Medicine®

Exploration of biomarkers in osteoarthritis based on bioinformatics

Tong Ye, MM, Zhou Haoyuan, MM, Zhou Bei, BM, Xu Kangyong, BM*

Abstract

Osteoarthritis (OA) seriously affects human health and brings a heavy social burden. This study aimed to identify new biomarkers involved in OA. Differential expression analysis and gene set enrichment analysis were performed on the microarray data set of OA. Identify key genes from immune-related DEGs and verify their expression in the validation set. CIBERSORT was used to analyze the infiltration of immune cells. The correlation between key genes and immune cells were conducted. A total of 1779 DEGs were identified in GSE82107. Gene set enrichment analysis results of top 4 for hallmark revealed the enrichment of DEGs were associated with genes in "HALLMARK_TNFA_SIGNALING_VIA_NFKB", "HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION", "HALL-MARK_INFLAMMATORY_RESPONSE" and "HALLMARK_HYPOXIA". A total of 108 immune-related DEGs were identified from the overlap between 2498 immune-related genes and 1779 DEGs. The expression of top 6 immune-related DEGs including ADIPOQ, FABP4, FOS, IGLC1, IGLV1–44 and leptin were measured in the validation set, the results shown that IGLC1 and IGLV1–44 might play a key role in the synovial membrane of OA. A total of 8 kinds of cells including B cells memory, Plasma cells, T cells CD4 memory resting, T cells gamma delta, natural killer cells activated, macrophages M0, Mast cells resting and Mast cells activated have significant differences in infiltration between the OA group and the control group. Besides, the expressions of IGLC1 and IGLV1–44 are highly correlated. Our results indicated that IGLC1 and IGLV1–44 may play the role of immune-related biomarkers in OA.

Abbreviations: ADIPOQ = Adiponectin, C1Q And Collagen Domain Containing, BMP7 = bone morphogenetic protein 7, COMP = cartilage oligomeric matrix protein, CXCL8 = C-X-C motif chemokine ligand 8, DEGs = differentially Expressed Genes, FABP4 = fatty acid-binding protein 4, FOS = Fos proto-oncogene, GEO = Gene Expression Omnibus database, GSEA = gene set enrichment analysis, IGLC1 = immunoglobulin lambda constant 1, IGLV1–44 = immunoglobulin lambda variable 1–44, MMP13 = matrix metallopeptidase 13, MMP3 = matrix metallopeptidase 3, MSigDB = molecular signatures database, NK = natural killer, OA = osteoarthritis, TGF-β = transforming growth factor beta 1.

Keywords: biomarkers, differentially expressed genes, immune infiltration, osteoarthritis

1. Introduction

Osteoarthritis (OA) is a chronic joint disease characterized by progressive deterioration of the hyaline cartilage concomitant with changes in the surrounding tissues, including the ligaments, synovium, and subchondral bone.[1,2] OA is a multifactorial disease, and both environmental and genetic factors play important roles in the etiology. Genetic biomarkers and improved phenotypic characterization are necessary to enhance our understanding of the genetic complexity underlying the pathogenesis and disease progression.[3–5]

Previous studies have shown that early diagnosis and treatment can help reduce the progressive joint damage in OA patients. Presently, there are several possible OA biomarkers, such as BMP7, COMP, MMP3, MMP13, CXCL8, and TGF-β.[6–8] Emerged in the past 10 years, the gene chip technology has been widely used to identify the gene expression profiles during diseases. Bioinformatics analysis of microarray data is extremely helpful in exploring the molecular mechanisms of diseases and uncovering disease markers.[9] However, most of the available data have not been effectively used because they were obtained for different research purposes and from diverse tissues, and the available data about immune-related biomarkers are even more limited.

This study used bioinformatics analysis to identify biomarkers that are highly associated with OA, with the hope that they may be used in diagnostic and prognostic prediction in OA.

2. Material and methods

In the present study, ethical approval was unnecessary because all analytical data were derived from publicly available database (https://www.ncbi.nlm.nih.gov/geo/).

2.1. Data collection and processing

Four microarray datasets including GSE82107, GSE55235, GSE55457, and GSE51588 were acquired from the GEO (https://www.ncbi.nlm.nih.gov/geo/) database. GSE82107 and

Brief information of the 4 datasets.

GSE ID	Platform	Sample	Tissue type
GSE82107	GPL 570	7 Control, 10 OA	Synovial membrane
GSE55235	GPL 96	10 Control, 10 OA	Synovial membrane
GSE55457	GPL 96	10 Control, 10 OA	Synovial membrane
GSE51588	GPL13497	10 Control, 40 OA	Subchondral bone

OA = osteoarthritis.

GSE51588 datasets were based on GPL570 and GPL13497 platforms, respectively. GSE55235 and GSE55457 datasets were both based on GPL96 platform. The information about all the datasets is briefly shown in Table 1. All raw data were processed with background correction and normalization by using the affy and oligo packages. Moreover, GSE55235 and GSE55457 datasets were merged after normalization, and the sva package was applied to remove the batch effect of the merged dataset, which was used as a validation set.

2.2. Identification of DEGs and GSEA

The limma package was used to perform differential gene expression analysis on the normalized data of GSE82107. $P < .05$ and the absolute value of log2 fold change ≥1 were set as the thresholds for DEGs. Subsequently, DEGs were subjected to gene set enrichment analysis (GSEA) by using the Clusterprofiler package.[10] The hallmarks in the molecular signatures database (MSigDB, http://software.broadinstitute.org/gsea/msigdb/index.jsp) were selected as reference. $P < 0.05$ was set as the cutoff criterion for GSEA.

2.3. Analysis of immune-related DEGs

A total of 2498 immune-related genes were obtained from the ImmPort database and intersected with DEGs. The intersecting genes were regarded as immune-related DEGs. Next, based on the absolute value of fold change, we selected the top 6 immune-related DEGs and checked their expression in the validation set (GSE55235 and GSE55457). Among these 6 immune-related DEGs, those that were significantly different in the validation set and under expressional regulation that was consistent with the previous ones were regarded as potential key genes. Additionally, the expression patterns of the key genes in GSE51588 dataset were confirmed in the subchondral bones of patients with OA.

2.4. Immune infiltration with CIBERSORT analysis

To estimate the immune cell composition in the synovial membrane, the CIBERSORT algorithm was used. The samples in the validation set were filtered with $P < .05$ set as the threshold based on the CIBERSORT results. The immune cells with significantly different infiltration levels between the OA patients and healthy controls were selected and further analyzed for Pearson correlation with the key genes. $P < .05$ indicates a significant correlation.

3. Results

3.1. Flow chart

The flow chart of the identification process of immune-related markers is shown in Figure 1.

3.2. Differential expression genes analysis

A total of 1779 DEGs were identified in GSE82107 (control and OA; n = 7 and 10, respectively), including 383 up-regulated and 1396 downregulated genes (Fig. 2A). The top 4 GSEA results for hallmark revealed that the DEGs were enriched for genes in "HALLMARK_TNFA_SIGNALING_VIA_NFKB", "HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION", "HALLMARK_INFLAMMATORY_RESPONSE" and "HALLMARK_ HYPOXIA" (Fig. 2B).
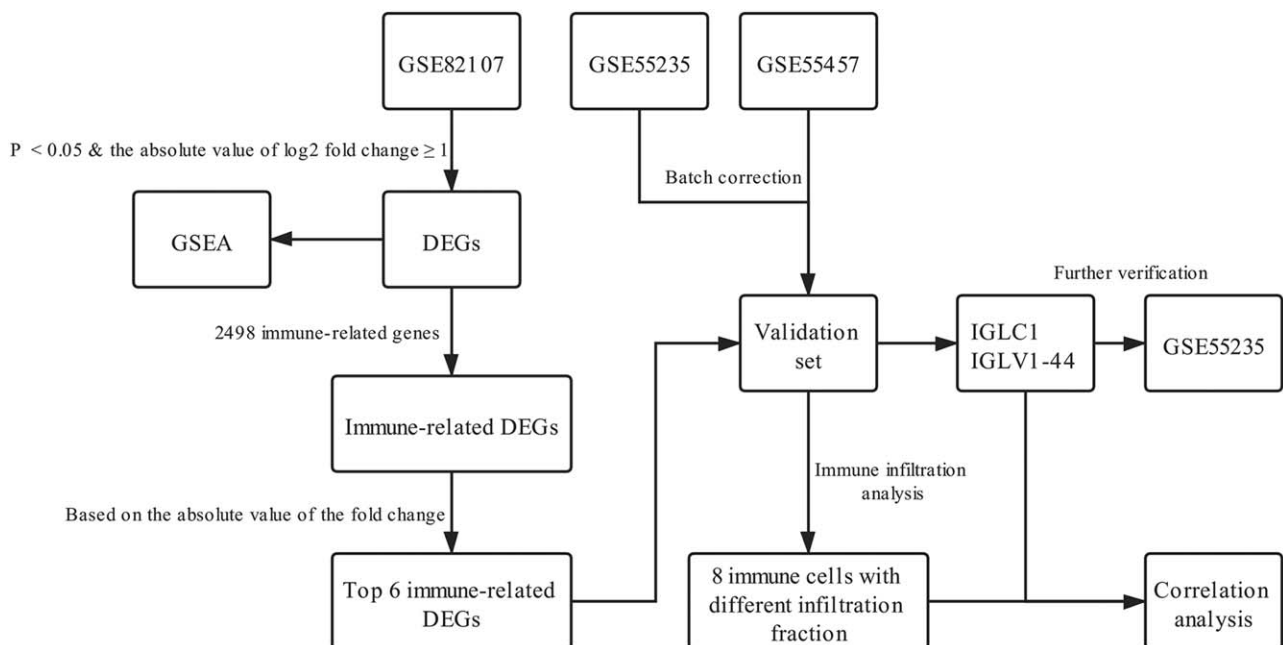

Figure 1. Flowchart for identification of immune-related markers in OA.
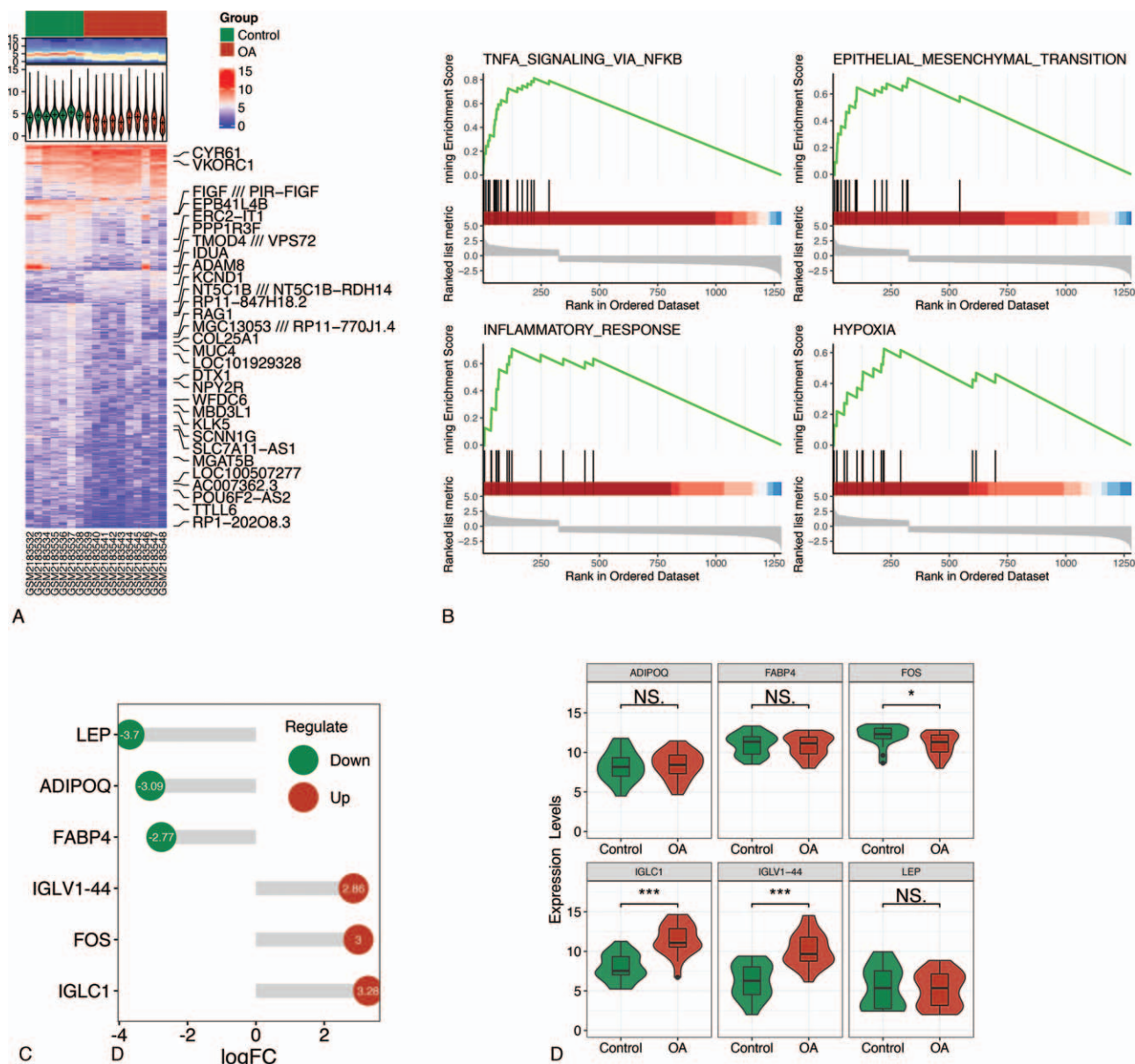
Figure 2. (A) Heatmap of DEGs. (B) GSEA of DEGs. (C) Expression regulation of top 6 immune-related DEGs. (D) Expression levels of top 6 immune-related DEGs in the validation set. DEG = differentially expressed genes, GSEA = gene set enrichment analysis.

3.3. Identification of key genes

A total of 108 immune-related DEGs were identified from the overlap between 2498 immune-related genes and 1779 DEGs. The expression regulation of the top 6 immune-related DEGs, namely ADIPOQ, FABP4, FOS, IGLC1, IGLV1–44, and leptin (LEP), are shown in Figure 2C. The expression of the above six genes was measured in the validation set (Fig. 2D); the results showed that IGLC1 and IGLV1–44 might play key roles in the synovial membrane of OA. To test the reliability of the key genes, we evaluated their expression levels in the subchondral bone. Compared with the levels in the control group, IGLC1 and IGLV1–44 were upregulated in the OA group. Especially, the expression level of IGLC1 was significantly different between the 2 groups (Fig. 3).

3.4. Assessment of immune infiltration

The level of immune cell infiltration for the validation set is shown in Figure 4A. Among all 22 immune cells, the infiltration levels of 8 types of cells (M0 macrophages, and memory B, plasma, resting CD4 memory T, gamma delta T, activated natural killer (NK), resting mast, and activated mast cells) significantly differed between the OA and control groups (Fig. 4B). Additionally, we performed a correlation analysis between the key genes and immune cells with differential infiltration scores to further elucidate the relationship between the key genes and cellular immunity. IGLC1 and IGLV1–44 showed significant positive correlation with memory B, plasma, and resting mast cells, but significant negative correlation with resting CD4 memory T, activated NK, and activated mast cells
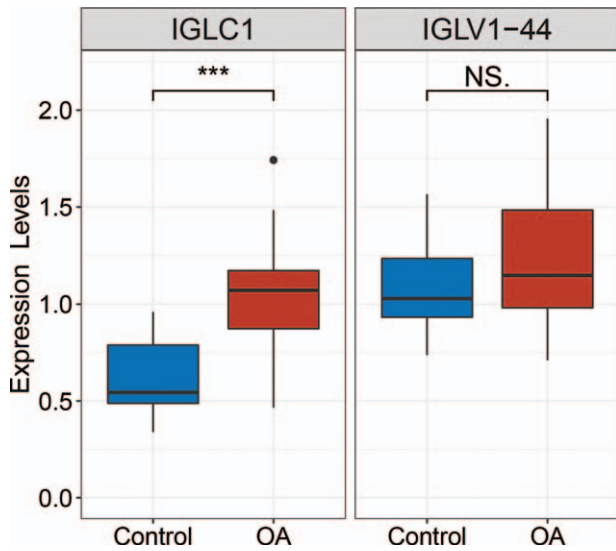
Figure 3. The expression levels of IGLC1 and IGLV1-44 in subchondral bone of OA patients.

(Fig. 4C and D). In addition, IGLC1 and IGLV1–44 highly significantly correlated with each other (Fig. 4E).

4. Discussion

Osteoarthritis is the most common chronic progressive joint disorder and is generally considered a whole-organ disease, causing organ dysfunction or joint failure.[11,12] Patients with OA usually experience joint pain and stiffness, and severe cases show limitation of movement.[13] Due to the lack of effective treatments and the heavy socioeconomic and physiological burden of the disease symptoms, identification of individuals at risk is critical for timely preventive measurements. However, the mechanism of OA pathogenesis remains unclear, and there are no effective biomarkers for early diagnosis.

In this study, the DEGs between OA patients and healthy individuals were identified using the available microarray data. A total of 1779 DEGs were observed. The results of GSEA indicated that the DEGs were closely related to the TNFA signaling via NFKB, epithelial-mesenchymal transition, inflammatory response, and hypoxia. Next, we obtained 108 immune-related DEGs by evaluating the intersection between immune-related genes and the DEGs. The top six immune-related DEGs, sorted by log2FC, were ADIPOQ, FABP4, FOS, IGLC1, IGLV1–44, and leptin. After analyzing the expression levels of these genes in the validation set, IGLC1 and IGLV1–44 were screened out. Furthermore, we employed the GSE51588 dataset to analyze the expression levels of
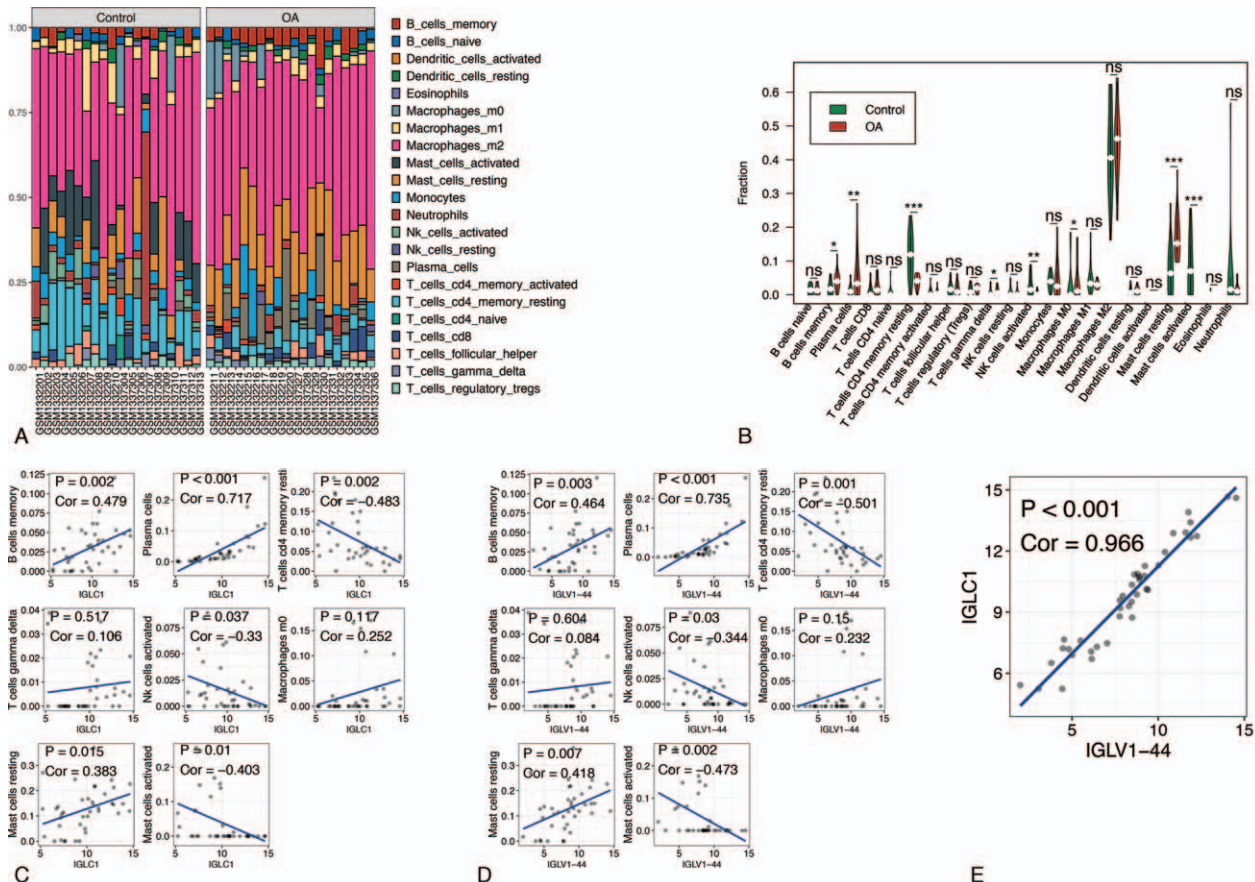

Figure 4. (A, B) Immune infiltration analysis between control and OA. (C) Correlation analysis between IGLC1 and 8 kinds of immune cells. (D) Correlation analysis between IGLV1-44 and 8 kinds of immune cells. (E) Correlation analysis between IGLC1 and IGLV1-44.

the key genes in the subchondral bone. As shown in Figure 2, the expression levels of the key genes do not differ between the subchondral bone and synovium of OA patients.

By using the CIBERSORT algorithm, we evaluated the infiltration levels of 22 immune cells in OA tissues. The results showed that M0 macrophages and memory B, plasma, gamma delta T, resting mast, and activated mast cells infiltrate OA tissues, whereas resting CD4 memory T and activated NK cells are underrepresented in OA tissues versus healthy controls. Finally, we used Pearson correlation analysis to assess for a correlation between these immune cells and the key genes. Surprisingly, the results showed that these eight immune cells are correlated with IGLC1 almost the same as they are correlated with IGLV1–44. Meanwhile, we found that the expression levels of IGLC1 and IGLV1–44 are very highly correlated with each other. Unfortunately, there are limited data about the involvement of IGLC1 and IGLV1–44 in human diseases, and there is no report on their roles in OA. We speculate that the synergistic effect of up-regulated IGLC1 and IGLV1–44 in OA has a certain effect on the infiltration of immune cells, but the underlying mechanism needs to be further studied.

5. Conclusions

In conclusion, IGLC1 and IGLV1–44 in OA tissues were upregulated compared with control. IGLC1 and IGLV1–44 are expected to become new biomarkers in OA.

Author contributions

Conceptualization: Zhou Bei, Xu Kang Yong.
Data curation: Tong Ye.
Investigation: Tong Ye, Zhou Hao Yuan.
Methodology: Tong Ye.
Validation: Tong Ye, Zhou Hao Yuan.
Writing – original draft: Tong Ye.
Writing – review & editing: Zhou Bei, Xu Kang Yong.

References

[1] Mobasheri A, Rayman MP, Gualillo O, et al. The role of metabolism in the pathogenesis of osteoarthritis. Nat Rev Rheumatol 2017;13: 302–11.
[2] Saberi Hosnijeh F, Bierma-Zeinstra SM, Bay-Jensen AC. Osteoarthritis year in review 2018: biomarkers (biochemical markers). Osteoarthritis Cartilage 2019;27:412–23.
[3] Roseti L, Desando G, Cavallo C, et al. Articular cartilage regeneration in osteoarthritis. Cells 2019;8.
[4] Xie C, Chen Q. Adipokines: new therapeutic target for osteoarthritis? Curr Rheumatol Rep 2019;21.
[5] Zhang M, Theleman JL, Lygrisse KA, et al. Epigenetic mechanisms underlying the aging of articular cartilage and osteoarthritis. Gerontology 2019;65:387–96.
[6] Alonso B, Bravo B, Mediavilla L, et al. Osteoarthritis-related biomarkers profile in chronic anterior cruciate ligament injured knee. Knee 2020; 27:51–60.
[7] Roberts HM, Law R-J, Thom JM. The time course and mechanisms of change in biomarkers of joint metabolism in response to acute exercise and chronic training in physiologic and pathological conditions. Eur J Appl Physiol 2019;119:2401–20.
[8] van Spil WE, Szilagyi IA. Osteoarthritis year in review 2019: biomarkers (biochemical markers). Osteoarthritis Cartilage 2020;28:296–315.
[9] Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. Nat Rev Genet 2019;20:631–56.
[10] Yu G, Wang L-G, Han Y, et al. clusterProfiler: an R Package for comparing biological themes among gene clusters. OMICS 2012;16: 284–7.
[11] Blanco FJ, Rego-Pérez I. Editorial: is it time for epigenetics in osteoarthritis? Arthritis Rheumatol 2014;66:2324–7.
[12] Rego-Pérez I, Durán-Sotuela A, Ramos-Louro P, et al. Mitochondrial genetics and epigenetics in osteoarthritis. Front Genet 2020;10.
[13] van der Spoel E, van Vliet NA, van Heemst D. Viewpoint on the role of tissue maintenance in ageing: focus on biomarkers of bone, cartilage, muscle, and brain tissue maintenance. Ageing Res Rev 2019;56:100964.