



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/245440/>

Version: Published Version

Article:

Stergiou, O.S., Bittner, N., Katsoula, G. et al. (2026) Molecular clustering in osteoarthritis primary tissues identifies shared inflammatory and tissue-specific pathway profiles.

Osteoarthritis and Cartilage. ISSN: 1063-4584

<https://doi.org/10.1016/j.joca.2026.08.008>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Osteoarthritis and Cartilage



Molecular clustering in osteoarthritis primary tissues identifies shared inflammatory and tissue-specific pathway profiles

Odysseas Sotirios Stergiou^{a b c 1}, Norbert Bittner^{b 2}, Georgia Katsoula^{a b 3}, Diane Swift^{d 4}, Jeremy Mark Wilkinson^{d * 5 7}, Eleftheria Zeggini^{b c ** 6 7}

^a Graduate School of Experimental Medicine, Technical University of Munich, München, Germany

^b Institute of Translational Genomics, Helmholtz Zentrum München Deutsches Forschungszentrum für Gesundheit und Umwelt, Neuherberg, Germany

^c Technical University of Munich (TUM), TUM University Hospital, TUM School of Medicine and Health, Munich, Germany

^d School of Medicine and Population Health, The University of Sheffield, Sheffield, UK

ARTICLE INFO

Article history:

Received 4 February 2026

Received in revised form 26 June 2026

Accepted 13 August 2026

Keywords:

Chondrocytes

Osteoarthritis

Stratification

Clustering

Knee osteoarthritis

Inflammation

ABSTRACT

Objectives: To disentangle the molecular heterogeneity of knee osteoarthritis (OA) through the classification and characterization of transcriptomic clusters in multiple joint tissues, and to uncover distinct biological pathways that will facilitate improved patient stratification.

Methods: We analyzed RNA sequencing data from 330 knee OA patients across low- and high-grade OA knee cartilage, synovium and infrapatellar fat pad tissues. We used unsupervised machine learning to identify distinct transcriptomic clusters and subsequently performed cluster-specific differential expression and pathway enrichment analyses. We applied multi-omics factor analysis in low-grade cartilage to construct a gene expression-based classifier for subtype prediction, which we validated in an independent knee OA RNA sequencing dataset.

Results: We identified robust clusters across all four joint tissues. In low-grade cartilage, we identified two patient groups separated by differences in inflammation and transcriptional regulation. A gene classifier distinguished these groups with a cross-validated accuracy of 94.5% (95% CI 91.1–96.6%). We also reproduced these subtypes in an external cohort in which the same axis similarly separated the subgroups. In high-grade cartilage there were three distinct clusters, characterized by inflammatory, neuroactive receptor-signaling, and housekeeping-transcriptional programs. Both synovium and infrapatellar fat pad showed two distinct subgroups. Despite the histological differences between these two tissues, subgrouping was based on shared biological functions related to immune activation, alongside disease tissue-specific ones.

Conclusions: Our findings identify gene expression-based patient clusters in different primary joint tissues and point to shared and disease tissue-specific molecular programs in OA, thus setting the foundation for transcription signature-based patient stratification.

© 2026 The Author(s). Published by Elsevier Ltd on behalf of Osteoarthritis Research Society International.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding author.

** Corresponding author at: Institute of Translational Genomics, Helmholtz Zentrum München Deutsches Forschungszentrum für Gesundheit und Umwelt, Neuherberg, Germany.

E-mail addresses: odysseas.stergiou@helmholtz-munich.de (O.S. Stergiou), norbert.bittner@helmholtz-munich.de (N. Bittner), georgia.katsoula@helmholtz-munich.de (G. Katsoula), dswift@sheffield.ac.uk (D. Swift), j.m.wilkinson@sheffield.ac.uk (J.M. Wilkinson), eleftheria.zeggini@helmholtz-munich.de (E. Zeggini).

¹ 0000-0001-7734-028X

² 0000-0002-8588-185X

³ 0000-0001-9611-9281

⁴ 0009-0008-9398-0293

⁵ 0000-0001-5577-3674

⁶ 0000-0003-4238-659X

⁷ Indicates equal contribution

<https://doi.org/10.1016/j.joca.2026.08.008>

1063-4584/© 2026 The Author(s). Published by Elsevier Ltd on behalf of Osteoarthritis Research Society International. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Key messages**What is already known on this topic:**

- Osteoarthritis (OA) is a complex and highly heterogeneous disease that exhibits significant variability in its molecular landscape, influenced by differences between tissue types and disease stages.
- Previous studies using clustering algorithms provide insights, but with limited clinical outcome.

What this study adds

- We carried out the largest multi-tissue transcriptomic study of knee OA heterogeneity to date, significantly increasing sample size compared to previous studies.
- We provide the first investigation of infrapatellar fat pad heterogeneity and identify for the first time discrete molecular subgroups in late-stage degeneration cartilage.
- We demonstrate that the subgroups identified in low-grade cartilage, together with their continuous molecular axis, are reproducible in an independent cohort.

How this study might affect research, practice or policy

- These findings provide a basis for the translation of molecular subgroups into clinical research and precision medicine in osteoarthritis.

Introduction

Osteoarthritis (OA) is a highly-prevalent degenerative disease characterized by the gradual deterioration of articular cartilage. It affects 606 million individuals, leading to chronic pain and disability, and its global burden is expected to impact 1 billion individuals by 2050 [1]. Among the different joints in which OA can occur, knee OA is the most common subtype [2]. In addition to age, obesity and mechanical stress, genetic predisposition plays a fundamental role in disease development and progression [3]. Transcriptomic and genomic studies have advanced our comprehension of OA aetiology through genome-wide association studies and molecular quantitative trait locus analyses [4–8].

OA is widely recognized as a clinically and molecularly heterogeneous disease, which involves diverse mechanisms that differ across patients [2]. This heterogeneity is not only reflected by the phenotypic variability of the affected individuals with respect to pain severity, joint involvement and disease progression, but also includes major differences at the molecular level. Such distinct patterns arise from differences in gene expression, immune signaling and extracellular matrix remodeling across patients and tissues [9]. Due to this complexity, traditional classification methods such as radiographic grading or symptom severity scores, have been unable to adequately disentangle the biological diversity of OA that would enable us to move away from therapeutic approaches that only offer symptom relief and no disease modification [10,11]. To tackle this issue, there is an increasing acknowledgement of the need

to stratify OA patients into biologically homogeneous subgroups that may reflect distinct disease endotypes [12]. A first study, by Fernández-Tajes et al. [13], identified a high-inflammation chondrocyte subset in articular cartilage, which was extended by Soul et al. [14] in a 60-patient RNA-sequencing study cohort. Subsequent efforts with increased sample size further helped to resolve these subtypes. Yuan et al. [15] analyzed 131 cartilage samples, and described a distinctive activated-sensory-neuron subtype enriched for neural and synaptic genes, while Steinberg et al. [16] profiled cartilage and synovium from 113 patients, providing the first molecular subgrouping of OA synovium and a gene classifier of cartilage heterogeneity.

Such efforts make use of transcriptomic and other omics-based data to define reproducible molecular subtypes, with the goal of enabling more targeted treatment strategies and improving trial design [14,16]. However, thus far, stratification efforts have focused primarily on cartilage and synovial tissues, leaving the infrapatellar fat pad underexplored, despite its established roles in joint inflammation and metabolic regulation [17].

In this study, we identify molecular subtypes across four primary tissues from up to 330 osteoarthritis patients, leveraging the largest transcriptomic OA cohort to date to improve statistical power for patient stratification. We evaluate the role of infrapatellar fat pad transcriptional heterogeneity in patient stratification.

Materials and methods**i) Patient samples and tissue collection:**

The patients and sample collections have been previously described [6]. Briefly, we recruited patients with osteoarthritis undergoing total knee replacement surgery and collected cartilage, synovium and infrapatellar fat pad tissues. Cartilage samples were categorized based on their degeneration stage using the ICRS scoring system: scores of 0–1 denote "low-grade" OA, and scores of 3–4 denote "high-grade" disease. A subset of this same cohort was also analyzed from Steinberg et al. [16] that included 106 patients (after QC), out of which 87 were low-grade and 95 high-grade cartilage. It also contained 77 synovium samples. All participants provided informed consent prior to the collection of their samples. Tissue samples were obtained under regulatory approval from the National Research Ethics Service (references 15/SC/0132 and 20/SC/0144) and with authorization from the Human Tissue Authority (license number 12182) via the South Yorkshire and North Derbyshire Musculoskeletal Biobank (University of Sheffield, UK).

ii) RNA sequencing and preprocessing

RNA extraction, library preparation, sequencing, and processing were performed as previously described [6,8,16]. Briefly, RNA was extracted using the Qiagen AllPrep RNA Mini Kit. Library preparation followed standard Illumina protocols (TruSeq RNA Sample Prep v.2), with SMART-Seq used for low-input samples. The samples were sequenced across five batches using Illumina HiSeq 2000/4000 and NovaSeq6000 platforms, generating a median of ~80.8 million reads per sample.

Alignment to Ensembl GRCh38 (release 105) genome was performed using STAR v2.7.9a [18], with gene quantification via featureCounts [19] and RSEM (https://github.com/broad-institute/gtex-pipeline/blob/master/rnaseq/src/aggregate_rsem_results.py) [20]. Optical duplicates were marked using Picard (Broad Institute, Picard Toolkit) and dupRadar [21], and extensive QC was conducted using RSeQC [22]. Samples were excluded if they failed any of the following criteria: RNA integrity number (RIN) < 5, < 70% uniquely mapped reads, < 40% reads mapping to genomic features, > 30% reads mapping to intronic regions or rRNA, > 30% strandness failure in RSeQC, or < 10 million total reads. Gene-level QC and sample outlier

detection were performed separately for each tissue type. Lowly expressed genes were filtered using the *filterByExpr* function from edgeR [23]. After quality control, we retained data for 493 cartilage samples, comprising 271 low-grade and 222 high-grade samples (16,991 genes), as well as 98 infrapatellar fat pad samples (15,726 genes) and 287 synovium samples (15,628 genes), resulting in a total of 878 samples across all tissues. Full processing details and data sources are available in the original publications [6,8].

iii) Molecular subgroups:

After quality control, gene counts were normalized using the Trimmed Mean of M-values (TMM) method through edgeR (v4.0.16). For downstream analysis, the data were transformed into log₂ counts per million (CPM). After batch correction, the expression matrix was used as input for ConsensusClusterPlus R (v1.68.0) [24]. The optimal number of clusters was determined using consensus cumulative distribution function (CDF) plots, proportion of ambiguous clustering (PAC) values, and average silhouette scores.

Differential expression analysis between the identified clusters was done with edgeR generalized linear model (GLM) framework. We filtered the lowly expressed genes with *filterByExpr*. TMM normalization factors and dispersions were estimated, and we fitted negative binomial GLMs with contrasts evaluated by likelihood ratio tests (*glmLRT*). As significantly differentially expressed genes we defined those with FDR < 0.05 and |log₂FC| > 0.58 (1.5-fold) to balance biological relevance and statistical stringency. To define distinct gene signatures in high-grade cartilage, we performed pairwise differential expression analysis across all combinations of the three clusters identified (i.e., Cluster 1 vs. Cluster 2, Cluster 1 vs. Cluster 3, and Cluster 2 vs. Cluster 3), and extracted genes uniquely upregulated in each group.

Functional enrichment analysis was performed using ConsensusPathDB (CPDB) (Release 35) [25] and Signaling Pathway Impact Analysis (SPIA) (v2.54.0) [26] using the sets of significantly upregulated genes per cluster as input. CPDB was used for the overrepresentation analysis (ORA) across major pathway databases (WikiPathways, Reactome, KEGG) and Gene Ontology (GO) categories of level 4 and 5. For these analyses, we required a minimum overlap of two genes between our list and the genes of each term and categories with a q-value < 0.05 were considered for further interpretation. SPIA complemented this analysis by assessing pathway perturbation significance, with results reported as FDR-adjusted p-values.

iv) Molecular heterogeneity and gene classifier

To further categorize the molecular heterogeneity beyond discrete clustering, we applied Multi-Omics Factor Analysis (MOFA) (v1.12.1) [27], to capture the major axes of variation within the samples. To check whether the primary factor separated the molecular subgroups we performed Welch's t-test (mean difference with 95% CI). To confirm this separation was due to biological variation, we re-tested it by linear regression adjusting for library size and number of expressed genes. For reproducibility, we independently performed clustering and MOFA on a publicly available RNA-seq dataset [14] and tested whether its primary factor likewise separated its subgroups. We then constructed a classifier using Prediction Analysis for Microarrays (PAMR; v1.57) [28] on the low-grade cartilage dataset, by selecting a minimal gene signature that could distinguish the molecular subgroups. The shrinkage threshold was based on the one-standard error rule, retaining the smallest gene set whose error was within one standard error of the minimum. The performance of the classifier was estimated by 10-fold cross-validation, with 95% confidence intervals, including area under the ROC curve with bootstrap 95% CIs.

v) Cross-study replication of cartilage subgroups

To compare the molecular subgroups we identified in cartilage with previously reported osteoarthritis subtypes, we evaluated their overlap with the cartilage signatures from Soul et al. [14] and a panel of previously reported inflammation markers [13,14,16]. For each subgroup, we tested for enrichment of its upregulated genes (FDR < 0.05, |log₂FC| > 0.58) within the corresponding Soul et al. subtype signature using a one-sided hypergeometric test, restricting the gene universe to genes present in both datasets. One-sided tests were used because the objective was to assess positive enrichment as evidence of cross-study concordance. As a specificity control, we also tested the mismatched pairings (each subgroup against the opposite Soul subtype). We then tested whether inflammation-associated marker genes reported by Fernández-Tajes et al. [13] and Soul et al. [14], as compiled by Steinberg et al. [16], showed the same direction of change in our data, at FDR < 5%.

vi) Within-patient concordance

To better characterize the patients who contributed samples from more than one tissue, we assessed whether molecular subgroups co-occur within individuals across tissues. For each of the six pairwise combinations, we used patients sampled for both tissues and tested using Fisher's exact test whether their subgroup assignments were related. We quantified the strength of the association by Cramér's V. For tissue pairs showing significant associations, we identified the subgroup combinations contributing most strongly to the association using standardized Pearson residuals. P-values for multiple comparisons were adjusted using the Benjamini-Hochberg procedure.

vii) Associations with demographic variables

We examined associations between the identified molecular subgroups and demographic variables (BMI, height, weight, sex, age, and smoking status). For continuous variables, we performed either a parametric test (Welch's t-test) in case of normal distribution, or a non-parametric test (Wilcoxon rank-sum or Kruskal-Wallis for high-grade cartilage) for non-normal distribution, based on the distribution of their quantile-quantile (QQ)-plots. Categorical variables were compared using chi-square tests. The covariates were tested as separate hypotheses with a significance threshold of $p < 0.05$.

viii) Statistical Analysis

This study is exploratory and hypothesis-generating. The proposed subgroupings should be interpreted as candidate molecular states, requiring further validation rather than established disease endotypes. The statistical approach specific to each analysis is described in the corresponding section. Across all analyses, assumptions of data distribution were based on QQ plots. For continuous comparisons, all effect estimates are reported with 95% confidence intervals, and p-values were adjusted by the Benjamini-Hochberg procedure where we tested the multiple hypotheses. All analyses were performed in R version 4.4.1; package versions are reported at first mention, and receiver-operating-characteristic analysis used pROC v1.18.5.

Results

Following quality control, we analyzed the transcriptomic profiles of 878 tissue samples from a total of 330 patients.

Differential OA clusters in cartilage, synovium, and infrapatellar fat pad tissues

After applying consensus clustering and evaluating cluster stability, we identified robust gene expression profile clusters, within each of the studied tissue types. Low-grade cartilage, synovium and

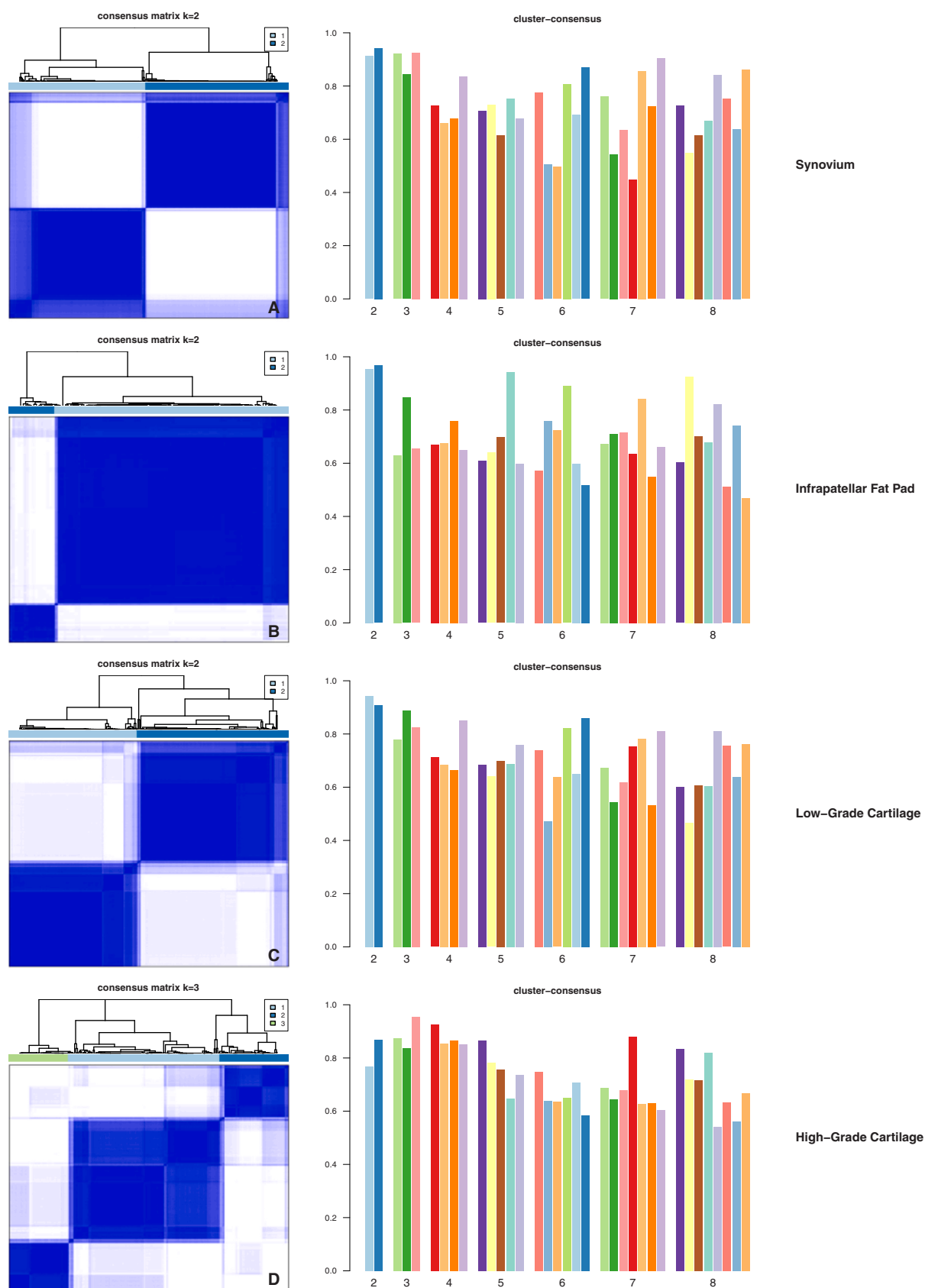


Fig. 1

Consensus clustering across four osteoarthritis tissues. A) Synovium ($k = 2$), B) Infrapatellar Fat Pad ($k = 2$), C) Low-Grade Cartilage ($k = 2$), and D) High-Grade Cartilage ($k = 3$). Left: Consensus matrices showing sample co-clustering stability (blue = high consensus). Right: Cluster-wise consensus scores across different values of k . Optimal cluster number was selected based on cluster stability and biological interpretability.

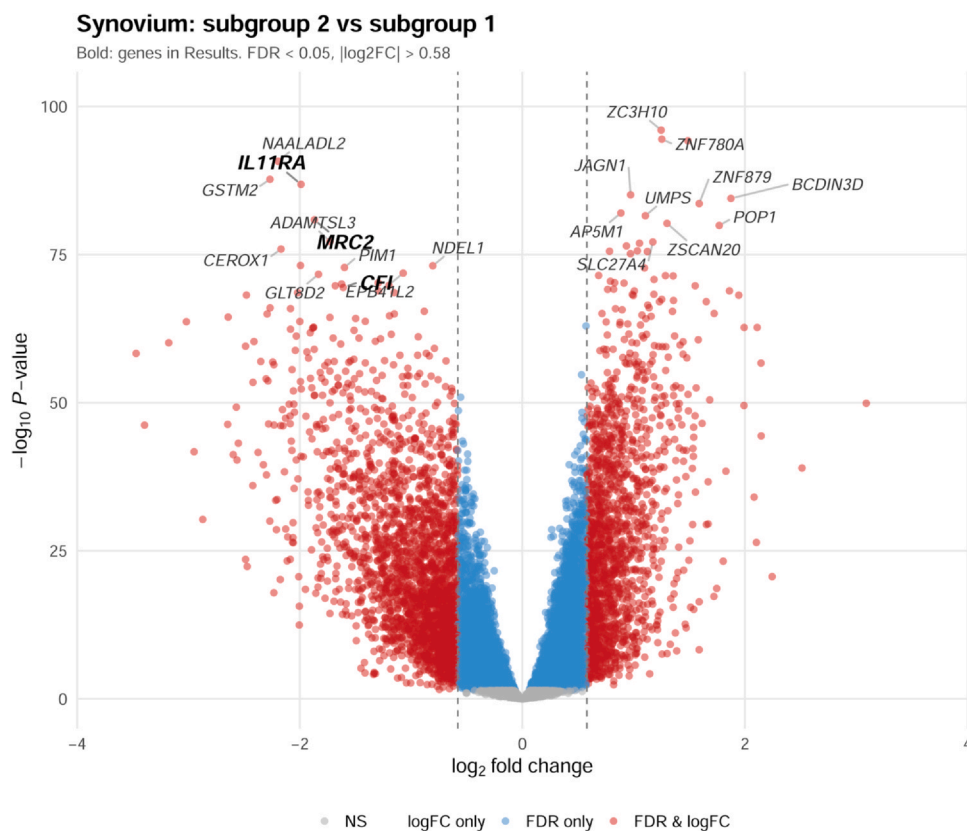
the infrapatellar fat pad were each found to be segregated into two transcriptionally distinct clusters, and high-grade cartilage was divided into three clusters (Fig. 1). The resulting clusters were supported by cluster stability (silhouette score > 0.85 , consensus values > 0.8 , Supplementary Table 31). We performed differential expression analysis (DEA) between clusters to identify differentially expressed genes defining their transcriptional differences (Fig. 2, Supplementary Figures 1,2). Full differential expression results can be found in Supplementary Tables 25–30.

The pathway- and GO-level enrichment of the DE genes revealed that the main biological axes separating the clusters involved shared inflammatory and extracellular matrix remodeling signatures, as well as tissue-specific patterns that reflect the differences in cellular stress, transcriptional and metabolic regulation, neuroactive signaling, and immune modulation (Fig. 3, Table 1). Complete GO-term,

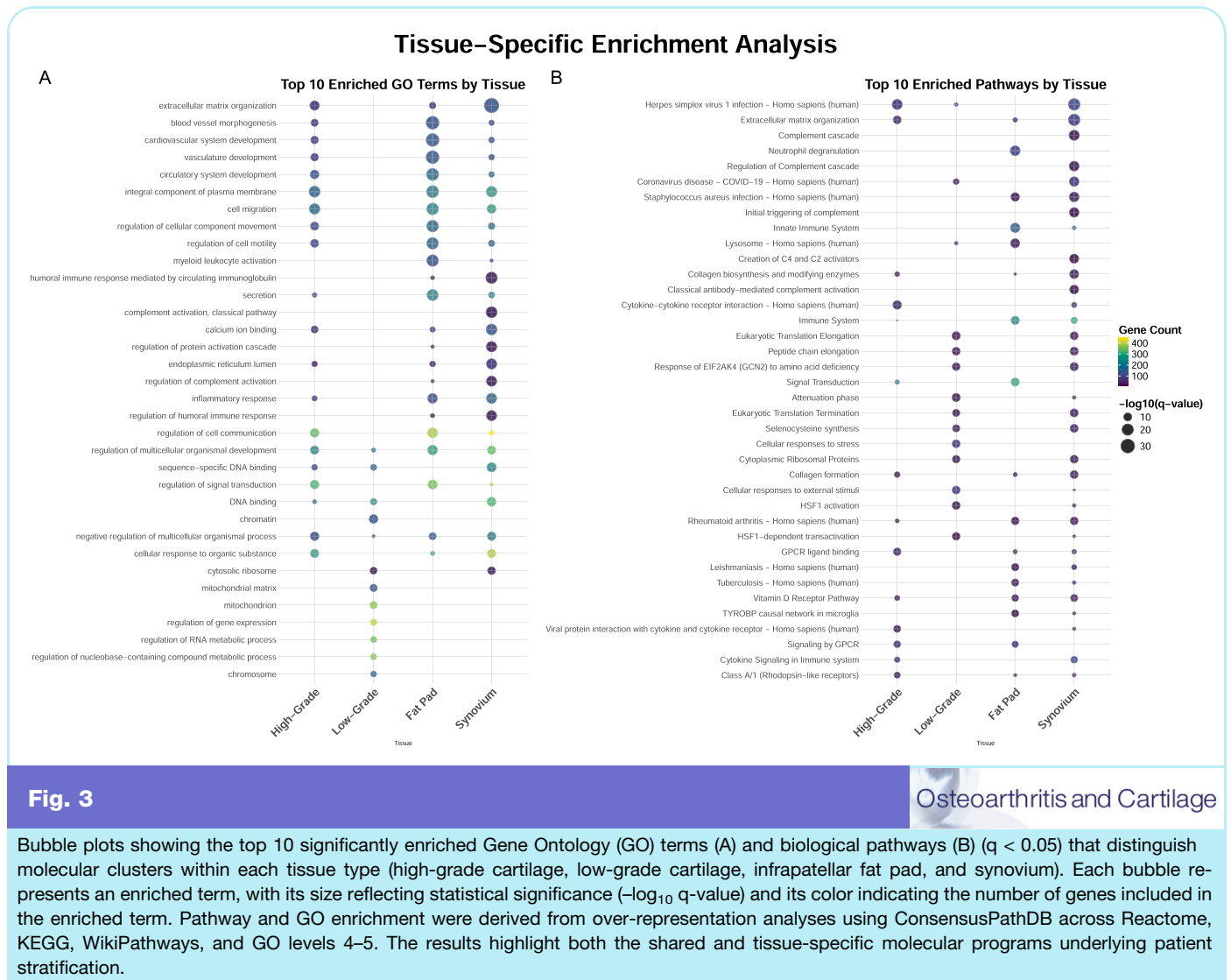
pathway enrichment and SPIA pathway perturbation results across all tissues, can be found in Supplementary Tables 1–24.

Low-grade cartilage

Unsupervised clustering analysis revealed two stable clusters. One cluster showed an inflammatory-catabolic state, whereas the other showed a non-inflammatory state, characterized by transcriptional and metabolic-regulatory activity. Specifically, the first subgroup showed elevated cytokine-cytokine receptor interaction ($FDR = 3.6 \times 10^{-4}$), chemokine signaling ($FDR = 1.4 \times 10^{-3}$) and Toll-like receptor signaling ($FDR = 1.9 \times 10^{-2}$), pointing to an activated inflammatory program, together with enrichment in lysosomal membrane ($q = 1.1 \times 10^{-3}$) and lytic-vacuole membrane ($q = 5.3 \times 10^{-4}$), indicating a lysosomal, degradative state. Representative

**Fig. 2**

Volcano plot of differentially expressed genes between synovium clusters. The x-axis shows $\log_2$ fold change (cluster 2 vs. cluster 1), and the y-axis shows $-\log_{10}(P\text{-value})$. Each dot represents a gene. Genes with $|\log_2FC| > 0.58$ and false discovery rate $FDR < 0.05$ are considered significantly differentially expressed and are color-coded accordingly: red (significant in both FDR and $\log_2FC$), blue (significant by FDR only), green (significant by $\log_2FC$ only), and grey (not significant). Labeled genes comprise those named in the Results (bold) together with the most significant differentially expressed genes on each side.



Bubble plots showing the top 10 significantly enriched Gene Ontology (GO) terms (A) and biological pathways (B) ($q < 0.05$) that distinguish molecular clusters within each tissue type (high-grade cartilage, low-grade cartilage, infrapatellar fat pad, and synovium). Each bubble represents an enriched term, with its size reflecting statistical significance ($-\log_{10} q$ -value) and its color indicating the number of genes included in the enriched term. Pathway and GO enrichment were derived from over-representation analyses using ConsensusPathDB across Reactome, KEGG, WikiPathways, and GO levels 4–5. The results highlight both the shared and tissue-specific molecular programs underlying patient stratification.

upregulated genes included the matrix regulator *IGFBP5*, the cytokine *IL11* and the cell-cycle regulator *CCND1*. The second cluster was enriched in transcriptional and metabolic programs such as regulation of cellular metabolic process ($q = 1.2 \times 10^{-4}$), polysomal ribosome ($q = 1.3 \times 10^{-4}$) and macromolecule biosynthetic process ($q = 1.7 \times 10^{-4}$), while also having stress and cell-cycle signatures (regulation of cell cycle $q = 6.0 \times 10^{-3}$, cellular response to heat $q = 3.2 \times 10^{-2}$), indicating a low-inflammatory state characterized by transcriptional and metabolic regulation (Table 1).

We applied Multi-Omics Factor Analysis (MOFA) to identify continuous axes of variation across samples. Factor 1 accounted for the largest share of transcriptomic variance (47.5%) and was associated with the separation between clusters (mean difference 1.95, 95% CI 1.54–2.35; Welch $t = 9.47$; $p = 1.2 \times 10^{-17}$), a separation that remained significant after accounting for sequencing depth (adjusted $\beta = -1.88$, 95% CI -2.28 to -1.48 ; $p = 1.3 \times 10^{-17}$). To examine the robustness of this axis, we repeated the clustering and MOFA analysis independently in an external low-grade cartilage RNA-seq dataset (Soul et al.; $n = 60$) [14]. Factor 1 in the external dataset explained 12.7% of variance and again separated the two subclusters (mean difference 2.02, 95% CI 1.68–2.36; $t = 11.84$; $p < 2.2 \times 10^{-16}$), supporting the reproducibility of this primary molecular axis in an independent cohort.

Lastly, we trained a PAMR classifier on the low-grade cartilage dataset. PAMR identified a 14-gene signature that effectively distinguished between the two subtypes (Supplementary Table 32) at 94.5% cross-validated accuracy (95% CI 91.1–96.6%). This gene set includes several with known roles in cartilage biology, such as *NR4A1*, a transcription factor involved in inflammatory signaling and chondrocyte apoptosis [29], *MAGED1*, a gene that putatively contributes to OA progression [30], *CBX4*, a chromatin regulator with roles in senescence and epigenetic modulation in OA [31], and the long non-coding RNAs *SNHG12* and *SNHG1*, which exert opposing effects on cartilage inflammation [32,33].

High-grade cartilage

In high-grade cartilage, we identified three patient clusters, each reflecting a unique molecular profile: inflammatory matrix remodeling, neuroactive receptor-signaling and housekeeping transcription (Table 1). To characterize cluster-specific profiles, we performed pairwise differential expression analysis across all cluster combinations and defined, for each cluster, the upregulated genes relative to the other two groups. One cluster was characterized by immune activation and matrix remodeling signatures, with pathway analysis identifying cytokine-cytokine receptor interaction ($FDR = 5.5 \times 10^{-17}$), extracellular matrix receptor interaction

Tissue Type	Proposed Clusters	Enriched Pathways, GO terms
Synovium	2	Cluster 1: Inflammatory, ECM remodeling, complement activation Cluster 2: RNA processing and methyltransferase
Fat Pad	2	Cluster 1: Vascular, endothelial, axon guidance Cluster 2: Innate immune activation with chondrogenic gene program
Cartilage:		
High Grade	3	Cluster 1: Housekeeping transcription with mild cytokine engagement Cluster 2: Neuroactive, GPCR, ion-channel, hormone signaling Cluster 3: Inflammatory, ECM remodeling, migratory
Low Grade	2	Cluster 1: Inflammatory, catabolic, lysosomal Cluster 2: Low-inflammatory, transcriptional/metabolic regulation

Table 1

Osteoarthritis and Cartilage

Overview of defined patient clusters across four osteoarthritis joint tissues, including the number of clusters identified per tissue and representative enriched biological signatures. Enrichment results reflect intra-tissue comparisons based on differential expression analysis between clusters, followed by pathway and Gene Ontology (GO) term over-representation analysis.

(FDR = 1.7×10^{-5}) and chemokine signaling (FDR = 1.3×10^{-4}) as elevated, alongside upregulation of genes such as *TNF* and the matrix remodeling enzymes *LOXL2* and *TIMP3*. Another cluster was associated with a neuroactive and ligand-receptor profile, as indicated by GO terms such as receptor ligand activity ($q = 4.3 \times 10^{-4}$), hormone activity ($q = 2.0 \times 10^{-3}$) and ion-gated channel activity ($q = 5.5 \times 10^{-3}$), and enrichment of the neuroactive ligand-receptor interaction pathway ($q = 8.6 \times 10^{-3}$). Key genes here included multiple neuroactive families, like the thyrotropin-releasing hormone (TRH) and the cannabinoid receptor *CNR2*. The third cluster was enriched in pathways related to metal-ion and nucleic-acid binding (metal ion binding $q = 1.3 \times 10^{-3}$; DNA binding $q = 1.4 \times 10^{-3}$), with cytokine-cytokine receptor interaction (FDR = 3.3×10^{-3}) as its only significant pathway.

Synovium

In synovial tissue, we identified two transcriptional clusters showing the previously reported inflammatory axis (Table 1) [16]. One cluster exhibited strong immune activation together with an extracellular matrix signature. Enrichment analysis pointed to extracellular matrix organization ($q = 7.5 \times 10^{-23}$), humoral immune response mediated by circulating immunoglobulin ($q = 7.6 \times 10^{-10}$) and complement activation through the classical pathway ($q = 2.7 \times 10^{-8}$), while pathway analysis also identified extracellular matrix-receptor interaction (FDR = 2.0×10^{-9}), complement and coagulation cascades (FDR = 3.3×10^{-6}) and cytokine-cytokine receptor interaction (FDR = 9.1×10^{-6}) as elevated. Representative upregulated genes include the macrophage receptor *MRC2*, the interleukin-11 receptor, *IL11RA* and complement factor I (*CFI*). In contrast, the second cluster was enriched in RNA-metabolic and methyltransferase processes (RNA metabolic process $q = 2.5 \times 10^{-4}$, methyltransferase activity $q = 2.9 \times 10^{-4}$, RNA methylation $q = 6.4 \times 10^{-4}$). No pathways in this cluster were significantly elevated.

Infrapatellar fat pad (IPFP)

Our study introduces the first transcriptomic stratification of the infrapatellar fat pad (IPFP), revealing two distinct patient subgroups characterized by divergent immune, chondrogenic and vascular signatures (Table 1). One cluster combined an innate-immune program with a chondrogenic signature, enriched in myeloid leukocyte activation ($q = 3.3 \times 10^{-10}$), neutrophil degranulation ($q = 4.7 \times 10^{-7}$) and inflammatory response ($q = 3.2 \times 10^{-4}$), and pathway analysis identified osteoclast differentiation (FDR = 1.29×10^{-4}), lysosome (FDR = 2.9×10^{-4}) and cytokine-cytokine receptor interaction (FDR = 3.8×10^{-4}) as elevated. The chondrogenic part was reflected in upregulation of the chondrogenic regulator *SOX5* and the cartilage matrix gene *PRG4*. The second cluster displayed a vascular and neural-guidance signature. This group was enriched in blood vessel morphogenesis ($q = 2.0 \times 10^{-15}$) and vasculature development ($q = 2.0 \times 10^{-15}$) and pathway analysis identified axon guidance (FDR = 1.2×10^{-5}) and extracellular matrix-receptor interaction (FDR = 3.3×10^{-5}) as elevated, with representative genes such as the endothelial transcription factors *ERG* and *TAL1*, as well as the VEGF receptors *FLT1* and *FLT4*.

Cross-study replication of the low-grade cartilage subgroups

The subgroup we labeled as inflammatory-catabolic significantly overlapped with the high-inflammation cartilage signature of Soul et al. [14] (378 of 451 genes; $p = 2.3 \times 10^{-22}$), while the transcriptional and metabolic subgroup overlapped with their non-inflammatory signature (102 of 262 genes; $p = 4.6 \times 10^{-7}$). The mismatched pairings were non-significant (both $p \approx 1.0$). Of nine inflammation markers tested, six (*KYNU*, *PLEK*, *LY96*, *LAPTM5*, *COL1A1* and *AQP1*) were concordantly upregulated in our inflammatory-catabolic subgroup

(FDR < 5%), two (*ITGAM*, *GPNMB*) showed the same direction without reaching significance, and one (*LCP2*) was not detected.

Within-patient concordance of subgroup membership across tissues

For patients contributing more than one tissue, subgroup membership co-occurred significantly only between low-grade and high-grade cartilage ($n = 196$; adjusted $p = 1.0 \times 10^{-13}$; Cramér's $V = 0.54$). The strongest association was found between the transcriptional-metabolic low-grade subgroup and the neuroactive high-grade one (47 patients; adjusted standardized residual +7.3). The remaining tissue pairs showed no significant association.

Molecular subgroupings and demographic characteristics

We detected no significant associations between the subclusters identified and demographic characteristics, including age, sex, BMI, and smoking status ($p > 0.05$), across the different tissues.

Discussion

We have conducted the most extensive analysis of osteoarthritis transcriptional subtypes to date, consisting of 878 samples across low-grade and high-grade cartilage, synovium, and, for the first time, the infrapatellar fat pad. Across these tissues, the previously-outlined heterogeneity points to a consistent structure. An inflammatory axis recurred as a defining feature of every tissue, in accordance with previous studies in cartilage and synovium [13,14,16], while each also had a tissue-specific second axis. Previous studies have proposed several molecular endotypes, mainly the inflammatory, low-repair, metabolic and bone-remodeling states [34,35].

In low-grade cartilage, the inflammatory pole of the axis corresponds to the inflammatory endotype of osteoarthritis, the most clinically advanced of the proposed endotypes and the one consistently linked to pain and progression [35]. Enrichment in inflammatory pathways is consistent with the inflammatory chondrocyte phenotype that single-cell atlases have resolved in degraded cartilage [36]. The second subgroup is defined by the absence of this inflammatory-catabolic signature. Such a non-inflammatory subtype has similarly emerged from single-cell data stratification [36]. A further study on blood biomarkers has identified a low tissue turnover endotype, and showed that disease was less likely to progress in these patients [34]. Whether this is genuinely protective cannot be resolved from bulk transcriptomics data alone, and will require single-cell resolution and linkage to clinical outcome data [35]. The robustness of the axis itself is supported by its reproducibility in an independent knee OA cohort [14] and by a classifier that distinguishes the subgroups with high accuracy.

In high-grade cartilage, we resolved three stable clusters, a within-grade stratification not achieved in earlier studies [16]. The existence of distinct transcriptional states indicates that chondrocytes do not converge into a single degenerative profile, but retain transcriptional heterogeneity between patients. Among the three subgroups, we defined a neuroactive state. A neural-associated subtype was also previously reported by Yuan et al., which they linked to pain [15]. The independent emergence of neural-related signaling in osteoarthritic cartilage points to it as a recurring axis of disease heterogeneity. This matters because pain, even though it is the dominant clinical burden of osteoarthritis, cannot be explained by structural damage alone [35]. Among the neuroactive receptors marking this state are the $\alpha 2$ -adrenergic receptors ADRA2A, ADRA2B and ADRA2C, which we find coordinately upregulated in this subgroup. Notably, $\alpha 2$ -adrenergic signaling has been recently shown to drive chondrocyte hypertrophic differentiation, and its inhibition is proposed as a cartilage regeneration strategy [37].

Our study provides the first molecular stratification of the infrapatellar fat pad in osteoarthritis. Although this tissue is considered an active player in joint inflammation [38], its transcriptomic features have only recently begun to be systematically studied [8,39]. In our analysis of nearly 100 infrapatellar fat pad samples, we identified two subgroups: one combining innate-immune activation with a chondrogenic signature, and a second defined by vascular and neural-guidance programs. The chondrogenic pairing is biologically plausible: the fat pad is an established reservoir of mesenchymal progenitors with chondrogenic potential [40], while cartilage-matrix genes including *PRG4* have been reported in a single-nucleus RNA sequencing study of osteoarthritis fat-pad fibroblasts [38]. The vascular-neural subgroup is consistent with the dense innervation and vascularization of the fat pad, features that have also been implicated in osteoarthritic knee pain [41].

Across the four tissues, a shared immune-inflammatory program was evident in osteoarthritis, alongside tissue-specific signatures: neuroactive and metabolic-regulatory states in cartilage, RNA-processing and methyltransferase activity in synovium, and vascular-neural and chondrogenic programs in the fat pad. Within patients, subgroup assignments were correlated between low- and high-grade cartilage but not across other tissues. These molecular endotypes appear to be properties of individual joint tissues rather than of the patient as a whole. This highlights the need to consider the local tissue microenvironment when designing biomarkers or developing targeted interventions for OA [42], since these differences most likely reflect their unique cellular architecture, exposure to mechanical stress, and the degree of vascular and immune system involvement [43,44].

While this study expands our understanding of osteoarthritis heterogeneity, some aspects may limit immediate clinical translation. Because samples were collected at a single surgical timepoint, we could not follow how subgroup membership relates to disease progression, and the variables available to us were demographic rather than clinical, showing no association with subgroup membership. Therefore, we are unable to establish whether these distinctions track with clinical outcomes or treatment responses. This would require longitudinal phenotyping, ideally through minimally invasive samples such as synovial fluid or blood. Validating these subgroups with external, independent datasets across all tissue types will be essential to assess their robustness and generalizability.

Bulk RNA sequencing averages over cell types, so states such as the fat-pad chondrogenic program cannot be assigned to specific cell populations. The increasing number of single-cell studies in cartilage and synovium [45,46] provide promising opportunities to deconvolute or validate these bulk-defined axes of heterogeneity. Integrating further data layers, including proteomics, spatial profiling, and epigenetic marks, will refine these classifications, and advance more precise, tissue-aware management of osteoarthritis.

Author contributions

Study design: E.Z., J.M.W.; Clinical collection: J.M.W., D.S.; Gene Expression data: G.K.; Data analysis: O.S.S.; Interpretation of results: O.S.S., E.Z., N.B., J.M.W.; Manuscript drafting: O.S.S., G.K., N.B., E.Z.; Manuscript reviewing and editing: all authors

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT in order to improve readability of the text. After using this tool/service,

the authors reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Declaration of Competing Interest

The authors declare no competing interests.

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, Dr. Gertrud Eckstein and Dr. Jenny Hankinson for their help in sample handling and additional processing of RNA-seq and WGS data. We also thank all study participants for their willingness to donate study samples. The work on osteoarthritis cartilage samples was funded by Wellcome Trust (206194).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.joca.2026.08.008.

References

- [1] Z. Wang, Z. Xiao, C. Sun, G. Xu, J. He, Global, regional and national burden of osteoarthritis in 1990–2021: a systematic analysis of the global burden of disease study 2021, *BMC Musculoskelet. Disord.* 25 (2024) 1021.
- [2] S. Tang, C. Zhang, W.M. Oo, K. Fu, M.A. Risberg, S.M. Bierma-Zeinstra, et al., Osteoarthritis, *Nat. Rev. Dis. Prim.* 11 (2025) 10.
- [3] M. Fernandez-Moreno, I. Rego, V. Carreira-García, F.J. Blanco, Genetics in osteoarthritis, *Curr. Genom.* 9 (2008) 542–547.
- [4] P. Kreitmaier, D. Swift, J.M. Wilkinson, E. Zeggini, Epigenomic differences between osteoarthritis grades in primary cartilage, *Osteoarthr. Cartil.* 32 (2024) 1126–1133.
- [5] J. Steinberg, L. Southam, T.I. Roumeliotis, M.J. Clark, R.L. Jayasuriya, D. Swift, et al., A molecular quantitative trait locus map for osteoarthritis, *Nat. Commun.* 12 (2021) 1309.
- [6] G. Katsoula, J.E.G. Lawrence, A.L. Arruda, M. Tutino, P. Balogh, L. Southam, et al., Primary cartilage transcriptional signatures reflect cell-type-specific molecular pathways underpinning osteoarthritis, *Am. J. Hum. Genet.* 111 (2024) 2735–2755.
- [7] G. Katsoula, J. Steinberg, M. Tuerlings, R. Coutinho de Almeida, L. Southam, D. Swift, et al., A molecular map of long non-coding RNA expression, isoform switching and alternative splicing in osteoarthritis, *Hum. Mol. Genet.* 31 (2022) 2090–2105.
- [8] G. Katsoula, A.L. Arruda, M. Tutino, et al., Transcriptome analysis in osteoarthritis primary tissues identifies high-confidence effector genes, *Nat Commun* 17 (2026) 8128.
- [9] L. Tong, H. Yu, X. Huang, J. Shen, G. Xiao, L. Chen, et al., Current understanding of osteoarthritis pathogenesis and relevant new approaches, *Bone Res.* 10 (2022) 60.
- [10] W.M. Oo, C. Little, V. Duong, D.J. Hunter, The development of disease-modifying therapies for osteoarthritis (DMOADs): the evidence to date, *Drug. Des. Devel. Ther.* 15 (2021) 2921–2945.
- [11] M.D. Brandt, J.B. Malone, T.J. Kean, Advances and challenges in the pursuit of disease-modifying osteoarthritis drugs: a review of 2010–2024 clinical trials, *Biomedicine* 13 (2025) 355.
- [12] J.B. Driban, M.R. Sitler, M.F. Barbe, E. Balasubramanian, Is osteoarthritis a heterogeneous disease that can be stratified into subsets? *Clin. Rheumatol.* 29 (2010) 123–131.
- [13] J. Fernández-Tajes, A. Soto-Hermida, M.E. Vázquez-Mosquera, E. Cortés-Pereira, A. Mosquera, M. Fernández-Moreno, et al., Genome-wide DNA methylation analysis of articular chondrocytes reveals a cluster of osteoarthritic patients, *Ann. Rheum. Dis.* 73 (2014) 668–677.
- [14] J. Soul, S.L. Dunn, S. Anand, F. Serracino-Inglott, J.-M. Schwartz, R.P. Boot-Handford, et al., Stratification of knee osteoarthritis: two major patient subgroups identified by genome-wide expression analysis of articular cartilage, *Ann. Rheum. Dis.* 77 (2018) 423.
- [15] C. Yuan, Z. Pan, K. Zhao, J. Li, Z. Sheng, X. Yao, et al., Classification of four distinct osteoarthritis subtypes with a knee joint tissue transcriptome atlas, *Bone Res.* 8 (2020) 38.
- [16] J. Steinberg, L. Southam, A. Fontalis, M.J. Clark, R.L. Jayasuriya, D. Swift, et al., Linking chondrocyte and synovial transcriptional profile to clinical phenotype in osteoarthritis, *Ann. Rheum. Dis.* 80 (2021) 1070–1074.
- [17] B. Pereira Herrera, K. Emanuel, P.J. Emans, M. van Griensven, B. Cillero-Pastor, Infrapatellar fat pad as a source of biomarkers and therapeutic target for knee osteoarthritis, *Arthritis Res. Ther.* 27 (2025) 81.
- [18] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, et al., STAR: ultrafast universal RNA-seq aligner, *Bioinformatics* 29 (2013) 15–21.
- [19] Y. Liao, G.K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, *Bioinformatics* 30 (2014) 923–930.
- [20] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, *BMC Bioinforma.* 12 (2011) 323.
- [21] S. Sayols, D. Scherzinger, H. Klein, dupRadar: a Bioconductor package for the assessment of PCR artifacts in RNA-Seq data, *BMC Bioinforma.* 17 (2016) 428.
- [22] L. Wang, S. Wang, W. Li, RSeQC: quality control of RNA-seq experiments, *Bioinformatics* 28 (2012) 2184–2185.
- [23] Y. Chen, L. Chen, A.T.L. Lun, P.L. Baldoni, G.K. Smyth, edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets, *Nucleic Acids Res.* 53 (2025) gkaf018.
- [24] M.D. Wilkerson, D.N. Hayes, ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking, *Bioinformatics* 26 (2010) 1572–1573.
- [25] R. Hervig, C. Hardt, M. Lienhard, A. Kamburov, Analyzing and interpreting genome data at the network level with ConsensusPathDB, *Nat. Protoc.* 11 (2016) 1889–1907.
- [26] A.L. Tarca, S. Draghici, P. Khatri, S.S. Hassan, P. Mittal, J.-S. Kim, et al., A novel signaling pathway impact analysis, *Bioinforma. Oxf. Engl.* 25 (2009) 75–82.
- [27] R. Argelaguet, B. Velten, D. Arnol, S. Dietrich, T. Zenz, J.C. Marioni, et al., Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets, *Mol. Syst. Biol.* 14 (2018) e8124.
- [28] R. Tibshirani, T. Hastie, B. Narasimhan, G. Chu, Diagnosis of multiple cancer types by shrunken centroids of gene expression, *Proc. Natl. Acad. Sci.* 99 (2002) 6567–6572.
- [29] X. Shi, H. Ye, X. Yao, Y. Gao, The involvement and possible mechanism of NR4A1 in chondrocyte apoptosis during osteoarthritis, *Am. J. Transl. Res.* 9 (2017) 746–754.
- [30] D. Zhao, L. Zeng, G. Liang, M. Luo, J. Pan, Y. Dou, et al., Transcriptomic analyses and machine-learning methods reveal dysregulated key genes and potential pathogenesis in human osteoarthritic cartilage, *Bone Jt. Res.* 13 (2024) 66–82.
- [31] Y.-H. Chen, X. Zhang, D. Attarian, V.B. Kraus, Synergistic roles of CBX4 chromo and SIM domains in regulating senescence of primary human osteoarthritic chondrocytes, *Arthritis Res. Ther.* 25 (2023) 197.
- [32] X. Yang, H. Chen, H. Zheng, K. Chen, P. Cai, L. Li, et al., LncRNA SNHG12 Promotes Osteoarthritis Progression Through Targeted Down-Regulation of miR-16-5p, *Clin. Lab.* 68 (2022).
- [33] J. Lei, Y. Fu, Y. Zhuang, K. Zhang, D. Lu, LncRNA SNHG1 alleviates IL-1 β -induced osteoarthritis by inhibiting miR-16-5p-mediated p38 MAPK and NF- κ B signaling pathways, *Biosci. Rep.* 39 (2019) BSR20191523.
- [34] F. Angelini, P. Widera, A. Mobasheri, J. Blair, A. Struglics, M. Uebelhoefer, et al., Osteoarthritis endotype discovery via clustering of biochemical marker data, *Ann. Rheum. Dis.* 81 (2022) 666–675.
- [35] M.A. Karsdal, L.C. Rovati, J. Tambiah, O. Kubassova, C. Ladel, F. Berenbaum, et al., The inflammatory endotype in osteoarthritis: Reflections from the 2024 OARSI clinical trials symposium (CTS) with a special emphasis on feasibility for clinical development, *Osteoarthr. Cartil. Open.* 7 (2025) 100572.
- [36] Y. Fan, X. Bian, X. Meng, L. Li, L. Fu, Y. Zhang, et al., Unveiling inflammatory and prehypertrophic cell populations as key contributors to knee cartilage degeneration in osteoarthritis using multi-omics data integration, *Ann. Rheum. Dis.* 83 (2024) 926–944.
- [37] X. Wei, J. Qiu, R. Lai, T. Wei, Z. Lin, S. Huang, et al., A human organoid drug screen identifies α 2-adrenergic receptor signaling as a therapeutic target for cartilage regeneration, *Cell. Stem Cell.* 31 (2024) 1813–1830.e8.
- [38] H. Peters, P. Potla, J.S. Rockel, T. Tockovska, C. Pastrello, I. Jurisica, et al., Cell and transcriptomic diversity of infrapatellar fat pad during knee osteoarthritis, *Ann. Rheum. Dis.* 84 (2025) 351–367.
- [39] R. Gandhi, M. Takahashi, C. Virtanen, K. Syed, J.R. Davey, N.N. Mahomed, Microarray analysis of the infrapatellar fat pad in knee osteoarthritis: relationship with joint inflammation, *J. Rheumatol.* 38 (2011) 1966–1972.
- [40] P. Hindle, N. Khan, L. Biant, B. Péault, The infrapatellar fat pad as a source of perivascular stem cells with increased chondrogenic potential for regenerative medicine, *Stem Cells Transl. Med.* 6 (2017) 77–87.
- [41] M.G. Wang, P. Seale, D. Furman, The infrapatellar fat pad in inflammation, knee joint health, and osteoarthritis, *npj Aging* 10 (2024) 34.
- [42] D.J. Hunter, S. Bierma-Zeinstra, Osteoarthritis, *Lancet* 393 (2019) 1745–1759.
- [43] C.R. Scanzello, S.R. Goldring, The role of synovitis in osteoarthritis pathogenesis, *Bone* 51 (2012) 249–257.
- [44] L. Zhang, J. Hu, K.A. Athanasiou, The role of tissue engineering in articular cartilage repair and regeneration, *Crit. Rev. Biomed. Eng.* 37 (2009) 1–57.
- [45] C. Huang, B. Zeng, B. Zhou, G. Chen, Q. Zhang, W. Hou, et al., Single-cell transcriptomic analysis of chondrocytes in cartilage and pathogenesis of osteoarthritis, *Genes. Dis.* 12 (2025) 101241.
- [46] Z.Y. Huang, Z.Y. Luo, Y.R. Cai, C.-H. Chou, M.L. Yao, F.X. Pei, et al., Single cell transcriptomics in human osteoarthritis synovium and *in silico* deconvoluted bulk RNA sequencing, *Osteoarthr. Cartil.* 30 (2022) 475–480.