identification of the specific pathway biomarkers distinguishing the gut microbiome of the children from those of adults, we performed a linear discriminant effect size (LEfSe) analysis[47] with the default settings: α (ANOVA) = 0.05 and logarithmic LDA (linear discriminant analysis) score = 2.0.

Results

Stool microbiota 16S rRNA data generation in the GenR and RS cohorts

Selection of subjects

For GenR 4,959, and for RS 2,440 participants were invited to provide a stool sample. In total, 2,921 (response rate = 59%) and 1,691 (response rate = 69%) were received at Erasmus MC for GenR and RS stool samples, respectively (**Supplementary Figure 1**). Excluding antibiotic users resulted in exclusion of 196 samples from GenR and 7 samples from RS. In GenR, individuals who used antibiotics in the last year had a significantly lower microbiome diversity and altered microbiome composition (**Supplementary Table 1**). For probiotic use and recent travelling activity outside the Netherlands we could not detect significant effects on diversity or composition in the two cohorts (**Supplementary Table 1**). After quality control (**Supplementary Figure 1**), 16S rRNA data of 2,214 subjects in the initial dataset of GenR and 1,544 subjects in the initial dataset of RS were included.

Assessing the influence of technical co-variates and sample exclusion

Since stool samples were collected at ambient temperature via regular mail, we assessed the effects of time in the mail (TIM) on the microbiome profiles. Furthermore, since DNA-yields varied across different DNA isolation batches (**Supplementary Figure 3**), we assessed its effect on the microbiome profiles as well. Upon longer periods of TIM (except for day 7 in GenR), an increase in the relative abundance of phylum *Proteobacteria* was observed in the average profiles of both cohorts (**Supplementary Figure 4**). For RS only, an increase in phylum Bacteroidetes was observed between days 6 and

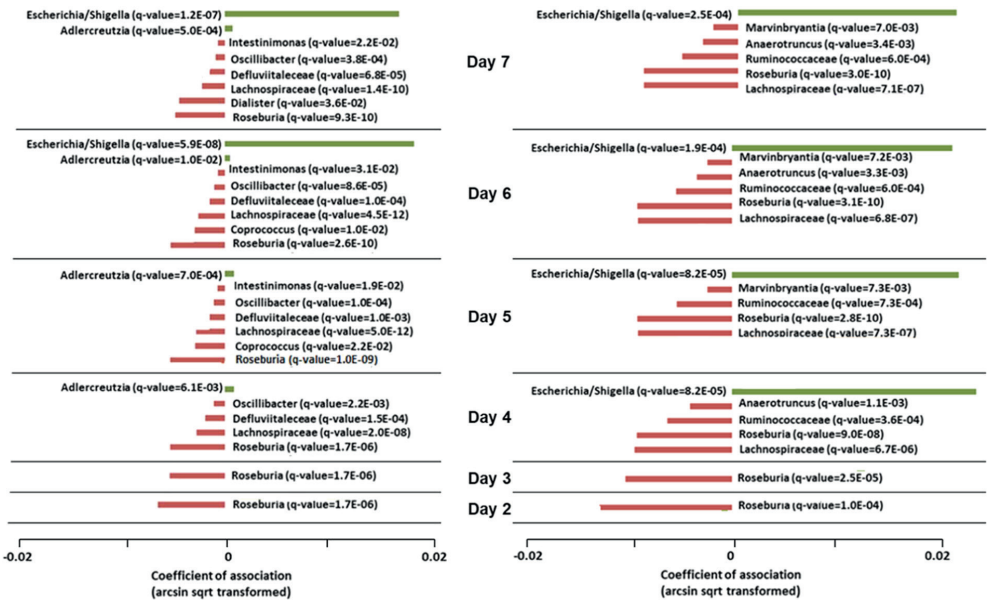
7. In contrast, we did not observe any substantial changes in phylum-level profiles with respect to the DNA-Batch variable (**Supplementary Figure 4**). Moreover, TIM had a small but significant negative effect on α-diversity in GenR and RS (beta = −0.02 alpha units/day, p-value = $9.3x10^{-03}$ and beta = −0.03, p -value= $4.6x10^{-03}$, respectively). Again, we did not observe an effect of Batch on α-diversity. Correlations of TIM and Batch with overall composition (β-diversity) in both cohorts were small but significant (R^2 = 0.004, p-value = 0.001 and R^2 = 0.002, p-value = 0.005 for TIM and R^2 = 0.01, p-value = 0.001 and R2 = 0.005, p-value = 0.001 for Batch in GenR and RS, respectively). In taxonomy-based analyses (MaAsLin) at genus-level of both datasets for TIM, we observed increased abundances of genus *Escherichia/Shigella* upon prolonged times in the mail (**Figure 1**). However, these differences were only significant after 3 days in the RS cohort and after 5 days in the GenR cohort (**Figure 1**). Furthermore, we observed smaller decreases in the abundances of a number of genera, including *Roseburia* and *Coprococcus*, upon prolonged TIM.

In order to further evaluate the importance of TIM and Batch effects, we added them to linear models for the analysis of well-established association of α-diversity with BMI[11,13,40] in both GenR and RS. Although, the estimates (betas) o f the exposure (i.e., BMI) remained similar after various levels of adjustment, the model including sex, age and TIM (model 3) as co-variates provided a better fit to the data than the model including only sex and age (model 2; likelihood ratio test: p-value = $5.8x10^{-03}$ and p -value= $5.5x10^{-03}$ for GenR and RS, respectively). Including Batch, within model 1 (model 2), resulted in a better fit to the data in RS only (**Table 1**; p-value = $1.9x10^{-04}$).

Table 2: association of BMI with Shannon diversity in the 16S datasets of GenR and RS cohorts.

Model	Linear model: α-diversity ~BMI + covar.	R^2	Estimate	P-value	R^2	Estimate	P-value
0	BMI	0.008	-0.031	$9.6x10^{-06}$	0.015	-0.021	$9.9x10^{-07}$
1	BMI + sex	0.008	-0.031	$7.3x10^{-06}$	0.014	-0.021	$1.0x10^{-06}$
2	BMI + sex + age	0.008	-0.032	$7.0x10^{-06}$	0.015	-0.021	$7.8x10^{-07}$
3	BMI + sex + age + TIM	0.011	-0.032	$1.0x10^{-05}$	0.019	-0.021	$9.2x10^{-07}$
4	BMI + sex + age + TIM + Batch	0.011	-0.032	$1.0x10^{-05}$	0.027	-0.020	$1.6x10^{-06}$
5	BMI + sex + age + TIM + Batch + ethnicity	0.38	-0.021	$3.2x10^{-03}$			

Stepwise linear model used for each covariates in the analysis of association of microbial diversity with BMI (TIM: time in mail).



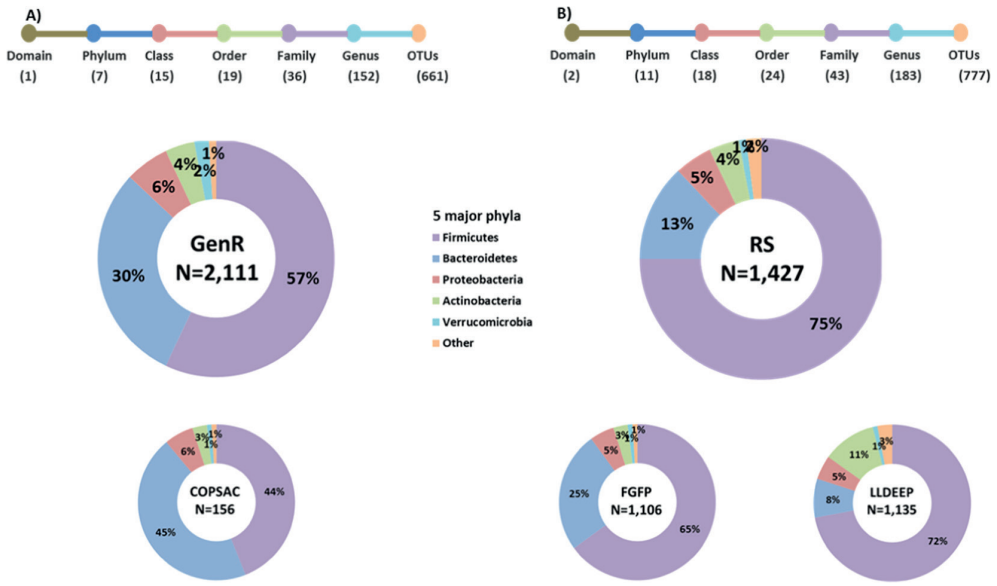
▲ **Figure 1:** Effect of ambient temperature on individual OTUs. Regression analysis of individual OTUs with time in mail (TIM) for samples in GenR (**A**) and RS (**B**). At each TIM. the initial OTU table was subsampled to contain only samples up to that TIM. Red bars indicate bacteria that decreased in abundance and green bars bacteria that increased upon increasing TIM. Q-values are indicated; only significantly abundant OTUs are presented.

5.1

Description and validation of final 16S rRNA datasets of the GenR and RS cohorts

After quality control and sample exclusion, the final 16 S datasets comprised 2,111 individuals in GenR and 1,427 subjects in RS. The average age of GenR children was 10 (range = 9 to 12) years and included participants from 13 self-reported ethnicities with Dutch being the most frequent one (62%; **Table 2 and Supplementary Figure 5**). The average age of RS participants in the final 16S dataset was 57 years (range = 46 to 88) and 82% were Dutch based on self-reported ethnicity (**Table 2 and Supplementary Figure 5**).

A total number of 661 OTUs were identified in the GenR cohort and 777 OTUs in the RS cohort (**Figure 2**). Of these, 656 OTUs overlapped between the two cohorts, leaving 5 OTUs specific for GenR and 42 OTUs specific for RS. For both datasets, variations in overall microbiome profiles were driven by the relative abundances of the four major phyla *Firmicutes, Bacteroidetes, Proteobacteria* and *Actinobacteria* (**Supplementary Figure 6**). We observed higher Shannon α-diversities in RS (mean 4.02; SD = 0.50) than



▲ **Figure 2: Characteristics of the final datasets of the two cohorts.** GenR (A) and RS (B). Number of observed taxa at each taxonomy level Top: indicates the number of unique OTUs identified in each tax-onomic clade, top A: RS cohort, top B: GenR cohort. Bottom: Donut plots indicate the average relative abundances of the top major phyla in each cohort. Donut plots of the COPSAC cohort (children aged 6 years) and doughnut plots of FGFP and LLD cohorts (adults) are plotted for comparison with the abun-dance in GenR and RS.

in the GenR cohort (mean 3.81; SD = 0.57; **Figure 2**). Furthermore, we observed the absence of kingdom Archaea and a lower abundance of phylum *Firmicutes* in the GenR cohort as compared to RS (**Figure 2**).

To validate our datasets, we compared the averaged phylum level profiles of our cohorts with those of other cohorts of similar characteristics (i.e., age range, ethnic background). In addition, we tested the well-established association between BMI and alpha diversity[11,13,40], and we determined the association between single OTUs and BMI and compared these results with other cohorts.

To compare GenR and RS 16S datasets with other populations, we used the 16S datasets of the COPSAC39, LLD[40,41] and FGFP[42] cohorts. The average phylum-level profiles of GenR were similar to those of COPSAC, and the average phylum-level profiles of RS were similar to those of LLD and FGFP (**Figure 2**). The lower abundance of phylum Firmicutes in the GenR cohort was also observed in the COPSAC cohort[39]. Further, in order to validate our microbiome datasets, we investigated the association of gut microbiome with BMI. We excluded subjects of non-Northern European origin from both GenR and RS, which resulted in a dataset of 1,712 GenR samples and a dataset of 1,371 RS samples.

α-diversity was negatively associated with BMI after adjusting for age, sex and technical covariates in both cohorts (beta = −0.019, SE = $2.3x10^{-03}$, p-value = $1.1*10^{-03}$ for GenR and beta = −0.015, SE = $2.8x10^{-03}$, p-value = $1.8x10^{-06}$ for RS). We compared this association with observations reported in similar cohorts of Northern European origin (LLD and FGFP). The negative correlation of α-diversity with BMI was in line with findings reported previously[40]. At genus-level, we identified 27 BMI-associated genera in GenR and 33 BMI-associated genera in RS (**Table 3, Supplementary Table 2**), of which 14 genera overlapped across both cohorts. In the RS cohort, we confirmed 6 genera found to be associated with BMI in the LLD and FGFP cohorts (**Table 3**). In GenR we only confirmed two genera (*Alistipes* and *Barnesiella*) to be associated with BMI in RS, LLD and FGFP cohorts. As previously reported40,48 we also observed the association between increased abundance of genus *Akkermansia* and lower BMI.

Comparison of RS and GenR stool microbiome diversities, compositions and functions

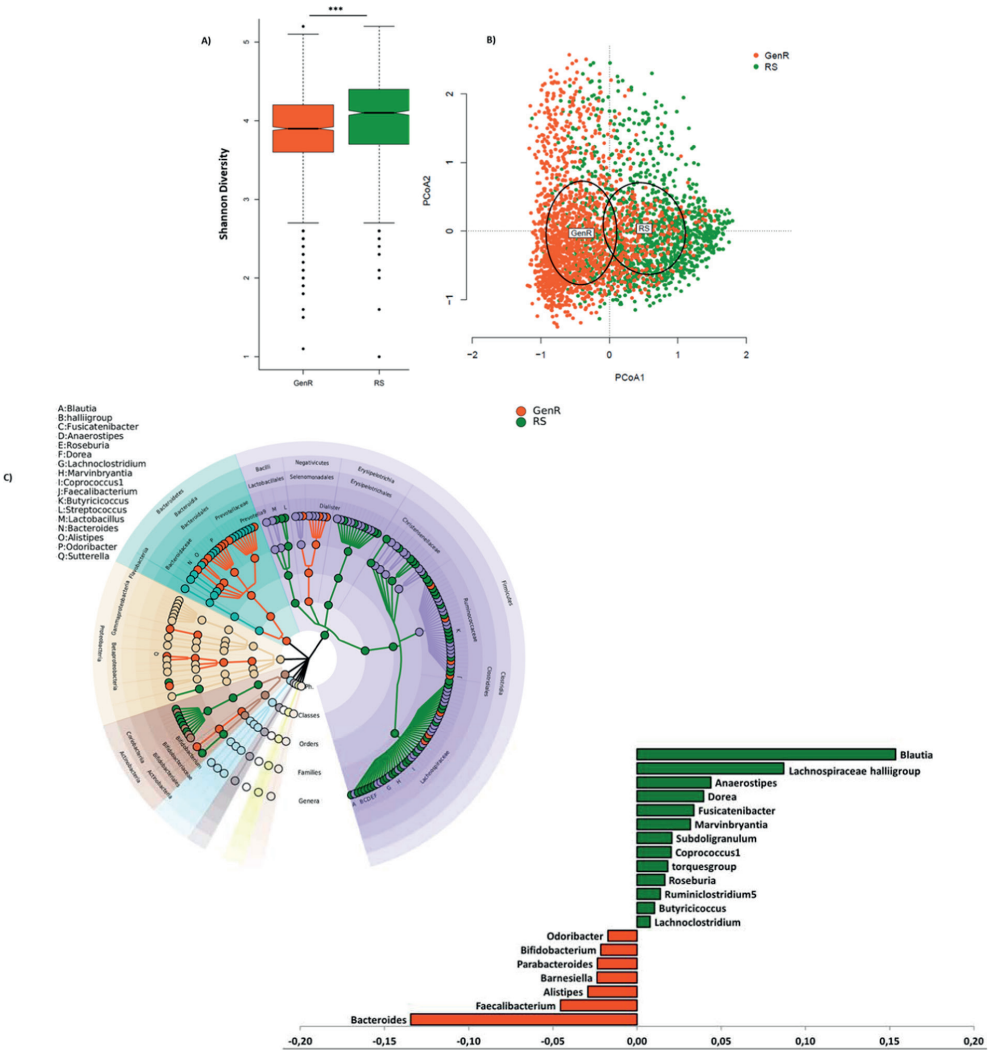
To analyze differences in gut microbiome compositions between children and adults, both datasets were combined after exclusion of the non-Northern European samples. The combined dataset included 3,082 samples: 1,371 from RS and 1,712 from GenR and the final OTU table contained 173 genera. Although Shannon α-diversity was not associated with age in each cohort separately (linear model; GenR: beta = −0.003, p-value = 0.91; RS: beta = 0.001, p-value = 0.63), Shannon α-diversity was significantly higher in the RS cohort than in the GenR cohort in the combined dataset (p-value < $2.2x10^{-16}$, mean = 5.9, sd = 0.74 for RS and mean = 5.58, sd = 0.82 for GenR). Also, the overall compositions (Bray Curtis distances) differed significantly between both groups (**Figure 3B**; PERMANOVA; R^2 = 0.06; p-value = 0.001). We observed a 2 to 3 times lower abundance of phylum Firmicutes in the GenR cohort as compared with RS (p-value < $2.2x10^{-16}$). Furthermore, we observed higher abundances of the gram-negative classes (Bacteroidia, Negativicutes and some classes from phylum Proteobacteria) in children. Regression analysis, in MaAsLin, on the relative abundances of the individual genera showed higher relative abundances of 59 genera in RS compared to GenR and higher relative abundances of 20 genera in GenR compared to the RS cohort (q-value < 0.05; **Supplementary Table 3**). The largest differences were observed for genera from family *Lachnospiraceae* including *Blautia, Lachnospiraceae, Anaerostipes, Dorea, Fusicatenibacter, Coprococcus, Roseburia, Ruminoclosteridium, Butyricicoccus* and *Lachnoclostridium* that were more abundant in RS, and genera from families *Ruminococcaceae and Bacteroidaceae* including *Bateroides, Faecalibacterium, Alistipes, Barnesiella, Parabacteroides, Bifidobacterium and Odoribacter* that were more abundant in GenR (**Figure 3C**). The most abundant genera were Bacteroides in GenR (children) cohort and Blautia in the RS (adults) cohort (**Figure 3D**).

Table 3: Single OTU associations with BMI in the GenR and RS cohorts and further replication in FGFP and LLD

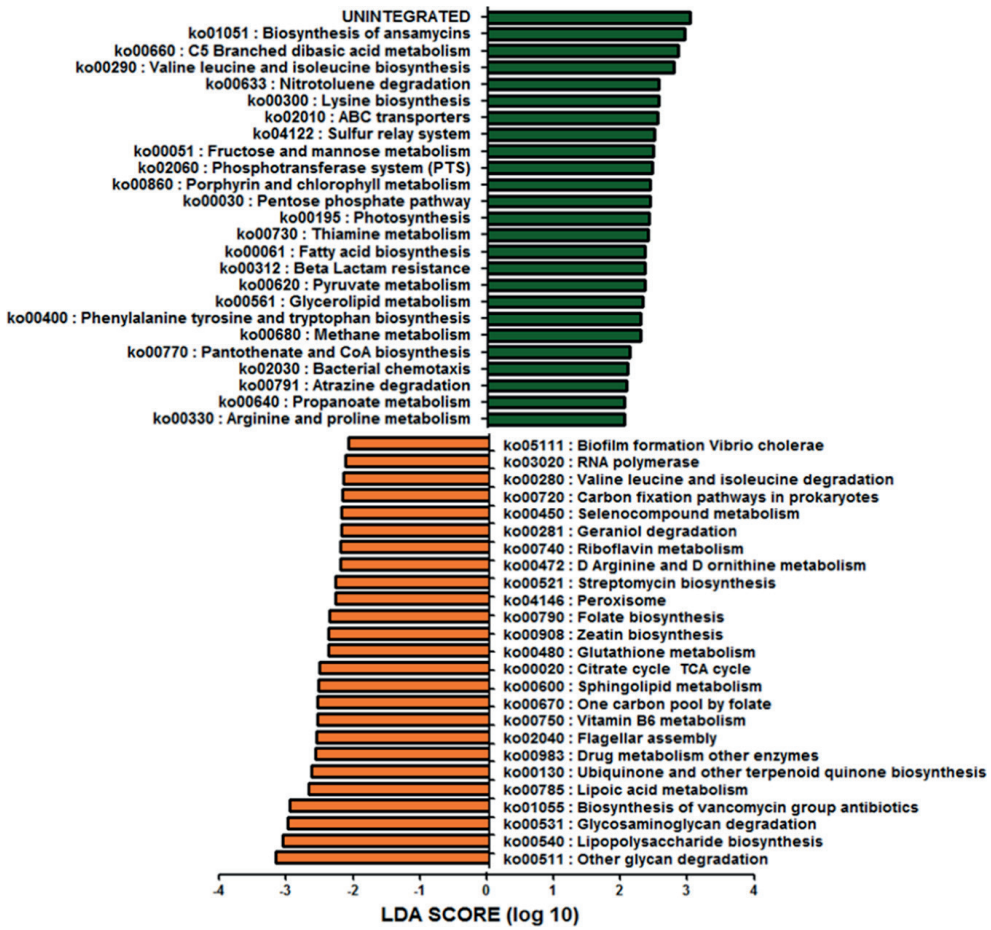
Genus	GenR (n=1,712)		RS (n=1,371)		FGFP (n=1,106)		LLD (n=1,135)	
	CoE	q-value	CoE	q-value	CoE	q-value	CoE	q-Value
Christensenellaceae R7 group	−0.0044	$8.3x10^{-04}$	−0.0034	$6.5x^{-08}$				
Alistipes	−0.0024	$8.1x10^{-03}$	−0.0015	$4.2x10^{-04}$	−0.1326	$9.7x10^{-06}$	−0.0026	$4.3x10^{-05}$
Ruminococcus 1			−0.0014	$2.5x10^{-03}$				
Coprococcus 2			−0.0014	$1.7x10^{-02}$				
Ruminiclostridium 6	−0.0019	$3.4x10^{-04}$	−0.0012	$1.1x10^{-03}$				
Anaerotruncus			−0.0008	$1.0x10^{-03}$			−0.0016	$8.5x10^{-02}$
Barnesiella	−0.0016	$9.2x10^{-03}$	−0.0007	$2.9x10^{-03}$	−0.0761	$1.1x10^{-02}$	−0.0030	$5.3x10^{-02}$
Akkermansia			−0.0006	$3.8x10^{-02}$				
Ruminiclostridium9	−0.0005	$6.4x10^{-03}$	−0.0004	$4.3x10^{-03}$				
Odoribacter	−0.0010	$1.6x10^{-03}$	−0.0004	$1.1x10^{-02}$			−0.0019	$8.1x10^{-02}$
Oscillospira			−0.0004	$2.4x10^{-03}$	−0.0933	$1.9x10^{-03}$		
Butyricicoccus			0.0006	$1.2x10^{-02}$				
Lactobacillus			0.0006	$8.5x10^{-04}$				
Lachnoclostridium			0.0006	$4.5x10^{-02}$				
Coprococcus 3			0.0008	$3.3x10^{-02}$				
Dorea			0.0014	$2.3x10^{-03}$	0.1408	$2.6x10^{-06}$		
Blautia			0.0021	$4.6x10^{-02}$				
Streptococcus			0.0026	$1.1x10^{-05}$				
Parabacteroides	−0.0015	$2.6x10^{-02}$						
Oscillibacter	−0.0005	$3.1x10^{-02}$						
Terrisporobacter	0.0005	$4.2x10^{-02}$						
Intestinibacter	0.0014	$9.7x10^{-04}$						
Romboutsia	0.0020	$2.6x10^{-03}$						
Bifidobacterium	0.0046	$3.2x10^{-04}$						

Bacterial associations with BMI in the RS, GenR and other cohort studies. The +/− sign of the coefficient values indicate the direction of the correlation of the genus with BMI. q-value = FDR corrected P-value. Only known bacteria are presented.

To assess functional differences in the gut microbiome between children and adults, we predicted the functional content based on the 16S rRNA data using PICRUSt. Analyses of the differences in the predicted functional metagenomics data between children and adults showed significant overrepresentation of 25 pathways in children and 25 pathways in adults (**Figure 4, Supplementary Table 4**). A remarkable difference between GenR and RS was the predominance of catabolic pathways in GenR (Val, Leucine, iso-Leucine degradation; ko00280) and its opposite in RS (Val, Leucine, iso-Leucine biosynthesis; ko00290). Putative colonization-related pathways like biofilm formation (ko05111), flagellar assembly (ko02040) and LPS biosynthesis (ko00540) were only enriched in GenR cohort.



▲ Figure 3: Comparison of the gut microbiome diversity and composition between adults (RS) and children (GenR). (A) boxplots of the Shannon diversity Index. (B) ordination plot of the gut microbiome composition in the two cohorts based on Bray-Curtis dissimilarities. The centroid and dispersion of each cohort is represented by the cohort name and ellipses, respectively. Clustering of RS and GenR was tested for significance using PERMANOVA. (C) Circular representation of the taxonomic tree of the microbiome compositions of the two cohorts. Each node represents one taxon at different taxonomic level. Orange nodes are the taxa that were observed with higher abundance in the GenR cohort and green nodes represent the taxa that were higher abundant in the RS cohort. (D) The genera represented the most in each cohort. On the x-axis the arcsine squared root transformed coefficients of the most significantly abundant genera in each cohort are shown. Orange bars represent GenR and green bars represent RS. Minus signs in the x-axis are used only for visualization.



▲ **Figure 4:** Predicted functional composition of metagenomes based on 16S rRNA gene sequencing data from GenR and RS cohorts. LEfSe based on the PICRUSt dataset revealed differentially enriched metabolic pathways associated with GenR (orange) or RS (green).

Discussion

In this publication we report on the 16S stool microbiota profiles of 3,538 subjects from two large, deeply phenotyped and well-characterized population-based cohorts: the Generation R (GenR) Study and the Rotterdam Study (RS). The 16S microbiome datasets originated from both children (GenR: n = 2,111) and adults (RS: n = 1,427) populations. In general, the stool microbiota within RS had similar profiles as profiles in the previously published LLD and FGFP cohorts[40-42]. Similarly, our GenR stool microbiota possessed similar profiles as those observed in the COPSAC cohort[39]. Further, both of

our cohorts replicated the previously reported negative association between α-diversity and BMI[40]. Additionally, BMI was also associated with community composition (Bray-Curtis) in both RS and GenR cohorts, as reported for the LLD and FGFP cohorts. Also, Alistipes and Barnesiella were reported as negatively associated with BMI in all four of these cohorts. Interestingly, the lower Firmicutes abundance observed in the GenR and COPSAC children cohorts compared to RS, LLD and FGFP adult cohorts suggests an age-specific phenomenon is responsible for this difference. However, understanding these differences merit further investigation. Thus, as both our RS and GenR datasets contain many characteristics that are similar to previously published large cohort studies, we conclude that our datasets allow valid investigations into the composition and variation of the gut microbiota across child and adult subjects. Comparing the gut microbiome composition between children and adults in a combined dataset of 3,083 children and adults showed significant clustering of the cohorts based on overall composition (PCoA beta diversity) and significant different α-diversities. The lower diversity in GenR suggested a lack of 'maturation' of the gut microbiome in children. However, we did not observe a significant age-related change in alpha diversity within each cohort separately. This might be due to the narrow age-range in both cohorts (9.8 ± 0.32 in GenR, 56.8 ± 5.9 in RS, **Table 2**). At genus level, the relative abundances of genus *Bacteroides* were higher in GenR than in RS. Members of the genus Bacteroides are specialized at utilizing both plant and host-derived polysaccharides[49-52].

As compared to other genera, *Bacteroides* have a large number of genes specialized to metabolize various glycans. They also have environmental sensors that control their expression upon exposure to glycans. The predicted metagenomics data showed indeed higher abundances of glycan degradation pathways in children's gut microbiome than in adults. In contrast, the relative abundances of the genus *Blautia* were higher in RS than in GenR. Blautia digest complex carbohydrates like whole-grains[53] and its abundance has been shown to be reduced in patients with colorectal cancer[54] and in children with type 1 diabetes[55]. Its abundance has also been reported to be inversely related to bone mineral density in human[56]. In fact, the relative abundance of *Blautia* and *Bacteroides,* has been implicated as a determinant of the so-called 'healthy microbiome' due to differences in the metabolic functions of these genera[54,55,57,58].

With respect to metabolic functions, we identified many predicted carbohydrate pathways that were significantly enriched in adult versus child metagenome datasets, with many of these carbohydrate pathways involving the metabolism of simple sugars such as fructose and mannose. However, whether these differences in microbiota profiles can be explained by possible differences in energy demand and metabolic programming between adults and children remains to be proven. Other interesting differ-

ences in the predicted functional pathways between children and adults, were observed in antibiotics synthesis pathways. The vancomycin biosynthesis pathways were significantly higher in the gut microbiome of children as compared to adults. Vancomycin is a glycopeptide with activity against gram-positive bacteria, while gram-negative bacteria are resistant to this drug because of their distinct cell-wall[59,60]. On the other hand, biosynthesis of ansamycin was higher in adults than in children. Ansamycin is another antibiotic with activity against gram-positive, as well as gram-negative bacteria[61]. These findings indicate potential differential anti-bacterial activities between the microbiota of children as compared to adults, which appears to be in line with our observation of higher relative abundances of gram-negative bacteria in the gut microbiota of children. However, the exact role of antibiotic production in modulating the microbiota profiles of children and adults remains to be shown.

Another difference identified by predictive functional pathway analysis was the differential biosynthesis of vitamin B classes by the gut microbiota of children and adults. Children's gut microbiomes are associated with pathways related to increased biosynthesis of vitamins B2, B6, and B9 (folate), while in adults vitamin B1 and B5 biosynthesis pathways appear to be increased. This predictive finding could be linked to age-dependent differences in, for example, the requirement for folate, which is particularly important for cell division and growth during pregnancy and childhood.

In children, the enrichment of putative colonization-related pathways involving biofilm formation, flagella assembly and LPS biosynthesis, may provide insights into the development of intra-microbiome and host-microbiome interactions, and also be linked to our observation of higher relative abundances of gram-negative bacteria in children - LPS is a major constituent of the cell wall of gram-negative bacteria. Finally, observing photosynthesis pathway (ko00195) in the functional data could be explained by the presence of Cyanobacteria in our datasets, as Cyanobacteria are photosynthetic bacteria. Likely, the photosynthesis pathways detected in the PICRUSt analysis were derived from these Cyanobacteria. The presence of Cyanobacteria in the gut may have a dietary origin[62].

As with all studies, the current study was not free of limitations. One of the main discussion points involves the fact that stool samples were not directly frozen and immediately stored at −80 °C. For logistical reasons relating to the large scale cohorts recruited, we depended on postal delivery of home-collected stool samples to the research laboratory. Several studies have addressed the effects on microbiota composition when sample collection is performed at room temperature[63-68], and observed that longer periods of storage at ambient temperature may affect microbiota profile composition

and decrease diversity in the samples. We, therefore, analyzed the effects of stool sample collection at ambient temperature in both cohorts. Within our own initial datasets, we addressed this issue by using several approaches and observed increased abundances of *Escherichia/Shigella* when stool samples were mailed to the research laboratory and received after 3 days (RS) or after 5 days (GenR) in the post. Therefore, to avoid the influence of any delayed processing time, we excluded samples that had been in the mail for longer than 3 days in RS and 5 days in GenR cohorts.

Next to the increase in abundance of *Escherichia/Shigella* upon prolonged times in the mail, we observed several taxa that decreased over time (*Coprococcus* and *Roseburia*). These decreases were, however, relatively low and likely a consequence of the compositionality of the data: if one OTU increases, other OTUs decrease. We decided to adjust for these technical artifacts in further analyses by including time in mail as a technical cofactor. The actual explanation for the difference in microbiota composition stability over time between RS and GenR cohorts is likely due to the fact that GenR participants were asked to keep their samples in their home fridge at 4 °C (for posting on Monday) if they were produced at the weekend, allowing better preservation of samples compared to RS participants who mailed their samples over the weekend. As well as the exclusion of 3 day and 5 day samples, we also included TIM as a technical covariate in all analyses.

Next to excluding samples that had been in the mail for too long, we also excluded recent antibiotic users from our datasets (196 samples from GenR and 7 samples from RS). Antibiotic use had a significant effect on alpha diversity and overall composition in GenR (**Supplementary Table 1**), whereas in RS the number of users was too small to have a detectable effect. Probiotic use and travel abroad were also recorded, but we could not detect significant effects on alpha diversity and composition in both cohorts. Given the small effect size of probiotic use, this study might be underpowered to conclude a lack of effect.

Recent studies indicate that the method of DNA isolation is the main source of technical variance in microbiome studies[69-71]. We, therefore, analyzed the effect of DNA isolation throughout the 391 runs (134 runs for RS and 257 runs for GenR) that were performed in this study and observed a batch effect causing a reduction in average DNA yield per sample in a proportion of the hundreds of runs performed. As we could not trace any clear cause for this technical artifact we introduced a “Batch” variable that allowed us to discriminate between low yield and the high yield runs. This “Batch” variable was significantly associated with overall profiles in the GenR and RS cohorts and was included as technical covariate in all analyses. Another, more general

limitation when comparing different cohort datasets, is the accuracy of replication of 16S profiles, since sample collection, sample storage, DNA isolation, PCR amplification, amplified 16S rRNA variable region, the sequencing technique used and downstream bioinformatics may differ between different cohorts. For example, out of 33 associated OTUs with BMI, we could only replicate 6 in LLD and FGFP. In addition, geographical differences and lack of repetition may limit the accuracy of replicating microbiota profiles from different countries. In addition, although (besides using harmonized sample collection and dataset generation) we adjusted for the potential confounders in the analyses comparing children and adults, we should be aware that we cannot totally discard the presence of residual stratification as a consequence of the fact that both populations were derived from different cohorts.

To conclude, in this study we performed microbiota profiling on the stools of 2,111 children in the age-range of 9 to 12 years and 1,427 adult individuals in the range of 46 to 88 years of age. We observed a clear distinction between the gut microbiomes of children as compared to adults, including differences in microbiota diversity and Firmicutes and Bacteroidetes abundances. These changes were associated with predicted shifts in functional properties, including in energy metabolism, antibiotic production and the production of essential B-vitamins. These observations are likely due to the development of human-gut microbiome interactions with age. As both GenR and the RS cohort have been deeply characterized[24,25]. We here presented two valuable datasets for studying the possible association of the human stool microbiome and, life style, environmental factors, and health and disease outcomes. In addition, as the participants of both cohorts have been genotypes, also association between host genetics and host microbiome could be investigated.

Acknowledgements

The generation and management of stool microbiome data for the Generation R Study (Focus @9) and the Rotterdam Study (RSIII-2) was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. We thank Nahid El Faquir and Jolande Verkroost-Van Heemst for their help in sample collection and registration, and Kamal Arabe, Hedayat Razawy and Karan Singh Asra for their help in DNA isolation and sequencing. Furthermore, we thank drs. Jeroen Raes and Jun Wang (KU Leuven, Belgium) for their guidance in 16S rRNA profiling and dataset generation. Djawad Radjabzadeh was funded by an Erasmus MC mRACE grant “Profiling of the human gut microbiome”. The authors are very grateful to the study participants of the Rotterdam Study, the staff from the Rotter-

dam Study (particularly L. Buist and J.H. van den Boogert) and the participating general practitioners and pharmacists. 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 GWAS datasets are supported by the Netherlands Organization of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project nr. 050-060-810. FR and CMG received funding from ZonMW-VIDI, project nr. 016.136.367. The Generation R Study is conducted by the Erasmus Medical Center in close collaboration with the School of Law and Faculty of Social Sciences of the Erasmus University Rotterdam, the Municipal Health Service Rotterdam area, Rotterdam, the Rotterdam Homecare Foundation, Rotterdam and the Stichting Trombosedienst & Artsenlaboratorium Rijnmond (STAR-MDC), Rotterdam. We acknowledge the contribution of children and parents, general practitioners, hospitals, midwives and pharmacies in Rotterdam. The general design of Generation R Study was made possible by financial support from the Erasmus Medical Center, Rotterdam, the Erasmus University Rotterdam, the Netherlands Organization for Health Research and Development (ZonMW), the Netherlands Organization for Scientific Research (NWO), the Ministry of Health, Welfare and Sport and the Ministry of Youth and Families. This project also received funding from the European Union's Horizon 2020 research and innovation programme under the following grant agreements: [No. 633595 (DynaHEALTH) and No. 733206 (LIFECYCLE)]. Furthermore, Generation R received additional funding from the European Research Council (ERC Consolidator Grant, ERC-2014-CoG-648916).

Supplementary Material

Additional material, as well as a funding statement, is published online only, to view please visit the paper online at the journal website:

https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1006260

References

1. Carroll IM, Chang Y-H, Park J, Sartor RB, Ringel Y. Luminal and mucosal-associated intestinal microbiota in patients with diarrhea-predominant irritable bowel syndrome. Gut Pathog. 2010;2:19. doi: 10.1186/1757-4749-2-19.- DOI- PMC- PubMed
2. Carroll IM, et al. Molecular analysis of the luminal- and mucosal-associated intestinal microbiota in diarrhea-predominant irritable bowel syndrome. Am. J. Physiol. Gastrointest. Liver. Physiol. 2011;301:G799–G807. doi: 10.1152/ajpgi.00154.2011.- DOI- PMC- PubMed
3. Frank DN, et al. Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. Inflamm. Bowel Dis. 2011;17:179–184. doi: 10.1002/ibd.21339.- DOI- PMC- PubMed
4. Frank DN, et al. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. Proc. Natl. Acad. Sci. USA. 2007;104:13780–13785. doi: 10.1073/pnas.0706625104.- DOI- PMC- PubMed
5. Kassinen A, et al. The fecal microbiota of irritable bowel syndrome patients differs significantly from that of healthy subjects. Gastroenterology. 2007;133:24–33. doi: 10.1053/j.gastro.2007.04.005.- DOI- PubMed
6. Mai V, et al. Fecal microbiota in premature infants prior to necrotizing enterocolitis. PloS One. 2011;6:e20647–e20647. doi: 10.1371/journal.pone.0020647.- DOI- PMC- PubMed
7. Malinen E, et al. Analysis of the fecal microbiota of irritable bowel syndrome patients and healthy controls with real-time PCR. Am. J. Gastroenterol. 2005;100:373. doi: 10.1111/j.1572-0241.2005.40312.x.- DOI- PubMed
8. Ringel Y, Carroll IM. Alterations in the intestinal microbiota and functional bowel symptoms. Gastrointest. Endosc. Clin. N. Am. 2009;19:141–150. doi: 10.1016/j.giec.2008.12.004.- DOI- PubMed
9. Swidsinski A, Loening-Baucke V, Verstraelen H, Osowska S, Doerffel Y. Biostructure of fecal microbiota in healthy subjects and patients with chronic idiopathic diarrhea. Gastroenterology. 2008;135:568–579. doi: 10.1053/j.gastro.2008.04.017.- DOI- PubMed
10. Kalliomäki M, Carmen Collado M, Salminen S, Isolauri E. Early differences in fecal microbiota composition in children may predict overweight. Am. J. Clin. Nutr. 2008;87:534–538. doi: 10.1093/ajcn/87.3.534.- DOI- PubMed
11. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Human gut microbes associated with obesity. Nature. 2006;444:1022–1023. doi: 10.1038/4441022a.- DOI- PubMed
12. Penders J, et al. Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study. Gut. 2007;56:661–667. doi: 10.1136/gut.2006.100164.- DOI- PMC - PubMed
13. Turnbaugh PJ, et al. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006;444:1027–1031. doi: 10.1038/nature05414.- DOI- PubMed
14. Vael C, Desager K. The importance of the development of the intestinal microbiota in infancy. Curr. Opin. Pediatr. 2009;21:794–800. doi: 10.1097/MOP.0b013e328332351b.- DOI- PubMed
15. Wang M, et al. Reduced diversity in the early fecal microbiota of infants with atopic eczema. J. Allergy Clin. Immunol. 2008;121:129–134. doi: 10.1016/j.jaci.2007.09.011.- DOI- PubMed
16. Floch MH. Advances in Intestinal Microecology. Nutr. Clin. Pract. 2012;27:193–194. doi: 10.1177/0884533612439708.- DOI- PubMed
17. Ringel-Kulka T. Targeting the intestinal microbiota in the pediatric population. Nutr. Clin. Pract. 2012;27:226–234. doi: 10.1177/0884533612439895.- DOI- PubMed
18. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. PLoS Biol. 2007;5:e177–e177. doi: 10.1371/journal.pbio.0050177.- DOI- PMC- PubMed
19. Tiihonen K, Ouwehand AC, Rautonen N. Human intestinal microbiota and healthy ageing. Ageing Res. Rev. 2010;9:107–116. doi: 10.1016/j.arr.2009.10.004.- DOI- PubMed

20. Yatsunenko T, et al. Human gut microbiome viewed across age and geography. Nature. 2012;486:222–227. doi: 10.1038/nature11053.- DOI- PMC- PubMed
21. Kozyrskyj AL, Ernst P, Becker AB. Increased risk of childhood asthma from antibiotic use in early life. Chest. 2007;131:1753–1759. doi: 10.1378/chest.06-3008.- DOI- PubMed
22. Risnes KR, Belanger K, Murk W, Bracken MB. Antibiotic exposure by 6 months and asthma and allergy at 6 years: findings in a cohort of 1,401 US children. Am. J. Epidemiol. 2011;173:310–318. doi: 10.1093/aje/kwq400.- DOI- PMC- PubMed
23. Hviid A, Svanström H, Frisch M. Antibiotic use and inflammatory bowel diseases in childhood. Gut. 2011;60:49–54. doi: 10.1136/gut.2010.219683.- DOI- PubMed
24. Kooijman MN, et al. The Generation R Study: design and cohort update 2017. Eur. J. Epidemiol. 2016;31:1243–1264. doi: 10.1007/s10654-016-0224-9.- DOI- PMC- PubMed
25. Ikram MA, et al. The Rotterdam Study: 2018 update on objectives, design and main results. Eur. J. Epidemiol. 2017;32:807–850. doi: 10.1007/s10654-017-0321-4.- DOI- PMC- PubMed
26. Fadrosh DW, et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome. 2014;2:6. doi: 10.1186/2049-2618-2-6.- DOI- PMC - PubMed
27. Caporaso JG, et al. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods. 2010;7:335–336. doi: 10.1038/nmeth.f.303.- DOI- PMC- PubMed
28. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods. 2013;10:996–998. doi: 10.1038/nmeth.2604.- DOI- PubMed
29. Schmieder R, Lim YW, Rohwer F, Edwards R. TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. BMC bioinformatics. 2010;11:341–341. doi: 10.1186/1471-2105-11-341.- DOI- PMC- PubMed
30. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 2014;30:614–620. doi: 10.1093/bioinformatics/btt593.- DOI- PMC- PubMed
31. Quast C, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41:D590–D596. doi: 10.1093/nar/gks1219.- DOI- PMC - PubMed
32. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 2007;73:5261–5267. doi: 10.1128/AEM.00062-07.- DOI- PMC- PubMed
33. Benson AK, et al. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. Proc. Natl. Acad. Sci. USA. 2010;107:18933–18938. doi: 10.1073/pnas.1007028107.- DOI- PMC- PubMed
34. Bokulich NA, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat. Methods. 2013;10:57–59. doi: 10.1038/nmeth.2276.- DOI- PMC- PubMed
35. R Foundation for Statistical Computing, Vienna, Austria. A language and environment for statistical computing, https://www.R-project.org (2010).
36. Oksanen, J. et al. Vegan: community ecology package, http://CRAN.R-project.org/package=vegan (2013).
37. McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS One. 2013;8:e61217–e61217. doi: 10.1371/journal.pone.0061217. - DOI- PMC- PubMed
38. Morgan XC, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol. 2012;13:R79–R79. doi: 10.1186/gb-2012-13-9-r79.- DOI- PMC- PubMed
39. Stokholm J, et al. Maturation of the gut microbiome and risk of asthma in childhood. Nat. Commun. 2018;9:141. doi: 10.1038/s41467-017-02573-2.- DOI- PMC- PubMed
40. Fu J, et al. The gut microbiome contributes to a substantial proportion of the variation in blood lipids. Circ. Res. 2015;117:817–824. doi: 10.1161/CIRCRESAHA.115.306807.- DOI- PMC- PubMed

41. Zhernakova A, et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science. 2016;352:565–569. doi: 10.1126/science.aad3369.- DOI- PMC - PubMed

42. Falony G, et al. Population-level analysis of gut microbiome variation. Science. 2016;352:560–564. doi: 10.1126/science.aad3503.- DOI- PubMed

43. Medina-Gomez C, et al. Bone mass and strength in school-age children exhibit sexual dimorphism related to differences in lean mass: The Generation R Study. J. Bone Miner. Res. 2016;31:1099–1106. doi: 10.1002/jbmr.2755.- DOI- PubMed

44. Medina-Gómez C, et al. BMD loci contribute to ethnic and developmental differences in skeletal fragility across populations: assessment of evolutionary selection pressures. Mol. Biol. Evol. 2015;32:2961–2972. doi: 10.1093/molbev/msv170.- DOI- PMC- PubMed

45. Langille MGI, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat. Biotechnol. 2013;31:814–821. doi: 10.1038/nbt.2676.- DOI- PMC- PubMed

46. Abubucker S, et al. Metabolic reconstruction for metagenomic data and its application to the human microbiome. PLoS Comput. Biol. 2012;8:e1002358–e1002358. doi: 10.1371/journal.pcbi.1002358.- DOI- PMC- PubMed

47. Segata N, et al. Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12:R60. doi: 10.1186/gb-2011-12-6-r60.- DOI- PMC- PubMed

48. Everard A, et al. Cross-talk between Akkermansia muciniphila and intestinal epithelium controls diet-induced obesity. Proc. Natl. Acad. Sci. USA. 2013;110:9066–9071. doi: 10.1073/pnas.1219451110.- DOI- PMC- PubMed

49. Anderson KL, Salyers AA. Biochemical evidence that starch breakdown by Bacteroides thetaiotaomicron involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes. J. Bacteriol. 1989;171:3192–3198. doi: 10.1128/JB.171.6.3192-3198.1989.- DOI- PMC- PubMed

50. Chung WSF, et al. Modulation of the human gut microbiota by dietary fibres occurs at the species level. BMC Biol. 2016;14:3. doi: 10.1186/s12915-015-0224-3.- DOI- PMC- PubMed

51. Koropatkin NM, Cameron EA, Martens EC. How glycan metabolism shapes the human gut microbiota. Nat. Rev. Microbiol. 2012;10:323–335. doi: 10.1038/nrmicro2746.- DOI- PMC- PubMed

52. Xu J, et al. Evolution of symbiotic bacteria in the distal human intestine. PLoS Biol. 2007;5:e156–e156. doi: 10.1371/journal.pbio.0050156.- DOI- PMC- PubMed

53. Martínez I, et al. Gut microbiome composition is linked to whole grain-induced immunological improvements. ISME J. 2013;7:269–280. doi: 10.1038/ismej.2012.104.- DOI- PMC- PubMed

54. Chen W, Liu F, Ling Z, Tong X, Xiang C. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. PloS One. 2012;7:e39743–e39743. doi: 10.1371/journal.pone.0039743.- DOI- PMC- PubMed

55. Murri M, et al. Gut microbiota in children with type 1 diabetes differs from that in healthy children: a case-control study. BMC Med. 2013;11:46. doi: 10.1186/1741-7015-11-46.- DOI- PMC- PubMed

56. Hou AY, Kaczmarek JL, Khan NA, Holscher HD. Dietary fiber and the human gastrointestinal microbiota as predictors of bone health. FASEB J. 2017;31:lb322. doi: 10.1096/fj.201601082R.- DOI

57. Bajaj JS, et al. Colonic mucosal microbiome differs from stool microbiome in cirrhosis and hepatic encephalopathy and is linked to cognition and inflammation. Am. J. Physiol. Gastrointest. Liver Physiol. 2012;303:G675–G685. doi: 10.1152/ajpgi.00152.2012.- DOI- PMC- PubMed

58. Hong P-Y, Croix JA, Greenberg E, Gaskins HR, Mackie RI. Pyrosequencing-based analysis of the mucosal microbiota in healthy individuals reveals ubiquitous bacterial groups and micro-heterogeneity. PloS One. 2011;6:e25042–e25042. doi: 10.1371/journal.pone.0025042.- DOI- PMC- PubMed

59. Nikaido H. Outer membrane barrier as a mechanism of antimicrobial resistance. Antimicrob. Agents Chemother. 1989;33:1831–1836. doi: 10.1128/AAC.33.11.1831.- DOI- PMC- PubMed

60. Shivaramaiah HS, Relhan N, Pathengay A, Mohan N, Flynn HW. Endophthalmitis caused by gram-positive bacteria resistant to vancomycin: clinical settings, causative organisms, antimicrobial susceptibilities, and treatment outcomes. Am. J. Ophthalmol. Case Rep. 2018;10:211–214. doi: 10.1016/j.ajoc.2018.02.030.- DOI- PMC- PubMed

61. Wrona IE, Agouridas V, Panek JS. Design and synthesis of ansamycin antibiotics. C. R. Chim. 2008;11:1483–1522. doi: 10.1016/j.crci.2008.07.003.- DOI

62. Huttenhower C, et al. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486:207–214. doi: 10.1038/nature11234.- DOI- PMC- PubMed

63. Carroll IM, Ringel-Kulka T, Siddle JP, Klaenhammer TR, Ringel Y. Characterization of the fecal microbiota using high-throughput sequencing reveals a stable microbial community during storage. PloS One. 2012;7:e46953–e46953. doi: 10.1371/journal.pone.0046953.- DOI- PMC- PubMed

64. Dominianni C, Wu J, Hayes RB, Ahn J. Comparison of methods for fecal microbiome biospecimen collection. BMC Microbiol. 2014;14:103. doi: 10.1186/1471-2180-14-103.- DOI- PMC- PubMed

65. Gorzelak MA, et al. Methods for improving human gut microbiome data by reducing variability through sample processing and storage of stool. PloS One. 2015;10:e0134802–e0134802. doi: 10.1371/journal.pone.0134802.- DOI- PMC- PubMed

66. Lauber CL, Zhou N, Gordon JI, Knight R, Fierer N. Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. FEMS Microbiol. Lett. 2010;307:80–86. doi: 10.1111/j.1574-6968.2010.01965.x.- DOI- PMC- PubMed

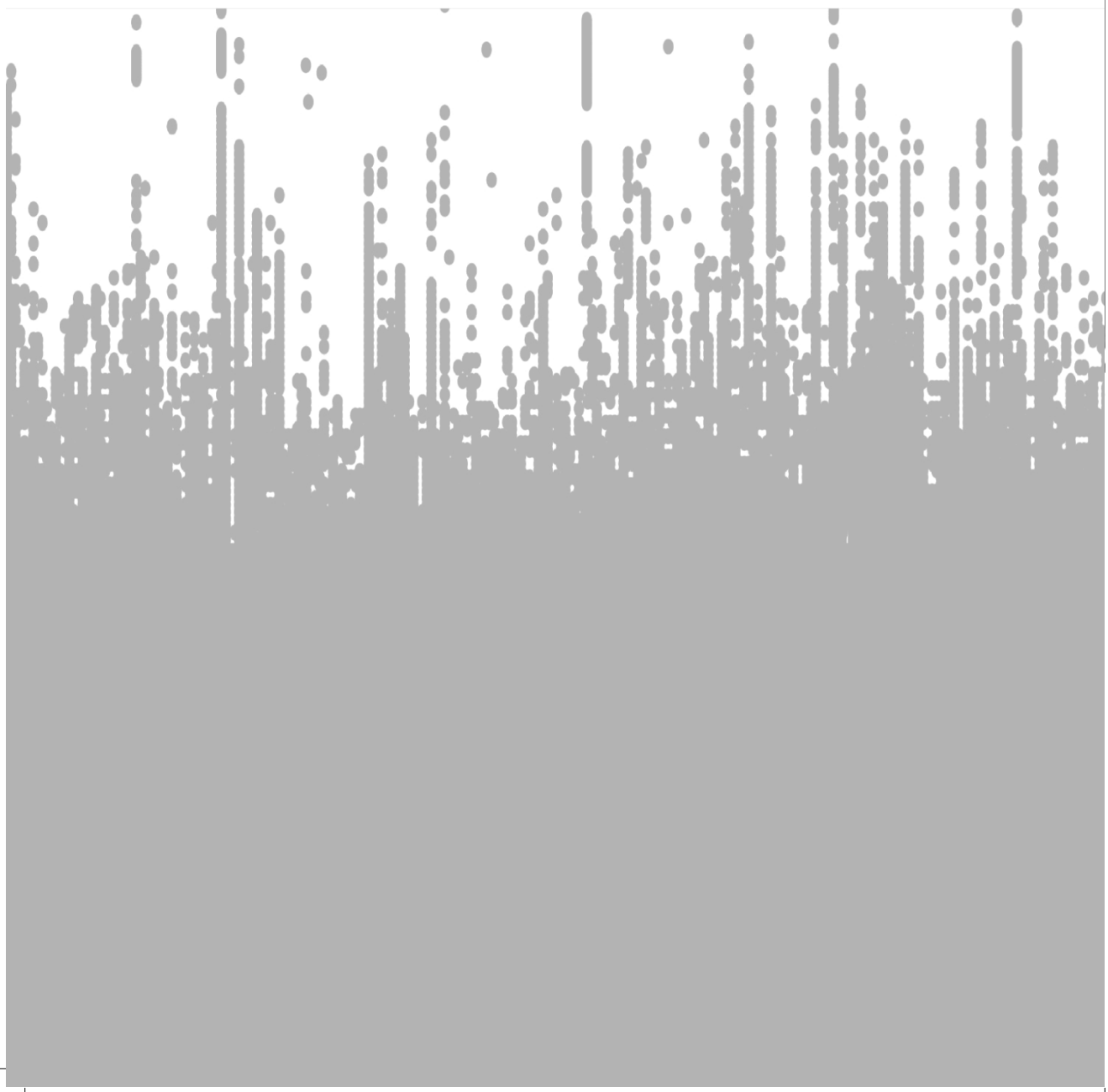
67. Sinha R, et al. Collecting fecal samples for microbiome analyses in epidemiology studies. Cancer Epidemiol. Biomarkers Prev. 2016;25:407–416. doi: 10.1158/1055-9965.EPI-15-0951.- DOI- PMC - PubMed

68. Tedjo DI, et al. The effect of sampling and storage on the fecal microbiota composition in healthy and diseased subjects. PloS One. 2015;10:e0126685–e0126685. doi: 10.1371/journal.pone.0126685. - DOI- PMC- PubMed

69. Costea PI, et al. Towards standards for human fecal sample processing in metagenomic studies. Nat. Biotechnol. 2017;35:1069–1076. doi: 10.1038/nbt.3960.- DOI- PubMed

70. Santiago A, et al. Processing faecal samples: a step forward for standards in microbial community analysis. BMC Microbiol. 2014;14:112. doi: 10.1186/1471-2180-14-112.- DOI- PMC- PubMed

71. Sinha R, et al. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat. Biotechnol. 2017;35:1077–1086. doi: 10.1038/nbt.3981.- DOI- PMC – PubMed



Chapter 5.2

Intestinal microbiome composition and its relation to joint pain and inflammation

Boer, C.G, Radjabzadeh, D., Medina-Gomez, C., Garmaeva, S., Schiphof, D., Arp, P., Koet, T., Kurilshikov, A., Fu, J., M. Ikram, M.A., Bierma-Zeinstra, S., Uitterlinden, A.G., Kraaij, R., Zhernakova, A. & van Meurs, J.B.J

Published in: Nature Communications volume 10, Article number: 4881 (2019)

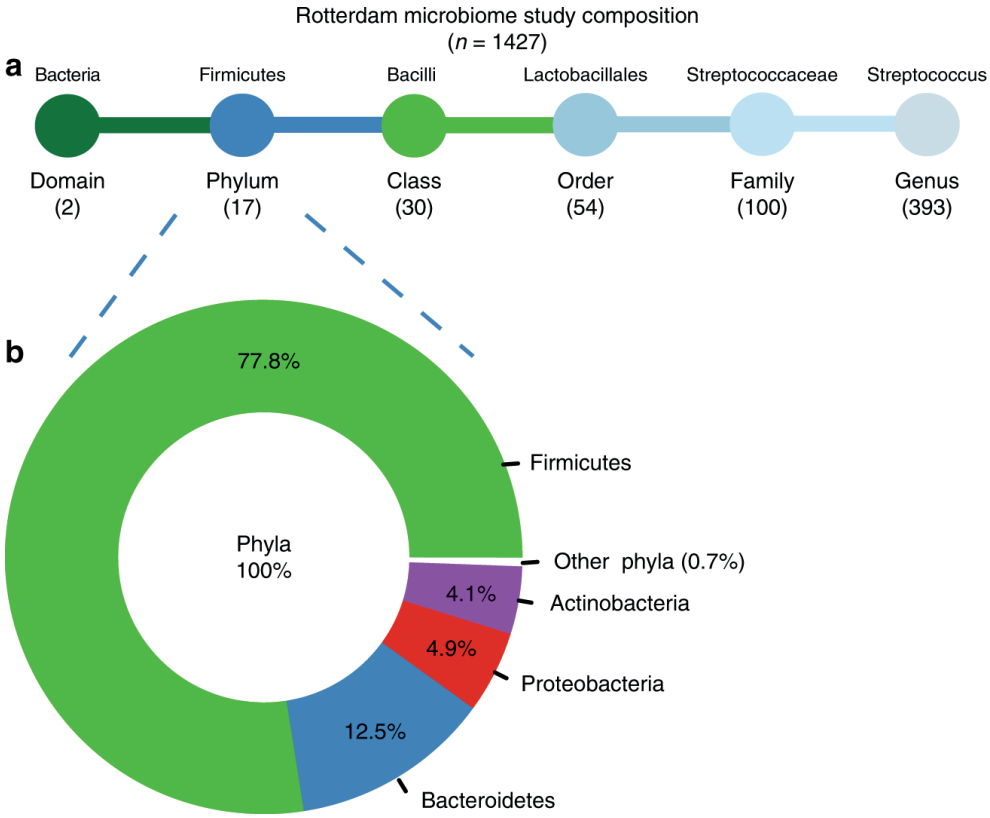
Abstract

Macrophage-mediated inflammation is thought to have a causal role in osteoarthritis-related pain and severity, and has been suggested to be triggered by endotoxins produced by the gastrointestinal microbiome. Here we investigate the relationship between joint pain and the gastrointestinal microbiome composition, and osteoarthritis-related knee pain in the Rotterdam Study; a large population based cohort study. We show that abundance of Streptococcus species is associated with increased knee pain, which we validate by absolute quantification of Streptococcus species. In addition, we replicate these results in 867 Caucasian adults of the Lifelines-DEEP study. Finally we show evidence that this association is driven by local inflammation in the knee joint. Our results indicate the microbiome is a possible therapeutic target for osteoarthritis-related knee pain.

Introduction

Osteoarthritis (OA) is a degenerative joint disease and the most common form of arthritis: an estimated 22% of the adult population has at least one joint affected by osteoarthritis and this prevalence increases to 49% in individuals over 65 years of age[1]. The hallmark clinical symptom of OA is pain, which is one of the leading causes of disability in OA[2]. Pathological changes in OA affect all joint tissues: degradation of cartilage and bone, abnormal bone formation (osteophytes), and inflammation of the synovial membrane, (synovitis). Although OA is often described as predominantly caused by mechanical factors and genetic predisposition, the existence of inflammation in OA, locally or systemic, is widespread[3-5]. In addition, it has become apparent that, by promoting or exacerbating OA symptoms, predominantly OA joint pain[4-7] this local or systemic inflammation has a causal role in OA pathology[3-5].

Obesity, a well-known risk factor for OA, is thought to increase OA risk through increased mechanical loading on weight-bearing joints. However, obesity also increases the risk for OA in non-weight-bearing joints[5,8]. The increased risk of OA in non-weight-bearing joints seen in obese individuals might be directed through low-grade systemic inflammation[9,10]. The gastrointestinal microbiome has emerged as one of the factors triggering obesity associated low-grade systemic inflammation[10-13]. Obesity is associated with changes in gastrointestinal-microbiome composition, which can lead to an increased intestinal absorption of immunogenic bacterial products[12,14,15]. Gastrointestinal bacteria produce a wide range of biologically active molecules, such as metabolites, short-chain fatty acids, proteins and enzymes, of which some are secreted in, outer membrane or membrane vesicles (OMVs/MVs). All gram-negative bacteria, archaea, fungi and several gram-positive bacteria can constitutively produce these vesicles[16-18]. These vesicles are insensitive to proteases, suggesting they can transport their content over long distances from their sites of origin. The content of these vesicles can be delivered to different organs in a concentrated manner[19]. Also, some of these biologically active molecules can affect intestinal mucosal permeability (short-chain fatty acids) or activate the immune system (lipopolysaccharide)[20]. Specifically, these molecules can affect macrophage activation and Toll-like-receptor (TLR) pathways[16,21], which have recently been shown to be the predominant inflammatory responses seen in OA[4]. Indeed, elevated levels of the bacterial endotoxin LPS (lipopolysaccharide), in the blood or in the synovium of OA patients, is associated with more severe knee OA, knee pain and inflammation[22]. This and other studies link gut-microbiome composition to low-grade systemic and local inflammation seen in OA[23-25].



▲**Figure 1: Schematic representation of the gut-microbiome taxonomic abundance in the Rotterdam Study cohort**. A) overview of the number unique taxonomies detected at each level, unknown and unclassified bacteria were excluded. A) Above, as an example, the taxonomic classification for Streptococcus is shown. B) Donut plot of the relative abundancy in percentage (%) of the different unique phyla present in the entire dataset (n=1427) unknown and unclassified bacteria were excluded

In the present study we examine stool microbiome as a proxy for the gastrointestinal-microbiome composition in relation to knee OA severity, OA-related knee pain, measured by the WOMAC-pain score, and obesity, in a large population-based cohort. We report a significant association between Streptococcus species (spp.) abundance in stool microbiome samples, knee WOMAC pain, knee inflammation, and replicate this finding in an independent cohort. Our results suggest a true relationship between the gastrointestinal microbiome, low-grade inflammation in the knee, and knee OA pain, independent of obesity.

Results

Rotterdam Study Microbiome cohort profile

For 1,427 participants from the Rotterdam Study (RSIII) we determined the gastrointestinal-microbiome composition by taking the stool microbiome as a proxy for the intestinal microbiome. In the Rotterdam Study Microbiome cohort, we sequenced two hypervariable regions of the bacterial 16S rRNA gene, hypervariable regions V3 and V4. After quality control, the 16S reads were directly mapped against the Silva 16S sequence database (v128) using RDP classifier for taxonomic classification. Classification prediction was done on multiple taxonomic levels: domain, phylum, class, order, family and genus. Gastrointestinal-microbiome composition of the 1,427 participants in the Rotterdam Study Microbiome cohort is schematically presented **Figure 1**. In total, there are 596 single taxonomies in our cohort, with unknown and unclassified bacteria excluded, as these could not be identified for clinical therapeutic relevance. At phylum level, the dominant phyla are Firmicutes (77.8%) and Bacteriodites (12.5%), followed by Proteobacteria (4.9%) and Actinobacteria (4.1%, Figure. 1). This is in concordance with other large-scale population-cohorts of Caucasian adults[26,27]. General characteristics of the Rotterdam Study Microbiome cohort are presented in **Table 1**. The study population (n = 1,427) consisted of 57.5% women (n = 821) and was slightly obese, with an average body mass index (BMI) of 27.5. A total of 124 individuals had radiographic knee OA, while 285 participants reported knee OA pain (WOMAC pain score > 0). The majority of participants reporting knee pain was female (n = 206). The average WOMAC pain score was also significantly higher in females compared to males (p-value = 1.1×10^{-07}, Student's T-test, **Table 1**).

Streptococcus abundance is associated with OA knee pain

First, we examined whether the overall microbiome composition was different across knee WOMAC pain scores and OA severity (Kellgren-Lawrence radiographic OA severity scores). We found that knee WOMAC pain significantly contributes to the intestinal microbiome β-diversity as evaluated at genus level (Aitchison distance, r2 = 0.0014, p-value = 0.005, PERMANOVA, **Supplementary Figure 1**). This association was attenuated when BMI was added to the model and significance was lost (r2 = 0.00088, p-value = 0.143, PERMANOVA). The intra-individual gastrointestinal-microbiome diversity (α-diversity, Shannon index and inverse Simpson) was not associated with knee WOMAC pain. For knee OA severity (KLsum, n = 941), we did not identify an association with gastrointestinal microbiome α- and β-diversity (**Supplementary Table 1**).

Table 1: General characteristics of the Rotterdam Study Microbiome cohort

Rotterdam Study Microbiome	Females	Males	Total
Cohort participants	821	606	1.427
Age (years)	56.8 (5.9)	56.9 (5.9)	56.9 (5.9)
BMI (kg/m2)	27.4 (4.9)	27.6 (4.0)	27.5 (4.5)
Alcohol (g/day)	1.3 (2.7)	1.3 (2.3)	1.3 (2.6)
Smoking (y/n)	98 smokers/721 non smokers	97 smokers/507 non smokers	195 current smokers
PPI (y/n)	182 users/638 non-users	114 users/492 non-users	296 current PPI users
NSAIDs (y/n)	127 users/693 non-users	51 users/555 non-users	178 current NSAID users
Knee phenotypes			
Knee OA (y/n)	84 cases/456 controls	40 cases/361 controls	124 cases/817 controls
KLSum score	1.0 (1.4)	0.7 (1.2)	0.8 (1.3)
WOMAC-Pain score	1.2 (2.6)	0.6 (1.9)	0.9 (2.3)
WOMAC-Pain score > 0	206	79	285
α-diversity metrics			
Shannon Index	4.0 (4.1)	4.0 (4.0)	4.0 (0.5)
Inverse Simpson Index	26.0 (12.1)	25.5 (12.2)	25.8 (12.2)

Depicted are the mean and the SD (standard deviation) in parenthesis
PPI oral use of proton pump inhibitors, NSAIDs oral use of non-steroidal anti-inflammatory drugs, OA osteoarthritis, WOMAC Western Ontario and McMaster Osteoarthritis Index

To gain more insight into which gut-microbiome taxonomies drive the association with knee WOMAC pain, we performed multivariate association analysis on the 256 taxonomies remaining after additional QC. After adjusting for age, sex and technical covariates, we found four microbiome abundances significantly (FDR < 0.05) associated with knee WOMAC-pain severity (**Table 2**). These were all in the same clade; from class, order, family leading to the bacterial genus of *Streptococcus* (genus: coefficient = 5.0×10^{-03}, FDR p-value = 1.2×10^{-05}, MaAsLin, **Table 2**; See **Supplementary Data 1** for the full summary statistics). After additional correction for possible confounders (smoking and alcohol consumption)[28], *Streptococcus spp.* abundance remained significantly associated with knee WOMAC pain (coefficient = 4.8×10^{-03}, FDR P-value = 3.8×10^{-05}, MaAsLin, **Supplementary Table 2**). This association is largely independent of BMI (coefficient = 4.1×10^{-03}, FDR p-value = 2.1×10^{-03}, MaAsLin, **Supplementary Table 2**). Ethnicity has recently been put forward as a possible confounder for gastrointestinal-microbiome composition26. However, after exclusion of individuals with non-European ancestry (n = 163) from our analysis, *Streptococcus spp.* remained associated with knee WOMAC pain (coefficient = 5.0×10^{-03}, FDR p-value = 5.8×10^{-05}, MaAsLin, **Supplementary Table 3**), also, after adjustment for confounders (coefficient = 4.3×10^{-03}, FDR p-value = 1.9×10^{-03}, MaAsLin). Altogether, we found that the gut microbiome, and in particular a greater relative abundance of *Streptococcus spp.*, is significantly associated with higher knee WOMAC pain independent of smoking, alcohol consumption and BMI.

Streptococcus association not driven by oral medication use

Proton Pump Inhibitors (PPI) are among the most widely used over-the-counter drugs in the world. They are used to treat gastro-esophageal reflux and prevent gastric-ulcers. Recent research has shown that the gastrointestinal-microbiome composition of PPI users is profoundly different from non-PPI users, mainly due to a strong increase in *Lacto*-bacilli abundance, driven by Streptococcus spp.[29]. Another potential over-the-counter medication frequently taken by patients with joint complaints are non-steroidal anti-inflammatory drugs (NSAIDs), which also affect the gastrointestinal microbiome[29]. To investigate, whether the association between Streptococcus spp. abundance and knee WOMAC pain is mediated by use of these drugs, we included PPI and NSAID use as covariates in our model (in addition to age, sex, technical covariates, smoking, alcohol consumption, and BMI). After adjustment for these covariates, the coefficient was attenuated, but the association remains significant (coefficient = 3.4×10^{-03}, p-value = 2.4×10^{-04}, MaAsLin, Supplementary Table 4). In line with these results, excluding all current PPI users (n = 265) from the analysis, greater Streptococcus spp. abundance was significantly associated with higher knee WOMAC pain (n = 1104, coefficient = 2.4×10^{-03}, p-value = 4.0×10^{-03}, MaAsLin). Thus, the association between, Streptococcus spp. and knee WOMAC pain was not due to the confounding effect of PPI use. Since knee WOMAC-pain scores were not normally distributed (left-skewed with an overabundance of zeros), we also assessed the association using a Poisson-regression model, which can account for the non-normal distribution. Here as well, a significant association with Streptococcus spp. abundance was observed (beta = 1.85, p-value = 1.4×10^{-04}, Poisson-regression). Further sensitivity analysis excluding individuals who did not report pain (n = 1,142) or whose knee WOMAC-pain scores were deemed as outliers (WOMAC-pain score >10, > 4 SD, n = 56) resulted in slight changes in the association coefficients. The association, however, remained significant (Supplementary Table 5). Altogether, these analyses show that the association between relative abundance of Streptococcus spp. and knee WOMAC pain is robust.

To adjust for possible spurious collinearity between microbe abundancies[31], we have used the isometric log-ratio transformation (ILR). Using ILR, we have compared the relative *Streptococcus spp.* abundancy against the geometric mean of the abundancy of all other genera. Results show that the ILR transformed *Streptococcus spp.* abundancy is associated with knee WOMAC pain (p-value = 9.9×10^{-06}, MaAsLin, **Supplementary Table 6**). The association remained significant after adjusting for smoking, alcohol consumption and BMI (p-value = 8.4×10^{-04}, MaAsLin, **Supplementary Table 6**), and NSAID and PPI use (p-value = 8.5×10^{-03}, MaAslin, **Supplementary Table 6**). The qPCR results and ILR transformation demonstrate that the association between Streptococcus spp. abundance and knee WOMAC pain is not an artefact of the 16S rRNA sequencing.

Table 2: Results of multivariate linear regression analysis of gut microbiome relative abundancies and knee WOMAC-pain scores, in the Rotterdam Study, LifeLines-DEEP

Rotterdam Study (RS)							
Taxonomy		**%**	**N**	**CoE**	**SE**	**p-value**	**FDR**
Class	***Bacilli***	27.3%	1,419	6.1×10^{-03}	1.0×10^{-03}	9.1×10^{-09}	1.0×10^{-06}
Order	***Lactobacillales***	100%	1,417	6.1×10^{-03}	1.1×10^{-03}	7.6×10^{-09}	1.0×10^{-06}
Family	***Streptococcaceae***	79.6%	1,402	4.9×10^{-03}	9.3×10^{-04}	1.5×10^{-07}	8.7×10^{-06}
Genus	***Streptococcus***	98.7%	1,396	5.0×10^{-03}	9.3×10^{-04}	7.3×10^{-08}	5.6×10^{-06}
		Life Lines Deep (LLD)					**Meta- Ana;ysis (RS + LLD)**
		%	**N**	**CoE**	**SE**	**p-value**	**N**	**p-value**
Class	***Bacilli***	27.3%	867	5.4×10^{-03}	1.3×10^{-03}	3.6×10^{-05}	2,286	1.1×10^{-12}
Order	***Lactobacillales***	100%	864	4.9×10^{-03}	1.3×10^{-03}	2.4×10^{-04}	2,281	8.3×10^{-12}
Family	***Streptococcaceae***	79.6%	863	2.9×10^{-03}	1.3×10^{-03}	2.3×10^{-02}	2,265	2.1×10^{-08}
Genus	***Streptococcus***	98.7%	860	3.3×10^{-03}	1.6×10^{-03}	3.7×10^{-02}	2,256	1.3×10^{-08}

Adjusted for age, sex and, technical covariates: DNA isolation batch and TimeInMail. RS: Rotterdam Study (n=1427) P-value, CoE and SE from MaAsLin, LLD: LifeLines-DEEP (n=867) P-value, CoE and SE from MaAsLin analysis, Meta: Rotterdam Study and LifeLines Deep meta-analyzed together, sample size weighted inverse-variance meta-analysis in METAL. Taxonomy%=percentage of taxonomy is from one taxonomy level higher, ex. 23.7% of all Firmicutes are Bacilli. N=number of individuals in cohort where microbial abundancy is not zero for that taxonomy. FDR: P-value adjusted for multiple testing, Benjamin-Hochberg false discovery rate. CoE coefficient, SE standard error

Quantitative determination of Streptococcus spp

As 16S rRNA-sequencing derived relative microbiome data cannot provide information about the extent or directionality of changes in microbiome taxa abundance[30], we determined the absolute amount of *Streptococcus spp.* in the individuals of our study population. For each sample in our cohort (n = 1,427), we quantified the number of *Streptococcus spp.* using genus specific qPCR and the total microbial load using 16S rRNA qPCR. We calculated the absolute quantity of *Streptococcus spp.* and normalized for the total bacterial load in each samples as measured by 16S rRNA qPCR. The 16S rRNA-sequencing results and qPCR *Streptococcus spp.* quantity yielded similar results (Spearman correlation, r = 0.80, p-value = 2.2×10^{-16}, **Supplementary Figure 2**). Using the absolute abundance of *Streptococcus* measured by qPCR instead of the relative abundance derived from the 16S rRNA-sequencing profiles, we again found a significant association between higher knee WOMAC pain and greater absolute *Streptococcus spp.* abundance (beta = 0.10, p-value = 7.4×10^{-03}, Poisson regression), also after adjustment for smoking, alcohol consumption, and BMI (beta = 0.074, p-value = 4.5×10^{-02}, Poisson regression).

Table 3: Results of the association analysis of Streptococcus and knee joint effusion

Taxonomy		N	Model 1 CoE	Model 1 P-value	Model 2 CoE	Model 2 P-value
Class	***Bacilli***	314	9.4×10^{-03}	3.4×10^{-02}	2.7×10^{-03}	3.5×10^{-01}
Order	***Lactobacillales***	314	9.8×10^{-03}	2.7×10^{-02}	2.7×10^{-03}	3.6×10^{-01}
Family	***Streptococcaceae***	310	9.6×10^{-03}	1.7×10^{-02}	3.0×10^{-03}	2.6×10^{-01}
Genus	***Streptococcus***	308	1.0×10^{-02}	1.3×10^{-02}	3.3×10^{-03}	2.1×10^{-01}

Knee joint inflammation was measured as severity of effusion as measured on knee MRI. Knee MRI's were only available for an all-female obese subgroup of the Rotterdam Study Microbiome dataset (n = 373). First model assessed the association of Knee effusion with the microbiome, adjusted for age, sex, DNA isolation batch and TimeInMail (technical covariates). Second model was WOMAC-pain score adjusted for age, sex, technical covariates and, effusion severity. P-values were determined by MaAsLin analysis. N = number of individuals in cohort where microbial abundancy is not zero for that taxonomy. CoE coefficient

Independent replication of Streptococcus association

We sought replication for all four associations with WOMAC pain, i.e., class, order, family and, the bacterial genus of *Streptococcus,* in an independent Dutch cohort, Life-Lines-DEEP (LLD) (**Supplementary Table 7**). LLD has a lower sample size (n = 867), younger age (mean age = 45.6) and fewer individuals with OA-related knee pain (WOMAC pain > 0, n = 197), however, the average knee WOMAC pain was very similar compared with the Rotterdam Study cohort (RS = 0.9 and LLD = 0.9, **Table 1** and **Supplementary Table 7**). Despite lower power in the replication cohort we observed a significant association (p-value < 0.05) between knee WOMAC-pain scores and all four taxonomies to the genus of *Streptococcus* ($\text{coefficient}_{\text{replication}} = 3.3 \times 10^{-03}$, $\text{p-value}_{\text{replication}} = 3.7 \times 10^{-02}$, MaAsLin, **Table 2**). Also, in the meta-analysis of RS and LLD, greater *Streptococcus spp.* abundance was significantly associated with higher knee WOMAC pain ($\text{p-value}_{\text{meta}} = 1.3 \times 10^{-08}$, MaAsLin, n = 2256, **Table 2**). After adjusting for BMI, we found significant replication on class, order and the family level of *Streptococcacea* ($\text{coefficient}_{\text{replication}} = 2.7 \times 10^{-03}$, $\text{p-value}_{\text{replication}} = 3.5 \times 10^{-02}$, MaAsLin, **Supplementary Table 8**). In the replication study, additional adjustment for BMI slightly attenuated the association, a 9.1% decrease in coefficient, for the genus of *Streptococcus* (**Supplementary Table 8**). This was in concordance with the 14.5% decrease in coefficient for streptococcus, seen in the discovery cohort after BMI adjustment (**Supplementary Table 2**). In the BMI adjusted meta-analysis for RS and LLD, the association between knee WOMAC pain and *Streptococcus spp.* was highly significant ($\text{p-value}_{\text{meta}} = 1.1 \times 10^{-06}$, METAL, n = 2,256, **Supplementary Table 8**).

Streptococcus association with knee joint inflammation

If *Streptococcus spp.* abundance is causally related to higher knee WOMAC pain, a possible mechanism might be through local priming of macrophages in the synovial lining resulting in inflammation[22]. For a random subset of the Rotterdam Study Microbiome cohort, knee magnetic resonance images (MRI) were available (n = 373, all females)[32] at the time of microbiome measurements. Knee joint inflammation was assessed by scoring the amount of effusion in the tibiofemoral joint and in the patellofemoral joint for both knees. We found that higher WOMAC-pain scores were significantly correlated with more knee effusion (Spearman correlation r = 0.14, p-value = 9.0×10^{-03}). In addition, we observed that greater *Streptococcus spp.* abundance was significantly associated with more effusion in the knee joints (coefficient = 1.0×10^{-02}, p-value = 1.3×10^{-02}, MaAsLin, **Table 3**). When WOMAC pain was added to the model, the association of knee effusion with *Streptococcus spp.* disappears (coefficient = 3.3×10^{-03}, p-value = 0.21, MaAsLin), indicating that the association of *Streptococcus spp.* with knee pain severity is driven by knee inflammation severity.

Discussion

Using a large, deeply phenotyped population-based cohort, we identified a significant association between greater relative and absolute *Streptococcus spp.* abundance and higher OA-related knee pain. These results were validated by replication in an independent cohort and by meta-analysis. Finally, we presented evidence that this association was driven by local inflammation in the joint.

We observed that the intestinal microbiome β-diversity was significantly associated with knee WOMAC scores. After testing 256 taxonomies individually, we found a microbiome-wide association with knee WOMAC pain and *Streptococcus spp*., where greater *Streptococcus spp.* relative abundance is associated with higher knee WOMAC pain. We found this association to be robust, not be caused by outlier observations, or due to the confounding effects of smoking, alcohol intake, oral medication usage or BMI[28,29]. Neither was the association an artefact of the microbiome profiles as relative fractions of the 16S rRNA sequencing, or due to possible co-linearity in the data[30,33].

Although, the results of the sensitivity analyses were in line with a true association between *Streptococcus spp.* and knee WOMAC pain in our cohort, validation in an independent cohort is essential. Replication of microbiome abundance, however, is difficult, because data might not be similar between studies if sample preparation and

data analysis are not done in the exact same way[34]. We sought replication in the LLD cohort, which has a different study population and data preparation method than our cohort does[27]. Despite these differences, we could replicate our association in LLD for all taxonomic levels, class, order, family and for the genus of *Streptococcus.* Also, in the meta-analysis of RS and LLD, Streptococcus spp. was significantly associated with knee WOMAC pain.

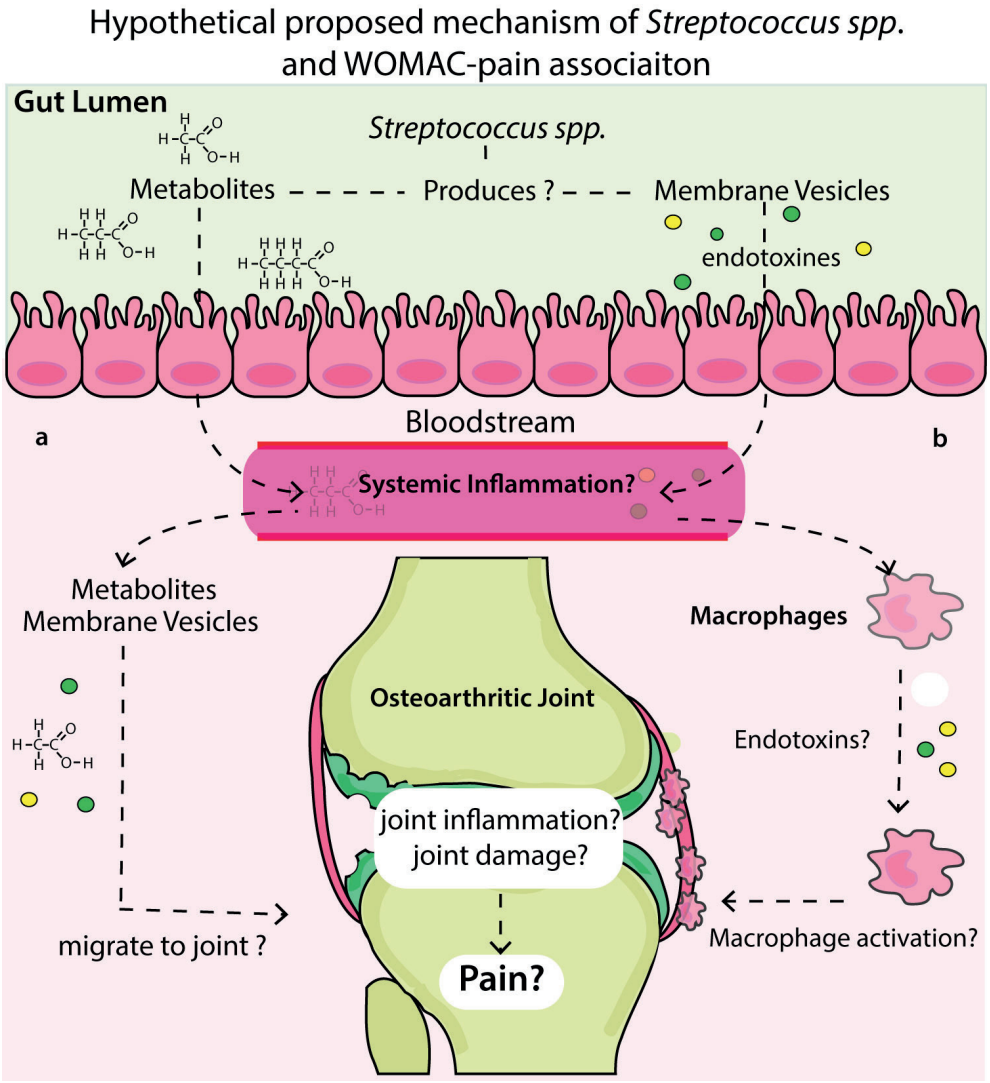
Obesity-mediated gastrointestinal-microbiome changes are postulated to affect low-grade systemic and local inflammation in OA[23-25]. Nevertheless, in our study the effect of *Streptococcus spp.* on knee WOMAC pain is not fully driven by BMI. This suggests a direct role for the gastrointestinal microbiome in OA-related knee pain and inflammation. We postulate that greater *Streptococcus spp.* abundance leads to higher knee WOMAC pain through local joint inflammation. This is in line with our observation that *Streptococcus spp.* abundance was significantly associated with effusion severity in the knee joints. This leads to believe that *Streptococcus spp.* might also be involved in other inflammatory joint pain disorders. This is not unlikely since several *Streptococcus spp.* have been linked to osteomyelitis[35,36], rheumatic fever[37,38] and, post-streptococcal reactive arthritis[39,40]. The last two are disorders in which due to molecular mimicry with group A *Streptococcus,* cross-reactive antibodies are produced against joint tissues, leading to rheumatic joint inflammation and damage[38]. However, these disorders involve mainly pathogenic species, such as *S. pyogenes*. Yet, most *Streptococcus spp.* are commensal species and have been found throughout the human oral-gastro-intestinal microbiome[41-43], still, these can produce immunogenic bacterial products. Several *Streptococcus spp.* have been shown to constitutively produce MVs[18]. These MVs may present *Streptococcus spp.* epitopes and/or may contain immunogenic products[17,24]. Such bacterial products can trigger macrophage activation through TLR pathways. This type of macrophage activation is predominantly seen in OA-related joint inflammation[16,21] and is thought to be related to pain[3-7]. We therefore propose that greater *Streptococcus spp.* abundance may lead to an increase of bacterial products in the circulation through increased production of metabolites that pass the gut-blood barrier or through immunogenic products that prime local or systemic macrophages. We have summarized this hypothetical model in **Figure 2**.

Our results show a difference in microbiome β-diversity in individuals with higher knee WOMAC-pain scores, which is driven by greater abundancy of *Streptococcus spp.* This association is highly robust. Moreover, we replicated our findings in an independent cohort. However, our study has some limitations. First, a cross-sectional study design was used and cannot unequivocally establish causality. For this, longitudinal studies are required. Second, other follow-up studies are needed to better elucidate

the molecular pathway connecting Streptococcus spp., knee inflammation and WOMAC pain, to validate or reject our postulated hypothesis (Figure. 2).

Blood and joint tissue could be examined for the presence of *Streptococcus* MVs, metabolites, and their possible association to knee inflammation and WOMAC pain severity. Third, to examine whether the association found in this study is unique for knee OA pain, other quantitative pain measurements should be examined, as well as measurements of pain at other joints. In, addition, other inflammatory joint disorders could also be examined. Forth, due to the limited resolution of 16S rRNA-sequencing methodology, we were unable to identify whether a specific *Streptococcus* species or strain was driving the association. Last, we find no association between knee OA severity (KLsum) and gastrointestinal-microbiome composition. Knee OA severity, however, was measured by radiographic OA severity, and this does not include the clinical symptom of joint pain. Although considering our proposed mechanism, an effect on knee OA severity could be expected, as inflammation can lead to joint damage. It is possible that our study currently lacks the power to detect such an association, or requires a longitudinal design to detect such effects on OA severity or disease progression.

In sum, we demonstrate an association between greater abundance of *Streptococcus spp.* and higher osteoarthritis-related knee pain, but the causality of this association needs to be established. A possible explanation for the found association is the induction or exacerbating of local joint inflammation by *Streptococcus spp.* The precise mechanisms by which *Streptococcus spp.* may trigger joint inflammation, are not known. We hypothesized that it may involve metabolites or MVs produced by *Streptococcus spp.* in the gastrointestinal tract. As joint pain is one of the most impacting clinical OA symptoms and pain in OA has a high socioeconomic burden, it is of crucial importance to identify novel treatments that reduce clinical symptoms, in particular pain, and decrease the socioeconomic burden. Our results point to the microbiome as a therapeutic target for OA-related joint pain and possibly for other inflammatory joint pain disorders. The gastrointestinal microbiome is a promising therapeutic target, because it is sensitive to change through (diet) interventions. This could offer an easily accessible and safe treatment options for OA associated joint pain. Therefore, it is pivotal to explore the causal mechanism and possible translation of the present findings into clinical practice.



▲Figure 2: Proposed hypothetical me pathophysiological mechanism explaining the association between the gut microbiome (Streptococcus spp.), knee WOMAC pain and knee effusion. No causality has been established between Streptococcus spp. abundance and OA-related knee pain, however, if such causality exists, we propose the following model: Members of the Streptococcus spp. are known to produce metabolites and membrane vesicles, which both may interact with host cells. These bacterial products can pass the gut-blood barrier, and possibly either a target the knee joint through activation of macrophages in the synovial lining, leading to joint inflammation and damage, or b enter the circulation, activate macrophages to pro-inflammatory macrophages, which may trigger a low-grade systemic inflammatory state, invoking or exacerbate joint inflammation and damage, leading to increased knee pain

Methods

Rotterdam study cohort

The Rotterdam study (RS) is a large population-based prospective study population (14,926 participants) ongoing since 1990 to study determinants of chronic disabling diseases. The first cohort RS-I, started in 1990 and includes individuals of 55 years and older living in the Ommoord district of Rotterdam in the Netherlands. In 2000, a second cohort was started RS-II with individuals who had become 55 years of age or moved into the study district since the start of the study. In 2006, the third cohort was initiated, RSIII, of individuals aged > 45, living in the Ommoord district. The cohorts are predominantly Caucasian and a further detailed description of the design and rationale of the Rotterdam Study has been published elsewhere[44]. The Rotterdam Study has been approved by the Medical Ethical Committee of the Erasmus MC, University Medical Center Rotterdam, the Netherlands (MEC 02.1015). All subjects provided written consent prior to participation in the Rotterdam Study. Stool sample collection started in 2012 during the second visit of the Rotterdam Study III population (RSIII-2). A random group of 2440 participants was invited to provide a stool sample. In total 1691 (response rate = 69%) stool samples were received via mail at the Erasmus MC for analysis of stool microbiome composition as a marker for gastrointestinal-microbiome composition. After quality control 1,427 samples remained for further analysis.

Taxonomic profiling of gastrointestinal microbiota in RS

Sample collection: Stool samples were collected at home by the participant using a Commode Specimen Collection System (Covidien, Mansfield, MA) and ~1 g aliquot was transferred to a 25 × 76 mm feces collection tube, which was sent via the regular mail to the Erasmus MC. Participants also filled out a short questionnaire addressing date and time of defecation, current or recent antibiotics use, current probiotics use, and recent travel activities. Upon arrival samples were stored (−20 °C), samples taking longer than 3 days to arrive at the Erasmus MC were excluded from further analysis, for the remaining samples TimeInMail (in days) was registered as a technical covariate. DNA isolation: For each participant, frozen stool samples were allowed to thaw for 10 min at room temperature prior to DNA isolation. An Aliquot of ~300 mg of stool was homogenized using 0.1 mm silica beads (MP Biomedicals, LLC, Bio Connect Life Sciences BV, Huissen, The Netherlands) and DNA was isolated from the samples using the Arrow stool DNA kit according to the manufacturers' protocol (Arrow Stool DNA; Isogen Life Science, de Meern, The Netherlands). 16S rRNA gene sequencing: The V3 and V4 hypervariable regions of the 16S rRNA gene were amplified using the 319F (ACTCCTACGGGAGGCAG-

CAG) –806R (GGACTACHVGGGTWTCTAAT) primer pair and dual indexing[45], a full list of all primers used can be found in Supplementary Table 9. Amplicons were normalized using the SequalPrep Normalization Plate kit (Thermo Fischer Scientific) and pooled. Amplicon pools were purified prior to sequencing (Agencourt AMPure XP, Beckman Coulter Life Science, Indianapolis, IN) and size and quantity was assessed (LabChip GX, PerkinElmer Inc., Groningen, The Netherlands). A Control library was added to ~10% of each amplicon pool as a positive control (PhiX Control v3 library, Illumina Inc., San Diego, CA). The hypervariable V3 and V4 regions of the 16S rRNA gene were sequenced in paired-end mode (2 × 300 bp) using the MiSeq platform (Illumina Inc., San Diego, CA) with an average depth of 50,000 paired reads per sample. Data processing and quality control: Sequence read quality control and taxonomic classification, was done using an in-house pipeline (µRAPtor) based on QIIME version 1.9.0[46] and UPARSE version 8.1[47]. Low-quality, merged, and chimeric reads were excluded. Duplicate samples, samples with <10,000 reads, and samples from participants that have used antibiotics (self-reported) in the 6 months prior to sample production were excluded. The remainder of the reads (~93%) were normalized using random 10,000 read subsampling (rarefication). To reconstruct taxonomic composition a direct classification of 16S sequencing reads using RDP classifier (2.12) and the SILVA 16S rRNA database (relase.128)[48]. This was done for each taxonomic level available: domain, phylum, class, order, family, and genus, with binning posterior probability cut-off of 0.8. The microbial Shannon diversity index was calculated on the taxonomic level of genera, using vegan package in R. Relative abundancies were calculated for each taxonomic level prior to any additional QC (domain, phylum, class, order, family and genus). Unknown or unassigned classifications were excluded from the final dataset (n = 69), as these currently cannot be used for clinical or therapeutic applications. In total, we were left with gut-microbiome taxonomies for 1427 individuals and 596 taxonomic classifications for analysis (Figure 1).

Phenotype descriptions

Osteoarthritic knee joint pain was determined by the WOMAC questionnaire, which is a disease specific questionnaire to assess the severity of hip and knee OA, consisting of 24 items covering three domains: pain, stiffness, and function[49]. The WOMAC includes five items that measure OA joint specific pain, scored on a 5-point Likert scale, 0–4. The five knee specific WOMAC-pain scores were summed to create the knee WOMAC pain score, ranging from 0 to 20, where higher scores represent worse OA-related knee joint pain. Knee OA severity was determined by the radiographic Kellgren and Lawrence score (KL-score)[50]. Using radiographs of both knee joints, left and right, the KL-score was determined for each joint. Scores were subsequently summed for the left and right knee to form the Knee KLsum score. Knee joint effusion could be determined,

for an all-female random subset of the Rotterdam Study III cohort (RSIII), by knee MRIs using a 1.5T MRI scanner (General Electric Healthcare, Milwaukee, Wisconsin, USA). For further detail of our used MRI protocol see ref. [51]. All MRI images were scored by a trained reader (blinded for clinical, radiographic and genetic data). Joint effusion was determined in the tibiofemoral joint (TFJ) and in the patellofemoral joint (PFJ) together (grade 0–3: 0 = no joint effusion, 1 = small joint effusion, 2 = moderate joint effusion, and 3 = severe joint effusion). The scores of the left and the right knee were summed, resulting in a score from 0 to 6, where higher scores represent more severe knee joint effusion. For all phenotypes, if data on either the left or right knee joint was missing, individuals were excluded. Oral medication usage such as, proton pump inhibitors (PPIs) and non-steroidal anti-inflammatory drugs (NSAIDs), were determined by questionnaire at the same time point as WOMAC scores, knee X-rays, knee MRIs, and stool sample collection.

Lifelines-DEEP replication cohort

The LifeLines-DEEP (LLD) cohort is a subcohort of the LifeLines cohort (167,729 participants) that employs a broad range of investigative procedures to assess the biomedical, socio-demographic, behavioral, physical and psychological factors that contribute to health and disease in the general Dutch population[52]. A subset of approximately 1500 participants was included in LLD subcohort. For these participants, additional biological materials were collected, including analysis of the gut-microbiome composition. The collection, phenotyping, and processing of LLD have been described in detail[27,53]. Briefly, microbiome data was generated for 1179 LLD samples. Fecal samples were collected at home within two weeks of blood sample collection and stored immediately at 20 °C. After transport on dry ice, fecal samples were stored at −80 °C. Aliquots were made and DNA was isolated with the AllPrep DNA/RNA Mini Kit (Qiagen; cat. #80204). The 16S rRNA gene of the isolated DNA was sequenced at the Broad Institute, Boston, using Illumina MiSeq pair-ends. Hypervariable region V4 was selected using forward primer 515F (GTGCCAGCMGCCGCGGTAA) and reverse primer 806 R (GGACTACHVGGGTWTCTAAT) (Supplementary Table 9)[54]. Closed-reference OUT picking has been done with 97% similarity cutoff using UCLUST[55] program and GreenGenes 13.5 reference database[56] from QIIME[46] software. Overall, for 878 samples, both WOMAC scores and microbiome information was available. The library size of microbial sequencing was rarefied to 10,000 read-depth using the rarefy function in R package vegan (version 2.5–2). At this depth, 11 subjects were excluded. After the exclusion step, we had 867 samples (362 men and 505 women) remained for the final analysis. Their characteristics are summarized in Supplementary Table 7. The LifeLines-DEEP study has been approved by the medical ethical committee of the University Medical Center Groningen, The Netherlands.

qPCR replication of 16S rRNA-sequencing results

To validate the 16S rRNA-sequencing results we determined the absolute quantitative amount of Streptococcus spp. using qPCR. We determined the amount of Streptococcus spp. for each fecal DNA sample using the BactoReal qPCR assay (Ingenetix GmbH, Vienna, Austria) based on the Streptococcus 23S rRNA gene. A standard curve of a Plasmid standard containing the 23S rRNA gene (Ingenetix GmbH, Vienna, Austria) was included in each plate to calculate the amount of Streptococcus spp. present in each sample. Each sample was run in duplo in a 384-wells PCR plate containing 40 ng fecal DNA, 1x primers and probes (Ingenetix GmbH, Vienna, Austria), 1x TaqMan Gene Expression Master Mix (Life technologies) in a total volume of 5 µl. The qPCR was performed in a QuantStudio 7 Flex (ThermoFisher) with an initial denaturation at 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 60 s. To normalize the amount of Streptococcus spp. for the total amount of bacteria in each sample, a 16 S qPCR was performed (U16SRT-F: ACTCCTACGGGAGGCAGCAGT and U16SRT-R: TATTACCGCGGCTGCTGGC, Supplementary Table 9)[57]. A standard curve of a Bacterial DNA Standard (Zymo Research) sample was included in each PCR plate to calculate the total amount of bacteria. Each sample was run in duplo in a 384-well plate containing 200 pg fecal DNA, 200 pmol forward and reverse primers, 1x SYBR Fast ABI PrismTM Mastermix (Kapa-Biosystems) in a total volume of 5 µl. The qPCR was performed in a QuantStudio 7 Flex (ThermoFisher) with an initial denaturation at 95 °C for 3 min, followed by 40 cycles of 95 °C for 5 s and 60 °C for 20 s. Absolute abundancies were calculated from the standard curves. We adjusted the total absolute abundancy of bacteria for the average number of 16S copies (4.2 copies per bacteria[58]). We adjusted the absolute abundance of Streptococcus spp. for the average number of 23S rRNA gene copies in Streptococcus spp. (3.36 copies per Streptococcus spp., based on S. pneumoniae, S. thermophiles, S. pyogens, S. parasanguinis, S. dysgalactiae, S. salivarius, S. suis, S. mutans, and S. agalactiae). We normalized the absolute abundance of Streptococcus spp. by calculating the log transformed value of the number of Streptococcus spp. per 1000 bacteria in each sample: log(Streptococcusabsoluteabundancetotalabundanceofbacteria/1000)

Statistical analysis

Statistical analyses were performed in R: A Language and Environment for Statistical Computing[59]. Inter-individual microbial composition (β-diversity) was calculated using the Aitchison distance calculated from the CLR (centered log-ratio) normalized data. CLR normalization was calculated on the counts of the directly taxonomic classified reads. The full dataset, including unknown and unclassified taxonomies was used for the CLR. To all read counts 1 was added to cope with the overabundance of zero's in the data, for CLR[33]. Subsequent statistical analysis was done through permutation

analysis of variance (PERMANOVA) to inspect the global effect of WOMAC-pain score and KLsum score on the overall microbiome profiles, using the adonis PERMANOVA function in VEGAN. PERMANOVA models included age, sex, TimeInMail, DNA isolation batch, BMI and WOMAC-pain score or KLsum, in that order. Results were visualized using a PCA plot. CLR method and PCA plot were based on Gloor et al.[33]. Intra-individual microbial composition metrics (α-diversity) used are the Shannon Index and Inverse Simpson Index. Association of α-diversity with WOMAC pain score was examined by Poisson-regression model adjusted for age, sex, batch and TimeInMail. To identify associated microbiological taxa with the investigated phenotypes, we have used the multivariate statistical linear regression analysis, R package: MaAsLin[60]. MaAsLin is a specialized statistical R package for the analysis of microbial community abundance data and clinical/phenotypic metadata. All unknown and unclassified taxonomies were excluded from this analysis (n = 63). In MaAsLin, we have used adjusted default settings, i.e., Linear model based, quality control (QC) and exclusion of outliers based on the Grubbs test on the microbiome data only, and arcsine-square-root transformation of the single taxonomies relative abundance table to normalize the microbiome profiles. After MaAsLin QC we were left with 256 taxonomies for analysis (2 domains, 12 phyla, 18 classes, 25 orders, 41 families, and 158 genera). Missing values were not imputed, nor was automatic QC of the metadata, WOMAC pain score and covariates, or boosting by excluding metadata from the association analyses performed. For each analysis we forced the following cofactors: age (years), sex (0/1), technical covariates: DNA isolation batch (0/1) and TimeInMail (days). Depending on the model we additionally forced the following cofactors: BMI (body mass Index), smoking (current smoker, y/n), daily alcohol consumption (glass/day), PPI use (y/n), and NSAID use (y/n). Statistical significance was determined by multiple testing correction, Benjamini-Hochberg False discovery rate (FDR) < 0.05. Removal of possible collinearity in the microbiome data was done by ILR (isometric log-ratio transformation) on the counts of the directly taxonomic classified reads. The full dataset including unknown and unclassified taxonomies was used. Multivariate linear regression model adjusted for age, sex, and cohort-specific technical covariates was performed on the ILR transformed data. Replication analysis in LifeLines-DEEP (LLD), used MaAsLin using similar settings, with the exception of the automatic QC in MaAsLin. In LLD the QC of the metadata was done manually. All analysis were also adjusted for age, sex, and cohort-specific technical covariates. Meta-analysis of RS and Lifelines was performed using inverse-variance weighting by METAL[61]. Correlation between 16S sequencing Streptococcus spp. abundancy and qPCR Streptococcus spp. abundance was done by Spearman correlation in R. Association of knee WOMAC pain and qPCR data were done by Poisson-regression models adjusted for age, sex, and qPCR technical covariates (plate number). All figures and graphs were made in R and adapted in Adobe Illustrator.

Supplementary Material

Additional material, as well as a funding statement, is published online only, to view please visit the paper online at the journal website:

https://www.nature.com/articles/s41467-019-12873-4#MOESM1

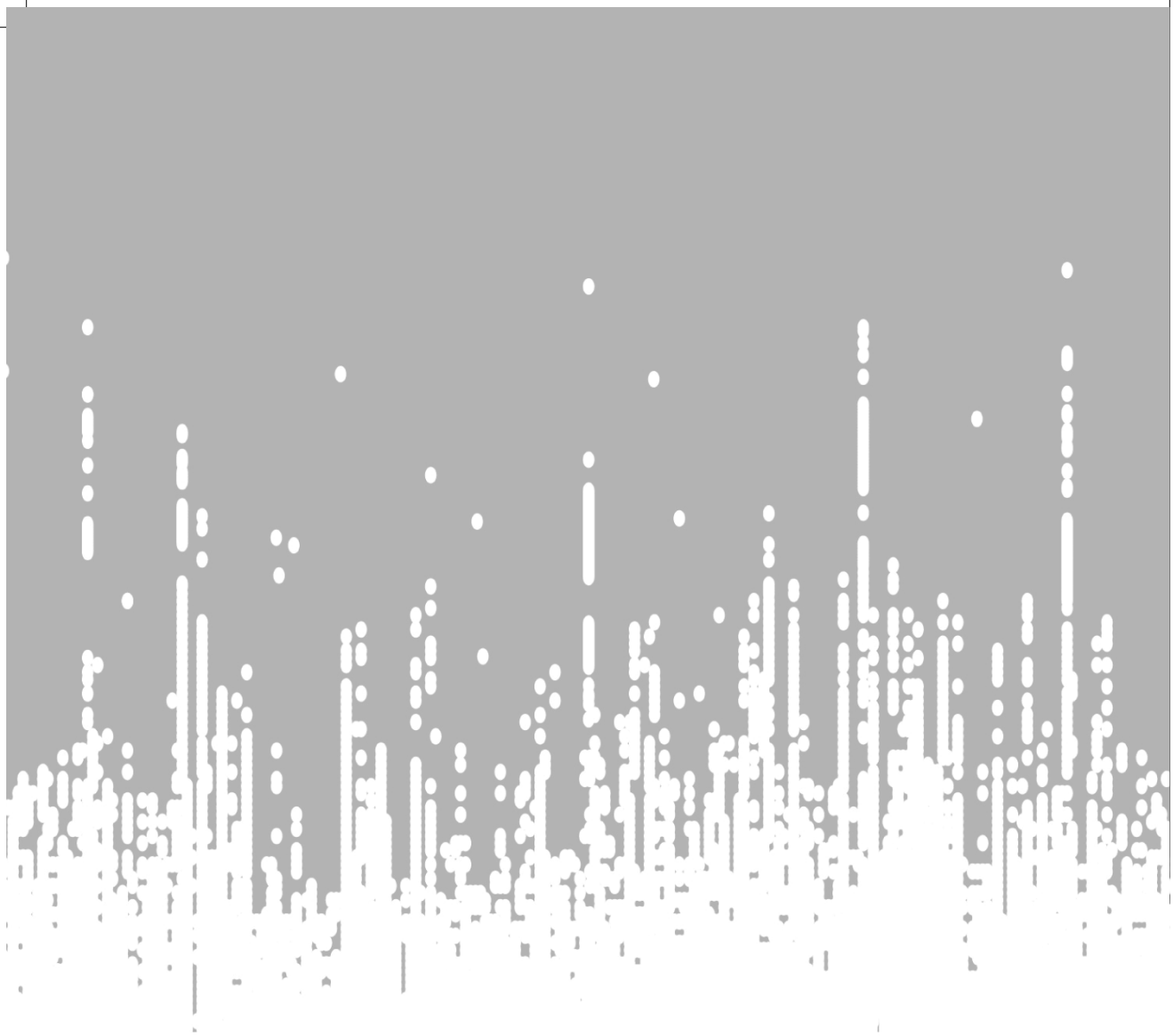
References

1. Centers for Disease Control and Prevention (CDC). Prevalence of doctor-diagnosed arthritis and arthritis-attributable activity limitation–United States, 2010–2012. Mmwr. Morb. Mortal. Wkly. Rep. 62, 869–873 (2013).
2. Centers for Disease Control and Prevention (CDC). Prevalence and most common causes of disability among adults–United States, 2005. Mmwr. Morb. Mortal. Wkly. Rep. 58, 421–426 (2009).
3. Ayral, X., Pickering, E. H., Woodworth, T. G., Mackillop, N. & Dougados, M. Synovitis: a potential predictive factor of structural progression of medial tibiofemoral knee osteoarthritis–results of a 1 year longitudinal arthroscopic study in 422 patients. Osteoarthr. Cartil. 13, 361–367 (2005).
4. Kraus, V. B. et al. Direct in vivo evidence of activated macrophages in human osteoarthritis. Osteoarthr. Cartil. 24, 1613–1621 (2016).
5. Dahaghin, S., Bierma-Zeinstra, S. M. A., Koes, B. W., Hazes, J. M. W. & Pols, H. A. P. Do metabolic factors add to the effect of overweight on hand osteoarthritis? The Rotterdam Study. Ann. Rheum. Dis. 66, 916–920 (2007).
6. Saberi Hosnijeh, F. et al. Association between biomarkers of tissue inflammation and progression of osteoarthritis: evidence from the Rotterdam study cohort. Arthritis Res. Ther. 18, 81 (2016).
7. Utomo, L., Bastiaansen-Jenniskens, Y. M., Verhaar, J. A. N. & van Osch, G. J. V. M. Cartilage inflammation and degeneration is enhanced by pro-inflammatory (M1) macrophages in vitro, but not inhibited directly by anti-inflammatory (M2) macrophages. Osteoarthr. Cartil. 24, 2162–2170 (2016).
8. Reyes, C. et al. Association between overweight and obesity and risk of clinically diagnosed knee, hip, and hand osteoarthritis: a population-based cohort study. Arthritis Rheumatol. 68, 1869–1875 (2016).
9. Berenbaum, F., Griffin, T. M. & Liu-Bryan, R. Review: metabolic regulation of inflammation in osteoarthritis. Arthritis Rheumatol. 69, 9–21 (2017).
10. Winer, D. A., Luck, H., Tsai, S. & Winer, S. The intestinal immune system in obesity and insulin resistance. Cell Metab. 23, 413–426 (2016).
11. Cani, P. D. et al. Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet-induced obesity and diabetes in mice. Diabetes 57, 1470–1481 (2008).
12. Cani, P. D. et al. Changes in gut microbiota control inflammation in obese mice through a mechanism involving GLP-2-driven improvement of gut permeability. Gut 58, 1091–1103 (2009).
13. Turnbaugh, P. J. et al. A core gut microbiome in obese and lean twins. Nature 457, 480–484 (2009).
14. Brun, P. et al. Increased intestinal permeability in obese mice: new evidence in the pathogenesis of nonalcoholic steatohepatitis. Am. J. Physiol. Liver Physiol. 292, G518–G525 (2007).
15. Luck, H. et al. Regulation of obesity-related insulin resistance with gut anti-inflammatory agents. Cell Metab. 21, 527–542 (2015).
16. Kaparakis-Liaskos, M. & Ferrero, R. L. Immune modulation by bacterial outer membrane vesicles. Nat. Rev. Immunol. 15, 375–387 (2015).
17. Bonnington, K. E. & Kuehn, M. J. Protein selection and export via outer membrane vesicles. Biochim. Biophys. Acta 1843, 1612–1619 (2014).
18. Brown, L., Wolf, J. M., Prados-Rosales, R. & Casadevall, A. Through the wall: extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi. Nat. Rev. Microbiol. 13, 620–630 (2015).
19. Bomberger, J. M. et al. Long-distance delivery of bacterial virulence factors by Pseudomonas aeruginosa outer membrane vesicles. PLoS Pathog. 5, e1000382 (2009).
20. Levy, M., Blacher, E. & Elinav, E. Microbiome, metabolites and host immunity. Curr. Opin. Microbiol. 35, 8–15 (2017).
21. Ohto, U., Fukase, K., Miyake, K. & Shimizu, T. Structural basis of species-specific endotoxin sensing by innate immune receptor TLR4/MD-2. Proc. Natl Acad. Sci. USA 109, 7421–7426 (2012).
22. Huang, Z. Y., Stabler, T., Pei, F. X. & Kraus, V. B. Both systemic and local lipopolysaccharide (LPS) burden are associated with knee OA severity and inflammation. Osteoarthr. Cartil. 24, 1769–1775

(2016).

23. Huang, Z. & Kraus, V. B. Does lipopolysaccharide-mediated inflammation have a role in OA? Nat. Rev. Rheumatol. 12, 123–129 (2015).
24. Metcalfe, D. et al. Does endotoxaemia contribute to osteoarthritis in obese patients? Clin. Sci. (Lond.). 123, 627–634 (2012).
25. Schott, E. M. et al. Targeting the gut microbiome to treat the osteoarthritis of obesity. JCI Insight 3 (2018).
26. Deschasaux, M. et al. Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography. Nat. Med. 24, 1526–1531 (2018).
27. Zhernakova, A. et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science 352, 565–569 (2016).
28. Capurso, G. & Lahner, E. The interaction between smoking, alcohol and the gut microbiome. Best. Pract. Res. Clin. Gastroenterol. 31, 579–588 (2017).
29. Imhann, F. et al. Proton pump inhibitors affect the gut microbiome. Gut 65, 740–748 (2016).
30. Vandeputte, D. et al. Quantitative microbiome profiling links gut community variation to microbial load. Nature 551, 507–511 (2017).
31. Gloor, G. B., Wu, J. R., Pawlowsky-Glahn, V. & Egozcue, J. J. It's all relative: analyzing microbiome data as compositions. Ann. Epidemiol. 26, 1–8 (2016).
32. Schiphof, D. et al. Sensitivity and associations with pain and body weight of an MRI definition of knee osteoarthritis compared with radiographic Kellgren and Lawrence criteria: a population-based study in middle-aged females. Osteoarthr. Cartil. 22, 440–446 (2014).
33. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome datasets are compositional: and this is not optional. Front. Microbiol. 8, 2224 (2017).
34. Sinha, R., Abnet, C. C., White, O., Knight, R. & Huttenhower, C. The microbiome quality control project: baseline study design and future directions. Genome Biol. 16, 276 (2015).
35. McGuire, T., Gerjarusak, P., Hinthorn, D. R. & Liu, C. Osteomyelitis caused by beta-hemolytic streptococcus group B. JAMA 238, 2054–2055 (1977).
36. Murillo, O. et al. Streptococcal vertebral osteomyelitis: multiple faces of the same disease. Clin. Microbiol. Infect. 20, O33–O38 (2014).
37. Cunningham, M. W. Rheumatic fever, autoimmunity, and molecular mimicry: the streptococcal connection. Int. Rev. Immunol. 33, 314–329 (2014).
38. Tandon, R. et al. Revisiting the pathogenesis of rheumatic fever and carditis. Nat. Rev. Cardiol. 10, 171–177 (2013).
39. Mackie, S. L. & Keat, A. Poststreptococcal reactive arthritis: what is it and how do we know? Rheumatology 43, 949–954 (2004).
40. Barash, J. Rheumatic fever and post-group A streptococcal arthritis in children. Curr. Infect. Dis. Rep. 15, 263–268 (2013).
41. Dewhirst, F. E. et al. The human oral microbiome. J. Bacteriol. 192, 5002–5017 (2010).
42. Pei, Z. et al. Bacterial biota in the human distal esophagus. Proc. Natl Acad. Sci. USA 101, 4250–4255 (2004).
43. Bik, E. M. et al. Molecular analysis of the bacterial microbiota in the human stomach. Proc. Natl Acad. Sci. USA 103, 732–737 (2006).
44. Ikram, M. A. et al. The Rotterdam Study: 2018 update on objectives, design and main results. Eur. J. Epidemiol. 32, 807–850 (2017).
45. Fadrosh, D. W. et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome 2, 6 (2014).
46. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7, 335–336 (2010).

47. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat. Methods 10, 996–998 (2013).
48. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41, D590–D596 (2012).
49. Bellamy, N. & Buchanan, W. W. A preliminary evaluation of the dimensionality and clinical importance of pain and disability in osteoarthritis of the hip and knee. Clin. Rheumatol. 5, 231–241 (1986).
50. Kellgren, J. H. & Lawrence, J. S. Radiological assessment of osteo-arthrosis. Ann. Rheum. Dis. 16, 494–502 (1957).
51. Peters, M. J. et al. Associations between joint effusion in the knee and gene expression levels in the circulation: a meta-analysis. F1000Research 5, 109 (2016).
52. Scholtens, S. et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. Int. J. Epidemiol. 44, 1172–1180 (2015).
53. Tigchelaar, E. F. et al. Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. BMJ Open 5, e006772 (2015).
54. Fu, J. et al. The gut microbiome contributes to a substantial proportion of the variation in blood lipids. Circ. Res. 117, 817–824 (2015).
55. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461 (2010).
56. DeSantis, T. Z. et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72, 5069–5072 (2006).
57. Clifford, R. J. et al. Detection of bacterial 16S rRNA and identification of four clinically important bacteria by real-time PCR. PLoS ONE 7, e48558 (2012).
58. Větrovský, T. & Baldrian, P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. PLoS ONE 8, e57923 (2013).
59. Team, R. R: A Language and Environment for Statistical Computing. (2013).
60. Morgan, X. C. et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol. 13, R79 (2012).
61. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191 (2010).



Chapter 6

General Discussion

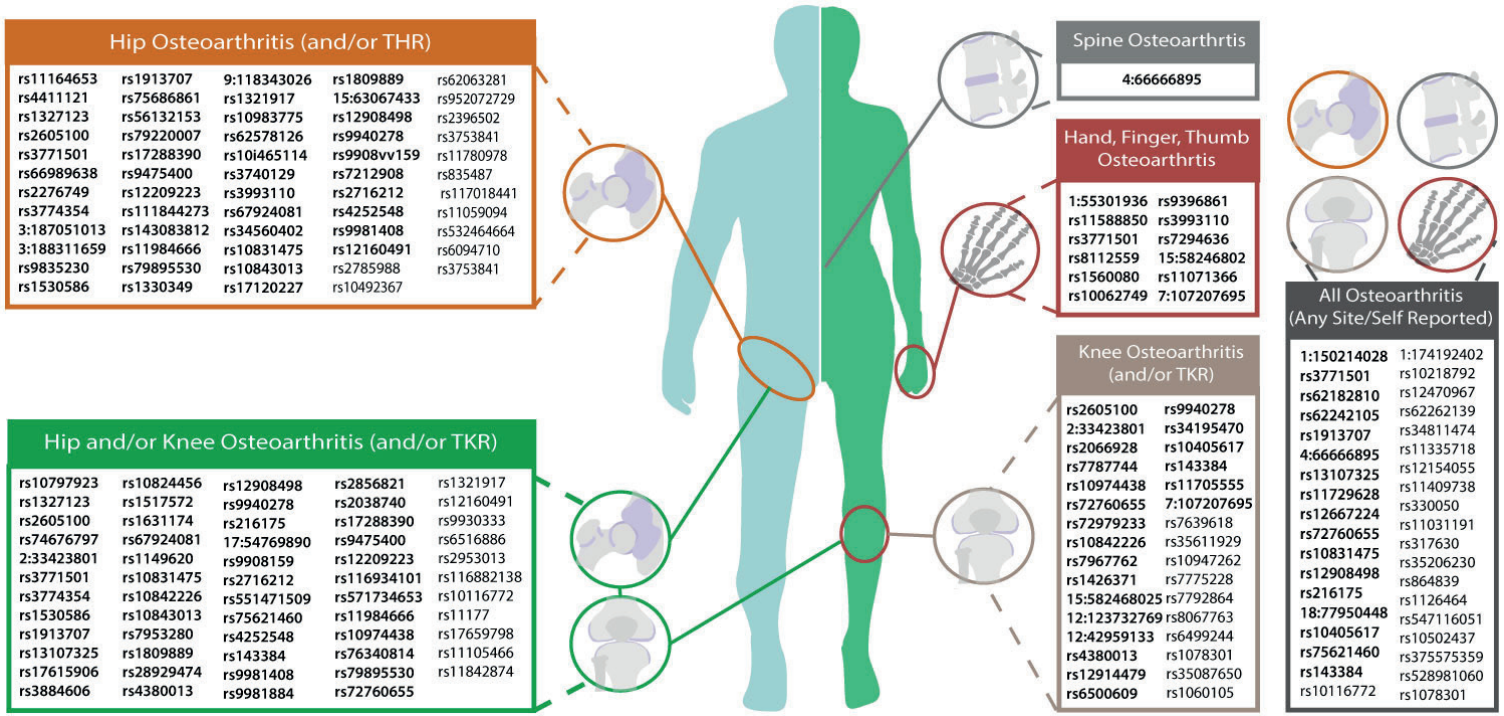
The overall objective of this thesis was to gain greater insight into the pathology of osteoarthritis, with the hope to bring possible preventive or curative treatments closer for one of the world's oldest diseases. In the previous chapters we have presented novel genetic loci, genes, biological pathways and risk factors for osteoarthritis. There the main findings, limitations and implications of each study were discussed. In this chapter, all findings presented in this thesis are brought together, considerations are addressed and suggestions for future osteoarthritis research are given.

Osteoarthritis associated genetic variation

In this thesis, genome-wide association studies (GWASs) were used to identify novel genetic loci associated with osteoarthritis, in order to identify novel genes and biological pathways involved in osteoarthritis pathology. Prior to the start of this thesis the number of identified osteoarthritis associated loci was severely limited, only 17 loci were known[1], a far cry of the hundreds predicted to be involved**.** Currently 148 osteoarthritis associated loci have been reported. In **Figure 1** all currently reported genome-wide osteoarthritis associated SNVs are listed, those mentioned in this thesis are highlighted.

GWAS size matters and more is needed

The success of GWASs to robustly identify associated genetic loci, is dependent on several factors; (i) the phenotype studied must have a genetic component, (ii) the quality and number of genetic variants that are used in the GWAS, (iii) the sample size (power) of the GWAS, (iv) and the heterogeneity of the studied phenotype[2]. Especially sample size is one of the most limiting factors for statistical power in GWAS[2]. This is demonstrated in **Chapter 4**. Before this thesis, the largest osteoarthritis GWAS consisted of ~7,000 osteoarthritis cases (~42,000 controls) and identified five loci[1]. Now the largest osteoarthritis GWAS effort, the Genetics of Osteoarthritis (GO) consortium, (~180.000 cases and ~600.000 controls) identified 100 osteoarthritis associated loci (**Chapter 4**). It is this increase in sample size, driven by the availability of large scale biobanks, such as the UKbiobank, and increased (inter)national collaboration efforts[9-11], that has caused the increase in associated osteoarthritis loci. Yet, despite these advances and progress made by the GO consortium, the proportion of genetic heritability explained by the found SNVs for osteoarthritis is still relatively low, ranging between 6% for hand osteoarthritis to 21% for hip osteoarthritis and total hip replacement (**Table 1**)**.** Indicating that still much of the osteoarthritis genetics, remains undiscovered.



▲ **Figure 1: All currently known osteoarthritis associated loci**. In bold are the variants mentioned in this thesis. For variants in high LD ($r^2 \geq 0.6$, $D' \geq 0.6$) only one variant was depicted in the figure. TKR: total knee replacement, THR: total hip replacement, TJR: total joint replacement.

Table 1: Osteoarthritis heritability and variance explained in the Genetics of Osteoarthritis Consortium

Phenotype	Genetic heritability* (H^2)	GWAS SNV heritability† (h^2)	Total genetic heritability explained‡
Hip OA	58% (29%-87%) [3]	12%	21% (14%-43%)
THR	73% (66%-78%)[4]	15%	21% (19%-23%)
Knee OA	44% (31%-56%)[5]	5%	11% (8%-15%)
TKR	53% (31%-75%)[6]	5%	10% (7%-18%)
Spine OA	75% (30%-100%) [7]	10%	13% (10%-14%)
Hand OA	56% (34%-76%) [7]	4%	6% (5%-11%)
Finger OA	42% (26%-58%) [8]	5%	12% (9%-20%)
Thumb OA	53% (37%-69%)[8]	4%	8% (6%-11%)

*Osteoarthritis genetic heritability as determined by twin studies.
† Predicted proportion of the SNV caused genetic heritability explained by the genome-wide significant SNVs in the GO Consortium
‡Percentage of the total genetic heritability explained by the genome-wide significant SNVs of the GO Consortium and the 95% Confidence Interval(CI)(H^2/h^2).
OA: osteoarthritis, SNV: Single Nucleotide Polymorphism, THR: Total Hip Replacement, TKR: total Knee Replacement

The genetic architecture of osteoarthritis may explain the current lack of explained genetic heritability. Osteoarthritis genetic architecture can be described by the polygenic model. Under the assumption of polygenic genetic architecture many genetic variants collectively contribute to disease risk[12]. Rare and/or uncommon variations with very large effect sizes cause more severe and/or early-onset diagnoses, in line with the monogenetic(early-onset) forms of osteoarthritis. Meaning that the effect sizes for common and rare variants linked with the non-monogenetic forms of osteoarthritis are likely to be small to very small[12]. This also means that on an individual level osteoarthritis genetic risk is caused by a (near) unique combination of many rare and common risk variations, each contributing a small effect to the over genetic risk for osteoarthritis. Thus, in order to identify these many rare and common variants with small effect sizes, massive and population specific sample sizes are needed to sufficiently increase GWAS power to identify all of them and to fully explain all of the osteoarthritis genetic risk.

Phenotyping is essential for GWAS

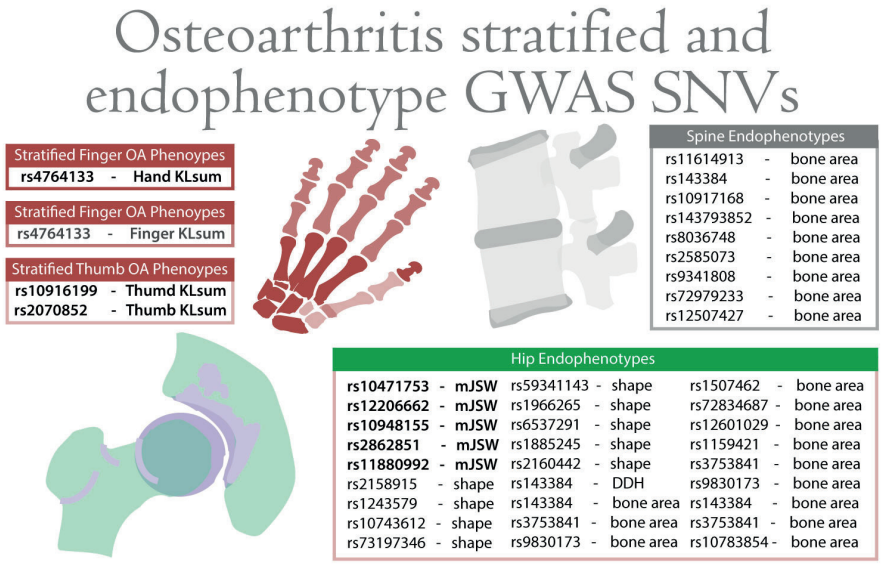
However, increasing sample sizes for GWAS is a double edged sword: it generates more power through larger sample sizes but also more noise due to heterogeneity in the diverse cohorts[13]. Increasing sample sizes in GWAS also increases the burden and costs of phenotyping. For very large studies (e.g., UK biobank, 23andMe) extensive phenotyping is not feasible, thus more easily and cost effective phenotypes are collected (hospital records, self-reported diagnosis and complaints). Such “minimal phenotyping”[14] has

several advantages: such as minimizing phenotyping costs and reducing phenotypes to single or few data points, allowing for easy large scale data collection. However, minimal phenotypes can introduce significant bias in the phenotype and reduce GWAS power[14]. For example individuals with the same hospital diagnosis of osteoarthritis (cases) may not have the same symptoms, pathology or underlying causal pathways, as osteoarthritis is a highly heterogenous disease [15, 16]. This heterogeneity also occurs for self-reported osteoarthritis, with an additional bias of uncertainty if the self-reported diagnosis is an accurate one.

In **Chapter 4**, not only different osteoarthritis case definitions were used in the different cohorts, with the majority being the minimal phenotype of self-reported/hospital diagnosed osteoarthritis. This heterogeneity between phenotype definitions and bias from the minimal phenotypes significantly introduces noise into the GWAS results[17]. Specifically, some of the identified loci could be nonspecific for osteoarthritis[14], i.e., misdiagnosis in self-reported osteoarthritis, which would have significant implications for follow-up molecular characterization and pathway identification. Additionally, osteoarthritis was defined as a dichotomous phenotype: present or absent. However, this black and white presentation of osteoarthritis completely ignores the underlying genetic, biological and pathological complexity of this disease[13, 18, 19]. Thus these biases may explain the low genetic osteoarthritis heritability found in **Chapter 4**. Also, they illustrate the need for phenotypic heterogeneity reduction methods and the use of phenotypes that do account for the genetic, biological and pathological complexity of osteoarthritis.

Phenotyping: Quantitative phenotypes

One way to increase GWAS power and use phenotypes that can more accurately represent the range of osteoarthritis disease severity, is to use quantitative phenotypes [17]. In **Chapter 2.2** and **3.1**, the radiographic summarized osteoarthritis phenotype of KLsum was used. The advantage of this phenotype is that it measures osteoarthritis severity across multiple joints at the same time. Creating a semi-quantitative measure, more accurately representing the severity of osteoarthritis in these joints than a dichotomous trait could do. Thus improving GWAS power, as is evident by the thumb KLsum and osteoarthritis associated SNV rs10916199 in the *WNT9A* locus. This SNV was found to be genome-wide associated (p-value=$2.*10^{-12}$) with thumb KLsum in a modest total sample size of ~8,700 (**Chapter 3.2**), however this variant was only found to be genome-wide significant associated ($3.5*10^{-10}$) with thumb OA in the GO consortium with a sample size of 10,536 cases and 236,919 controls (**Chapter 4**). Similar results were found for the osteoarthritis of the finger and hand KLsum associated SNV, rs4764133 located near *MGP* (**Chapter 3 and Chapter 4**)



▲Figure 2: All known osteoarthritis stratified and endophenotype associated loci. In bold are the variants mentioned in this thesis. Only genome-wide significant associated (p-value≤$5*10^{-08}$) are included, for variants in high LD(r^2≥0.6, D'≥0.6) only one variant is depicted. KLsum: sum of Kellgren-Lawrence OA severity score, mJSW: minimum joint space width, DDH: developmental dysplasia of the hip, bone area as measured by DXA scan, shape: shape of the joint measured on radiographs.

Phenotyping: Endophenotypes

Another method to increase GWAS power and interpretation is the use of endophenotypes, phenotypes more closely related to the underlying genetics and/or pathology than to the end-stage disease itself. In **Chapter 2.1** cartilage thickness in the hip joint, measured by minimum joint space with (mJSW), was used as an endophenotype for osteoarthritis. Using mJSW over the dichotomous definition of hip osteoarthritis has as main advantage that mJSW focuses on a single part of the joint, the joint space width, as a measure of cartilage thickness. The mJSW can be measured in all individuals, whether or not they have osteoarthritis creating a quantitative measure with greater power than that of the dichotomous trait[17].

Additionally, as GWAS interpretation is dependent on phenotype definition[13, 14, 19], endophenotypes, by their focus on a single underlying characteristic or tissue linked to the disease, helps with interpretation of the observed association with the minimal phenotypes. Also, for known disease variants, endophenotypes can help to elucidate their biological role: in which pathway or pathology these variants may play a role. However, this also means that possibly not all SNVs identified with an endophenotype are automatically associated with end-stage clinical disease. Still, endophenotypes are a very successful method to increase GWAS power, to identify underlying mecha-

nisms, pathways and pathology of osteoarthritis[20-23]. Another strategy to decrease phenotype heterogeneity, and thus increase power, is the use of stratified osteoarthritis phenotypes. For example, grouping individuals based on specific osteoarthritis symptoms (pain vs. no pain, the amount of pain), based on pathological characteristics (osteophytes and joint space narrowing) or joint function (range of motion, disability) or using osteoarthritis endophenotypes: cartilage thickness or the shape of the joint or bone. An overview of all osteoarthritis associated stratified and endophenotype SNVs is given in **Figure 2**, and variants mentioned in this thesis are highlighted.

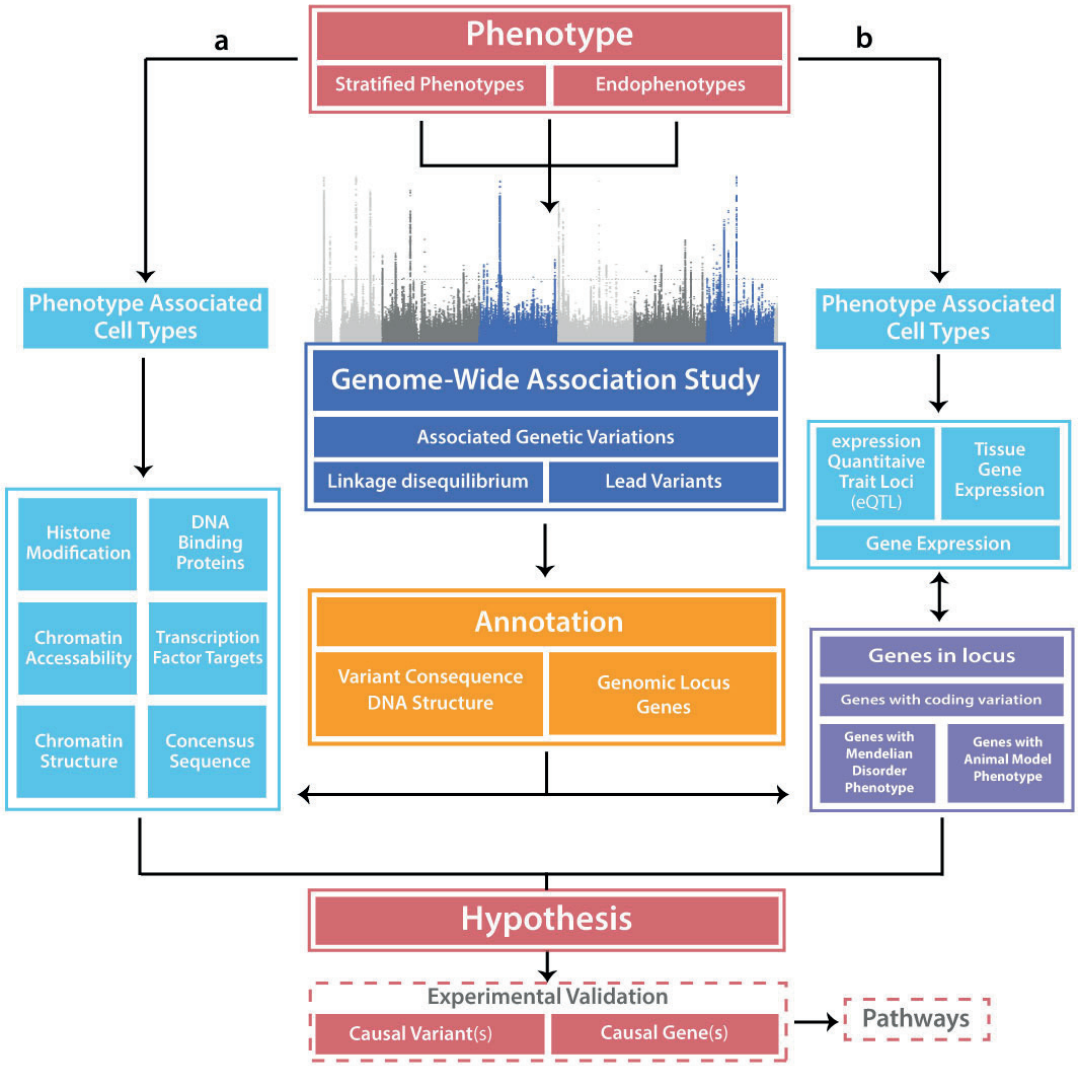
From (osteoarthritis) GWAS to (osteoarthritis) Gene

Finding novel osteoarthritis, or osteoarthritis stratified/endophenotype associated loci, is not the sole purpose of a GWAS. The ultimate goal of performing osteoarthritis GWAS is: to identify associated genetic variation to (i) elucidate genes and biological pathways involved in osteoarthritis pathology. Clearly indicating further steps after GWAS analysis; to identify genes and pathways associated with osteoarthritis, arguably more important that just identifying osteoarthrosis associated genetic variation. However, due to the fundamental design of a GWAS, only an associated genomic locus can be identified directly, never directly a disease associated gene or pathway, without additional evidence. This path from GWAS associated SNV to disease gene is, unfortunately, neither simple nor straightforward. It is a time and effort consuming endeavour, requiring extensive post-GWAS analysis, interdisciplinary knowledge and collaborations. In **Figure 3** an overview of the possible steps that need to be taken to get from osteoarthritis GWAS to novel disease insight, is presented. All, or several, of these steps were performed for each osteoarthritis GWAS in this thesis (**Chapters 2- 4**).

Elucidating GWAS: Causal variant(s)

In order to find an answer to the main GWAS question: how does the GWAS associated variant effects the target gene and how does this contribute to disease risk? First three major difficulties in GWAS interpretation need to be addressed: *(i)*what is/are the causal variants(s), *(ii)* what is the target gene (causal gene) and *(iii)* in which cell type or cell state does the causal variant exert its effect on the causal gene?

First, due to the design of GWAS studies, not a single SNV is associated, but a whole linkage block of SNVs in high LD ($r^2 \geq 0.8$), which can span a region over 100kb containing hundreds of linked SNVs. Any one or combination of these may exert the disease causing risk[24]. Although, it is not easy to identify a causal variant (or always necessary), identification of the causal variant can lead to identification of the causal



▲Figure 3: The possible roads for osteoarthritis GWAS translation. first a phenotype needs to be defined, this could also include an endophenotype or a stratified phenotype. GWAS, finds an associated lead variant and it's linked linkage disequilibrium block. Annotation provides the link between GWAS results and biological interpretation: translational research a) this side presents a schematic overview for the possible steps taken to identify possible causal variants and their effects. b) this side presents the possible steps taken to identify possible causal genes. For both the identification of a causal variant and gene all data needs to be examined in the context of the phenotype associated target cell types. All steps taken provide a line of evidence for or against possible causal variant(s) and gene(s). They provide a basis for further experimental validation to prove causality. Identification of causal variants, may also lead to the identification of causal genes, and identification of causal genes can be used for the identification of pathways involved.

gene. If the causal SNV is located in a coding region of a gene and affects protein translation, folding or function it is likely that that gene would also be the causal gene. However, many of these variants are located in the non-coding regions of the genome. It is vastly more difficult to predict the possible biological consequences of these non-coding variants than it is for coding variants, where much more is known about the consequences of sequence variation[25]. Yet, even non-coding causal SNVs can help identification of the causal gene.

Elucidating GWAS: Causal Gene(s)

GWAS associated SNVs are more likely to reside in gene regulatory regions, and thus are thought to exert their disease risk via affecting gene expression[26]. However, identifying which gene or genes might be targeted is not straightforward. It is a critical mistake to assume that noncoding variations, located intronic in a gene sequence will exert their function on that same gene, since only 14% of SNVs in non-coding regions target nearest genes[27]. Causal SNVs can be located as far as 2MB away from their target gene[28], due to the 3D structure of DNA, which connects DNA elements that are apparently far away from each other. Even larger distances between regulatory region and gene are possible as even larger topologically associated domains (TADs), have been observed [29]. These are large (megabase scale) genomic regions of interconnectivity, in which enhancers usually contact to genes also located within the TAD, but not to genes outside of the TAD. Fortunately, public databases exist of the 3D chromatin structure of multiple cell types and tissues to help identify such gene regulatory region - gene DNA loops. In **chapter 3.1** and **3.2** evidence for such long-range interactions of ~200-700 kb between regulatory regions and target gene has been presented: rs10948155 with *RUNX2 (~700 Kb distance)* and rs10916199 with *WNT9A (~200 Kb distance)*.

Elucidating GWAS: Causal Cell Type, Fate or State

Finally, all annotation and functional genomic data should be examined in the context of cell types or cell states associated with the GWAS phenotype. Here, endophenotypes or stratified phenotypes can directly indicate the target cell type. For example, for mJSW, cartilage cells (chondrocytes) are the target cell type in which to examine the GWAS findings (**Chapter 2.1**). In addition,, the GWAS associated SNVs, can indicate which cell types to examine. GWAS SNVs tend to be particularly enriched in enhancers that are active in the target cell types [26, 30]. By examining this enrichment for the osteoarthritis associated SNVs in the GO consortium, we identified osteoarthritis target tissues (**Chapter 4**). These tissues included the known target tissues (cartilage and bone), but also indicated possible roles for muscle, endocrine and neurological tissues. Which in turn can be used to define novel osteoarthritis endophenotypes or stratified phenotypes, based

on properties or pathological features of these tissues. Although interpretation of GWAS findings is time consuming, laborious, and requires additional experimental validation, it is essential if we want to use GWAS to learn more about osteoarthritis aetiology and eventually identify novel therapeutic targets.

From osteoarthritis GWAS to Clinic

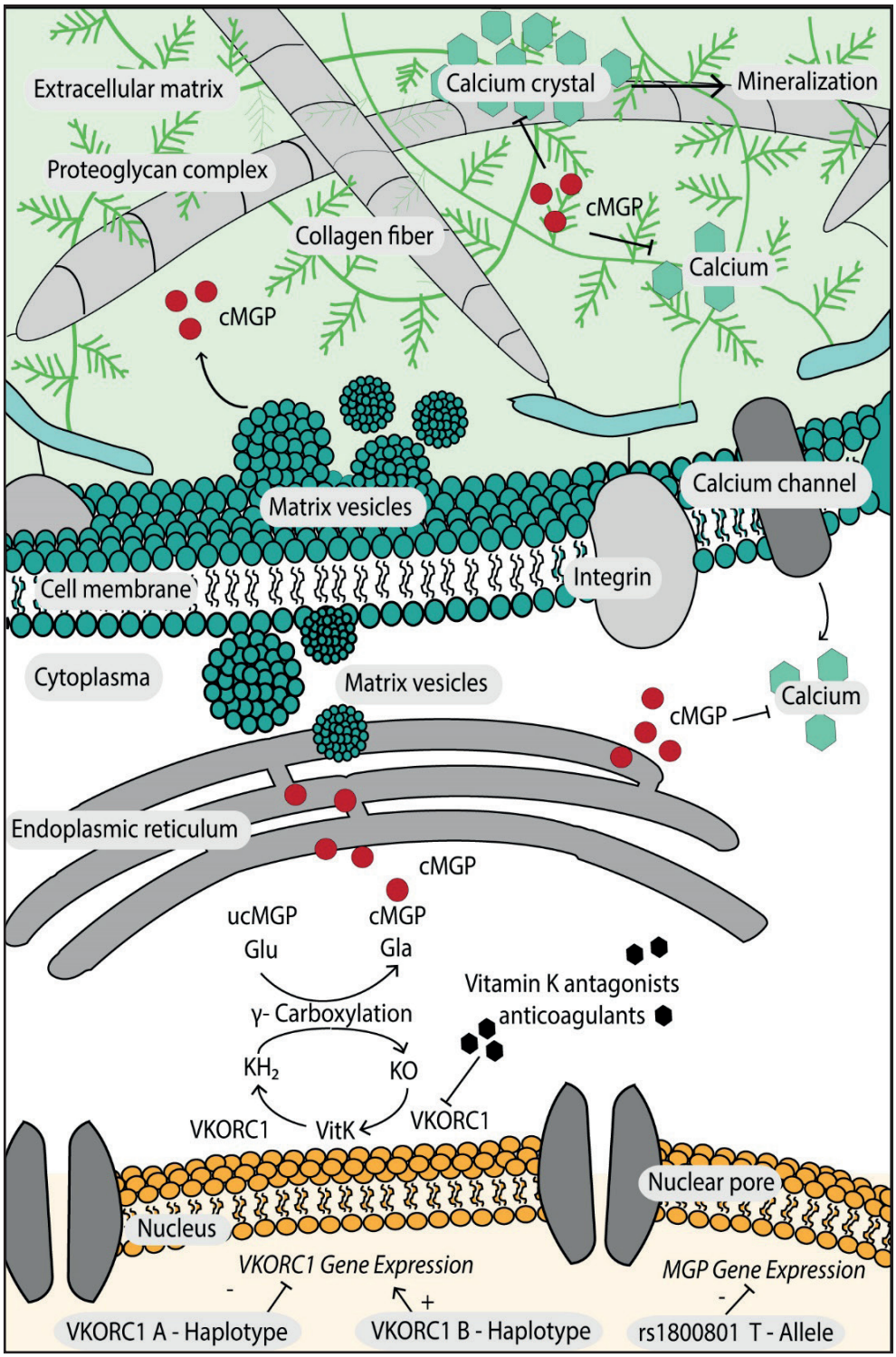
For drug development, treatments based on human genetic evidence of disease association are twice as likely to become approved[31]. When for a GWAS association the causal gene is clearly defined, the approval for drug development increases to four-fold[32]. Therefore, the steps given in the previous paragraph are absolutely essential if we wish to advance our understanding of osteoarthritis pathology and hope to translate this into knowledge or treatments beneficial for patients. **Chapter 3**, represents the path from GWAS finding to identification of a causal gene and pathways, to clinical implication and possible therapeutic intervention.

MGP, Vitamin K, anticoagulants and Osteoarthritis

Using a summarized quantitative osteoarthritis phenotype, SNVs in the *Matrix Gla-Protein*(*MGP*) gene were genome-wide associated with hand KLsum (**Chapter 3.1**). Translational research indicated *MGP* as the causal gene, where the GWAS associated SNVs reduce *MGP* expression in cartilage and bone. *MGP*, is an essential inhibitor of calcification in soft tissues such as the cartilage, it binds to Ca^{2+} preventing the formation of calcium crystals (basic calcium phosphate crystals)[33] (**Figure 4**). MGP is dependent on γ -carboxylation by vitamin K for its function, validating low vitamin K level as risk factors of osteoarthritis[34-36]. Linking a modifiable risk factor, vitamin K levels, with a genetic risk factor for osteoarthritis. In **Chapter 3.2** , the interaction between vitamin K and *MGP* risk variants was investigated. Vitamin K antagonists anticoagulants (VKAs) inhibit the functioning of the VKORC1 protein in the vitamin K cycle, thereby inhibition the γ –carboxylation and thus functioning of MGP (**Figure 4**). There was a two-fold increased risk for osteoarthritis incidence and progression in VKA users, compared to non-users. When the underlying genetics were taken into account, *MGP* risk allele carrier and *VKORC1 B*-haplotype carrier, VKA users had a three-fold increased risk of osteoarthritis progression (**Chapter 3.2**).

MGP, VKORC1 genetics and Osteoarthritis

In the populations of European ancestry, 48% are carriers of the *MGP* risk allele and the *VKORC1 B*-haplotype, exposing half of the VKA users to very high risk of osteoarthritis incidence and progression. Additionally, VKAs are most commonly prescribed to the


Extracellular matrix
Calcium crystal
Mineralization
Proteoglycan complex
cMGP
Collagen fiber
Calcium
cMGP
Calcium channel
Matrix vesicles
Integrin
Cell membrane
Cytoplasma
Matrix vesicles
cMGP
Calcium
Endoplasmic reticulum
cMGP
ucMGP
Glu
cMGP
Gla
Vitamin K antagonists
anticoagulants
γ- Carboxylation
KH2
KO
VKORC1
VitK
VKORC1
Nuclear pore
Nucleus
VKORC1 Gene Expression
MGP Gene Expression
-
+
-
VKORC1 A - Haplotype
VKORC1 B - Haplotype
rs1800801 T - Allele

◀**Figure 4: The biological role of MGP and Vitamin K.** A graphical summary of the functions of MGP, vitamin K in calcification, the effect underlying genetic variation and Vitamin K antagonist anticoagulant (VKA) use. Inactive MGP or uncarboxylized MGP, is γ –carboxylized by vitamin K into active of carboxylized MGP (cMGP), which is transported to the extracellular matrix to bin Ca^{2+}, to prevent the formation of calcium crystals. Variants in *MGP* result is less expression of MGP and are associated with higher risk of osteoarthritis. VKORC1 B haplotype causes increased expression of VKORC1, compared to VKORC1 A-haplotype. VKORC1 B-haplotype associated with hight maintenance dosages of VKAs, conversely VKOC1 A-haplotype is associated with decreases VKA dosages. VKAs inhibit VKORC1 function, and thereby also γ –carboxylation of MGP, increasing osteoarthritis risk.

elderly[37], a group already at higher risk of osteoarthritis, and VKA users are advised to limited their dietary vitamin K intake, resulting in additional risk for low vitamin K status and therefore, possibly osteoarthritis. Interestingly, the *MGP* risk variant poses a large risk for osteoarthritis with environmental interaction (VKA use), while in the GWAS, the effect size was small (**Chapter 4**).

This Indicates that GWAS effect size does not predict the clinical or therapeutic impact of the associated SNV[2]. It also underlines the necessity to understand the biological mechanism behind the genetic association, in order to fully recognize its clinical impact. In addition, although first only found to be associated with hand and finger osteoarthritis and later nominal significant with knee osteoarthritis (**Chapter 4**), the *MGP* risk variants pose risk for hip and knee osteoarthritis in VKA users. Indicating non joint specific role for *MGP* and vitamin K in osteoarthritis, further increasing the therapeutic potential importance of the vitamin K-cycle.

Osteoarthritis clinical implications of MGP

These genome-wide hand osteoarthritis associated variants in *MGP* have illustrated the a possible novel pathway associated with osteoarthritis, i.e., chondrocalcinosis. In addition, they have elucidated underlying fundamental biological mechanisms and confirmed prior epidemiological observations on vitamin K and osteoarthritis[34, 35, 38, 39]. More importantly, these variants and subsequent follow-up characterization of these variants have identified vitamin K as possible target for pharmacotherapy and pharmacogenetic clinical trials. This was the result of only a single osteoarthritis associated locus and only a handful of other osteoarthritis loci have been deeply translational investigated[11, 40-42]. Tus the potential impact of deep translation research on those many remaining variants, for example the 52 novel reported in the GO-consortium, could be significant on osteoarthritis understanding and therapeutic research.

Novel risk factors for osteoarthritis: the Microbiome

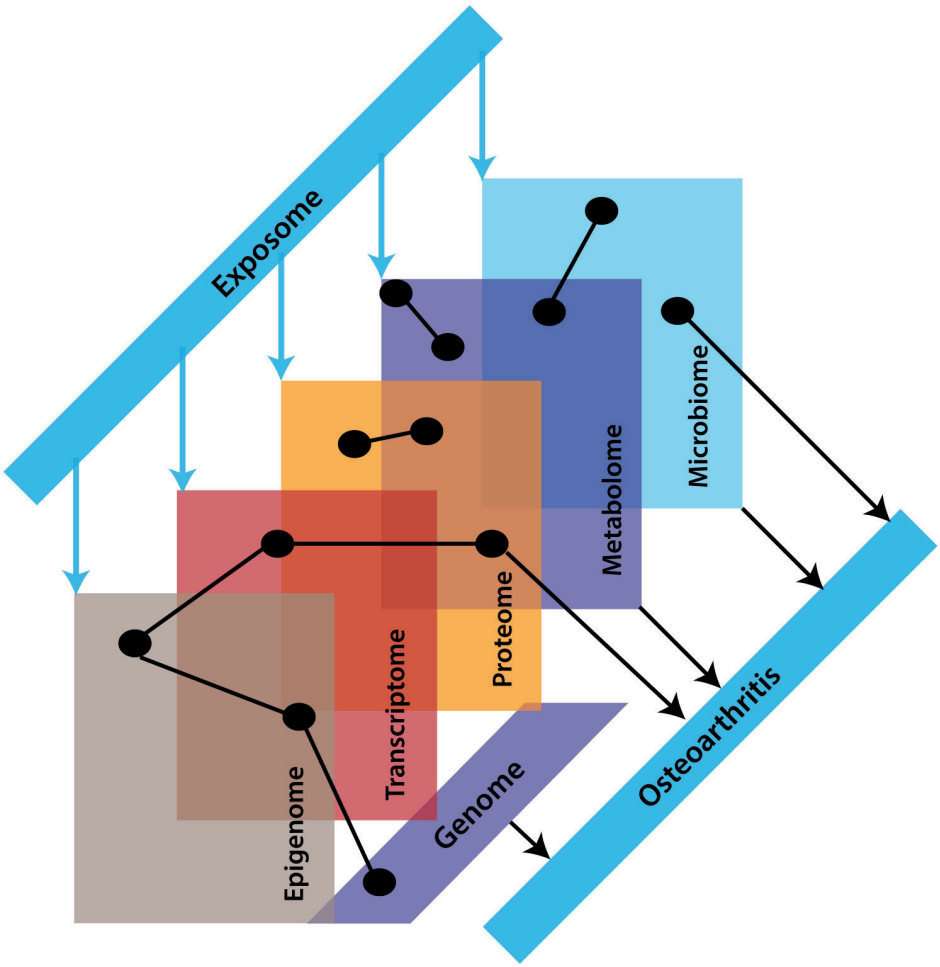
As a complex disease, osteoarthritis risk is determined by a combination of genetic and environmental risk factors. Each of these risk factors, however does not exist alone in a vacuum, all are interconnected, forming complex layers of interactions (**Figure 5**). A "novel" layer in this complex system is the microbiome. C**hapter 5** investigated the gastrointestinal microbiome composition as a possible risk factor for osteoarthritis. Changes in composition and function of the gastrointestinal microbiome are affected by a large number of host and environmental factors and have been associated with many common diseases and health deficits[43, 44]. Osteoarthritis is no exception to this, and in **Chapter 5.1** increased abundance of *Streptococcus spp.* in the gastrointestinal microbiome was associated with increased inflammation and osteoarthritis related pain and inflammation in the knee joint. If the association is causal, it would be a significant finding for possible therapeutic intervention, as pain is the main disabling symptom in osteoarthritis[45]. Both for symptomatic relief, as pain elevation, as well as long-term disease modification (reduced inflammation damage). Some preliminary studies have shown that antibiotic use could reduce chronic low back pain, postulated to be caused by a low grade inflammation caused by the gastrointestinal microbiome[46, 47]. Other possible methods exist and/or are under development to alter the composition of the gastrointestinal microbiome (faecal transplant and probiotics, diet intervention) and thus may also be investigated as therapeutic strategies for osteoarthritis.

What does the future hold for osteoarthritis?

It has been more than a decade ago that the first osteoarthritis GWAS was performed and since this first effort our knowledge of osteoarthritis genetics has significantly expanded, as has been presented in this thesis. Despite the increase in associated genetic loci, pathways and risk factors, much remains unknown on osteoarthritis pathology. The following recommendations can be implemented to aid the elucidation of the genetic architecture of osteoarthritis and disease pathology: *(i)* expansion of osteoarthritis GWASs, *(ii) Re-defining osteoarthritis* , *(iii) Microbiomics: It's all in the gut* and *(vi) Post GWAS era: move to the clinic* . All four are interconnected, and will be needed to bring preventive or curative treatments closer for one of the world's oldest diseases

Expansion of osteoarthritis GWAS: method, size, population, and phenotype

In the last few years we have seen an effective sevenfold increase of reported osteoarthritis associated loci, predominantly driven by large scale osteoarthritis GWAS. For



▲ **Figure 5: the multiple omics layers and interactions leading to osteoarthritis.** Schematically several omics- layers are represented. The exposome (environment) influences/interacts with all omics, except for the genome. The genome does underly the epigenome/transcriptome and proteome directly, but only indirect the other omics. Black dots and lines represent possible pathological pathways and interactions within and between omics-layers resulting in osteoarthritis pathology.

other complex diseases similar success of ever larger GWAS have been reported, with some reaching sample sizes of over 1 million[48, 49]. With the formation and ongoing efforts of the Genetics of Osteoarthritis (GO) consortium, such large sample sizes are also becoming possible for osteoarthritis GWAS. Although the first success has been made by the GO consortium, several issues need to be addressed. Much of the heritability (~94-79%) of osteoarthritis remains unexplained, although the low predicted heritability (8%-22%) seen in the GO consortium could be large explained by bias/heterogeneity of the used osteoarthritis phenotypes. In general (**Table 1**) this "miss-

ing heritability" is thought to be partly explained by rare SNVs (minor allele frequency <0.01) as the vast majority of human genetic variation is rare. In addition, other types of genetic variations besides SNVs exist, such as structural variation, which may explain a portion of the "missing heritability". As GWAS predominantly evaluates common SNVs (minor allele frequency >0.01), a proportion of the heritability will remain undiscovered. In addition, rare genetic variation is population specific, and the majority of (~80%) GWASs were performed in individuals of European descent[50]. Differences in disease incidence, prevalence and severity between populations can in part be attributed to underlying genetic variation and environmental interaction[51]. It is therefore not surprising that the results of European ancestry based GWAS are not directly transferable to other populations, as the underlying causal variants and effect sizes differ across populations[51]. Thus more globally inclusive data containing more genetic variation are needed, for example by whole genome sequencing globally inclusive populations.

Although whole genome-sequencing data is the golden standard for identifying all possible genetic variations, it is considerably more expensive and computational intensive than genotyping arrays. The newest generation arrays, such as the Global Screening Array(GSA) or the PMDA (Thermofisher), are not only cheap, but they also increasingly cover more of the genome for different ethnic populations, including clinically valuable rare variants[52]. Moreover, new imputation panels (TOPMED) and imputation methods are improving the calling of rare and population specific variants[53]. However, there still is a lack of true population specific reference panels i.e., for the Rotterdam Study this means a "Dutch ancestry" whole genome sequencing reference panel, rather than a whole genome sequencing reference panel of "European Ancestry US Citizens". Considering that many existing cohorts already have genotyping data available and could improve quality simply by re-imputing data, genotyping arrays remain highly cost effective for GWASs. Regardless of the method, for successful GWAS very large sample sizes are needed to study rare genetic variation. The costs saved by using array-based genotyping instead of whole-genome-sequencing could thus be better used to increase the sample size.

However, is larger always better? Large scale GWASs have undoubtedly been successful and sample size is usually the key limiting factor in GWAS discovery. However, small scale GWAS using "smart phenotyping" such as stratified or endophenotype strategies have also been very successful, and have as extra benefit that they reduce heterogeneity and improve GWAS interpretation. Although, such smaller scale studies would be even more successful with larger sample sizes. The identification of a single robust GWAS finding could already lead to significant therapeutic impact. Thus also in GWAS: quality (robust and well validated association) rules over quantity (hundreds of

non-robust signals), especially for therapeutic development. Therefore, investment in deep and accurate phenotyping over ever increasing sample sizes, may be more therapeutically beneficial in the long run.

Thus, if we want to fully understand the genetic architecture and pathology of osteoarthritis, future studies will need to invest in large scale genotyping *and* deep phenotyping in globally diverse populations. Together with creating globally inclusive whole genome sequencing reference panels for increased imputation quality. Such investments will be needed if we want to develop novel treatment strategies against osteoarthritis from which the whole global population could benefit. The first steps toward this goal have already been taken within the GO consortium. The consortium is ever increasing in size and including more and more diverse global populations. Fortunately, it also include cohorts with deep phenotyping information. Allowing for the development and use of phenotypes more accurately representing the deep complexity of osteoarthritis.

Re-definition of osteoarthritis

As stated above and throughout this thesis, the phenotyping of osteoarthritis is critical for future osteoarthritis GWAS and research success. Especially if we want to move *"from bench to bedside"*. How we define osteoarthritis as a disease strongly determines the interpretation and power of a GWAS. Current and past osteoarthritis GWAS have predominantly used "minimal phenotype" definitions, i.e. hospital diagnosed and self-reported osteoarthritis, possibly only identifying the "low-hanging fruit" of osteoarthritis GWAS[13, 17]. Exploring a wider range of biological/clinical relevant disease phenotypes, i.e., stratified and endophenotypes, will most likely lead to novel discoveries and/or a more complete understanding of osteoarthritis genetics. Future osteoarthritis definition and phenotypes need to embrace the heterogeneity in osteoarthritis. Although, this heterogeneity is well recognized[54], osteoarthritis is still described as a singular disease. Osteoarthritis should be referred to as a syndrome; multiple causes, pathways, tissues and pathologies all leading to the same clinical outcome: osteoarthritis[54]. The re-definition of osteoarthritis as syndrome can also be incorporated in GWAS phenotype definitions: stratifying osteoarthritis symptoms (pain, stiffness, swelling, disability), main affected tissue (cartilage, bone, muscle etc.), or by underlying cause (traumatic, obesity) or predominant molecular pathway (vitamin K pathway).

In addition, future phenotypes should also acknowledge the impact of other large osteoarthritis risk factors. The largest risk factor of osteoarthritis, is age, if the controls consist of young individuals (<50 years) they may still develop osteoarthritis at a later age. For GWAS research this might mean some genetic loci may be missed. However, age could also be incorporated into the osteoarthritis phenotype in GWAS for example by

using age of osteoarthritis onset as phenotype rather than osteoarthritis status. Other examples of (environmental) risk factor incorporation into osteoarthritis phenotypes might be to examine disease progression over time, changes of osteoarthritis risk with changes in environmental risk factors, such as diet (gut microbiome) and drug use, over time.

Acknowledging osteoarthritis as a syndrome will emphasize phenotype definition as the key importance in osteoarthritis research. Accurate identification of osteoarthritis phenotypes, will increase GWAS power by reducing heterogeneity. Follow-up research can then link underlying mechanisms to the specific phenotypes, possibly leading to the identification of specific clinical subsets of osteoarthritis patients. Possibly improving patient care by providing the possibility of future targeted/personalized therapeutic interventions based on the clinical subset of osteoarthritis. Thus by re-defining osteoarthritis as a syndrome, we can improve osteoarthritis research, improve patient care by possibly future development of targeted treatments.

Microbiomics: It's all in the gut

The human microbiome, specifically the human gastrointestinal microbiome, has been shown to play an important role in health and disease[43, 44], possibly including osteoarthritis (**Chapter 5**). Although, the microbiome is an easy and attractive therapeutic target for osteoarthritis, no conclusive evidence of causality is established. Partially, this can be attributed to the novelty of large scale microbiomics data: the limited availability of data, lack of standardization in data processing, methodology and analysis. Moreover, much is unknown on the exact molecular mechanisms or pathways by which the human microbiome functions in health and disease. This also means that for most microbiome and disease associations causality is not known. However, for future research it is absolutely essential that the causality and the directionality (does *Streptococcus spp.* cause pain or does pain lead to *Streptococcus spp.* colonialization) is established.

There are several possible research methods by which causality of *Streptococcus spp.* on osteoarthritis could be investigated. These would include longitudinal analysis of gastrointestinal composition and osteoarthritis related pain, in order to examine if an increase in *Streptococcus spp.* is followed by an increase in joint inflammation and pain. The detection of *Streptococcus* (or bacterial) metabolites and excreted membrane vesicles in the blood and joint tissue (e.g. synovium) and if these can trigger joint inflammation and/or pain, would also be evidence of a causal relationship. In addition, to establishing causality, identifying the exact species or strain of *Streptococcus* bacteria is of therapeutic importance. If the association is causal and decrease of *streptococcus spp.*

could prevent or decrease joint pain and inflammation, possible therapeutic strategies include antibiotics, diet intervention, faecal transplants or even microbial metabolites [55, 56]. However, these therapies also affect the rest of the gastrointestinal microbiome. If the exact *Streptococcus* species or strain is known specific antibiotics, probiotics or introduction of competitor species could be used, without affecting the entire microbiome composition. Thus, still quite some research is needed before the microbiome is truly part of the osteoarthritis risk factors and possible therapeutic targets.

Post GWAS era: move to the clinic

The increase in osteoarthritis genetic loci over the last decade has not only increased the understanding of osteoarthritis genetics, but also understanding of osteoarthritis pathology. Especially, translational research into GWAS results has demonstrated novel genes, pathways and mechanisms to be involved in osteoarthritis pathology. Some are now considered or in the development process for novel treatment strategies: *MGP(osteoarthritis in any joint, Vitamin K)(***Chapter 3***)[21, 36]*, *TGFB1 (knee osteoarthritis, INVOSSA)[57]*, *CSTK (extracellular matrix turnover/inflammation inhibitor, CSTK inhibitor) [58]* and *FGF18(cartilage regeneration for knee osteoarthritis, Sprifermin)[59]*. However, a true move from GWAS to clinical treatment has not (yet) been made for osteoarthritis, for this to happen considerable more functional follow-up and translational studies are needed. Particularly, bioinformatic investigation of omics, multi-omics and integrative omics is needed (**Figure 5**). Although for osteoarthritis, some omics resources exist (cartilage and bone gene expression, methylation, histone markers , ATAC-seq), the data is scarce and is measured in few individuals or only in specific states (healthy or diseased). What is needed, is a large scale integrative multi-omics data resource: measuring multiple osteoarthritis relevant tissues, tissue states and omics-layers in one large group of individuals. Several biobanks and groups are working towards such a goal (UK-biobank, ROADMAP, ENCODE), albeit not specifically for osteoarthritis or osteoarthritis relevant tissue. Imagine the possibilities of such an osteoarthritis specific resource, considering the immense value and insight the current limited osteoarthritis specific omics data have provided so far. Such a data resource will be invaluable for the identification of novel osteoarthritis associated genes, pathways and therapeutic development.

Osteoarthritis associated SNVs could potentially also directly be used in a clinical setting. These SNVs could be used to create an individual level genetic risk score or polygenic risk score (PRS) for osteoarthritis. These scores reflect an individual's genetic risk for developing a certain disease[60]. For osteoarthritis, currently 148 genomic loci

have been reported. Although, these only explain 6-21% of the genetic heritably for osteoarthritis, they may explain sufficient genetic risk to be successfully used in a PRS, for example hip osteoarthritis where 21% of the genetic heritability is explained. We might be able to identify individuals at high risk of developing hip osteoarthritis, preferably at an early age before osteoarthritis onset, and take preventative measures[60]. These would include the prevention of environmental osteoarthritis risk factors, such as obesity. No treatment strategies exist (for now) for osteoarthritis, thus such preventative measure could be extremely valuable, especially in individuals with a high genetic risk of osteoarthritis. By lowering environmental risk factors such individuals could either prevent osteoarthritis or extend the age of onset of osteoarthritis.

In addition, PRS could also be used in osteoarthritis patients to identify which underlying pathological pathway are driving their osteoarthritis. Using the translational knowledge of , i.e., of which genes and pathways, of the SNVs included in the PRS, the pathological pathways involved can be identified. Allowing for a "personalized medicine" approach for identifying subgroups of osteoarthritis patients and adjusting therapeutic strategies accordingly Here, it becomes evident why GWAS needs to be done on globally diverse populations. If we do not take into consideration differences between populations, the possibility exists that the developed PRS is only effective in one specific population. Possibly of widening existing disparities and excluding minority populations of valuable osteoarthritis clinical care.

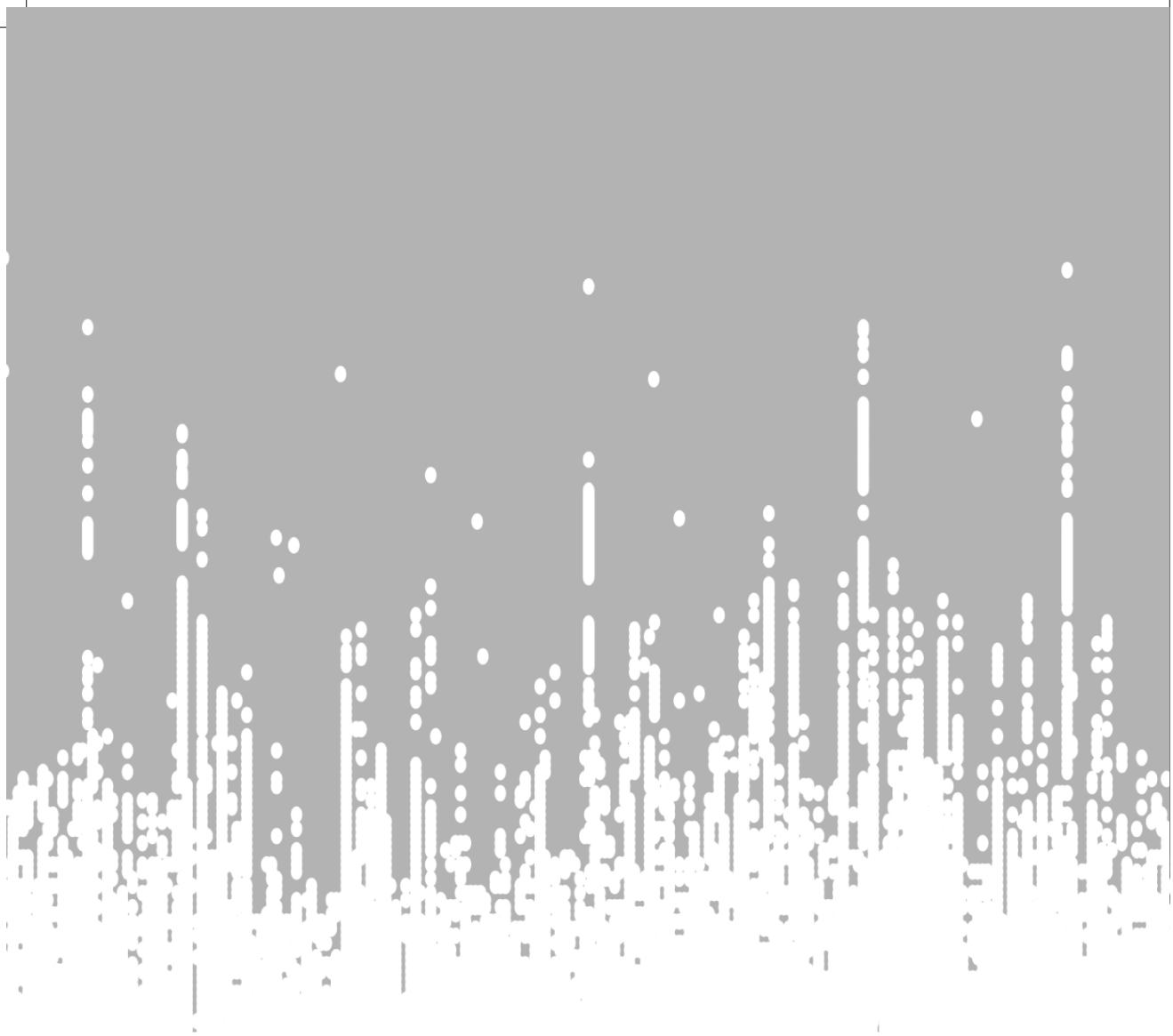
Although seemingly much needs to be done to move from "bench to bedside", the current SNV array prices have reached a record low (~€30,- per sample). Combined with the rapid increase in understanding of osteoarthritis genetics, clinical implications might reach the "bedside" soon. Hopefully the research presented here and suggestions for future research will contribute to help lower the burden of osteoarthritis, which has been plaguing synovial joints for hundreds of millions years.

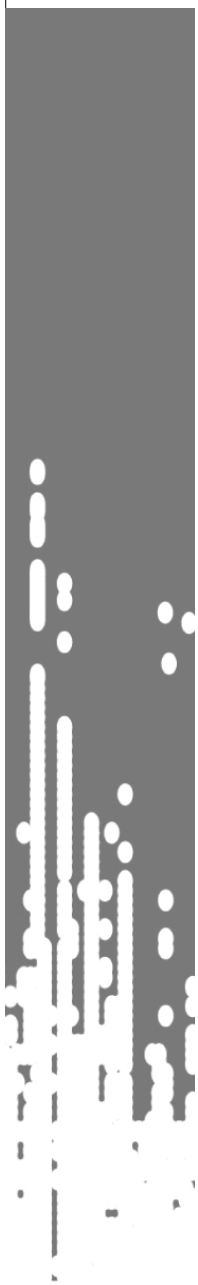
References

1. Zeggini, E., et al., Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. Lancet, 2012. 380(9844): p. 815-23.
2. Visscher, P.M., et al., 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet, 2017. 101(1): p. 5-22.
3. MacGregor, A.J., et al., The genetic contribution to radiographic hip osteoarthritis in women: results of a classic twin study. Arthritis Rheum, 2000. 43(11): p. 2410-6.
4. Magnusson, K., et al., Genetic factors contribute more to hip than knee surgery due to osteoarthritis- a population-based twin registry study of joint arthroplasty. Osteoarthritis Cartilage, 2017. 25(6): p. 878-884.
5. Spector, T.D., et al., Genetic influences on osteoarthritis in women: a twin study. Bmj, 1996. 312(7036): p. 940-3.
6. Magnusson, K., A. Turkiewicz, and M. Englund, Nature vs nurture in knee osteoarthritis- the importance of age, sex and body mass index. Osteoarthritis Cartilage, 2019. 27(4): p. 586-592.
7. Bijkerk, C., et al., Heritabilities of radiologic osteoarthritis in peripheral joints and of disc degeneration of the spine. Arthritis Rheum, 1999. 42(8): p. 1729-35.
8. Ishimori, M.L., et al., Heritability patterns in hand osteoarthritis: the role of osteophytes. Arthritis Res Ther, 2010. 12(5): p. R180.
9. Zengini, E., et al., Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. Nat Genet, 2018. 50(4): p. 549-558.
10. Tachmazidou, I., et al., Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. Nat Genet, 2019. 51(2): p. 230-236.
11. Styrkarsdottir, U., et al., Meta-analysis of Icelandic and UK data sets identifies missense variants in SMO, IL11, COL11A1 and 13 more new loci associated with osteoarthritis, in Nat Genet. 2018: United States. p. 1681-1687.
12. Wray, N.R., et al., Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. Cell, 2018. 173(7): p. 1573-1580.
13. Brzustowicz, L.M. and A.S. Bassett, Phenotype Matters: The Case for Careful Characterization of Relevant Traits. Am J Psychiatry, 2008. 165(9): p. 1096-8.
14. Cai, N., et al., Minimal phenotyping yields genome-wide association signals of low specificity for major depression. Nat Genet, 2020. 52(4): p. 437-447.
15. Hunter, D.J. and S. Bierma-Zeinstra, Osteoarthritis. Lancet, 2019. 393(10182): p. 1745-1759.
16. Kerkhof, H.J., et al., Recommendations for standardization and phenotype definitions in genetic studies of osteoarthritis: the TREAT-OA consortium. Osteoarthritis Cartilage, 2011. 19(3): p. 254-64.
17. Manchia, M., et al., The Impact of Phenotypic and Genetic Heterogeneity on Results of Genome Wide Association Studies of Complex Diseases, in PLoS One. 2013.
18. Bierma-Zeinstra, S.M. and M. van Middelkoop, Osteoarthritis: In search of phenotypes. Nat Rev Rheumatol, 2017. 13(12): p. 705-706.
19. MacRae, C.A. and R.S. Vasan, Next Generation GWAS: Time to Focus on Phenotype? Circ Cardiovasc Genet, 2011. 4(4): p. 334-6.
20. Styrkarsdottir, U., et al., GWAS of bone size yields twelve loci that also affect height, BMD, osteoarthritis or fractures. Nat Commun, 2019. 10(1): p. 2054.
21. den Hollander, W., et al., Genome-wide association and functional studies identify a role for matrix Gla protein in osteoarthritis of the hand. Ann Rheum Dis, 2017. 76(12): p. 2046-2053.
22. Castano-Betancourt, M.C., et al., Novel Genetic Variants for Cartilage Thickness and Hip Osteoarthritis. PLoS Genet, 2016. 12(10): p. e1006260.

23. Hatzikotoulas, K., et al., Genome-wide association study of developmental dysplasia of the hip identifies an association with GDF5. Commun Biol, 2018. 1: p. 56.
24. Visscher, P.M., et al., 10 Years of GWAS Discovery: Biology, Function, and Translation, in Am J Hum Genet. 2017. p. 5-22.
25. Gabriel, S.B., et al., The structure of haplotype blocks in the human genome. Science, 2002. 296(5576): p. 2225-9.
26. Maurano, M.T., et al., Systematic localization of common disease-associated variation in regulatory DNA. Science, 2012. 337(6099): p. 1190-5.
27. Mumbach, M.R., et al., Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nat Genet, 2017. 49(11): p. 1602-12.
28. Brodie, A., J.R. Azaria, and Y. Ofran, How far from the SNP may the causative genes be? 2016.
29. Ibrahim, D.M. and S. Mundlos, Three-dimensional chromatin in disease: What holds us together and what drives us apart? Curr Opin Cell Biol, 2020. 64: p. 1-9.
30. Kundaje, A., et al., Integrative analysis of 111 reference human epigenomes. Nature, 2015. 518(7539): p. 317-30.
31. Nelson, M.R., et al., The support of human genetic evidence for approved drug indications. Nat Genet, 2015. 47(8): p. 856-60.
32. King, E.A., J.W. Davis, and J.F. Degner, Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. PLoS Genet, 2019. 15(12): p. e1008489.
33. Schurgers, L.J., E.C. Cranenburg, and C. Vermeer, Matrix Gla-protein: the calcification inhibitor in need of vitamin K. Thromb Haemost, 2008. 100(4): p. 593-603.
34. Misra, D., et al., Vitamin K deficiency is associated with incident knee osteoarthritis. Am J Med, 2013. 126(3): p. 243-8.
35. Neogi, T., et al., Low vitamin K status is associated with osteoarthritis in the hand and knee. Arthritis Rheum, 2006. 54(4): p. 1255-61.
36. Neogi, T., et al., Vitamin K in hand osteoarthritis: results from a randomised clinical trial. Ann Rheum Dis, 2008. 67(11): p. 1570-3.
37. Zhu, J., et al., Trends and Variation in Oral Anticoagulant Choice in Patients with Atrial Fibrillation, 2010-2017. Pharmacotherapy, 2018. 38(9): p. 907-20.
38. Shea, M.K., et al., Association of Vitamin K Status Combined With Vitamin D Status and Lower-Extremity Function: A Prospective Analysis of Two Knee Osteoarthritis Cohorts. Arthritis Care Res (Hoboken), 2018. 70(8): p. 1150-1159.
39. Shea, M.K., et al., The association between vitamin K status and knee osteoarthritis features in older adults: the Health, Aging and Body Composition Study. Osteoarthritis Cartilage, 2015. 23(3): p. 370-8.
40. Styrkarsdottir, U., 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(5): p. 498-502.
41. Meulenbelt, I., et al., Identification of DIO2 as a new susceptibility locus for symptomatic osteoarthritis. Hum Mol Genet, 2008. 17(12): p. 1867-75.
42. Capellini, T.D., et al., Ancient selection for derived alleles at a GDF5 enhancer influencing human growth and osteoarthritis risk. Nat Genet, 2017. 49(8): p. 1202-1210.
43. Falony, G., et al., Population-level analysis of gut microbiome variation. Science, 2016. 352(6285): p. 560-4.
44. Jackson, M.A., et al., Gut microbiota associations with common diseases and prescription medications in a population-based cohort. Nat Commun, 2018. 9(1): p. 2655.
45. Neogi, T., The epidemiology and impact of pain in osteoarthritis. Osteoarthritis Cartilage, 2013. 21(9): p. 1145-53.

46. Albert, H.B., et al., Antibiotic treatment in patients with chronic low back pain and vertebral bone edema (Modic type 1 changes): a double-blind randomized clinical controlled trial of efficacy. Eur Spine J, 2013. 22(4): p. 697-707.
47. Bråten, L.C.H., et al., Efficacy of antibiotic treatment in patients with chronic low back pain and Modic changes (the AIM study): double blind, randomised, placebo controlled, multicentre trial. Bmj, 2019. 367: p. l5654.
48. Jansen, P.R., et al., Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nat Genet, 2019. 51(3): p. 394-403.
49. Lee, J.J., et al., Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet, 2018. 50(8): p. 1112-1121.
50. Peterson, R.E., et al., Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. Cell, 2019. 179(3): p. 589-603.
51. Bien, S.A., et al., The Future of Genomic Studies Must Be Globally Representative: Perspectives from PAGE. Annu Rev Genomics Hum Genet, 2019. 20: p. 181-200.
52. Tozzi, V., et al., Global, pathway and gene coverage of three Illumina arrays with respect to inflammatory and immune-related pathways. Eur J Hum Genet, 2019. 27(11): p. 1716-1723.
53. Brody, J.A., et al., Analysis commons, a team approach to discovery in a big-data environment for genetic epidemiology. Nat Genet, 2017. 49(11): p. 1560-1563.
54. Deveza, L.A. and R.F. Loeser, Is osteoarthritis one disease or a collection of many? Rheumatology (Oxford), 2018. 57(suppl_4): p. iv34-iv42.
55. Thaiss, C.A. and E. Elinav, The remedy within: will the microbiome fulfill its therapeutic promise? J Mol Med (Berl), 2017. 95(10): p. 1021-1027.
56. Descamps, H.C., et al., The path toward using microbial metabolites as therapies. EBioMedicine, 2019. 44: p. 747-754.
57. Cherian, J.J., et al., Preliminary results of a phase II randomized study to determine the efficacy and safety of genetically engineered allogeneic human chondrocytes expressing TGF-β1 in patients with grade 3 chronic degenerative joint disease of the knee. Osteoarthritis Cartilage, 2015. 23(12): p. 2109-2118.
58. Pan, W., et al., Inhibition of Ctsk alleviates periodontitis and comorbid rheumatoid arthritis via downregulation of the TLR9 signalling pathwa





Chapter 7

Summary

Summary

Osteoarthritis is the most common chronic degenerative joint disease worldwide, and has been present throughout history, evidenced by fossilized findings of osteoarthritic joints. Making Osteoarthritis one of the world's oldest diseases, and yet no curative treatments are available. The aim of this thesis is to identify novel genes involved in osteoarthritis, in order to gain greater insight into the pathology of osteoarthritis and bring possible preventive or curative treatments closer for one of the world's oldest disease.

Chapter 1 gave an introduction of osteoarthritis and it's long history. Also, key concepts in genetics, genetic epidemiology research and epigenetics were covered. **Chapter 1.2**, delved deeper into the role that epigenetics play in osteoarthritis research and pathology.

In **Chapter 2** results of two genome-wide association studies using two different types of osteoarthritis phenotypes were presented. Although, these studies had a relatively small sample size (n≤10,000) both studies robustly identify novel osteoarthritis associated loci. Using follow-up translational research, novel candidate genes for osteoarthritis therapeutic development were identified. Demonstrating and confirming the potential for novel insight into the genetic architecture of osteoarthritis by using biologically meaningful phenotypes (stratified or endophenotypes).

Chapter 2.1 showed the use of minimal joint space width (mJSW) as an endophenotype for hip cartilage thickness. Using this osteoarthritis endophenotype four novel genetic loci associated with cartilage thickness and hip osteoarthritis were identified. In addition, two known osteoarthritis associated loci were replicated, (rs2864419) DOT1L and (rs109481947) *SUPT3H-RUNX2*. Systematic prioritization using diverse lines of evidence pointed to likely causal genes for the identified loci: *RUNX2, TGFα, PIK3R1, SLBP/FGFR3, DOT1L/GADD45B* and *TREH/DDX6*. Of these TGFα, PIK3R1 and FGFR3 were differentially expressed between osteoarthritis lesioned and non-lesioned cartilage.

In **Chapter 2.2** three hand osteoarthritis phenotypes were created based on the pattern and severity of radiographic osteoarthritis in the hand joints: thumb osteoarthritis severity (klsum), finger osteoarthritis severity (KLsum) and hand osteoarthritis severity (KLsum). Two known hand osteoarthritis associated loci were replicated and

two novel thumb osteoarthritis loci were identified. One of the thumb osteoarthritis severity osteoarthritis loci (rs10916199) was also associated with clinical thumb osteoarthritis. Using multiple approaches that leverage different levels of information and functional data, the underlying biological mechanism and candidate gene for this locus (rs10916199) was investigated. *WNT9A* was identified as possible novel gene involved in osteoarthritis pathogenesis, with rs1158850 as potential causal variant in this locus. In addition, several previously identified genetic loci for osteoarthritis were shown to also confer risk for osteoarthritis across multiple joints: *TGFα, RUNX2, COL27A1. ASTN2. IL11* and the *GDF5*-loci.

Chapter 3 continued the use of this hand osteoarthritis severity phenotype to identify the second hand osteoarthritis associated loci. Again, using follow-up translational research a novel osteoarthritis associated gene was discovered. However, in the second part of this chapter, the implications of this novel osteoarthritis gene association are put to the test and revealed some serious clinical implications. Demonstrating the importance of in depth follow-up research and value of genetic studies for osteoarthritis research.

In **Chapter 3.1** an osteoarthritis phenotype, which quantified the severity of radiographic osteoarthritis, was used to identify the second genetic loci associated with hand osteoarthritis. A coding variant in the *MGP* gene was associated with increased risk of hand osteoarthritis. This variant showed allele specific gene expression in cartilage tissue, the risk variant significantly reduced *MGP* gene expression. MGP is an essential inhibitor of cartilage calcification, thus reduced *MGP* expression could lead to increase cartilage calcification and higher risk of osteoarthritis. In addition, MGP can only inhibit cartilage calcification if γ-carbolized by vitamin K. Demonstrating the link between low vitamin K status and high risk of osteoarthritis.

Using this new knowledge, **Chapter 3.2** investigated the possible role of vitamin K antagonist anticoagulants in osteoarthritis risk and progression. Vitamin K mediated γ-carboxylation is also essential for several blood coagulation factors, thus most anticoagulant medication inhibits this process. The vitamin K antagonist anticoagulants inhibit the function of *VKORC1*, an essential protein in the Vitamin K mediated γ-carboxylation pathway. Individuals taking the vitamin K antagonist anticoagulant acenocoumarol, were found to have a twofold greater risk of incidence and progression of knee and/or hip osteoarthritis then those not using acenocoumarol. Next, individuals whom were carrier of the MGP osteoarthritis risk allele and the *VKORC1* B-haplotype, associated with a higher dose of acenocoumarol, had a more than

threefold increased risk of osteoarthritis when using acenocoumarol, compared to non-users. In a popu-lation with European ancestry ~50% of the population is carrier of both these alleles, thus potentially exposing half of the acenocoumarol users in that population to a signif-icant increased risk of osteoarthritis. Thus, if the clinical indication allows, individuals in need of anticoagulation might be prescribed non-vitamin K inhibiting anticoagulants. In addition, the results of this chapter, validated the importance of MGP and vitamin K in osteoarthritis pathology, opening up the possibility to investigate vitamin K as thera-peutic target for osteoarthritis.

In genome-wide association studies, the most important factors for success are phenotype definition and sample size. Chapters 2 to 3, examined the effect of phenotype definition, in **Chapter 4** sample size is examined. This was done by collaborating in the to date largest osteoarthritis genetics consortium: Genetics of Osteoarthritis (GO) consortium. The GO-consortium consists of over 20 different population studies from across the world, although predominantly of European ancestry, and contained in total 177,517 osteoarthritis cases and 649,173 controls.

Chapter 4.1 presents the first results of the 11 osteoarthritis genome-wide association studies performed in the GO-consortium. In total 100 genetic loci associated with osteoarthritis were discovered, of which 52 were novel and 48 were known osteoarthritis loci. Among these loci are, rare variants with large effects, variants with opposite direction of effects between phenotypes, characterise differences between weight-bearing and non-weight bearing joints and loci associated with all osteoarthritis phenotypes. Novel possible osteoarthritis target tissues could be identified as well as, genetic correlation of osteoarthritis with neuronal pathways and pain. Functional genomics data from primary patient tissues identified possible causal gene(s) for each locus and revealed possible shared underlying molecular mechanisms for comorbid cardiovascular diseases and osteoarthritis. All providing a new wealth of insight in to osteoarthritis pathology and highlight attractive targets for drug discovery and development.

Osteoarthritis is a complex disease, and thus environmental factors also play a major role in the pathology. A novel environmental factor associated with disease is the human microbiome. The gut microbiota has been shown to play diverse roles in human health and disease although the underlying mechanisms have not yet been fully elucidated. Thus **Chapter 5** investigated the possible role that the trillions of bacteria living inside

our digestive tract may play in osteoarthritis. Recent studies had shown an important role for the microbiome in obesity, and demonstrated that the microbiome changes with age. As, obesity and age are most well characterized risk factors for osteoarthritis.

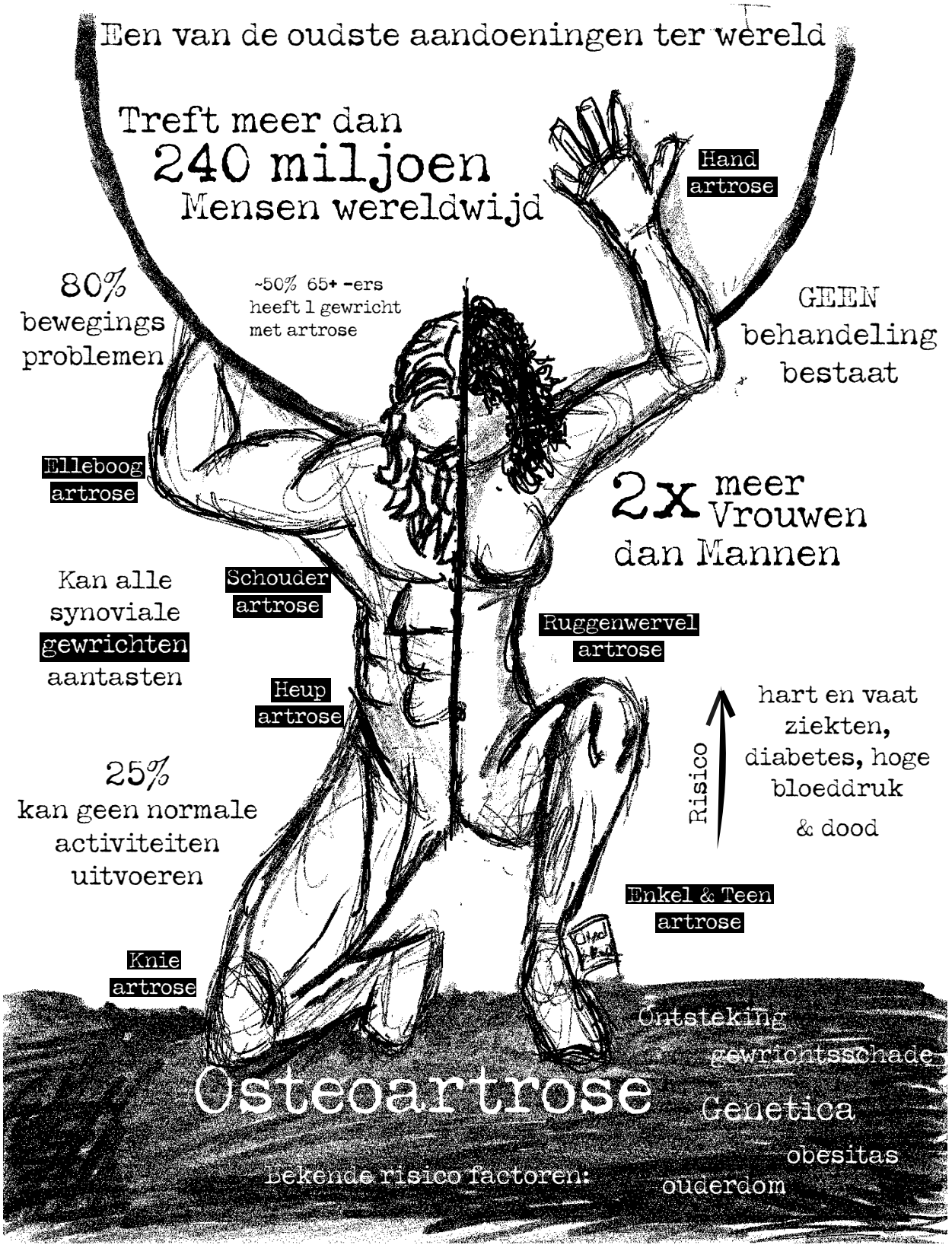
Chapter 5.1 we presented two novel large scale population bases microbiome data sets to be used for microbiome research: Generation R (n=2,111) and Rotterdam Study microbiome (n=1,427) datasets. Using these datasets we identified compositional and functional differences in gut microbiome between children and adults. Moreover, we confirm that these microbiome profiles can serve as reference for future studies on specific human disease susceptibility in childhood, adulthood and specific diseased populations.

Using the Rotterdam Study microbiome dataset as presented in the previous chapter, in **Chapter 5.2**, the possible relationship between osteoarthritis, joint pain and gastrointestinal microbiome was investigated. Previous research had suggested that osteoarthritis-related joint pain and inflammation may be triggered or exacerbated by endotoxins produced by the gastrointestinal microbiome. Using the stool microbiome data set of the Rotterdam Study (n=1,427), an association between relative abundance of Streptococcus spp. and osteoarthritis-related knee pain was identified. These results were replicated in an independent cohort (Life-Lines Deep, n=867). Increased abundance of Streptococcus spp. in the gastro-intestinal microbiome was associated with increased knee joint pain, validated by absolute quantification of Streptococcus species. In addition, evidence was presented that this association is driven by local inflammation in the knee joint. Indicating the microbiome as a possible therapeutic target for osteoarthritis-related knee pain.

Last, in **Chapter 6**, all the findings in this thesis are combined and put into a broader perspective. Finally, recommendations are given the next steps that need to be taken in order find preventive or curative treatments for one of the world's oldest diseases.

Despite the many advances in osteoarthritis genetics in the recent years, ~148 reported associated loci, much remains to be discovered. Many hundreds gene and thousands of genetics variants are probably involved in genetic osteoarthritis risk. Next, much of the interaction between genetics and the environment remains unknown in osteoarthritis pathology and risk, e.g., gut microbiome, osteoarthritis and genetics. In sum, much remains unknown about osteoarthritis pathology and risk. However, despite these missing pieces, many hopeful discoveries have been made, which in the future may lead to novel therapeutic and preventative strategies. This includes the novel

osteoarthritis associated genes, the microbiome and vitamin K, with further research these discoveries may one day end up in clinical applicates for one of the world's oldest disease, osteoarthritis.


Een van de oudste aandoeningen ter wereld
Treft meer dan
240 miljoen
Mensen wereldwijd
Hand artrose
80%
bewegings problemen
~50% 65+ -ers heeft 1 gewricht met artrose
GEEN behandeling bestaat
Elleboog artrose
2x meer Vrouwen dan Mannen
Kan alle synoviale gewrichten aantasten
Schouder artrose
Ruggenwervel artrose
Heup artrose
Risico
hart en vaat ziekten, diabetes, hoge bloeddruk & dood
25%
kan geen normale activiteiten uitvoeren
Enkel & Teen artrose
Knie artrose
Ontsteking
gewrichtsschade
Genetica
obesitas
ouderdom
Bekende risico factoren:
Osteoartrose

NEDERLANDSE SAMENVATTING

"De meest fundamentele ideeën van de wetenschap zijn in wezen eenvoudig en kunnen in de regel worden uitgedrukt in een taal die voor iedereen begrijpelijk is."
~ Albert Einstein, The Evolution of Physics (1938)

"laten we een voorbeeld nemen aan Annie M.G. Schmidt met Jip en Janneke"
~ Bas Eenhoorn pleit voor "Jip-en-Janneke Taal" in de Tweede Kamer (2002)

"Een beetje onnauwkeurigheid scheelt soms heel veel uitleg"
~ H. H. Munro, Clovis On The Alleged Romance Of Business (1924)

Het doel van de wetenschap is eigenlijk samen te vatten tot twee punten: *i)* het systematisch verweven van kennis en *ii)* deze kennis overdragen. Nu gaat dat eerste punt vrij aardig in de huidige academische wetenschap, maar het probleem zit hem in het tweede punt, die kennis overdracht. Tussen wetenschappers (en andere vakidioten) is er een heel specialistisch systeem van publicaties, posters, orals (presentaties), peer review en congressen om kennis over te dragen. Waar deze methoden (over het algemeen) goed werken tussen wetenschappers, is het compleet alien voor de buitenstaander, "It's not life as we know or understand it"~Lt.Com. Spock. De drempel is (te) hoog voor buitenstaanders om mee te kunnen doen aan dit systeem van kennis overdracht. De wetenschap en de kennis die daarbij verworven wordt, moet toegankelijk zijn voor iedereen. Wetenschappers hebben de morele en ethische plicht om hun kennis ook effectief over te dragen aan niet-wetenschappers. Dit is geen gemakkelijke opgave, maar dat is kennisoverdracht nooit. Dus zie hier mijn poging om meer dan 7 jaar artrose onderzoek in "Jip-en-Janneke taal" samen te vatten.

Nederlandse Samenvatting

Osteoartritis (artrose), is de meest voorkomende chronische gewrichtsaandoening wereldwijd. In Nederland alleen al zijn er ongeveer 1 miljoen mensen met artrose in één of meerdere gewrichten. Niet alleen komt artrose veel voor, het is ook één van de oudste aandoeningen ter wereld; er zijn meerdere gefossiliseerde dinosaurus gewrichten met artrose gevonden. Toch bestaan er momenteel geen genezende behandelingen voor artrose. Dus het wordt hoog tijd dat daar verandering in gaat komen! Het doel van dit proefschrift was dan ook om beter inzicht te krijgen in het ontstaan en de verloop (pathologie) van artrose. Dit proefschrift was met name gericht om genen te ontdekken die betrokken zijn bij de pathologie van artrose, om hopelijk voorkomende en genezende behandelingen te kunnen ontwikkelen.

In **Hoofdstuk 1** werd een inleiding gegeven van artrose en de lange geschiedenis van deze ziekte, van de eerste beschrijvingen van de symptomen door Hippocrates (460-375 B.C.E.) tot de start van dit proefschrift. De wetenschappelijke naam voor artrose, osteoartritis, komt van het oude Grieks ὀστέον (ostéon): "bot" en artritis van de combinatie artr- ἄρθρον (árthron): "gewricht" en -itis ῖτις (îtis): "met betrekking tot" ofwel "bot met betrekking tot het gewricht". Ondanks deze naam, tast artrose niet alleen het bot in het gewricht aan, maar het hele gewricht en alle weefsels daarin. Kenmerkend voor artrose zijn de (onomkeerbare) afbraak van het gewrichtskraakbeen en de vorming van benige (bot) uitgroeisels aan de rand van het gewricht (osteofyten). Artrose is een complexe ziekte, in de wetenschappelijke betekenis van complex: artrose wordt veroorzaakt door een samenwerking tussen omgevingsfactoren en genetica. Hierbij wordt de betrokkenheid van genetica voor artrose geschat op ~39% - 65%, afhankelijk van het gewricht (hand, knie of heup). Dus een groot gedeelte van het risico voor het krijgen van artrose wordt bepaald door de genetica. Hoe en welke genetische variatie en genen betrokken zijn bij artrose was het hoofddoel van dit proefschrift.

Het DNA (Desoxyribonucleïnezuur) is een groot en ingewikkeld molecuul dat alle genetische informatie bevat: de "blauwdruk" voor de ontwikkeling, functioneren en onderhoud van een organisme. Het DNA bestaat uit twee om elkaar gewikkelde strengen (dubbele-helix) van tegenover elkaar liggende nucleotiden (baseparen): adenine (A) met thymine (T) en cytosine (C) met guanine (G). Deze streng van baseparen (sequentie), is in mensen ~3.2 miljard baseparen lang en verdeeld over 24 chromosomen. Deze sequentie codeert voor allerlei erfelijke informatie, bijvoorbeeld, haar en oog kleur, maar ook mindere zichtbare zaken als hoe het immuunsysteem reageert. De humane DNA sequentie is nagenoeg identiek en toch ook uniek voor ieder mens op aarde. Gemiddeld genomen komt het DNA tussen twee willekeurige mensen voor ~99.5% overeen.

De ~0.5% variatie bepaald o.a. de verschillen: blauwe of bruine ogen? Krullend of stijl haar? Maar ook een individu's risico voor bepaalde ziekten, als artrose. Verreweg de meeste genetische variaties zijn "Single Nucelotide Variations/ Polymorphisms (SNVs of SNPs)", of wel variaties van één enkele nucleotide.

Dit proefschrift gebruikte voornamelijk genoomwijde associatie studies (GWAS) om te ontdekken welke genetische variaties (SNVs of SNPs, spreek uit als "snip") gelinkt (ge-associeerd) zijn aan het hebben of krijgen van artrose. Bij een GWAS wordt er gekeken of er genetische variaties zijn die statistisch gezien, vaker voorkomen bij mensen met de onderzochte eigenschap dan bij mensen die die eigenschap niet hebben. Bijvoor-beeld: welke genetische variatie(s) komen vaker voor bij mensen met artrose dan bij de mensen die geen artrose hebben? Voor GWAS zijn er dus DNA en uiterlijke, wel of geen artrose(fenotypen), gegevens nodig. Van de resultaten van de vele GWAS studies die er gedaan zijn sinds de allereerste GWAS (in 2002) weten we dat de gevonden varianten vaak maar een klein beetje bijdragen aan de onderzochte eigenschap. Dit betekend dat er gegevens van heel veel mensen nodig zijn voor een succesvolle GWAS studie, denk aan enkele duizenden voor een "kleine" studie en tienduizenden tot honderdduizen-den mensen voor "gemiddelde" tot "grote" studies. Gelukkig kon voor dit proefschrift gebruik gemaakt worden van data van de "*Rotterdam Study (RS)*" ofwel het "*Erasmus Rotterdam Gezondheid Onderzoek, ERGO*".

De *Rotterdam Study* is een langlopend bevolkingsonderzoek gericht op gezond-heidsproblemen die zich vooral voordoen op oudere leeftijd. Dit bevolkingsonderzoek was gestart in 1990 en bestaat inmiddels uit bijna 20.000 mensen van 40 jaar en ouder uit de Rotterdamse wijk Ommoord. De deelnemers worden iedere 3 tot 4 jaar uitgen-odigd in het ERGO onderzoekscentrum voor een uitgebreid en uitvoerig gezondheid-sonderzoek; waaronder DNA gegevens, röntgenfoto's, MRI, medicijn gebruik, vragen-lijsten etc. Door deze grootschaligheid en uitgebreide opzet is de *Rotterdam Study* een zeer wel bekend begrip in de medisch-wetenschappelijke wereld.

In **hoofdstuk 2** werd gebruikt gemaakt van verschillende soorten artrose kenmerken (fenotypen) om via genoomwijde associatie studies nieuwe genetische associaties te vinden met artrose. Door gebruik te maken van deze artrose fenotypen laat dit hoofd-stuk zien dat je zelfs met een kleine studie (voor genetisch onderzoek "kleine" studies) toch genetische varianten kan ontdekken die gelinkt zijn aan artrose.

In **hoofdstuk 2.1**, is de minimale afstand tussen de heup gewrichtsdelen (*mini-mum joint space width, mJSW*) gebruikt als maat voor de ernst van de kraakbeen afname

in het heup gewricht. Kraakbeen afname is één van de kenmerken van artrose, maar kraakbeen afname alleen betekend niet dat er artrose in het gewricht is. Deze maat meet dus niet direct de ernst van de (mogelijke) artrose in het gewricht, maar meet een "tussenliggend" (intermediair) kenmerk (fenotype) voor artrose. Zulke intermediaire fenotypen, worden ook wel endofenotypen genoemd. Dit zijn kenmerken (fenotypen) die dichter gerelateerd zijn met de onderliggende erfelijkheid (groei van kraakbeen) dan de ziekte zelf (artrose). Omdat een endofenotype dichter bij de genetica ligt kan ook onze "kleine" (< 10.000 mensen) genetische studie meerdere genetische variaties identificeren die gelinkt waren aan kraakbeen dikte: vier nieuwe associaties en twee bekende associaties werden gevonden. Met behulp van verschillende biologische data; kraakbeen gen expressie (welke genen "aan" en "uit" staan in kraakbeen), kraakbeen eQTL data (of een genetische variant een gen "aan" of "uit" kan zetten) en literatuur data (wat is er al bekend over de genetische variant en de genen); werden er genen gevonden die mogelijk betrokken zijn bij het ziekte proces van artrose: *RUNX2, TGFα, PIK3R1, SLBP/FGFR3, DOT1L/GADD45B en TREH/DDX6*. Van deze lijst van genen (in hun afkortingsnamen) waren er twee, *TGFα* en *FGFR3,* die hoger tot expressie kwamen (meer "aan" stonden) in kraakbeen aangetast door artrose dan in kraakbeen niet aangetast door artrose.

In **hoofdstuk 2.2**, zijn er drie hand-artrose fenotypen gemaakt gebaseerd op de ernst en het patroon van de artrose in de hand. Eén hand (vingers en duim) bevat maar liefst 16 gewrichten, en al deze gewrichten kunnen aangetast worden door artrose. Dit maakt het lastig om de hand te onderzoeken: om één studie te doen naar ieder hand gewricht apart voor beiden handen (32 gewrichten), is wel erg veel werk. Maar, bij hand artrose is er vaak niet één gewricht aangedaan maar meerdere. Door te onderzoeken welke gewrichten vaak samen of tegelijk zijn aangedaan door artrose, werd de hand opgedeeld in: alleen de vinger gewrichten, de duim gewrichten en heel de hand (alle gewrichten samen). De ernst van de artrose in alle hand gewrichten werd gemeten via de Kellgren en Lawrence (KL) schaal. Deze meet op röntgen foto's o.a. de mate van de gewrichtsspleet vernauwing en aantal osteofyten. Door de KL-schaal (van 0 tot 4) op te tellen voor ieder gewricht werd er een "KL-som score" gemaakt die de ernst van de artrose weer geeft: duim artrose ernst, vinger artrose ernst en hand artrose ernst.

Wederom, door gebruik te maken van specifieke artrose fenotypen, werd in een "kleine" (< 10.000 mensen) genetische studie nieuwe varianten gelinkt aan artrose gevonden. Twee nieuwe genetische links met duim artrose en twee bekende hand/vinger artrose werden gevonden. Eén van deze genetische variaties met duim KLsom (ernst van de duim artrose) werd ook gelinkt aan het hebben van klinisch aangetoonde duim artrose. Dit is belangrijk, want het geeft aan dat deze genetische variatie ook echt

te maken heeft met artrose en niet alleen met het "intermediair" artrose fenotype. Voor deze genetische variant, rs10916199 (het SNV identificatie nummer) werd onderzocht hoe deze genetische variant kan leiden tot een groter risico op duim artrose. Dit werd gedaan door middel van meerdere benaderingen die gebruikmaken van verschillende biologische data. Uit de onderzochte data blijkt dat de genetische variant mogelijk in een "regel knop gebied" (regulatory region) voor het *WNT9A* gen ligt en op die manier zorgt dat dit gen "te veel" of op het verkeerde moment "aan" staat. Daarnaast vonden we ook, dat veel van de bekende genetische variaties gelinkt aan knie of heup artrose ook risico kunnen geven voor hand artrose.

Hoofdstuk 3 ging verder met het gebruik van het hand KLsom fenotype. Er werd één nieuwe genetische associatie gevonden, ook weer in een "kleine" (< 10.000 mensen) genetische studie. Wederom werd met behulp van vervolg onderzoek een nieuw artrose gelinkt gen ontdekt. In het tweede deel van dit hoofdstuk worden de implicaties van dit nieuwe artrose gen onderzocht en de mogelijke consequenties hiervan voor artrose preventie en behandeling. Deze hoofdstukken laten de waarde zien van genetische studies naar artrose en het belang van diepgaand vervolg onderzoek hierbij.

In **hoofdstuk 3.1** werd ook gebruikt gemaakt van hand KLsom (de ernst van de artrose op röntgen foto's) om de tweede genetische associatie te ontdekken voor hand artrose. Een genetische variant in het *MGP* gen werd gelinkt aan een hoger risico voor hand artrose. Deze genetische variant zorgde voor minder *MGP* expressie in kraakbeen, het *MGP* gen stond minder "aan". *MGP* zorgt ervoor dat o.a. het kraakbeen niet verkalkt, en kraakbeen verkalking gebeurt ook bij artrose gewrichten. Dus verminderde expressie van *MGP* kan leiden tot een verhoogde verkalking van het kraakbeen en daarmee een hoger risico op artrose. Bovendien kan MGP alleen de verkalking van kraakbeen remmen als het wordt "geactiveerd " (γ-carboxylatie) door vitamine K. Dit kan de verklaring zijn voor het verband tussen lage vitamine K bloed waarden en een hoog risico op artrose.

Met behulp van deze nieuwe kennis werd in **hoofdstuk 3.2** de mogelijke rol van vitamine K remmende bloedverdunners in artrose ontwikkeling onderzocht. Verschillende bloedstollingsfactoren, stofjes (eiwitten) die betrokken zijn bij het helpen stollen van het bloed, hebben ook de vitamine K "activering"(γ-carboxylatie) nodig om te functioneren. Daarom remmen de meeste, en meest gebruikte, bloedverdunners dit activeringsproces van vitamine K, met als gevolg dat de activering van MGP ook geremd wordt. Op zijn beurt, lijdt een verminderde MGP activiteit weer naar een verhoogd risico op artrose. Mensen die dus deze bloedverdunners (acenocoumarol) gebruikten,

bleken een tweemaal groter risico te hebben op het krijgen van of het verergeren van knie en heup artrose, dan de mensen die geen bloedverdunners gebruikten. Daarbij hadden mensen die drager waren van de genetische *MGP* variatie gelinkt aan artrose, en een genetische variatie (*VKORC1*) gelinkt aan een hogere dosis bloedverdunners, een meer dan viervoud verhoogd risico op artrose bij gebruikt van bloedverdunners. In een populatie bestaande uit mensen van Europese afkomst is ~30% van de populatie drager van beide genetische variaties. Wat dus inhoudt dat een derde van alle gebruikers van deze bloedverdunners (acenocoumarol) word blootgesteld aan een zeer verhoogd risico op het krijgen of verergeren van artrose. Er bestaan bloedverdunners die niet de vitamine K activatie remmen, dus als de klinische indicatie het toestaat zouden deze bloedverdunners voorgeschreven kunnen worden. Bovendien bevestigde de resultaten van dit hoofdstuk het belang van MGP en vitamine K in het ziekte proces van artrose, waardoor vitamine K verder onderzocht moet worden als mogelijk doelwit voor artrose behandelingen.

In genoomwijde associatie studies (GWAS) zijn de belangrijkste succesfactoren: het fenotype dat gebruikt wordt en grootte van de studie. In de hoofdstukken 2 tot en met 3 is het effect van fenotype onderzocht, in **hoofdstuk 4** werd de grootte van de studie onderzocht. Dit werd gedaan door samen te werken in het tot nu toe grootste consortium (samenwerking) voor artrose-genetica: Genetics of Osteoarthritis (GO) consortium. Het GO-consortium bestaat uit meer dan 20 verschillende bevolkingsonderzoeken van over de hele wereld, voornamelijk Europese bevolkingen, en bevatte in totaal 177.517 gevallen van artrose en 649.173 controles.

Hoofdstuk 4.1 presenteerde de eerste resultaten van de 11 osteoartritis genoom-wijde associatie studies (GWAS) uitgevoerd in het GO-consortium. In totaal werden 100 genetische varianten ontdekt die gelinkt zijn aan artrose, waarvan 52 nieuw waren en 48 bekende. Hiervan zijn er heel veel gelinkt aan zowel gewicht-dragende(heup, knie, rug) als niet gewicht-dragende gewrichten (hand, vinger, duim). Daarbij werd ook de eerste genetische link gevonden voor ruggenwervel artrose. Door ook in deze studie gebruikt te maken van vervolg onderzoek en biologische gegevens konden er niet alleen nieuwe artrose genen worden gevonden maar ook mogelijke genetische links met (gewrichts-)pijn, ontsteking en zenuwstelsel ontwikkeling. Al deze data en resultaten geven een schat weer aan mogelijke nieuwe inzichten in artrose, en bieden nieuwe mogelijkheden aan voor het ontdekken en ontwikkelen van geneesmiddelen voor artrose.

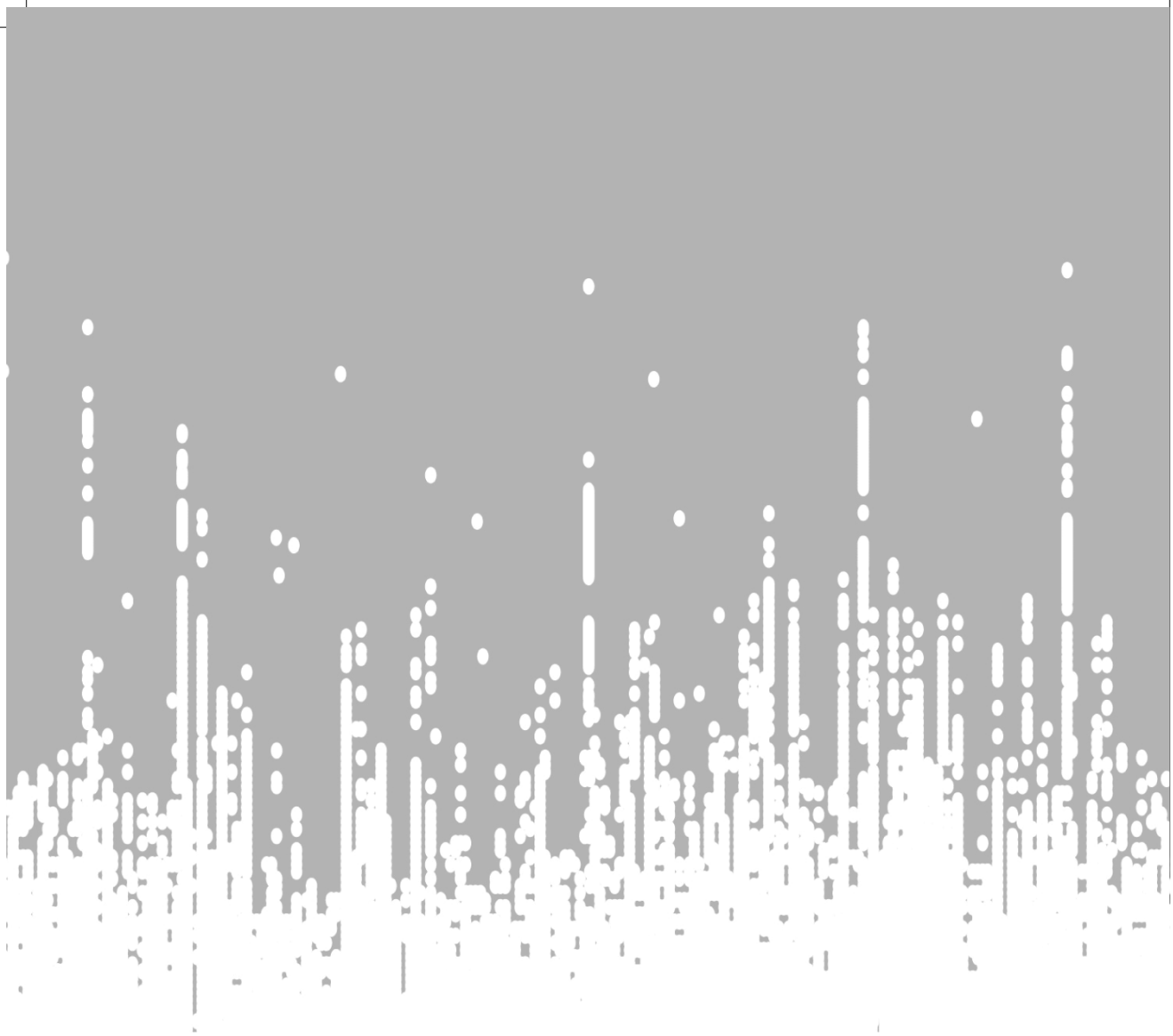
Artrose is een complexe ziekte en dus niet alleen de genetica speelt een rol in het krijgen van artrose maar ook vele andere factoren. Een nieuw ontdekte mogelijke factor voor artrose is de darmflora: de bacteriën en micro-organismen in leven in het spijsverteringskanaal. In **hoofdstuk 5** werd er gekeken naar de mogelijke relatie tussen de biljoenen bacteriën in ons spijsverteringskanaal leven en artrose.

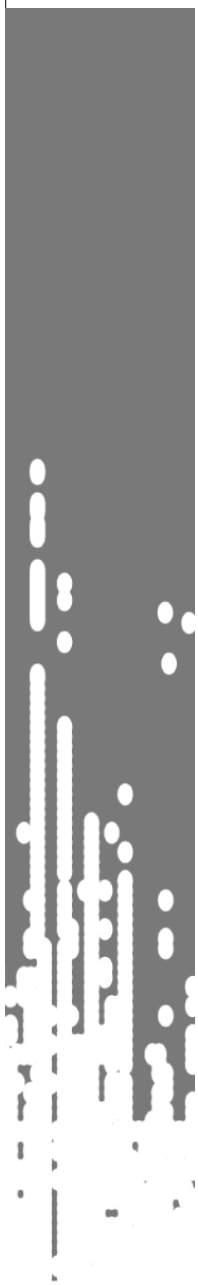
In **Hoofdstuk 5.1** wordt de eerste data over de darmflora van o.a. de *Rotterdam Study* gepresenteerd. Deze dataset bevat informatie over de compositie (welke bacterie soorten en hoeveel) van de darmflora van maar liefst 1427 mensen uit de *Rotterdam Study*. Daarnaast werd ook de darmflora data van de *Generation R* studie gepresenteerd, dit is ook een populatie studie. Maar anders dan de *Rotterdam study* richt *Generation R* zich op kinderen en de *Rotterdam Study* op ouderen (45-plussers). Door gebruikt te maken van deze twee datasets tonen we aan dat er verschillen bestaan in de compositie van de darmflora tussen kinderen en volwassenen. Volwassenen hebben een veel diverser darmflora dan kinderen.

In **hoofdstuk 5.2** werd de mogelijke relatie tussen artrose, gewrichtspijn en de darmflora onderzocht. Eerder onderzoek had gesuggereerd dat artrose-gerelateerde gewrichtspijn en -ontsteking kan worden veroorzaakt of verergerd door stofjes (endotoxinen) die worden geproduceerd door de darmflora. Dit werd onderzocht met behulp van gegevens over de samenstelling van de darmflora van 1.427 mensen uit de Rotterdam Studie. Een verband werd gevonden tussen het relatieve aantal *Streptococcus spp.* bacteriën in de darmflora en artrose gerelateerde knie pijn. Deze bevinding werd ook gerepliceerd in een onafhankelijke studie (Life-Lines Deep, n = 867) en bevestigd door absolute kwantificatie (meten van de absolute aantallen bacteriën) van *Streptococcus spp.* in de *Rotterdam Studie*. Een verhoogd aantal *Streptococcus spp.* bacteriën in de darmflora werd geassocieerd met verhoogd kniegewrichtspijn. Daarnaast werd verhoogde aantallen *Streptococcus spp*. geassocieerd met ernstigere lokale ontsteking in het kniegewricht. Mogelijk produceren *Streptococcus spp.* endotoxines, die ontstekingsopwekkend zijn en zo ontsteking, schade en pijn in het knie gewricht veroorzaken. Omdat de darmflora beïnvloedbaar is (v.b.: antibiotica, dieet etc.) zou dit een nieuwe mogelijkheid vormen voor het ontwikkelen van mogelijke artrose behandel methoden.

Ten slotte worden in **hoofdstuk 6** alle bevindingen in dit proefschrift gecombineerd en in een breder perspectief geplaatst en aanbevelingen gegeven voor de toekomst van artrose onderzoek. Ondanks dat er in de afgelopen jaren ontzettend veel nieuwe kennis

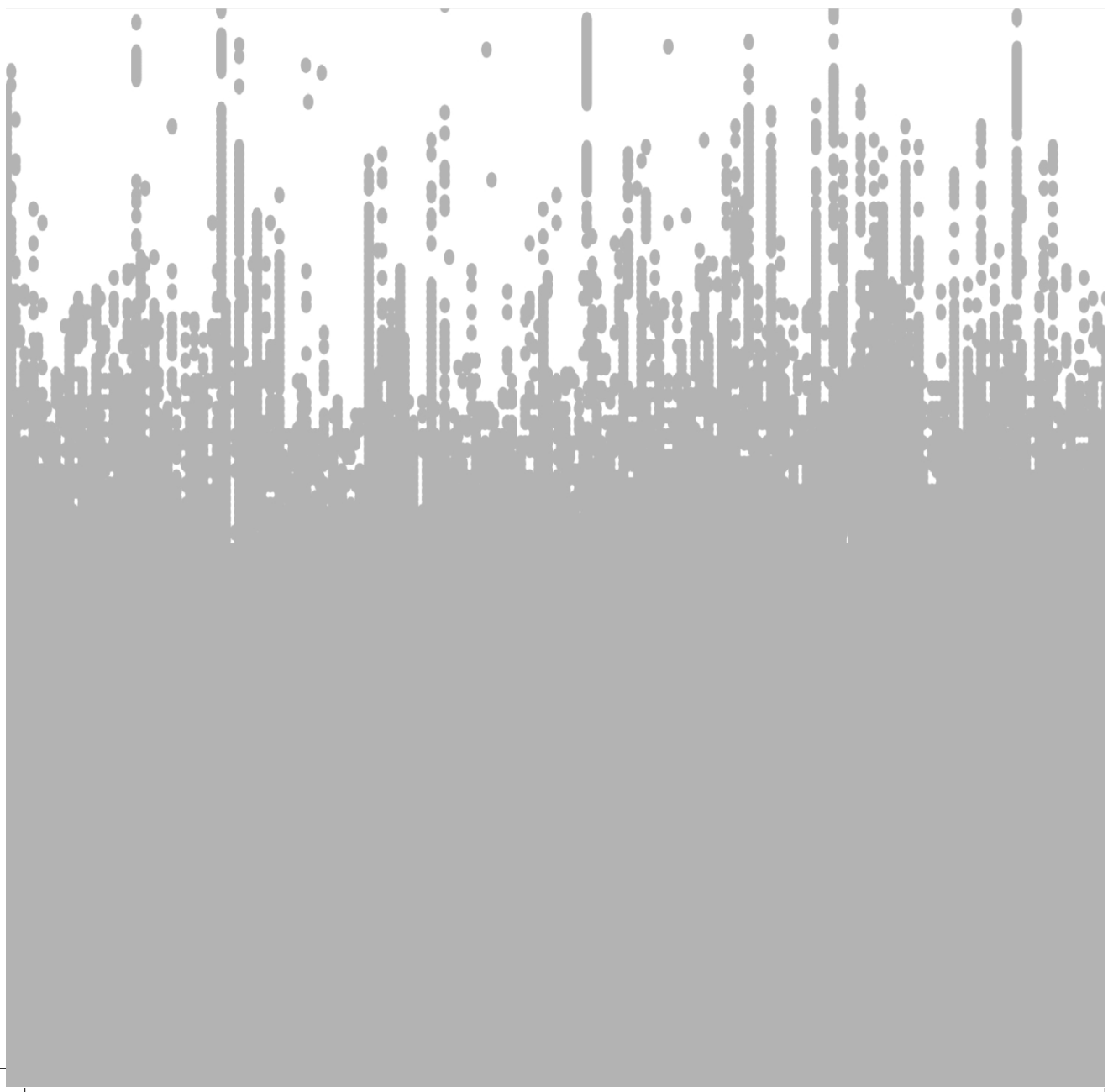
is opgedaan over artrose, met name over de genetica achter artrose (er zijn nu ~148 genetische varianten bekend) is dit nog maar het topje van de ijsberg. Vele honderden genen en nog veel meer genetische variaties zijn mogelijk betrokken bij artrose. Daarnaast is ook nog niet altijd even duidelijk hoe de omgevingsfactoren en genetische factoren samen leiden tot artrose, of is er nog te weinig kennis over hoe bepaalde factoren (v.b. de darmflora) bijdragen aan het risico op artrose. Er is en blijft nog veel onbekend over artrose, ondanks de incomplete informatie over artrose zijn er wel al hoopvolle ontdekkingen en ontwikkelingen die kunnen leiden tot behandelingen of preventie van artrose (v.b. Vitamine K en bloedverdunners). Hopelijk is de wetenschap met dit proefschrift weer enkele stappen dichterbij een oplossing gekomen voor het behandelen van een van 's werelds oudste ziekten.





Chapter 8

Appendices



About the Author

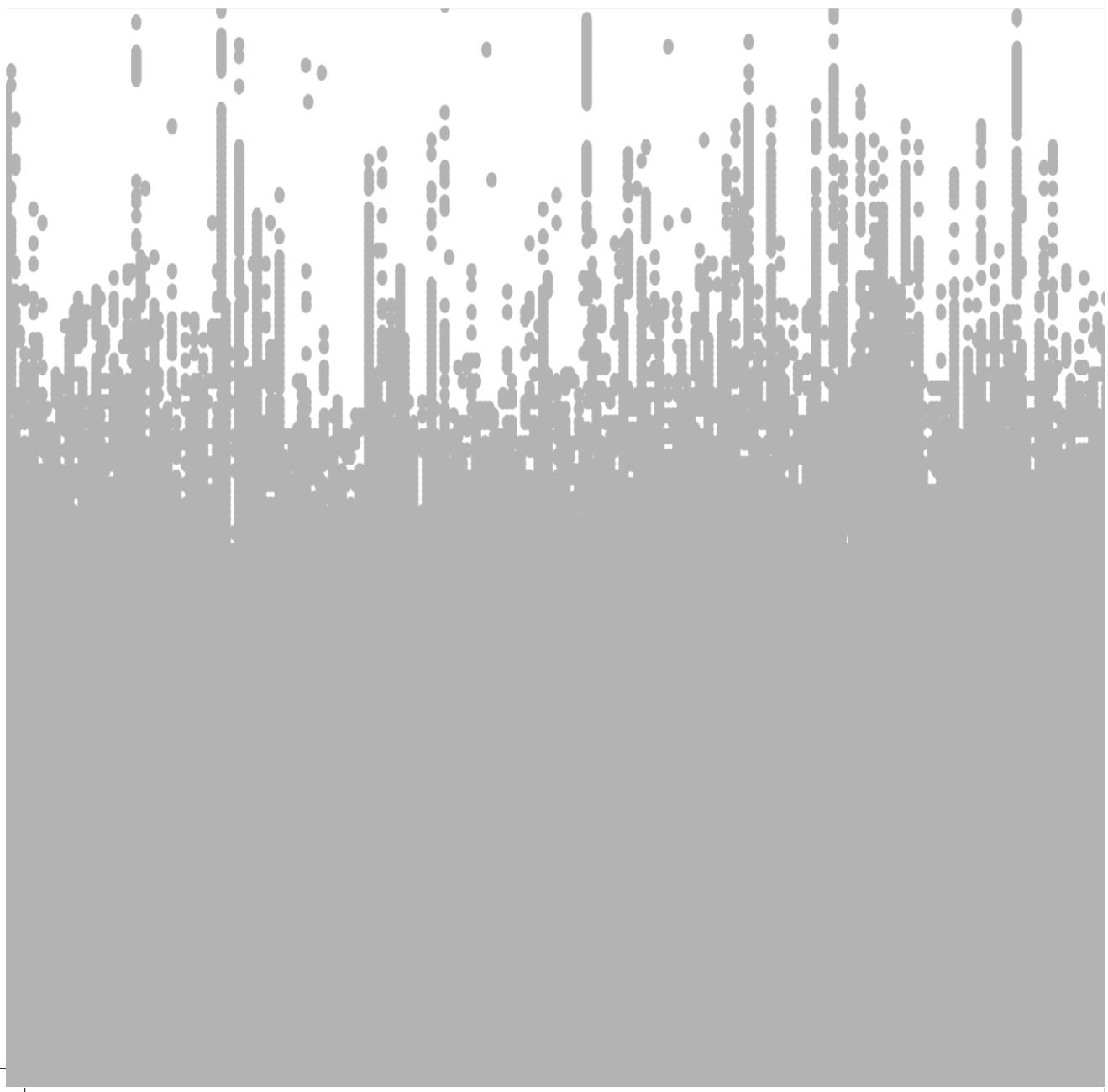
About the author

Cindy Germaine Boer was born on the 29th of May 1989 in Sliedrecht, the Netherlands. After completing her secondary school at *CSG De Lage Waard* in 2007, she tarted her Bachelor degree in Biology at the University of Utrecht. During her Bachelor degree, Cindy focused on molecular and cell biology, with a minor in virology. In 2013, she received her Master of Science degree in Biomedical Sciences at the University of Utrecht in Cancer Genomics and Developmental Biology. During her Master degree, Cindy completed internships at the University Medical Center Utrecht, department of Medical Oncology on bipolar spindle assembly in *C. elegans,* and at INSERM(*Institut national de la santé et de la recherche médicale*), department of Human Virology in Lyon, France on immune modulation of the measles vaccine.

In may of 2013, Cindy started as PhD candidate at the department of Internal Medicine in the Erasmus MC, Rotterdam, under the supervision of Prf. Dr. André Uitterlinden and dr. Joyce B.J. van Meurs. In 2017 and 2019, she received the *Osteoarthritis Research Society International*(OARSI) Young Investigator- Highest rated abstract award, for her work on osteoarthritis and the gut-microbiome. In 2018 Cindy joined the *Genetics of Osteo-arthritis*(GO) consortium effort and and visted the research group of prof. dr. Elefthe-ria Zeggini in Cambridge (Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cam-bridge, UK) and in Neuherberg (Institute of Translational Genomics, Helmholtz Zentrum München, Neuherberg, Gemany).

In 2017, Cindy received the diagnosis of Usher syndrome and became involved in patient outreach groups for retinal disorders at the Dutch eye society(Oogvereniging). where she was also involved in lobying the Dutch government for approval of retinal gene-therapy(LUXTURNA). Cindy also serves on the medical advisory council of the Usher Syndrome Foundation(Stichting Usher Syndroom).

In her spare time Cindy likes to practice Historic European Martial Arts (HEMA), focussing on the longsword and joins in 15th century Historical re-enactment.



PhD portfolio

PhD Portfolio

Name:	Cindy Germaine Boer
Erasmus MC Department:	Internal Medicine
Research school:	Molecular Medicine Post Graduate School
PhD period:	May 13th 2013 – December 15th 2020
Promoter:	Prof. dr. André A.G. Uitterlinden
Supervisor[co-promoter]:	Dr. Joyce B.J. van Meurs
ORCID:	orcid.org/0000-0003-4809-0044
Publons:	publons.com/a/1367544/

PhD training

Courses	**Year**
Principles of Genetic Epidemiology (ESP43)	2013
Genomics in Molecular Medicine (ESP 57)	2013
Genome Wide Association Analysis (ESP29)	2013
Population Genomics of Complex Traits and Disease (SNP Course)	2013
RNA-seq data analysis (NBIC, LUMC Leiden)	2013
The R-course (MolMed)	2013
Research Management for PhD-students (MolMed)	2013
Research Integrity (Erasmus MC)	2014
Osteoarthritis Epigenetics Workshop(KNAW)	2015
Course on Python (MolMed)	2015
Biomedical English writing (MolMed)	2017
R and Big Data (Wageningen)	2019

Inter(national) Meetings	**Presentation**	**Year**
"*Novel variants for cartilage thickness and hip osteoarthritis: revealing genes implicated in cartilage and bone development*",		
- OARSI, Paris, France	Oral presentation	2014
"Discovery of rare coding variants for hip OA by exome-sequencing"		
- OARSI, Paris, France	Poster	2014
"Genetic variants in the SUPT3H-RUNX2 locus confer susceptibility for bone and cartilage related disorders via long-range regulation of RUNX2",		
- Molecular Medicine day, Rotterdam, the Netherlands	Poster	2014

Inter(national) Meetings (continued)	Presentation	Year
- Science days, Antwerp, Belgium	Poster	2015
- ECTS-IBMS, Rotterdam, the Netherlands	Oral presentation	2015
- OARSI, Seattle, Washington, USA	Oral presentation	2015
- Epigenetics of Common Diseases, Hinxton,		
- Cambridge, UK	Poster	2015
"Annotating hits from EWAS".		
- PACE consortium, invited webinar speaker	Oral presentation	2015
"Coding variants in the MGP gene are involved in Osteoarthritis of the hand",		
- OARSI, Seattle, Washington, USA	Poster	2015
- OARSI, Amsterdam , the Netherlands	Poster	2015
- Science days, Antwerp, Belgium	Oral presentation	2016
- Molecular Medicine Day, Rotterdam, the Netherlands	Oral presentation	2016
- ESHG, Barcelona, Spain	Poster	2016
"Rotterdam Study Exome: Genetic Variation in Skeletal Dysplasia Genes"		
- OARSI, Las Vegas, Nevada, USA	Poster	2017
"The role of the gut microbiome in osteoarthritis and joint pain"		
- Molecular Medicine day, Rotterdam, the Netherlands	Poster	2017
- OARSI, Las Vegas, Nevada, USA	Oral presentation	2017
- Science days, Antwerp, Belgium	Oral presentation	2018
- CHARGE investigator meeting, Rotterdam,		
- the Netherlands	Oral presentation	2018
- ESHG, Milaan, Italy	Poster	2018
- OARSI, Toronto, Canada	Oral presentation	2019
"Genetic exploration of osteoarthritis (endo)phenotypes provides novel biological insight into osteoarthritis"		
- OARSI, Liverpool, United Kingdom	Oral Presentation	2018
"Large-scale global multi-ethnic GWAS doubles the number of osteoarthritis loci and identifies new treatment targets"		
- ASHG, Houston, Texas, USA	Oral Presentation	2019
"The role of Vitamin K and MGP in osteoarthritis"		
- OARSI, Toronto, Canada	Poster	2019
- Science Days, Sint-Michielsgestel, the Netherlands	Oral Presentation	2020

Other	Year
Co-chair bi-weekly meetings "MoleEpi"	2016-Present

Teaching – Courses and Practicals		Year
ESP57	Genomics in Molecular Medicine	2014-2018
	Lecturing	
	Designing of practical	
	Coach at computer practical(s)	
SNP course	SNPs and human Diseases	2014-2020
	Lecturing	
	Designing of practical	
	Coach at computer practical(s)	
GEO3	Genome-wide association studies	2016
	Lecturing	
	Designing of practical	
	Coach at computer practical	
GWAS-course	Genome-wide association studies Lecture	2020

Teaching – Supervising Students			Year
Idries Habib	Bachelor Thesis	Hogeschool Rotterdam	2015
Dennis Boone	Internship	Avans Hogeschool	2016
Kamal Arabe	Internship	University of Utrecht	2016
Michiel van Berkel	Internship	Avans Hogeschool	2018
Justin D. Boer	Research Internship	Erasmus MC	2019-2020
Segio chaves-Chavez	Master Mentor	HAN Nijmegen	2018-2020

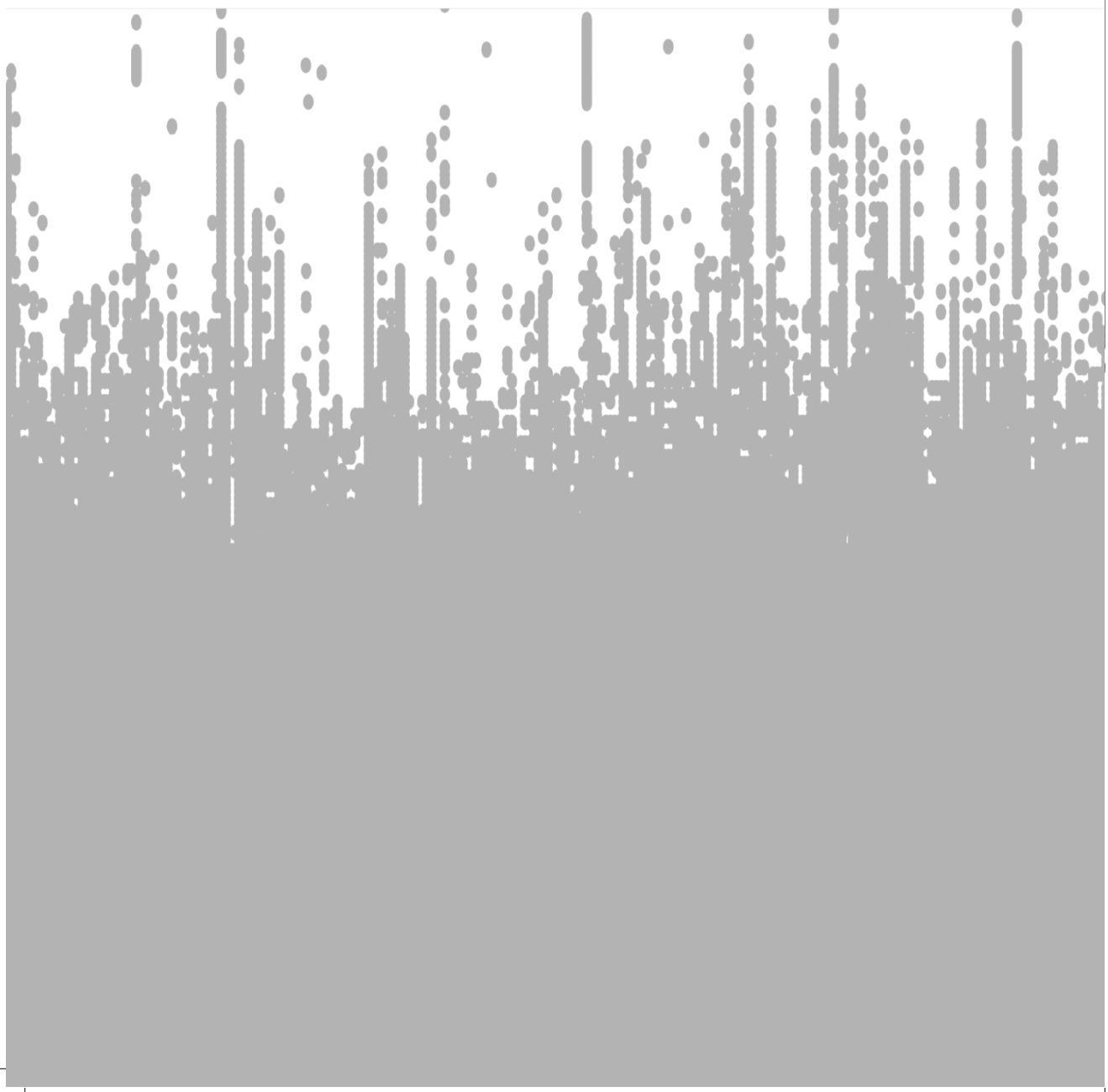
Peer review (referee activities)	Year
- Annals of the Rheumatic Diseases	2016
- Cartilage	2017-2018
- Osteoarthritis & Cartilage	2018-2019
- Nature communications	2020

Grants and Awards	Year
- Rheumafonds – OARSI travel grant	2015
- Best oral presentation, Molecular Medicine Day, Rotterdam, the Netherlands.	2016

- Rheumafonds – OARSI travel grant 2016
- Young investigator award - highest rated abstract, OARSI, Las Vegas Nevada 2017
- Young investigator award - Highest rated abstract, OARSI, Toronta, Canada 2018

International Research Visits	**Year**
Visiting worker – (1 month) Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK	2018
Visiting worker – (1 week) Helmholtz Insitute, Munchen Germany	2019

Volunteer work	**Year**
Dutch eye society (Oogvereniging)	2018-present
- Retina group core-member	
- Luxturna/Gene therapy advisory board	
Patiënt Panel "Uitzicht" Foundation	2018-2020
Usher Syndrome Foundation (Stichting Usher Syndroom)	
Scientific advisory board member (WAR)	2018-present



LIST OF PUBLICATIONS

Publications in this thesis

van Meurs J.B., **Boer C.G.**, Lopez-Delgado L, Riancho JA., **Role of Epigenomics in Bone and Cartilage Disease***, J Bone Miner Res. 2019 Feb;34(2):215-230. doi: 10.1002/jbmr.3662. Epub 2019 Feb 4.*

Castaño-Betancourt M.C.*, Evans D.S.*, Ramos Y.F.M.*, **Boer C.G.***, Metrustry S., Liu Y., den Hollander W., van Rooij J., Kraus V.B., Yau M.S., Mitchell B.D., Muir K., Hofman A., Doherty M., Doherty S., Zhang W., Kraaij R., Rivadeneira F., Barrett-Connor E., Maciewicz R.A., Arden N., Nelissen R.G.H.H, Kloppenburg M., Jordan J.M., Nevitt M.C., Slagboom E.P., Hart D.J, Lafeber F., Styrkarsdottir U., Zeggini E., Evangelou E., Spector T.D., Uitterlinden A.G., Lane N.E.†, Meulenbelt I.†, Valdes A.M.† & van Meurs J.B.J.†, **Novel Genetic Variants for Cartilage Thickness and Hip Osteoarthritis**, *PLoS Genet. 2016 Oct 4;12(10):e1006260. doi: 10.1371/journal.pgen.1006260. eCollection 2016 Oct.*

Boer C.G, Yau M.S, Rice S.J., Coutinho de Almeida R., Cheung K., Styrkarsdottir U., Southam L., Broer L., Wilkinson J.M., Uitterlinden A.G., Zeggini E., Felson D.T., Loughlin J., Young M., Capellini T., Meulenbelt I. & van Meurs J.B.J., **Hand phenotypes identify WNT9A as novel gene associated with thumb osteoarthritis**, *Accepted in Annals of Rheumatic Diseases (2020).*

8

den Hollander W.*, Boer C.G.*, Hart D.J., Yau M.S, Ramos Y.F.M., Metrustry S., Broer L., Deelen J., Cupples L.A., Rivadeneira F., Kloppenburg M., Peters M., Spector T.D., Hofman A., Slagboom P.E., Nelissen R.G.H.H., Uitterlinden A.G., Felson D.T., Valdes A.M., Meulenbelt I.† & van Meurs J.B.J.†, **Genome-wide association and functional studies identify a role for matrix Gla- protein in osteoarthritis of the hand**, *Ann Rheum Dis. 2017 Dec;76(12):2046-2053. doi: 10.1136/annrheumdis-2017-211214. Epub 2017 Aug 30.*

Boer C.G., Szilagyi I., Nguyen N.L., Neogi T., Meulenbelt I., Ikram M.A., Uitterlinden A.G., Bierma-Zeinstra S., Stricker B. & van Meurs J.B.J., **Vitamin K antagonist anticoagulant usage is associated with increased incidence and progression of osteoarthritis**, *Manuscript submitted to Annals of Rheumatic Diseases*

Boer C.G.*, Hatzikotoulas K.*, Southam L.*, Stefánsdóttir L., Zhang Y., Teder-Laving M., Coutinho de Almeida R., Zheng J., Hartley A., Wu T.T., Skogholt A., Terao C., Gabrielson M.E., Hveem K., Johnsen M.B., Nakajima M., Reimann E., Sanderson E., Sham P.C., Takuwa H., Thomas L, Winsvold B., Zengini E., Zwart J.A., Barysenka A., *[Genetics of Osteoarthritis Consortium], [arcOGEN Consortium]*, Tsezou A., Yin Cheung J.P., Mägi R., Thorleifsson G., Lotta L.A., Chan D., Cheah K., Smith G.D., Valdes A.M., Ikegawa S., Tobias J., Esko T., Gaunt T., Meulenbelt I., Styrkársdóttir U., Lee M.T.M. , van Meurs J.B.J. & Zeggini E. **Deciphering Osteoarthritis Genetics Across 826,690 Individuals from 9 populations,** *Manuscript submitted to Cell*

Radjabzadeh D., **Boer C.G.**, Beth S.A., van der Wal P., Kiefte-De Jong J.C., Jansen M.A.E., Konstantinov S.R., Peppelenbosch M.P., Hays J.P., Jaddoe V.W.V., Ikram M.A., Rivadeneira F., van Meurs J.B.J., Uitterlinden A.G., Medina-Gomez C., Moll H.A. & Kraaij R., **Diversity, compositional and functional differences between gut microbiota of children and adults**, *Sci Rep. 2020 Jan 23;10(1):1040. doi: 10.1038/s41598-020-57734-z.*

Boer C.G., Radjabzadeh D., Medina-Gomez C., Garmaeva S., Schiphof D., Arp P., Koet T., Kurilshikov A., Fu J., Ikram M.A., Bierma-Zeinstra S., Uitterlinden A.G., Kraaij R., Zhernakova A. & van Meurs J.B.J., **Intestinal microbiome composition and its relation to joint pain and inflammation**, *Nat Commun. 2019 Oct 25;10(1):4881. doi: 10.1038/s41467-019-12873-4.*

* These authors contributed equally as first authors
† These authors contributed equally as senior authors

Other Publications

Gourru-Lesimple G., Mathieu C., Thevenet T., Guillaume-Vasselin V., Jégou J-F., **Boer C.G.**, Tomczak K., Bloyet L-M., Giraud C., Grande S., Goujon C., Cornu C & Horvat B., **Measles Virus Infection of Human Keratinocytes: Possible Link Between Measles and Atopic Dermatitis**. *J Dermatol Sci. 2017 May;86(2):97-105. doi: 10.1016/j.jdermsci.2017.01.015. Epub 2017 Feb 10.*

Medina-Gomez C., Kemp J.P., Dimou N.L., Kreiner E., Chesi A., Zemel B.S., Bønnelykke K., **Boer C.G.**, Ahluwalia T.S., Bisgaard H., Evangelou E., Heppe D.H.M., Bonewald L.F., Gorski J.P., Ghanbari M., Demissie S., Duque G., Maurano M.T., Kiel D.P., Hsu Y.H., van der Eerden B.C.J., Ackert-Bicknell C., Reppe S., Gautvik K.M., Raastad T., Karasik D., van de Peppel J., Jaddoe V.W.V, Uitterlinden A.G., Tobias J.H., Grant S.F.A., Bagos P.G., Evans D.M. & Rivadeneira F., **Bivariate Genome-Wide Association Meta-Analysis of Pediatric Musculoskeletal Traits Reveals Pleiotropic Effects at the SREBF1/TOM1L2 Locus**. *Nat Commun. 2017 Jul 25;8(1):121. doi: 10.1038/s41467-017-00108-3.*

Medina-Gomez C., Kemp J.P., Trajanoska K., Luan J., Chesi A., Ahluwalia T.S., Mook-Kanamori D.O., Ham A, Hartwig F.P., Evans D.S., Joro R., Nedeljkovic I., Zheng H.F., Zhu K., Atalay M., Liu C.T., Nethander M., Broer L., Porleifsson G., Mullin B.H., Handelman s.K., Nalls M.A., Jessen L.E., Heppe D.H.M., Richards J.B., Wang C., Chawes B., Schraut K.E., Amin N., Wareham N., Karasik D., Van der Velde N., Ikram M.A., Zemel B.S., Zhou Y., Carlsson C.J., Liu Y., McGuigan F.E., **Boer C.G.**, Bønnelykke K., Ralston S.H., Robbins J.A., Walsh J.P., Zillikens M.C., Langenberg C., Li-Gao R., Williams F.M.K., Harris T.B., Akesson K., Jackson R.D., Sigurdsson G., den Heijer M., van der Eerden B.C.J., van de Peppel J., Spector T.D., Pennell C., Horta B.L., Felix J.F., Zhao J.H., Wilson S.G., de Mutsert R., Bisgaard H., Styrkársdóttir U, Jaddoe V.W., Orwoll E., Lakka T.A., Scott R, Grant S.F.A., Lorentzon M., van Duijn C.M., Wilson J.F., Stefansson K., Psaty B.M., Kiel D.P., Ohlsson C., Ntzani E., van Wijnen A.J., Forgetta V., Ghanbari M., Logan J.G., Williams G.R., Bassett J.H.D., Croucher P.I., Evangelou E., Uitterlinden A.G., Ackert-Bicknell C.L., Tobias J.H., Evans D.M. & Rivadeneira F., **Life-Course Genome-wide Association Study Meta-analysis of Total Body BMD and Assessment of Age-Specific Effects.** *Am J Hum Genet. 2018 Jan 4;102(1):88-102. doi: 10.1016/j.ajhg.2017.12.005***.**

Ghanbari M., Peters M.J., de Vries P.S., **Boer C.G.**, van Rooij J.G.J., Lee Y.C., Kumar V., Uitterlinden A.G., Ikram M.A., Wijmenga C., Ordovas J.M., Smith C.E., van Meurs J.B.J., Erkeland S.J., Franco O.H. & Dehghan A., **A Systematic Analysis Highlights Multiple Long Non-Coding RNAs Associated With Cardiometabolic Disorders**, *J Hum Genet. 2018 Apr;63(4):431-446. doi: 10.1038/s10038-017-0403-x. Epub 2018 Jan 30.*

Saberi Hosnijeh F., Kavousi M, **Boer C.G.**, Uitterlinden A.G., Hofman A., Reijman M., Oei E.H.G., Bierma-Zeinstra S.M. & van Meurs J.B.J., **Development of a Prediction Model for Future Risk of Radiographic Hip Osteoarthritis**, *Osteoarthritis Cartilage. 2018 Apr;26(4):540-546. doi: 10.1016/j.joca.2018.01.015. Epub 2018 Jan 31*.

Liu C., Marioni R.E, Hedman Å.K., Pfeiffer L, Tsai P-C., Reynolds L.M., Just A.C., Duan Q., **Boer C.G.**, Tanaka T., Elks C.E., Aslibekyan S., Brody A.J., Kühnel B., Herder C., Almli L.M., Zhi D., Wang Y., Huan T., Yao C., Mendelson M.M., Joehanes R., Liang L., Love S-A., Guan W., Shah S., McRae A.F., Kretschmer A., Prokisch H., Strauch K., Peters A., Visscher P.M., Wray N.R., Guo X., Wiggins K.L., Smith A.K., Binder E.B., Ressler K.J., Irvin M.R., Absher D.M., Hernandez D., Ferrucci L., Bandinelli S., Lohman K., Ding J., Trevisi L., Gustafsson S., Sandling J.H., Stolk L., Uitterlinden A.G., Yet I., Castillo-Fernandez J.E., Spector T.D., Schwartz J.D., Vokonas P., Lind L., Li Y., Fornage M., Arnett D.K., Wareham N.J. , Sotoodehnia N., Ong K.K., van Meurs J.B.J., Conneely N.K., Baccarelli A.A., Deary I.J., Bell J.T., North K.E., Liu Y., Waldenberger M., London S.J., Ingelsson E. & Levy D., **A DNA methylation biomarker of alcohol consumption**. *Mol Psychiatry. 2018 Feb;23(2):422-433. doi: 10.1038/mp.2016.192.*

Zengini E., Hatzikotoulas K., Tachmazidou I., Steinberg J., Hartwig F.P., Southam L., Hackinger S., **Boer C.G.**, Styrkarsdottir U., Gilly A., Suveges D., Killian B., Ingvarsson T., Jonsson H., Babis G.C., McCaskie A., Uitterlinden A.G., van Meurs J.B.J., Thorsteinsdottir U., Stefansson K., Smith G.D., Wilkinson J.M. & Zeggini E., **Genome-wide Analyses Using UK Biobank Data Provide Insights Into the Genetic Architecture of Osteoarthritis**, *Nat Genet. 2018 Apr;50(4):549-558. doi: 10.1038/s41588-018-0079-y. Epub 2018 Mar 20.*

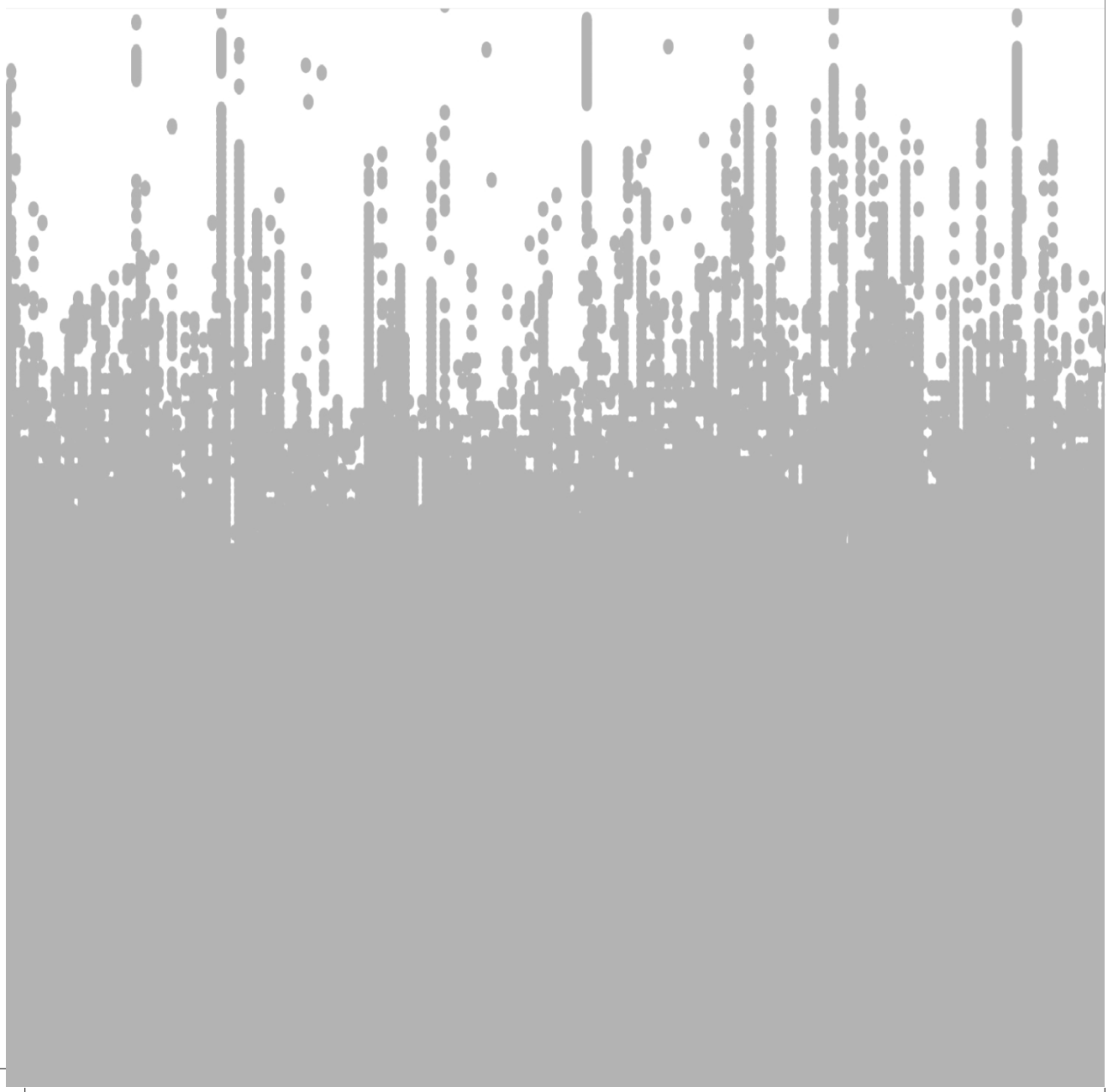
Warrington N.M., Shevroja E., Hemani G., Hysi P.G., Jiang Y., Auton A., **Boer C.G.**, Mangino M., Wang C.A., Kemp J.P., McMahon G., Medina-Gomez C., Hickey M, Trajanoska K, Wolke D., Ikram M.A, [*23andMe Research Team*], Montgomery G.W., Felix J.F., Wright M.J., Mackey D.A., Jaddoe V.W., Martin N.G., Tung J.Y., Smith G.D., Pennell C.E., Spector T.D., van Meurs J.B.J., Rivadeneira F., Medland S.E., & Evans D.M., **Genome-wide Association Study Identifies Nine Novel Loci for 2D:4D Finger Ratio, a Putative Retrospective Biomarker of Testosterone Exposure in Utero**, *Hum Mol Genet. 2018 Jun 1;27(11):2025-2038. doi: 10.1093/hmg/ddy121.*

Suri P., Palmer M.R., Tsepilov Y.A., Freidin M.B., **Boer C.G.**, Yau M.S., Evans D.S., Gelemanovic A., Bartz T.M., Nethander M., Arbeeva L, Karssen L., Neogi T., Campbell A., Mellstrom D., Ohlsson C., Marshall L.M., Orwoll E, Uitterlinden A.G., Rotter J.I., Lauc G., Psaty B.M., Karlsson M.K., Lane N.E., Jarvik G.P., Polasek O., Hochberg M, Jordan J.M., van Meurs

J.B.J., Jackson R., Nielson C.M., Mitchell B.D., Smith B.H., Hayward C., Smith N.L., Aulchenko Y.S. & Williams F.M.K., **Genome-wide Meta-Analysis of 158,000 Individuals of European Ancestry Identifies Three Loci Associated With Chronic Back Pain,** *PLoS Genet. 2018 Sep 27;14(9):e1007601. doi: 10.1371/journal.pgen. 1007601. eCollection 2018 Sep*

Styrkarsdottir U., StefanssonO.A., Gunnarsdottir K., Thorleifsson G., Lund S.H., Stefansdottir L., Juliusson K., Agustsdottir A.B., Zink F., Halldorsson G.H., Ivarsdottir E.V., Benonisdottir S., Jonsson H., Gylfason A., Norland K., Trajanoska K., **Boer C.G.**, Southam L., Leung J,C,S., Tang N.L.S., Kwok T.C.Y., Lee J.S.W., Ho S.C., Byrjalsen I., Center J.R., Lee S.H., Koh J.M., Lohmander L.S., Ho-Pham L.T., Nguyen T.V., Eisman J.A., Woo J., Leung P.C., Loughlin J., Zeggini E., Christiansen C., Rivadeneira F., van Meurs J.B.J., Uitterlinden A.G., Mogensen B., Jonsson H., Ingvarsson T., Sigurdsson G., Benediktsson R., Sulem P., Jonsdottir I., Masson G., Holm H., Norddahl G.L., Thorsteinsdottir U., Gudbjartsson D.F. & Stefansson K., **GWAS of bone size yields twelve loci that also affect height, BMD, osteoarthritis or fractures.** *Nat Commun. 2019 May 3;10(1):2054. doi: 10.1038/s41467-019-09860-0.*

Hollander W., Pulyakhina I., **Boer C.G.**, Bomer N., van der Breggen R., Arindrarto W., Couthino de Almeida R., Lakenberg N., Sentner T., Laros J.F.J., 't Hoen P.A.C., Slagboom E.P.E., Nelissen R.G.H.H., van Meurs J.B.J., Ramos Y.F.M., Meulenbelt I., **Annotating Transcriptional Effects of Genetic Variants in Disease-Relevant Tissue: Transcriptome-Wide Allelic Imbalance in Osteoarthritic Cartilage**, *Arthritis Rheumatol. 2019 Apr;71(4):561-570. doi: 10.1002/art.40748. Epub 2019 Feb 23.*



Dankwoord
Acknowledgments

By mutual confidence, and mutual aid,
Great deeds are done, and great discoveries made;
~ Iliad, Homer

Ik heb vaak genoeg gezegd dat het mij niet zoveel uitmaakte hoe lang ik over mijn PhD zou doen, zolang ik maar onderzoek kon doen!Toch ben ik nu wel erg blij dat mijn PhD nu af is en het boekje hier ligt! Maar, niets in de wetenschap doe je alleen, ook promoveren niet! Dus wil ik hier iedereen bedenken die, in welke vorm dan ook, heeft bijgedragen aan dit boekje!

Om te beginnen wil ik mijn fantastische promotor en co-promotor bedanken, **André en Joyce**, zonder jullie zou ik niet de wetenschapper zijn die ik nu ben. Ik wil jullie beiden van uit de grond van mijn hart bedanken voor jullie ontzettende steun en vertrouwen in mij, hoe jullie klaar staan om mij te helpen ook op persoonlijk vlak.

Beste **Joyce**, Ik had geen betere co-promotor kunnen wensen dan jou, en ik kan me ook geen andere co-promotor voorstellen. Vanaf de eerste dag, heb jij mij de kansen en ruimte gegeven om mijzelf te ontwikkelen. Jij hebt altijd naar mijn (soms wilde) ideeën geluisterd, mijn (maffe) uitspattingen toegestaan (van flauwe presentatie grappen tot kerstkaarten) en ik kijk er dan ook erg naar uit om ook het volgende avontuur met jou aan te gaan: mijn post-doc!

Beste **André**, ik denk nog steeds dat onze gedeelde voorliefde voor de Kantonese keuken de doorslag is geweest om mij aan te nemen als PhD student. En gelukkig maar! Jouw passie voor de wetenschap is enorm aanstekelijk, en dankzij jou aandringen op replicatie, replicatie, en replicatie heb ik niet alleen goed gerepliceerde resultaten gekregen maar ook zoveel mooie contacten en samenwerkingen. Dus, bedankt voor al jouw steun, vertrouwen en wijze lessen (of goede film aanbevelingen. i.e., *GATTACA*).

Ook wil ik mijn (lees)commissie leden bedanken, I also would like to thank my thesis committee members: **Prof. dr. Gerjo van Osch, Prof. dr. Eleftheria Zeggini, Prof.dr. Arfan Ikram, Prof. dr. Ingrid Meulenbelt, Prof. dr. Sita Bierma-Zeinstra and dr. Unnur Styrkarsdottir.**

Beste Gerjo, jij bent een van mijn eerste samenwerkingscontacten geweest in het Erasmus MC. Dat maakt het voor mij extra leuk dat jij nu ook in mijn lees commissie zit. Dankzij jou, de fantastische postdocs en (phd-studenten) in jouw groep heb ik veel geleerd (en blijf ik leren) over alle kraakbeen en OA biologie. Ik kijk uit naar toekomstige samenwerkingen!

Dear Ele, I feel so privileged to have worked with you and even more privileged to have you in my committee. When I first saw you present your massive OA GWAS results on the OARSI, I could not have imaged that I would be working with you on an even bigger GWAS. It has been such a pleasure and informative experience to have worked with you. Thank you so much for welcoming me in you group in Cambridge and Munich. I felt so very welcome and I look forward to future collaborations with you!

Beste **Professor Arfan Ikram**, ik voel me vereerd dat u onderdeel bent van mijn leescommissie, ik kijk er naar uit om met u over mijn werk te discussiëren (of over de mooie Rotterdam Study).

Beste **Ingrid,** vanaf het begin van mij PhD hebben wij heel veel samen gewerkt, ik heb zelfs nog even op een blauwe maandag bij jou in het lab gestaan! En wat een super samenwerking is dit altijd geweest en blijft dit zijn. Ik kijk erg uit naar toekomstige projecten, zeker waar we gaan eindigen met *MGP!*

Dear **Unnur**, your GWAs paper on severe hand osteoarthritis where you found *ALDH1A2* as a novel hand OA gene, was my very first OA GWAS paper I've ever read. It was an immediate inspiration for me. Thus for me, it is really quite special that you are now part of my thesis committee, thank you.

Beste **Sita**, Ik vind het super leuk dat jij deelt uitmaakt van mijn promotie commissie. Ik vond en vind het altijd super fijn om met jou samen te werken, ik heb echt veel van jou werk geleerd over epidemiologie en artrose. Ik hoop dat we nog veel kunnen samen werken en met name over vitamine K!

Mijn paranimfen, **Justin en Joost**, wil ik graag bedanken voor al jullie hulp en steun tijdens mijn PhD, en met name tijdens de laatste lootjes.

Mijn grote broertje, **Justin**, ik vind het heel speciaal dat jij mijn paranimf bent. Ik vind het zo ontzettend gaaf en leuk dat ik met jou samen hebt gewerkt, dat ik met jou kan sparren over mijn onderzoek aan de eettafel (ook al worden onze ouders daar soms helemaal knetter gek van). Dus op naar die *Boer and Boer* et al., publicatie!

Beste **Joost**, ik wist zeker al 4 jaar dat ik jou als paranimf wilde, en volgens mij wist jij dat ook net zo lang (ik heb geen pokerface). Jouw IT hulp heeft menig analyse gered! En ook al ben jij beroemd om de uitspraak "wij zijn collega's geen vrienden", voor mij ben jij zeker een goede vriend. Onze koffie pauzes waren niet alleen gezellig, maar ze waren zeker ook het gene wat mij door mijn Phd heen geholpen heeft.

Natuurlijk, wil ook alle collega's (en vrienden) van het genetisch lab bedanken, jullie hebben allemaal bijgedragen aan mijn boekje, in data, onderzoek, hulp of gewoon gezelligheid.

Allereerst wil ik mijn kamer genoten bedanken: Beste **Linda**, vanaf de eerste dag dat jij bij mij op het kantoor kwam werken zat het gelijk goed (en was het gelijk feest). Jij bent niet alleen een fantastische wetenschapper, collega, docent, mens, maar ook een goede vriendin. Ik kan altijd bij jou terecht met mijn vragen, zorgen en blijdschappen. En ik zou half de genetische kennis niet hebben dankzij jouw. Ik wordt altijd blij als ik het kantoor inloop en jij er ook bent! Dear **Vid**, ik ben zo blij dat jij naast mij in het kantoor bent komen zitten. I love our scientific discussion which start so serious ("quick question") but end-up so completely absolutely insane, that most (fictional) "mad" scientists look completely sane. **Carolina** "beauty" Medine-Gomez, you I have always learned a lot from you (including but not limited to: "colourful" Spanish words) and hope to continue to do so. I have always enjoyed your our time together.

Beste **Mila, Pascal, Michael, Ramazan, Pelle, Thomas, Pawan, Marijn, Zara, Gaby**, **Costanza en Yard**, zonder jullie continue werk zouden er geen datasets zijn voor mij om mee te werken, zonder jullie constante inzet in het lab zou hier geen boekje hebben gelegen. En niet alleen was jullie hulp en werk onmisbaar, ook jullie gezelligheid en collegialiteit! Beste **Pascal, Thomas en Zara**, en zonder jullie hulp had ik ~1600 qPCRs zelf moeten doen (en laten we eerlijk zijn, dat was drama geweest). Beste **Pelle Ramazan**, en **Yard** jullie laten me altijd lachen, jullie hebben altijd de beste verhalen, stunts en (photoshop) fotos (Ramazan, jij hebt voor mij de beste skills hier). Maar vooral ben ik jullie dankbaar voor de dozen die voor mij bewaard worden en dan klaar staan bij mijn kantoor!

Beste **Michael**, beste Mike, jij maakt de sfeer! Ik vergeet nooit het moment dat jij mij vroeg: "Ow dus je vind Star Trek leuk? Trek je dan ook wel eens van die outfits aan?". Dit was hat moment, dat ik mij thuis voelde op onze afdeling, Toen, kon ik jouw vraag met nee beantwoorden en samen met je lachen, maar inmiddels hebben wij samen wel vreemdere outfits aangehad. Ik kijk er dan ook (met angst) naar uit wat wij in de toekomst gaan verzinnen. En voor ik het vergeet: Hey Mike, alles pluis?,

Beste **Eline**, super bedankt voor al je hulp! Zeker in deze laatste stressvolle tijd,. jij bent echt super Eline!

To all other (ex) PhD candidates, who are on their own PhD quests or have started a new adventure. All of your discussions and talks (scientific or otherwise) have contributed to my PhD: **Bahar** (You are a true Rockstar), **Ingrid** (looking forward to more OARSI conference fun together!), **Marjolein de Kruif** (een fantasche eerste OARSI de geflambeerde pannekoeken zijn nog steeds niet overtroffen), **Marjolein Peters** (wat was ik zenuwachtig tijdens mijn sollicitatie, dat jij daar was heeft heel veel geholpen!), Martha (For helping me embark on my osteoarthritis quest), **Pooja** (my fellow horror movie enthusiast), **Jeroen** (altijd feest bij Illumina), **Sergio** (vind het nog steeds een eer om jou mentor te zijn), **Lieke** (ik kijk uit naar jou "presentatie" word grappen). **Elizabeth**, **Jinluan, Vivi, Ariadne, Komal, Denise en Annlies** (It was always so nice and fun to talk you all in the short time before the workdiscussions and on the social lab outings!). "The girls": **Enisa and Olja** (the time we went to "pop" some bubble-wrap!), **Kate** (for the best movie/series/book tips), **Fjorda** (you are such a kind person). **Ling** (het was super leuk om samen met jou paranimf te zijn). **Djawad** (success in je Postdoc!), **Natialia** (I liked our end of the day talks in the office), **Ruolin**, **Artemis, Zografia, Tracey and Sam,** (I hope to get to know you all more!),

I also want to thank all of our senior members and professors in our group! **Carola** (bedankt voor de altijd, super interessante klinische blik op mijn onderzoek)**, Fernando** (Bedankt voor het altijd stellen van goede en kritische vragen op mijn onderzoek)**, Annemieke** (Jij bent zo ontzettend creatief, vind het altijd gaaf als je een van je papier creaties laat zien), **Robert** (Dank voor al jouw werk met het maken van de microbiome data).

I want to thank all of collaborators, I have enjoyed working with each of you and have learned something new for every one of you! In particular I want to thank **Lorraine Southam** and **Konstantinos Hatzikotoulas**. Dear Loz and Kostas, I am so happy to be able to had the opportunity to work with you, and even more so to get to know you. Thank you so much for making mee feel at home in Cambridge and Munich. I am looking forward to have a beer with you both again (and with Fin), although I don't know a place which could match to your beer fridges Loz and Finn!

To all of the MolEpi meeting group: I thank all of you for your presentations, discussions, stories and friendship. I love our meeting, I love our interdisciplinary knowledge exchange, and how each and all of you help each other, and have helped me. I feel that these meetings have let me grow as a scientist and as a person. Met deze wil ik jou **Janine** daarom ook persoonlijk bedanken, ik voelde me, en voel me nog steeds vereerd om samen met jou deze meeting te leiden.

No quest would be complete without some good friends and brave warriors (Nerds & Geeks) by my side: alle vrienden en vriendinnen met wie ik mijn hoogte punten en (diepte punten) van mij PhD samen heb en mocht delen, ook dankzij jullie persoonlijk steun aan mij ligt deze PhD-thesis er. Aan alle Zwaard en Steen-ers (& Hollandse Bastaarden): bedankt dat ik altijd mijn (werk) frustraties op jullie mocht uitslaan met een zwaard en ik dat ik deze slagen (vaker) terug kreeg.

Als laatste, wil ik mijn familie bedanken. Ook zonder jullie steun, liefde en vertrouwen was dit boekje er niet geweest.

Beste **Marissa**, beste (schoon) zus, ik kijk altijd uit naar de zondagen om gezellig met jou en Justin (bord)spellen te spelen om weer een beetje te ont-stressen van mijn PhD

Lieve **papa en mama**, ik kan me geen betere ouders voorstellen dan jullie. Jullie zijn er altijd voor mij geweest, mij altijd zo ontzettend gesteund en gestimuleerd in alles! Ik kon en mocht altijd mijzelf zijn. Zonder jullie hulp, steun en vertrouwen was ik nooit zo ver gekomen. Ik hou van jullie.

Oma Mieke, u bent mijn persoonlijke heldin en voorbeeld. U heeftt mij altijd gesteund in mijn zoektocht naar kennis, Hoe ik u altijd de laatste weetjes die ik had geleerd kon vertellen, en u altijd aandachtig luisterde en vragen stelde. Hierin heeft u mij altijd gestimuleerd om meer kennis op doen. Maar, nog meer dan dit, heeft u mij geleerd en laten zien wat belangrijk is in het leven, wat er allemaal wel kan.

Mijn liefste **Jeroen**, jij bent echt een steun voor mij geweest tijdens deze laatste lootjes. Jij bent mij rust, mijn liefde en maakt mij compleet, ik hou van je.

To all of you, friends, family, co-workers, and collaborators, whom I have mentioned. Not mentioned or have forgotten to mention; all of you have made my PhD quest the best one! Thank you all, and I look forward to more adventures and quests with all of you.

Aan jullie allen, vrienden, familie, collega's, die ik heb genoemd, niet genoemd, of ver-geten ben te noemen: Dankzij jullie was mijn PhD een geweldig avontuur! Iedereen be-dankt, en ik kijk er naar uit om nog meer avonturen met jullie allemaal te beleven.

Cindy G. Boer