



Universiteit
Leiden
The Netherlands

The bone and cartilage interplay in osteoarthritis: key to effective treatment strategy

Tuerlings, M.

Citation

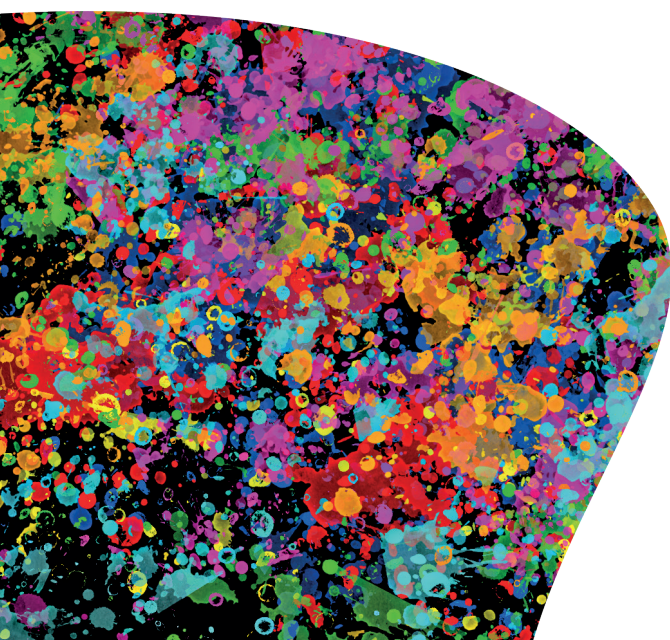
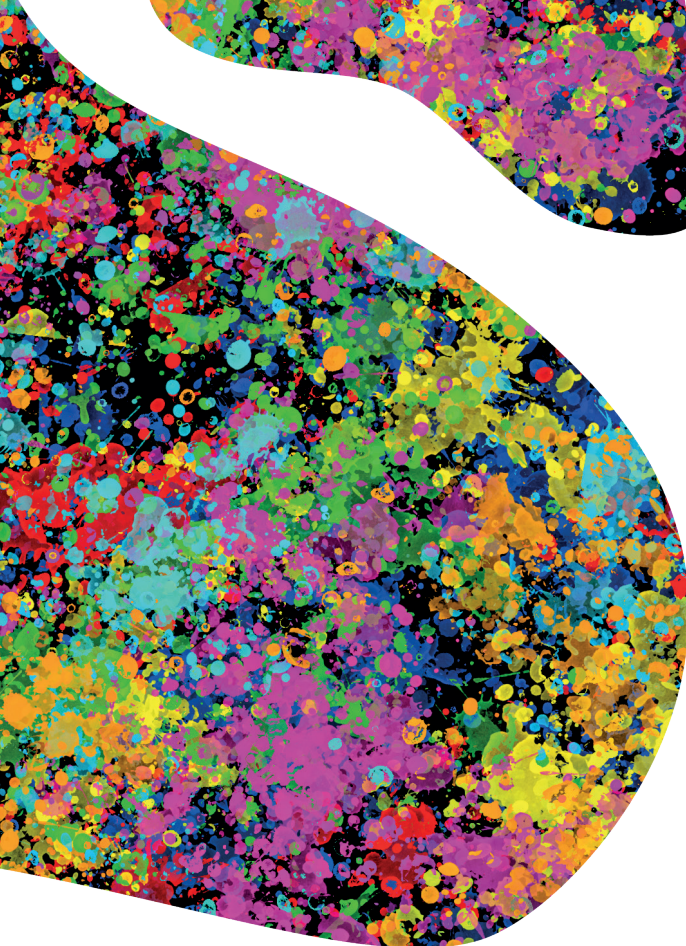
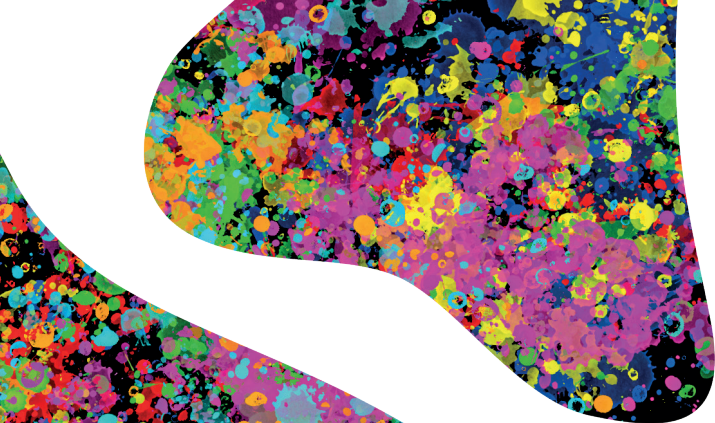
Tuerlings, M. (2023, September 27). *The bone and cartilage interplay in osteoarthritis: key to effective treatment strategy*. Retrieved from <https://hdl.handle.net/1887/3642518>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3642518>

Note: To cite this publication please use the final published version (if applicable).



CHAPTER 4



Identification of circulating microRNAs predicting osteoarthritis molecular endotypes and matching druggable targets

Margo Tuerlings¹, Ilja Boone¹, H. Eka D. Suchiman¹, Nico Lakenberg¹, Robert J.P. van der Wal², Rob G.H.H. Nelissen², Yolande F.M. Ramos¹, Rodrigo Coutinho de Almeida¹, Ingrid Meulenbelt¹

¹ Dept. of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands.

² Dept. Orthopaedics Leiden University Medical Center, Leiden, The Netherlands.

Abstract

Objective: To identify circulating micro RNAs (miRNAs) that could serve as biomarkers allowing for effective personalized treatment strategies.

Methods: Previously generated datasets of articular cartilage (mRNA-sequencing, N=56 patients) and plasma (miRNA-sequencing, N=56 patients) were integrated (N=20 patients of whom both cartilage mRNA-seq and plasma miRNA-seq were available). Generalized estimating equations and LASSO regression were applied to identify miRNAs and mRNAs marking previously identified OA endotype A and B. To identify potential druggable targets for OA molecular endotypes, we combined previously reported differentially expressed (DE) genes between preserved and lesioned OA cartilage exclusive for endotype A or B, recent GWAS data and the drug-gene interaction database.

Results: We identified miR-6804-5p, miR-182-3p, let-7e-3p, and miR-3179 expressed in plasma that together with sex and age were able to distinguish OA molecular endotype A and B. To validate predictive capacity of these four miRNAs for molecular endotype, we first identified mRNAs expressed in cartilage marking OA endotypes. Combining plasma miRNA-seq data and articular cartilage mRNA RT-qPCR data showed that prediction of OA endotypes coincided for 86% of additional patients. To match OA endotypes to druggable targets, we filtered exclusive DE genes for each endotype on OA risk genes. We identified MAP2K6 and HLA-DPA1 as druggable targets specific for endotype A and B, respectively.

Conclusion: We here showed that plasma expression levels of miR-6804-5p, miR-182-3p, let-7e-3p, and miR-3179 might be used to distinguish OA endotype A and B, which then could be used to treat patients in an OA endotype specific manner. Use of circulating miRNAs as biomarkers provides a window of opportunities for effective personalized OA treatment strategies.

Introduction

Osteoarthritis (OA) represents multiple subtypes of a degenerative joint disease, in which progressive and irreversible degeneration of the articular cartilage, structural changes in the subchondral bone, inflammation, and loss of joint space is seen [1, 2]. With total joint replacement surgery and pain relief treatment being the only treatment options for OA, there is an unmet desire for disease modifying treatments that target underlying pathophysiological processes [3]. Failure of disease modifying drug development, so far, is partly caused by the fact that it has followed a “one-drug-fits-all-patients” approach, in which OA heterogeneity is ignored [4, 5]. To address heterogeneity in OA pathophysiology, multiple studies have focused on the identification of OA molecular endotypes based on gene expression cluster analysis [6]. To this end, Soul et al. [7] identified two cluster analysis-based OA endotypes using RNA-seq data of knee articular cartilage. These two molecular endotypes were associated to changes in inflammasome activation and innate immune responses, and changes from chondrogenic to a more osteogenic phenotype, respectively. Recently, we also reported on the identification of two OA molecular endotypes in hip and knee articular cartilage in an independent dataset [8]. These endotypes represented similar pathways as reported by Soul et al. [7], indicating their consistency and robustness. Moreover, we showed that these patients showed clinical phenotypic differences. Endotype B patients, having an inflammatory driven disease process, showed increased joint space narrowing comparing to endotype A patients, having a hypertrophy driven disease process [8]. Together these data confirmed that OA may be amenable to tailored treatments targeting these unique molecular endotypes. Nonetheless, to enable molecular endotype-based stratification of patients before treatment in clinical practice, easily accessible and non-invasive biomarkers are required reflecting ongoing processes in articular cartilage. Hereto, Soul et al. [7] reported a set of proteins which were predicted to be secreted in the synovial fluid and could potentially serve as a biomarker for OA endotype. However, collecting synovial fluid is still an invasive procedure and therefore not optimal. For that matter, studies implicate circulating microRNAs (miRNAs) as novel promising biomarkers in numerous diseases, as they are stable in plasma and serum and relatively easily accessible [9-11]. More related to OA, Beyer et al. [12] found *let-7e* as a negative dose-dependent predictor for severe OA and Ntounou et al. [13] identified circulating miRNAs predicted to regulate metabolic processes and could serve as biomarker for OA. Likewise, Murata et al. [14] identified miR-132 predictive for rheumatoid arthritis and OA. Recently, we showed for the first time that circulating miRNAs were able to mark disease related mRNA expression patterns in articular cartilage with early OA [15]. To our knowledge, however, there are no studies yet identifying circulating miRNAs as biomarker for OA molecular endotypes.

In the current study, we used our previously described circulating miRNA-seq dataset of plasma [15] and mRNA-seq dataset of OA articular cartilage [16] (N=20 overlapping patients), to identify miRNAs that could serve as biomarkers for our previously reported OA molecular endotypes A and B [8]. Moreover, to identify OA endotype specific potential druggable targets, we combined our previously reported differentially expressed genes between macroscopically preserved and lesioned OA articular cartilage exclusive for either endotype A or endotype B with recent largest GWAS meta-analysis so far [17].

Methods

Sample description

The current study includes 68 patients, who underwent a joint replacement surgery due to OA, as part of the RAAK study (**Supplementary Table 1**). Macroscopically preserved OA cartilage was collected from all joints as described previously [18]. Plasma was collected from 56 patients. Informed consent was obtained from all participants and ethical approval for the RAAK study was given by the medical ethics committee of the Leiden University Medical Center (P08.239/P19.013).

miRNA and mRNA sequencing

miRNA and mRNA sequencing were performed on Illumina HiSeq 2500 and HiSeq 2000/4000, respectively, as described previously [15, 16]. More information on alignment, mapping and quality control is available in **supplementary methods**.

Principal component analysis

As described previously, we selected the 1000 genes with highest coefficient of variation (COV) [8]. Then, we performed principal component analysis (PCA) on the samples selected for these 1000 genes using the FactoMineR 1.42 [19] package in R. The threshold of 0.45 was set based on the average factor loading score of two samples with the smallest difference in PC1. More information is available in **supplementary methods**.

Prediction models

Spearman correlations between factor loading scores and expression levels of either miRNA or mRNA were calculated using the Hmisc 4.2-0 package in R. Both generalized estimating equations and elastic net regularization were performed. The model showing the highest number of correct predictions with the least number of variables was selected. Additional information is available in **supplementary methods**.

RT-qPCR

RT-qPCR was performed to quantitatively determine the mRNA expression levels. The

relative gene expression was evaluated by the $-\Delta CT$ values, using *LRRC41* and *U2AF2* as internal controls.

Results

Identification of plasma miRNAs predicting OA molecular endotype

To identify blood-based miRNAs able to distinguish between previously identified OA molecular endotype A (hypertrophy pathway) and B (immune response) [8], we integrated previously generated datasets of plasma (miRNA-sequencing, dataset 1, **Figure 1**) and articular cartilage (mRNA-sequencing, dataset 2, **Figure 1**), which were partially overlapping (N=20 patients, subset 1, **Figure 1**) [15, 16]. Since these 20 overlapping patients were unevenly distributed over OA endotype A (N=17) and B (N=3), we converted binominal “endotype A” and “endotype B” back to a quantitative contribution per patient to either endotype A or B by using principal component analysis (PCA) (**Supplementary Figure 1**), as reported previously [7]. Patients characteristics of the data- and the subsets are listed in **Supplementary Table 1**. To identify readily detectable miRNAs in plasma (dataset 1), we selected for highest quartile of expression levels (N=663 miRNAs out of 2652 miRNAs expressed in total). Subsequently, to find plasma miRNAs that mark OA molecular endotype, we correlated these 663 miRNAs to quantitative factor loading scores of OA endotypes of subset 1 patients (N= 20 patients with complete data, **Figure 1**). In total 21 significant correlating miRNAs were detected. Among the highest correlating miRNAs, we found miR-195-5p ($\rho=-0.61$), miR-182-3p ($\rho=0.60$) and miR-4665-5p ($\rho=0.60$) (**Supplementary Table 2**). Subsequently, to identify the minimal number of miRNAs with highest predictive value for OA endotype, we used miRNAs with a correlation of $|\rho|>0.5$ (N=9 miRNAs) and performed generalized estimating equations (GEE) with quantitative factor loading scores of OA endotypes as dependent variable and miRNAs as covariates while adjusting for age and sex. Upon selecting for significant variables, we found that miR-6804-5p, miR-182-3p, let-7e-3p,

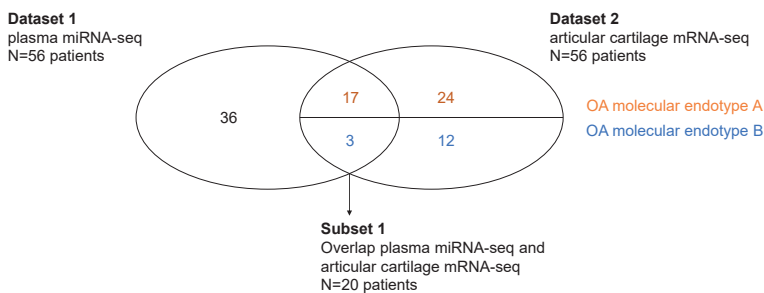


Figure 1 – Venn diagram of samples used in this study.

Dataset 1 consists of 56 patients of whom miRNA-seq of plasma was available and dataset 2 consists of 56 patients of whom mRNA-seq of cartilage was available. Subset 1 refers to the overlap between plasma miRNA-seq data and articular cartilage mRNA-seq data.

and miR-3179 together explained quantitative OA molecular endotypes of patients of subset 1 (**Equation 1**, **Supplementary Table 3**, **Supplementary Figure 2**).

Equation 1

Quantitative factor loading scores of OA molecular endotypes = $-0.45 * \text{miR_6804_5p} + 0.52 * \text{miR_182_3p} - 0.50 * \text{let_7e_3p} + 0.16 * \text{miR_3179} + 0.04 * \text{age} - 0.56 * \text{sex} - 1.95$

Confirmation of OA molecular endotype predicting plasma miRNAs

To allow validation of the capability of the four identified miRNAs in plasma to predict OA endotypes using **Equation 1** in an independent dataset, we had available subset 2 (**Figure 2**). Subset 2 consisted of 7 patients of whom plasma miRNA-seq and articular cartilage cDNA to measure limited number of genes by RT-qPCR, but no information on OA endotype data was available. Therefore, we first set out to identify mRNA markers in articular cartilage that together enable prediction of OA endotype of patients in subset 1 (**Figure 1**). These markers eventually allow delineation of OA endotypes based on cartilage mRNA expression levels in subset 2 (**Figure 2**).

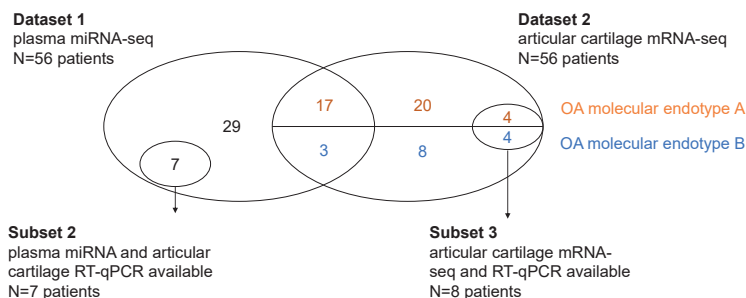


Figure 2 – Venn diagram of samples used in this study.

Dataset 1 consists of 56 patients of whom miRNA-seq of plasma was available and dataset 2 consists of 56 patients of whom mRNA-seq of cartilage was available. Subset 2 refers to the patients of whom we had plasma miRNA-seq data available and articular cartilage cDNA available to perform RT-qPCR. Subset 3 refers to the patients of whom we had articular cartilage mRNA-seq available and articular cartilage cDNA available to perform RT-qPCR.

To predict OA endotype of patients based on mRNA-seq data of articular cartilage in subset 1, we selected for previously reported FDR significantly differentially expressed genes between endotype A and B (N=2967 genes) [8]. To select for genes that are readily detectable, we further stratified for genes that were among the two highest expression level quartiles in articular cartilage (N=500 genes). Subsequently, we performed Spearman correlation between expression levels of these 500 genes and quantitative factor loading scores of OA endotypes (**Supplementary Table 4**). To identify the minimal number of genes with highest predictive value for OA endotype,

we selected genes highly correlating to OA endotypes ($|\rho|>0.9$, $N=13$ genes) and we performed LASSO regression. We found that expression levels of *MMP19*, *SLCO2B1* and *CDH11* in articular cartilage together could predict OA endotype (**Equation 2**), with 97.5% accuracy in the 36 non-overlapping samples (dataset 2, **Supplementary Figure 3B**). Together, these data confirm that mRNA expression levels of specified three genes in cartilage are highly predictive of OA molecular endotypes.

Equation 2

$$\text{Predicted factor loading score} = 0.31 * \text{MMP19} + 0.20 * \text{SLCO2B1} + 0.0047 * \text{CDH11} - 3.13$$

Being able to predict OA endotype based on mRNA levels in cartilage, allowed us to perform validation of predictive values of the four miRNAs in the independent subset 2 as articular cartilage RT-qPCR data and plasma miRNA-seq was available herein. However, prior to this analyses we needed to address conversion of RNA-seq to RT-qPCR derived mRNA expression data. Therefore we used subset 3 (**Figure 2**), consisting of 8 patients of whom RNA-seq data and articular cartilage cDNA was available. The RT-qPCR threshold was set based on the average difference between normalized read counts (VST) and $-\Delta\text{CT}$ values. Using this threshold, all 8 patients were assigned correctly to their OA endotype (**Figure 3**).

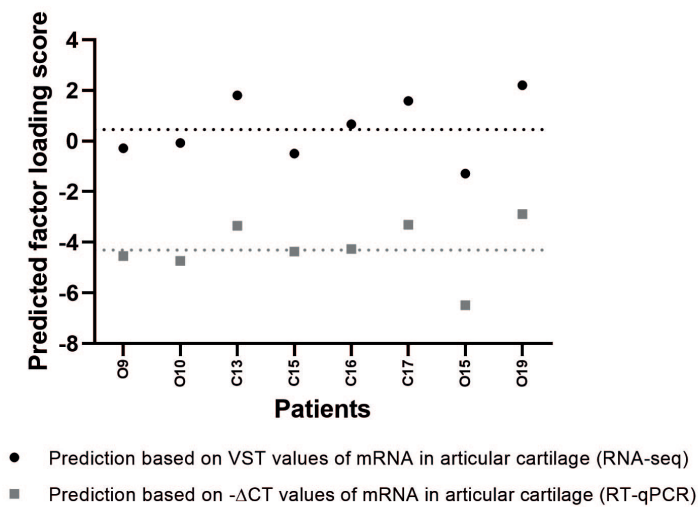


Figure 3 – Prediction of OA molecular endotype using VST expression levels of articular cartilage mRNA (RNA-sequencing data) and prediction of OA molecular endotype using $-\Delta\text{CT}$ values of articular cartilage mRNA (RT-qPCR) in 8 patients. Threshold of RNA-seq prediction: 0.45. Threshold of RT-qPCR prediction: -4.31. Patients indicated with O: patients of whom we have plasma miRNA-seq data and articular cartilage mRNA-seq data. Patients indicated with C: patients of whom we only have articular cartilage mRNA data.

Next, we used subset 2 (**Figure 2**), to assign patients to either OA endotype A or B based on plasma miRNA-seq data by applying **Equation 1** or based on articular cartilage RT-qPCR data by applying **Equation 2**. As shown in **Figure 4**, all patients were assigned to the same OA molecular endotype by both prediction methods, except for P6 (86% accuracy).

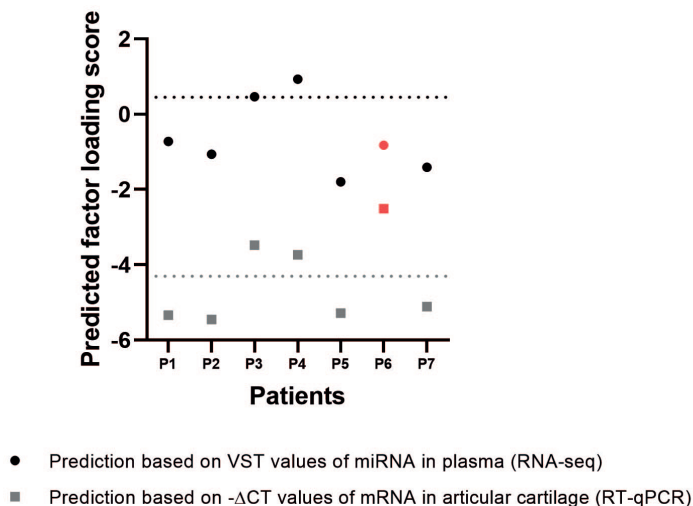


Figure 4 – Prediction based on VST expression levels of plasma miRNA (RNA sequencing data) and prediction based on $-\Delta\text{CT}$ values of articular cartilage mRNA (RT-qPCR) in 7 patients. Patients indicated with P: patients of whom we have plasma miRNA-seq data and articular cartilage RT-qPCR data.

Potential therapeutics for OA molecular endotype

Now that we have identified biomarkers that enable stratification of patients based on their OA molecular endotype, the next step was to identify potential druggable targets able to specifically treat endotype A or B patients. In this regard, we used our previously reported differentially expressed genes between macroscopically preserved and lesioned OA articular cartilage unique for endotype A (N=1114 genes) or B (N=72 genes) [8]. To select for genes that were most likely causal to OA pathophysiology and therefore could target underlying OA pathophysiological process, we filtered these differentially expressed genes for genes with SNPs that were recently identified in the largest genome-wide meta-analysis so far [17]. In doing so, we identified *POLD3*, *ERG*, *MAP2K6*, and *MN1* as differentially expressed OA risk genes with unique expression patterns in endotype A patients, making these four genes attractive potential druggable targets for patients with endotype A OA (**Supplementary Table 5A**). Similarly, we identified *HLA-DPA1* as attractive potential druggable target for patients with endotype B OA (**Supplementary Table 5B**). Subsequently, we used our previous studies on allelic

imbalanced expression (AIE) in articular cartilage and subchondral bone to identify the direction of effect of these five potential targets [20, 21]. We found SNPs located in *MAP2K6* and *HLA-DPA1* showing AIE, which were in high LD ($R^2 = 0.84$ and $R^2 = 0.71$, respectively) with the identified OA risk SNPs. Based on this AIE we could make a firm hypothesis that increased levels of *MAP2K6* and *HLA-DPA1* confer risk to OA, suggesting that inhibiting these genes could be a potential treatment strategy. To explore whether there are already approved drugs available to target these genes, we used the online drug gene interaction database (DGIdb 4.0)[22]. In total we found 11 drug-gene interactions for OA molecular endpoint A, including 3 drug-gene interactions of *MAP2K6*, while we did not find any drug-gene interactions for endpoint B (**Supplementary Table 5**).

Discussion

The aim of the current study was to identify non-invasive biomarkers able to classify patients according to previously identified robust OA molecular endpoints in articular cartilage. By combining miRNA-seq data of plasma and mRNA-seq data of articular cartilage of the same patients, we identified four miRNAs (miR-6804-5p, miR-182-3p, let-7e-3p, and miR-3179) expressed in plasma that were able to classify our previously reported OA molecular endpoints in articular cartilage. In an additional dataset of patients without previously assessed OA endpoint, we showed that prediction of OA molecular endpoints coincided using miRNA-seq data of plasma and mRNA RT-qPCR data of articular cartilage in 86% of patients. Moreover, we identified *MAP2K6* and *HLA-DPA1* as potential druggable targets for specific treatment endpoint A or B patients, respectively. Therefore, we advocate that expression levels of these four miRNAs in plasma could be used to stratify patients into OA molecular endpoints prior to treatment during clinical trials for more effective treatment response. Hypothetically, to limit failure of clinical trials, patients with inflammatory-driven OA (endpoint B) should be enriched in clinical trials using anti-inflammatory drugs or treatment with *MAP2K6* inhibitors, while patients with hypertrophy-driven OA (endpoint A) should be enriched in clinical trials using *HLA-DPA1* inhibitors. To our knowledge, we here showed for the first time that circulating miRNAs can be used as biomarker for OA molecular endpoints, as such providing a novel window of opportunities for effective personalized OA treatment strategies.

Mitogen-activated protein kinase kinase 6, encoded by *MAP2K6*, is an intracellular signaling protein which is activated by stress signals and inflammation [23]. Map kinases are known to regulate, amongst others, pain mediators and cartilage degrading enzymes such as matrix metalloproteinases and are therefore formerly suggested as therapeutic targets for OA [24]. Based on allelic imbalanced expression in articular cartilage of allele rs1133228-A [20], located in *MAP2K6* and in high LD with the identified OA risk SNP

rs2716212 [17], we hypothesize that increased expression of *MAP2K6* confers risk to OA. Therefore, inhibiting *MAP2K6* expression in patients with OA endotype A might be a potential therapeutic strategy. *HLA-DPA1*, encoding major histocompatibility complex class II DP alpha 1, plays a role in the immune system and is upregulated in rheumatoid arthritis [25]. Allelic imbalanced expression in subchondral bone of rs1126506-T [21], located in *HLA-DPA1* and in high LD with OA risk SNP rs2856821, we hypothesize that increased expression confers risk to OA. Based on this hypothesis inhibition of *HLA-DPA1* expression could be a therapeutic strategy specific for patients with OA endotype B. Functional studies investigating underlying biological mechanisms of both genes are necessary to confirm their potential as druggable target.

To find the minimal number of miRNAs and mRNAs with highest predictive value for OA endotypes, we performed both LASSO regression and generalized estimating equations. We selected the method that showed the lowest number of variables with highest predicting capacity. In doing so, four circulating miRNAs were identified of which the expression levels together were able to predict OA molecular endotype with 86% accuracy in replication. Nonetheless, given the relatively small sample size, replication in a larger dataset would be required to confirm. To our knowledge, these four miRNAs in plasma were not previously linked to OA pathophysiology, except for let-7e. Beyer et al. showed that expression levels of let-7e were significantly different between plasma of OA patients and healthy controls [12]. Moreover, it has been shown that let-7e could be used as a negative dose dependent predictor of OA [26]. In our dataset, let-7e-3p was shown to negatively correlate with quantitative factor loading scores of OA endotypes ($\rho=-0.55$), with let-7e-3p being higher expressed in OA endotype A representing chondrocyte hypertrophy. MiR-182-3p expression was previously shown to have a potential regulatory role in osteosarcoma [27] and plasma levels of miR-182 were previously associated to various other types of cancer [28, 29]. Plasma exosome levels of miR-3179 were previously shown to be associated with low bone mineral density in postmenopausal women [30]. To our knowledge, associations between miR-6804 and the musculoskeletal system have not yet been reported.

Although sample sizes of our discovery and validation datasets were relatively small (N=20 overlapping patients, subset 1), the large consistent and robust differences between the two OA molecular endotypes allowed for the detection of four predicting miRNAs. Notably, miR-195-5p and miR-4665-5p were among the highest correlating miRNAs to factor loading score ($\rho=-0.61$ and $\rho=0.60$, respectively), while these miRNAs were not included in the final predicting model.

Altogether, we here showed that miRNA expression levels in plasma could reflect

ongoing processes in articular cartilage, making them attractive, easily accessible, non-invasive biomarkers which could further advance the development of personalized medicine of OA and could lead to a higher clinical trial success rate.

Declarations

Acknowledgements

We thank all the participants of the RAAK study. The LUMC has and is supporting the RAAK study. We thank Enrike van der Linden, Demiën Broekhuis, Anika Rabelink-Hoogenstraaten, Peter van Schie, Shaho Hasan, Maartje Meijer, Daisy Latijnhouwers and Geert Spierenburg for collecting the RAAK material. We thank the Sequence Analysis Support Core (SASC) of the Leiden University Medical Center for their support. Data is generated within the scope of the Medical Delta programs Regenerative Medicine 4D: Generating complex tissues with stem cells and printing technology and Improving Mobility with Technology.

Funding

The study was funded by the Dutch Scientific Research council NWO /ZonMW VICI scheme (nr 91816631/528), Foundation for Research in Rheumatology (FOREUM), BBMRI-NL complementation project CP2013-83, Ana Fonds (O2015-27), and TreatOA (grant 200800 from the European Commission Seventh Framework Program).

Disclosures

The authors have declared no conflicts of interest.

References

1. Kean, W.F., R. Kean, and W.W. Buchanan, *Osteoarthritis: symptoms, signs and source of pain*. Inflammopharmacology, 2004. **12**(1): p. 3-31.
2. Loeser, R.F., et al., *Osteoarthritis: a disease of the joint as an organ*. Arthritis Rheum, 2012. **64**(6): p. 1697-707.
3. Ghouri, A. and P.G. Conaghan, *Update on novel pharmacological therapies for osteoarthritis*. Ther Adv Musculoskelet Dis, 2019. **11**: p. 1759720x19864492.
4. Grässel, S. and D. Muschter, *Recent advances in the treatment of osteoarthritis*. F1000Research, 2020. **9**: p. F1000 Faculty Rev-325.
5. Angelini, F., et al., *Osteoarthritis endotype discovery via clustering of biochemical marker data*. Annals of the Rheumatic Diseases, 2022. **81**(5): p. 666-675.
6. Yuan, C., et al., *Classification of four distinct osteoarthritis subtypes with a knee joint tissue transcriptome atlas*. Bone Research, 2020. **8**(1): p. 38.
7. Soul, J., et al., *Stratification of knee osteoarthritis: two major patient subgroups identified by genome-wide expression analysis of articular cartilage*. Ann Rheum Dis, 2018. **77**(3): p. 423.
8. Coutinho de Almeida, R., et al., *Identification and characterization of two consistent osteoarthritis subtypes by transcriptome and clinical data integration*. Rheumatology (Oxford), 2020.
9. Cruz, M.S., et al., *miRNAs emerge as circulating biomarkers of post-myocardial infarction heart failure*. Heart Failure Reviews, 2020. **25**(2): p. 321-329.
10. Vanhie, A., et al., *Plasma miRNAs as biomarkers for endometriosis*. Hum Reprod, 2019. **34**(9): p. 1650-1660.
11. Bouvy, C., et al., *Circulating MicroRNAs as Biomarkers in Diffuse Large B-cell Lymphoma: A Pilot Prospective Longitudinal Clinical Study*. Biomark Cancer, 2018. **10**: p. 1179299x18781095.
12. Beyer, C., et al., *Signature of circulating microRNAs in osteoarthritis*. Ann Rheum Dis, 2015. **74**(3): p. e18.
13. Ntoumou, E., et al., *Serum microRNA array analysis identifies miR-140-3p, miR-33b-3p and miR-671-3p as potential osteoarthritis biomarkers involved in metabolic processes*. Clin Epigenetics, 2017. **9**: p. 127.
14. Murata, K., et al., *Plasma and synovial fluid microRNAs as potential biomarkers of rheumatoid arthritis and osteoarthritis*. Arthritis Res Ther, 2010. **12**(3): p. R86.
15. Ramos, Y.F.M., et al., *Circulating MicroRNAs Highly Correlate to Expression of Cartilage Genes Potentially Reflecting OA Susceptibility-Towards Identification of Applicable Early OA Biomarkers*. Biomolecules, 2021. **11**(9): p. 1356.

16. Coutinho de Almeida, R., et al., *RNA sequencing data integration reveals an miRNA interactome of osteoarthritis cartilage*. Ann Rheum Dis, 2019. **78**(2): p. 270-277.
17. Boer, C.G., et al., *Deciphering osteoarthritis genetics across 826,690 individuals from 9 populations*. Cell, 2021.
18. Ramos, Y.F., et al., *Genes involved in the osteoarthritis process identified through genome wide expression analysis in articular cartilage; the RAAK study*. PLoS One, 2014. **9**(7): p. e103056.
19. Lê, S., J. Josse, and F. Husson, *FactoMineR: An R Package for Multivariate Analysis*. Journal of Statistical Software, 2008. **25**(1): p. 1 - 18.
20. den Hollander, W., et al., *Annotating Transcriptional Effects of Genetic Variants in Disease-Relevant Tissue: Transcriptome-Wide Allelic Imbalance in Osteoarthritic Cartilage*. Arthritis Rheumatol, 2019. **71**(4): p. 561-570.
21. Coutinho de Almeida, R., et al., *Allelic expression imbalance in articular cartilage and subchondral bone refined genome-wide association signals in osteoarthritis*. medRxiv, 2022: p. 2022.04.07.22273552.
22. Freshour, S.L., et al., *Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsourcing efforts*. Nucleic Acids Research, 2020. **49**(D1): p. D1144-D1151.
23. Jia, Z. and Q.J. Wei, *CircRNA-MSR Regulates LPS-Induced C28/I2 Chondrocyte Injury through miR-643/MAP2K6 Signaling Pathway*. Cartilage, 2021. **13**(2_suppl): p. 785s-795s.
24. Loeser, R.F., E.A. Erickson, and D.L. Long, *Mitogen-activated protein kinases as therapeutic targets in osteoarthritis*. Current opinion in rheumatology, 2008. **20**(5): p. 581-586.
25. Li, X., et al., *Promising targets and drugs in rheumatoid arthritis*. Bone & Joint Research, 2020. **9**(8): p. 501-514.
26. Feng, L., et al., *Circulating microRNA let-7e is decreased in knee osteoarthritis, accompanied by elevated apoptosis and reduced autophagy*. Int J Mol Med, 2020. **45**(5): p. 1464-1476.
27. Chen, G., et al., *Potential Regulatory Effects of miR-182-3p in Osteosarcoma via Targeting EBF2*. BioMed Research International, 2019. **2019**: p. 4897905.
28. Liu, X., et al., *Elevated circulating miR-182 acts as a diagnostic biomarker for early colorectal cancer*. Cancer Manag Res, 2018. **10**: p. 857-865.
29. Song, C., et al., *High expression of microRNA-183/182/96 cluster as a prognostic biomarker for breast cancer*. Scientific Reports, 2016. **6**(1): p. 24502.
30. Kong, D., et al., *Comparative profile of exosomal microRNAs in postmenopausal women with various bone mineral densities by small RNA sequencing*. Genomics, 2021. **113**(3): p. 1514-1521.

Supplementary data

Supplementary methods

miRNA sequencing

Small RNAs were isolated from 200 μ l plasma using the Qiagen miRNAeasy Serum/Plasma Kit (Qiagen, Germany). The TruSeq rapid SBS kit (Illumina, USA) was used to generate small RNA sequencing libraries and RNAs were separated on 4-20% SDS-PAGE. Sequencing was performed on the Illumina HiSeq2500. Alignment to GRCh37/hg19 reference genome was done using Bowtie [1]. HTseq count v0.11.1 [2] was used to estimate the read abundances per sample and were assigned to miRbase v21 [3]. In total, 2652 miRNAs were mapped. Since miRNAs generally show low expression levels and in the current study we were aiming to identify biomarkers, we selected the upper expression quartile for further analysis (N=663 miRNAs), to only include miRNAs that are readily measurable.

mRNA sequencing

Total RNA was isolated from preserved OA articular cartilage using Qiagen RNeasy Mini Kit (Qiagen, GmbH, Hilden, Germany). Paired-end 2 $\times$ 100 bp RNA-sequencing (Illumina TruSeq RNA Library Prep Kit, Illumina HiSeq2000 and Illumina HiSeq4000) was performed. Strand specific RNA-seq libraries were generated which yielded a mean of 20 million reads per sample. Data from both Illumina platforms were integrated and analyzed with the same in-house pipeline. RNA-seq reads were aligned using GSNAP [4] against GRCh37/hg19 using default parameters. Read abundances per sample was estimated using HTseq count v0.11.1 [2]. Only uniquely mapping reads were used for estimating expression. The quality of raw reads was checked using MultiQC v1.7 [5]. The adaptors were clipped using Cutadapt v1.1 [6] applying default settings (min overlap 3, min length).

Define previously assigned subtypes

Since the N=20 overlapping patients were unevenly distributed over subtype A (N=17 patients) and subtype B (N=3 patients) and a binominal prediction requires at least five observations per group, we converted binomial variables subtype A and subtype B to a continuous variable. First, we selected the 1000 genes with highest coefficient of variation (COV) as described previously [7]. Then, we performed principal component analysis (PCA) on the samples selected for these 1000 genes using the FactoMineR 1.42 [8] package in R. To validate the assignment of subtype A and B using the factor loading scores, we performed hierarchical clustering on the PCA map, using the FactoMineR package. Subsequently, the factor loading scores of the patients were used as continuous variable for further analysis. The threshold of 0.45 was set based on the average factor

loading score of two samples with the smallest difference in PC1.

Generalized Estimating Equations

Generalized Estimating equations (GEE) was performed using IBM SPSS statistics v25. A linear model was applied and an independent structure was used. The factor loading score was set as the dependent variable and mRNAs, miRNAs, sex and/or age were set as covariates. For prediction of factor loading scores using miRNAs in plasma, only miRNAs showing correlation of $|\rho| > 0.5$ with factor loading score were included as covariate. For prediction of factor loading score using mRNAs in articular cartilage, only mRNAs showing correlation of $|\rho| > 0.9$ with factor loading score were included as covariate.

Elastic net regularization

Elastic net regularization was performed using the glmnet 4.1 [9] package in R, with 80% of the dataset as training dataset and 20% of the dataset as test dataset. The lambda resulting in the minimum mean cross-validated error was selected (miRNAs in plasma: lambda = 0.1133088, mRNAs in articular cartilage: lambda = 0.2766053). The alpha was selected based on the mean squared error (as low as possible), the minimum number of variables included in the model, and the highest correlation between predicted factor loading score and actual factor loading score (alpha = 1, for both miRNAs in plasma and mRNAs in articular cartilage).

RT-qPCR

RNA was isolated from the cartilage as described above. cDNA synthesis was done using Transcriptor First Strand cDNA Synthesis Kit (Roche, Switzerland), using 100 ng of RNA. RT-qPCR was performed to quantitatively determine the mRNA expression levels. The relative gene expression was evaluated by the $-\Delta\Delta C_T$ values, using LRRC41 and U2AF2 as internal controls. The housekeeping genes were identified by selecting for low coefficient of variance across all patients and by selecting for minimal difference in expression level between subtype A and subtype B (LRRC41: FC=0.983, U2AF2: FC=1.019).

References supplementary methods

1. Langmead, B., et al., Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 2009. 10(3): p. R25.
2. Anders, S., P.T. Pyl, and W. Huber, HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 2015. 31(2): p. 166-9.
3. Kozomara, A. and S. Griffiths-Jones, miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*, 2014. 42(Database issue): p. D68-73.
4. Wu, T.D. and C.K. Watanabe, GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*, 2005. 21(9): p. 1859-75.
5. Ewels, P., et al., MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 2016. 32(19): p. 3047-8.
6. Martin, M., Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011, 2011. 17(1): p. 3 %J EMBnet.journal.
7. Coutinho de Almeida, R., et al., Identification and characterization of two consistent osteoarthritis subtypes by

transcriptome and clinical data integration. Rheumatology (Oxford), 2020.

8. Lê, S., J. Josse, and F. Husson, FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software, 2008. 25(1): p. 1 - 18.

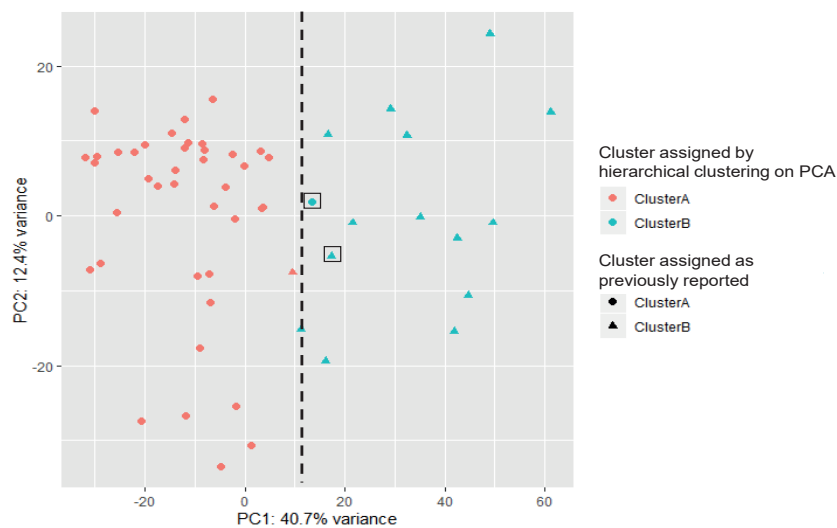
9. Friedman, J.H., T. Hastie, and R. Tibshirani, Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 2010. 33(1): p. 1 - 22.

10. Coutinho de Almeida, R., et al., Allelic expression imbalance in articular cartilage and subchondral bone refined genome-wide association signals in osteoarthritis. 2022: p. 2022.04.07.22273552.

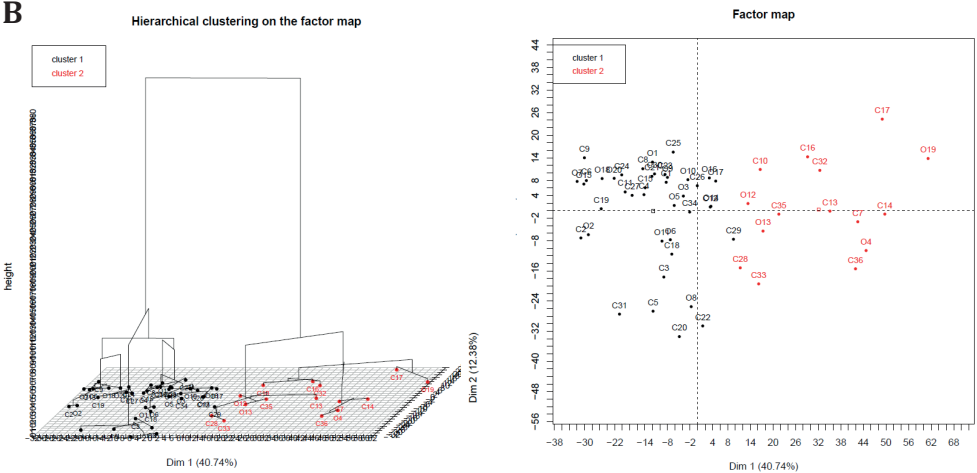
11. den Hollander, W., et al., Annotating Transcriptional Effects of Genetic Variants in Disease-Relevant Tissue: Transcriptome-Wide Allelic Imbalance in Osteoarthritic Cartilage. Arthritis Rheumatol, 2019. 71(4): p. 561-570.

Supplementary figures

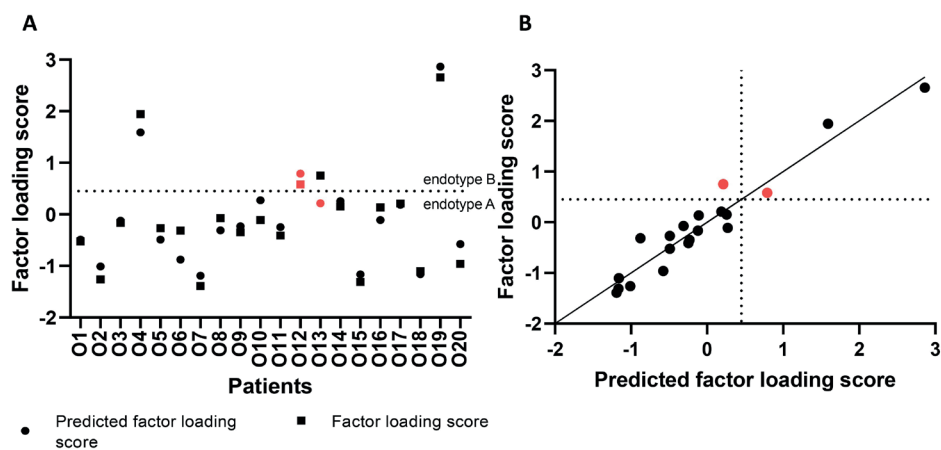
A



B

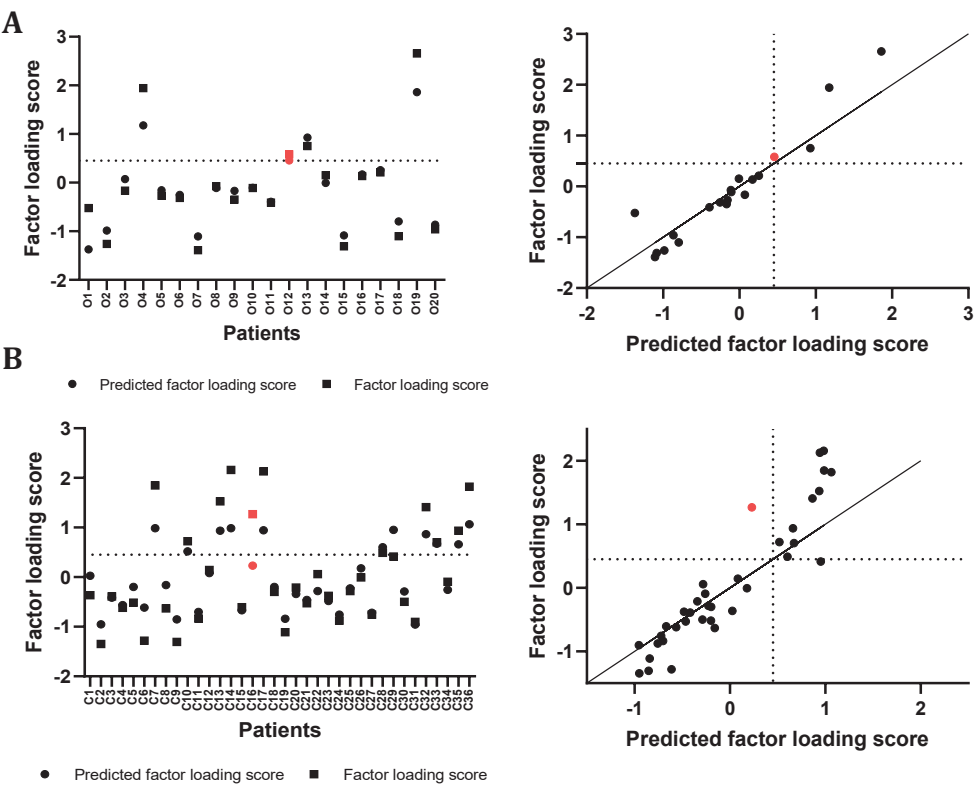


Supplementary Figure 1- Visualisation of cluster analysis.
(A) PCA on the samples using the 1000 genes showing the highest COV showing the separation of the clusters mainly described in PC1. The shapes represent OA subtypes as reported previously, while the colors represent OA subtype based on hierarchical clustering on the PCA. Patient O12 and C29 were not assigned similarly by the two clustering methods. The dotted line represents the threshold of separation, corresponding to a factor loading score of 0.45. (B) Hierarchical clustering on the PCA factor map to confirm whether patients are correctly assigned to an OA subtype.



Supplementary Figure 2 – Factor loading scores and predicted factor loading scores.

(A) Factor loading score vs. Predicted factor loading score. (B) Correlation between factor loading score and predicted factor loading score ($\rho = 0.92$) (right). In red the patients are shown that are incorrectly predicted. The dotted line represents the factor loading score threshold of 0.45. Patients indicated with O: patients of whom we have miRNA in plasma data and mRNA in articular cartilage data.



Supplementary Figure 2 – Factor loading score vs. Predicted factor loading score (left) and correlation between factor loading score and predicted factor loading score ($\rho = 0.96$ and $\rho = 0.93$, respectively) (right).

(A) Predictions are based on mRNA in articular cartilage and shown for the 20 overlapping patients and (B) 36 non-overlapping patients. In red patients are shown that are incorrectly predicted. The dotted line represents the factor loading score threshold of 0.45, i.e. patients with factor loading score above 0.45 are assigned to subtype B, while patients with factor loading scores below 0.45 are assigned to subtype A. Patients indicated with O: patients of whom we have miRNA in plasma data and mRNA in articular cartilage data. Patients indicated with C: patients of whom we only have mRNA in articular cartilage data.

Supplementary tables

Supplementary Table 1 – Information on samples used in this study
Supplementary Table 1A - Distribution of patients over datasets

	dataset 1	dataset 2
Goal	1. Define quantitative mRNA regression scores marking contribution of patient to OA subtype 2. Define mRNAs in articular cartilage that mark OA subtype	Define miRNAs in plasma that mark OA subtype
Number of patients	N=56	N=56
Cartilage mRNA data	N=56	N=20
Plasma miRNA data	N=20	N=56
OA subtype	N=56	N=20
Subtype A	N=41	N=17
Subtype B	N=15	N=3

Supplementary Table 1B - Sample characteristics

	Cartilage samples in cluster analysis (N=56)	Overlap between cartilage and miRNA samples (N=20)	Additional dataset for replication (N=7)
mean age (stdev)	68.0 (8.4)	72.8 (6.1)	67.9 (9.8)
female (male)	45 (11)	16 (4)	6 (6)
knees (hips)	35 (21)	15 (5)	8 (4)

Chapter 4

Supplementary Table 2 (partially) - Spearman correlations between top 25% highest expressed miRNA in plasma (N=663 miRNAs) and factor loading score.
The top 50 highest correlations between miRNAs and factor loading scores are shown here.

miRNA	ρ	$ \rho $	Pval	Padj
hsa-miR-195-5p	-0.61	0.61	4.13E-03	9.97E-01
hsa-miR-182-3p	0.60	0.60	4.89E-03	9.97E-01
hsa-miR-4665-5p	0.60	0.60	5.60E-03	9.97E-01
hsa-miR-3179	-0.58	0.58	7.90E-03	9.97E-01
hsa-miR-1307-3p	0.57	0.57	8.28E-03	9.97E-01
hsa-let-7e-3p	-0.55	0.55	1.15E-02	9.97E-01
hsa-miR-193b-3p	-0.55	0.55	1.26E-02	9.97E-01
hsa-miR-195-3p	-0.54	0.54	1.43E-02	9.97E-01
hsa-miR-6804-5p	-0.53	0.53	1.67E-02	9.97E-01
hsa-miR-500b-3p	-0.49	0.49	2.76E-02	9.97E-01
hsa-miR-378e	0.49	0.49	2.78E-02	9.97E-01
hsa-miR-1270	-0.49	0.49	2.81E-02	9.97E-01
hsa-miR-1343-3p	0.48	0.48	3.05E-02	9.97E-01
hsa-miR-11401	0.48	0.48	3.11E-02	9.97E-01
hsa-miR-4504	-0.48	0.48	3.20E-02	9.97E-01
hsa-miR-296-3p	0.47	0.47	3.68E-02	9.97E-01
hsa-miR-6810-5p	0.46	0.46	4.26E-02	9.97E-01
hsa-miR-301b-3p	-0.45	0.45	4.49E-02	9.97E-01
hsa-miR-132-5p	-0.45	0.45	4.51E-02	9.97E-01
hsa-miR-671-5p	0.45	0.45	4.66E-02	9.97E-01
hsa-miR-6767-5p	-0.45	0.45	4.74E-02	9.97E-01
hsa-miR-4651	0.44	0.44	5.04E-02	9.97E-01
hsa-miR-3909	-0.43	0.43	5.61E-02	9.97E-01
hsa-miR-4510	-0.42	0.42	6.28E-02	9.97E-01
hsa-miR-5698	-0.42	0.42	6.29E-02	9.97E-01
hsa-miR-4661-5p	0.42	0.42	6.39E-02	9.97E-01
hsa-miR-1227-3p	-0.42	0.42	6.62E-02	9.97E-01
hsa-miR-4646-5p	0.42	0.42	6.77E-02	9.97E-01
hsa-miR-4755-5p	-0.41	0.41	6.95E-02	9.97E-01
hsa-miR-548j-3p	-0.41	0.41	7.20E-02	9.97E-01
hsa-miR-1-3p	-0.41	0.41	7.33E-02	9.97E-01
hsa-miR-4473	-0.40	0.40	8.01E-02	9.97E-01
hsa-miR-1255b-5p	-0.40	0.40	8.06E-02	9.97E-01

Circulating miRNAs to predict OA molecular endotype

miRNA	ρ	$ \rho $	Pval	Padj
hsa-miR-6813-5p	0.40	0.40	8.44E-02	9.97E-01
hsa-miR-4488	0.40	0.40	8.44E-02	9.97E-01
hsa-miR-450b-5p	-0.39	0.39	8.56E-02	9.97E-01
hsa-miR-574-5p	-0.39	0.39	8.78E-02	9.97E-01
hsa-miR-190b-5p	0.39	0.39	8.83E-02	9.97E-01
hsa-miR-339-5p	0.39	0.39	8.83E-02	9.97E-01
hsa-miR-1296-5p	-0.39	0.39	8.88E-02	9.97E-01
hsa-miR-320c	0.39	0.39	9.10E-02	9.97E-01
hsa-miR-30c-1-3p	0.39	0.39	9.23E-02	9.97E-01
hsa-miR-500a-3p	-0.38	0.38	9.37E-02	9.97E-01
hsa-miR-410-3p	0.38	0.38	9.86E-02	9.97E-01
hsa-miR-532-3p	0.38	0.38	9.94E-02	9.97E-01
hsa-miR-129-5p	-0.38	0.38	1.01E-01	9.97E-01
hsa-miR-6726-3p	-0.38	0.38	1.02E-01	9.97E-01
hsa-miR-6812-3p	-0.37	0.37	1.08E-01	9.97E-01
hsa-miR-411-5p	0.37	0.37	1.08E-01	9.97E-01

Supplementary Table 3 - GEE, with factor loading score as dependent variable and age, sex and correlating miRNAs as covariates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-1.95	0.74	-3.39	-0.50	6.95	1	8.37E-03
Age	0.04	0.01	0.02	0.06	16.20	1	5.70E-05
Sex	-0.56	0.21	-0.97	-0.15	7.20	1	7.31E-03
hsa_miR_6804_5p	-0.45	0.06	-0.55	-0.34	65.14	1	6.99E-16
hsa_miR_182_3p	0.52	0.05	0.42	0.62	105.91	1	0.00E+00
hsa_let_7e_3p	-0.50	0.08	-0.65	-0.34	41.59	1	1.12E-10
hsa_miR_3179	0.16	0.04	0.08	0.24	16.03	1	6.25E-05

Chapter 4

Supplementary Table 4 (partially) - Correlations between well-expressed (highest two expression quartiles) FDR significantly differentially expressed mRNAs between cluster A and B in articular cartilage and the factor loading score.

The top 50 highest correlations between mRNAs and factor loading scores are shown here.

Ensembl ID	Gene name	ρ	$ \rho $	Pval	Padj
ENSG00000101347	SAMHD1	0.94	0.94	5.32E-10	2.66E-07
ENSG00000140937	CDH11	0.92	0.92	9.20E-09	1.95E-06
ENSG00000123342	MMP19	0.91	0.91	1.80E-08	1.95E-06
ENSG00000172061	LRRCL15	0.91	0.91	2.01E-08	1.95E-06
ENSG00000177575	CD163	0.91	0.91	2.01E-08	1.95E-06
ENSG00000144810	COL8A1	0.91	0.91	2.34E-08	1.95E-06
ENSG00000196735	HLA-DQA1	0.91	0.91	2.81E-08	1.95E-06
ENSG00000150687	PRSS23	0.91	0.91	3.12E-08	1.95E-06
ENSG00000159189	C1QC	0.91	0.91	3.59E-08	1.99E-06
ENSG00000154096	THY1	0.90	0.90	4.72E-08	2.15E-06
ENSG00000160255	ITGB2	0.90	0.90	4.72E-08	2.15E-06
ENSG00000137491	SLCO2B1	0.90	0.90	5.40E-08	2.25E-06
ENSG00000158710	TAGLN2	0.90	0.90	6.16E-08	2.37E-06
ENSG00000187653	TMSB4XP8	0.90	0.90	8.25E-08	2.95E-06
ENSG00000122861	PLAU	0.90	0.90	9.35E-08	3.12E-06
ENSG00000162745	OLFML2B	0.89	0.89	1.16E-07	3.61E-06
ENSG00000162511	LAPTM5	0.89	0.89	1.47E-07	4.31E-06
ENSG00000155659	VSIG4	0.88	0.88	2.59E-07	7.19E-06
ENSG00000196126	HLA-DRB1	0.88	0.88	3.57E-07	9.40E-06
ENSG00000128294	TPST2	0.88	0.88	3.97E-07	9.92E-06
ENSG00000100292	HMOX1	0.88	0.88	4.40E-07	1.03E-05
ENSG00000107438	PDLIM1	0.87	0.87	4.53E-07	1.03E-05
ENSG00000011600	TYROBP	0.87	0.87	4.87E-07	1.06E-05
ENSG00000159713	TPPP3	0.87	0.87	5.39E-07	1.12E-05
ENSG00000204287	HLA-DRA	0.87	0.87	5.95E-07	1.19E-05
ENSG00000075223	SEMA3C	0.87	0.87	7.24E-07	1.39E-05
ENSG00000143320	CRABP2	0.87	0.87	8.19E-07	1.52E-05
ENSG00000141480	ARRB2	0.86	0.86	1.06E-06	1.75E-05
ENSG00000137507	LRRCL32	0.86	0.86	1.08E-06	1.75E-05
ENSG00000019582	CD74	0.86	0.86	1.16E-06	1.75E-05
ENSG00000121281	ADCY7	0.86	0.86	1.16E-06	1.75E-05
ENSG00000166825	ANPEP	0.86	0.86	1.16E-06	1.75E-05

Ensembl ID	Gene name	ρ	$ \rho $	Pval	Padj
ENSG00000186340	THBS2	0.86	0.86	1.16E-06	1.75E-05
ENSG00000203747	FCGR3A	0.86	0.86	1.42E-06	2.09E-05
ENSG00000108821	COL1A1	0.85	0.85	1.65E-06	2.34E-05
ENSG00000133110	POSTN	0.85	0.85	1.80E-06	2.34E-05
ENSG00000136167	LCP1	0.85	0.85	1.80E-06	2.34E-05
ENSG00000168398	BDKRB2	0.85	0.85	1.80E-06	2.34E-05
ENSG00000074410	CA12	0.85	0.85	1.82E-06	2.34E-05
ENSG00000116741	RGS2	0.85	0.85	1.96E-06	2.39E-05
ENSG00000173369	C1QB	0.85	0.85	1.96E-06	2.39E-05
ENSG00000157613	CREB3L1	0.85	0.85	2.13E-06	2.48E-05
ENSG00000183160	TMEM119	0.85	0.85	2.13E-06	2.48E-05
ENSG00000261371	PECAM1	0.85	0.85	2.52E-06	2.87E-05
ENSG00000167460	TPM4	0.85	0.85	2.74E-06	3.04E-05
ENSG00000136235	GNPMB	0.84	0.84	3.48E-06	3.77E-05
ENSG00000205403	CFI	0.84	0.84	3.56E-06	3.77E-05
ENSG00000099953	MMP11	0.84	0.84	3.77E-06	3.77E-05
ENSG00000129038	LOXL1	0.84	0.84	3.77E-06	3.77E-05

Supplementary Table 5 - OA risk genes differentially expressed between preserved and lesioned OA articular cartilage in either molecular endotype A or endotype B.
Supplementary Table 5A - OA risk genes differentially expressed between preserved and lesioned OA tissue of patients with OA molecular endotype A

Gene	Fold change ^[7]	P-value ^[7]	Padj ^[7]	Previously reported allelic imbalance	OA susceptibility based on AI	Potential therapeutics
POLD3	0.85	6.28E-03	4.18E-02	-	-	-
ERG	0.82	1.27E-03	1.25E-02	-	-	Sotalol hydrochloride (inhibitor), Amiodarone hydrochloride (inhibitor), Dofetilide (inhibitor), Idarubicin hydrochloride, Medroxyprogesterone acetate, Daunorubicin hydrochloride, Guanidine hydrochloride (inhibitor), Dalfampridine (inhibitor)
MAP2K6	0.70	6.51E-03	4.28E-02	cartilage [20]	Increased expression	Cobimetinib (inhibitor), Binimetinib (inhibitor), Trametinib (inhibitor)
MN1	0.69	2.28E-03	1.96E-02	-	-	-

Supplementary Table 5B - OA risk genes differentially expressed between preserved and lesioned OA tissue of patients with OA molecular endotype B

Gene	Fold change ^[7]	P-value ^[7]	Padj ^[7]	Previously reported allelic imbalance	OA susceptibility based on AI	Potential therapeutics
HLA-DPA1	0.33	1.05E-06	4.55E-04	subchondral bone [21]	Increased expression	-