7-29-2020

Sequencing identifies a distinct signature of circulating microRNAs in early radiographic knee osteoarthritis

Shabana A. Ali

R Gandhi

P Potla

S Keshavarzi

O Espin-Garcia

See next page for additional authors

Authors

Shabana A. Ali, R Gandhi, P Potla, S Keshavarzi, O Espin-Garcia, K Shestopaloff, C Pastrello, D Bethune-Waddell, S Lively, A V. Perruccio, Y R. Rampersaud, C Veillette, J S. Rockel, I Jurisica, T Appleton, and M Kapoor

Osteoarthritis and Cartilage xxx (xxxx) xxx

Osteoarthritis and Cartilage

OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

Sequencing identifies a distinct signature of circulating microRNAs in early radiographic knee osteoarthritis

S.A. Ali † ‡, R. Gandhi † §, P. Potla †, S. Keshavarzi †, O. Espin-Garcia †, K. Shestopaloff †, C. Pastrello †, D. Bethune-Waddell †, S. Lively †, A.V. Perruccio † § ||, Y.R. Rampersaud † §, C. Veillette † §, J.S. Rockel †, I. Jurisica †, T. Appleton ¶, M. Kapoor † § # *

† *Arthritis Program, Krembil Research Institute, University Health Network, Toronto, ON, Canada*
‡ *Bone & Joint Center, Department of Orthopaedic Surgery, Henry Ford Health System, Detroit, MI, USA*
§ *Department of Surgery, Faculty of Medicine, University of Toronto, ON, Canada*
|| *Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Canada*
¶ *Department of Medicine and Department of Physiology and Pharmacology, Western University, London, ON, Canada*
# *Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada*

ARTICLE INFO

Article history:
Received 9 April 2020
Accepted 20 July 2020

Keywords:
MiRNome
Osteoarthritis
Knee joint
Circulating MicroRNA
High-throughput nucleotide sequencing
Computational biology

SUMMARY

Objective: MicroRNAs act locally and systemically to impact osteoarthritis (OA) pathophysiology, but comprehensive profiling of the circulating miRNome in early vs late stages of OA has yet to be conducted. Sequencing has emerged as the preferred method for microRNA profiling since it offers high sensitivity and specificity. Our objective is to sequence the miRNome in plasma from 91 patients with early [Kellgren–Lawrence (KL) grade 0 or 1 ($n = 41$)] or late [KL grade 3 or 4 ($n = 50$)] symptomatic radiographic knee OA to identify unique microRNA signatures in each disease state.
Design: MicroRNA libraries were prepared using the QIAseq miRNA Library Kit and sequenced on the Illumina NextSeq 550.Counts were produced for microRNAs captured in miRBase and for novel microRNAs. Statistical, bioinformatics, and computational biology approaches were used to refine and interpret the final list of microRNAs.
Results: From 215 differentially expressed microRNAs (FDR < 0.01), 97 microRNAs showed an increase or decrease in expression in ≥85% of samples in the early OA group as compared to the median expression in the late OA group. Increasing this threshold to ≥95%, seven microRNAs were identified: hsa-miR-335-3p, hsa-miR-199a-5p, hsa-miR-671-3p, hsa-miR-1260b, hsa-miR-191-3p, hsa-miR-335-5p, and hsa-miR-543. Four novel microRNAs were present in ≥50% of early OA samples and had 27 predicted gene targets in common with the prioritized set of predicted gene targets from the 97 microRNAs, suggesting common underlying mechanisms.

Abbreviations: OA, Osteoarthritis; KL, Kellgren–Lawrence; BMI, body mass index.

* Address correspondence and reprint requests to: M. Kapoor, Arthritis Program, Krembil Research Institute, University Health Network, 60 Leonard Avenue, Toronto, Ontario, M5T 2S8, Canada. Tel.: 1-416-603-5800x4796.

E-mail addresses: s.amanda.ali@gmail.com (S.A. Ali), Rajiv.Gandhi@uhn.ca (R. Gandhi), Pratibha.Potla@uhnresearch.ca (P. Potla), biostat.keshavarzi@gmail.com (S. Keshavarzi), Osvaldo.Espin-Garcia@uhnresearch.ca (O. Espin-Garcia), konstantin.shestopaloff@utoronto.ca (K. Shestopaloff), chiara.pastre@gmail.com (C. Pastrello), dylan.bethune.waddell@gmail.com (D. Bethune-Waddell), Starlee.Lively@uhnresearch.ca (S. Lively), Anthony.Perruccio@uhnresearch.ca (A.V. Perruccio), raja.rampersaud@uhn.ca (Y.R. Rampersaud), christian.veillette@uhn.ca (C. Veillette), jason.rockel@utoronto.ca (J.S. Rockel), juris@ai.utoronto.ca (I. Jurisica), Tom.Appleton@sjhc.london.on.ca (T. Appleton), mkapoor@uhnresearch.ca (M. Kapoor).

https://doi.org/10.1016/j.joca.2020.07.003

Conclusion: Applying sequencing to well-characterized patient cohorts produced unbiased profiling of the circulating miRNome and identified a unique panel of 11 microRNAs in early radiographic knee OA.

Introduction

The complex pathophysiology of osteoarthritis (OA) results from an interplay between systemic and local joint factors[1]. There are few large-scale studies with clearly defined patient cohorts profiling molecular differences in circulating factors in OA, particularly in early stages of disease. Next generation sequencing has emerged as the gold standard approach for profiling molecular targets in biospecimens of interest in both health and disease. Sequencing offers the sensitivity and specificity to detect novel transcripts and low abundance isoforms that may be unique to particular disease stages[2]. Decreases in the cost of sequencing have made this approach more feasible for analyzing large datasets, providing the statistical power needed to consider multiple patient factors alongside molecular factors. As the technology advances, applications are expanding, and now include sequencing of small RNAs such as microRNAs.

MicroRNAs are small non-coding ribonucleic acid (RNA) molecules that inhibit expression of their gene targets[3]. Potentially important biomarkers for disease, peripheral blood microRNAs are abundant, relatively stable, and detectable through minimally invasive means[4,5]. MicroRNAs play key roles in OA pathophysiology[6], having been identified in several human joint tissues and fluids including knee and hip cartilage[7], synovium[8], spine cartilage[9], synovial fluid[10] and plasma[11]. Previous studies have used real-time PCR[11], real-time PCR array[12], microarray[13–15], and next generation sequencing[16,17] to identify circulating microRNAs in OA vs control, but the only two studies to use sequencing did not report significant differences[16,17]. To the best of our knowledge, sequencing has never been applied to identify circulating microRNA signatures between early and late symptomatic radiographic knee OA.

Here we use sequencing to profile circulating microRNAs in blood plasma from 91 patients with symptomatic radiographic knee OA[18,19]. The early radiographic knee OA cohort ($n = 41$) included patients with Kellgren–Lawrence (KL) grades 0 or 1, and the late radiographic knee OA cohort ($n = 50$) included patients with KL grades 3 or 4. Patients with KL grade 2 were excluded in order to clearly differentiate the early OA group from the late OA group. We identified a distinct signature of plasma microRNAs in early OA, consisting of seven microRNAs and four novel microRNAs. With an integrated computational biology approach, we identified putative downstream gene targets and pathways modulated by this microRNA signature in early OA.

Methods

Early and late symptomatic radiographic knee OA cohorts

In this study, we utilized blood plasma samples from three knee OA cohorts (Fig. 1). Our first cohort, The Arthritis Program Early Knee OA cohort, at the University Health Network (UHN; Toronto, Canada), comprised plasma samples ($n = 22$) obtained from patients with KL grade 1 and persistent pain for at least 4 weeks. These patients were diagnosed by orthopaedic surgeon R.G. Our second cohort, the Western Ontario Registry for Early Osteoarthritis (WOREO) Knee Study, at the University of Western Ontario (London, Canada), included plasma samples ($n = 19$) from patients referred to the St. Joseph's Health Care London Rheumatology Centre with symptomatic early OA (KL grade 0 or 1) who answered 'yes' to having frequent knee symptoms on most days for at least 4 weeks. These patients were diagnosed by rheumatologist T.A. Our third cohortincluded plasma samples ($n = 50$) from our Longitudinal Evaluation in the Arthritis Program (LEAP) cohort (UHN; Toronto, Canada) comprised of patients exhibiting KL grades 3 and 4. All patients were symptomatic and fulfilled the American College of Rheumatology clinical criteria for knee OA classification[18]. Blood samples were collected in K2-EDTA tubes, centrifuged at 4,000 rpm for 10 min at 4 °C, aliquotted into 250 μL per cryovial, flash frozen with liquid nitrogen, and stored in a cryo freezer until use.

Sequencing of microRNAs

RNA was extracted from 200 μL of human plasma ($n = 99$) using the QIAcube semi-automated system following manufacturer's instructions for the miRNeasy Serum/Plasma Advanced Kit (Qiagen, Hilden, Germany). Low concentrations prevent quality and quantity checks of RNA eluate (18 μL in nuclease-free water), so volumetric measures are used. cDNA libraries ($n = 50$ late OA; $n = 49$ early OA) were prepared from 5 μL RNA using the QIAseq miRNA Library Kit according to manufacturer's recommendations for biofluid samples. A detailed protocol is provided in Supplementary File 1. A key advantage of this method is that unique molecular indexes (UMIs) are incorporated during cDNA synthesis, uniquely tagging each strand prior to amplification, which more accurately reflects endogenous microRNA levels by controlling for library amplification bias.

Individual libraries were quantified using a fluorometric high sensitivity dsDNA assay (DeNovix, Wilmington, DE, USA). Library quality was assessed on a high sensitivity DNA chip on the Bioanalyzer (Agilent, Santa Clara, CA, USA). On the day of sequencing, library batches were pooled, re-quantified (DeNovix), denatured and diluted to 1.5 pM, and spiked with 1% PhiX Sequencing Control V3 library, which acts as a quality metric for cluster generation, sequencing, and alignment. Spiked library pools were sequenced on Illumina's NextSeq 550 system using a high-output kit v2 following a 76-base single-end read protocol (Illumina, San Diego, CA, USA) at the Arthritis Centre for Diagnostic and Therapeutic Innovation (Krembil Research Institute, Toronto, ON, Canada). Data were deposited in the Gene Expression Omnibus database under accession number GSE151341.

Sequencing data alignment and counts generation

An overview of this method is provided in Supplementary Figure 3. Demultiplexing of.bcl files and conversion into Fastq files was performed, followed by flagging quality parameters such as insufficient number of reads (<400,000 reads), low sequence quality scores (<Q20), low percentage duplication (20%), and poor 3' adapter contamination (<40%) in reads. Random UMIs of 12 bp length were extracted for later deduplication of reads. Regular expression pattern was used in UMI-tools software to select reads

Please cite this article as: Ali SA et al., Sequencing identifies a distinct signature of circulating microRNAs in early radiographic knee osteoarthritis, Osteoarthritis and Cartilage, https://doi.org/10.1016/j.joca.2020.07.003

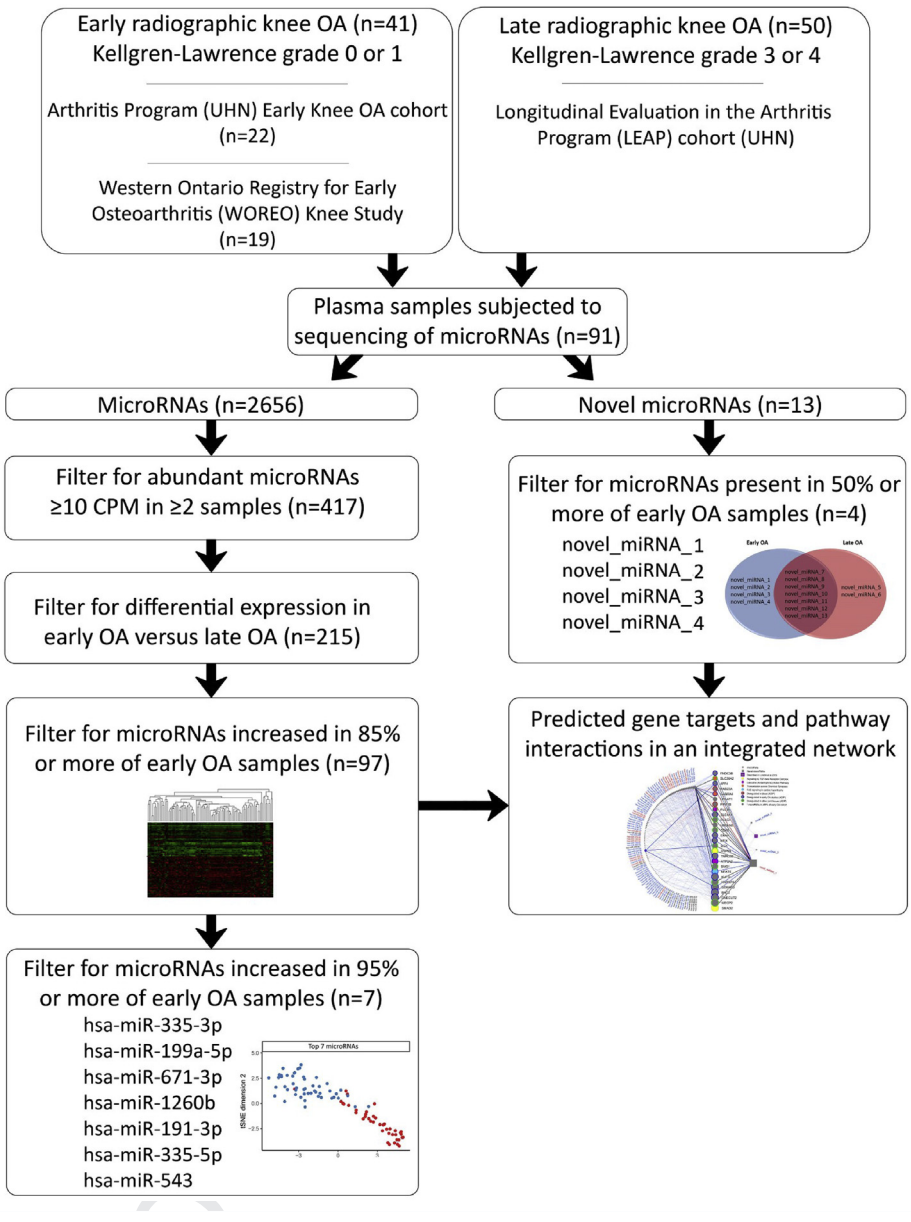


Fig. 1 **Overview of experimental workflow**. Cohorts of early OA (***n*** = 41) and late OA (***n*** = 50) were subjected to sequencing of microRNAs isolated from blood plasma. Sequential filtration of microRNAs resulted in the final lists used for computational analyses to predict gene targets and pathways.

containing less than two mismatches in the 3' adapter sequence mentioned, followed by a 12 bp random UMI tag sequence, and the remaining part of the read was discarded. We discarded reads of less than 18bp as too short and reads greater than 30 bp as too long given that microRNAs are typically 22–25 bp in length.

Alignment of reads was carried out against mature microRNA sequences from miRBase v22.1 with the following parameters: no mismatches in a seed read length of 30 bp, no alignment to forward/reverse-complement reference strand, and reporting only the best hits found in a strata of reads. Unaligned reads were kept separate from this step. Deduplication of aligned reads based on UMIs was performed to correctly assign the number of reads for each mature microRNA in miRBase v22.1. The second round of alignment involved using the unaligned reads against the human reference genome (vGRCh38) with the same parameters except allowing one mismatch instead of zero. Annotation was performed only for those regions which aligned between the chromosomal coordinates of human mature microRNAs from miRBase v22.1 andcounts were generated. Finally, each microRNA from miRBase v22.1 was assigned a combined total count of mature microRNA

Please cite this article as: Ali SA et al., Sequencing identifies a distinct signature of circulating microRNAs in early radiographic knee osteoarthritis, Osteoarthritis and Cartilage, https://doi.org/10.1016/j.joca.2020.07.003

Covariate	Total ($n = 91$)	KL0/1 ($n = 41$)	KL3/4 ($n = 50$)	p-value
SEX				0.55
Female	49 (54)	24 (59)	25 (50)	
Male	42 (46)	17 (41)	25 (50)	
AGE				0.001
Mean (SD)	59.5 (13.6)	50.4 (13.6)	66.9 (8.1)	
Median (min, max)	61 (24, 85)	52 (24, 76)	67 (51, 85)	
BMI				0.35
Mean (sd)	27.7 (4.5)	27.3 (4.9)	28 (4.2)	
Median (min,max)	26.7 (19.9, 44.6)	26 (19.9, 44.6)	26.9 (21, 38.4)	
BMI Category				0.51
Normal	29 (32)	15 (37)	14 (28)	
Overweight	37 (41)	17 (41)	20 (40)	
Obese	25 (27)	9 (22)	16 (32)	
RACE				0.022
Non-White	20 (22)	14 (34)	6 (12)	
White	71 (78)	27 (66)	44 (88)	

Table I

Patient characteristics. Frequencies (%) are provided for categorical variables while mean (standard deviation = SD) and median (min, max) are presented for continuous variables in the full cohort (***n*** = 91) and for the early OA vs late OA groups. Non-parametric Kruskal–Wallis tests are applied for comparisons of continuous variables. Chi-square tests are applied for comparisons of categorical variables. Shown in bold are statistically significant ***p***-values

Osteoarthritis and Cartilage

reads as well as genome based mature microRNA region reads mapping to 2,656 mature microRNAs.

Statistics

Samples with low numbers of aligned reads were removed from the analysis based on having log-reads below the 0.025 normal quantile of the log-reads distribution across all samples ($n = 8$ early OA samples). MicroRNAs were filtered to have at least 10 counts-per-million (CPM) of classified sequences in at least two samples. For principal component analysis, counts were transformed using variance stabilizing transformation before plotting[20]. Inference was performed on the raw counts using a negative binomial regression estimated by quasi-likelihood[21] with trended dispersion and normalized by total aligned sequences. The trimmed mean of m-values (TMM) method was not applied as a large number of the microRNAs were differentially expressed between the early OA group vs the late OA group, violating the TMM assumptions[22]. Covariates potentially affecting differential expression between microRNAs were screened using univariate models[23]. A final list of differentially expressed microRNAs between early OA and late OA samples was obtained using a multivariate model adjusted for age, sex, BMI, race, collection site, and batch, and filtered using a false discovery rate (FDR) of 0.01, estimated by the Benjamini-Hochberg (BH) method. The analysis was implemented in R 3.6.1 using the edgeR package (v3.28.0).

Novel microRNA analysis

Discovery of novel microRNAs was performed using miRDeep v2[24]. This did not include processing of UMI tags, instead duplicate reads were removed based on the read alone. Reads were aligned to the human reference genome (vGRCh38) and used by miRDeep2.pl script, where novel microRNAs were discovered by providing human mature and hairpin microRNA sequences along with mouse mature microRNA sequences for homology assessment. Using custom scripts, only those mature microRNAs were selected which had a significant randfold score and no homology with mouse mature microRNAs.

Gene target and pathway enrichment analysis

MicroRNA gene targets were predicted using mirDIP 4.1[25], and the top 1% of targets for each microRNA were used. Novel microRNA gene targets were predicted using BiTargeting[26], miRanda v3.3a[27], and miRAW[28] using the default parameters specific to each tool, and the top 1% of targets from each prediction tool was used. Potential human 3'UTR targets were downloaded from the AURA v2.4.4 database website[29]. Targets from both predictions were used to calculate pathway enrichment, using pathDIP 4.0.21.2[30] and "Literature curated" pathways excluding Kyoto Encyclopedia of Genes and Genomes (KEGG) due to its lack of specificity to OA[31]. Only pathways with q-value (FDR: BH-method) <0.05 were retained. Protein-protein interactions (PPIs) among targets were obtained from IID version 2018-11[32]. A network was built using NAViGaTOR 3.0.8, including microRNA to target interactions and PPIs. Genes and microRNAs from the network were used to query the Arthritis Data Integration Portal (ADIP; http://ophid.utoronto.ca/adip/), and if present in the database, were highlighted in the network according to the type of study in which they were identified.

Study approval

Written informed consent was obtained from all participants prior to inclusion in the study. All studies were performed under research ethics board (REB) approval from the UHN (REB # 16-5969 and REB # 14-7592) and the University of Western Ontario (REB # 109255).

Please cite this article as: Ali SA et al., Sequencing identifies a distinct signature of circulating microRNAs in early radiographic knee osteoarthritis, Osteoarthritis and Cartilage, https://doi.org/10.1016/j.joca.2020.07.003

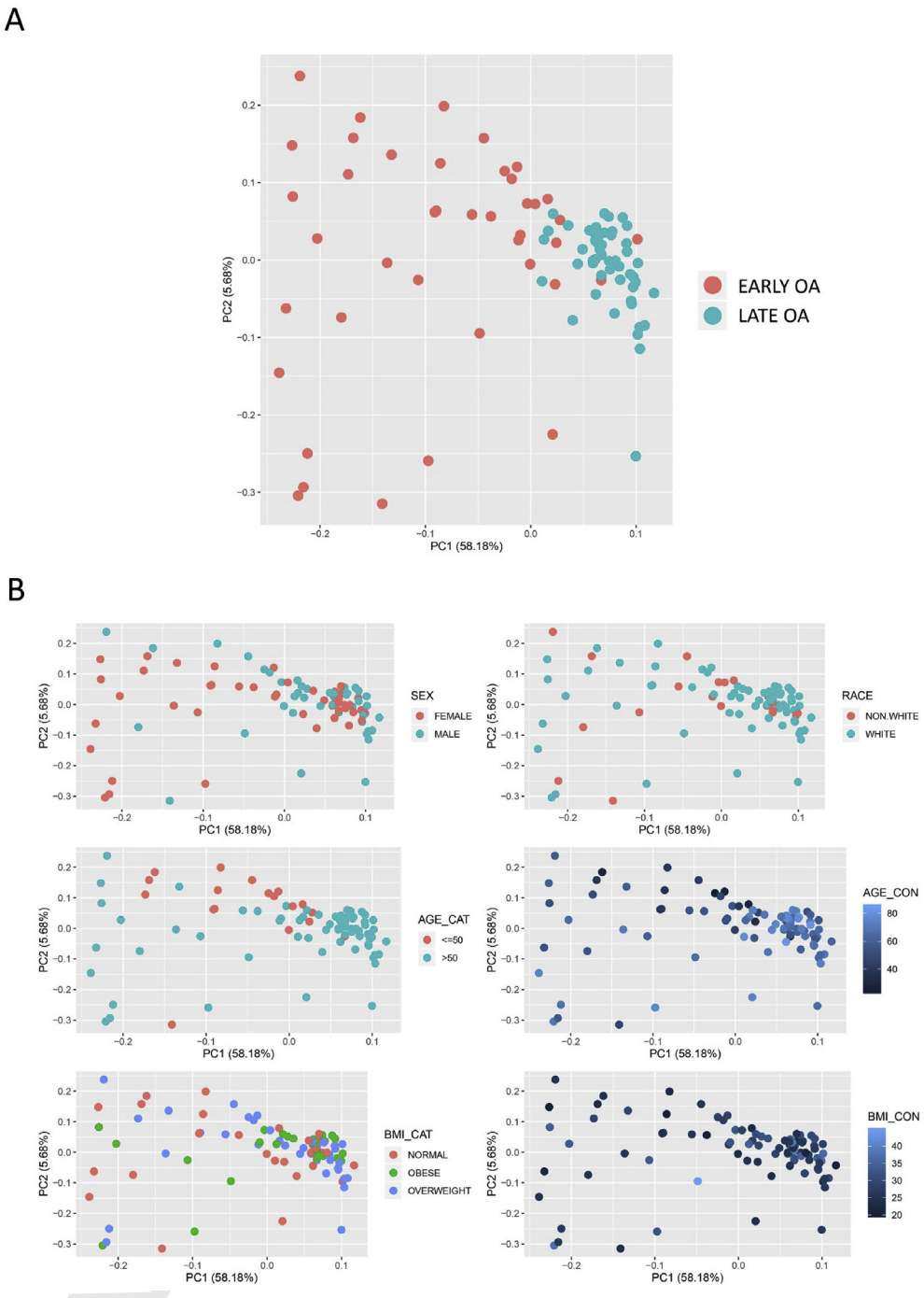


Fig. 2 **Clustering analysis of microRNA sequencing data.** (A) Principal component analysis plot with points coloured according to early OA or late OA. (B) Principal component analysis plot with points coloured according to (L–R) sex, race, age categorical (AGE_CAT), age continuous (AGE_CON), BMI categorical (BMI_CAT; 18.5–24.9 = normal; 25–29.9 = overweight; 30+ = obese), BMI continuous (BMI_CON).

Osteoarthritis and Cartilage

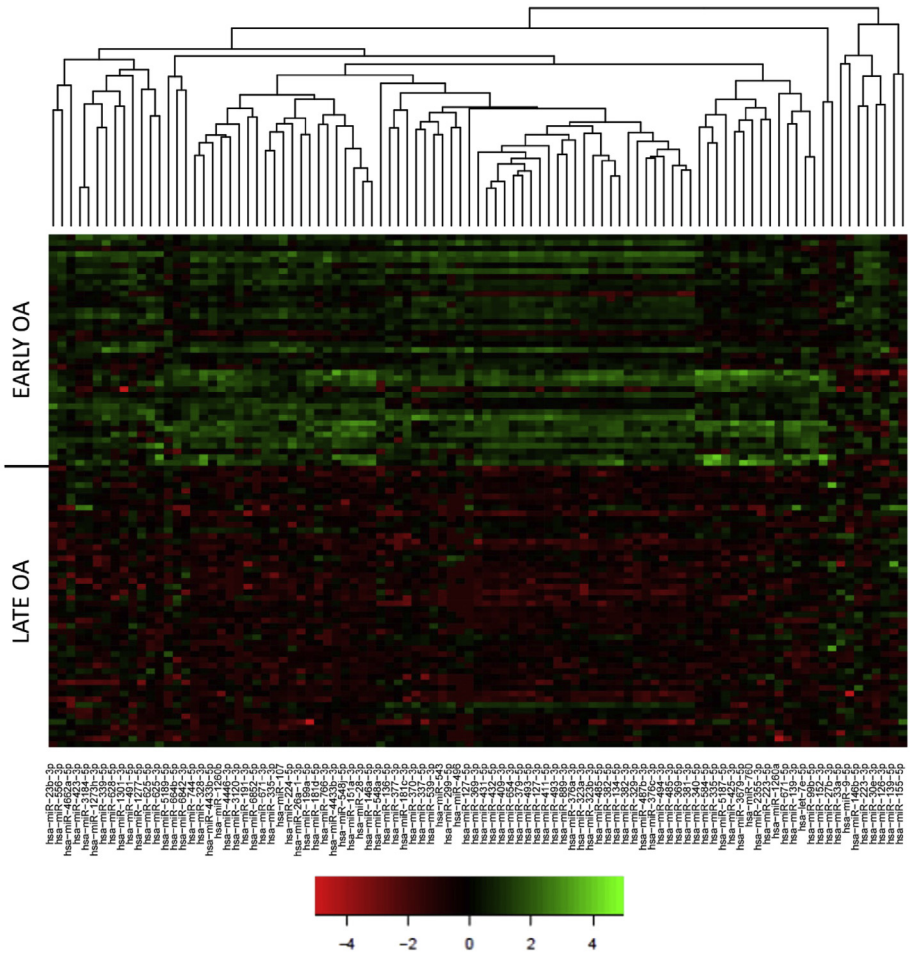
Results

Establishing a pipeline for sequencing microRNAs in human OA plasma

The microRNA-sequencing pipeline we used was carefully documented to support reproducibility (Fig. 1). A limitation of existing sequencing studies is the lack of granularity in reporting sequencing methods and data analysis used to produce final results. In addition to describing our cohorts (Table I), we provide detailed methods for blood sample acquisition, microRNA isolation, library preparation, quality control assessment, sequencing, read alignment, UMI analysis, counts generation, normalization, and statistical and bioinformatics approaches applied. RNA containing

Please cite this article as: Ali SA et al., Sequencing identifies a distinct signature of circulating microRNAs in early radiographic knee osteoarthritis, Osteoarthritis and Cartilage, https://doi.org/10.1016/j.joca.2020.07.003

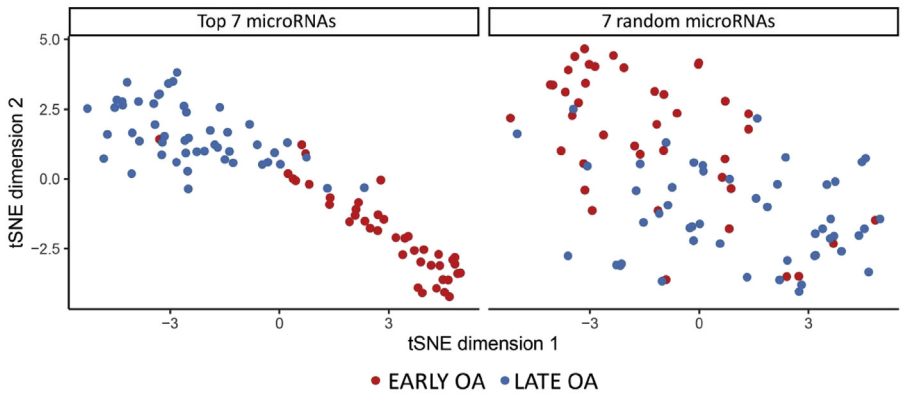
A



B

microRNA (n=7)	logFC	logCPM	FDR	Proportion
hsa-miR-335-3p	3.28	5.60	0.000	1.00
hsa-miR-199a-5p	3.17	6.67	0.000	0.98
hsa-miR-671-3p	3.50	4.43	0.000	0.98
hsa-miR-1260b	3.62	4.25	0.000	0.95
hsa-miR-191-3p	2.41	3.77	0.000	0.95
hsa-miR-335-5p	1.91	9.31	0.000	0.95
hsa-miR-543	3.59	2.90	0.000	0.95

C



microRNAs was isolated from blood plasma of 99 patients with OA and subjected to microRNA library preparation (see Supplementary File 1 for detailed protocol). Fluorometric quantification determined the mean concentration (±SD) of early OA libraries ($n = 49$) was 18.6 ± 4.4 ng/μL and late OA libraries ($n = 50$) was 19.1 ± 4.0 ng/μL. Library integrity assessment determined the average library size of early OA samples was 182 ± 3 bp and late OA samples was 181 ± 2 bp (Supplementary Figure 1).

Please cite this article as: Ali SA et al., Sequencing identifies a distinct signature of circulating microRNAs in early radiographic knee osteoarthritis, Osteoarthritis and Cartilage, https://doi.org/10.1016/j.joca.2020.07.003

Fig. 3

Differentially expressed microRNAs in early OA plasma as compared to late OA plasma. (A) Heatmap showing log fold change in expression of 215 microRNAs with FDR<0.01 in early OA samples (***n*** = 41) as compared to late OA samples (***n*** = 50). (B) MicroRNAs (***n*** = 7) selected based on higher expression across 95% or more of samples in the early OA group as compared to the median expression in the late OA group. logFC = log fold change. logCPM = log counts-per-million. FDR = false discovery rate. (C) T-distributed stochastic neighbour embedding plot, which is a non-linear dimensionality-reduction method, using the seven microRNAs shown in (B; left graph) as variables to group early OA and late OA samples, as compared to seven randomly selected microRNAs (right graph).

Osteoarthritis and Cartilage

Libraries were sequenced and quality assessment of the runs revealed high quality (mean Q30 of 94.4% ± 1.2 SD) and high yield (35G ± 3 SD passing filter; Supplementary Figure 2). This resulted in >13 ± 2 SD million raw reads per sample and an average of 2.0 ± 1.2 SD million mapped reads per sample. Eight samples from the early OA group yielded fewer than 400,000 mapped reads and were excluded, resulting in $n = 41$ early OA samples for subsequent analysis. UMI analysis was performed to account for library amplification bias (Supplementary Figure 3). Counts generated for all 2,656 microRNAs from miRBase v22.1 were normalized by total counts and filtered for microRNAs with at least 10 counts-per-million (CPM) reads in at least two samples, resulting in 417 microRNAs (Supplementary File 2).

Identifying a panel of circulating microRNAs in early radiographic knee OA

Early and late OA cohorts were carefully characterized by sex, age, BMI, and race, where BMI was considered as both a continuous and categorical variable (Table I). As the first step, overall patterns in the sequencing data from early and late OA cohorts were visualized in an unbiased manner through dimensionality reduction and principal component analysis. From this, a distinct cluster of late OA samples emerged, clearly separating from the early OA samples [Fig. 2(A)]. Separation to this degree was not observed for any of the other variables considered, including sex, race, age, and BMI, whether considered as a categorical or continuous variable [Fig. 2(B)]. Age and race were identified as significantly different between groups (Table I), and in the second step were included as factors in multivariate models used to determine differential expression of microRNAs in early OA vs late OA. Using a FDR threshold of less than 0.01, 215 microRNAs were identified.

As the third step, the list of 215 microRNAs was filtered in a biologically relevant manner, selecting microRNAs expressed consistently higher or lower across samples in the early OA group as compared to the median expression level in the late OA group using progressing stringency. This filtering yielded 97 microRNAs that showed higher expression across 85% or more of samples in the early OA group as compared to the late OA group (Fig. 3(A); Supplementary File 2). Among these, seven microRNAs (hsa-miR-335-3p, hsa-miR-199a-5p, hsa-miR-671-3p, hsa-miR-1260b, hsa-miR-191-3p, hsa-miR-335-5p, and hsa-miR-543) had higher expression in 95% or more of samples in the early OA group, including hsa-miR-335-3p which was consistently elevated in 100% of early OA samples, showing a 3.28 log fold difference in early OA vs late OA [Fig. 3(B)]. When a t-distributed stochastic neighbour embedding plot was used to visualize the early OA and late OA clusters generated using these seven microRNAs as variables, clear separation of groups was observed [Fig. 3(C)].

Identifying a panel of novel circulating microRNAs in early radiographic knee OA

Having identified a panel of seven microRNAs in the early OA cohort, we next investigated whether there were novel microRNAs present in early OA, as discovery of novel microRNAs is a major advantage of using sequencing technology. Novel microRNAs were detected based on their predicted secondary structure and lack of homology with murine microRNAs (Supplementary Figure 3). Using progressing stringencies, this list was filtered for novel microRNA sequences that consistently appeared across samples within a cohort (early OA or late OA). Filtering for novel microRNAs that were expressed in ≥85% of samples yielded one novel microRNA in the early OA cohort (novel_miRNA_11) that was also found in ≥85% of samples in the late OA cohort, and four novel microRNAs in the late OA cohort (novel_miRNA_7, 8, 9, 11) that were also found in ≥50% of samples in the early OA cohort. The sequence for each novel microRNA is shown in Fig. 4.

To identify novel microRNAs that were potentially associated with early OA but not late OA and vice versa, we filtered for novel microRNAs that were expressed in ≥50% of samples within a cohort. This revealed 11 novel microRNAs in the early OA cohort and nine microRNAs in the late OA cohort, seven of which were overlapping, leaving four novel microRNAs in the early OA cohort and two novel microRNAs in the late OA cohort (Fig. 4). The four novel microRNAs identified in ≥50% of samples in the early OA cohort were found in 36% or less of samples in the late OA cohort, demonstrating that these four novel microRNAs are more commonly found in patients with early OA than late OA (Fig. 4). Finally, the novel microRNA sequences were compared to a previous list of novel microRNAs generated from sequencing 13 different human tissue types[33]. As shown in Fig. 4, the sequences for four of the microRNAs (novel_miRNA_3, 5, 7, 9) were previously reported, providing independent evidence of the existence of these novel microRNAs, and supporting the method we used for discovery of novel microRNAs.

Identifying a panel of circulating microRNAs in late radiographic knee OA

To identify circulating microRNAs that were associated with late OA, the list of 215 differentially expressed microRNAs was filtered for microRNAs which were consistently higher or lower in expression across samples in the late OA group, as compared to the median expression level in the early OA group. Filtering for microRNAs that were expressed in ≥85% of samples in the late OA group yielded no microRNAs, but filtering at a minimum threshold of 50% yielded three microRNAs (Supplementary File 2). The three microRNAs showed increased expression in late OA as compared to early OA, with hsa-miR-193b-5p, hsa-miR-193a-5p, and hsa-miR-455-5p showing higher expression in 62%, 52%, and 50% of late OA samples, respectively (Supplementary Figure 4). Characterizing novel microRNAs, novel_miRNA_5 and 6 were identified in 50% of

Please cite this article as: Ali SA et al., Sequencing identifies a distinct signature of circulating microRNAs in early radiographic knee osteoarthritis, Osteoarthritis and Cartilage, https://doi.org/10.1016/j.joca.2020.07.003

Novel miRNA	Sequence	Early OA (n=41)	Late OA (n=50)	Londin et al., 2015
novel_miRNA_1	gucuggcucaggguuggg	56 (N=23)	36 (N=18)	Absent
novel_miRNA_2	ucccuguucgggcgccacu	68 (N=28)	20 (N=10)	Absent
novel_miRNA_3	uguuuagcauccuguagccugc	59 (N=24)	4 (N=2)	Present
novel_miRNA_4	uaguggguuaucagaacu	66 (N=27)	14 (N=7)	Absent
novel_miRNA_5	acugaggggaugaaggaucagg	31 (N=13)	50 (N=25)	Present
novel_miRNA_6	caugaauggauuaaugag	44 (N=18)	50 (N=25)	Absent
novel_miRNA_7	ugguccaacgacaggaguagg	58 (N=24)	86 (N=43)	Present
novel_miRNA_8	gaugccugggaguugcgaucug	51 (N=21)	92 (N=46)	Absent
novel_miRNA_9	uuaguggcucccucugccugca	80 (N=33)	86 (N=43)	Present
novel_miRNA_10	aggaagguggggaugacg	66 (N=27)	80 (N=40)	Absent
novel_miRNA_11	uugaggucggacaugguggcu	93 (N=38)	92 (N=46)	Absent
novel_miRNA_12	acuagggauuggggggaau	58 (N=24)	60 (N=30)	Absent
novel_miRNA_13	aggagugggaggagaaug	66 (N=27)	66 (N=33)	Absent

Fig. 4

Novel microRNAs identified in early OA and late OA plasma. Sequences for 13 novel microRNAs that were expressed in 50% or more of samples within each group (Early OA or Late OA). Frequencies (%) and ***n*** values are provided for each group and each microRNA, with only those in 50% or more of samples shown in black. Sequences for novel microRNAs were cross-referenced to those reported by Londin ***et al.***[33] and indicated as 'Present' if previously reported or 'Absent' if not.

late OA samples, with novel_miRNA_5 being previously described[33]. Because novel_miRNA_6 was also present in 44% of samples in the early OA cohort, it is unlikely to be unique to late OA samples (Fig. 4). Since a larger and more consistently expressed panel of microRNAs was identified in early OA samples, subsequent analyses focused on this group.

MicroRNAs, gene targets, and pathway network in early radiographic knee OA

Having identified a panel of microRNAs in early OA, we next performed integrated computational analysis to predict downstream gene targets and pathways modulated by these microRNAs. To minimize filtering and maximize discovery of potential interactions, we chose to use the list of 97 microRNAs that were identified in ≥85% of early OA samples. Gene targets were predicted using mirDIP 4.1[25] and the top 1% of ranked targets for each microRNA were selected (Supplementary File 3). MicroRNA-gene pairs that were validated in previous studies according to TarBase v.7.0[34] and NPinter v.3.0[35] were annotated. Using the four novel microRNAs that were present in ≥50% of early OA samples, gene targets were predicted and the top 1% of ranked targets from each tool were selected (Supplementary File 3). A list of 27 genes was predicted to be common targets of our four novel microRNAs and of at least 10 microRNAs from our list of 97 microRNAs in miRbase (Supplementary File 3). We show that some of these predicted targets contained the seed sequence for binding by the novel microRNAs (Supplementary Figure 5).

The 27 genes that were common were subjected to pathway enrichment analysis using pathDIP 4.0.21.2[30], and 14 pathways were identified that met a q-value (FDR: BH-method) cutoff of <0.05 (Supplementary File 4). To visualize relationships between the microRNAs, gene targets, and pathways, network analysis was performed using NAViGaTOR 3.0.11 (Supplementary File 5)[36]. MicroRNA hsa-miR-548j-5p and novel_miRNA_1 had the greatest number of predicted gene targets (Supplementary File 3) and *SMAD2* had the greatest number of interactions, including microRNAs targeting it and protein-protein interactions (Fig. 5). It was

Please cite this article as: Ali SA et al., Sequencing identifies a distinct signature of circulating microRNAs in early radiographic knee osteoarthritis, Osteoarthritis and Cartilage, https://doi.org/10.1016/j.joca.2020.07.003

present in six of the 14 significant pathways, including the most significant pathway 'Signaling by TGF-beta Receptor Complex' ($p < 0.01$; Fig. 5, Supplementary File 4). Results from Arthritis Data Integration Portal (ADIP) revealed that among our 27 predicted genes, two were previously identified in the blood of OA patients as compared to healthy controls[37], 13 were previously identified when comparing early to late stages of OA synovial fluid and synovial membrane[38], and nine were dysregulated in other joint tissues (Fig. 5).

Discussion

By sequencing circulating microRNAs, we identified a distinct panel of 11 microRNAs in early symptomatic radiographic knee OA, which included four novel microRNAs. We used biologically meaningful prioritization methods to identify 215 microRNAs that were differentially expressed between early and late OA cohorts, 97 of which were consistently higher in ≥85% of early OA samples as compared to late OA samples. Further filtering revealed seven microRNAs that were able to separate early OA and late OA samples based on their presence in ≥95% of early OA samples. We also found 13 novel microRNAs that were present in ≥50% of samples within a group, nine of which had not been previously described[33], and four of which were associated with early OA. Since there remain uncharacterized tissue-specific microRNAs[33], the novel microRNAs identified here merit further investigation as OA-specific markers.

Unlike studies that use arrays or sequencing to identify candidates of interest in only a subset of their cohort, a major strength of our study was the use of sequencing to profile the entire cohort of 91 samples. As described by Kok *et al.*, small sample sizes (<25 per group) in high-throughput microRNA screens are a common pitfall for the identification of biomarkers given the limited reproducibility[39]. Our first filtering step was to select microRNAs with high abundance (≥10CPM in ≥two samples), to increase the likelihood

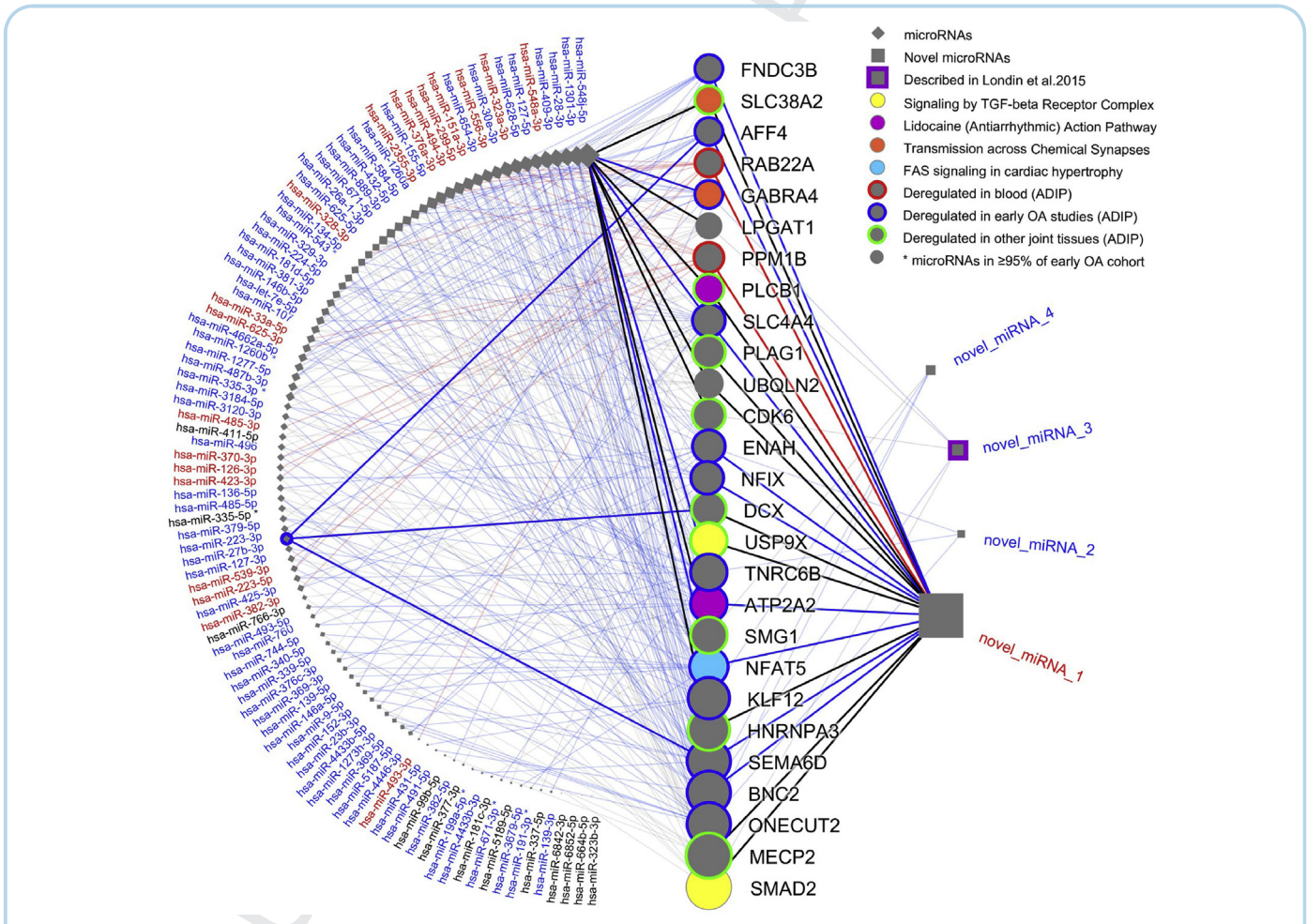


Fig. 5

Network analysis showing microRNAs and predicted gene targets. MicroRNAs (***n*** = 97) are shown on the left, novel microRNAs are shown on the right (***n*** = 4), and putative common gene targets are shown with connecting edges. Pathway membership is shown using node colour as per legend. Genes frequently deregulated in OA according to ADIP are highlighted with node outline as per legend. MicroRNAs targeting them are connected by edges and have labels of the same color. Purple outline highlights the novel microRNA previously described by Londin ***et al.***[33]. Asterisk (*) denotes microRNAs identified in at least 95% of early OA samples.

Osteoarthritis and Cartilage

Please cite this article as: Ali SA et al., Sequencing identifies a distinct signature of circulating microRNAs in early radiographic knee osteoarthritis, Osteoarthritis and Cartilage, https://doi.org/10.1016/j.joca.2020.07.003

of detection in future studies. We then normalized the results by total counts rather than using traditional methods such as TMM, which can mask differences between groups by assuming that the majority of microRNAs are not differentially expressed[22]. These methodological decisions led us to discover a panel of 11 microRNAs with increased likelihood for reproducibility.

To interpret the biological relevance of our panel of microRNAs, including predicting potential functions of novel microRNAs, computational biology approaches were leveraged. Integration of predicted gene targets between the 97 microRNAs and 4 novel microRNAs in early OA revealed 27 common gene targets and multiple interactions. Demonstrating relevance to OA pathophysiology, several matrix metalloproteinases ($n = 9$), aggrecanases ($n = 7$), and collagens ($n = 16$) were predicted as targets of the 97 microRNAs. The four novel microRNAs also showed disease relevance, with *MMP13*, *COL6*, and *COL27* as predicted targets (Supplementary File 3) which could potentially contribute to cartilage catabolism and anabolism.

Annotation with ADIP revealed potential connections to the blood, synovial fluid, and synovial membrane (Fig. 5). *RAB22A* and *PPM1B* were previously identified as significantly upregulated in blood from OA patients as compared to healthy controls[37], and in our prediction results were targeted by 10 microRNAs each (only one of which was common, hsa-miR-556-3p) and novel_miRNA_1. Several microRNAs, genes, and connections were annotated based on reported differences between early stages and late stages of OA in synovial fluid and synovial membrane[38]. Among these was hsa-miR-27b-3p, which was previously shown to be increased in late-stage vs early-stage knee OA synovial fluid[10]. Hsa-miR-548j-5p and novel_miRNA_1 had the most predicted gene targets, including *SMAD2* as a common target, and 'Signaling by TGF-beta Receptor Complex' as the top pathway. This demonstrates the utility of ADIP in predicting disease-relevant microRNA interactions, where in this case, downregulation of TGF-beta signaling at the joint level would be expected to reduce the regenerative potential of cartilage through suppression of type II collagen, lubricin, and aggrecan genes that are required for joint homeostasis[40].

The circulating microRNAs we identified may reflect an early signal that induces a cascade of events that contribute to OA. Experimental validation of our findings is required to identify the source and target tissue(s) of the microRNAs, whether it be the cartilage, synovium, bone, adipose tissue, or others. Among the top 11 microRNAs identified, hsa-miR-335-3p and -5p were highly expressed in 100% and 95% of early OA samples, respectively. Arms of miR-335 are expressed in osteoblasts[41] and mesenchymal stem cells[42], with a role in promoting chondrogenesis[43]. The presence of these microRNAs in the blood plasma suggests that they may act systemically to impact more than one joint or target tissue. Animal models, where it is possible to modulate microRNAs genetically or pharmacologically, are required to investigate the potential impact of these microRNAs in OA pathophysiology. Elucidating the biological mechanism of these microRNAs may have therapeutic relevance since microRNAs can be targeted with antisense oligonucleotides[44]. Given our well-defined cohort of patients and the optimized microRNA sequencing methodology that was applied, our results reflect a distinct panel of circulating microRNAs in early radiographic knee OA. Future studies should be directed towards understanding the role and mechanism of action of the identified microRNAs in OA pathology, and to determine if any of these microRNAs have potential as biomarkers in OA.

Author contributions

SAA was involved in the conception and design of the study, acquisition of sequencing data, analysis and interpretation of data, writing the article, revising it for important intellectual content and approved final version of the manuscript. RG was involved in the creation of early OA cohort, revised the article for important intellectual content and approved final version of the manuscript. PP performed bioinformatics analysis, drafted and revised the article for important intellectual content and approved final version of the manuscript. SK, OEG and KS performed statistical analysis, drafted and revised the article for important intellectual content and approved final version of the manuscript. CP, DBW and IJ performed computational biology analyses, drafted and revised the article for important intellectual content and approved final version of the manuscript. AP, YRR, CV were involved in creating LEAP OA data and biobank, drafted and revised the article for important intellectual content and approved final version of the manuscript. SL and JSR drafted and revised the article for important intellectual content and approved final version of the manuscript. TA provided early OA samples, drafted and revised the article for important intellectual content and approved final version of the manuscript.MK was involved in the creation of early OA cohort, conception and design of the study, interpretation of data, drafting the article, revising it critically for important intellectual content and approved final version of the manuscript.

Competing interest

US Provisional Patent Application No. 63/033,463 titled "Circulating MicroRNAs in Knee Osteoarthritis and Uses Thereof" has been filed.

Funding

This work is supported in part by the Canadian Institute of Health Research Operating grant (# 156299), The Krembil Foundation grant, Canada Research Chairs Program, The Arthritis Society, Canadian Foundation for Innovation, Ontario Research Fund, Ian Lawson Van Toch Fund, IBM and The Arthritis Program via The Toronto General and Western Hospital Foundation, University Health Network.

Acknowledgements

The authors would like to thank Kim Perry, Kala Sundararajan, Mary Nasim and Garth Blackler for their help with clinical samples and data, and Madison McIntyre for her help with illustrations.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.joca.2020.07.003.

References

1. Ratneswaran A, Rockel JS, Kapoor M. Understanding osteoarthritis pathogenesis: a multiomics system-based approach. Curr Opin Rheumatol 2020;32:80–91.
2. Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: approaches and considerations. Nat Rev Genet 2012;13:358–69.
3. Nugent M. MicroRNAs: exploring new horizons in osteoarthritis. Osteoarthritis Cartilage 2016;24:573–80.
4. Wang Z, Lu Y, Han J. Peripheral blood microRNAs: a novel tool for diagnosing disease? Intractable Rare Dis Res 2012;1:98–102.
5. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, *et al.* Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci U S A 2008;105:10513–8.
6. Endisha H, Rockel J, Jurisica I, Kapoor M. The complex landscape of microRNAs in articular cartilage: biology, pathology, and therapeutic targets. JCI Insight 2018;3.

7. Coutinho de Almeida R, Ramos YFM, Mahfouz A, den Hollander W, Lakenberg N, Houtman E, *et al.* RNA sequencing data integration reveals an miRNA interactome of osteoarthritis cartilage. Ann Rheum Dis 2019;78:270–7.
8. Bratus-Neuenschwander A, Castro-Giner F, Frank-Bertoncelj M, Aluri S, Fucentese SF, Schlapbach R, *et al.* Pain-associated transcriptome changes in synovium of knee osteoarthritis patients. Genes 2018;9.
9. Nakamura A, Rampersaud YR, Sharma A, Lewis SJ, Wu B, Datta P, *et al.* Identification of microRNA-181a-5p and microRNA-4454 as mediators of facet cartilage degeneration. JCI Insight 2016;1, e86820.
10. Li YH, Tavallaee G, Tokar T, Nakamura A, Sundararajan K, Weston A, *et al.* Identification of synovial fluid microRNA signature in knee osteoarthritis: differentiating early- and late-stage knee osteoarthritis. Osteoarthritis Cartilage 2016;24:1577–86.
11. Murata K, Yoshitomi H, Tanida S, Ishikawa M, Nishitani K, Ito H, *et al.* Plasma and synovial fluid microRNAs as potential biomarkers of rheumatoid arthritis and osteoarthritis. Arthritis Res Ther 2010;12:R86.
12. Borgonio Cuadra VM, Gonzalez-Huerta NC, Romero-Cordoba S, Hidalgo-Miranda A, Miranda-Duarte A. Altered expression of circulating microRNA in plasma of patients with primary osteoarthritis and in silico analysis of their pathways. PloS One 2014;9, e97690.
13. Beyer C, Zampetaki A, Lin NY, Kleyer A, Perricone C, Iagnocco A, *et al.* Signature of circulating microRNAs in osteoarthritis. Ann Rheum Dis 2015;74:e18.
14. Ntoumou E, Tzetis M, Braoudaki M, Lambrou G, Poulou M, Malizos K, *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 Epigenet 2017;9:127.
15. Kong R, Gao J, Si Y, Zhao D. Combination of circulating miR-19b-3p, miR-122-5p and miR-486-5p expressions correlates with risk and disease severity of knee osteoarthritis. Am J Transl Res 2017;9:2852–64.
16. Rousseau JC, Millet M, Croset M, Sornay-Rendu E, Borel O, Chapurlat R. Association of circulating microRNAs with prevalent and incident knee osteoarthritis in women: the OFELY study. Arthritis Res Ther 2020;22:2.
17. Aae T, Karlsen T, Haugen I, Risberg M, Lian O, Brinchmann J. Evaluating plasma extracellular vesicle microRNAs as possible biomarkers for osteoarthritis. Osteoarthritis Cartilage Open 2020:1.
18. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, *et al.* Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis Rheum 1986;29:1039–49.
19. Kellgren JH, Lawrence JS. Radiological assessment of osteoarthrosis. Ann Rheum Dis 1957;16:494–502.
20. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 2010;11:R106.
21. Lund SP, Nettleton D, McCarthy DJ, Smyth GK. Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. Stat Appl Genet Mol Biol 2012;11.
22. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 2010;11:R25.
23. Ameling S, Kacprowski T, Chilukoti RK, Malsch C, Liebscher V, Suhre K, *et al.* Associations of circulating plasma microRNAs with age, body mass index and sex in a population-based study. BMC Med Genom 2015;8:61.
24. Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res 2012;40:37–52.
25. Tokar T, Pastrello C, Rossos AEM, Abovsky M, Hauschild AC, Tsay M, *et al.* mirDIP 4.1-integrative database of human microRNA target predictions. Nucleic Acids Res 2018;46:D360–70.
26. Veksler-Lublinsky I, Shemer-Avni Y, Kedem K, Ziv-Ukelson M. Gene bi-targeting by viral and human miRNAs. BMC Bioinf 2010;11:249.
27. Betel D, Koppal A, Agius P, Sander C, Leslie C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. Genome Biol 2010;11:R90.
28. Pla A, Zhong X, Rayner S miRAW. A deep learning-based approach to predict microRNA targets by analyzing whole microRNA transcripts. PLoS Comput Biol 2018;14, e1006185. Q3
29. Dassi E, Re A, Leo S, Tebaldi T, Pasini L, Peroni D, *et al.* AURA 2: empowering discovery of post-transcriptional networks. Translation 2014;2, e27738.
30. Rahmati S, Abovsky M, Pastrello C, Kotlyar M, Lu R, Cumbaa CA, *et al.* pathDIP 4: an extended pathway annotations and enrichment analysis resource for human, model organisms and domesticated species. Nucleic Acids Res 2019. Q4
31. Huang H, Zheng J, Shen N, Wang G, Zhou G, Fang Y, *et al.* Identification of pathways and genes associated with synovitis in osteoarthritis using bioinformatics analyses. Sci Rep 2018;8:10050.
32. Kotlyar M, Pastrello C, Malik Z, Jurisica I. IID 2018 update: context-specific physical protein-protein interactions in human, model organisms and domesticated species. Nucleic Acids Res 2019;47:D581–9.
33. Londin E, Loher P, Telonis AG, Quann K, Clark P, Jing Y, *et al.* Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. Proc Natl Acad Sci U S A 2015;112:E1106–15.
34. Sethupathy P, Corda B, Hatzigeorgiou AG. TarBase: a comprehensive database of experimentally supported animal microRNA targets. RNA 2006;12:192–7.
35. Hao Y, Wu W, Li H, Yuan J, Luo J, Zhao Y, *et al.* NPInter v3.0: An Upgraded Database of Noncoding RNA-Associated Interactions. Oxford: Database; 2016.
36. Brown KR, Otasek D, Ali M, McGuffin MJ, Xie W, Devani B, *et al.* NAViGaTOR: network analysis, visualization and graphing Toronto. Bioinformatics 2009;25:3327–9.
37. Ramos YF, Bos SD, Lakenberg N, Bohringer S, den Hollander WJ, Kloppenburg M, *et al.* Genes expressed in blood link osteoarthritis with apoptotic pathways. Ann Rheum Dis 2014;73:1844–53.
38. Wang Q, Rozelle AL, Lepus CM, Scanzello CR, Song JJ, Larsen DM, *et al.* Identification of a central role for complement in osteoarthritis. Nat Med 2011;17:1674–9.
39. Kok MGM, de Ronde MWJ, Moerland PD, Ruijter JM, Creemers EE, Pinto-Sietsma SJ. Small sample sizes in high-throughput miRNA screens: a common pitfall for the identification of miRNA biomarkers. Biomol Detect Quantif 2018;15:1–5.
40. Shen J, Li S, Chen D. TGF-beta signaling and the development of osteoarthritis. Bone Res 2014;2.
41. Chen YJ, Chang WA, Huang MS, Chen CH, Wang KY, Hsu YL, *et al.* Identification of novel genes in aging osteoblasts using next-generation sequencing and bioinformatics. Oncotarget 2017;8:113598–613.

Please cite this article as: Ali SA et al., Sequencing identifies a distinct signature of circulating microRNAs in early radiographic knee osteoarthritis, Osteoarthritis and Cartilage, https://doi.org/10.1016/j.joca.2020.07.003

42. Tornero-Esteban P, Rodriguez-Rodriguez L, Abasolo L, Tome M, Lopez-Romero P, Herranz E, *et al.* Signature of microRNA expression during osteogenic differentiation of bone marrow MSCs reveals a putative role of miR-335-5p in osteoarthritis. BMC Muscoskel Disord 2015;16:182.
43. Lin X, Wu L, Zhang Z, Yang R, Guan Q, Hou X, *et al.* MiR-335-5p promotes chondrogenesis in mouse mesenchymal stem cells and is regulated through two positive feedback loops. J Bone Miner Res 2014;29:1575–85.
44. Nakamura A, Rampersaud YR, Nakamura S, Sharma A, Zeng F, Rossomacha E, *et al.* microRNA-181a-5p antisense oligonucleotides attenuate osteoarthritis in facet and knee joints. Ann Rheum Dis 2019;78:111–21.

Please cite this article as: Ali SA et al., Sequencing identifies a distinct signature of circulating microRNAs in early radiographic knee osteoarthritis, Osteoarthritis and Cartilage, https://doi.org/10.1016/j.joca.2020.07.003