Journal of Biosciences and Medicines, 2026, 14(6), 345-353
https://www.scirp.org/journal/jbm
ISSN Online: 2327-509X
ISSN Print: 2327-5081

Exploratory Transcriptomic Analysis of a Selected Comparison in the GSE206848 Synovium Microarray Dataset

Pengyuan He[1,2,3], Junxiu Zhou[4], Lizhu Lu[4], Haidong Zhou[2,3], Jihua Wei[2,3*]

[1]Clinical Medical College of Youjiang Medical University for Nationalities, Baise, China
[2]Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, China
[3]Guangxi Key Laboratory for Preclinical and Translational Research on Bone and Joint Degenerative Diseases, Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, China
[4]Baise Municipal Maternal and Child Health Hospital, Baise, China
Email: *1261290953@qq.com

How to cite this paper: He, P.Y., Zhou, J.X., Lu, L.Z., Zhou, H.D. and Wei, J.H. (2026) Exploratory Transcriptomic Analysis of a Selected Comparison in the GSE206848 Synovium Microarray Dataset. *Journal of Biosciences and Medicines*, **14**, 345-353.
https://doi.org/10.4236/jbm.2026.146023

Received: May 20, 2026
Accepted: June 22, 2026
Published: June 25, 2026

Abstract

Background: Osteoarthritis (OA) is a common degenerative joint disease in which synovial inflammation contributes to pain and structural progression. The public GSE206848 dataset contains normal, OA, and rheumatoid arthritis (RA) synovium samples and can be used for exploratory transcriptomic screening. During revision, the sample labels in the uploaded online analysis report were checked against the GEO record. The online report used seven normal synovium samples as the reference group and two test samples; these test samples correspond to RA synovium rather than the seven OA synovium samples. Therefore, the findings are presented only as exploratory results from the selected comparison and should not be interpreted as OA diagnostic biomarkers. **Methods:** This study did not reprocess raw CEL files or independently redownload a new raw expression matrix from GEO. Instead, the revision was based on the processed expression matrix and differential-expression output provided in the uploaded online analysis report for GSE206848. According to the report documentation, processed expression values were retrieved from GEO by GEOquery and analyzed with limma in R. The dataset is listed on platform GPL570, Affymetrix Human Genome U133 Plus 2.0 Array. Quantile normalization was applied in the online workflow to align sample expression distributions. Differentially expressed genes (DEGs) were screened using |log2 fold change| > 1 and nominal $P < 0.05$; Benjamini-Hochberg adjusted P values were also reported. **Results:** The online report included seven reference samples and two test samples. Box plots showed broadly comparable

expression distributions after normalization, and PCA suggested separation between the two selected groups. A total of 1082 genes met the nominal screening threshold. Only CYP2C19 remained significant after multiple-testing correction among the representative genes listed in the original table. Other genes such as LEP, ADIPOQ, GPD1, PLIN1, CRP, PTX3, KRT6A, and KLF9 should be described as nominally different or representative genes rather than FDR-significant genes. **Conclusion:** This exploratory analysis identified genes showing differential expression in the selected GSE206848 comparison included in the online report. Because the test group contains only two samples and corresponds to RA rather than OA, the results cannot support OA diagnostic-biomarker or therapeutic-target claims without reanalysis of the seven OA samples and independent validation.

Keywords

GSE206848, Synovium, Microarray, Exploratory Analysis, Limma, Osteoarthritis, Rheumatoid Arthritis

1. Introduction

Osteoarthritis (OA) is a highly prevalent joint disease characterized by cartilage degeneration, subchondral bone remodeling, osteophyte formation, and variable synovial inflammation. Synovitis is increasingly recognized as an important contributor to pain, functional impairment, and disease progression [1]-[3]. Transcriptomic analysis of synovial tissue can help identify molecular alterations associated with joint inflammation and tissue remodeling, but interpretation depends strongly on sample selection, group definition, multiple-testing correction, and validation strategy [4]-[6].

GSE206848 is a public synovium microarray dataset deposited in the Gene Expression Omnibus (GEO), a public archive for functional genomics data [7] [8]. The GEO record describes 16 samples in total: seven normal synovium samples, seven OA synovium samples, and two RA synovium samples [8]. In the original manuscript, the dataset was described as an OA versus normal comparison. However, review of the online analysis report and the sample labels shown in the provided figures indicates that the report compared seven reference samples with two test samples. According to the GEO sample labels, the two-sample test group corresponds to RA synovium samples rather than OA synovium samples [8].

For this reason, the present revised manuscript corrects the overinterpretation in the original version. The analysis is retained as an exploratory differential-expression screen based on the selected online-report comparison. Claims regarding OA biomarkers, lipid-metabolism pathways, energy balance, inflammatory pathways, and therapeutic targets are softened because no formal GO/KEGG enrichment analysis or independent validation was performed.

2. Materials and Methods

2.1. Data Source and Sample Definition

The present study did not reprocess raw CEL files from GEO. Instead, analyses were based on the processed expression matrix and differential-expression results provided in the uploaded online analysis report associated with GSE206848. GEO describes this dataset as human synovium expression profiling by array and lists platform GPL570, Affymetrix Human Genome U133 Plus 2.0 Array [8]. The dataset contains normal, OA, and RA synovium samples. In this revision, the exact samples used by the uploaded online report were explicitly listed to improve reproducibility (Table 1).

Table 1. Sample definition and status in the uploaded online analysis report.

Sample ID	GEO label	Analysis group	Tissue/source label	Status in online report
GSM6265690	NS1	Reference	Normal/non-OA synovium	Included
GSM6265691	NS2	Reference	Normal/non-OA synovium	Included
GSM6265692	NS3	Reference	Normal/non-OA synovium	Included
GSM6265693	NS4	Reference	Normal/non-OA synovium	Included
GSM6265694	NS5	Reference	Normal/non-OA synovium	Included
GSM6265695	NS8	Reference	Normal/non-OA synovium	Included
GSM6265696	NS10	Reference	Normal/non-OA synovium	Included
GSM6265697-GSM6265703	OAS1, OAS2, OAS6, OAS7, OAS8, OAS9, OAS10	OA	OA synovium	Not used in the uploaded online report
GSM6265704	RAS5	Test	RA synovium	Included
GSM6265705	RAS7	Test	RA synovium	Included

Normal synovium was defined according to GEO sample labels beginning with NS and described as non-OA/normal synovium. RA samples were not appropriate for an OA-versus-normal analysis. They were included in the uploaded online report as the two test samples, which is why the present revision avoids presenting the results as OA-specific. To support an OA-specific conclusion, the analysis should be rerun using GSM6265697-GSM6265703 as the OA group and excluding GSM6265704-GSM6265705.

2.2. Data Preprocessing and Normalization

According to the documentation of the uploaded online report, processed expression values were retrieved from GEO through GEOquery and analyzed using the limma package in R [9] [10]. The present revision evaluated and interpreted the processed expression matrix reported by that workflow rather than generating a new expression matrix from raw microarray CEL files. Because the data were from Affymetrix GPL570 and the plotted sample values were on a typical microarray

log2 scale, the expression matrix was treated as log2-transformed or already transformed processed expression data. Quantile normalization was used in the online workflow to align the distributions of expression values across samples. This step should not be described as removing batch effects unless an explicit batch variable and batch-correction method are used.

Probe annotation followed the GPL570 platform annotation file. Probes without gene annotation were removed. Where multiple probes mapped to the same gene symbol, the probe with the highest expression signal was retained as the representative probe, as described in the online report.

2.3. Differential Expression Analysis

Differential expression analysis was performed using limma with group as the explanatory variable [10]. The design and contrast can be written as follows:

```
group <- factor(c(rep("ref", 7), rep("test", 2)), levels = c("ref", "test"))
design <- model.matrix(~0 + group)
colnames(design) <- c("ref", "test")
contrast.matrix <- makeContrasts(test_vs_ref = test - ref, levels = design)
fit <- lmFit(exprSet, design)
fit2 <- contrasts.fit(fit, contrast.matrix)
fit2 <- eBayes(fit2)
```

DEGs were screened using $|\log2FC| > 1$ and nominal $P < 0.05$. Adjusted P values were calculated using the Benjamini-Hochberg method. Because most representative genes did not pass $FDR < 0.05$, nominally significant genes are described as exploratory or representative genes rather than confirmed significant biomarkers.

2.4. Visualization

Box plots were used to assess sample expression distributions after normalization. PCA was used to visualize global expression differences between groups. Volcano plots summarized the relationship between log2FC and statistical significance. A heatmap displayed the expression patterns of selected DEGs and hierarchical clustering across samples; the heatmap visualization followed standard approaches for complex expression matrices [11].

3. Results

3.1. Data Quality and Sample Distribution

As shown in **Figure 1**, the median expression values of the nine selected samples were broadly aligned after normalization, suggesting that the distributions were comparable. **Figure 2** shows that the reference and test groups were separated in the reduced-dimensional PCA space. These observations support the use of the normalized matrix for exploratory differential expression analysis. However, comparable box-plot medians do not prove that batch effects were removed.

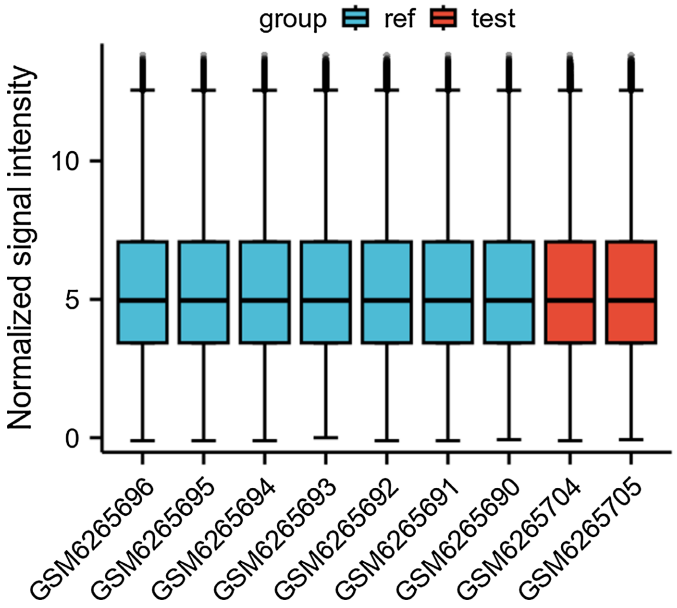

Figure 1. Box plot of normalized signal intensity in the selected samples. The medians are broadly comparable after normalization. The analysis included 7 normal (reference) and 2 RA (test) samples.

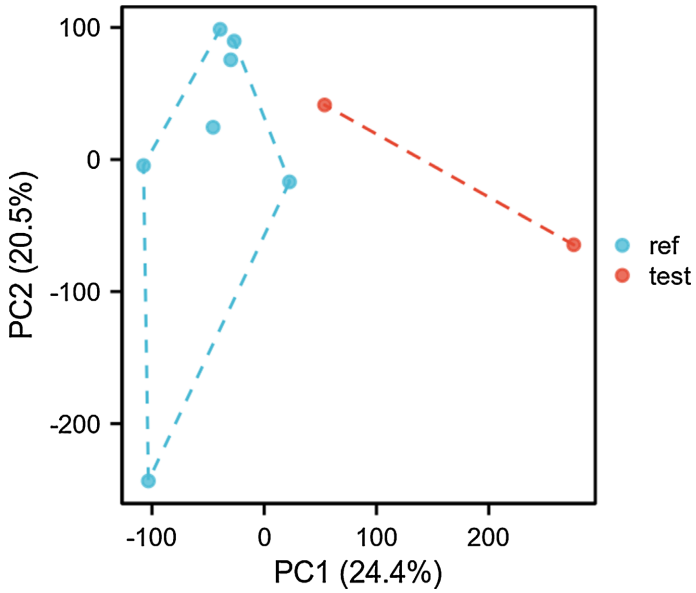

Figure 2. PCA plot of the selected samples. The reference and test groups show separation in the reduced-dimensional space. The analysis included 7 normal (reference) and 2 RA (test) samples.

3.2. Differential Expression Statistics

Using $|\log2FC| > 1$ and nominal $P < 0.05$, the online report identified 1,082 genes in the selected comparison (**Table 2**). The number of genes changed substantially under different fold-change thresholds, indicating that the results are sensitive to threshold choice. **Figure 3** summarizes the genes that passed the nominal screening threshold in the volcano plot.

Table 2. Number of genes identified under different nominal screening thresholds.

Screening criteria	Number of genes
|log2FC| > 2 and P < 0.05	103
|log2FC| > 1 and P < 0.05	1,082
|log2FC| > 0.58 and P < 0.05	2,536

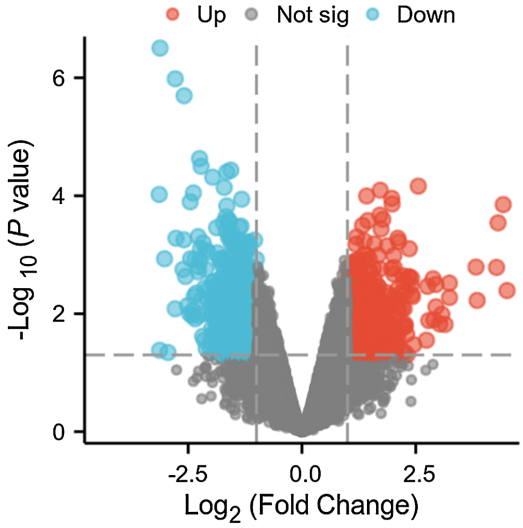

Figure 3. Volcano plot showing genes that passed the nominal screening threshold in the selected comparison. The analysis included 7 normal (reference) and 2 RA (test) samples.

3.3. Representative Genes

Representative genes from the original table are shown in Table 3. Because only CYP2C19 reached FDR significance among these listed genes, the other entries should be interpreted as nominally different exploratory genes in this selected comparison. The heatmap of selected DEGs is shown in Figure 4.

Table 3. Representative genes listed in the selected comparison from the online report.

Gene	log2FC	P value	adj. P value	Direction	Known function
GPD1	4.51	0.0040	0.231	Up	Glycerol-3-phosphate dehydrogenase; glycerol metabolism
LEP	4.31	0.0003	0.125	Up	Leptin; adipokine and inflammatory mediator
PLIN1	4.27	0.0016	0.175	Up	Lipid droplet-associated protein
ADIPOQ	3.14	0.0152	0.305	Up	Adiponectin; metabolic and immunomodulatory functions
SCD	2.35	0.0024	0.200	Up	Fatty acid desaturation
CYP2C19	−3.12	3.13e−07	0.017	Down	Cytochrome P450 enzyme; drug metabolism
PTX3	−3.13	0.0416	0.376	Down	Acute-phase/inflammatory response protein
CRP	−2.59	0.0006	0.144	Down	C-reactive protein; systemic inflammatory marker
KRT6A	−3.02	0.0012	0.158	Down	Keratin; epithelial differentiation
KLF9	−2.47	0.0100	0.282	Down	Transcription factor

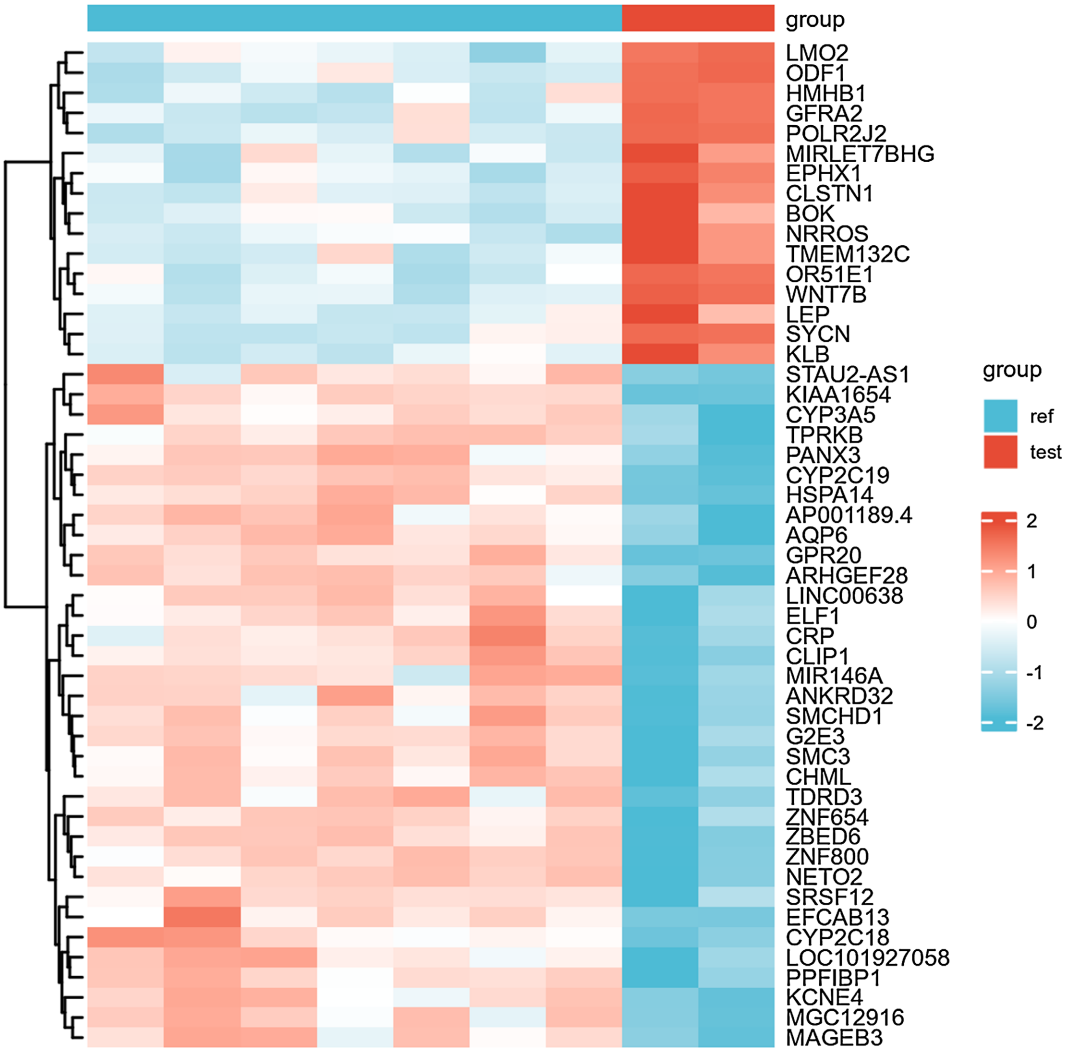

Figure 4. Heatmap of selected DEGs. The red annotation bar represents the two test samples used by the online report. The analysis included 7 normal (reference) and 2 RA (test) samples.

4. Discussion

This revised analysis addresses several methodological issues raised during review. First, the platform information has been standardized throughout the manuscript. GSE206848 is listed on platform GPL570, Affymetrix Human Genome U133 Plus 2.0 Array, and this designation is used consistently in all sections of the revised manuscript [8]. Second, the sample composition is explicitly reported. The online report used seven reference samples and two test samples, and the two test samples correspond to RA labels. This prevents the results from being interpreted as a direct OA-versus-normal comparison.

Third, the statistical interpretation has been corrected. The original wording described several genes as significantly differentially expressed. In the revised version, genes are described according to the evidence level. CYP2C19 passed the Benjamini-Hochberg correction in the representative table, whereas LEP, ADIPOQ, GPD1, PLIN1, SCD, CRP, PTX3, KRT6A, and KLF9 did not reach FDR

significance in that table. These genes may be useful for generating hypotheses, but they require reanalysis in the correct OA sample set and independent validation before biomarker claims can be made [5] [6].

Fourth, pathway-level claims have been softened. Although several representative genes have known roles in lipid metabolism, adipokine signaling, or inflammation, no GO or KEGG enrichment analysis was performed in the uploaded report. Therefore, this manuscript no longer claims that lipid metabolism, energy balance, or inflammatory response pathways were statistically enriched. These biological themes are discussed only as possible functional annotations of representative genes in the context of synovial inflammation and the synovium-synovial fluid microenvironment [1]-[3].

The most important limitation is the sample selection problem. The dataset itself contains seven OA samples, but the uploaded online report used two test samples that appear to correspond to RA synovium. Therefore, the present results should be treated as an exploratory analysis of selected synovial samples rather than an OA-specific biomarker study. A corrected OA study should rerun the pipeline with seven normal samples and seven OA samples, exclude the two RA samples, and then validate candidate genes in an independent cohort or by qPCR/immunohistochemistry.

5. Conclusion

This exploratory analysis identified a set of genes showing differential expression in the selected comparison included in the uploaded online report. Because the comparison involved seven normal synovium samples and two RA synovium samples, and because most representative genes did not remain significant after multiple-testing correction, the findings should be regarded as hypothesis-generating observations rather than validated disease-associated biomarkers. These data do not support diagnostic-biomarker or therapeutic-target claims for OA. Further analyses using the complete OA sample set and independent validation are required before biological or clinical conclusions can be drawn.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Sanchez-Lopez, E., Coras, R., Torres, A., Lane, N.E. and Guma, M. (2022) Synovial Inflammation in Osteoarthritis Progression. *Nature Reviews Rheumatology*, **18**, 258-275. https://doi.org/10.1038/s41584-022-00749-9

[2] Ingale, D., Kulkarni, P., Electricwala, A., Moghe, A., Kamyab, S., Jagtap, S., *et al.* (2021) Synovium-Synovial Fluid Axis in Osteoarthritis Pathology: A Key Regulator of the Cartilage Degradation Process. *Genes*, **12**, Article 989. https://doi.org/10.3390/genes12070989

[3] Smolinska, V., Klimova, D., Danisovic, L. and Harsanyi, S. (2024) Synovial Fluid Markers and Extracellular Vesicles in Rheumatoid Arthritis. *Medicina*, **60**, Article

1945. https://doi.org/10.3390/medicina60121945

[4] Moore, L., Pan, Z. and Brotto, M. (2022) RNAseq of Osteoarthritic Synovial Tissues: Systematic Literary Review. *Frontiers in Aging*, **3**, Article ID: 836791. https://doi.org/10.3389/fragi.2022.836791

[5] He, M., Yu, Q., Xiao, H., Dong, H., Li, D. and Gu, W. (2024) Screening and Validation of Key Genes Associated with Osteoarthritis. *BMC Musculoskeletal Disorders*, **25**, Article No. 954. https://doi.org/10.1186/s12891-024-08015-7

[6] Liao, C.S., He, F.Z., Li, X.Y., Zhang, Y. and Han, P.F. (2024) Analysis of Common Differential Gene Expression in Synovial Cells of Osteoarthritis and Rheumatoid Arthritis. *PLOS ONE*, **19**, e0303506. https://doi.org/10.1371/journal.pone.0303506

[7] Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., *et al.* (2013) NCBI GEO: Archive for Functional Genomics Data Sets—Update. *Nucleic Acids Research*, **41**, D991-D995. https://doi.org/10.1093/nar/gks1193

[8] GEO Accession GSE206848 (2022) Dysregulated Gene Expression in Human Osteoarthritic (OA) and Rheumatoid Arthritis (RA) Synovium. Gene Expression Omnibus, National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE206848

[9] Davis, S. and Meltzer, P.S. (2007) GEOquery: A Bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*, **23**, 1846-1847. https://doi.org/10.1093/bioinformatics/btm254

[10] Smyth, G.K. (2005) Limma: Linear Models for Microarray Data. In: Gentleman, R., Carey, V.J., Huber, W., Irizarry, R.A. and Dudoit, S., Eds., *Statistics for Biology and Health*, Springer-Verlag, 397-420. https://doi.org/10.1007/0-387-29362-0_23

[11] Gu, Z., Eils, R. and Schlesner, M. (2016) Complex Heatmaps Reveal Patterns and Correlations in Multidimensional Genomic Data. *Bioinformatics*, **32**, 2847-2849. https://doi.org/10.1093/bioinformatics/btw313