Technische Universität München

TUM School of Medicine and Health

Profiling DNA methylation in primary osteoarthritis tissues

Peter Kreitmaier

Vollständiger Abdruck der von der TUM School of Medicine and Health der Technischen

Universität München zur Erlangung eines Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Radu Roland Rad

Prüfende der Dissertation:

    1.   Prof. Dr. Eleftheria Zeggini
    2.   Prof. Dr. Julien Gagneur

Die Dissertation wurde am 14.08.2024 bei der Technischen Universität München eingereicht

und durch die TUM School of Medicine and Health am 04.12.2024 angenommen.

Abstract

Osteoarthritis is a complex joint disease, caused by interplay between environmental and genetic factors, that affects more than 300 million people worldwide. Its impact on public health systems is estimated to increase further due to aging populations. Current treatment methods are limited to pain management and joint replacement, underlining the need to develop novel, personalised treatment strategies. Thus, it is important to extend our knowledge of the genomic basis of osteoarthritis.

Here, we examined DNA methylation profiles of four primary tissue types, namely primary, macroscopically intact (low-grade) and degraded (high-grade) cartilage, synovium as well as infrapatellar fat pad tissue in 170 patients, who underwent total knee replacement due to late-stage osteoarthritis.

We performed epigenome-wide association studies (EWAS) to compare the DNA methylation profiles between low-grade and high-grade osteoarthritis cartilage and revealed wide-spread epigenetic markers of cartilage degeneration (146,777 differentially methylated sites). We further built a prediction model that could distinguish low-grade and high-grade osteoarthritis cartilage samples with high accuracy (mean accuracy: 90.69%, standard deviation: 4.08), which we could validate in an external validation set (82.35% accuracy).

By combining methylation profiles with matched genotype data from the same patients, we generated the first genome-wide cis methylation quantitative trait locus (mQTL) maps of low- and high-grade osteoarthritis cartilage, synovium, and infrapatellar fat pad tissue and identified QTL-targeted methylation sites (low-grade cartilage: 73,836, high-grade cartilage: 52,819, synovium: 40,361, fat pad: 35,948, FDR < 0.05).

We further integrated mQTL maps of the four osteoarthritis primary tissue types with large genome-wide association study (GWAS) results for osteoarthritis using colocalisation and causal inference analysis to investigate causal roles of methylation in osteoarthritis and resolve osteoarthritis GWAS signals.

Together, this work presents the largest EWAS for cartilage degeneration and the first genome-wide mQTL maps in cartilage, synovium and fat pad. We reveal robust epigenetic markers of cartilage degeneration and propose a validated model that can distinguish cartilage

samples in different osteoarthritis stages. We further highlight the causal roles of epigenetic mechanisms across osteoarthritis tissues. Together, our results enhance insights into the epigenetics underlying osteoarthritis.

Zusammenfassung

Arthrose ist eine komplexe Gelenkskrankheit, welche durch das Zusammenspiel von Umwelt- und genetischen Faktoren verursacht wird und mehr als 300 Millionen Menschen weltweit betrifft. Es wird angenommen, dass seine Auswirkungen auf öffentliche Gesundheitssysteme aufgrund der alternden Bevölkerung weiter zunehmen wird. Aktuelle Behandlungsmöglichkeiten sind auf Schmerztherapie und Gelenksersatz begrenzt, welches die Notwendigkeit unterstreicht, neuartige und personalisierte Therapiestrategien zu entwickeln. Hierfür ist es wichtig, unser Wissen über die genomische Basis von Arthrose zu erweitern.

Hier untersuchen wir DNA Methylierungsprofile von vier primären Gewebsarten, nämlich primäres, mikroskopisch intaktes (niedriger Arthrosegrad) und degradiertes (hoher Arthrosegrad) Knorpelgewebe, der Synovialmembran wie auch dem infrapatellaren Fettpolster in bis zu 170 Patienten, die sich aufgrund von Arthrose im Spätstadium einem totalen Knieersatz unterzogen haben.

Wir haben epigenomweite Assoziationsstudien (EWAS) durchgeführt, um die DNA-Methylierungsprofile von Knorpel mit geringgradiger und hochgradiger Arthrose zu vergleichen, und dabei weit verbreitete epigenetische Marker der Knorpeldegeneration (146.777 differentiell methylierte Stellen) festgestellt. Darüber hinaus erstellten wir ein Vorhersagemodell, das Knorpelproben mit geringgradiger und hochgradiger Arthrose mit hoher Genauigkeit unterscheiden konnte (mittlere Genauigkeit: 90,69 %, Standardabweichung: 4,08), was wir in einem externen Validierungsset replizieren konnten (82,35 % Genauigkeit).

Durch die Kombination von Methylierungsprofilen mit Genotypdaten derselben Patienten erstellten wir die ersten genomweiten cis-Methylierungskarten (mQTL) von Knorpel mit niedrig und hochgradiger Arthrose, Synovium und infrapatellarem Fettpolstergewebe und identifizierten QTL-gerichtete Methylierungsstellen (niedriggradiger Knorpel: 73.836, hochgradiger Knorpel: 52.819, Synovium: 40.361, Fettpolster: 35.948, FDR < 0,05).

Darüber hinaus haben wir mQTL-Karten der vier primären Gewebetypen für Arthrose mit den Ergebnissen großer genomweiter Assoziationsstudien (GWAS) für Arthrose integriert, indem

wir Kolokalisierungs- und kausale Inferenzanalysen durchgeführt haben, um die kausale Rolle der Methylierung bei Arthrose zu untersuchen und GWAS-Signale für Arthrose aufzulösen.

Zusammengenommen stellt diese Arbeit die größte EWAS für Knorpeldegeneration und die ersten genomweiten mQTL-Karten in Knorpel, Synovium und Fettpolster vor. Wir zeigen robuste epigenetische Marker für die Knorpeldegeneration auf und schlagen ein validiertes Modell vor, mit dem Knorpelproben in verschiedenen Arthrosestadien unterschieden werden können. Darüber hinaus heben wir die kausale Rolle epigenetischer Mechanismen in verschiedenen Arthrose-Geweben hervor. Zusammengenommen verbessern unsere Ergebnisse den Einblick in die Epigenetik, die der Arthrose zugrunde liegt.

Acknowledgements

Thank you to…

…Ele for your guidance, encouragement, and constant support in the last few years! Thank you for enabling all these opportunities and that I could work on these projects with you.

…my TAC members Caroline Relton and Julien Gagneur for their feedback in our committee meetings.

…Loz and Kostas for being wonderful office mates! Thank you for your companionship and for sharing your knowledge and experience.

…all previous and current Zeggini group members (Georgia, Grace, Xenofon, Ana, Ozvan, Archit, Shibo, Yue, Odysseas, Arthur, Mauro, Norbert, Jenny, Ope, Will, Andrei, YC, and all others) for great companionship and support.

…the ITG ops team (Iris, Anne, Bahar and others) for helping me get all the administrative work done.

…the Cai, Kim-Hellmuth, and Wang-Sattler group members for our lunch breaks, sharing ideas, and contributing to a positive work environment.

…all collaborators (including Mark Wilkinson, Diane Swift, Caroline Relton, Matthew Suderman, Ingrid Meulenbelt, Rodrigo Coutinho de Almeida, and Julia Steinberg) for their valuable input.

…my previous supervisors (Lukas Simon, Tania Carillo-Roa, Fabian Theis, Elisabeth Binder, Martin Preusse, Nikola Mueller), who made it possible for me to gain my first experience in the academic world before starting my PhD.

….my family and friends for their constant support!

Contents

1. List of publications

1.1. Publications contributing to this thesis

1.1.1. **Kreitmaier P**, Suderman M, Southam L, Coutinho de Almeida R, Hatzikotoulas K, Meulenbelt I, Steinberg J, Relton CL, Wilkinson JM*, Zeggini E*. An epigenome-wide view of osteoarthritis in primary tissues. Am J Hum Genet. 2022 Jul 7;109(7):1255-1271. doi: 10.1016/j.ajhg.2022.05.010. Epub 2022 Jun 8. PMID: 35679866; PMCID: PMC9300761.

1.1.2. **Kreitmaier P**, Park YC, Swift D, Gilly A, Wilkinson JM*, Zeggini E*. Epigenomic profiling of the infrapatellar fat pad in osteoarthritis. Hum Mol Genet. 2024 Feb 28;33(6):501-509. doi: 10.1093/hmg/ddad198. PMID: 37975894

1.1.3. **Kreitmaier P**, Swift D, Wilkinson JM*, Zeggini E*. Epigenomic differences between osteoarthritis grades in primary cartilage. Osteoarthritis Cartilage. 2024 Jul 23:S1063-4584(24)01314-1. doi: 10.1016/j.joca.2024.07.008. Epub ahead of print. PMID: 39053729

1.2. Other publications

1.2.1. Katsoula G, **Kreitmaier P**, Zeggini E. Insights into the molecular landscape of osteoarthritis in human tissues. Curr Opin Rheumatol. 2022 Jan 1;34(1):79-90. Doi: 10.1097/BOR.0000000000000853. PMID: 34750308; PMCID: PMC8635257.

1.2.2. **Kreitmaier P**, Katsoula G, Zeggini E. Insights from multi-omics integration in complex disease primary tissues. Trends Genet. 2023 Jan;39(1):46-58. Doi: 10.1016/j.tig.2022.08.005. Epub 2022 Sep 19. PMID: 36137835.

1.2.3. Penner-Goeke S, Bothe M, Rek N, **Kreitmaier P**, Pöhlchen D, Kühnel A, Glaser LV, Kaya E, Krontira AC, Röh S, Czamara D, Ködel M, Monteserin-Garcia J, Diener L, Wölfel B, Sauer S, Rummel C, Riesenberg S, Arloth-Knauer J, Ziller M, Labeur M, Meijsing S, Binder EB. High-throughput screening of glucocorticoid-induced enhancer activity reveals mechanisms of stress-related psychiatric disorders. Proc Natl Acad Sci U S A. 2023 Dec 5;120(49):e2305773120. Doi: 10.1073/pnas.2305773120. Epub 2023 Nov 27. PMID: 38011552; PMCID: PMC10710077.

1.2.4. Bittner N, Shi C, Zhao D, Ding J, Southam L, Swift D, **Kreitmaier P**, Tutino M, Stergiou O, Cheung JTS, Katsoula G, Hankinson J, Wilkinson JM, Orozco G, Zeggini E. Primary osteoarthritis chondrocyte map of chromatin conformation reveals novel candidate effector genes. Ann Rheum Dis. 2024 Mar 13:ard-2023-224945. Doi: 10.1136/ard-2023-224945. Epub ahead of print. PMID: 38479789.

*equal contribution

2. Introduction

Osteoarthritis is a complex and degenerative joint disorder that affects all tissues of diarthrodial joints. Osteoarthritis is highly prevalent, affecting more than 300 million people worldwide[1]. Epidemiological studies further observed a multitude of potential risk factors, including sex (women tend to develop osteoarthritis more frequently and with higher severity, such as more pain)[2], age (osteoarthritis affects older people, e.g. 40 % of people over 70)[3] and genetics[4].

Despite its high prevalence, no curative therapies are known, and treatment possibilities are limited to pain management and joint replacements. Osteoarthritis constitutes a challenge for public health systems, and its burden may increase further due to aging populations[1].

The high prevalence of osteoarthritis, combined with limited treatment possibilities, emphasise the urgent need for improved preventive measures as well as novel, personalised treatment possibilities for which an enhanced understanding of the genetic and genomic architecture of osteoarthritis is required.

Joint tissue changes during osteoarthritis

Osteoarthritis affects all tissues of diarthrodial joints, including the cartilage, synovium, and infrapatellar fat pad. In diarthrodial joints, the cartilage is located on the surface of adjacent bones[5] and enables smooth joint movements as well as absorbing and distributing compressive load[6]. Healthy cartilage is avascular and aneural, with chondrocytes being the only cell type. During osteoarthritis, affected cartilage is traversed by nerves and blood vessels. Healthy cartilage has a smooth surface, but develops fissures during osteoarthritis, ultimately resulting in its degeneration, which is the most prominent feature of the disease.

The synovium is a connective tissue that lines the joint capsule and separates the synovial cavity from neighboring tissues[7]. Synovium of osteoarthritis-affected joints show increased inflammation levels, referred to as synovitis[8].

The infrapatellar fat pad is an adipocyte-rich tissue located inferior to the patella in the anterior part of the knee joint[9]. It protects and stabilises other knee components when the joint is exposed to mechanical forces during movement. Osteoarthritis-affected infrapatellar fat pad undergoes disease-related alterations, including fibrosis, inflammation, and vascularisation. The infrapatellar fat pad is highly innervated and therefore a major source for osteoarthritis-related pain, which emphasizes its clinical relevance.

Together, osteoarthritis leads to a multitude of changes in joint tissues. Studying its genetics and genomics can reveal biological mechanisms that underlie these tissue changes.

Genetic architecture of osteoarthritis

Genome-wide association studies (GWAS) have provided valuable insights into the complex genetic architecture of osteoarthritis[4]. Boer et al. performed the largest GWAS meta-analysis to date, across 826,690 individuals (including 177,517 osteoarthritis cases) for 11 osteoarthritis traits. They identified 100 independent risk variants, of which 52 are novel, and further highlighted different genetic effects between weight- and non-weight bearing joints as well as between sexes.

Although GWAS revealed relevant parts of the polygenic architecture of osteoarthritis, its signals can be hard to interpret (such as to estimate the effector genes), because genetic risk variants mostly reside in non-coding regions[10]. Therefore, combining genetic results with molecular profiles of affected patient tissues can extend our knowledge beyond genetics, by

identifying the tissue- or context-specific molecular mechanisms of osteoarthritis, e.g. resolve the effector gene through which genetic risk loci exert osteoarthritis-promoting effects.

Advancement in technologies enable the generation of molecular profiles of human tissues across several levels at the genome-wide scale. Large consortia such as GTEx[11], Roadmap[12], ENCODE[13] and the Human Cell Atlas[14] have made these publicly available, thus providing established resources for human tissue-specific molecular maps. However, these exclude osteoarthritis-affected joint tissues.

Transcriptomics and proteomic profiles of osteoarthritis tissues

A number of studies have generated transcriptomic and proteomic profiles of osteoarthritis[15,16]. Steinberg et al.[17] combined genotype with transcriptomics and proteomics data of three joint tissue types from 115 patients that underwent total joint replacement (hip or knee) due to late-stage osteoarthritis: macroscopically intact (low-grade) and degraded (high-grade) osteoarthritis cartilage, as well as from synovial tissue. For each tissue, they associated genetic variants with expression levels of genes and proteins in cis (i.e., within the vicinity of the genetic variant), thus resulting in the first genome-wide expression quantitative trait locus (eQTL) and protein quantitative trait locus (pQTL) maps in these tissues. Integrating these maps with GWAS results using colocalisation revealed five genes (*ALDH1A2, NPC1, SMAD3, FAM53A, and SLC44A*) through which genetic loci might exert their osteoarthritis-promoting effect, so called effector genes. Furthermore, the authors carried out differential analyses of molecular profiles between low- and high-grade osteoarthritis cartilage, thus identifying 409 genes[17] that were differentially expressed on the transcriptomics and proteomics level. In a follow-up analysis, these 409 genes were integrated in a candidate

therapeutic compound analyses which revealed 19 compounds that could reverse osteoarthritis-progression related expression changes.

Katsoula et al. generated enhanced, better-powered transcriptomic profiles of cartilage degeneration by comparing low- and high-grade osteoarthritis cartilage profiles of 124 patients[18]. They reported 365 genes, of which 38 could be replicated in an external validation set, and further revealed 33 osteoarthritis-related long non-coding RNA genes, of which 25 were linked to osteoarthritis for the first time. Furthermore, they studied the osteoarthritis-related cartilage spliceosome by investigating differential transcript usage (identification of 89 osteoarthritis-linked isoforms, corresponding to 82 genes) and performing the first genome-wide differential splicing scan (209 differentially spliced genes between low- and high-grade osteoarthritis cartilage).

Steinberg et al.[19] used transcriptomics profiles from primary cartilage (low- and high-grade osteoarthritis cartilage) and synovium samples to investigate molecular subtypes among 113 osteoarthritis patients. They identified two osteoarthritis patient subgroups in the synovium and related subgroup differences to inflammation, extracellular matrix and to cell adhesion. Similarly, they found two patient clusters in low-grade osteoarthritis cartilage with gene expression differences being linked with inflammation, extracellular matrix-related and cell adhesion pathways. The high-inflammation subgroup showed an overrepresentation of women. They further constructed a seven-gene based classifier that can distinguish between low-grade cartilage subgroups, which was replicated in an external data set.

Coutinho de Almeida et al. investigated interactions between mRNA and miRNA in 63 patients[20]. They reported widespread transcriptomic differences between low- and high-

grade osteoarthritis patients and further combined mRNA and miRNA data to generate maps of the interactome between these transcriptome layers.

Other studies have investigated transcriptomic and proteomic profiles at the single-cell level [21–26]. For example, Wang et al.[22] generated single-cell transcriptomics data from 5 osteoarthritis patients, 5 Kashin–Beck patients and 5 healthy controls and particularly highlighted a chondrocyte subpopulation, namely mitochondrial chondrocytes, which were found in diseased cartilage samples (osteoarthritis and Kashin–Beck affected cartilage), but not in healthy tissue, which suggest mitochondrial dysfunction in osteoarthritis and Kashin–Beck disease. Grandi et al.[24] profiled proteomics of cartilage samples from 20 osteoarthritis patients as well as five normal samples at the single cell level. They reported 20 cell subpopulations in cartilage and observed three patient clusters based on relative proportions of cell subpopulations.

Epigenetic profiles of osteoarthritis tissues

Gene and protein expression is, in part, regulated by epigenetic processes. Thus, investigating the epigenetic maps in relevant tissues can extend our understanding beyond genetics, transcriptomics and proteomics. A small number of studies have investigated epigenetic profiles in osteoarthritis-relevant tissues at the genome-wide scale. Liu et. al[27] applied Assay for Transposase-Accessible Chromatin with high throughput sequencing (ATAC-seq) to generate a genome-wide open chromatin landscape in primary knee cartilage of osteoarthritis patients. They reported 109,215 accessible chromatin regions and, by integrating genotype, DNA methylation and gene expression data, highlighted osteoarthritis-relevant enhancer regions and enrichments in relevant pathways, such as ossification or mesenchymal stem cell differentiation.

Other efforts have performed whole genome chromosome conformation (Hi-C) analyses to study the three-dimensional chromosome organization in chondrocytes[28,29]. Bittner et al.[28] studied affected cartilage samples to generate the first Hi-C map in primary chondrocytes from eight osteoarthritis patients. They reported 345 genetic variants in chromatin loop anchors to be linked with 77 osteoarthritis GWAS risk signals. These signals included active enhancer-promoter loops which pointed to two candidate effector genes for osteoarthritis (*SPRY4* and *PAPPA*).

DNA methylation in osteoarthritis tissues

A highly relevant epigenetic mark is DNA methylation, which refers to the covalent binding of a methyl group to the DNA. In eukaryotes, this epigenetic modification primarily occurs at cytosine bases followed by a guanine (Cpgs sites) in the DNA strand[30]. The addition of a methyl group to the cytosine and its maintenance is performed by specialised enzymes, referred to as DNA methyltransferases[31]. The removal of this methyl group, DNA demethylation, is primarily regulated through TET methylcytosine dioxygenases, which oxidise cytosine-attached methyl-groups[32]. This can lead to demethylation during DNA replication (passive demethylation: oxdised methyl group is not recognised by the DNA methylation maintenance complex) or through DNA replication independent processes (active demethylation: thymine DNA glycosylase involving process in which a cytosine with an oxidised methyl group is actively replaced by an unmethylated cytosine)[32].

Higher methylation levels at sites in promoter regions and particularly around the transcription start site are linked to downregulated gene expression, whereas methylation effects in gene bodies or other regulatory regions remain less predictable. Current array-based technologies measure up to ~ 900,000 methylation sites [33–35], thus enabling genome-wide

14

studies of DNA methylation profiles in a cost- and time-efficient manner. DNA methylation is tissue-specific and interacts with internal and external factors (such as genetics, environment, including lifestyle, etc.), highlighting its relevance to better understand the molecular basis of complex diseases[36].

DNA methylation profiles of primary, osteoarthritis-affected joint tissue types have not been included in large consortia (such as GTEx[37], Roadmap[12], ENCODE[13]), underlining the need for additional, joint tissue-specific molecular studies.

A number of DNA methylation studies in osteoarthritis have focused on genomic regions around osteoarthritis risk variants to better understand associations between genetic variants and methylation sites[38–40]. Rice et al. [40] collected samples from osteoarthritis cartilage from patients that underwent total joint replacement due to knee osteoarthritis (n = 57), hip osteoarthritis (n = 14) or neck-of-femur (NOF) fracture (n = 16). By combining genotype and methylation data, they identified effects of genetic variants on methylation in 10 out of 42 tested osteoarthritis risk loci. Follow-up analyses, such as integration of chromatin state and enhancer data, revealed putative target genes of the identified methylation-associated variants, including *COLGALT2*, *COL11A2* and *WWP2*.

Other studies have characterised genome-wide DNA methylation profiles of primary cartilage. A common approach is to compare methylation profiles between low- and high-grade osteoarthritis cartilage samples to identify methylation sites that are associated with cartilage degeneration, as epigenetic markers of osteoarthritis progression in the cartilage [41–45].

Zhang et al.[42] examined DNA methylation data from 12 knee osteoarthritis patients that underwent total joint replacement. They generated data for three different cartilage regions

with varying osteoarthritis-related degradation: (1) Outer region of the lateral tibial plateau (oLT) with a visibly smooth cartilage surface (early osteoarthritis stage), (2) the inner region of lateral tibial plateau (iLT) with sufficient cartilage to detect visible fissures on tissue cross-section (intermediate osteoarthritis stage) and (3) the inner region of medial tibial plateau (iMT) with visible loss of articular cartilage (late osteoarthritis stage). The authors highlighted 519 differentially methylated sites when comparing iMT vs. oLT cartilage samples, thus suggesting methylation differences between early and late-stage osteoarthritis.

Den Hollander et al.[44] combined DNA methylation and gene expression data from 31 osteoarthritis-affected individuals (17 knee and 14 hip osteoarthritis patients). They found 5,282 cartilage-degeneration related methylation sites, 87 of which were significantly correlated with 70 cartilage degeneration-related gene, thus highlighting their link with cartilage degeneration across these two molecular levels.

A few other studies have extended to non-cartilage joint tissues and investigated specific candidate genes or regions [46–50]. For example, Kehayova et al.[46] investigated the effect of the genetic variants rs11583641 and rs1046934 on *COLGALT2* methylation and gene expression in synovium. They found associations of rs11583641 with *COLGALT2* expression and methylation, with directions of effect opposite to those previously estimated in cartilage, highlighting the tissue specificity of rs11583641 effects on the molecular profile.

Additional, comprehensive and better-powered studies that focus on the primary joint tissues of osteoarthritis patients are required. Some studies have generated DNA methylation profiles of osteoarthritis-affected tissues of patients but are limited regarding (1) sample sizes and (2)

tissue types. Furthermore, (3) genome-wide methylation QTL maps that enable resolution of osteoarthritis GWAS signals have not been generated to date.

Aims

In this thesis, I present studies that (1) significantly increase sample sizes of osteoarthritis tissues, (2) extend epigenome-wide profiles to synovium as well as infrapatellar fat pad and (3) generate genome-wide methylation QTL maps in osteoarthritis joint tissues. This thesis includes the following publications:

- Study 1: I present a large EWAS (90 patients) for cartilage degeneration (Publication 1 in Appendix)[7]. Furthermore, we provide the first genome-wide mQTL maps in low- and high-grade osteoarthritis cartilage as well as in synovium.

- Study 2: I describe the first epigenetic profile and mQTL map of osteoarthritis-affected infrapatellar fat pad (Publication 2 in Appendix)[51].

- Study 3: I further increase sample size for a cartilage degeneration EWAS, now including 170 patients (Publication 3 in Appendix)[52]. This large cohort further enable sex-specific epigenome-wide profiles of osteoarthritis progression in cartilage.

3. Materials and methods

Patient cohort and sample collection

Samples from patients undergoing total knee replacement surgery due to late-stage osteoarthritis were collected. From weight-bearing areas of affected joints, two types of cartilage samples were collected: Low-grade osteoarthritis cartilage (Study 1: 97 patients[7], Study 3: 170 patients), which refers to macroscopically intact cartilage, and high-grade osteoarthritis cartilage (Study 1: 90 patients[7], Study 3: 170 patients), describing

macroscopically degraded cartilage samples. The cartilage degradation status was quantified by the OARSI[53] and/or ICRS [54] grading systems.

Samples from the synovial membrane (from the suprapatellar region from 78 patients)[17] as well as adipose tissue from the infrapatellar fat pad (by sharp dissection of the fat tissue from the surface of the patellar ligament from 70 patients)[55] were also collected.

All patients provided written, informed consent prior to participation in the study. This work was approved by Oxford NHS REC C (10/H0606/20, SC/15/0132 and SC/20/0144) and samples were collected under Human Tissue Authority license 12182, South Yorkshire and North Derbyshire Musculoskeletal Biobank, University of Sheffield, UK.

Cell isolation and DNA extraction

We isolated the relevant cell type (chondrocytes, synoviocytes and adipocytes from cartilage, synovium and infrapatellar fat pad, respectively) from relevant joint tissues by applying published protocols[17,41,51], and used Qiagen AllPrep DNA Min Kit for DNA extraction.

DNA methylation measurement

We applied two Illumina array-based platforms to measure DNA methylation. HumanMethylation450 Bead Chip (450K) measures methylation levels at 485,512 methylation sites[56], prioritising sites in genes (covering 21,231 out of 21,474 of genes in The Reference Sequence database) or regulatory elements (26,658 regions with high CpG site frequencies, referred to as CpG islands).

Methylation EPIC BeadChip (EPIC) extended to more than 850,000 methylation sites [34]. It substantially overlaps with the older 450k array (450,161 CpG sites are shared), and also

contain 413,743 new CpG sites. These newly included methylation sites have higher proportions in gene bodies (32% of EPIC-only vs. 27% of shared EPIC/450k array in gene bodies), intergenic regions (20% of EPIC-only vs. 13% of shared EPIC/450k array in intergenic regions) or enhancer regions (21,070 EPIC-only sites vs. 7,763 sites shared EPIC/450k sites are in FANTOM5 enhancer regions), thus providing better insights into these genomic elements.

To measure samples on these array technologies, we performed bisulfite conversion on isolated DNA by treating it with sodium bisulfite, which converts unmethylated cytosine to thymine, while methylated cytosine remains unchanged[57]. The bisulfite-converted genomic DNA is then subjected to the DNA methylation array, which includes two query probes per interrogated methylation site that hybridise with either the methylated or unmethylated version of the site. The intensities of each probe pair are related to each other by calculating the *beta* value (using R package meffil [58]) to quantify the methylation level of a specific methylation site[59]:

$$beta_i = \frac{\max{(y_{i,unmeth}, 0)}}{\max(y_{i,unmeth}, 0) + \max(y_{i,meth}, 0) + alpha}$$

with $y_{i,meth}$ and $y_{i,unmeth}$ denoting the intensities measured by the methylated and unmethylated probe for a methylation site *i*, respectively. Furthermore, *alpha* denotes a recommended constant (default: 100) to regularise very low intensities[59]. *Beta* values range from 0 (completely unmethylated) to 1 (completely methylated).

Since Beta values may show heteroscedasticity, particularly for highly methylated or unmethylated sites, these are often converted to M-values for downstream analyses[59]:

$$M\text{-}value_i = log_2{\left(\frac{\max{(y_{i,meth}, 0)} + alpha}{\max{(y_{i,un\ meth}, 0)} + alpha}\right)}$$

Positive and negative M-values denote methylation sites that are mostly methylated and unmethylated, respectively.

Together, we applied array-based technologies to measure DNA methylation profiles on genome-wide scale.

Epigenome-wide association study for cartilage degeneration

We conducted epigenome-wide association studies (EWAS) to compared DNA methylation between low- and high-grade osteoarthritis cartilage samples matched from the same patients to identify methylation sites with different methylation levels, referred to as differentially methylated sites (DMS). More specifically, we performed a test per methylation site by building paired (to compare samples from the same patient) linear models. We constructed these models using the R package limma[60], particularly by using lmFit (build models) and eBayes function. To account for technical variations (such as batch effects) and minimise the influence of potential confounding factors, we included surrogate variables (SV) from DNA methylation data as covariates in the model:

$$M\text{-}value \sim cartilage\_type + patient\_id + SV$$

Here, *cartilage_type* denotes whether a sample is of low- or high-grade ostearothrits cartilage. The variable *patient_id* denotes the individual, thus ensures paired modeling. Furthermore, *M-value* refers to the methylation levels (in M-value unit as described in the Methods section "DNA methylation measurement"). To correct for multiple testing, we conservatively applied Bonferroni correction (significance threshold: $p < 0.05$ / number_tests).

We further tested whether there are differentially methylated regions (DMR) between low- and high-grade osteoarthritis cartilage samples. We applied the R package dmrff (function dmrff with default parameter setting[61]; DMR are composed of more than one methylation site and achieve a Bonferroni-adjusted p < 0.05).

Pathway enrichment analysis

To biologically interpret DMS, we performed pathway enrichment analysis for genes that are annotated to differentially methylated sites. Here, we used publicly available annotation files (https://zwdzwd.github.io/InfiniumAnnotation/).

We performed this analysis used the gometh function from the R package missMethyl[62,63] which accounts for biases that are introduced when mapping DNA methylation sites to genes. More specifically, bias can be introduced since a gene can contain several DNA methylation sites ("probe-number bias"[62]) as well as a DNA methylation site can be mapped to several genes ("multi-gene bias"[62]). Genes harbouring more methylation sites are more likely to be identified as differentially methylated[62]. Furthermore, genes can contain the same methylation sites, thus violating independence assumptions[62]. The gometh function from the R package missMethyl[62,63] implements a statistical testing procedure (Wallenius' noncentral hypergeometric distribution) which models probe-number and multi-gene bias as biased sampling problem.

Technically, we used DMS as query (gometh parameter "sig.cpg") and methylation sites that passed the QC (all methylation sites that could have potentially been identified as DMS in the EWAS) as background set (gometh parameter "all.cpg").

Methylation QTL analysis

We integrated matched methylation data and genotype data from the same patients to generate genome-wide mQTL maps in primary low- and high-grade osteoarthritis cartilage, synovium and fat pad. We sought to identify common genetic variants (minor allele frequency > 5%) that are associated with methylation levels of close methylation (within 1 Mb), referred to as cis mQTL. Here, we tested genetic variant-methylation site pairs by using linear models:

$$M\text{-}value \sim genetic\_variant + age + sex + technical\_covs$$

We included *age* and *sex* in these models to account for potential biological confounding. Furthermore, we account for technical confounders, here denoted by *technical_cov*. We either directly added known information (such as sequencing batch information per sample) or estimate these from data, for example by including PEER (probabilistic estimation of expression residuals) factors[7,64].

We performed cis-mQTL analysis by applying established software, such as Matrix eQTL[65] or FastQTL[66]. When using R package Matrix eQTL (function Matrix_eQTL_main)[65], we corrected for multiple testing by conservatively applying the Bonferroni correction or false discovery rate (FDR) correction.

FastQTL[66] enables more sophisticated multiple testing correction for cis-QTL mapping across genetic variants and methylation sites. Per methylation site, it corrects for tested genetic variants by applying an adaptive permutation scheme to generate empirical p values. It further corrects across tested methylation sites by applying the Storey-Tibshirani FDR procedure[67] (with FDR < 0.05 denoting QTL-targeted methylation sites). These FDR-corrected pvalues are in turn used to calculate methylation site-specific significance thresholds. For each

methylation sites, variants with a nominal pvalue below these thresholds are defined as mQTLs.

Colocalisation analysis

We integrated osteoarthritis GWAS results of large meta-analyses[4,68] and mQTL maps on summary statistics level by performing colocalisation analysis to test whether a GWAS risk locus and a mQTL signals for mQTL targeted methylation sites statistically overlap ('colocalise'), suggesting a shared causal variant for the respective methylation site and GWAS trait. We performed colocalisation analysis[69] when the index variants of a GWAS risk signal was in proximity (e.g. within 100 kilobases) to a mQTL-targeted methylation site and considered all cis variants (within 1 Mb either side of a methylation site) for the actual colocalization test. We applied the coloc.fast function (default parameter settings; https://github.com/tobyjohnson/gtx/blob/526120435bb3e29c39fc71604eee03a371ec3753/R/coloc.R)[69] to estimate a posterior probability for having a shared causal variant. We used threshold of 80% denoting colocalisation ('PP4 ≥ 0.8').

Mendelian randomisation analysis

We supplemented these causal insights by applying causal inference analysis. More specifically, we applied two-sample Mendelian randomisation (MR) analysis by integrating GWAS [4,68] and mQTL summary statistics. For mQTL-targeted methylation sites, we used mQTL as instruments to estimate a causal effect of methylation (exposure) on osteoarthritis (outcome). We applied the R package TwoSampleMR[70] to conduct MR analysis for every mQTL-tageted methylation site. We further performed clumping (r2 < 0.01) to filter for independent mQTLs per methylation site. For methylation sites with exactly 1 independent

23

mQTL, we applied the Wald-ratio, otherwise the inverse-variance-weighted (IVW) method. We corrected for multiple testing using Bonferroni correction.

Study design 1: An epigenome-wide view of osteoarthritis in primary tissues

We collected samples from primary chondrocytes from low- (macroscopically intact) and high-grade (macroscopically degraded) osteoarthritis cartilage as well as synoviocytes from the synovium from 98 patients (Figure 1; Publication 1 in Appendix)[52]. We measured DNA methylation using Illumina 450k and EPIC array. We further genotyped patients using the Illumina CoreExome chip and performed imputation. We performed an EWAS by comparing low- and high-grade osteoarthritis cartilage in to find epigenetic markers for cartilage degeneration which we characterized in a follow-up pathway enrichment analysis. We further generated mQTL maps in low- and high-grade osteoarthritis cartilage and synovium by integrating genotype data from the same patients. We integrated these mQTL maps with osteoarthritis GWAS results by applying colocalisation and causal inference methods to examine the causal role of methylation in osteoarthritis in a disease-stage sand tissue-specific manner.
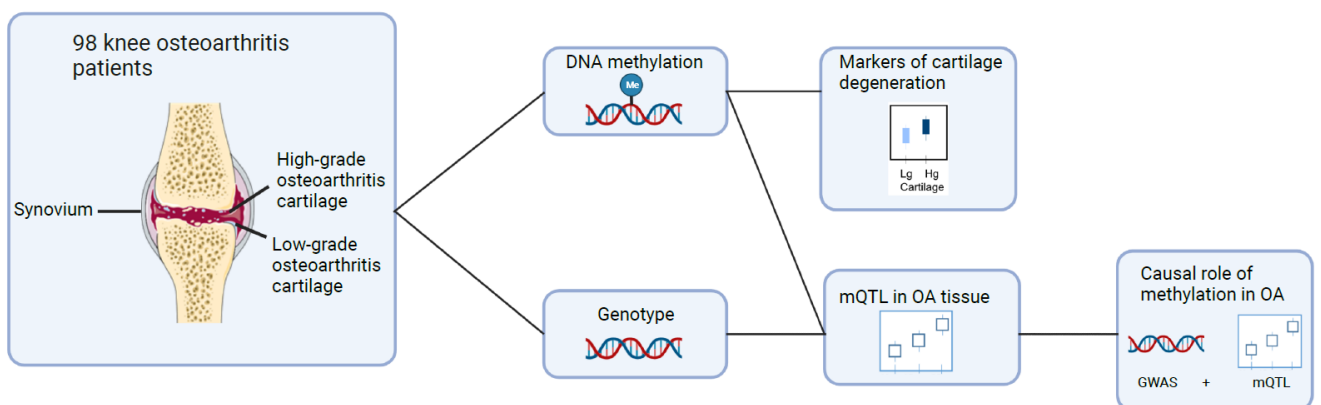
Figure 1: We use cartilage and synovium methylation profiles to perform EWAS for cartilage degeneration as well as generate mQTL maps to resolve osteoarthritis GWAS signals. This figure was created with BioRender.com.

Study design 2: Epigenomic profiling of the infrapatellar fat pad in osteoarthritis

We collected blood as well as infrapatellar fat pad samples from 70 patients (Figure 2; Publication 2 in Appendix)[51]. We profiled fat pad and blood methylation profiles using Illumina EPIC array and estimated patient genotypes with whole-genome sequencing. We compared blood and fat pad methylation profiles to investigate their distinctness. We combined fat pad methylation and genotype data matched from the same patient to generate a infrapatellar fat pad mQTL map. Furthermore, we combined this map with GWAS results and performed colocalisation and causal inference analysis to investigate the causal role of fat pad methylation in osteoarthritis.
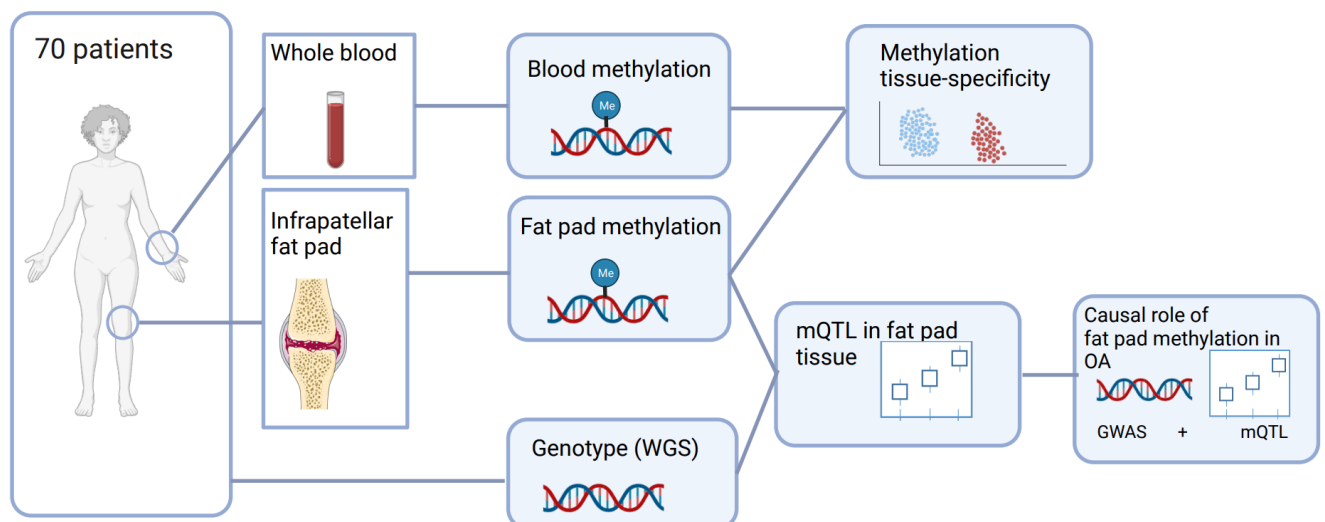

Figure 2: We compare infrapatellar fat pad with blood methylation and generate a fat pad mQTL map using methylation profiles of 70 patients. This figure was created with BioRender.com.

Study design 3: Epigenomic differences between osteoarthritis grades in primary cartilage

We collected samples from primary chondrocytes from 170 patients across two degradation stages: low- (macroscopically intact) and high-grade (macroscopically degraded) osteoarthritis cartilage (Figure 3, Publication 3 in Appendix)[52]. These methylation profiles were measured using Illumina EPIC array. We performed an EWAS for cartilage degeneration by compared low- and high-grade osteoarthritis cartilage methylation We then conducted EWAS stratified by sex (in 96 women and 74 men) to generate sex-specific epigenetic profiles of cartilage degeneration. We further compared EWAS in men and women to find common and sex-specific epigenetic markers for cartilage degeneration.
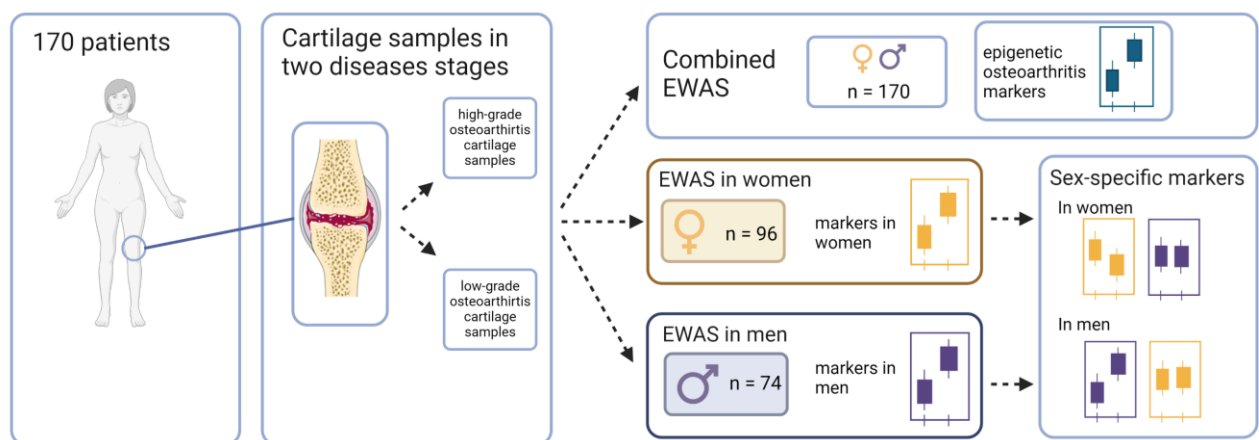

Figure 3: We perform sex-combined and sex-specific EWAS for cartilage degeneration using cartilage methylation profiles of 170 patients. This figure was created with BioRender.com.

4. Results

4.1. An epigenome-wide view of osteoarthritis in primary tissues

We examined genotype and DNA methylation data of primary chondrocytes macroscopically intact (low-grade) and degraded (high-grade) cartilage and primary synoviocytes of 98 patients who underwent total knee replacement due to late-stage osteoarthritis (Figure 1; Publication 1 in Appendix)[7].

We performed an epigenome-wide association study (EWAS) to compare DNA methylation profiles between low-grade and high-grade osteoarthritis cartilage and identified widespread epigenetic markers of cartilage degeneration (15,328 differentially methylated sites), of which we were able to replicate 46.5 % in a smaller, independent validation set (17 patients), thus highlighting the robustness of these signals.

We built random forest-based classifiers that could distinguish low-grade and high-grade osteoarthritis cartilage samples. In our own patient cohort, these classifiers achieved high accuracies (mean accuracy: 90.69%, standard deviation: 4.08), which we were able to validate in an external validation set (82.35% accuracy).

We estimated genome-wide cis methylation quantitative trait locus (mQTL) maps of low- and high-grade osteoarthritis cartilage, as well as synovium, and identified a multitude of QTL targeted methylation sites (low-grade cartilage: 73,836, high-grade cartilage: 52,819, synovium: 40,361, FDR < 0.05). By comparing the mQTL profiles between low- and high-grade osteoarthritis cartilage, we identified 18 genetic variants that exert an effect on a specific methylation site only in low-, but not in high-grade osteoarthritis cartilage or vice versa, suggesting genetic effects on methylation that are switched on/off during osteoarthritis-related cartilage degeneration processes.

We integrated mQTL maps of the three osteoarthritis primary tissues with GWAS results for three osteoarthritis traits, namely osteoarthritis at any site, knee osteoarthritis, and total knee replacement. Using Mendelian randomisation, we identified methylation sites with a putative causal effect on osteoarthritis (low-grade cartilage: six, high-grade cartilage: eight, synovium: eight). Furthermore, colocalisation resolved GWAS signals for osteoarthritis at any site (13 of

33 tested GWAS signals), knee osteoarthritis (six of 12 tested GWAS signals), and total knee replacement (one of five tested GWAS signals).

In summary, we have performed the largest EWAS for cartilage degeneration, generated the first genome-wide mQTL maps in cartilage and synovium and suggest epigenetic mechanisms underlying osteoarthritis in cartilage and synovium. We identified robust epigenetic markers of cartilage degeneration, and propose a validated model than can distinguish cartilage samples in different osteoarthritis stages.

Author contribution:

- Data analysis (DNA methylation data preprocessing, differential analyses, classifier construction, mQTL analyses, Mendelian randomisation, colocalisation, replication analyses)
- Interpretation of results (with Eleftheria Zeggini, Matthew Suderman, J Mark Wilkinson and Caroline L Relton)
- Manuscript drafting (with Eleftheria Zeggini)
- Manuscript reviewing and editing (with all other authors)

4.2. Epigenomic profiling of the infrapatellar fat pad in osteoarthritis

We characterised the methylation profile of primary infrapatellar fat pad tissue at the epigenome-wide level in osteoarthritis-affected knees. We investigated genome-wide DNA methylation profiles of primary infrapatellar from 70 osteoarthritis patients undergoing total knee replacement surgery (Figure 2; Publication 2 in Appendix)[51]. From a subset of these patients, matching blood DNA methylation profiles (n = 58) and whole genome sequencing data (n = 68 patients) were available.

We conducted an epigenome-wide association study (EWAS) to compare the DNA methylation profiles between the infrapatellar fat pad and blood and identified extensive differences (84,973 differentially methylated sites, $p < 6.4 \times 10^{-08}$). We further generated a cis methylation quantitative trait locus (mQTL) map in fat pad tissue by integrating fat pad methylation and genotype data matched from the same patients, and identified 35,948 mQTL-targeted methylation sites. This constitutes the first genome-wide mQTL map of fat pad which we make available to the wider community. Using two-sample Mendelian randomisation and colocalisation analyses, we resolve eleven osteoarthritis GWAS signals and provide insights into the molecular mechanisms underlying osteoarthritis aetiopathology.

Our study generates the first epigenetic landscape of infrapatellar fat pad, highlights its distinctness from blood, and provides the first genome-wide mQTL map of primary infrapatellar fat pad tissues of osteoarthritis-affected knees.

Author contribution:

- Data analysis (DNA methylation data preprocessing, differential analyses, mQTL analyses, Mendelian randomisation, colocalisation)
- Interpretation of results (with Eleftheria Zeggini)
- Manuscript drafting (with Eleftheria Zeggini)
- Manuscript reviewing and editing (with Eleftheria Zeggini, J Mark Wilkinson and Diane Swift)

4.3. Epigenomic differences between osteoarthritis grades in primary cartilage

We investigated the genome-wide DNA methylation profiles of primary chondrocytes from macroscopically intact (low-grade) and degraded (high-grade) cartilage from osteoarthritis patients undergoing total knee replacement for osteoarthritis (Figure 3; Publication 3 in Appendix)[52].

We conducted an epigenome-wide association study (EWAS) for cartilage degeneration by comparing low- and high-grade osteoarthritis cartilage samples from within-individual matched samples across 170 patients and identified 146,777 DMS ($p < 6.41 \times 10 \times 10^{-8}$). Biological characterisation of these markers revealed enrichment of a wide biological spectrum, including apoptosis- and neuron-related terms.

We further generated sex-specific epigenetic profiles of cartilage degeneration by performing separate EWAS in 96 women and 74 men, and identified widespread epigenetic osteoarthritis markers (women: 62,313 DMS, men: 61,513 DMS, $p < 6.41 \times 10^{-8}$).

Comparing these EWAS results identified substantial overlaps (43,152 DMS identified in women and men) as well as sex-specific markers (women: 361 DMS; men: 480 DMS; with $p < 6.41 \times 10^{-8}$ in one sex but $p > 0.05$ in the other). Furthermore, we identified pathways that may be osteoarthritis-related in a sex-specific manner (FDR < 0.05 in one sex, but $p > 0.05$ in the other) (women: 19 GO terms, men: 51), including the immune system as well as nervous system-related terms.

We provide the largest genome-wide methylation profile of primary osteoarthritis cartilage to date, which almost doubles the size of the next largest study, enabling enhanced and sex-specific insights into epigenetic processes underlying osteoarthritis progression.

Author contribution:

- Data analysis (DNA methylation data preprocessing, differential analyses, sex-specificity analyses)
- Interpretation of results (with Eleftheria Zeggini)
- Manuscript drafting (with Eleftheria Zeggini)
- Manuscript reviewing and editing (with Eleftheria Zeggini and J Mark Wilkinson)

5. Discussion

Osteoarthritis is a complex and prevalent joint disease. GWAS have generated insights into its polygenic architecture, but genome-wide DNA methylation profiles of affected joint tissues are excluded from large consortia (such as GTEx[71], Roadmap[12], ENCODE[13]). In this work, I have generated genome-wide DNA methylation profiles across four joint tissue types, detect and characterise epigenetic osteoarthritis markers, and generate genome-wide mQTL maps, which I have used to identify mechanistically relevant methylation sites and genes.

Methylation profiles reveal extensive differences between tissues, for instance, global methylation differences between low- and high-grade cartilage and synovium[7]. To further refine epigenetic insights into osteoarthritis progression in cartilage, I conducted the largest EWAS for cartilage degeneration to date[7,52] which revealed widespread differences on methylation site and region level. I found strong evidence for replication in an independent dataset[7], despite its smaller sample size and lower power. Together, these results underline the cell-type and disease-grade specificity of methylation in primary tissues, and thus the need to profile all components of the affected joint to generate a full molecular picture of osteoarthritis.

Comparing joint with blood methylation data emphasise the importance of examining joint primary tissues. Differences comprised distinct methylation profiles between fat pad and blood methylation in the same patient cohort[52], mQTLs with opposite directions of effect when comparing joint and blood maps as well as colocalisation signals in joint, that were absent in blood[7]. The latter propose that at least a subset of the regulatory effects mediated by osteoarthritis-linked variants through close methylation sites are specific to osteoarthritis-affected tissues. Together, these findings underline the value of studying disease-relevant

31

tissues rather than solely molecular data from peripheral tissue types. Biological characterisation of osteoarthritis-progression related epigenetic markers in cartilage, through pathway enrichment analysis, suggests a multitude of osteoarthritis-associated processes, such as previously-reported external matrix organization, skeletal system development, and signaling pathways, reflecting a wide spectrum of involved biological mechanisms during osteoarthritis-related cartilage degeneration. Of note, I identified processes related to the epithelium (such as "positive regulation of epithelial cell migration")[7], nervous system (e.g. "dendrite morphogenesis", "regulation of synaptic plasticity") or neurotransmission related terms (e.g. "synaptic vesicle cycle", "neurotransmitter secretion"), for the first time, which may be linked with angiogenesis or innervation in affected cartilage, and ultimately be associated with pain development in affected knees[72]. Other pathways are related to apoptosis (such as "regulation of extrinsic apoptotic signaling pathway") which has been linked with cartilage breakdown previously[73]. Together, these results suggest that DNA methylation is involved in regulatory changes of these biological pathways during osteoarthritis.

For the first time, I provided sex-specific epigenetic insights into osteoarthritis progression[52]. I performed EWAS for cartilage degeneration separately in women (n = 96) and men (n = 74) and found widespread epigenetic markers, respectively. Of note, I identified a number of sex-specific markers (women: 413 markers, men: 539 markers), proposing a few sex-specific molecular mechanisms. I have also identified pathways that are enriched in epigenetic osteoarthritis markers in only one sex (women: 19 pathways, men: 51 pathways). Some were related to innervation and neurotransmission which may be linked with more pain in affected joints in women[74]. Other sex-specific pathways were immune system-related which is

potentially associated with proinflammatory factor level differences between sexes in chondrocyte cell cultures (*IL1A*, *IL6*, and *IL8* expression levels in cultured chondrocytes of low-grade osteoarthritis cartilage are increase in women [75]) or overrepresentation of women in a high inflammation-linked osteoarthritis patient cluster[19]. Together, these results suggest sex-specific roles of some epigenetic markers or mechanisms which are potentially linked with higher osteoarthritis prevalence and severity in women. Our studies present the first genome-wide mQTL maps of affected joint tissues, across disease-stages (low- and high-grade osteoarthritis cartilage) as well as joint tissues (cartilage, synovium and infrapatellar fat pad). These results are publicly available (https://msk.hugeamp.org/downloads.html), and thus constitute a highly relevant source for the osteoarthritis research community and beyond. These maps revealed 18 differential mQTLs between low- and high-grade osteoarthritis cartilage samples, proposing different regulatory effects of genetic variants on methylation sites at early and later disease stages. This suggests altering genetic effects on epigenetic profiles during osteoarthritis progression.

Integrating these mQTL maps with osteoarthritis GWAS results, such as through colocalisation and MR analysis identified methylation sites with a putative causal role in low- (36 methylation sites) and high-grade (37 methylation sites) osteoarthritis cartilage, synovium (24 methylation sites) and infrapatellar fat pad (37 methylation sites). These signals comprised methylation sites located in genes which are involved in osteoarthritis-related pathways.

For example, I found cartilage and synovium methylation in *WWP2* to be causally involved in osteoarthritis. *WWP2* encodes a ligase that contributes to protein ubiquitination. It is linked with micro-RNA140, a regulator in chondrocytes. Mouse models revealed a regulatory role of *Wwp2* in cartilage homeostasis through *Adamts5*[76]. In cartilage, I further estimated

33

methylation in *ALDH1A2* or *LTBP1* with a causal role in osteoarthritis. ALDH1A2 is an enzyme that catalyses the synthesis of retinoic acid (activated form of vitamin A). It has been associated with bone collagen degeneration[77] and is used to induce matrix degeneration in cartilage samples[78]. LTBP1 is involved in regulating transforming growth factor (TGF) betas, a cytokine class with a role in extracellular matrix synthesis and maintenance as well as inflammation and chondrocytes hypertrophy control[79]. Chondrocytes generate inactivated TGF betas which are bound by LTBP1 to the extracellular matrix in cartilage[80]. In synovium, I also found *BSN, CRADD* and *MFHAS1* methylation to be putatively causally linked with osteoarthritis. The gene *MFHAS1* is involved in the regulation of Toll-like receptors TLR2 and TLR4[81,82]. These receptors, presented by cells in the synovium, interact with matrix molecules released from degraded cartilage, triggering the synthesis of chemokines and cytokines, and ultimately the inflammatory cell infiltration of the synovium. The adapter protein CRADD contributes to the synthesis of the apoptosis-related PIDDosome-complex, which in turn activates *CASP2*[83]. This result proposes that apoptosis, which has been related to osteoarthritis in the synovium previously[84], is regulated through DNA methylation in the synovium. BSN is a component of the presynaptic skeleton complex (a structure that assists in the vesicle fusion of synaptic vesicles and presynaptic membranes), thus is involved in neurotransmission[85]. Furthermore, genes with potential osteoarthritis relevant methylation sites in the infrapatellar fat pad comprised *USP8, TSKU* and *FER1L4. USP8* encodes a protein that contributes to the regulation of epidermal growth factor receptor which has been linked to inflammation and angiogenesis[86]. The gene *TSKU* inhibits the osteoarthritis-linked Wnt signalling pathway[87]. Furthermore, *FER1L4* modulates osteoarthritis-associated factors IL-6[88] and VEGF[89]. Together, these mQTL maps reveal causal association of methylation with osteoarthritis.

My efforts enhance our understanding of the molecular profiles of osteoarthritis. Emerging sequencing technologies will increase resolution into the molecular profile of osteoarthritis-relevant tissues, including at the single-cell and spatial omics levels. For example Perturb-seq[90] investigates genetic variations at the single-cell transcriptomics level, thus better elucidating the functional genomics of complex diseases. Long-read sequencing technologies profile less well-studied genomic regions, such as genomic structural variation or highly repetitive regions (including telomeres or centromeres)[91]. The most common DNA methylation profiling technology is currently Illumina's methylation array set, which target 450,000 (450k) or 900,000 (EPIC) methylation sites. Sequencing-based technologies, such as whole genome bisulfate sequencing[92] may extend methylation profiling to roughly 28 million methylation sites in the human genome. Application of these novel technologies may improve molecular insights to osteoarthritis tissues.

Genetic and molecular studies of osteoarthritis have primarily focused on individuals of European ancestry, which bias our understanding of the disease. Extending studies to non-European populations can generate more comprehensive genetic and molecular profiles of osteoarthritis, ensuring that the results are both accessible and applicable to individuals worldwide.

In summary, my work highlights the tissue as well as disease-grade specificity of the epigenetic profile in osteoarthritis-affected joint tissues. I identify a multitude of epigenetic osteoarthritis markers, reveal disease-associated biological pathways and identify likely effectors genes of osteoarthritis. My findings highlight the relevant role of DNA methylation in osteoarthritis aetiopathogenesis.

References

1. Safiri, S. *et al.* Global, regional and national burden of osteoarthritis 1990-2017: a systematic analysis of the Global Burden of Disease Study 2017. *Ann Rheum Dis* **79**, 819–828 (2020).

2. Srikanth, V. K. *et al.* A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. *Osteoarthritis Cartilage* **13**, 769–781 (2005).

3. Vos, T. *et al.* Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2163–2196 (2012).

4. Boer, C. G. *et al.* Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. *Cell* **184**, 4784-4818.e17 (2021).

5. Goldring, S. R. & Goldring, M. B. Changes in the osteochondral unit during osteoarthritis: structure, function and cartilage-bone crosstalk. *Nat Rev Rheumatol* **12**, 632–644 (2016).

6. Roughley, P. J. & Mort, J. S. The role of aggrecan in normal and osteoarthritic cartilage. *J Exp Orthop* **1**, 8 (2014).

7. Kreitmaier, P. *et al.* An epigenome-wide view of osteoarthritis in primary tissues. *The American Journal of Human Genetics* **0**, (2022).

8. Mathiessen, A. & Conaghan, P. G. Synovitis in osteoarthritis: current understanding with therapeutic implications. *Arthritis Res Ther* **19**, 18 (2017).

9. Eymard, F. *et al.* Induction of an inflammatory and prodegradative phenotype in autologous fibroblast-like synoviocytes by the infrapatellar fat pad from patients with knee osteoarthritis. *Arthritis Rheumatol* **66**, 2165–2174 (2014).

10. Zeggini, E., Gloyn, A. L., Barton, A. C. & Wain, L. V. Translational genomics and precision medicine: Moving from the lab to the clinic. *Science* **365**, 1409–1413 (2019).

11. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).

12.   Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

13.   Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

14.   Regev, A. *et al.* The Human Cell Atlas. *Elife* **6**, e27041 (2017).

15.   Katsoula, G., Kreitmaier, P. & Zeggini, E. Insights into the molecular landscape of osteoarthritis in human tissues. *Curr Opin Rheumatol* **34**, 79–90 (2022).

16.   Kreitmaier, P., Katsoula, G. & Zeggini, E. Insights from multi-omics integration in complex disease primary tissues. *Trends Genet* **39**, 46–58 (2023).

17.   Steinberg, J. *et al.* A molecular quantitative trait locus map for osteoarthritis. *Nature Communications* **12**, 1309 (2021).

18.   Katsoula, G. *et al.* A molecular map of long non-coding RNA expression, isoform switching and alternative splicing in osteoarthritis. *Hum Mol Genet* **31**, 2090–2105 (2022).

19.   Steinberg, J. *et al.* Linking chondrocyte and synovial transcriptional profile to clinical phenotype in osteoarthritis. *Ann Rheum Dis* **80**, 1070–1074 (2021).

20.   Coutinho de Almeida, R. *et al.* RNA sequencing data integration reveals an miRNA interactome of osteoarthritis cartilage. *Ann Rheum Dis* **78**, 270–277 (2019).

21.   Pu, H. *et al.* Single cell transcriptome profiling of infrapatellar fat pad highlights the role of interstitial inflammatory fibroblasts in osteoarthritis. *Int Immunopharmacol* **131**, 111888 (2024).

22.   Wang, X. *et al.* Comparison of the major cell populations among osteoarthritis, Kashin-Beck disease and healthy chondrocytes by single-cell RNA-seq analysis. *Cell Death Dis* **12**, 551 (2021).

23.   Chou, C.-H. *et al.* Synovial cell cross-talk with cartilage plays a major role in the pathogenesis of osteoarthritis. *Sci Rep* **10**, 10868 (2020).

24.   Grandi, F. C. *et al.* Single-cell mass cytometry reveals cross-talk between inflammation-dampening and inflammation-amplifying cells in osteoarthritic cartilage. *Sci Adv* **6**, eaay5352 (2020).

25.   Sun, H. *et al.* Single-cell RNA-seq analysis identifies meniscus progenitors and reveals the progression of meniscus degeneration. *Ann Rheum Dis* **79**, 408–417 (2020).

26.   Ji, Q. *et al.* Single-cell RNA-seq analysis reveals the progression of human osteoarthritis. *Ann Rheum Dis* **78**, 100–110 (2019).

27.   Liu, Y. *et al.* Chromatin accessibility landscape of articular knee cartilage reveals aberrant enhancer regulation in osteoarthritis. *Scientific Reports* **8**, 15499 (2018).

28.   Bittner, N. *et al.* Primary osteoarthritis chondrocyte map of chromatin conformation reveals novel candidate effector genes. *Ann Rheum Dis* ard-2023-224945 (2024) doi:10.1136/ard-2023-224945.

29.   Thulson, E. *et al.* 3D chromatin structure in chondrocytes identifies putative osteoarthritis risk genes. *Genetics* **222**, iyac141 (2022).

30.   Brena, R. M., Huang, T. H.-M. & Plass, C. Toward a human epigenome. *Nature Genetics* **38**, 1359–1360 (2006).

31.   Lyko, F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat Rev Genet* **19**, 81–92 (2018).

32.   Onodera, A. *et al.* Roles of TET and TDG in DNA demethylation in proliferating and non-proliferating immune cells. *Genome Biol* **22**, 186 (2021).

33.   Kaur, D. *et al.* Comprehensive Evaluation of The Infinium Human MethylationEPIC v2 BeadChip. *Epigenetics Commun* **3**, 6 (2023).

34.   Pidsley, R. *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol* **17**, 208 (2016).

35.   Sandoval, J. *et al.* Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* **6**, 692–702 (2011).

36.   Petronis, A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature* **465**, 721–727 (2010).

37.	DNA methylation QTL mapping across diverse human tissues provides molecular links between genetic variation and complex traits - PubMed. https://pubmed.ncbi.nlm.nih.gov/36510025/.

38.	Kehayova, Y. S., Wilkinson, J. M., Rice, S. J. & Loughlin, J. Mediation of the Same Epigenetic and Transcriptional Effect by Independent Osteoarthritis Risk-Conferring Alleles on a Shared Target Gene, COLGALT2. *Arthritis Rheumatol* **75**, 910–922 (2023).

39.	Kehayova, Y. S., Watson, E., Wilkinson, J. M., Loughlin, J. & Rice, S. J. Genetic and Epigenetic Interplay Within a COLGALT2 Enhancer Associated With Osteoarthritis. *Arthritis Rheumatol* (2021) doi:10.1002/art.41738.

40.	Rice, S. J., Cheung, K., Reynard, L. N. & Loughlin, J. Discovery and analysis of methylation quantitative trait loci (mQTLs) mapping to novel osteoarthritis genetic risk signals. *Osteoarthritis Cartilage* **27**, 1545–1556 (2019).

41.	Steinberg, J. *et al.* Integrative epigenomics, transcriptomics and proteomics of patient chondrocytes reveal genes and pathways involved in osteoarthritis. *Sci Rep* **7**, 8935 (2017).

42.	Zhang, Y. *et al.* Genome-wide DNA methylation profile implicates potential cartilage regeneration at the late stage of knee osteoarthritis. *Osteoarthritis and Cartilage* **24**, 835–843 (2016).

43.	Bonin, C. A. *et al.* Identification of Differentially Methylated Regions in New Genes Associated with Knee Osteoarthritis. *Gene* **576**, 312–318 (2016).

44.	den Hollander, W. *et al.* Transcriptional associations of osteoarthritis-mediated loss of epigenetic control in articular cartilage. *Arthritis Rheumatol* **67**, 2108–2116 (2015).

45.	Moazedi-Fuerst, F. C. *et al.* Epigenetic differences in human cartilage between mild and severe OA. *J Orthop Res* **32**, 1636–1645 (2014).

46.	Kehayova, Y. S., Wilkinson, J. M., Rice, S. J. & Loughlin, J. Osteoarthritis genetic risk acting on the galactosyltransferase gene COLGALT2 has opposing functional effects in articulating joint tissues. *Arthritis Res Ther* **25**, 83 (2023).

47.   Parker, E. *et al.* Multi-Tissue Epigenetic and Gene Expression Analysis Combined With

Epigenome Modulation Identifies RWDD2B as a Target of Osteoarthritis Susceptibility. *Arthritis &*

Rheumatology **73**, 100–109 (2021).

48.   Sorial, A. K. *et al.* Multi-tissue epigenetic analysis of the osteoarthritis susceptibility locus

mapping to the plectin gene PLEC. *Osteoarthritis Cartilage* (2020) doi:10.1016/j.joca.2020.06.001.

49.   Rice, S. J. *et al.* Identification of a novel, methylation-dependent, RUNX2 regulatory region

associated with osteoarthritis risk. *Hum Mol Genet* **27**, 3464–3474 (2018).

50*.*   Rushton, M. D. *et al.* Methylation quantitative trait locus analysis of osteoarthritis links

epigenetics with genetic risk. *Hum Mol Genet* **24**, 7432–7444 (2015).

51.   Kreitmaier, P. *et al.* Epigenomic profiling of the infrapatellar fat pad in osteoarthritis. *Hum*

Mol Genet **33**, 501–509 (2024).

52.   Kreitmaier, P., Swift, D., Wilkinson, J. M. & Zeggini, E. Epigenomic differences between

osteoarthritis grades in primary cartilage. *Osteoarthritis Cartilage* S1063-4584(24)01314–1 (2024)

doi:10.1016/j.joca.2024.07.008.

53.   Pearson, R. G., Kurien, T., Shu, K. S. S. & Scammell, B. E. Histopathology grading systems for

characterisation of human knee osteoarthritis--reproducibility, variability, reliability, correlation,

and validity. *Osteoarthritis Cartilage* **19**, 324–331 (2011).

54.   Mainil-Varlet, P. *et al.* Histological assessment of cartilage repair: a report by the Histology

Endpoint Committee of the International Cartilage Repair Society (ICRS). *J Bone Joint Surg Am* **85-**

A Suppl 2, 45–57 (2003).

55.   Epigenomic profiling of the infrapatellar fat pad in osteoarthritis - PubMed.

https://pubmed.ncbi.nlm.nih.gov/37975894/.

56.   Bibikova, M. *et al.* High density DNA methylation array with single CpG site resolution.

Genomics **98**, 288–295 (2011).

57.   Li, Y. & Tollefsbol, T. O. DNA methylation detection: bisulfite genomic sequencing analysis.

Methods Mol Biol **791**, 11–21 (2011).

58.   Min, J. L., Hemani, G., Davey Smith, G., Relton, C. & Suderman, M. Meffil: efficient

normalization and analysis of very large DNA methylation datasets. *Bioinformatics* **34**, 3983–3989

(2018).

59.   Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation

levels by microarray analysis. *BMC Bioinformatics* **11**, 587 (2010).

60.   Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and

microarray studies. *Nucleic Acids Res* **43**, e47 (2015).

61.   Suderman, M. *et al.* dmrff: identifying differentially methylated regions efficiently with power

and control. *bioRxiv* 508556 (2018) doi:10.1101/508556.

62.   Maksimovic, J., Oshlack, A. & Phipson, B. Gene set enrichment analysis for genome-wide DNA

methylation data. *Genome Biology* **22**, 173 (2021).

63.   Phipson, B., Maksimovic, J. & Oshlack, A. missMethyl: an R package for analyzing data from

Illumina's HumanMethylation450 platform. *Bioinformatics* **32**, 286–288 (2016).

64.   Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of

expression residuals (PEER) to obtain increased power and interpretability of gene expression

analyses. *Nat Protoc* **7**, 500–507 (2012).

65.   Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations.

Bioinformatics **28**, 1353–1358 (2012).

66.   Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL

mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).

67.   Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad

Sci U S A* **100**, 9440–9445 (2003).

68.   Tachmazidou, I. *et al.* Identification of new therapeutic targets for osteoarthritis through

genome-wide analyses of UK Biobank data. *Nature Genetics* **51**, 230–236 (2019).

69.   Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association

studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).

70.   Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the

human phenome. *eLife* **7**, e34408 (2018).

71.   Oliva, M. *et al.* DNA methylation QTL mapping across diverse human tissues provides

molecular links between genetic variation and complex traits. *Nat Genet* **55**, 112–122 (2023).

72.   Suri, S. *et al.* Neurovascular invasion at the osteochondral junction and in osteophytes in

osteoarthritis. *Annals of the Rheumatic Diseases* **66**, 1423–1428 (2007).

73.   Hwang, H. S. & Kim, H. A. Chondrocyte Apoptosis in the Pathogenesis of Osteoarthritis. *Int J

Mol Sci* **16**, 26035–26054 (2015).

74.   McAlindon, T. E., Cooper, C., Kirwan, J. R. & Dieppe, P. A. Knee pain and disability in the

community. *Br J Rheumatol* **31**, 189–192 (1992).

75.   Pan, Q. *et al.* Characterization of osteoarthritic human knees indicates potential sex

differences. *Biol Sex Differ* **7**, 27 (2016).

76.   Mokuda, S. *et al.* Wwp2 maintains cartilage homeostasis through regulation of Adamts5. *Nat

Commun* **10**, 2429 (2019).

77.   Varghese, S., Rydziel, S., Jeffrey, J. J. & Canalis, E. Regulation of interstitial collagenase

expression and collagen degradation by retinoic acid in bone cells. *Endocrinology* **134**, 2438–2444

(1994).

78.   Shlopov, B. V. *et al.* Osteoarthritic lesions: involvement of three different collagenases.

Arthritis Rheum **40**, 2065–2074 (1997).

79.   Thielen, N. G. M., van der Kraan, P. M. & van Caam, A. P. M. TGFβ/BMP Signaling Pathway in

Cartilage Homeostasis. *Cells* **8**, (2019).

80.   Saharinen, J., Taipale, J. & Keski-Oja, J. Association of the small latent transforming growth

factor-beta with an eight cysteine repeat of its binding protein LTBP-1. *EMBO J* **15**, 245–253

(1996).

81.   Shi, Q. *et al.* MFHAS1 suppresses TLR4 signaling pathway via induction of PP2A C subunit cytoplasm translocation and inhibition of c-Jun dephosphorylation at Thr239. *Mol Immunol* **88**, 79–88 (2017).

82.   Zhong, J. *et al.* MFHAS1 Is Associated with Sepsis and Stimulates TLR2/NF-κB Signaling Pathway Following Negative Regulation. *PLoS One* **10**, e0143662 (2015).

83.   Ahmad, M. *et al.* CRADD, a novel human apoptotic adaptor molecule for caspase-2, and FasL/tumor necrosis factor receptor-interacting protein RIP. *Cancer Res* **57**, 615–619 (1997).

84.   Huang, H. *et al.* Identification of pathways and genes associated with synovitis in osteoarthritis using bioinformatics analyses. *Sci Rep* **8**, 10050 (2018).

85.   Südhof, T. C. The presynaptic active zone. *Neuron* **75**, 11–25 (2012).

86.   Berlin, I., Schwartz, H. & Nash, P. D. Regulation of epidermal growth factor receptor ubiquitination and trafficking by the USP8·STAM complex. *J Biol Chem* **285**, 34909–34921 (2010).

87.   Wang, Y., Fan, X., Xing, L. & Tian, F. Wnt signaling: a promising target for osteoarthritis therapy. *Cell Commun Signal* **17**, 97 (2019).

88.   He, J., Wang, L., Ding, Y., Liu, H. & Zou, G. lncRNA FER1L4 is dysregulated in osteoarthritis and regulates IL-6 expression in human chondrocyte cells. *Sci Rep* **11**, 13032 (2021).

89.   Hamilton, J. L. *et al.* Targeting VEGF and Its Receptors for the Treatment of Osteoarthritis and Associated Pain. *J Bone Miner Res* **31**, 911–924 (2016).

90.   Schraivogel, D. *et al.* Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nat Methods* **17**, 629–635 (2020).

91.   Logsdon, G. A., Vollger, M. R. & Eichler, E. E. Long-read human genome sequencing and its applications. *Nat Rev Genet* **21**, 597–614 (2020).

92.   Stirzaker, C., Taberlay, P. C., Statham, A. L. & Clark, S. J. Mining cancer methylomes: prospects and challenges. *Trends in Genetics* **30**, 75–84 (2014).

Appendix

Publication 1

Kreitmaier P, Suderman M, Southam L, Coutinho de Almeida R, Hatzikotoulas K, Meulenbelt I, Steinberg J, Relton CL, Wilkinson JM*, Zeggini E*. An epigenome-wide view of osteoarthritis in primary tissues. Am J Hum Genet. 2022 Jul 7;109(7):1255-1271. doi: 10.1016/j.ajhg.2022.05.010. Epub 2022 Jun 8. PMID: 35679866; PMCID: PMC9300761.

An epigenome-wide view of osteoarthritis in primary tissues

Authors

Peter Kreitmaier, Matthew Suderman,
Lorraine Southam, ..., Caroline L. Relton,
J. Mark Wilkinson, Eleftheria Zeggini

Correspondence

j.m.wilkinson@sheffield.ac.uk (J.M.W.),
eleftheria.zeggini@helmholtz-muenchen.de (E.Z.)

CellPress

An epigenome-wide view of osteoarthritis in primary tissues

Peter Kreitmaier,[1,2] Matthew Suderman,[3] Lorraine Southam,[1] Rodrigo Coutinho de Almeida,[4] Konstantinos Hatzikotoulas,[1] Ingrid Meulenbelt,[4] Julia Steinberg,[1,5] Caroline L. Relton,[3] J. Mark Wilkinson,[6,8,*] and Eleftheria Zeggini[1,7,8,*]

Summary

Osteoarthritis is a complex degenerative joint disease. Here, we investigate matched genotype and methylation profiles of primary chondrocytes from macroscopically intact (low-grade) and degraded (high-grade) osteoarthritis cartilage and from synoviocytes collected from 98 osteoarthritis-affected individuals undergoing knee replacement surgery. We perform an epigenome-wide association study of knee cartilage degeneration and report robustly replicating methylation markers, which reveal an etiologic mechanism linked to the migration of epithelial cells. Using machine learning, we derive methylation models of cartilage degeneration, which we validate with 82% accuracy in independent data. We report a genome-wide methylation quantitative trait locus (mQTL) map of articular cartilage and synovium and identify 18 disease-grade-specific mQTLs in osteoarthritis cartilage. We resolve osteoarthritis GWAS loci through causal inference and colocalization analyses and decipher the epigenetic mechanisms that mediate the effect of genotype on disease risk. Together, our findings provide enhanced insights into epigenetic mechanisms underlying osteoarthritis in primary tissues.

Introduction

Osteoarthritis (MIM: 165720) is a complex degenerative joint disease characterized by chronic pain and stiffness. It affects more than 40% of people over the age of 70 and is a leading cause of disability worldwide.[1] In spite of its high prevalence, treatment methods are limited to pain management and total joint replacement (TJR). To drive the development of novel and personalized treatments, it is necessary to understand the genetic and genomic architecture underlying osteoarthritis. Genome-wide association studies (GWASs) have determined around 150 independent osteoarthritis-linked single-nucleotide variants.[2] For the most part, it is unknown which variants and genes at these loci are causal to disease development and along which molecular pathways they exert their osteoarthritis-promoting effect. To identify these mechanisms, studies using relevant tissues are necessary, and TJR surgeries provide an opportunity to molecularly profile relevant tissues from osteoarthritis-affected individuals.[3]

DNA methylation in promoter regions and particularly around the transcription start site is strongly associated with gene downregulation, whereas its effect in gene bodies or other regulatory regions remains less predictable. DNA methylation is dynamic, with highly tissue-specific patterns,[4] and can interact with a multitude of factors such as genotype, age, sex, or environment.[5] The methylation profiles of relevant tissues and cell types in complex diseases

can further our understanding of disease etiology, for example by generating insights into perturbed regulatory mechanisms and by revealing epigenetic markers of disease development or progression. Given the importance of tissue-specific molecular patterns, initiatives such as GTEx,[6] ENCODE,[7] ROADMAP,[8] and BLUEPRINT[9] have generated large publicly available resources that have made molecular datasets broadly accessible. However, these datasets do not include osteoarthritis-affected tissues.

To fill this gap, a small number of studies have investigated DNA methylation profiles of articular cartilage, typically comparing methylation profiles between macroscopically intact (low-grade) and degraded (high-grade) osteoarthritis cartilage to identify epigenetic markers of cartilage degeneration. Previous epigenome-wide association studies (EWASs) of this type have been limited in size, with a maximum of 17 knee osteoarthritis-affected individuals studied to date.[10–14] There is a need for better powered studies to improve our understanding of the role of DNA methylation in osteoarthritis (supplemental note S1).

Combining DNA methylation data with matched genotypes enables the detection of genetic variants associated with differential methylation levels at cytosine-guanine dinucleotides (CpGs), i.e., methylation quantitative trait loci (mQTLs). Characterizing these associations can help elucidate effector genes through which disease-associated genetic risk variants may exert their biological effect. To date, studies seeking to investigate mQTL effects in joint

tissues have mostly focused on a candidate gene[15] or single genetic variants previously linked to osteoarthritis.[16–19] One study has investigated the association of genome-wide methylation with gene expression in osteoarthritis-affected cartilage in 31 osteoarthritis-affected individuals (17 knee and 14 hip osteoarthritis patients).[11] They reported 87 methylation sites that were correlated with the expression of 70 genes, where both gene and methylation site were linked to cartilage degeneration. Of these, 36 were targeted by *cis*-mQTLs. There remains a need to comprehensively map the mQTL landscape on a genome-wide scale and, in better-powered sample sizes, to generate comprehensive insights into the interplay between genetic variation and epigenetic changes in osteoarthritis tissues, and to provide a resource to help elucidate the mechanism for novel genetic risk loci discovered in GWASs.

To date, molecular studies of osteoarthritis have mainly focused on articular cartilage, the most prominent osteoarthritis-affected tissue. However, osteoarthritis is regarded as a disease of the whole joint, affecting multiple tissues within the synovial joint. Therefore, expanding genomic analyses to include other joint tissues has the potential to reveal novel insights into disease progression. The synovium, a connective tissue that lines the joint capsule separating the synovial cavity from neighboring tissues, undergoes pathological alterations during osteoarthritis. There is well-documented evidence of synovial inflammation in osteoarthritis-affected joints, referred to as synovitis.[20] Several studies have compared mQTL effects across tissues but have focused only on specific, osteoarthritis-linked loci.[16,19,21,22]

In this study, we have analyzed genome-wide methylation profiles from up to 98 osteoarthritis-affected individuals undergoing TJR due to knee osteoarthritis (matched low-grade and high-grade cartilage and synovium). We enhance our understanding of osteoarthritis aetiopathogenesis by (1) identifying methylation markers for cartilage degeneration, (2) building machine-learning-based models to distinguish between low-grade and high-grade osteoarthritis cartilage samples, (3) determining genome-wide methylation quantitative trait loci (mQTLs) in osteoarthritis tissues (cartilage and synovium), and (4) resolving high-confidence effector genes for osteoarthritis GWAS signals.

Subjects and methods

For full details of methods, see supplemental subjects and methods.

Osteoarthritis-affected individuals and study samples
Samples from osteoarthritis-affected knees were collected in 101 osteoarthritis-affected individuals that underwent total knee replacement due to late-stage osteoarthritis. Cartilage samples were graded with the OARSI cartilage classification system (cohort1) or International Cartilage Repair Society (ICRS) scoring system (cohort2 and cohort3). This work was approved by Oxford NHS REC C (10/H0606/20 and 15/SC/0132), and samples were collected under Human Tissue Authority license 12182, Sheffield Musculoskeletal Biobank, University of Sheffield, UK. Before

participating in the study, all osteoarthritis-affected individuals provided written, informed consent.

Sample extraction
A previous study[3] reported the isolation of the chondrocytes (section "Isolation of chondrocytes"), the isolation of synoviocytes (section "Isolation of synoviocytes"), and DNA extraction (section "DNA, RNA and protein extraction") in its methods part.

DNA methylation data
Genome-wide DNA methylation was measured with the Illumina 450k or EPIC array in three sequencing batches. We used the R package minfi to read idat files.[23,24] We removed samples of three ethnicity outliers, gender mismatches (two samples), X-Y ratio outliers, and samples with unbalanced ratios between methylated and unmethylated signals (ten samples). To normalize methylation signals, we applied functional normalization.[25] We removed probes on sex chromosomes, probes with detection p values of $p > 0.01$ in more than 5% of the samples, and previously reported cross-reactive probes.[26–28] Furthermore, we excluded probes that had been reported to overlap with common genetic variants, as the signal of these probes might solely reflect genetic variation rather than true methylation signal.[26] The resulting data comprised 401,870 methylation loci and 266 samples from 98 osteoarthritis-affected individuals (56 female and 42 male patients, age range: 38–88, age mean: 69.6, age sd: 9.72, Table S1), including 98, 90, and 78 samples from low-grade osteoarthritis cartilage, high-grade osteoarthritis cartilage, and synovium, respectively. We conducted downstream statistical analyses on M values as recommended.[29]

DNA methylation data (replication set)
We used published methylation data for low-grade and high-grade osteoarthritis cartilage to replicate the findings of the EWAS and the machine-learning-based classifiers.[30]

The data is publicly available in the Gene Expression Omnibus database[31] and accessible through the entry number GEO: GSE63106. The replication data comprises methylation data of matching low-grade and high-grade osteoarthritis cartilage samples from 31 patients who underwent total joint replacement to treat primary osteoarthritis (knee: 17 osteoarthritis-affected individuals, hip: 14 osteoarthritis-affected individuals).

Genotype data
Genotypes were measured with the InfiniumCoreExome-12v1-1_A array or the InfiniumCoreExome-24v1-1_A array (supplemental subjects and methods). Genotype data were preprocessed as previously described.[3]

Sample stratification with multivariate modelling
To investigate differences between tissues on a global level, we used DNA methylation data (including 98, 90, and 78 samples from low-grade and high-grade osteoarthritis cartilage and synovium, respectively) corrected for batch effects with the ComBat function[32] from the R package sva and considered these corrected methylation. We applied (1) principal-component analysis (R function prcomp) and (2) a follow-up hierarchical clustering (R package FactoMineR).[33]

Differential methylation analysis (discovery)
To identify differentially methylated sites (DMSs) in pairs of low-grade and high-grade osteoarthritis cartilage samples from 90

osteoarthritis-affected individuals, we performed linear modeling by using the function lmFit and eBayes function of limma.[34] We added the factor variable patient ID to ensure paired analysis design and 18 surrogate variables (SVs) to account for technical confounders as covariates. To assess genome-wide significance in the EWAS, we applied Bonferroni correction considering the number of tested methylation sites: $0.05/401,870 = 1.24 \times 10^{-7}$. To identify differentially methylated regions (DMRs), we applied the R package dmrff.[35] Regions were defined as differentially methylated when composed of more than one methylation site and achieving a Bonferroni-adjusted $p < 0.05$. To identify sex-specific markers of cartilage degeneration, we used a similar approach as in the combined analysis (supplemental subjects and methods).

Differential methylation analysis (replication)

We performed an EWAS on knee samples of the replication data (17 low-grade and high-grade osteoarthritis cartilage samples, respectively) to validate our findings. To determine DMSs between low- and high-grade osteoarthritis cartilage, we applied a mixed-effect model, which is similar to what has been applied to this dataset previously[11] (supplemental subjects and methods). Replicated DMSs are defined as (1) showing the same direction of effect in the replication set (2) at nominal significance ($p < 0.05$). We performed the EWAS on a regional level in the replication dataset with dmrff (default settings analog to the discovery analysis). We defined DMRs as replicated when they are composed of exactly the same methylation sites in the replication set and show the same direction of effect on nominal significance.

Pathway enrichment analysis

We used the gometh and goregion functions (available through R package missMethyl) to identify enrichments among DMSs and DMRs.[36,37] We considered pathways consisting of between 20 and 200 genes.

Distinguishing cartilage grades with machine learning

We constructed classifiers that distinguish cartilage grades. More specifically, we trained and tested random forest (RF)-based classifiers repeatedly in 5-fold cross validations (cv) in 25 iterations (R package caret). In total, we trained and tested 125 RF models (25 iterations × 5-fold cv) (supplemental subjects and methods). To validate our approach, we trained RF-, support-vector-machine-, and gradient-boosting-machine-based classifiers on our entire dataset and tested the prediction quality of the resulting classifiers on the validation dataset. We then applied the classifiers and assessed their prediction quality separately in hip and knee samples. Prediction accuracies and their 95% confidence intervals were calculated with carets ConfusionMatrix-function.

Identification of methylation quantitative trait loci

We performed genome-wide *cis*-methylation quantitative trait locus (mQTL) analysis in low-grade (97 samples) and high-grade osteoarthritis cartilage (89 samples) as well as in synovium (78 samples), thus including only samples for which complete covariate information was available. We restricted our analyses to SNPs with a minor allele frequency > 0.05. Furthermore, we defined the *cis*-distance with 1 Mb. We conducted the mQTL analysis by using the R package MatrixEQTL.[38] We applied linear models and corrected for age, sex, and batch effects (supplemental subjects and methods). We defined two thresholds to identify genome-wide-significant methylation QTL effects.

(1) Bonferroni threshold: genome-wide significance defined by $p < 0.05/$number of tested SNP-methylation site pairs (low-grade osteoarthritis cartilage: $p < 3.05 \times 10^{-11}$, high-grade osteoarthritis cartilage: $p < 3.03 \times 10^{-11}$, synovium: $p < 3.03 \times 10^{-11}$).

(2) False discovery rate (FDR): we estimated the FDR of mQTL effects by using the MatrixEQTL package. It calculates the FDR considering the total number of tested *cis*-pairs per tissue.

To characterize mQTL architecture in osteoarthritis tissues, we used methylation site annotations of Illumina's annotation file (version 1.2). For the enrichment approaches, we applied hypergeometric tests (R function phyper). To identify sex-specific *cis*-mQTLs, we applied MatrixEQTL by using an interaction model (supplemental subjects and methods).

Differential mQTL effects in low-grade and high-grade osteoarthritis cartilage

To calculate differential mQTL effects between low-grade and high-grade osteoarthritis cartilage, we used the software MetaTissue v0.5 (see web resources).[39] Analogously to our genome-wide, tissue-specific approach to identify mQTLs, we included sex, age, and sequencing batches as covariates in these models. We used the MetaTissue software to calculate posterior probabilities (m values) and focused on genetic variant-methylation site pairs with a significant effect in one tissue (m value > 0.9) but not in the other (m value < 0.1).

Comparing joint with whole blood methylation QTLs

We compared mQTL effects (Bonferroni correction) of joint tissues (low-grade osteoarthritis cartilage, high-grade osteoarthritis cartilage, and synovium) with the corresponding effects (mQTL effect between the same variant-methylation site pairs) of a mQTL meta-analysis (Genetics of DNA Methylation Consortium, see web resources) of 36 cohorts in whole blood.[40] We considered results from the fixed-effect models from the whole blood mQTL meta-analysis.

Summary statistics of GWASs

For the MR approach and the colocalization analysis, we included summary statistics from three osteoarthritis-related phenotypes: (1) osteoarthritis at any site (all OA) and (2) knee osteoarthritis (knee OA) and (3) total knee replacement (TKR). Summary statistics for all OA and knee OA were previously published[41] and downloaded from the GWAS Catalog. We calculated summary statistics for TKR by meta-analyzing the arcOGEN and UKBB data with the METAL software.[42]

Two-sample Mendelian randomization

To estimate putative causal effects of methylation, we applied two-sample Mendelian randomization (2SMR) by integrating mQTL data of the three examined joint tissues and GWAS data from three osteoarthritis traits (all OA, knee OA, and TKR). We performed 2SMR following the workflow implemented in the R package TwoSampleMR (version 0.4.25).[43] In low-grade osteoarthritis cartilage, we tested 3,378 methylation sites for their putative causal effect on osteoarthritis (all OA = 3,378 methylation sites, knee OA = 3,378, and TKR = 3,343). In high-grade osteoarthritis cartilage, we considered 2,042 methylation sites (all OA = 2,042, knee OA = 2,042, and TKR = 2,026). In synovium, we investigated the effect of 1,561 methylation sites (all OA = 1,560, knee OA = 1,560, and TKR = 1,542). In total, we tested 10,099, 6,110, and 4,662

methylation site-osteoarthritis trait combinations in low-grade and high-grade osteoarthritis cartilage and synovium, respectively. Per tissue, we applied the Bonferroni method to correct for the number of performed tests (low-grade osteoarthritis cartilage: $p < 4.95 \times 10^{-6}$, high-grade osteoarthritis cartilage: $p < 8.18 \times 10^{-6}$, synovium: $p < 1.07 \times 10^{-5}$).

We investigated the opposite direction of effect (osteoarthritis causal for methylation changes) for every tested methylation site-osteoarthritis trait combination (using R package TwoSampleMR). We used 27, 10, and 4 SNPs as instrumental variable (IV) for all OA, knee OA, and TKR, respectively. Here, we applied the inverse-variance-weighted (IVW) method.

Colocalization analysis

We applied colocalization analysis to statistically estimate the overlap of mQTL signals in the three osteoarthritis tissues and GWAS signals.[44] We examined genome-wide signals for osteoarthritis at any site (all OA, 33 risk loci), knee osteoarthritis (knee OA, 12 risk loci), and total knee replacement (TKR, 5 risk loci). We performed colocalization by applying coloc.fast function (web resources). We conducted the colocalization analysis separately for each GWAS osteoarthritis trait and each tissue (supplemental subjects and methods). We used a posterior probability threshold for having a shared causal variant ("PP4") of ≥80% (thus indicating colocalization) as previously applied.[3] Annotated genes and locations of colocalized GWAS signals were extracted from Ensembl Variant Effect Predictor (see web resources).

Combining colocalization results with eQTL and gene expression data

We combined these colocalization results with previously estimated eQTL data from the same patient cohort.[3] More specifically, we tested whether the lead SNP of colocalized GWAS OA signals show an eQTL effect on local genes at nominal significance ($p < 0.05$). We used previously published, matching expression data (low-grade osteoarthritis cartilage: 75 osteoarthritis-affected individuals, high-grade osteoarthritis cartilage: 76, synovium: 70)[3] of the same osteoarthritis-affected individuals in the same tissue types to test associations between osteoarthritis-linked methylation sites and genes in the same region. We estimated associations between methylation and gene expression by using linear models. We estimated putative causal effects of methylation on gene expression by using one-sample MR with the R package ivreg (supplemental subjects and methods).

Comparative analysis of colocalization in joint and blood

We tested whether osteoarthritis-risk variant-methylation site pairs that colocalize using joint mQTL data also colocalize when overlapping osteoarthritis GWASs with whole blood mQTL data. For this colocalization approach, we applied the same colocalization method as performed on joint mQTL data. We applied a threshold of PP4 ≥ 80% and PP4 < 20% indicating colocalization and no colocalization, respectively.

Results

Methylation profiles differ between tissue types and disease grades

To describe distinct methylation profiles in three tissue types (low-grade and high-grade osteoarthritis cartilage and synovium), we first assessed whether tissue and osteoarthritis grade have strong, systematic effects on global variations in the epigenome. We used principal-component analysis to examine variation in global methylation profiles and observed a clear separation between synovium and cartilage samples along the first principal component (PC) and partly overlapping clustering between low-grade and high-grade osteoarthritis cartilage along the second PC (Figure 1A). A linear model confirmed the significant association between the second PC and cartilage grades ($p = 1.51 \times 10^{-16}$, beta = 253.79, SE = 27.93). Using a hierarchical clustering approach, we observed stratification by tissue type and cartilage degradation state (Figure 1B).

EWAS reveals widespread, robustly replicating signals

To identify DNA methylation markers of cartilage degeneration, we performed an EWAS on paired low-grade and high-grade osteoarthritis cartilage samples from 90 osteoarthritis-affected individuals across 401,870 methylation sites (supplemental note S2). We identified 15,328 differentially methylated sites (DMSs) distributed across the whole genome (Figures 2A [upper panel], 2B, and 2C and Table S2) by using a significance threshold of $p < 1.24 \times 10^{-7}$ (subjects and methods). Furthermore, we identified 2,477 differentially (Bonferroni-adjusted $p < 0.05$) methylated regions (DMRs) (Figure 2A [bottom panel], Figure S1, and Table S3).

To biologically characterize the DMS, we performed enrichment analyses and identified 29 and 4 Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) terms, respectively (Figure 2D, Figure S2, and Table S4), including pathways linked to osteoarthritis, e.g., terms associated with external matrix organization[45] and skeletal system development,[10,13,14,45] as well as the epithelium-related term "positive regulation of epithelial cell migration" in articular cartilage. This term showed limited overlap with other enriched pathways on the constituent gene level (e.g., extracellular matrix structural constituent: two of 93 annotated, differentially methylated genes are also annotated to "positive regulation of epithelial cell migration," collagen fibril organization: one of 34, integrin-mediated signaling pathway: 11 of 58, cartilage development: seven of 104, chondrocyte differentiation: five of 62) suggesting its distinctness, e.g., to pathways that are linked to the extracellular matrix or cartilage development. This pathway may point to an epithelium-related etiological mechanism.

We used an independent dataset from 17 knee osteoarthritis patients to replicate the epigenetic differences between low-grade and high-grade osteoarthritis cartilage.[11,30] We replicated 7,192 DMSs and 105 DMRs (Tables S2, S3, and S5). The effect sizes of replicated DMSs (Pearson r = 0.96, $p < 2.2 \times 10^{-16}$) and DMRs (Pearson r = 0.95, p value $< 2.2 \times 10^{-16}$) in the discovery and replication datasets were highly correlated (Figures 3A and 3B). These results point to the robustness of the identified methylation changes.

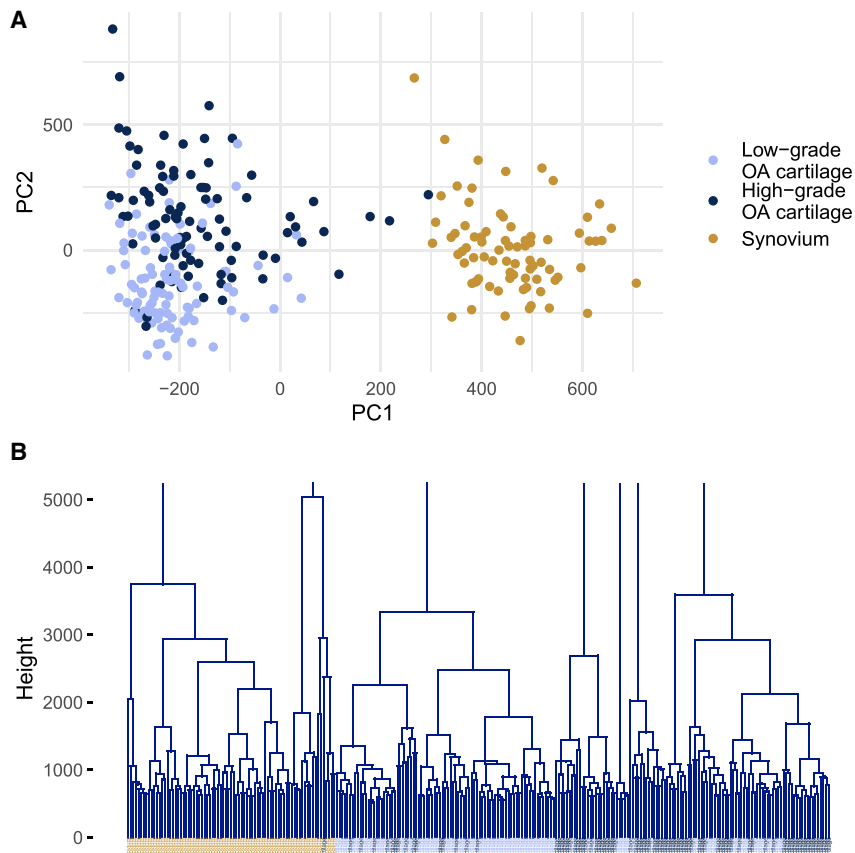We further performed EWAS separately on paired low-grade and high-grade cartilage in female (n = 52) and

Figure 1. Multivariate analyses of methylation profiles distinguish between different tissues and disease grades

(A) In a principal-component analysis, the first PC separates cartilage from synovium, while the second PC is associated with cartilage grades (with overlapping clusters from low-grade and high-grade osteoarthritis cartilage samples).

(B) Hierarchical clustering shows a separation of global methylation profiles by tissue type. "Height" on the y axis denotes the distance between clusters. OA denotes osteoarthritis.

male (n = 38) osteoarthritis-affected individuals and identified female- (n = 1,338) and male- (n = 3,316) specific DMSs in cartilage, suggesting sex-specific markers (supplemental note S2).

Machine-learning models distinguish cartilage grades with high accuracy

Next, we sought to test whether epigenetic changes in different cartilage grades can be harnessed to develop a model that robustly distinguishes low-grade from high-grade osteoarthritis cartilage. First, we constructed RF-based classifiers in the discovery knee osteoarthritis-affected individual cohort in a repeated 5-fold cross-validation approach. Here, we achieved high prediction accuracies (mean accuracy: 90.69%; standard deviation: 4.08, 95% confidence interval [CI] 89.98–91.41). Furthermore, the resulting receiver operating characteristic (ROC) curve revealed an area under the curve of 0.97 (Figure S3), highlighting the high sensitivity and specificity of these classifiers.

To validate these findings, we trained the final RF-based classifier on our entire patient cohort (subjects and methods) and evaluated its accuracy in an external dataset composed of 17 knee and 14 hip osteoarthritis-affected individuals.[30] In this replication cohort, we achieved an accuracy of 82.35% (95% CI 65.47–93.24) for knee samples, whereas in hip samples the achieved accuracy was lower at 64.29% (95% CI 44.07–81.36). We also observed these

differences when using support vector machines (knee: 85.29%, 95% CI 68.94–95.05; hip: 57.14%, 95% CI 37.18–75.54) and gradient-boosting machines (knee: 76.47%, 95% CI 58.83–89.25; hip: 50.00%, 95% CI 30.65–69.35). The lower accuracy achieved in hip samples supports the effect of methylation joint specificity within osteoarthritis.[30,45] GO enrichment analysis of the 300 most important methylation sites in the final RF model did not identify any significant enrichments. Of these 300 methylation sites, 99.3% (n = 298) and 77.7% (n = 233) were among the DMS identified in the discovery and replication analysis, respectively. This suggests that epigenetic markers for cartilage degeneration are prioritized predictors in the classifier. External validation of the classifier was somewhat limited by the small sample size of the replication set, resulting in wide confidence intervals. Hence, validation in larger datasets is further warranted.

This model shows that epigenetic differences can be used to distinguish disease stages in cartilage. Samples of more accessible tissue types (such as blood and synovial fluid) need to be included in the model training and testing to develop a clinically relevant tool.

Genome-wide mQTL maps in osteoarthritis-relevant tissues

We combined DNA methylation data with matching genotype data from the same osteoarthritis-affected individuals to identify genetic variants that are significantly associated with methylation levels of proximal methylation sites (<1 Mb; cis-mQTLs). We performed this analysis at the genome-wide scale in low-grade (n = 97) and high-grade (n = 89) osteoarthritis cartilage samples as well as in synovium (n = 78 samples), and identified widespread signal in every tissue (Figure 4, Figure S4, and Table S6). Applying a conservative Bonferroni threshold to correct for the number of tested genetic variant-methylation site pairs per tissue ($p < 1 \times 10^{-11}$), we identified 10,639, 6,785, and 4,493
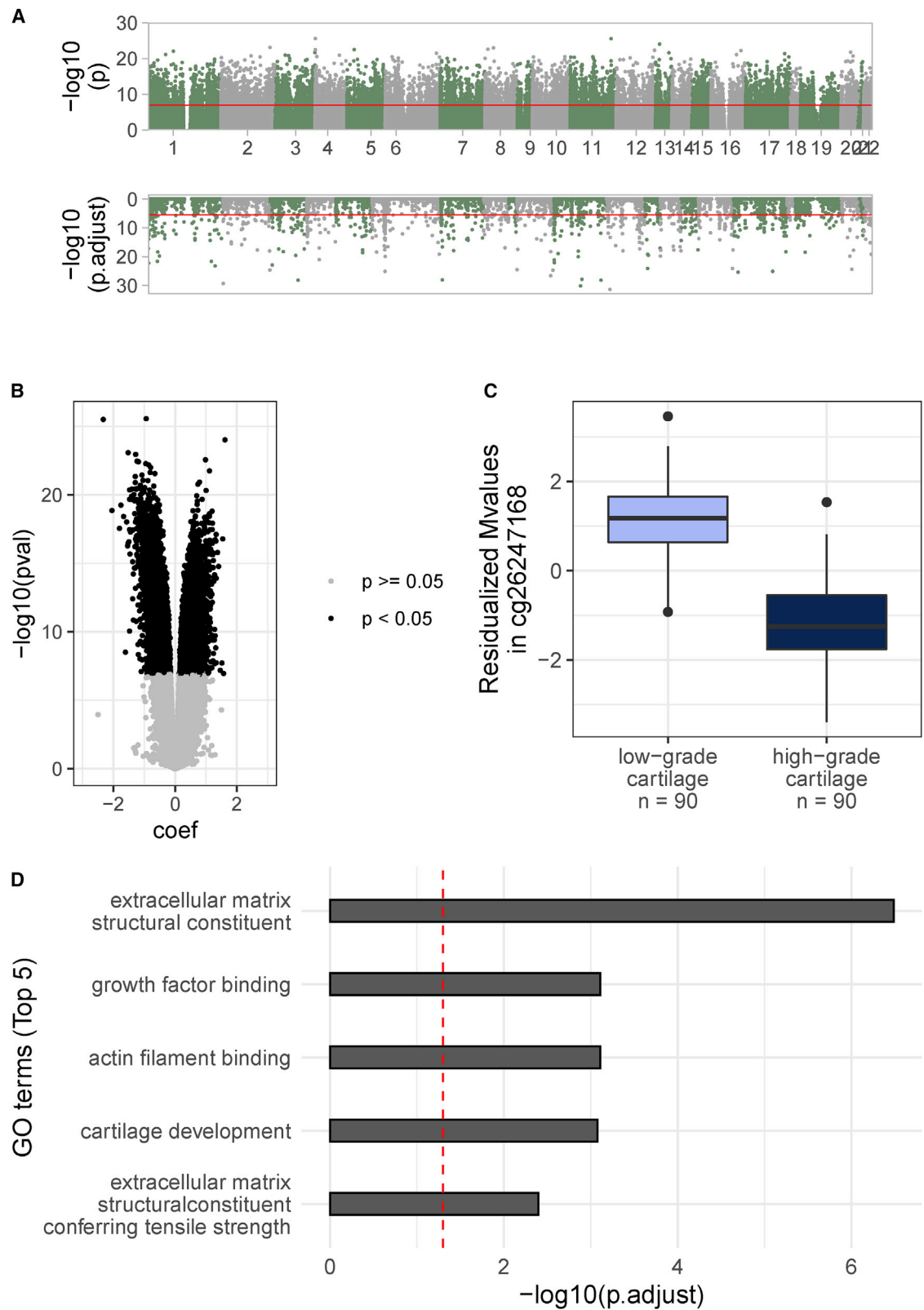
Figure 2. Differential methylation between low-grade and high-grade osteoarthritis cartilage

(A) Genome-wide signals for differential methylation sites (top) and regions (bottom) between low-grade and high-grade osteoarthritis cartilage. Red lines indicate genome-wide significance (top: nominal $p < 1.24 \times 10^{-7}$, bottom: Bonferroni-adjusted $p < 0.05$).

(B) Volcano plot showing hyper- and hypomethylated sites.

(legend continued on next page)

methylation sites significantly associated with at least one mQTL in low-grade osteoarthritis cartilage, high-grade osteoarthritis cartilage, and synovium, respectively, and also included genetic-epigenetic effects in loci of previously reported mQTLs in osteoarthritis-relevant tissue (supplemental note S3). This represents a genome-wide map of mQTLs in osteoarthritis tissues. These data are made publicly available (see data and code availability).

Next, we further characterized the architecture of these mQTL maps. In low-grade and high-grade osteoarthritis cartilage, 66.93% (7,121 of 10,639) and 66.44% (4,508 of 6,785) methylation sites with at least one mQTL were annotated to a gene, respectively (Figures S5 and S6). In both cartilage tissue types, we identified significant (Bonferroni p < 0.05) over-representation of intergenic methylation sites or sites within gene bodies and under-representation of methylation sites close to transcription start sites ("TSS200," "TSS1500"), in untranslated regions ("3′ UTR," "5′ UTR") and first exons ("1st exon").

In synovium, 67.44% (n = 3,030 of 4,493) methylation sites with at least one mQTL were annotated to a gene (Figure S7). Here, we found significant (Bonferroni p < 0.05) over-representation of intergenic methylation sites and under-representation of methylation sites that are within 200 bp to a transcription start site or in untranslated regions or first exons. These results suggest similar mQTL architectures across osteoarthritis tissues.

Furthermore, we tested whether mQTL effects differ between osteoarthritis-affected individuals of different sexes and identified methylation sites targeted by sex-specific mQTLs (FDR < 0.05) in low-grade (n = 282) and high-grade (n = 337) osteoarthritis cartilage as well as in synovium (n = 874) (Figure S8 and Tables S7, S8, and S9). This suggests sex-specific genetic effects on methylation in osteoarthritis tissues.

Comparing mQTLs in cartilage and synovium with whole blood

Next, we asked whether the mQTL profiles of primary osteoarthritis tissues differ to those of more easily accessible, peripheral tissue samples. We compared the cis-mQTL effects of each of the three examined joint tissues with those of whole blood, which is the most commonly examined tissue type for DNA methylation. To maximize the number of identifiable osteoarthritis-tissue-specific effects, we compared cis-mQTL effects in joint-tissue to those of a publicly available, large-scale whole blood meta-analysis including 36 studies (27,750 European ancestry participants).[40]

Because a mQTL can be associated with more than one methylation site (and vice versa), we use the term "mQTL-site pair" to indicate the association between a spe-

cific mQTL and a specific methylation site. Of the 482,751 mQTL-site pairs in low-grade osteoarthritis cartilage, information of 365,411 were available in whole blood. Of these, 88.6% (n = 323,863) were significant in blood (p < 10^{-11}) with a concordant direction of effect. Notably, 9.61% of overlapping mQTL-site pairs (n = 35,117) were both significant and had an opposite direction of effect in blood. Similarly, we compared 219,661 (of 286,558) mQTL-sites pair identified in high-grade osteoarthritis cartilage and found that 90.53% (n = 198,867) of these had a significant (p < 10^{-11}) effect in the same direction in blood. Notably, 7.87% (n = 17,297) of present mQTL-site pair had a significant but opposing direction of effect in blood. In synovium, 78.88% (n = 156,931) of 198,958 mQTL-site pair were available in whole blood, of which 96.87% (n = 152,023) had a significant mQTL effect in blood (p < 10^{-11}) in the same direction. 2.29% (n = 3,594) of overlapping mQTL-site pairs showed a significant effect in the opposite direction in blood. In summary, we found that the majority of mQTL effects identified in osteoarthritis-related tissues show the same direction in whole blood but also observed effects in opposing directions in all tested joint tissue types. The latter indicates non-negligible differences in the mQTL profile between osteoarthritis-relevant joint tissues and whole blood.

Identification of grade-specific mQTLs in cartilage

We compared mQTL effects between the osteoarthritis joint tissues. Comparing low-grade cartilage with synovium, we found 143,258 mQTL-site pairs to be significant in both. The effects showed high correlation (Pearson r = 0.97, p < 10^{-16}) and 33 mQTL-site pairs showed an effect in the opposite direction. The effects of the 122,378 mQTL-site pairs that were significant in both high-grade cartilage and synovium also showed high correlation (Pearson r = 0.97, p < 2.2 × 10^{-16}), and their effect directions were all concordant. In low-grade and high-grade osteoarthritis cartilage, the effect sizes of mQTL-site pairs were highly correlated (comparing 256,036 mQTL-site pairs that were significant in both low-grade and high-grade osteoarthritis cartilage: Pearson r = 0.99, p < 2.2 × 10^{-16}) and showed only concordant effect directions. Overall, our findings point to broadly concordant mQTL effects across osteoarthritis tissues.

We subsequently sought to identify differential mQTLs, i.e., mQTLs that are present in either low-grade or high-grade cartilage but not in both. This can help identify mQTL effects that are potentially "switched on/off" with increasing cartilage degeneration grade, i.e., with disease stage. To this end, we applied a meta-analysis approach,[39] which improves power in identifying differential mQTLs by estimating a posterior probability of >0.9 and <0.1

(C) An example of hypomethylation in high-grade osteoarthritis cartilage at cg26247168 (beta: −2.32, p = 3.05 × 10^{-26}, SE = 0.14). The boxplots represent 25th, 50th, and 75th percentiles, and whiskers extend to 1.5 times the interquartile range.
(D) Most significant Gene Ontology gene annotations enriched in 15,328 DMSs. Red dashed lines indicate the significance threshold (Benjamini-Hochberg-adjusted p < 0.05).
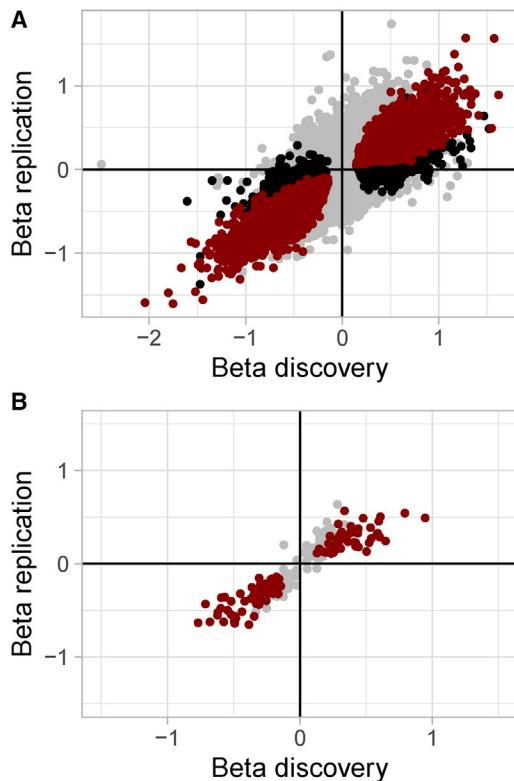
Figure 3. Replication of EWAS results in an independent dataset
(A and B) Effects of (A) methylation sites (n = 346,288) and (B) methylation regions (n = 271) present in both of the discovery and replication datasets. Black dots refer to DMSs/DMRs of the discovery set, red dots to DMSs/DMRs that additionally show an effect at nominal significance (nominal p < 0.05) in concordant direction in the replication set.

indicating the presence and absence of a mQTL effect, respectively. In total, we identified 195 genetic variants that show a differential mQTL effect on 18 methylation sites (Table S10). Following clumping, one independent differential mQTL was retained per methylation site (Figure 5). Of the 18 targeted methylation sites, 14 and 4 were mQTLs in low-grade and high-grade osteoarthritis cartilage only, respectively. Genes annotated to these methylation sites are linked to osteoarthritis-relevant terms in cartilage, e.g., they encode a matrix metalloproteinase (*MMEL1*) or are involved in cell adhesion (*CDH23* and *PARVA*).

Assessing the causal role of methylation in osteoarthritis
To identify methylation sites that play a causal role in osteoarthritis progression, we applied two-sample Mendelian randomization (MR) to the methylation sites associated with cartilage degeneration (exposure) and the mQTLs we identified in osteoarthritis-relevant tissues, together with genetic associations from three GWASs: knee osteoarthritis (knee OA), osteoarthritis at any site (all OA), and total knee replacement (TKR). We used the mQTLs as instrumental variables (Figure S9) in the MR analysis.

We identified 6, 8, and 11 significant osteoarthritis trait-methylation site combinations in low-grade and high-

grade osteoarthritis cartilage and synovium, respectively. When performing an MR approach to examine causality in the opposite direction, namely the effect of osteoarthritis on methylation (Figure S10), we could not find any evidence for a significant effect for these osteoarthritis trait-methylation site combinations, thus providing further evidence for the causal role of these methylation sites on osteoarthritis (and not vice versa).

In total, we identified 19 methylation sites with a putative causal effect on osteoarthritis (Figure 6 and Table S11). In low-grade osteoarthritis cartilage, we identified six methylation sites with a potential causal effect (Bonferroni correction, $p < 4.95 \times 10^{-6}$). Four of these showed association with hypermethylation, and two showed associations with hypomethylation and osteoarthritis development. Among the annotated genes is *WWP2* (cg26736200 in gene body), a key regulator in chondrocytes (discussion).

In high-grade cartilage, eight methylation sites were causally linked to osteoarthritis (Bonferroni correction, $p < 8.18 \times 10^{-6}$). Of these, five sites showed association of hypermethylation with a protective effect against osteoarthritis development, whereas the other three sites were associated with higher risk. Annotated genes include *COLGALT2* (cg18131582 in gene body), a transferase that catalyzes the transfer of galactose to collagen during collagen synthesis.[46] A previous study suggests that the expression of this gene in cartilage is influenced by an osteoarthritis-risk variant.[47]

In synovium, we identified 11 significant methylation site-trait combinations, involving eight unique methylation sites ($p < 1.07 \times 10^{-5}$). In five of these eight sites, increased methylation levels showed a protective effect against osteoarthritis development, whereas in three sites hypermethylation was associated with higher risk. Annotated genes include *MFHAS1* (cg01784220 in the 1st exon), a gene involved in Toll-like receptor signaling,[48,49] which is thought to be centrally involved in the osteoarthritis-related immune response in synovial joints.[50]

We identified one methylation site (cg26736200) in low-grade osteoarthritis cartilage and two methylation sites (cg17551891 and cg00076555) in high-grade osteoarthritis cartilage that were also identified as potentially causal for osteoarthritis in the synovium. For these three methylation sites, the direction of effect was concordant across tissues. Cg26736200 is annotated to the gene body of *WWP2*. Cg1755189 is located in the gene body of *MAD1L1*, a gene involved in cell-cycle regulation, which may point to cell senescence of chondrocytes in osteoarthritis articular cartilage.[51] Cg00076555 is located in the 3′ UTR of *BSN* (discussion).

Resolution of GWAS signals
We performed a colocalization analysis to determine whether osteoarthritis-linked genetic risk variants exert their effect through the regulation of nearby methylation sites. For all OA, 13 of 33 tested GWAS signals colocalized
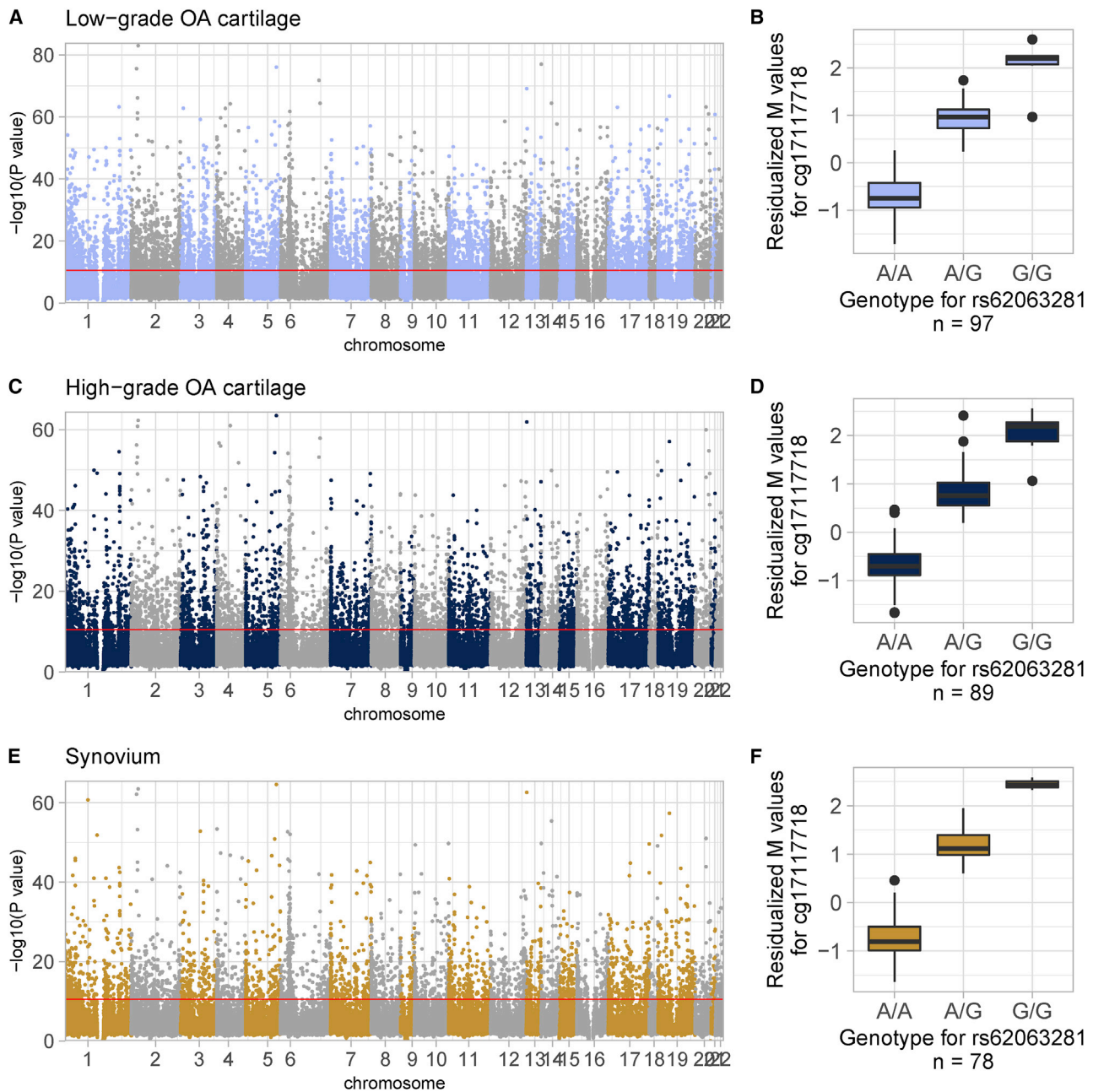
Figure 4. The mQTL landscape in cartilage and synovium

(A–F) Manhattan plots depicting the negative log of the p value of the most significant association per methylation site across all variants within 1 Mb in (A) low-grade osteoarthritis cartilage, (C) high-grade osteoarthritis cartilage, and (E) synovium. Red lines indicate genome-wide significance (Bonferroni correction). The boxplots describe the effect of rs62063281 on methylation site cg17117718 in (B) low-grade osteoarthritis cartilage (beta = 1.65, p = 1.19 × 10^{-48}, SE = 0.06), (D) high-grade osteoarthritis cartilage (beta = 1.60, p = 3.55 × 10^{-37}, SE = 0.07), and (F) synovium (beta = 1.95, p = 2.39 × 10^{-44}, SE = 0.06), as an example. The boxplots represent 25th, 50th, and 75th percentiles, and whiskers extend to 1.5 times the interquartile range.

with mQTLs (ten in low-grade osteoarthritis cartilage, seven in high-grade osteoarthritis cartilage, and six in synovium; example in Figure 7B). For knee OA, six of 12 tested GWAS signals colocalized with mQTL signals (five in low-grade osteoarthritis cartilage, four in high-grade osteoarthritis cartilage, and four in synovium; example in Figure 7A). For TKR, one of five tested GWAS signals colocalized with mQTL signals (in low-grade osteoarthritis cartilage). Overall, osteoarthritis-related GWAS signals co-

localized with mQTL signals of 32 unique methylation sites in low-grade osteoarthritis cartilage, 29 in high-grade osteoarthritis cartilage, and 17 in synovium. In total, we colocalized mQTL signals of 56 unique methylation sites with osteoarthritis-risk variants across the three affected individual tissues (Table S12).

By comparing the findings from colocalization and causal inference analysis (in the previous section), we identified two methylation sites in low-grade osteoarthritis

| variant	Gene	Msite	Gene (Msite)	Beta (L−G)	Beta Standard error (L−G)	Posterior Prob (L−G)	Beta (H−G)	Beta Standard error (H−G)	Posterior Prob (H−G)	Tissue with QTL effect
rs62175812	CERS6	cg04746649	LASS6	−0.927	0.157	1.000	−0.082	0.151	0.004	L−G
rs10055514	TMED7	cg05930586		0.893	0.139	1.000	−0.012	0.130	0.000	L−G
rs793442	FILIP1L CMSS1	cg06077670	C3orf26 FILIP1L MIR548G	0.886	0.140	1.000	−0.086	0.130	0.000	L−G
rs13148169	intergenic	cg06759629		0.998	0.159	1.000	0.187	0.154	0.090	L−G
rs2072184	TWISTNB MIR3146	cg07541023	TWISTNB	0.924	0.150	1.000	0.017	0.145	0.001	L−G
rs2956078	APIP	cg11058730	PDHX APIP	0.845	0.132	1.000	0.161	0.126	0.021	L−G
rs56918106	SEMA3D	cg11878016		1.016	0.151	1.000	0.248	0.137	0.033	L−G
rs4789346	MXRA7	cg13533061	JMJD6	−0.872	0.133	1.000	−0.186	0.120	0.052	L−G
rs10760430	MVB12B NRON	cg13960660	FAM125B	1.053	0.170	1.000	0.181	0.158	0.089	L−G
rs3823467	LINC00574 LINC00242	cg16413842	C6orf122 C6orf208	1.136	0.186	1.000	0.172	0.182	0.062	L−G
rs4988988	RP11−219G17.4 THA1P	cg16810279		0.987	0.158	1.000	0.224	0.151	0.095	L−G
rs4708476	SMOC2	cg18005896	SMOC2	−1.006	0.133	1.000	−0.261	0.121	0.057	L−G
rs79031158	TTC34	cg21389723	MMEL1	0.763	0.140	1.000	0.081	0.138	0.016	L−G
rs2104958	intergenic	cg25193276		−0.990	0.154	1.000	−0.216	0.144	0.080	L−G
rs113856858	SGPL1	cg09598552	CDH23	0.200	0.280	0.003	−1.978	0.328	1.000	H−G
rs7481217	TEAD1	cg18778433	PARVA	0.211	0.242	0.044	1.812	0.298	1.000	H−G
rs2095124	RASSF5 RP11−534L20.5	cg19452316	RASSF5	0.279	0.167	0.074	1.257	0.194	1.000	H−G
rs73088790	WDR82	cg23168339	WDR82	−0.007	0.228	0.001	−1.868	0.286	1.000	H−G
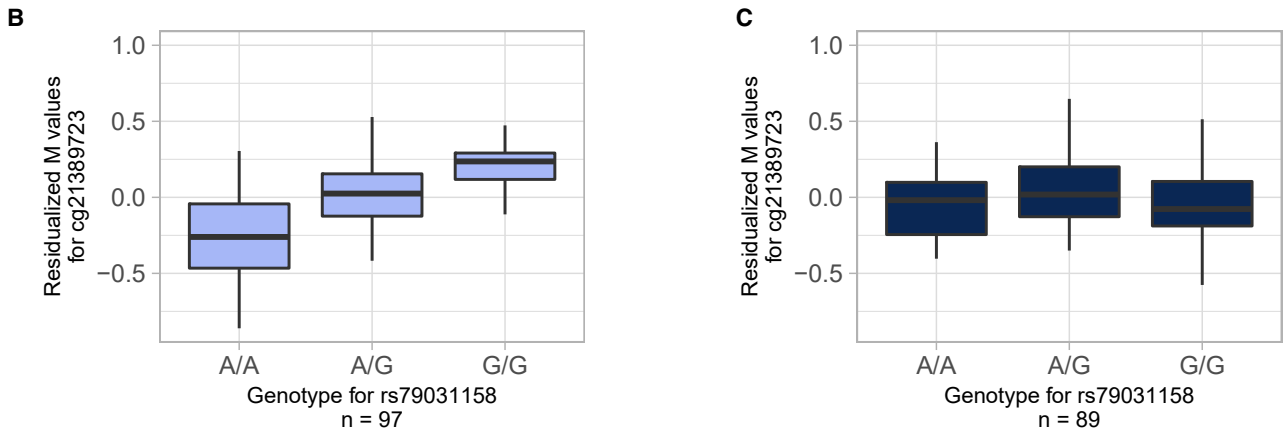
Figure 5. Differential mQTLs

(A–C) Each row refers to a variant with a differential mQTL effect. The table reports the genetic variant and the targeted methylation site as well as annotated genes, effect sizes with corresponding standard errors, and posterior probabilities (>0.9 indicate an effect, <0.1 indicate no effect) for the effects in low-grade or high-grade osteoarthritis cartilage. The reported effects were estimated by a meta-analysis approach (subjects and methods). Boxplots (B) and (C) exemplify a differential mQTL: rs79031158 is associated with methylation of cg21389723 in low-grade (B) but not in high-grade osteoarthritis cartilage (C). The boxplots represent 25th, 50th, and 75th percentiles, and whiskers extend to 1.5 times the interquartile range. Msite, methylation site; L-G, low-grade osteoarthritis cartilage; H-G, high-grade osteoarthritis cartilage; Posterior Prob, posterior probability.

cartilage (cg17125990 and cg26736200) and one methylation site in synovium (cg26736200) across both approaches, providing further evidence that these methylation sites play a causal role in osteoarthritis in the respective joint tissue.

Next, we combined these findings with results from eQTL data[3] generated in the same patient cohort. When osteoarthritis GWAS signals colocalized with mQTL data, we tested whether the GWAS signal index variant exerted an effect on the expression levels of any gene close to the relevant methylation site. We found such an eQTL effect below nominal significance levels for five genes in low-grade osteoarthritis cartilage (ALDH1A2, CHMP1A, FAM53A, RPP25, and TGFA), two genes in high-grade
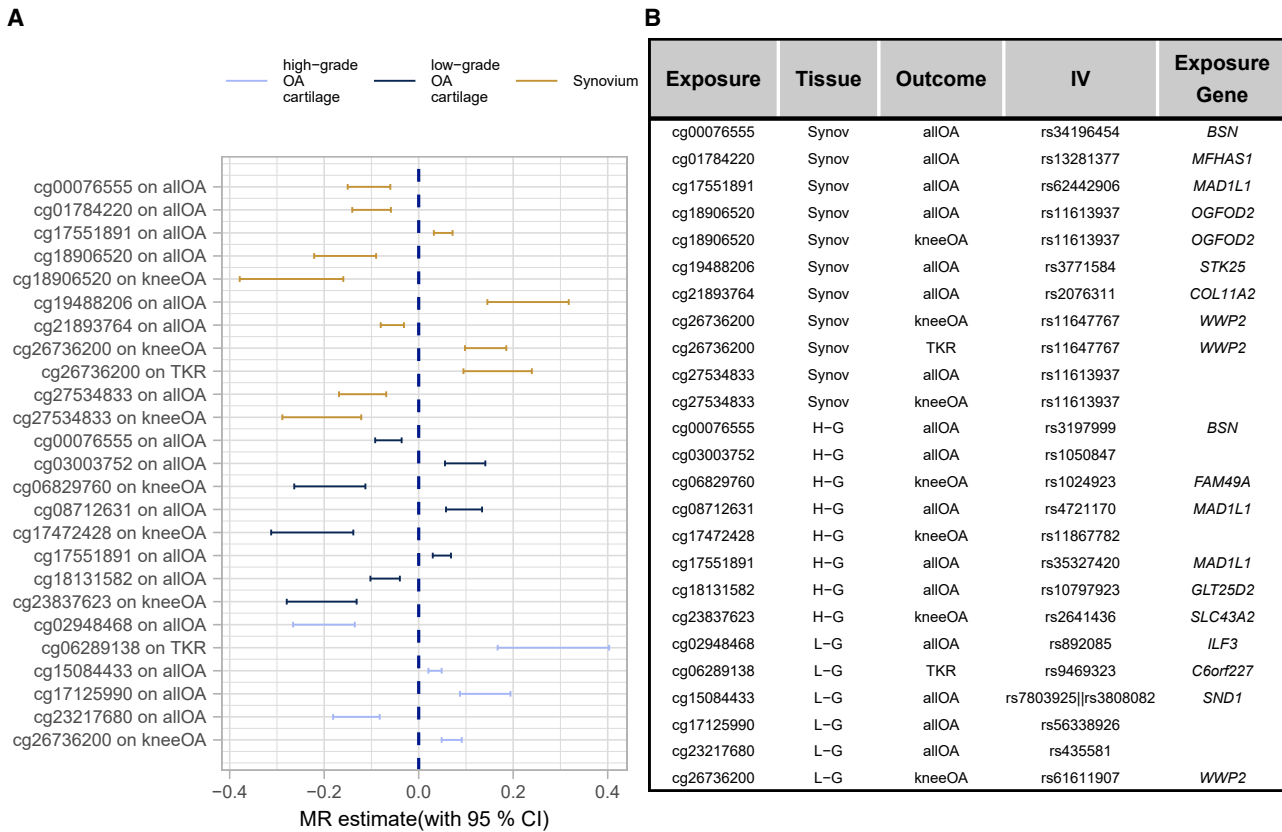
Figure 6. Overview of putative causal effects of methylation on osteoarthritis-related traits
(A and B) Forest plot (A) describing the putative causal effect (with 95% confidence interval) of increasing methylation levels in the respective sites on osteoarthritis-related traits. Only significant exposure-outcome associations exceeding tissue-specific Bonferroni thresholds are reported (low-grade osteoarthritis cartilage: $p < 4.95 \times 10^{-6}$, high-grade osteoarthritis cartilage: $p < 8.18 \times 10^{-6}$, synovium: $p < 1.05 \times 10^{-5}$). The table (B) reports the instrumental variable(s) (IV[s]) and annotated genes. We applied the Wald-ratio test in cases of one IV; otherwise the inverse-variance-weighted method was applied.

osteoarthritis cartilage (*FAM53A* and *LTBP1*), and one gene in synovium (*CRADD*) (Figure 7C). In total, we identified seven genes linked to an osteoarthritis-risk locus. Given their link to osteoarthritis-risk variants across two molecular layers, these genes are high-confidence effector genes at these osteoarthritis GWAS loci in the respective tissue.

We compared these results with findings from a recent differential expression analysis.[3] Two high-confidence effector genes were shown to be differentially expressed in high-grade compared to low-grade osteoarthritis cartilage (in high-grade osteoarthritis cartilage, *ALDH1A2* is overexpressed with FDR = 0.0017 and logFC = 0.38 and *CRADD* is underexpressed with FDR = 0.00067 and logFC = −0.24), thus providing additional supportive evidence for a role in osteoarthritis.

Next, we tested whether high-confidence effector genes correlate with nearby methylation sites, which in turn putatively mediate the effect of osteoarthritis-risk variants. Using expression and methylation data of the same osteoarthritis-affected individuals in the same tissue, we identified such expression quantitative trait methylation (eQTM) effects at nominal significance (p < 0.05) for three genes (*ALDH1A2*, *FAM53A*, and *RPP25*) in low-grade cartilage and one gene (*LTBP1*) in high-grade cartilage

(Figure 7C). To assess whether these observed associations are solely correlations, or whether methylation levels do have a causal effect on gene expression (by mediating the genetic effect on gene expression), we performed one-sample MR (supplemental subjects and methods). We found evidence (MR p < 0.05) for a causal effect of methylation on gene expression levels for two genes (*ALDH1A2* and *RPP25*) in low-grade osteoarthritis cartilage and one gene (*LTBP1*) in high-grade osteoarthritis cartilage (Table S13). These findings suggest that methylation mediates the effect of genetic variants on expression for these high-confidence effector genes.

Comparing colocalization of osteoarthritis loci in joint and whole blood mQTL data

To investigate the joint tissue specificity of colocalizing joint mQTL and osteoarthritis GWAS data, we asked whether these results could also be identified in whole blood (supplemental note S4). This would allow us to better understand whether the regulatory effects of osteoarthritis-risk loci mediated by proximal methylation sites are exclusive to disease-affected joint tissues or also observed in peripheral tissues. We tested whether the pairs of risk variant-methylation sites that colocalize in at least
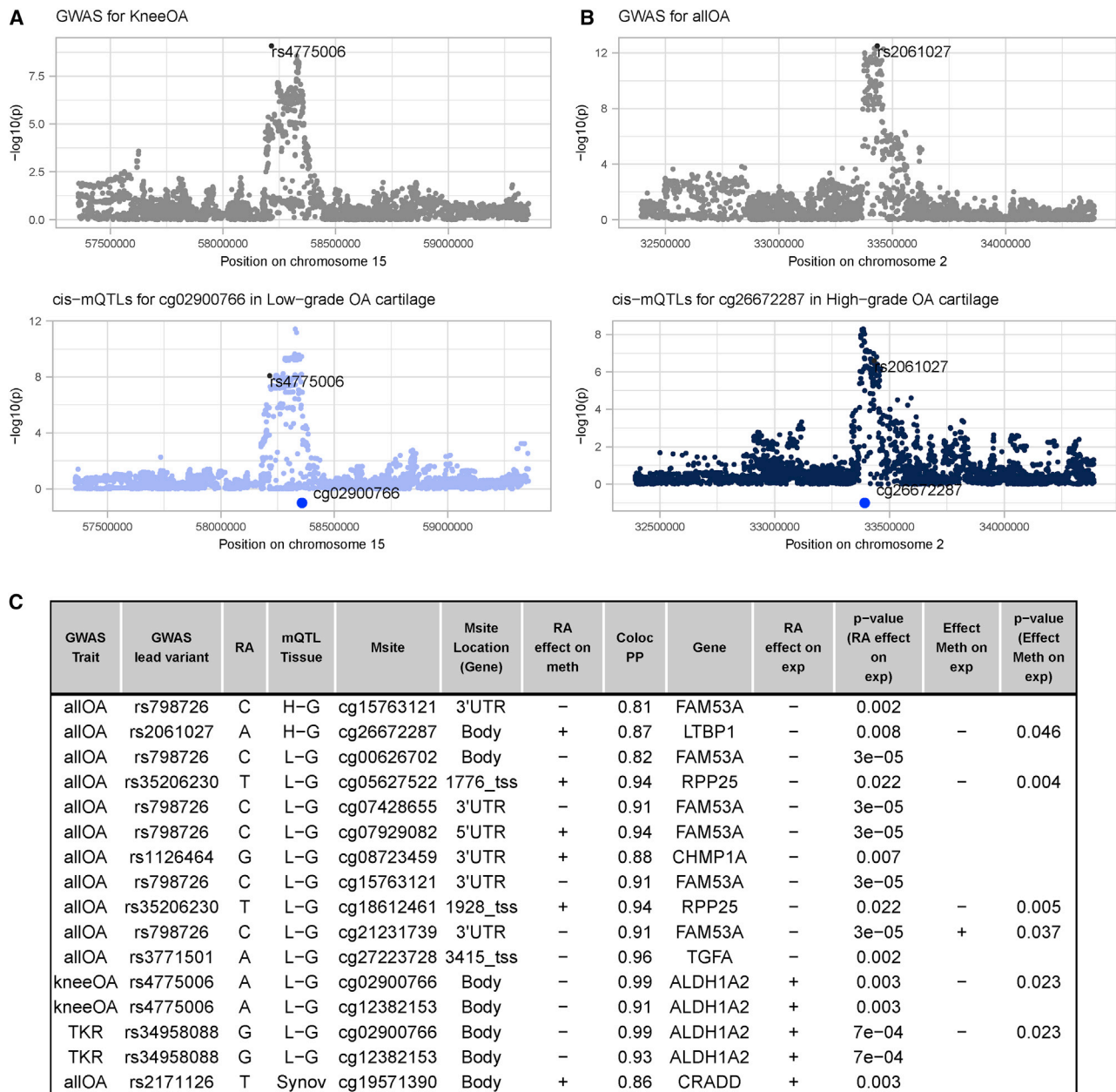
A GWAS for KneeOA

B GWAS for allOA

cis-mQTLs for cg02900766 in Low-grade OA cartilage

cis-mQTLs for cg26672287 in High-grade OA cartilage

C

GWAS Trait	GWAS lead variant	RA	mQTL Tissue	Msite	Msite Location (Gene)	RA effect on meth	Coloc PP	Gene	RA effect on exp	p-value (RA effect on exp)	Effect Meth on exp	p-value (Effect Meth on exp)
allOA	rs798726	C	H–G	cg15763121	3'UTR	−	0.81	FAM53A	−	0.002		
allOA	rs2061027	A	H–G	cg26672287	Body	+	0.87	LTBP1	−	0.008	−	0.046
allOA	rs798726	C	L–G	cg00626702	Body	−	0.82	FAM53A	−	3e−05		
allOA	rs35206230	T	L–G	cg05627522	1776_tss	+	0.94	RPP25	−	0.022	−	0.004
allOA	rs798726	C	L–G	cg07428655	3'UTR	−	0.91	FAM53A	−	3e−05		
allOA	rs798726	C	L–G	cg07929082	5'UTR	+	0.94	FAM53A	−	3e−05		
allOA	rs1126464	G	L–G	cg08723459	3'UTR	+	0.88	CHMP1A	−	0.007		
allOA	rs798726	C	L–G	cg15763121	3'UTR	−	0.91	FAM53A	−	3e−05		
allOA	rs35206230	T	L–G	cg18612461	1928_tss	+	0.94	RPP25	−	0.022	−	0.005
allOA	rs798726	C	L–G	cg21231739	3'UTR	−	0.91	FAM53A	−	3e−05	+	0.037
allOA	rs3771501	A	L–G	cg27223728	3415_tss	−	0.96	TGFA	−	0.002		
kneeOA	rs4775006	A	L–G	cg02900766	Body	−	0.99	ALDH1A2	+	0.003	−	0.023
kneeOA	rs4775006	A	L–G	cg12382153	Body	−	0.91	ALDH1A2	+	0.003		
TKR	rs34958088	G	L–G	cg02900766	Body	−	0.99	ALDH1A2	+	7e−04	−	0.023
TKR	rs34958088	G	L–G	cg12382153	Body	−	0.93	ALDH1A2	+	7e−04		
allOA	rs2171126	T	Synov	cg19571390	Body	+	0.86	CRADD	+	0.003		

Figure 7. Colocalization reveals overlapping signals in GWAS and mQTL data
(A–C) (A) and (B) exemplify colocalization events. In (A), we colocalized signals of *cis*-mQTL for the methylation site cg02900766 (A, bottom) with the GWAS for knee OA in the same genomic region (A, top). Here, we observed a posterior probability (PP) for a shared causal variant of 98.6%. Similarly, (B) visualizes the colocalization (PP = 86.5%) of *cis*-mQTL signals targeting cg26672287 in high-grade osteoarthritis cartilage (bottom) with GWAS signals for all OA (top). The highlighted variant (black) refers to the GWAS index variant in the respective genetic locus. (C) outlines osteoarthritis-linked genetic variants that colocalize with a methylation site and additionally show an eQTL effect at nominal significance (nominal p < 0.05) on the gene annotated to the respective methylation site in the same tissue. For four genes, we also identified an association (at nominal significance) with methylation sites for which *cis*-mQTLs, in turn, colocalize with a GWAS signal. RA, risk allele; Msite, methylation site; Coloc PP, posterior probability for colocalization; exp, gene expression; meth, methylation; _tss (in column Msite Location), methylation sites that are close to a transcript start site of the respective gene. The preceding number refers to the distance in bp.

one joint tissue also colocalize in a large whole blood mQTL meta-analysis.[40]

Considering all OA-risk variants, we found 15 risk variant-methylation site pairs for which we estimated colocalizing GWAS signals and mQTL in at least one joint tissue (ten and five pairs in low-grade and high-grade osteoar-thritis cartilage, respectively), but not in whole blood. These pairs involved eight all OA-risk variants in total (seven and three all OA signals in low-grade and high-grade osteoarthritis cartilage, respectively).

For the knee OA-risk variants, we identified five risk variant-methylation site pairs with colocalizing GWAS

signals and mQTL in at least one joint tissue (two, two, and one pairs in low-grade and high-grade osteoarthritis cartilage and synovium, respectively) but not in whole blood. These pairs involved two unique knee OA-risk variants (rs9277552 in low-grade and high-grade osteoarthritis cartilage and rs56116847 in synovium). For the TKR-risk variants, we did not find evidence for joint-tissue-specific colocalizations.

Discussion

Osteoarthritis is a common disease with a complex polygenic architecture. In this study, we analyzed the genome-wide methylation profile of low-grade osteoarthritis cartilage, high-grade osteoarthritis cartilage, and synovium at unprecedented scale and depth. We identified and biologically characterized DNA methylation markers of osteoarthritis grade and generated genome-wide maps of mQTLs in three understudied osteoarthritis-relevant tissues, which we used to identify mechanistically relevant genes.

Our data revealed global differences in the methylation profile between tissue types (cartilage versus synovium) and cartilage degeneration states (low-grade versus high-grade osteoarthritis cartilage), with robust evidence for replication in an independent dataset despite lower power due to smaller replication sample size. This study represents a large EWAS for knee cartilage degeneration, increasing the number of studied knee osteoarthritis-affected individuals by almost 6-fold, thus providing substantially higher power compared to previous studies. Together, our findings underline the cell type and osteoarthritis-grade specificity of DNA methylation in primary tissues, thus highlighting the importance of expanding molecular studies of complex diseases to multiple relevant tissues and cell types.

Indeed, comparison of our findings with methylation data available in peripheral blood further underlined the value of analyzing primary tissues. Observed differences included mQTLs with opposite directions of effect and evidence for colocalization in joint tissue, but not in whole blood, for genetic variants linked with osteoarthritis. These findings suggest that at least a subset of the regulatory effects conferred by osteoarthritis-linked variants through proximal methylation sites are specific to osteoarthritis-affected tissue. More generally, they emphasize the value of investigating disease-relevant tissues rather than solely relying on molecular data in peripheral tissue types.

Characterization of knee cartilage degeneration methylation markers revealed the involvement of biological processes such as external matrix organization, skeletal system development, and signaling pathways, which mirror the broad spectrum of physiological mechanisms observed during cartilage degeneration.[52] Our results indicate that the aetiology of osteoarthritis is partly regulated through aberrant DNA methylation. Notably, we report an enrichment of the epithelium-related term "positive regulation of epithelial cell migration." Given the role of epithelial cells in lining body cavities, in particular blood vessels, this finding may suggest that methylation is involved in the pathogenic release of pro-angiogenic factors. Our findings provide evidence that epithelium-linked mechanisms are relevant in osteoarthritic changes of the articular cartilage in affected joints.

Our study presents a genome-wide map of mQTLs in low-grade and high-grade osteoarthritis cartilage as well as in the synovium of osteoarthritis-affected knees. We identified 18 differential mQTLs between low-grade and high-grade osteoarthritis cartilage. This finding suggests distinct regulatory effects of genetic variants on methylation early and late in the cartilage degeneration process, thus proposing changing genetic influences on epigenetic profiles during osteoarthritis progression.

We identified methylation sites that play a putative causal role in osteoarthritis, for example for the WWP2, BSN, and MFHAS1 genes. WWP2 codes for WW domain-containing E3 ubiquitin protein ligase 2, which is involved in protein ubiquitination. WWP2 is the host gene of micro RNA 140, a key regulator in chondrocytes, which is targeted by methylation in that region. Wwp2 has previously been implicated in cartilage homeostasis through regulation of Adamts5, a gene encoding an aggrecanase. In addition, WWP2 demonstrates decreased expression levels in osteoarthritis-affected articular cartilage derived from samples of affected individuals.[53] Our findings indicate that methylation may be driving this aberrant mechanism. Previously, a study identified an mQTL that targets methylation sites in WWP2.[17] Another study found WWP2 expression to be significantly associated with proximal genetic variants and methylation levels of close methylation sites.[11] Together, these results support a role for genetically determined methylation for WWP2 regulation in osteoarthritis. BSN encodes a protein involved in neurotransmission. In the active zone of the synapse, BSN is part of the scaffold of the presynaptic skeleton complex, a structure that assists in the vesicle fusion of synaptic vesicles and presynaptic membranes.[54] This finding may point to innervation in cartilage and synovium during osteoarthritis. MFHAS1 plays a role in controlling Toll-like receptors TLR2 and TLR4,[48,49] which in turn promote inflammation of the synovium. Toll-like receptors are exposed by cells in the synovium. They bind released matrix molecules of degraded cartilage, which leads to the formation of chemokines and cytokines, in turn leading to the inflammatory cell infiltration of the synovium.[55]

We found evidence for 56 methylation sites mediating the effects of proximal osteoarthritis-linked genetic variants in osteoarthritis-relevant tissue. For seven genes (ALDH1A2, CHMP1A, CRADD, FAM53A, LTBP1, RPP25, and TGFA), we found evidence that GWAS signals for osteoarthritis colocalize with mQTLs in these genes and are additionally associated with gene expression levels in the same tissue. Four of these genes (ALDH1A2, FAM53A, LTBP1, and RPP25) showed an association between

expression and methylation. Together, our results provide support for a regulatory role of the associated genetic variants across two molecular layers, and altered gene expression is modulated through genetically determined DNA methylation levels.

CRADD is an adapter protein involved in apoptosis and plays a role in the formation of the PIDDosome-complex, which in turn triggers *CASP2*.[56] A role for dysregulated apoptosis in osteoarthritis synovial tissue has been previously suggested.[57] Our findings indicate that the apoptosis-contributing factor *CRADD* is regulated through DNA methylation in synovium. *ALDH1A2* codes for an enzyme that catalyzes the reaction from retinaldehyde to retinoic acid, an activated form of vitamin A. Retinoic acid has been linked to the degeneration of collagen in bone[58] and is further used as an agent to induce matrix degeneration in cartilage samples.[59] LTBP1 plays an essential role in the regulation of transforming growth factor (TGF) betas, a cytokine class that has been involved in extracellular matrix synthesis and maintenance, but also moderates the effects of inflammation and controls hypertrophy of chondrocytes.[60] TGF betas are produced by chondrocytes in their inactivated form. LTBP1 binds these inactive TGF-betas to the extracellular matrix in cartilage.[61]

Together, causal inference and colocalization analyses point to methylation sites that putatively contribute to osteoarthritis in synovium as well as in early (low-grade osteoarthritis cartilage) and late disease stages (high-grade cartilage).

In summary, our results highlight the cell type as well as disease-grade specificity of the methylome in osteoarthritis-relevant tissue. We identify evidence for the involvement of epithelium-related pathways and identify likely effector genes for hitherto unresolved osteoarthritis GWAS signals. In several cases, we are able to decipher the molecular mechanism underpinning these associations and demonstrate an important role for DNA methylation in the aetiopathogenesis of this debilitating disease.

Data and code availability

Methylation QTL, Mendelian randomization, and differential methylation results can be obtained online (hmgubox and the Downloads page of the Musculoskeletal Knowledge Portal, see web resources). All software used in this study is available from free repositories or manufacturers as referenced in the web Resources and supplemental subjects and methods.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2022.05.010.

Acknowledgments

Author contributions

Study design: E.Z., J.M.W.; Clinical collection: J.M.W.; Data Analysis: P.K., K.H., L.S., J.S.; Interpretation of results: P.K., E.Z., M.S., J.M.W., C.L.R.; Replication data: R.C.A., I.M.; Manuscript drafting: P.K., E.Z.; Manuscript reviewing and editing: all authors.

Declaration of interests

The authors declare no competing interests.

Web resources

An epigenome-wide view of osteoarthritis in primary tissues, https://hmgubox.helmholtz-muenchen.de/d/a23fce319fd844d4b293/

coloc.fast function, https://github.com/tobyjohnson/gtx/blob/526120435bb3e29c39fc71604eee03a371ec3753/R/coloc.R

Ensembl Variant Effect Predictor, http://grch37.ensembl.org/Homo_sapiens/Tools/VEP/

Gene Expression Omnibus database, https://www.ncbi.nlm.nih.gov/geo/ (GEO: GSE63106)

Genetics of DNA Methylation Consortium, http://mqtldb.godmc.org.uk/

GWAS catalog, https://www.ebi.ac.uk/gwas

MetaTissue, http://genetics.cs.ucla.edu/metatissue/

Musculoskeletal Knowledge Portal, http://mskkp.org

References

1. Vos, T., Flaxman, A.D., Naghavi, M., Lozano, R., Michaud, C., Ezzati, M., Shibuya, K., Salomon, J.A., Abdalla, S., Aboyans, V., et al. (2012). Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet *380*, 2163–2196. https://doi.org/10.1016/S0140-6736(12)61729-2.

2. Boer, C.G., Hatzikotoulas, K., Southam, L., Stefánsdóttir, L., Zhang, Y., Coutinho de Almeida, R., Wu, T.T., Zheng, J., Hartley, A., Teder-Laving, M., et al. (2021). Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. Cell *184*, 4784–4818.e17. https://doi.org/10.1016/j.cell.2021.07.038.

3. Steinberg, J., Southam, L., Roumeliotis, T.I., Clark, M.J., Jayasuriya, R.L., Swift, D., Shah, K.M., Butterfield, N.C., Brooks, R.A., McCaskie, A.W., et al. (2021). A molecular quantitative trait locus map for osteoarthritis. Nat. Commun. *12*, 1309. https://doi.org/10.1038/s41467-021-21593-7.

4. Brena, R.M., Huang, T.H.-M., and Plass, C. (2006). Toward a human epigenome. Nat. Genet. *38*, 1359–1360. https://doi.org/10.1038/ng1206-1359.

5. van Dongen, J., Nivard, M.G., Willemsen, G., Hottenga, J.-J., Helmer, Q., Dolan, C.V., Ehli, E.A., Davies, G.E., van Iterson, M., Breeze, C.E., et al. (2016). Genetic and environmental influences interact with age and sex in shaping the human methylome. Nat. Commun. *7*, 11115. https://doi.org/10.1038/ncomms11115.

6. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330. https://doi.org/10.1126/science.aaz1776.

7. Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., et al. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74. https://doi.org/10.1038/nature11247.

8. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330. https://doi.org/10.1038/nature14248.

9. Stunnenberg, H.G., and Hirst, M. (2016). The international human epigenome Consortium: a blueprint for scientific collaboration and discovery. Cell *167*, 1897. https://doi.org/10.1016/j.cell.2016.12.002.

10. Bonin, C.A., Lewallen, E.A., Baheti, S., Bradley, E.W., Stuart, M.J., Berry, D.J., van Wijnen, A.J., and Westendorf, J.J. (2016). Identification of differentially methylated regions in new genes associated with knee osteoarthritis. Gene *576*, 312–318. https://doi.org/10.1016/j.gene.2015.10.037.

11. den Hollander, W., Ramos, Y.F.M., Bomer, N., Elzinga, S., van der Breggen, R., Lakenberg, N., de Dijcker, W.J., Suchiman, H.E.D., Duijnisveld, B.J., Houwing-Duistermaat, J.J., et al. (2015). Transcriptional associations of osteoarthritis-mediated loss of epigenetic control in articular cartilage. Arthritis Rheumatol. *67*, 2108–2116. https://doi.org/10.1002/art.39162.

12. Moazedi-Fuerst, F.C., Hofner, M., Gruber, G., Weinhaeusel, A., Stradner, M.H., Angerer, H., Peischler, D., Lohberger, B., Glehr, M., Leithner, A., et al. (2014). Epigenetic differences in human cartilage between mild and severe OA. J. Orthop. Res. *32*, 1636–1645. https://doi.org/10.1002/jor.22722.

13. Steinberg, J., Ritchie, G.R.S., Roumeliotis, T.I., Jayasuriya, R.L., Clark, M.J., Brooks, R.A., Binch, A.L.A., Shah, K.M., Coyle, R., Pardo, M., et al. (2017). Integrative epigenomics, transcriptomics and proteomics of patient chondrocytes reveal genes and pathways involved in osteoarthritis. Sci. Rep. *7*, 8935. https://doi.org/10.1038/s41598-017-09335-6.

14. Zhang, Y., Fukui, N., Yahata, M., Katsuragawa, Y., Tashiro, T., Ikegawa, S., and Michael Lee, M.T. (2016). Genome-wide DNA methylation profile implicates potential cartilage regeneration at the late stage of knee osteoarthritis. Osteoarthritis Cartilage *24*, 835–843. https://doi.org/10.1016/j.joca.2015.12.013.

15. Bomer, N., den Hollander, W., Ramos, Y.F.M., Bos, S.D., van der Breggen, R., Lakenberg, N., Pepers, B.A., van Eeden, A.E., Darvishan, A., Tobi, E.W., et al. (2015). Underlying molecular mechanisms of DIO2 susceptibility in symptomatic osteoarthritis. Ann. Rheum. Dis. *74*, 1571–1579. https://doi.org/10.1136/annrheumdis-2013-204739.

16. Rice, S.J., Aubourg, G., Sorial, A.K., Almarza, D., Tselepi, M., Deehan, D.J., Reynard, L.N., and Loughlin, J. (2018). Identification of a novel, methylation-dependent, RUNX2 regulatory region associated with osteoarthritis risk. Hum. Mol. Genet. *27*, 3464–3474. https://doi.org/10.1093/hmg/ddy257.

17. Rice, S.J., Cheung, K., Reynard, L.N., and Loughlin, J. (2019a). Discovery and analysis of methylation quantitative trait loci (mQTLs) mapping to novel osteoarthritis genetic risk signals. Osteoarthritis Cartilage *27*, 1545–1556. https://doi.org/10.1016/j.joca.2019.05.017.

18. Rice, S.J., Tselepi, M., Sorial, A.K., Aubourg, G., Shepherd, C., Almarza, D., Skelton, A.J., Pangou, I., Deehan, D., Reynard, L.N., and Loughlin, J. (2019b). Prioritization of PLEC and GRINA as Osteoarthritis Risk Genes Through the Identification and Characterization of Novel Methylation Quantitative Trait Loci. Arthritis Rheumatol. *71*, 1285–1296. https://doi.org/10.1002/art.40849.

19. Rushton, M.D., Reynard, L.N., Young, D.A., Shepherd, C., Aubourg, G., Gee, F., Darlay, R., Deehan, D., Cordell, H.J., and Loughlin, J. (2015). Methylation quantitative trait locus analysis of osteoarthritis links epigenetics with genetic risk. Hum. Mol. Genet. *24*, 7432–7444. https://doi.org/10.1093/hmg/ddv433.

20. Mathiessen, A., and Conaghan, P.G. (2017). Synovitis in osteoarthritis: current understanding with therapeutic implications. Arthritis Res. Ther. *19*, 18. https://doi.org/10.1186/s13075-017-1229-9.

21. Parker, E., Hofer, I.M.J., Rice, S.J., Earl, L., Anjum, S.A., Deehan, D.J., and Loughlin, J. (2021). Multi-tissue epigenetic and gene expression analysis combined with epigenome modulation identifies RWDD2B as a target of osteoarthritis susceptibility. Arthritis Rheumatol. *73*, 100–109. https://doi.org/10.1002/art.41473.

22. Sorial, A.K., Hofer, I.M.J., Tselepi, M., Cheung, K., Parker, E., Deehan, D.J., Rice, S.J., and Loughlin, J. (2020). Multi-tissue epigenetic analysis of the osteoarthritis susceptibility locus mapping to the plectin gene PLEC. Osteoarthritis Cartilage *28*, 1448–1458. https://doi.org/10.1016/j.joca.2020.06.001.

23. Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363–1369. https://doi.org/10.1093/bioinformatics/btu049.

24. Fortin, J.-P., Triche, T.J., and Hansen, K.D. (2017). Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics *33*, 558–560. https://doi.org/10.1093/bioinformatics/btw691.

25. Fortin, J.-P., Labbe, A., Lemire, M., Zanke, B.W., Hudson, T.J., Fertig, E.J., Greenwood, C.M., and Hansen, K.D. (2014). Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol. *15*, 503. https://doi.org/10.1186/s13059-014-0503-2.

26. Chen, Y., Lemire, M., Choufani, S., Butcher, D.T., Grafodatskaya, D., Zanke, B.W., Gallinger, S., Hudson, T.J., and Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the illumina infinium humanMethylation450 microarray. Epigenetics *8*, 203–209. https://doi.org/10.4161/epi.23470.

27. McCartney, D.L., Walker, R.M., Morris, S.W., McIntosh, A.M., Porteous, D.J., and Evans, K.L. (2016). Identification of polymorphic and off-target probe binding sites on the illumina infinium methylationEPIC BeadChip. Genom. Data *9*, 22–24. https://doi.org/10.1016/j.gdata.2016.05.012.

28. Pidsley, R., Zotenko, E., Peters, T.J., Lawrence, M.G., Risbridger, G.P., Molloy, P., Van Djik, S., Muhlhausler, B., Stirzaker, C., and Clark, S.J. (2016). Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol. *17*, 208. https://doi.org/10.1186/s13059-016-1066-1.

29. Du, P., Zhang, X., Huang, C.-C., Jafari, N., Kibbe, W.A., Hou, L., and Lin, S.M. (2010). Comparison of Beta-value and

M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinf. *11*, 587. https://doi.org/10.1186/1471-2105-11-587.

30. den Hollander, W., Ramos, Y.F.M., Bos, S.D., Bomer, N., van der Breggen, R., Lakenberg, N., de Dijcker, W.J., Duijnisveld, B.J., Slagboom, P.E., Nelissen, R.G.H.H., and Meulenbelt, I. (2014). Knee and hip articular cartilage have distinct epigenomic landscapes: implications for future cartilage regeneration approaches. Ann. Rheum. Dis. *73*, 2208–2212. https://doi.org/10.1136/annrheumdis-2014-205980.

31. Edgar, R., Domrachev, M., and Lash, A.E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. *30*, 207–210. https://doi.org/10.1093/nar/30.1.207.

32. Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics *8*, 118–127. https://doi.org/10.1093/biostatistics/kxj037.

33. Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: an R package for multivariate analysis. J. Stat. Software *25*, 1–18. https://doi.org/10.18637/jss.v025.i01.

34. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47. https://doi.org/10.1093/nar/gkv007.

35. Suderman, M., Staley, J.R., French, R., Arathimos, R., Simpkin, A., and Tilling, K. (2018). dmrff: identifying differentially methylated regions efficiently with power and control. Preprint at bioRxiv. https://doi.org/10.1101/508556.

36. Maksimovic, J., Oshlack, A., and Phipson, B. (2021). Gene set enrichment analysis for genome-wide DNA methylation data. Genome Biol. *22*, 173. https://doi.org/10.1186/s13059-021-02388-x.

37. Phipson, B., Maksimovic, J., and Oshlack, A. (2016). missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics *32*, 286–288. https://doi.org/10.1093/bioinformatics/btv560.

38. Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353–1358. https://doi.org/10.1093/bioinformatics/bts163.

39. Sul, J.H., Han, B., Ye, C., Choi, T., and Eskin, E. (2013). Effectively identifying eQTLs from multiple tissues by combining mixed model and meta-analytic approaches. PLoS Genet. *9*, e1003491. https://doi.org/10.1371/journal.pgen.1003491.

40. Min, J.L., Hemani, G., Hannon, E., Dekkers, K.F., Castillo-Fernandez, J., Luijk, R., Carnero-Montoro, E., Lawson, D.J., Burrows, K., Suderman, M., et al.; BIOS Consortium (2021). Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. Nat. Genet. *53*, 1311–1321. https://doi.org/10.1038/s41588-021-00923-x.

41. Tachmazidou, I., Hatzikotoulas, K., Southam, L., Esparza-Gordillo, J., Haberland, V., Zheng, J., Johnson, T., Koprulu, M., Zengini, E., Steinberg, J., et al. (2019). Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. Nat. Genet. *51*, 230–236. https://doi.org/10.1038/s41588-018-0327-1.

42. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191. https://doi.org/10.1093/bioinformatics/btq340.

43. Hemani, G., Zheng, J., Elsworth, B., Wade, K.H., Haberland, V., Baird, D., Laurin, C., Burgess, S., Bowden, J., Langdon, R., et al.

(2018). The MR-Base platform supports systematic causal inference across the human phenome. Elife *7*, e34408. https://doi.org/10.7554/eLife.34408.

44. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. *10*, e1004383. https://doi.org/10.1371/journal.pgen.1004383.

45. Rushton, M.D., Reynard, L.N., Barter, M.J., Refaie, R., Rankin, K.S., Young, D.A., and Loughlin, J. (2014). Characterization of the cartilage DNA methylome in knee and hip osteoarthritis. Arthritis Rheumatol. *66*, 2450–2460. https://doi.org/10.1002/art.38713.

46. Schegg, B., Hülsmeier, A.J., Rutschmann, C., Maag, C., and Hennet, T. (2009). Core glycosylation of collagen is initiated by two β(1-O)Galactosyltransferases. Mol. Cell Biol. *29*, 943–952. https://doi.org/10.1128/MCB.02085-07.

47. Kehayova, Y.S., Watson, E., Wilkinson, J.M., Loughlin, J., and Rice, S.J. (2021). Genetic and Epigenetic Interplay Within a COLGALT2 Enhancer Associated With Osteoarthritis. Arthritis Rheumatol. https://doi.org/10.1002/art.41738.

48. Shi, Q., Xiong, B., Zhong, J., Wang, H., Ma, D., and Miao, C. (2017). MFHAS1 suppresses TLR4 signaling pathway via induction of PP2A C subunit cytoplasm translocation and inhibition of c-Jun dephosphorylation at Thr239. Mol. Immunol. *88*, 79–88. https://doi.org/10.1016/j.molimm.2017.06.017.

49. Zhong, J., Shi, Q.-Q., Zhu, M.-M., Shen, J., Wang, H.-H., Ma, D., and Miao, C.-H. (2015). MFHAS1 is associated with sepsis and stimulates TLR2/NF-κB signaling pathway following negative regulation. PLoS One *10*, e0143662. https://doi.org/10.1371/journal.pone.0143662.

50. Liu-Bryan, R. (2013). Synovium and the innate inflammatory network in osteoarthritis progression. Curr. Rheumatol. Rep. *15*, 323. https://doi.org/10.1007/s11926-013-0323-5.

51. Price, J.S., Waters, J.G., Darrah, C., Pennington, C., Edwards, D.R., Donell, S.T., and Clark, I.M. (2002). The role of chondrocyte senescence in osteoarthritis. Aging Cell *1*, 57–65. https://doi.org/10.1046/j.1474-9728.2002.00008.x.

52. Goldring, S.R., and Goldring, M.B. (2016). Changes in the osteochondral unit during osteoarthritis: structure, function and cartilage-bone crosstalk. Nat. Rev. Rheumatol. *12*, 632–644. https://doi.org/10.1038/nrrheum.2016.148.

53. Mokuda, S., Nakamichi, R., Matsuzaki, T., Ito, Y., Sato, T., Miyata, K., Inui, M., Olmer, M., Sugiyama, E., Lotz, M., and Asahara, H. (2019). Wwp2 maintains cartilage homeostasis through regulation of Adamts5. Nat. Commun. *10*, 2429. https://doi.org/10.1038/s41467-019-10177-1.

54. Südhof, T.C. (2012). The presynaptic active zone. Neuron *75*, 11–25. https://doi.org/10.1016/j.neuron.2012.06.012.

55. Scanzello, C.R., and Goldring, S.R. (2012). The role of synovitis in osteoarthritis pathogenesis. Bone *51*, 249–257. https://doi.org/10.1016/j.bone.2012.02.012.

56. Ahmad, M., Srinivasula, S.M., Wang, L., Talanian, R.V., Litwack, G., Fernandes-Alnemri, T., and Alnemri, E.S. (1997). CRADD, a novel human apoptotic adaptor molecule for caspase-2, and FasL/tumor necrosis factor receptor-interacting protein RIP. Cancer Res. *57*, 615–619.

57. Huang, H., Zheng, J., Shen, N., Wang, G., Zhou, G., Fang, Y., Lin, J., and Zhao, J. (2018). Identification of pathways and genes associated with synovitis in osteoarthritis using

bioinformatics analyses. Sci. Rep. *8*, 10050. https://doi.org/10.1038/s41598-018-28280-6.

58. Varghese, S., Rydziel, S., Jeffrey, J.J., and Canalis, E. (1994). Regulation of interstitial collagenase expression and collagen degradation by retinoic acid in bone cells. Endocrinology *134*, 2438–2444. https://doi.org/10.1210/endo.134.6.8194470.

59. Shlopov, B.V., Lie, W.R., Mainardi, C.L., Cole, A.A., Chubinskaya, S., and Hasty, K.A. (1997). Osteoarthritic lesions: involvement of three different collagenases. Arthritis Rheum. *40*, 2065–2074. https://doi.org/10.1002/art.1780401120.

60. Thielen, N.G.M., van der Kraan, P.M., and van Caam, A.P.M. (2019). TGFβ/BMP signaling pathway in cartilage homeostasis. Cells *8*, 969. https://doi.org/10.3390/cells8090969.

61. Saharinen, J., Taipale, J., and Keski-Oja, J. (1996). Association of the small latent transforming growth factor-beta with an eight cysteine repeat of its binding protein LTBP-1. EMBO J. https://doi.org/10.1002/j.1460-2075.1996.tb00355.x.

Kreitmaier P, Park YC, Swift D, Gilly A, Wilkinson JM*, Zeggini E*. Epigenomic profiling of the infrapatellar fat pad in osteoarthritis. Hum Mol Genet. 2024 Feb 28;33(6):501-509. doi: 10.1093/hmg/ddad198. PMID: 37975894

Epigenomic profiling of the infrapatellar fat pad in osteoarthritis

Peter Kreitmaier[1,2,3], Young-Chan Park[3], Diane Swift[4], Arthur Gilly[3], J. Mark Wilkinson[4,*,†], Eleftheria Zeggini [1,3,*,†]

[1]Technical University of Munich (TUM) and Klinikum Rechts der Isar, TUM School of Medicine and Health, Ismaninger Str. 22, Munich 81675, Germany
[2]Graduate School of Experimental Medicine, TUM School of Medicine and Health, Technical University of Munich, Ismaninger Str. 22, Munich 81675, Germany
[3]Institute of Translational Genomics, Helmholtz Zentrum München, German Research Center for Environmental Health, Ingolstaedter Landstr. 1, Neuherberg 85764, Germany
[4]Department of Oncology and Metabolism, The University of Sheffield, Beech Hill Rd, Sheffield S10 2RX, United Kingdom
*Corresponding author. Institute of Translational Genomics, Helmholtz Zentrum München Deutsches Forschungszentrum für Gesundheit und Umwelt, Neuherberg 85764, Germany. E-mail: eleftheria.zeggini@helmholtz-munich.de and Metabolic Bone Unit, Sorby Wing Northern General Hospital Sheffield, S5 7AU United Kingdom. E-mail: j.m.wilkinson@sheffield.ac.uk
†J. Mark Wilkinson and Eleftheria Zeggini joint corresponding author.

Abstract

Osteoarthritis is a prevalent, complex disease of the joints, and affects multiple intra-articular tissues. Here, we have examined genome-wide DNA methylation profiles of primary infrapatellar fat pad and matched blood samples from 70 osteoarthritis patients undergoing total knee replacement surgery. Comparing the DNA methylation profiles between these tissues reveal widespread epigenetic differences. We produce the first genome-wide methylation quantitative trait locus (mQTL) map of fat pad, and make the resource available to the wider community. Using two-sample Mendelian randomization and colocalization analyses, we resolve osteoarthritis GWAS signals and provide insights into the molecular mechanisms underpinning disease aetiopathology. Our findings provide the first view of the epigenetic landscape of infrapatellar fat pad primary tissue in osteoarthritis.

Keywords: osteoarthritis; DNA methylation; infrapatellar fat pad; methylation QTL; EWAS

Introduction

Osteoarthritis is a complex joint disease that affects more than 300 million people [1]. In the face of aging populations, the impact of osteoarthritis on public health systems is estimated to increase further [1]. Current treatment methods are limited to pain management and total joint replacement, highlighting the need to develop novel, personalised treatment strategies. Therefore, it is important to enhance our understanding of the genetic and genomic basis of osteoarthritis.

To date, genome-wide association analyses (GWAS) have identified more than 150 genetic risk loci [2] of osteoarthritis, thus improving our understanding of its polygenic basis. Large-scale molecular datasets of relevant, primary cell types of osteoarthritis patients can reveal molecular mechanisms underlying disease and provide insights beyond genetic studies. Combining results from genetic and molecular studies can help pinpoint molecular mechanisms of disease development and progression, specifically the likely effector genes through which genetic risk variants exert their effect on osteoarthritis development in affected tissues.

Whilst a number of studies have investigated genome-wide molecular profiles of osteoarthritis-affected primary joint tissues [3, 4] the majority have focused on cartilage [5]. Osteoarthritis affects all joint tissues, and a small number of genome-wide molecular studies have extended molecular profiling to other primary joint tissues, such as the synovium [6, 7] or subchondral bone [8].

The infrapatellar fat pad, an adipocyte-rich tissue located inferior to the patella in the anterior part of the knee joint [9], has not been deeply studied in osteoarthritis to date. The fat pad is located among other joint tissues and protects knee components (by stabilising the patella) when exposed to mechanical stress, e.g. during exercise. In osteoarthritis-affected knees, the infrapatellar fat pad undergoes disease-related alterations, including fibrosis, inflammation and vascularization. Furthermore, it is traversed by nerves and therefore constitutes a source of knee osteoarthritis-related pain.

The fat pad may also interact with other joint tissues during osteoarthritis development and progression [9]. For example, it is proposed that the fat pad secretes pro-inflammatory and catabolic factors that promote cartilage degeneration and inhibit repair mechanisms [10]. Studies using chondrocyte cultures and fat pad-derived fat-conditioned media have provided some first insights into the potential cross-talk between the fat pad and cartilage [9].

Furthermore, the fat pad lies adjacent to the synovium, a connective tissue that lines the joint capsule. Both tissues undergo similar osteoarthritis-related changes, e.g. develop a similar immune cell profile [11]. Studies *in vitro* and in mouse

models suggest interactions between these tissues [12–14]. For example, Bastiaansen-Jenniskens *et al.* cultured fibroblast-like synoviocytes in fat-conditioned medium from fat pad samples of knee osteoarthritis patients, and suggest that fat pad induces fibrotic changes in synoviocytes by stimulating collagen synthesis as well as cell proliferation and migration [14].

Only a small number of studies have examined the profile of infrapatellar fat pad in osteoarthritis patients. Gandhi *et al.* characterised microarray-based gene expression profiles of the infrapatellar fat pad in 34 (29 and five in late and early stage knee osteoarthritis, respectively) individuals [15]. Other studies have investigated the molecular characteristics of osteoarthritis fat pad in genomic regions of osteoarthritis risk signals [16–18] or focused on cytokines and extracellular matrix genes [19].

In this study, we focus on DNA methylation, an epigenetic mark that refers to the covalent addition of a methyl-group to the DNA. Methylation is dynamic, tissue-specific, and plays a regulatory role in gene expression. In general, promoter methylation is negatively correlated with gene expression, whereas methylation in other parts of the genome, such as the gene body, remain less understood.

We examine the genome-wide DNA methylation profile of infrapatellar fat pad adipocytes of osteoarthritis-affected knees. We (1) compare fat pad and blood methylation profiles matched from the same patients, (2) generate a genome-wide methylation quantitative trait loci (mQTL) map in fat pad and (3) resolve osteoarthritis GWAS signals by integrating omics with genetic association data.

Results
Distinct epigenetic profiles in blood and fat pad adipocytes

We investigated global differences in the epigenetic profile between fat pad and peripheral blood samples for the first time. We performed PCA integrating infrapatellar fat pad samples from knee osteoarthritis patients (n = 70) and matched blood samples from a subset of these individuals (n = 58). We identified a separation of fat pad and blood samples along the first principal component, which was associated with tissue type (logistic regression p value: 2.7×10^{-7}, beta: -0.013, SE: 0.0026). This underlines the tissue-specificity of the epigenetic profile on a global level (Fig. 1A).

To characterise tissue-specificity on the methylation site level, we performed an epigenome-wide association study (EWAS) of matched fat pad and blood samples from the same patient (n = 58) and identified 84 973 (of 780 177 tested sites, 10.89%) strongly differentially methylated sites (DMS) between fat pad and whole blood samples ($P < 6.4 \times 10^{-08}$, beta > 2, Table S1). Of these, 33 391 and 51 582 showed hyper- and hypomethylation in fat pad tissue, respectively (Fig. 1B). Together, these results highlight extensive differences in the epigenetic profile of fat pad and peripheral blood.

Genome-wide mQTL map in fat pad adipocytes

We performed cis-mQTL analysis to estimate genetic variants that are associated with methylation levels of nearby methylation sites (<= 1 Mb). We identified 35 948 mQTL-targeted methylation sites (Fig. 2A, Methods), including cg20673407 (Fig. 2B) and cg14016568 (Fig. 2C). Together, this constitutes the first genome-wide mQTL map of infrapatellar fat pad adipocytes in knee osteoarthritis. The full summary statistics are publicly available in the Musculoskeletal Knowledge Portal (http://mskkp.org).

Osteoarthritis GWAS signal resolution

Next, we integrated the newly-generated fat pad mQTL map with GWAS results of two osteoarthritis traits, namely knee osteoarthritis and total knee replacement [2], to determine methylation sites with a putative causal role in osteoarthritis.

We applied colocalisation to estimate a probability for methylation mediating the osteoarthritis-promoting effect of risk variants. In total, we identified 16 methylation sites for which mQTL signals colocalised with 11 (of 25 tested, 44%) GWAS signals (Posterior probability for colocalisation > 80%) (Tables 1 and S2). For knee osteoarthritis, we resolved 9 (of 24 tested, 37.5%) GWAS signals that colocalised with mQTL of 13 methylation sites (Fig. 3A). Analogously, colocalising mQTL with GWAS results for total knee replacement resolved 5 (of 10 tested, 50%) GWAS signals and revealed 7 methylation sites with a potential causal role in osteoarthritis (Fig. 3B).

Next, we performed causal inference analysis by applying two-sample Mendelian randomization (MR) to estimate the putative causal effect of methylation on osteoarthritis. In these MR models, we used mQTL as instruments as well as mQTL-targeted methylation sites and osteoarthritis as exposure and outcome, respectively (Method). Here, we detected 36 methylation sites with a putative role ($P < 7.70 \times 10^{-07}$) in osteoarthritis (Fig. 3C), in total (Fig. S1, Table S3). For knee osteoarthritis, we identified 32 methylation sites, of which 15 and 17 revealed a link of hyper- and hypomethylation with osteoarthritis, respectively. For total knee replacement, we identified 15 methylation sites with a putative causal role (9 and 6 showing hyper- and hypomethylation in osteoarthritis, respectively). Eleven methylation sites were identified in both osteoarthritis-relevant traits, for which the direction of effect was concordant.

MR and colocalisation identified 37 putative causal methylation sites, in total. Of these, 15 were identified in both approaches, thus providing two lines of evidence for their respective causal involvement (Tables 1 and S4). Together, colocalisation and MR results suggest that these methylation marks mediate the regulation of genetic risk variants on effector genes in fat pad.

Annotated genes of the identified 37 methylation sites have been previously linked to osteoarthritis using causal approaches on genome-wide mQTL maps of cartilage or synovium. This includes *WWP2* (annotated to fat pad relevant methylation site cg04703221), a chondrocyte regulator [20] for which methylation has been causally linked to osteoarthritis in low disease-grade cartilage and synovium [7]. *ALDH1A2* (cg12031962, cg12031962 and cg08668585) has also previously been linked to osteoarthritis at the methylation [7] (in low-and high-grade osteoarthritis cartilage as well as synovium) as well as expression [6] (low-grade osteoarthritis cartilage) levels. Furthermore, we identified osteoarthritis-linked methylation in the collagen type *COL27A1* (cg21771125).

We also identified likely effector genes that were not previously resolved in molecular QTL maps of primary osteoarthritis cartilage and synovium [6, 7] including *USP8* (cg01701297 and cg05456662; involved in cell proliferation), *TSKU* (cg17107561; encodes development-linked extracellular matrix protein) and *FER1L4* (cg14387502 cg05220160; involved in plasma membrane organization) which can be linked to osteoarthritis-relevant mechanisms (Discussion). Together, integrating the fat pad mQTL profile with osteoarthritis GWAS results using colocalisation and MR identified 37 methylation sites with a potential causal involvement in osteoarthritis in fat pad tissue.
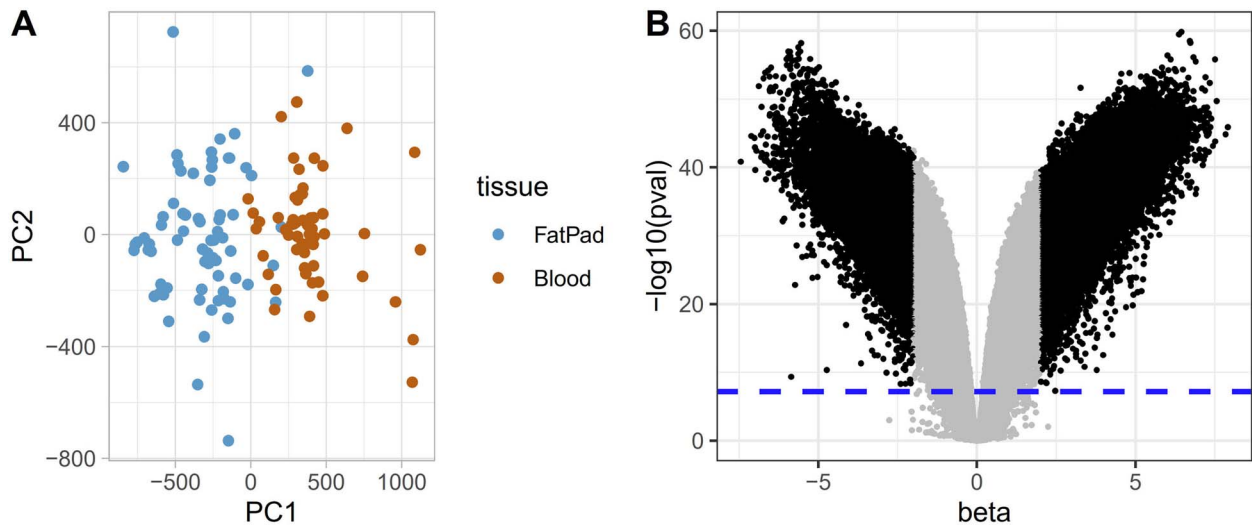
Figure 1. Distinct methylation profiles in blood and fat pad adipocytes. We investigated differences in the methylation profile between fat pad and blood. (A) On a global level, principal component analysis separates fat pad and blood samples along the first principal component. (B) On the methylation site level, a volcano plot demonstrates the multitude of differentially methylated sites. Sites with strong, differential methylation levels (beta > 2) exceeding the Bonferroni significance threshold ($P < 6.41 \times 10^{-08}$, dashed line) are shown in black, otherwise in grey.
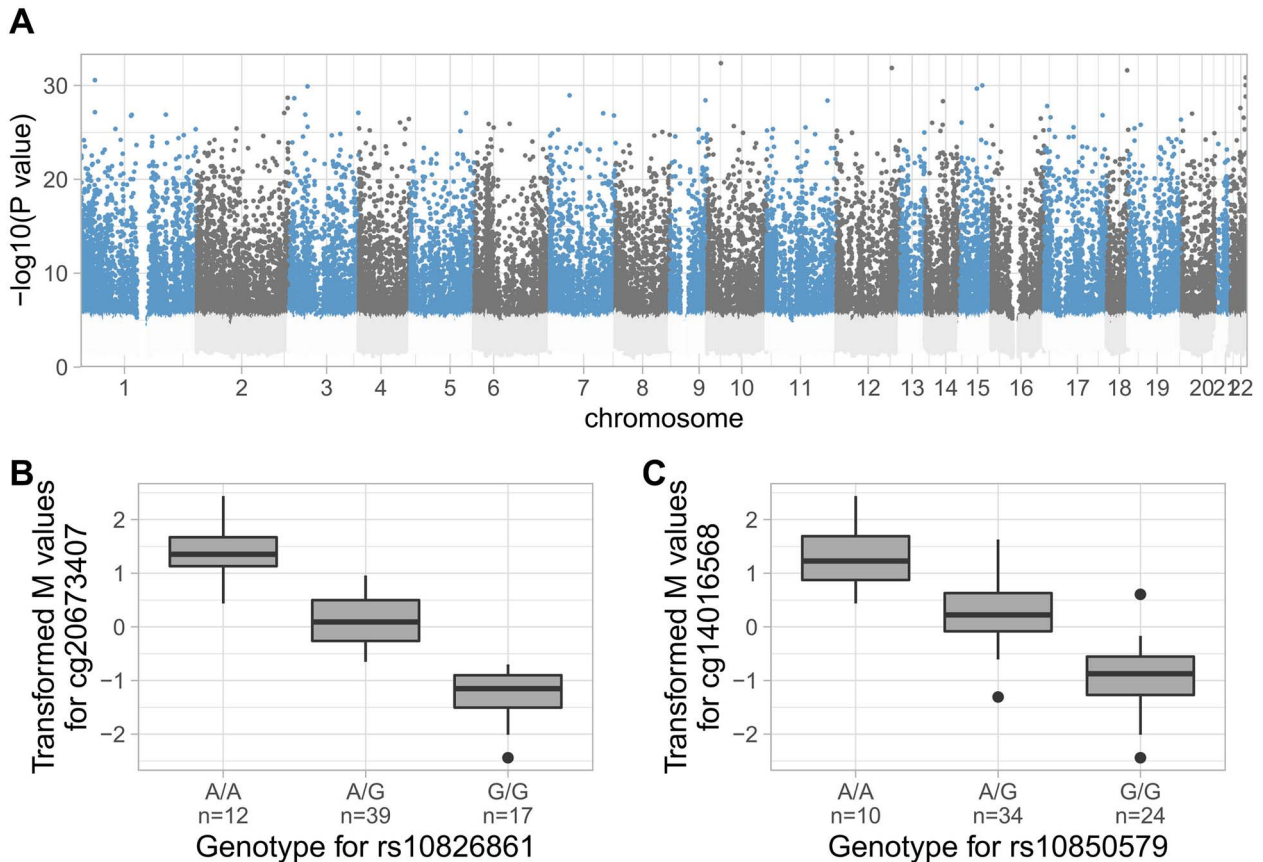

Figure 2. The mQTL map in fat pad adipocytes (A) Manhattan plots depicting the negative log of the *P* value of the most significant association per methylation site across all variants. QTL targeted methylation sites are shown in blue or dark grey, otherwise in light grey. As examples, the boxplots illustrate the effect (B) of rs10826861 on cg20673407 (beta = -1.40, SE = 0.05, $P = 4.15 \times 10^{-33}$) and (C) of rs10850579 on cg14016568 (beta = -1.20, SE = 0.05, $P = 1.10 \times 10^{-28}$). The boxplots represent 25th, 50th, and 75th percentiles, and whiskers extend to 1.5 times the interquartile range.

Discussion

Osteoarthritis is a common joint disorder with a polygenic architecture. Genome-wide molecular profiles of affected primary tissues remain understudied and excluded from large molecular data resources, such as GTEx [21], ENCODE [22] and RoadMap

[23]. In this study, we characterised the first epigenome-wide profile of osteoarthritis-affected infrapatellar fat pad. We identify extensive differences from the epigenetic profile of peripheral blood, generate the first genome-wide mQTL map in fat pad, and identify methylation sites with a likely causal role in osteoarthritis development and progression.

Table 1. Overview of colocalisation signals. Overview of 16 methylation sites for which fat pad mQTL colocalise with an osteoarthritis GWAS risk signal for koa and/or tkr (indicated by the column "Coloc GWAS trait"). For 15 of these methylation sites, we also identified a putative causal effect (column "MR effect": Positive and negative effects indicate links of hyper- and hypomethylation with osteoarthritis, respectively) on osteoarthritis using MR ($P < 7.70 \times 10^{-07}$). Abbreviations: Chr, chromosome (hg38); Pos, position (hg 38); Msite, methylation site; MR, Mendelian randomization; Coloc PP, posterior probability for colocalization; koa, knee osteoarthritis; tkr, total knee replacement.

Msite	Chr	Pos (Msite)	Gene (Msite)	GWAS lead snp	Coloc GWAS trait	Coloc PP	MR effect	MR pval	MR GWAS trait
cg01030629	5	142 425 831	SPRY4-AS1	rs10038860	koa	0.98	−0.04	1.9×10^{-08}	koa
cg01100316	4	1 744 409	TACC3	rs7680647	koa	0.96	0.06	1.5×10^{-08}	koa
				rs4865462	tkr	0.86	0.10	4.9×10^{-09}	tkr
cg01150736	1	219 476 017		rs2791549	koa	0.81			
cg01701297	15	50 462 696	RNA5SP395 USP8	rs4380013	koa	0.90	−0.05	3.3×10^{-09}	koa
cg04703221	16	69 933 160	MIR140 WWP2	rs34195470	koa	0.94	−0.08	5.7×10^{-07}	tkr
							−0.06	1.0×10^{-11}	koa
cg04878480	12	48 012 099	AC004801.4	rs7967762	koa	0.92	0.04	5.8×10^{-09}	koa
			AC004801.5	rs7967762	tkr	0.92	0.07	2.2×10^{-09}	tkr
cg05456662	15	50 424 073	USP8	rs4380013	koa	0.93	0.06	3.9×10^{-09}	koa
cg08668585	15	57 955 405	ALDH1A2	rs4144005	tkr	0.86	0.08	3.6×10^{-10}	tkr
cg10169515	12	123 222 989	MPHOSPH9	rs753350451	koa	0.86	−0.03	2.2×10^{-08}	koa
cg10239804	12	48 104 587	PFKM	rs7967762	koa	0.92	−0.05	5.8×10^{-09}	koa
			SENP1	rs7967762	tkr	0.92	−0.09	2.2×10^{-09}	tkr
cg12031962	15	58 061 651	ALDH1A2	rs4144005	tkr	0.90	0.07	3.6×10^{-10}	tkr
cg15373332	9	114 173 564	COL27A1	rs72760655	koa	0.99	0.06	4.4×10^{-10}	koa
				rs7023177	tkr	0.99	0.10	5.5×10^{-10}	tkr
cg15672022	5	142 426 207	SPRY4-AS1	rs10038860	koa	0.98	−0.08	1.9×10^{-08}	koa
cg16740022	5	142 426 441	SPRY4-AS1	rs10038860	koa	0.98	−0.04	1.9×10^{-08}	koa
cg17669802	20	35 387 551	UQCC1	rs143384	tkr	0.90	−0.08	9.41×10^{-21}	koa
							−0.11	2.2×10^{-14}	tkr
cg17729365	19	10 643 944	SLC44A2	rs2163832	koa	0.96	−0.06	1.0×10^{-08}	koa

Comparing fat pad and blood methylation profiles reveals abundant epigenetic differences underlining the epigenetic tissue-specificity of blood and joint tissues, thus highlighting the necessity to investigate disease-affected tissues.

We present the first genome-wide mQTL map for osteoarthritis-affected infrapatellar fat pad. Colocalising this mQTL map with osteoarthritis GWAS results resolved eleven genetic osteoarthritis risk signals, thus providing evidence for methylation mediating the genetic effect of these GWAS signals on osteoarthritis in fat pad.

We supplemented these causal insights using MR and, together with colocalisation, identified 37 methylation sites with a putative causal role in osteoarthritis in fat pad. Some methylation sites were close to genes (such as *WWP2*, *ALDH1A2*, and *COL27A1*) that have been previously causally linked to osteoarthritis using genome-wide molecular QTL maps of other primary joint tissues [6, 7], suggesting a disease-relevant role across joint tissues.

We also identify genes that have not been previously resolved in molecular QTL maps of primary osteoarthritis tissues [6, 7] such as *USP8, TSKU* and *FER1L4*. USP8 is involved in epidermal growth factor receptor regulation [24], a receptor linked to angiogenic and inflammatory mechanisms. *TSKU* is an inhibitor of Wnt signaling, a pathway which has been consistently linked to osteoarthritis across tissues, e.g. in cartilage, synovium and subchondral bone [25]. *FER1L4* regulates inflammatory factor IL-6 in osteoarthritis-affected cartilage [26] and is linked to VEGF, an osteoarthritis-linked angiogenic factor [27].

These signals can be related to signalling pathways that may contribute to osteoarthritis development in the infrapatellar fat pad and its interaction with other joint tissues.

For example, methylation-mediated upregulation of cytokines may be involved. Elevated levels of IL-6 and VEGF have been previously observed in fat pad samples of osteoarthritis patients [28]. Both factors are regarded to affect surrounding joint tissues. IL-6 is linked to protective, but also inflammatory and catabolic mechanisms in the cartilage [29]. Increased fat pad mRNA expression of VEGF has been associated with higher vascularisation of the neighbouring synovium [28], suggesting interactions between these tissues. The identification of a Wnt pathway regulator (TSKU) may relate to the production of WISP2, a target of the Wnt pathway, for which increased expression levels have been identified in osteoarthritis-affected fat pad [30].

Together, we have identified genes linked to processes that are observed in osteoarthritis-affected fat pad, such as inflammation or vascularization [28], and suggest an involvement of the detected methylation sites in disease-related alterations.

In this work, we investigate blood and fat pad methylation profiles of osteoarthritis patients, which reflect disease processes that could be cause or effect. The integration of genetic data, coupled to colocalisation and causal inference analyses, were all used to deconvolute the role of methylation in osteoarthritis. Obtaining healthy joint tissue as a control for the osteoarthritis-affected joint presents a major challenge, as removal of healthy structural human joint tissue is precluded on ethical and acceptability grounds. The fat pad mQTL map provides insights into genetic effects on infrapatellar fat pad methylation for the first time. Larger sample sizes will be required to achieve 80% power to detect mQTLs across the allele frequency spectrum (Fig. S2).

Our findings highlight differences in the epigenetic profile of fat pad tissue and blood and identify methylation sites that likely
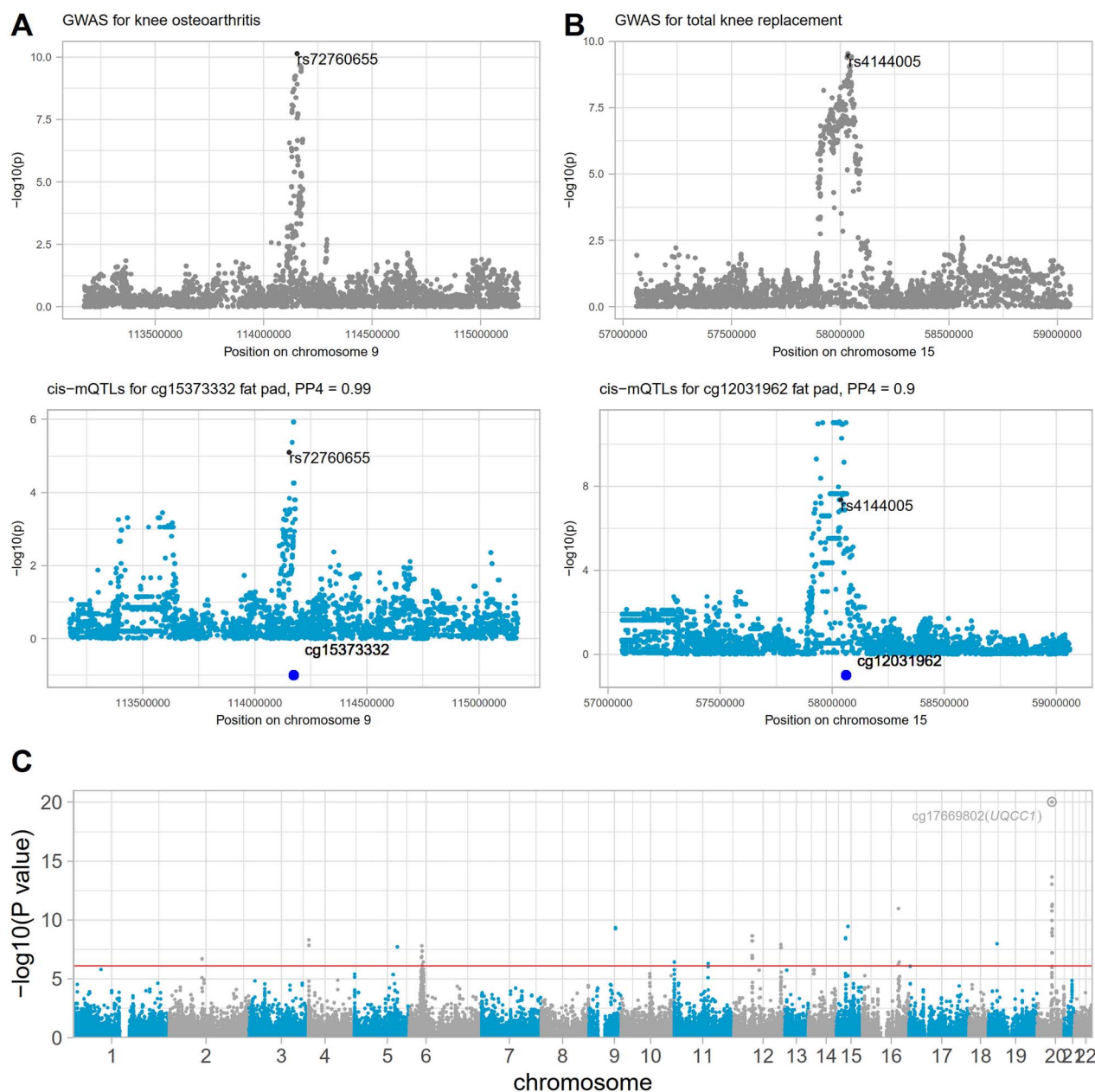
Figure 3. Osteoarthritis GWAS risk signals colocalise with mQTL two colocalisation events are exemplified in (A) and (B). In (A), we colocalised cis-mQTL for cg15373332 with a knee osteoarthritis GWAS signal (Posterior probability = 98.8%). Similarly, (B) shows cis-mQTL for cg12031962 colocalising with a total knee replacement GWAS signal (Posterior probability = 89.8%). (C) Manhattan plot depicting the Mendelian randomization *P* values to estimate the putative causal effects of methylation sites in fat pad on knee osteoarthritis or total knee replacement. The line indicates genome-wide significance applying the Bonferroni correction ($P < 8.31 \times 10^{-07}$).

exert the effect of GWAS risk signals in fat pad, shedding light on the mechanistically relevant role of fat pad methylation in osteoarthritis.

Materials and methods
Study participants
We have collected tissue samples from 210 patients undergoing total joint replacement surgery (111 women, 99 men, age 48–93 years, mean 71 years). All patients provided written, informed consent prior to participation in the study. Adipose tissue was collected from the infra-patellar fat pad by sharp dissection of the fat tissue from the surface of the patellar ligament to yield

not less than 1cm^3 of homogeneous adipose tissue. This work was approved by Oxford NHS REC C (10/H0606/20, SC/15/0132 and SC/20/0144), and samples were collected under Human Tissue Authority license 12 182, South Yorkshire and North Derbyshire Musculoskeletal Biobank, University of Sheffield, UK. We confirmed a joint replacement for osteoarthritis, with no history of significant knee surgery (apart from meniscectomy), knee infection, or fracture, and no malignancy within the previous 5 years. We further confirmed that no patient used glucocorticoid use (systemic or intra-articular) within the previous 6 months, or any other drug associated with immune modulation. We also obtained a peripheral blood sample to extract DNA from all patients.

Adipocyte and peripheral blood collection and processing

Adipose tissue samples were transported in Dulbecco's modified Eagle's medium (DMEM)/F-12 (1:1) (Life Technologies) supplemented with 2 mM glutamine (LifeTechnologies), 100 U/ml penicillin, 100 μg/ml streptomycin (Life Technologies), 2.5 μg/ml amphotericin B (Sigma-Aldrich) and 50 μg/ml ascorbic acid (Sigma-Aldrich) (serum free media). Next, the adipose tissue samples were cut into small pieces ($<2mm^3$) and digested in 3 mg/ml collagenase type I (Sigma-Aldrich) in serum free media for 1 h at 37°C on a flatbed shaker and resuspended in 2mls of PBS and passed through a 100 μm cell strainer (Fisher Scientific). Next, the eluent was made up to 10mls in PBS and centrifuged at 23 g for 5 min. Subsequently, the cell pellet was washed twice in PBS and centrifuged at 323 g for a further 5 min. Cells were counted using a haemocytometer and the viability checked using trypan blue exclusion (Invitrogen). The resulting cell pellet was resuspended in 650 μl of RLT buffer (Qiagen) and DTT Dithiothreitol (20ul DTT per 1 ml of RLT). The optimal cell number for spin column extraction from cells was between 4×10^6 and 1×10^7. Cells were then pelleted and homogenised. DNA extraction was carried out using Qiagen AllPrep DNA Mini Kit following the manufacturer's instructions. Samples were flash frozen in liquid nitrogen and stored at -80°C prior to assays. Peripheral blood was extracted for DNA using a Qiagen QIAamp DNA Blood Maxi kit, according to manufacturer's instructions. The whole blood DNA samples were frozen at -80°C prior to extraction.

Methylation data preprocessing

Genome-wide DNA methylation was measured using the Illumina EPIC array. We preprocessed methylation data using an R package meffil [31] based preprocessing pipeline (https://github.com/perishky/meffil/wiki).

We read and preprocessed blood DNA methylation data using the function meffil.qc, and removed ethnicity outliers, hip samples, samples with $> 10\%$ undetected (detection pvalue > 0.01) methylation values, sex outliers ($> 5 * sd$), methylated/unmethylated signal outliers ($> 3 * sd$) and control probe signal outlier ($> 5 * sd$). We then applied the same procedure (same R functions and thresholds) on DNA methylation samples from fat pad samples. Finally, we normalised methylation samples of all tissues together by applying meffil function meffil.normalize.quantiles (using 16 principal components) and meffil.normalize.samples.

We removed methylation sites with more than 10% of samples low bead number (< 3) or undetected methylation values (detection $P < 0.01$), non-autosomal methylation sites, methylation sites of cross-reactive probes and in close proximity (within 10 base pairs) to common SNPs (MAF > 0.05) in European population [32–34].

We converted initially generated beta values to Mvalues (beta2m function of R package lumi) [35] which we used for downstream analyses [36]. Per tissue, we further replaced strong outliers ($> 10 * sd$ from mean) with the methylation site-specific mean value. Based on a principal component analysis, we removed two outlier samples. The resulting fat pad methylation data comprised 780 177 methylation sites for 70 patients (46 women, 24 men, age 48–93 years, mean 71 years). For 58 of 70 patients, also methylation blood samples were available.

We used publicly available annotations (https://zwdzwd.github.io/InfiniumAnnotation/EPIC.hg38.manifest.tsv.gz and EPIC.hg38.manifest.gencode.v36.tsv.gz) to map probe identifier to the genomic location (hg38) and genes.

Whole-genome sequencing data generation and preprocessing

Whole-genome sequencing (WGS) samples were available for 68 of 70 patients with matching fat pad methylation data. They were measured in two sequencing batches. Of 68 WGS samples, 60 were measured in whole blood samples in the first sequencing batch. Furthermore, eight of 68 WGS samples were measured in cartilage samples in the second batch. In both batches, DNA samples were subjected to standard Illumina paired-end DNA library construction, amplified, and subjected to DNA sequencing using the NovaSeq platform.

Generated CRAM files were input into samtools (samtools conda version 1.14) to create bam files. Subsequently, "bedtools bamtofastq" (bedtools conda version 2.30.0) was applied to obtain data in the fastq format. Per sequencing batch, variant calling was performed using the publicly available pipeline Sarek from nf-core (version 2.7.1, https://nf-co.re/sarek/2.7.1) with the additional options "– tools HaplotypeCaller –generate_gvcf". This uses the GATK Haplotypecaller (GATK v4.1.7.0) and generates g.vcf files. For the genome "GRCh38" was used. For the joint variant calling we adapted a publicly available pipeline (https://github.com/IARCbioinfo/gatk4-GenotypeGVCFs-nf) and used it with GATK (docker container broadinstitute/gatk:4.2.5.0). Reference files for GRCh38 were used from GATK.

For QC on the variant level, we applied Variant Quality Score Recalibration tool using a tranche threshold of 99.5% for SNPs and the recommended 99% for INDELs. For SNPs, this produces an expected false positive rate of 2.5% and an expected sensitivity of 97%.

For QC on the sample-level, we removed strong outlier het rate (two samples), and non-reference allele concordance when compared to directly typed genotype data using variants MAF > 0.01 (one sample), and one sample being a moderate outlier in sequencing depth as well as het rate. No additional sample was excluded solely based on Ti/Tv or singletons.

Furthermore, we removed one sex mismatch and two samples to avoid the inclusion of any sample pair with a relatedness > 0.2. We further excluded two ethnicity outliers identified in an ethnicity check-up using Ancestry and kinship toolkit (based on 1000G data from phase three; https://github.com/Illumina/akt/tree/master) [37]. In total, we removed nine samples.

We excluded variants with MAF < 0.01, Hardy-Weinberg equilibrium $P < 10^{\wedge}-5$ and call rate $< = 0.99$. We then selected samples of individuals with matching fat pad methylation data (n = 68) and kept bi-allelic variants with MAF > 0.05. The resulting WGS data set used for the fat pad mQTL analysis comprised 68 samples and 6 395 994 variants.

Comparing DNA methylation of blood and fat pad tissue

We integrated 70 fat pad and 58 blood samples in a principal component analysis (PCA) to investigate global epigenetic differences between these tissue types. We regressed out known technical batches (slide, row, clinical cohort) by applying Combat from the R package sva [38] and performed PCA using prcomp function.

To compare methylation profiles on methylation site level, we performed differential methylation analysis between fat pad and blood samples paired from the same patients (n = 58). We performed linear modelling using the functions lmFit and eBayes from limma [39]. We added the patient ID to ensure paired analysis design and included 19 surrogate variables (SVs) to account for technical variants. These SVs were estimated using the num.sv

function from the sva package ('be'-method) by protecting the tissue information. We highlighted methylation sites that exceed genome-wide significance threshold (Bonferroni correction with $P < 6.41 \times 10^{-08}$ which corresponds to 0.05/780 177 methylation sites) with strong effect size (beta > 2).

Methylation quantitative trait locus analysis

For the methylation quantitative trait locus (mQTL) analysis, we included whole-genome-sequencing data and fat pad methylation data matching from the same patients (n = 68). We included 6 395 994 bi-allelic genetic variants with a MAF > 0.05 among these 68 patients. We further normalised methylation levels using inverse-normal transformation per methylation site and estimated PEER factors [40] (R package peer, default parameter setting) to infer hidden factors which we included to correct for technical variation. We performed cis-methylation QTL analysis (cis distance: 1 Mb either side of the tested methylation site) using FastQTL (https://github.com/francois-a/fastqtl/) [41]. We first estimated nominal p values for every tested methylation site-variant pair using linear regression with the following model:

$$Methylation\ values \sim genotype + age + sex + sequencing\_batch$$
$$+ row + 10\ PEER\_factors$$

Here, *row* refers to the sample location on the Illumina EPIC array chip. Since *row* can influence methylation levels [42], but did not significantly (ANOVA Bonferroni $P < 0.05$) associate with any of the ten PEER factors (Table S5), we conservatively added it to the model. The variable *sequencing_batch* accounts for WGS sequencing batches. Of 68 WGS samples, 60 and 8 were extracted in the first and second WGS sequencing batch, respectively (methods section "Whole-genome sequencing data generation and preprocessing"). To optimise the number of included PEER factors, we performed mQTL analysis with five, ten and 15 PEER factors and chose the number that maximises detected mQTL targeted methylation site (5 PEER factors: 34956 mQTL targeted methylation sites, 10 PEER factors: 35948, 15 PEER factors: 35808). Secondly, we applied an adaptive permutation scheme (implemented in FastQTL, parameter –permute 1000 10 000) to estimate a q value and nominal *P*-value threshold per methylation site. Methylation sites with a q value < 5% Storey-Tibshirani FDR are regarded as mQTL targeted. For each mQTL-targeted methylation site, significant QTL were variants with a nominal *P* value below the nominal *P* value threshold for that methylation site. Power analysis for the methylation QTL analysis was performed using the R package powerEQTL [43] (function powerEQTL.SLR) across MAF and sample sizes.

Colocalisation

We colocalised [44] fat pad methylation QTL with GWAS signals for knee osteoarthritis and total knee replacement [2]. For this analysis, we applied the coloc.fast function (https://github.com/tobyjohnson/gtx/blob/526120435bb3e29c39fc71604eee03a371ec3753/R/coloc.R) using default settings. We considered mQTL-targeted methylation sites located in the region that spans 100 kb either side of the GWAS signal index variant. For the colocalisation analysis, we included all variants that were included in the cis mQTL analysis for the tested methylation sites. We considered posterior probabilities ("PP4") > 80% as indicator for colocalisation.

Mendelian randomization

To estimate putative causal effects of QTL-targeted methylation sites in fat pad on osteoarthritis traits, we integrated the fat pad mQTL map with GWAS results for knee osteoarthritis and total knee replacement [2]. We applied two sample Mendelian randomization (MR) using the pipeline of the R package TwoSampleMR [45]. We considered methylation sites targeted by at least one mQTL. Per methylation site, we performed clumping (function clump_data, using the European reference panel and setting the R2 threshold to 0.01) to identify independent genetic variants which we included as instruments in the MR models. For methylation sites with one independent instrument, we applied the Wald-ratio, otherwise the inverse variance weighted method.

In total, we applied 64 898 MR models (32 448 and 32 450 for knee osteoarthritis and total knee replacement, respectively) to estimate the putative causal effect of 32 456 methylation sites. We applied the Bonferroni method to correct for multiple testing ($P < 7.70 \times 10^{-07}$).

Acknowledgements

We acknowledge the technical support of Core Facility Genomics at Helmholtz Munich. We thank Dr Inti Alberto de la Rosa Velasquez, Dr Peter Lichtner, Susanne Wittmann and Dr Thomas Walzthöni for help with DNA methylation and whole-genome sequencing data generation as well as whole-genome sequencing data preprocessing.

Author contributions

Study design: E.Z., J.M.W.; Clinical collection: J.M.W., D.S.; WGS data preprocessing: A.G., Y.C.P.; Data analysis: P.K.; Interpretation of results: P.K., E.Z.; Manuscript drafting: P.K., E.Z.; Manuscript reviewing and editing: P.K., E.Z., J.M.W., D.S.

Supplementary data

Supplementary data is available at *HMG Journal* online.

Conflict of interest statement: None declared.

Funding

Data availability

Full summary statistics will be made openly available through the MSK portal (http://mskkp.org) upon manuscript acceptance. All software used in this study is available from free repositories or manufacturers as referenced in the Materials and Methods section.

References

1. Safiri S, Kolahi A-A, Smith E. *et al.* Global, regional and national burden of osteoarthritis 1990-2017: a systematic analysis of the global burden of disease study 2017. *Ann Rheum Dis* 2020;**79**: 819–28.

2. Boer CG, Hatzikotoulas K, Southam L. *et al.* Deciphering osteoarthritis genetics across 826 690 individuals from 9 populations. *Cell* 2021;**184**:4784–4818.e17.

3. Kreitmaier P, Katsoula G, Zeggini E. Insights from multi-omics integration in complex disease primary tissues. *Trends Genet* 2023;**39**:46–58.

4. Katsoula G, Kreitmaier P, Zeggini E. Insights into the molecular landscape of osteoarthritis in human tissues. *Curr Opin Rheumatol* 2022;**34**:79–90.

5. Katsoula G, Steinberg J, Tuerlings M. *et al.* A molecular map of long non-coding RNA expression, isoform switching and alternative splicing in osteoarthritis. *Hum Mol Genet* 2022;**31**:2090–105.

6. Steinberg J, Southam L, Roumeliotis TI. *et al.* A molecular quantitative trait locus map for osteoarthritis. *Nat Commun* 2021;**12**:1309.

7. Kreitmaier P, Suderman M, Southam L. *et al.* An epigenome-wide view of osteoarthritis in primary tissues. *Am J Hum Genet* 2022;**30**:S48.

8. Tuerlings M, van Hoolwerff M, Houtman E. *et al.* RNA sequencing reveals interacting key determinants of osteoarthritis acting in subchondral bone and articular cartilage: identification of IL11 and CHADL as attractive treatment targets. *Arthritis Rheumatol* 2021;**73**:789–99.

9. Zeng N, Yan Z-P, Chen X-Y. *et al.* Infrapatellar fat pad and knee osteoarthritis. *Aging Dis* 2020;**11**:1317–28.

10. Clockaerts S, Bastiaansen-Jenniskens YM, Runhaar J. *et al.* The infrapatellar fat pad should be considered as an active osteoarthritic joint tissue: a narrative review. *Osteoarthr Cartil* 2010;**18**:876–82.

11. Klein-Wieringa IR, de Lange-Brokaar BJE, Yusuf E. *et al.* Inflammatory cells in patients with endstage knee osteoarthritis: a comparison between the synovium and the infrapatellar fat pad. *J Rheumatol* 2016;**43**:771–8.

12. Clements KM, Ball AD, Jones HB. *et al.* Cellular and histopathological changes in the infrapatellar fat pad in the monoiodoacetate model of osteoarthritis pain. *Osteoarthr Cartil* 2009;**17**:805–12.

13. Eymard F, Pigenet A, Citadelle D. *et al.* Induction of an inflammatory and prodegradative phenotype in autologous fibroblast-like synoviocytes by the infrapatellar fat pad from patients with knee osteoarthritis. *Arthritis Rheumatol* 2014;**66**:2165–74.

14. Bastiaansen-Jenniskens YM, Wei W, Feijt C. *et al.* Stimulation of fibrotic processes by the infrapatellar fat pad in cultured synoviocytes from patients with osteoarthritis: a possible role for prostaglandin f2α. *Arthritis Rheum* 2013;**65**:2070–80.

15. Gandhi R, Takahashi M, Virtanen C. *et al.* Microarray analysis of the infrapatellar fat pad in knee osteoarthritis: relationship with joint inflammation. *J Rheumatol* 2011;**38**:1966–72.

16. Sorial AK, Hofer IMJ, Tselepi M. *et al.* Multi-tissue epigenetic analysis of the osteoarthritis susceptibility locus mapping to the plectin gene PLEC. *Osteoarthr Cartil* 2020;**28**:1448–58.

17. Parker E, Hofer IMJ, Rice SJ. *et al.* Multi-tissue epigenetic and gene expression analysis combined with epigenome modulation identifies RWDD2B as a target of osteoarthritis susceptibility. *Arthritis Rheum* 2021;**73**:100–9.

18. 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**:7432–44.

19. Belluzzi E, Macchi V, Fontanella CG. *et al.* Infrapatellar fat pad gene expression and protein production in patients with and without osteoarthritis. *Int J Mol Sci* 2020;**21**:6016.

20. Mokuda S, Nakamichi R, Matsuzaki T. *et al.* Wwp2 maintains cartilage homeostasis through regulation of Adamts5. *Nat Commun* 2019;**10**:2429.

21. GTEx Consortium. The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science* 2020;**369**:1318–30.

22. Dunham I, Kundaje A, Aldred SF. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;**489**:57–74.

23. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* 2015;**518**:317–30.

24. Berlin I, Schwartz H, Nash PD. Regulation of epidermal growth factor receptor ubiquitination and trafficking by the USP8·STAM complex. *J Biol Chem* 2010;**285**:34909–21.

25. Wang Y, Fan X, Xing L. *et al.* Wnt signaling: a promising target for osteoarthritis therapy. *Cell Commun Signal* 2019;**17**:97.

26. He J, Wang L, Ding Y. *et al.* lncRNA FER1L4 is dysregulated in osteoarthritis and regulates IL-6 expression in human chondrocyte cells. *Sci Rep* 2021;**11**:13032.

27. Hamilton JL, Nagao M, Levine BR. *et al.* Targeting VEGF and its receptors for the treatment of osteoarthritis and associated pain. *J Bone Miner Res* 2016;**31**:911–24.

28. Favero M, El-Hadi H, Belluzzi E. *et al.* Infrapatellar fat pad features in osteoarthritis: a histopathological and molecular study. *Rheumatology (Oxford)* 2017;**56**:1784–93.

29. Wiegertjes R, van de Loo FAJ, Blaney Davidson EN. A roadmap to target interleukin-6 in osteoarthritis. *Rheumatology (Oxford)* 2020;**59**:2681–94.

30. Conde J, Scotece M, Abella V. *et al.* Identification of novel adipokines in the joint. Differential expression in healthy and osteoarthritis tissues. *PLoS One* 2015;**10**:e0123601.

31. Min JL, Hemani G, Davey Smith G. *et al.* Meffil: efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics* 2018;**34**:3983–9.

32. McCartney DL, Walker RM, Morris SW. *et al.* Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genom Data* 2016;**9**:22–4.

33. Pidsley R, Zotenko E, Peters TJ. *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol* 2016;**17**:208.

34. Chen Y, Lemire M, Choufani S. *et al.* Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* 2013;**8**:203–9.

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

36. Du P, Zhang X, Huang C-C. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 2010;**11**:587.

37. Arthur R, Schulz-Trieglaff O, Cox AJ. *et al.* AKT: ancestry and kinship toolkit. *Bioinformatics* 2017;**33**:142–4.

38. Leek JT, Johnson WE, Parker HS. *et al.* The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012;**28**:882–3.

39. Ritchie ME, Phipson B, Wu D. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;**43**:e47.

40. Stegle O, Parts L, Piipari M. *et al.* Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* 2012;**7**:500–7.

41. Ongen H, Buil A, Brown AA. *et al.* Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* 2016;**32**:1479–85.

42. Price EM, Robinson WP. Adjusting for batch effects in DNA methylation microarray data, a lesson learned. *Front Genet* 2018;**9**:83.

43. Dong X, Li X, Chang T-W. *et al.* powerEQTL: an R package and shiny application for sample size and power calculation of bulk tissue and single-cell eQTL analysis. *Bioinformatics* 2021;**37**: 4269–71.

44. Giambartolomei C, Vukcevic D, Schadt EE. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014;**10**: e1004383.

45. Hemani G, Zheng J, Elsworth B. *et al.* The MR-base platform supports systematic causal inference across the human phenome. *elife* 2018;**7**:e34408.

Kreitmaier P, Swift D, Wilkinson JM*, Zeggini E*. Epigenomic differences between osteoarthritis grades in primary cartilage. Osteoarthritis Cartilage. 2024 Jul 23:S1063-4584(24)01314-1. doi: 10.1016/j.joca.2024.07.008. Epub ahead of print. PMID: 39053729.

Osteoarthritis and Cartilage

Epigenomic differences between osteoarthritis grades in primary cartilage

Peter Kreitmaier # † ‡, Diane Swift §, J. Mark Wilkinson §, Eleftheria Zeggini # ‡ *

Technical University of Munich (TUM) and Klinikum Rechts der Isar, TUM School of Medicine and Health, Ismaninger Str. 22, 81675 Munich, Germany
† Graduate School of Experimental Medicine, TUM School of Medicine and Health, Technical University of Munich, 81675 Munich, Germany
‡ Institute of Translational Genomics, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany
§ Division of Clinical Medicine, School of Medicine and Population Health, University of Sheffield, Sheffield S10 2RX, UK

ARTICLE INFO

SUMMARY

Objective: Osteoarthritis is a common and complex joint disorder that shows higher prevalence and greater disease severity in women. Here, we investigate genome-wide methylation profiles of primary chondrocytes from osteoarthritis patients.
Design: We compare genome-wide methylation profiles of macroscopically intact (low-grade) and degraded (high-grade) osteoarthritis cartilage samples matched from osteoarthritis patients undergoing knee replacement surgery. We perform an epigenome-wide association study for cartilage degeneration across 170 patients and separately in 96 women and 74 men.
Results: We reveal widespread epigenetic differences with enrichments of nervous system and apoptosis-related processes. We further identify substantial similarities between sexes, but also sex-specific markers and pathways.
Conclusions: Together, we provide the largest genome-wide methylation profiles of primary cartilage to date with enhanced and sex-specific insights into epigenetic processes underlying osteoarthritis progression.

Introduction

Osteoarthritis is a prevalent joint disorder, affecting more than 300 million people worldwide.[1] Existing treatment approaches are limited to pain management and replacement surgery of affected joints. Due to increasingly older populations, the impact of osteoarthritis on public health systems will increase. Together, this highlights the need for novel, personalised treatment approaches that require an enhanced understanding of the genetic and genomic basis of osteoarthritis.

To date, genome-wide association studies have identified over 150 genetic risk loci for osteoarthritis,[2] shedding insights into its complex architecture. Integration of genetic data with molecular profiles of osteoarthritis-affected tissues accessible at the point of joint replacement surgery can help identify effector genes and their mechanisms of action. DNA methylation, an epigenetic mark that describes the covalent attachment of a methyl group to the DNA, is a useful molecular tool in this regard. DNA methylation is associated with gene expression regulation, for example elevated methylation levels close to the transcription start site (particularly in promotor regions) can be associated with reduced gene expression.

DNA methylation studies have generated valuable profiles of osteoarthritis tissues,[3] such as cartilage, synovium[4] and subchondral bone.[5] In cartilage, epigenome-wide association studies (EWAS) have been conducted to compare macroscopically intact (low-grade) and degraded (high-grade) osteoarthritis cartilage samples to study epigenetic markers of cartilage degeneration.[4,6–10] However, these studies have included small numbers of patients and have thus been limited in power.

Furthermore, most methylation osteoarthritis studies combine samples of both sexes. However, osteoarthritis prevalence and incidence are higher among women[11] and female osteoarthritis patients show more osteoarthritis-related pain and disability,[12–14] suggesting potential sex-specific etiological mechanisms. A methylation study has identified a small number of sex-specific cartilage degeneration markers but was limited in sample size (52 women and 38 men).[4]

Together, there is an urgent need for better-powered epigenetic studies of primary osteoarthritis tissues. In this study, we performed EWAS for cartilage degeneration in 170 patients (96 women and 74 men) to characterise common and sex-specific epigenetic markers of osteoarthritis.

Methods

Osteoarthritis-affected individuals and study samples

In this study, we examine cartilage samples from osteoarthritis-affected knees that were collected in 170 osteoarthritis patients (age 38–89 years, mean 70.86 years). These patients included 96 women (age 38–85 years, mean 70.68 years) and 74 men (age 50–89 years, mean 71.11 years) (Fig. S1). These individuals underwent total knee replacement due to late-stage osteoarthritis. Cartilage samples were graded agnostically to sex using the International Cartilage Repair Society (ICRS)[15] macroscopic scoring system (low-grade osteoarthritis cartilage: ICRS score 0 or 1, high-grade cartilage osteoarthritis: ICRS score 3 or 4). This work was approved by Oxford NHS REC C (10/H0606/20 and 15/SC/0132), and samples were collected under Human Tissue Authority license 12182, Sheffield Musculoskeletal Biobank, University of Sheffield, UK. Before participating in the study, all osteoarthritis-affected individuals provided written, informed consent.

Sample extraction

Knee chondrocytes were isolated by following a protocol reported in a previous study (Methods, section "Isolation of chondrocytes").[16]

DNA methylation preprocessing

DNA methylation was measured using the Illumina EPICv1 array. We used a R package meffil-based preprocessing pipeline (https://github.com/perishky/meffil/wiki).[17] We further tested for ethnicity outliers using the Illumina ancestry and kinship toolkit[18] (Fig. S2) and excluded samples with > 10% undetected (detection pvalue > 0.01) methylation values, sex outliers (> 5 * sd), methylated/unmethylated signal outliers (> 3* sd) and control probe signal outlier (> 5 * sd). We normalised methylation samples with the meffil function meffil.normalize.quantiles (including 16 principal components) and meffil.normalize.samples.

We excluded methylation probes with more than 10% of samples low bead number (< 3) or undetected methylation values (detection p < 0.01), probes of non-autosomal methylation sites, cross-reactive probes and probes of methylation sites that are close (within 10 base pairs) to common single nucleotide polymorphisms (minor allele frequency > 0.05) in European population.[19–21]

For downstream analysis, generated beta values were converted to Mvalues (negative M-value: more unmethylated DNA at a particular DNA methylation site; M-value is 0: equal amount of methylated and unmethylated DNA at a particular DNA methylation site; positive M-value: more methylated DNA at a particular DNA methylation site) using the beta2m function of R package lumi.[22] The resulting methylation data comprised 780,181 methylation sites for 170 patients, including 96 women and 74 men. For all 170 patients, matched low-grade and high-grade osteoarthritis samples were available.

We extracted the genomic location (hg38) and annotated genes from publicly available annotation files https://github.com/zhou-lab/InfiniumAnnotationV1/raw/main/Anno/EPIC/EPIC.hg38.manifest.tsv.gz and https://github.com/zhou-lab/InfiniumAnnotationV1/raw/main/Anno/EPIC/EPIC.hg38.manifest.gencode.v36.tsv.gz.

Differential methylation analysis

To compare epigenetic profiles between low- and high-grade osteoarthritis cartilage, we conducted principal component analysis (PCA) using the prcomp function. We then quantified association significances between cartilage types and principal components 1 and 2 by performing ANOVA (R function aov).

Next, we performed three EWAS for cartilage degeneration: One combined (170 patients) as well as in two sex-specific analyses (96 women, 74 men) to identify methylation sites associated with osteoarthritis-related cartilage degeneration. More specifically, we compared high- with low-grade osteoarthritis cartilage samples matched from the same patient. We applied functions from the R package limma (lmFit and eBayes function) to generate paired linear models to enable matched comparisons between low- and high-grade osteoarthritis cartilage samples. We further added surrogate variables (SVs) to account for technical confounders (combined analysis: 31 SV, women: 23, men: 19; these numbers were estimated using the num.sv function with the 'be' procedure).[23] These SVs also capture sequencing batches (Supplementary Note 1). This resulted in the following model:

$$M\text{-}values \sim cartilage\_type + patient\_id + SVs$$

Here, *patient ID* refers to the patient identifier (ensures paired modelling) and *cartilage_type* denotes the cartilage degradation status (low- vs high-grade osteoarthritis). We applied Bonferroni correction per EWAS to correct for multiple testing (threshold: $0.05/780{,}181$ methylation sites $= 6.41 \times 10^{-08}$). Methylation sites achieving significance below this threshold were regarded as differentially methylated sites (DMS). To identify differentially methylated regions (DMRs), we applied the R package dmrff using default parameter settings (maxgap = 500, p.cutoff = 0.05).[24] Regions are DMRs when consisting of more than one methylation site and achieving a Bonferroni-adjusted $p < 0.05$.

Replication analysis

To replicate our DMS results, we compared these findings with a previous EWAS (n = 90 patients) for cartilage degeneration.[4] We regarded DMS as replicated when showing the same direction of effect at nominal significance ($p < 0.05$) in the replication set.

Gene Ontology analysis

We performed Gene Ontology (GO) analysis to biologically characterise DMS of the combined as well as sex-specific EWAS. We applied the gometh function from the missMethyl package (version 1.24.0; we used R package GO.db 3.12.1 to load GO information).[25,26] We used gene annotations from the file EPIC.hg38.manifest.gencode.v36.txt.gz (column "genesUniq"). We included 780,181 methylation sites that passed the preprocessing procedure as background set ("all.cpg") and lists of DMS as query ("sig.cpg"). We only considered GO terms composed of between 20 and 200 genes and applied a Benjamini-Hochberg correction to account for multiple testing.

Comparing combined and sex-specific EWAS

To estimate sex-specific epigenetic markers of cartilage degeneration, we compared the results of sex-specific EWAS on a summary statistics level. Sex-specific DMS were methylation sites that (1) exceed genome-wide significance ($p < 6.41 \times 10^{-08}$) in one sex, but (2) not nominal significance in the other ($p < 0.05$). Furthermore, we compared GO analysis results of DMS identified in women and men on summary statistics level. Here, we defined sex-

specific cartilage degeneration-related GO terms as being significantly (false discovery rate (FDR) < 0.05) enriched in DMS in one sex, but not achieving nominal significance (p < 0.05) in DMS of the other.

Results

Widespread epigenetic markers for cartilage degeneration

We performed principal component analysis for macroscopically intact (low-grade) and degraded (high-grade) osteoarthritis cartilage samples. We identified significant differences along the first (ANOVA $p = 1.04 \times 10^{-13}$) and second principal component (ANOVA $p < 2 \times 10^{-16}$) (Fig. 1A), indicating pronounced global differences in the epigenetic profiles.

Next, we performed an EWAS for cartilage degeneration by comparing paired low-grade and high-grade cartilage samples from 170 patients. Of 780,181 tested methylation sites, 146,777 (18.8%) were differentially methylated (Bonferroni correction, $p < 6.41 \times 10^{-08}$) (Fig. 1B, exemplified by the most significantly DMS cg20482832 in Fig. 1C and Fig. S3, Table S1). Of these, 56,726 and 90,051 showed hyper- and hypomethylation in high-grade cartilage, respectively. We further found 4644 DMS with large methylation differences (Supplementary Note 2, Table S2 and S3). On the region level, we identified 18,661 regions to be differentially methylated between low- and high-grade osteoarthritis cartilage (Supplementary Note 3, Table S4).
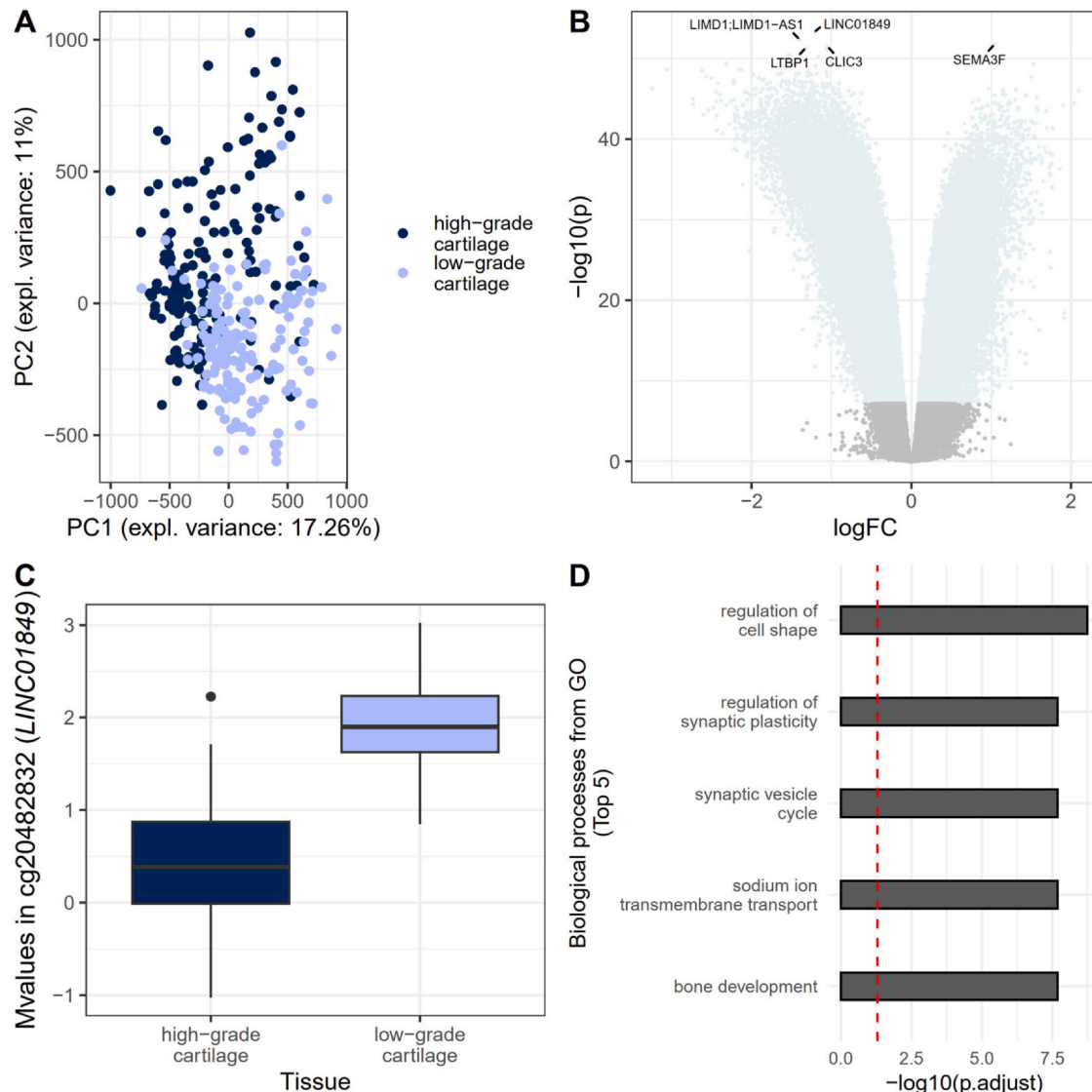

Fig. 1

Osteoarthritis and Cartilage

Methylation differences between low-grade and high-grade osteoarthritis cartilage. (A) Principal component analysis reveals global differences between low-grade and high-grade osteoarthritis cartilage samples. (B) A volcano plot visualises 146,777 DMS ($p < 6.41 \times 10^{-08}$), of which 56,726 and 90,051 are hyper- and hypomethylated, respectively. (C) The most significant DMS is cg20482832 (logfc: −1.22, $p = 5.8 \times 10^{-54}$, SE = 0.049). Whiskers extend to 1.5 times the interquartile range. (D) The five most significantly GO terms from the ontology *biological processes*. The red line indicates statistical significance (FDR < 0.05).

Comparing the DMS with a previous EWAS (n = 90 patients)[4] showed that 92% of comparable sites showed the same direction of effect at nominal significance p < 0.05, suggesting that the identified markers are robust (Supplementary Note 4, Table S5).

To biologically characterise these 146,777 DMS, we performed GO analyses and identified 1660 GO terms (at FDR < 0.05) (Table S6, Fig. 1D). These included terms that were previously associated with cartilage degeneration, such as musculoskeletal tissue development (e.g. "bone development", "cartilage development", "muscle cell development"), cytoskeletal structure (e.g. "actomyosin structure organisation", "actin filament bundle organisation"), extracellular matrix (e.g. "regulation of cell-matrix adhesion") or the epithelium (e.g. "morphogenesis of a branching epithelium", "branching morphogenesis of an epithelial tube").

Notably, we also detected enrichment for nervous system ("dendrite morphogenesis", "regulation of synaptic plasticity", "neuron projection organisation"), neurotransmission ("synaptic vesicle cycle", "neurotransmitter secretion", "signal release from synapse", "positive regulation of synaptic transmission") and apoptosis-related terms (e.g. "regulation of extrinsic apoptotic signalling pathway").

Together, these results suggest methylation sites and biological pathways that are associated with osteoarthritis progression in cartilage.

Next, we generated epigenetic profiles of cartilage degeneration stratified by sex. We performed EWAS separately in women (n = 96) and men (n = 74), again by comparing matched low-grade and high-grade cartilage samples from the same patient.

In women, we identified 62,313 DMS (Bonferroni correction, $p < 6.41 \times 10^{-8}$) (8.41% of tested methylation sites) (Fig. 2A, most significant DMS cg01931614 in Fig. 2B and Fig. S4, Table S7). Of these, 20,345 and 41,968 showed hyper- and hypomethylation in high-grade cartilage, respectively. These differential methylated sites were overrepresented in 361 GO terms (Table S8). On the region level, we identified 19,734 DMR in women (Supplementary Note 3, Table S9).

In men, we detected 61,513 DMS (Bonferroni correction, $p < 6.41 \times 10^{-8}$) (7.88% of tested methylation sites) (Fig. 2C, most significant DMS cg20482832 in Fig. 2D and Fig. S5, Table S10). Of these, 20,295 and 41,218 showed hyper- and hypomethylation in high-grade cartilage, respectively. These signals were enriched in 480 GO terms (Table S11). On the region level, we found 24,064 DMR in men (Supplementary Note 3, Table S12).

Together, the sex-stratified analyses also reveal widespread methylation differences between low-grade and high-grade osteoarthritis cartilage.

Sex-specific markers of cartilage degeneration

Next, we tested whether epigenetic markers for osteoarthritis are common across sexes or sex-specific (Fig. 3A). Of 62,313 and 61,513 that were identified in women and men, respectively, 43,152 overlapped (women: 69.2%, men: 70.1%). Effects of these were in concordant direction and highly correlating (Pearson r = 0.98, $p < 2.2 \times 10^{-16}$). Together, this suggests that a substantial part of epigenetic osteoarthritis markers in cartilage are shared between men and women.

We further detected sex-specific DMS, which are methylation sites associated with cartilage degeneration in one sex but not in the other (Method). We identified 413 (142 hyper- and 271 hypo-methylated in high-grade osteoarthritis cartilage, Fig. 3B and Fig. S6, Table S13) and 539 (259 hyper- and 280 hypomethylated in high-grade osteoarthritis cartilage, Fig. 3C and Fig. S7, Table S14) DMS that are specific for women and men, respectively. Furthermore, we found DMS with larger effect sizes (> = 1.5 x |logfc|) in men (n = 2224

methylation sites) and women (n = 74 methylation sites) when compared to the respective other sex, suggesting effect size magnitude differences of DMS between sexes (Table S15).

A subset of these sex-specific methylation markers (167 of 413 women specific DMS, 215 of 539 men specific DMS) do not achieve nominal significance in the combined analysis, suggesting that some markers are unidentifiable when samples of both sexes are analysed together.

On the biological pathway level, we compared 361 and 480 GO terms enriched (FDR < 0.05) among 62,313 and 61,513 DMS in women and men, respectively, and found 19 (of 361, 5.26%, Table S16) and 51 (of 480, 10.62%, Table S17) women- and men-specific GO terms which are enriched among DMS in one sex, but not in the other (Method). These sex-specific GO terms included terms related to the immune system (for example in men: "positive regulation of lymphocyte differentiation"; women: "phagocytic cup"), the nervous system (men: "regulation of synaptic vesicle cycle", "regulation of postsynaptic membrane neurotransmitter", "neuron migration"; women: "regulation of axon guidance") and hormone regulation (men: "negative regulation of hormone secretion"), suggesting sex-specific methylation changes in these pathways during osteoarthritis degeneration in cartilage.

We found 258 overlapping GO terms (71.47% and 53.75% of identified GO terms in women and men, respectively). Furthermore, these terms also overlapped with the 1660 GO terms of the combined analysis (women: 354 of 361 GO terms, men: 447 of 480), indicating substantial overlap between sex-specific and combined analyses on the biological pathway level.

Discussion

Here, we have generated the largest genome-wide methylation profile of low-grade and high-grade osteoarthritis cartilage to date. We estimate common and sex-specific epigenome-wide profiles of cartilage degeneration and identify DNA methylation markers for osteoarthritis progression across sexes and in a sex-specific manner.

We conducted the largest sex-combined (170 patients) and sex-specific (96 women, 74 men) EWAS for cartilage degeneration which almost doubles the sample size of the next largest study.[4] These analyses identify widespread epigenetic markers of cartilage degeneration, highlighting the distinctness of the methylation profile between early and late cartilage degeneration grades.

We compared sex-specific EWAS results and found osteoarthritis-related epigenetic markers and pathways to be largely overlapping between sexes. This suggests that the molecular processes contributing to osteoarthritis in cartilage are largely the same. We further identified a small number of epigenetic markers solely identified in men (n = 539 DMS) and women (n = 413 DMS), which suggests a few sex-specific epigenetic mechanisms.

GO analysis of combined EWAS results revealed biological processes previously associated with cartilage degeneration in genome-wide methylation studies,[4,6–9] including musculoskeletal tissue development, cytoskeletal structure, or extracellular matrix.

Notably, we identified apoptosis-related GO terms (such as "regulation of extrinsic apoptotic signalling pathway") which may point to chondrocyte apoptosis that has been associated with cartilage matrix breakdown.[27]

The nervous system (e.g. "dendrite morphogenesis", "regulation of synaptic plasticity", "neuron projection organisation") and neurotransmission-related (e.g. "synaptic vesicle cycle", "neurotransmitter secretion", "signal release from synapse", "positive regulation of synaptic transmission") terms were strongly represented, thus confirming a small number of nervous system-related signals in smaller osteoarthritis cartilage EWAS.[7–9] These signals may point to the innervation in the diseased cartilage,

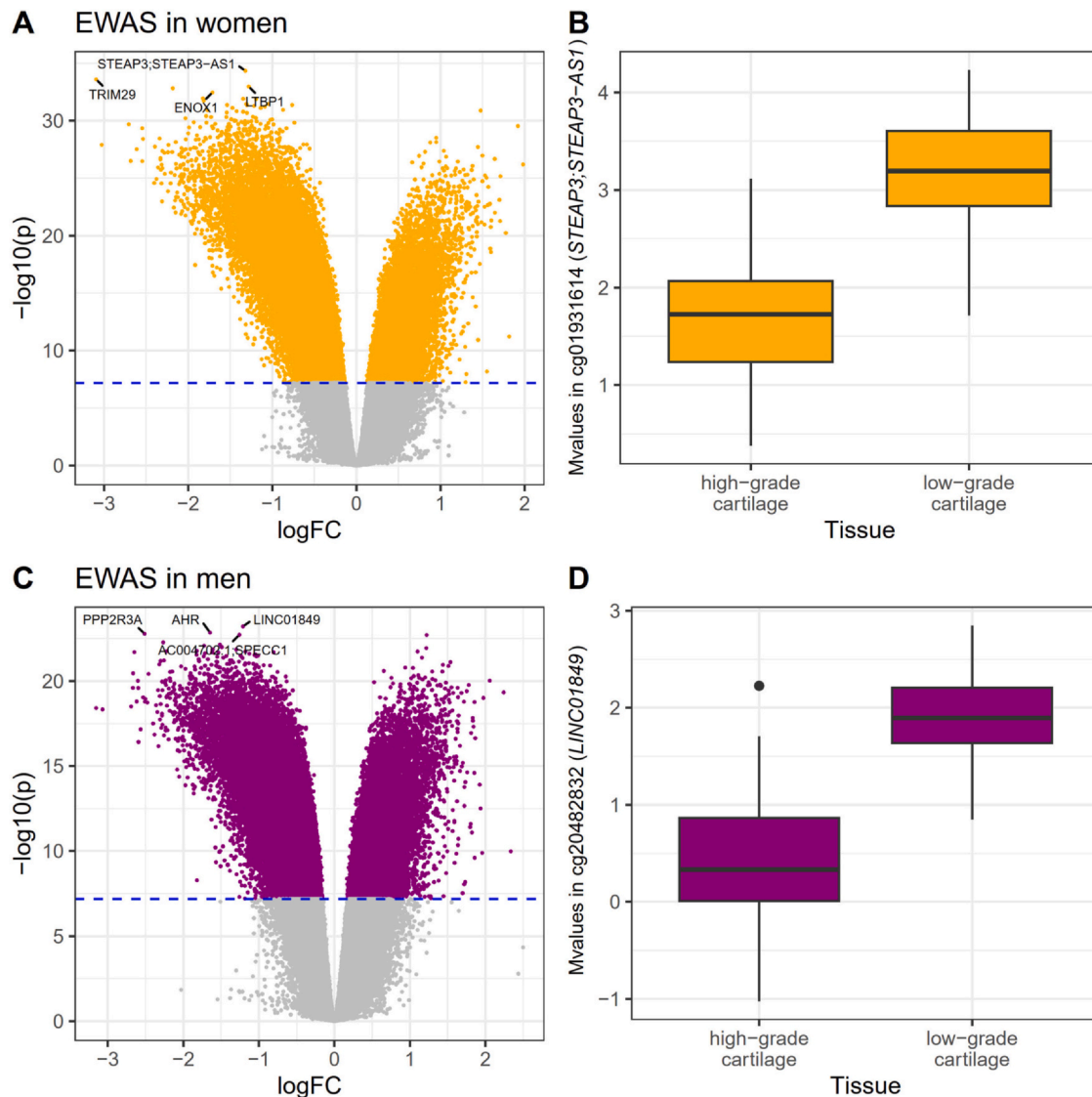**Fig. 2**   Osteoarthritis and Cartilage

Methylation profiles of cartilage degeneration in women and men. (A) A volcano plot visualises 62,313 DMS (20,345 and 41,968 are hyper- and hypomethylated, respectively) in women. (B) Cg01931614 is the most significant DMS in women (logfc: −1.32, p = 4.63 × 10^(−35), SE = 0.06). (C) In Men, 61,513 DMS (20,295 and 41,218 are hyper- and hypomethylated, respectively) are detected. (D) Cg20482832 is the most significant DMS in men (logfc: −1.21, p = 6.17 × 10^(−24), SE = 0.073). Blue lines indicate genome-wide significance (p < 6.41 × 10^−08).

potentially contributing to pain sensation in affected joints.[28] Notably, some nervous system-related terms were identified in only one sex, suggesting sex differences in the innervation and neurotransmission during osteoarthritis progression. This may be related with women being more likely to develop pain in osteoarthritis joints.[13]

Other sex-specific GO terms are related to the immune system, which may be linked with higher pro-inflammatory factor levels in chondrocyte cell cultures (*IL1A*, *IL6*, and *IL8* expression levels in cultured chondrocytes of low-grade osteoarthritis cartilage are higher in women[29]) or an overrepresentation of female patients in a high inflammation cluster.[30] Our results suggest a sex-specific regulative role of cartilage DNA methylation on parts of the immune system during osteoarthritis.

A hormone-related term ("negative regulation of hormone secretion") was only enriched in osteoarthritis markers in men, indicating sex differences in hormone regulation. Previous studies have observed associations between sex hormones and osteoarthritis.[31,32] Furthermore, cultured chondrogenic progenitor cells of osteoarthritis knees have been shown to demonstrate sex-dependent effects of sex hormones on gene expression.[33] Together, this suggests a sex-dependent role of some hormones in osteoarthritis.

Altogether, we compare low-grade (early degeneration state) and high-grade (late degeneration state) osteoarthritis cartilage samples matched from the same patients. By using the largest cohort of its kind, we generate insights at unprecedented power, in turn enabling enhanced insights into the osteoarthritis-related epigenetic signature in cartilage.
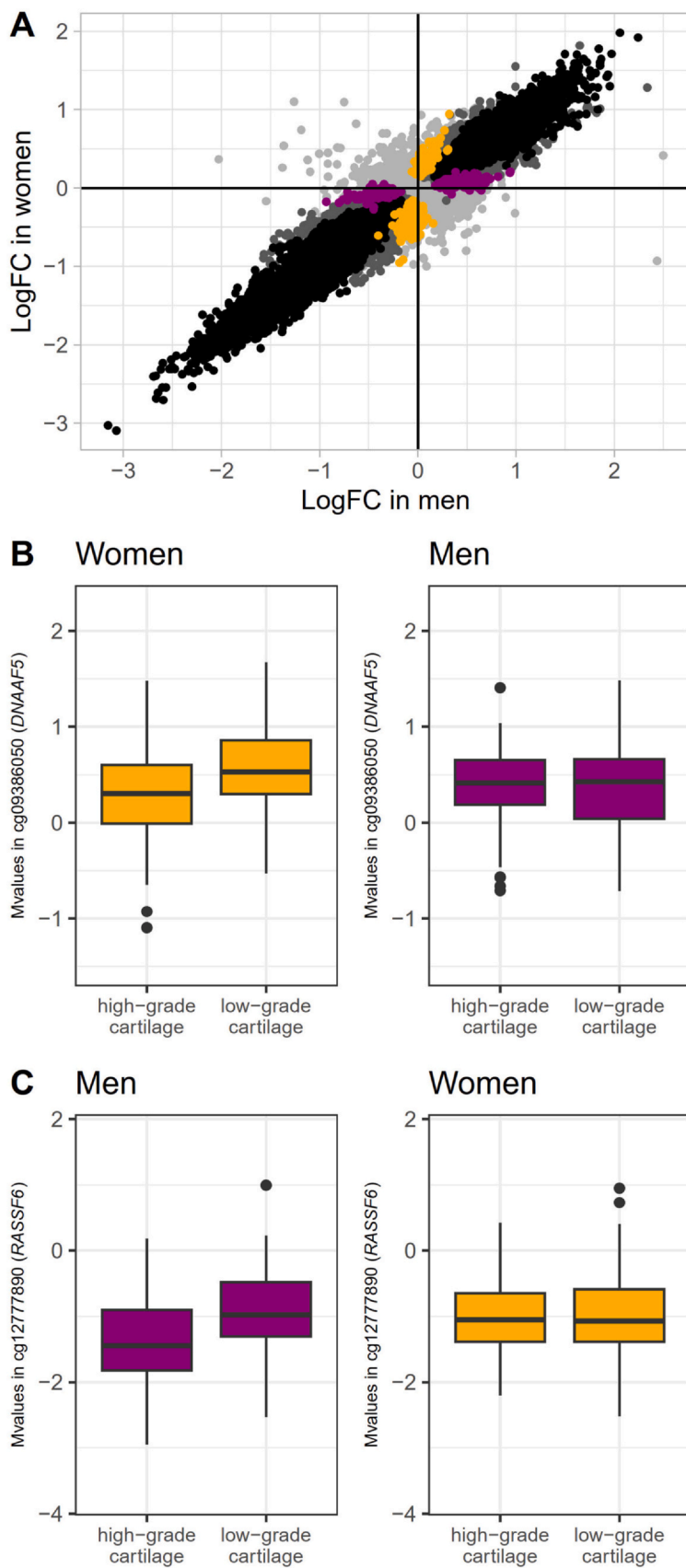
(caption on next page)

Fig. 3

Osteoarthritis and Cartilage

Identification of sex-specific epigenetic markers for cartilage degeneration. (A) Scatterplot comparing effects of DMS in women and men. Orange and purple dots refer to 413 and 539 women- and men-specific DMS, respectively, which are defined by achieving genome-wide significance ($p < 6.41 \times 10^{-08}$) in one sex, but not nominal significance ($p < 0.05$) in the other. Black dots are DMS identified in both sexes. Dark grey dots refer to methylation sites that pass genome-wide significance in only one sex. Other methylation sites are light grey. (B) cg09386050 is a DMS in women (logfc: -0.30, $p = 4.24 \times 10^{-9}$, SE = 0.046), but does not exceed nominal significance in men ($p = 0.83$). (C) Similarly, cg12777890 (logfc: -0.67, $p = 3.35 \times 10^{-14}$, SE = 0.07) is a DMS in men, but not nominal significant in women ($p = 0.13$).

We identify a multitude of methylation sites across the genome that are differentially methylated between these two osteoarthritis stages. These sites are associated ('markers') with osteoarthritis-related cartilage degeneration, thus are linked with osteoarthritis progression rather than disease onset.

In this work, we have studied the methylation profile of primary cartilage from osteoarthritis patients at the point of knee replacement surgery. Therefore, the identified DMS could be a consequence, rather than a cause, of osteoarthritis development. However, access to age-matched healthy cartilage tissue can be challenging. To reveal causal links between osteoarthritis and cartilage methylation, it is necessary to generate methylation quantitative trait locus maps and integrate these with osteoarthritis GWAS results using colocalisation or causal inference analyses.

Our study highlights widespread epigenetic markers for cartilage degeneration linked to a large spectrum of biological pathways, including apoptosis and neuronal development. We reveal large similarities in the epigenetic signature of osteoarthritis across sexes, but also find a number of sex-specific markers, thus providing enhanced insights into the osteoarthritis related epigenetic signature in cartilage.

Author contributions

Study design: E.Z., J.M.W.; Clinical collection: D.S., J.M.W.; Data analysis: P.K.; Interpretation of results: P.K., E.Z.; Manuscript drafting: P.K., E.Z.; Manuscript reviewing and editing: P.K., E.Z., J.M.W.

Conflict of interest

The authors declare no competing interests.

Data Availability

All software used in this study is available from free repositories or manufacturers, as referenced in the Materials and Methods section. Full summary statistics can be obtained online (https://hmgubox2.helmholtz-muenchen.de/index.php/s/pCbEFC9oHNdpNkA).

Acknowledgements

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.joca.2024.07.008.

References

1. Safiri S, Kolahi AA, Smith E, *et al.* Global, regional and national burden of osteoarthritis 1990-2017: a systematic analysis of the Global Burden of Disease Study 2017. Ann Rheum Dis 2020;79(6): 819–28. https://doi.org/10.1136/annrheumdis-2019-216515
2. Boer CG, Hatzikotoulas K, Southam L, *et al.* Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations. Cell 2021;184(18):4784–4818.e17. https://doi.org/10.1016/j.cell.2021.07.038
3. Katsoula G, Kreitmaier P, Zeggini E. Insights into the molecular landscape of osteoarthritis in human tissues. Curr Opin Rheumatol 2022;34(1):79–90. https://doi.org/10.1097/BOR.0000000000000853
4. Kreitmaier P, Suderman M, Southam L, *et al.* An epigenome-wide view of osteoarthritis in primary tissues. Am J Hum Genet 2022;109(7):1255–71. https://doi.org/10.1016/j.ajhg.2022.05.010
5. 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, 34460. https://doi.org/10.1038/srep34460
6. 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. https://doi.org/10.1002/art.39162
7. Zhang Y, Fukui N, Yahata M, *et al.* Genome-wide DNA methylation profile implicates potential cartilage regeneration at the late stage of knee osteoarthritis. Osteoarthritis Cartilage 2016;24(5):835–43. https://doi.org/10.1016/j.joca.2015.12.013
8. 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. https://doi.org/10.1002/jor.22722
9. Bonin CA, Lewallen EA, Baheti S, *et al.* Identification of differentially methylated regions in new genes associated with knee osteoarthritis. Gene 2016;576(0):312–8. https://doi.org/10.1016/j.gene.2015.10.037
10. 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(1):8935. https://doi.org/10.1038/s41598-017-09335-6
11. Srikanth VK, Fryer JL, Zhai G, Winzenberg TM, Hosmer D, Jones G. A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. Osteoarthritis Cartilage 2005;13(9): 769–81. https://doi.org/10.1016/j.joca.2005.04.014
12. de Kruijf M, Stolk L, Zillikens MC, *et al.* Lower sex hormone levels are associated with more chronic musculoskeletal pain in community-dwelling elderly women. Pain 2016;157(7):1425–31. https://doi.org/10.1097/j.pain.0000000000000535
13. McAlindon TE, Cooper C, Kirwan JR, Dieppe PA. Knee pain and disability in the community. Br J Rheumatol 1992;31(3):189–92. https://doi.org/10.1093/rheumatology/31.3.189
14. Odding E, Valkenburg HA, Algra D, Vandenouweland FA, Grobbee DE, Hofman A. Associations of radiological osteoarthritis of the hip

and knee with locomotor disability in the Rotterdam Study. Ann Rheum Dis 1998;57(4):203–8. https://doi.org/10.1136/ard.57.4.203

15. Mainil-Varlet P, Aigner T, Brittberg M, et al. Histological assessment of cartilage repair: a report by the histology endpoint committee of the International Cartilage Repair Society (ICRS). JBJS. 2003;85(suppl_2):45.

16. Steinberg J, Southam L, Roumeliotis TI, et al. A molecular quantitative trait locus map for osteoarthritis. Nat Commun 2021;12(1):1309. https://doi.org/10.1038/s41467-021-21593-7

17. Min JL, Hemani G, Davey Smith G, Relton C, Suderman M. Meffil: efficient normalization and analysis of very large DNA methylation datasets. Bioinformatics 2018;34(23):3983–9. https://doi.org/10.1093/bioinformatics/bty476

18. Arthur R, Schulz-Trieglaff O, Cox AJ, O'Connell J. AKT: ancestry and kinship toolkit. Bioinformatics 2017;33(1):142–4. https://doi.org/10.1093/bioinformatics/btw576

19. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. Genom Data 2016;9:22–4. https://doi.org/10.1016/j.gdata.2016.05.012

20. Pidsley R, Zotenko E, Peters TJ, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol 2016;17(1):208. https://doi.org/10.1186/s13059-016-1066-1

21. Chen Y an, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 2013;8(2):203–9. https://doi.org/10.4161/epi.23470

22. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics 2008;24(13):1547–8. https://doi.org/10.1093/bioinformatics/btn224

23. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 2012;28(6):882–3. https://doi.org/10.1093/bioinformatics/bts034

24. Suderman M, Staley JR, French R, Arathimos R, Simpkin A, Tilling K. dmrff: identifying differentially methylated regions efficiently with power and control. bioRxiv 2018. https://doi.org/10.1101/508556. Published online December 31.

25. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics 2016;32(2):286–8. https://doi.org/10.1093/bioinformatics/btv560

26. Maksimovic J, Oshlack A, Phipson B. Gene set enrichment analysis for genome-wide DNA methylation data. Genome Biol 2021;22(1):173. https://doi.org/10.1186/s13059-021-02388-x

27. Hwang HS, Kim HA. Chondrocyte apoptosis in the pathogenesis of osteoarthritis. Int J Mol Sci 2015;16(11):26035–54. https://doi.org/10.3390/ijms161125943

28. Suri S, Gill SE, Camin SM, de, McWilliams DF, Wilson D, Walsh DA. Neurovascular invasion at the osteochondral junction and in osteophytes in osteoarthritis. Ann Rheum Dis 2007;66(11):1423–8. https://doi.org/10.1136/ard.2006.063354

29. Pan Q, O'Connor MI, Coutts RD, et al. Characterization of osteoarthritic human knees indicates potential sex differences. Biol Sex Differ 2016;7:27. https://doi.org/10.1186/s13293-016-0080-z

30. Steinberg J, Southam L, Fontalis A, et al. Linking chondrocyte and synovial transcriptional profile to clinical phenotype in osteoarthritis. Ann Rheum Dis 2021;80(8):1070–4. https://doi.org/10.1136/annrheumdis-2020-219760

31. Freystaetter G, Fischer K, Orav EJ, et al. Total serum testosterone and western ontario and mcmaster universities osteoarthritis index pain and function among older men and women with severe knee osteoarthritis. Arthritis Care Res 2020;72(11):1511–8. https://doi.org/10.1002/acr.24074

32. Roman-Blas JA, Castañeda S, Largo R, Herrero-Beaumont G. Osteoarthritis associated with estrogen deficiency. Arthritis Res Ther 2009;11(5):241. https://doi.org/10.1186/ar2791

33. Koelling S, Miosge N. Sex differences of chondrogenic progenitor cells in late stages of osteoarthritis. Arthritis Rheum 2010;62(4):1077–87. https://doi.org/10.1002/art.27311