The comprehensive analysis of hub genes and pathways in osteoarthritis

Hong Ren[1], Zhendi Song[1], Chen Chao[2], Weizheng Mao[1]*

1, Department of General Surgery, Qingdao Municipal Hospital, Qingdao University School of Medicine, 5 Donghai Zhong Road, 266000, Qingdao, Shandong, China.

2, Department of General Surgery, Jintan Hospital, Jiangsu University School of Medicine, 16 South Gate Street, Changzhou, Jiangsu, 213200, China.

***Correspondence:**

Weizheng Mao, Department of General Surgery, Qingdao Municipal Hospital, Qingdao University School of Medicine, 5 Donghai Zhong Road, Qingdao, Shandong, 266000, People's Republic of China.

E-mail: greenwood8838@163.com;

Abstract

Purpose: Until now, the pathogenesis of osteoarthritis (OA) is yet unclarified. Thus, it is needed to uncover OA pathogenesis for developing efficacious treatment against OA.

Materials and Methods: We downloaded the expression profile GSE117999 from the National Center for Biotechnology Information (NCBI) database. The differentially expressed genes (DEGs) were screened by the R limma software package. Metascape was applied for Genetic Oncology (GO) analysis, containing biological process, cell component and molecular function. Protein-protein Interaction (PPI) enrichment analysis was conducted by Metascape. We utilized Gene Set Enrichment Analysis (GSEA) to investigate potential signaling pathways associated with OA. Finally, the relationship between the expression of the hub gene and human OA cells was explored through the Cell Counting Kit -8 (CCK-8) experiment.

Result: By processing the GSE117999 database, we acquired 1,318 differentially expressed genes (DEGs) in total, comprising 893 genes with up-regulated expression

and 425 genes with down-regulated expression. We utilized Metascape to assess the DEGs, so as to obtain pivotal pathways related to OA, consisting of autophagy and muscle structure development. Meanwhile, the Metascape was utilized to construct a PPI network to obtain hub genes TAS2R40, PDYN, KALRN, PRRC2A and SIN3A, which were all related to OA. GSEA was applied to obtain key pathways, containing 16p11.2 distal deletion, pyrimidine metabolism disorders biomarkers, pyrimidine metabolism and related diseases and base excision repair. Besides, the experimental results showed that knocking down the expression of the TAS2R40 gene significantly inhibited the proliferation of OA cells.

Conclusion: This study uses microarray technology and bioinformatics analysis to determine hub genes and vital pathways related to OA, providing new insights into exploring OA pathogenesis and biomarkers.

Keywords: Osteoarthritis; DEGs; Bioinformatics analysis; Hub genes; Signaling pathways

Background

Degenerative osteoarthritis is also called osteoarthritis (OA), degenerative arthritis, senile arthritis and hypertrophic arthritis[1], which is usually induced by age, obesity, strain, trauma, joint congenital abnormalities, joints degeneration and articular cartilage damage, joint edges along with reactive hyperplasia of subchondral bone induced by deformities and others[2]. OA is generally occurred in the middle-aged and elderly, and also widely exists in weight-bearing joints and joints with the cervical vertebra, lumbar spine, knee, hip and so on[3]. Overweight-bearing or over-employing joints also lead to the degeneration of articulation[4]. Joint pain, tenderness, stiffness, joint swelling, restricted mobility and joint deformities were the chronic clinical performance[5]. OA is the primary one to cause the whole disability, followed by bringing about severe morbidity and the cost of the healthcare system. Annually, 17.9 out of 1,000 patients are newly generated OA cases. The incidence ratio of the female is 23.7/1000, and 72.6/1000 among them is over the age of 60

years[6]. Additionally, knee OA accounts for nearly four-fifths of the whole OA cases, and it is still increasing with obesity and age[7].

Relieving pain and maintaining function are the principles of treating OA. Age, weight and joint trauma caused by repeated exercise, especially squatting and kneeling postures are common risk factors for OA[2]. As OA pathogenesis is complicated and unclear, there is no effective treatment method at present. Physicotherapy, analgesia and surgical interventions including joint replacement are the present recommended strategies towards OA[8]. Herein, uncovering the important genes and pathways related to OA is thus essential, which is conducive to clarifying OA pathogenesis and giving a new-split orientation for treating OA. Understanding the pathogenesis of OA is of great significance for its early prevention and treatment.

Here, we employ microarray technology and bioinformatics to investigate the differentially expressed genes (DEGs) based on the GSE117999 dataset, and further assess the transcription and biological processes in the articular cartilage between OA patients and normal people. Finally, we obtain the hub gene and signaling pathways related to OA and verify the effect of hub gene expression on human OA cells (HASMC) through experiments, which is of great significance for exploring the pathogenesis, biomarkers and treatments of OA.

Material and Methods

Gene expression microarray data acquisition

We downloaded gene expression profile GSE117999 from NCBI (https://pubmed.ncbi.nlm.nih.gov/). Gene Expression Omnibus (GEO)[9] database (http://www.ncbi.nlm.nih.gov/geo) is a public repository responsible for archiving and freely distributing the full set of microarrays, second-generation sequencing data and other high-throughput functional genomic data submitted by the scientific community for free. GEO database assists users in querying and downloading studies and gene expression patterns stored in GEO via offering a series of web-based interfaces and

applications, in addition to data storage. The gene expression profile GSE117999 came from GPL20844 Agilent-072363 SurePrint G3 Human GE v3 8x60K Microarray 039494. Gene transcripts in cartilage were from 12 patients with and without OA.

Identification of DEGs

We used the limma (linear models for microarray data) package in the *R* software (version 3.3.2, https://www.r-project.org/) to screen the DEGs between patients with and without OA in GSE117999. The genes with *P*-value<0.001 and fold change (FC)>1 were DEGs with up-regulated expression, whereas those with FC<1 were DEGs with down-regulated expression. Heat map package in *R* software was utilized to generate the heat maps of DEG in these 3 samples.

GO enrichment analyses of DEGs

As a powerful tool for gene function annotation analysis, Metascape (http://metascapeorg/) could employ the current prestige bioinformatics analysis to analyze the expression profiles of genes and proteins in batches, and explore gene or protein functions. In order to further understand the function of DEG, we carried out Metascape for Gene Ontology (GO)[10] analysis, containing biological process (BP) and cell component (CC), molecular function (MF) enrichment terms.

Protein-protein Interaction (PPI) Enrichment Analysis

The Metascape tool utilized the Molecular Complex Detection (MCODE) algorithm to explore dense networks of PPI. Individual MCODE component was marked with a different color and each component's biological significance was characterized. We employed functional enrichment analysis to investigate individual MCODE components. And the three functional terms with the highest p-score were regarded as the function with regard to MCODE components.

Gene Set Enrichment Analysis (GSEA)

As a calculation method, GSEA[11] was used to determine whether a set of a priori defined genes display statistically significant and consistent differences existing in two biological states. GSEA was conducted by The Molecular Signatures Database (MSigDB) (https://www.gsea-msigdb.org/gsea/msigdb).

Cell culture and transfection

We placed the human OA cells (HASMC) purchased from American Type Culture Collection in Dulbecco's modified Eagle medium containing 10% fetal bovine serum at 37°C with 5% CO_2.

For cell transfection experiments, we first purchased si-TAS2R40 and its negative control (si-NC) from Gemma Gene (Shanghai, China), a Shanghai biotechnology company. Afterwards, the HASMC cells were transfected with Lipofectamine 3000 (Lipo3000, Invitrogen, Carlsbad, CA, USA) according to the instructions provided by the manufacturer.

Cell proliferation assay

For cell proliferation experiments, Cell Counting Kit-8 (CCK-8) was the first choice for cell verification. First, we added HASMC cells to a 96-well plate at a density of 2×10^3/well, and then injected 10μl of CCK-8 reagent into each well. After that, on the first day, the second day, the third day and the fourth day, the optical density (OD) of the cells at 450 nm was measured with a microplate reader, and the proliferation curve was drawn.

Statistical analysis

SPSS 22.0 (Chicago, USA) was applied to analyze the data. The differences existing in two groups or multiple groups were assessed by Student's t test or ANOVA. $P<0.05$ meant there was a significant difference.

Result

Identification of the DEGs

We analyzed samples from 12 patients with OA and 12 patients without OA in the gene expression profile GSE117999. When the $P<0.001$ and FC>1, it was considered to be significantly up-regulated gene; when the $P<0.001$ and FC<1, it was considered to be significantly down-regulated gene. Through the processing of the limma package in the *R* software, we obtained a total of 1,318 DEGs, of which 893 were up-regulated and 425 were down-regulated. Figure 1 heatmap displayed distribution and clustering of 1,318 DEGs in GSE117999.

GO enrichment analysis of DEGs

Metascape was employed to perform GO enrichment analysis of DEGs. For up-regulated DEGs, the 10 most significantly enriched gene sets were detection of stimulus involved in sensory perception, antimicrobial peptides, GPCR ligand binding, Class C/3 (Metabotropic glutamate/pheromone receptors), rab protein signal transduction, visual perception, phase II-Conjugation of compounds, keratinocyte differentiation, water-soluble vitamin metabolic process and galanin receptor pathway (Fig. 2A). To further obtain the relationships between the enriched terms, top significantly enriched terms were selected to construct a network diagram (Fig. 2B). Besides, GO enrichment analysis of down-regulated DEGs showed that the 19 most significantly enriched gene sets were peptidyl-threonine phosphorylation, attachment of spindle microtubules to kinetochore, regulation of Rac protein signal transduction, autophagy, PCNA-Dependent Long Patch Base Excision Repair, production of miRNAs involved in gene silencing by miRNA, hindbrain development, progesterone-mediated oocyte maturation, meiotic cell cycle, nucleic acid phosphodiester bond hydrolysis, Glioma, pid netrin pathway, protein deacylation, hippocampus development, negatively regulated protein catabolic process, muscle structure development, microtubule cytoskeleton organization and retina homeostasis (Fig. 3A). Likewise, top significantly enriched terms were selected to construct a network diagram (Fig. 3B)

GSEA of the genes in OA tissues

GSEA showed the significantly enriched gene sets of 16p11.2 distal deletion (Fig. 4A), biomarkers for pyrimidine metabolism disorders (Fig. 4B), pyrimidine metabolism and related diseases (Fig. 4C) and base excision repair (Fig. 4D) were significantly enriched in patients with OA. It implied that these pathways perhaps functioned importantly in OA occurrence and prognosis, which was essential for exploring the pathogenesis and biomarkers of OA.

Module analysis of up-regulated PPI network

The PPI network of up-regulated DEGs was constructed through Metascape software (Fig. 5A). According to the MCODE method, 7 protein subclusters were identified, namely MCODE1, MCODE2, MCODE3, MCODE4, MCODE5, MCODE6 and MCODE7 (Fig. 5B). The red module contained 13 genes, they were TAS2R41, TAS2R40, TAS2R60, TAS2R4, CHRM2, TAS2R3, PDYN, CCL4L2, CCR8, CXCR6, CX3CL1, PPBP and CCR10. The blue module contained 8 genes, they were KALRN, EDN3, QRFPR, GRM5, CCKAR, UTS2, CYSLTR and NPSR1. The green module contained 4 genes, they were KRT1, KRT23, KRT37 and KRT34. The purple module contained 4 genes, they were DEFA3, DEFA4, DEFB136 and DEFB108B. The yellow module contained 3 genes, they were RAB9B, RAB39A and RAB39B. The brown module contained 3 genes, they were OPCML, TECTA and OTOA.

Module analysis of up-regulated PPI network

The PPI network of down-regulated DEGs was presented in Figure 6A. According to the MCODE method, 3 protein subclusters were identified, namely MCODE1, MCODE2 and MCODE3 (Fig. 6B). The red module contained 21 genes, they were SNRNP70, PKP2, PRRC2C, HNRNPA3, SNAPC4, KIF14, LARP4, RBM6, YTHDC2, PABPC1L, WDR1, FGF3, NOLC1, CALML3, PTK7, PTK2, TSC1, MAP2K1, GPD1, AGPAT5 and PLPP5. The blue module contained 8 genes, they were ZNF426, ZNF793, ZNF814, ZNF17, MZF1, ZNF107, SIN3B and ZNF404. The green module contained 3 genes, they were COG6, NAPB and BET1.

Knockdown of TAS2R40 gene hindered the proliferation of HASMC cells

In order to explore the effect of hub gene expression on the biological activities of OA cells, we used siRNA to knock down the expression of TAS2R40, and then compared with si-NC, we observed that the low expression of TAS2R40 significantly inhibited the proliferation of HASMC cells (Figure 7).

Discussion

OA is degenerative articular cartilage associated with inflammation, which is characterized by joint stiffness and pain[12]. OA is the most common among middle-aged and elderly people, and its incidence has been on the rise in recent years. At present, there is no cure for osteoarthritis, which has caused a heavy economic burden on patients, families and society[13]. In addition, the molecular mechanism of OA is still unclear, and there is an urgent need for new and effective biomarkers to improve the diagnosis and prevention of OA occurrence and progression[14]. Bioinformatics is a newly developing subject field and applied to analyze genomic DNA sequence information as the source, thus obtaining the information of protein-coding region, simulating and predicting the spatial structure of protein, and then carrying out the necessary drug design according to specific protein function[15]. The later three mentioned parts constitute important components regarding bioinformatics. As far as the specific content of bioinformatics research, bioinformatics comprises new algorithms and statistical methods research, data analysis and interpretation, as well as data use and management. Niraj Shenoy, et al. show that HIF1α is not targeted by 14q deletion in clear cell renal carcinoma, and not an inhibitor of neoplasm in this carcinoma by bioinformatics analysis[16]. Lizhe Zhu et al. demonstrate through bioinformatics data that PINK 1 affects the prognosis of cancer patients, possibly because of its interaction with infiltrating immune cells[17].

We carried out Metascape software to conduct GO enrichment analysis of DEGs with up-regulated and down-regulated expressions, respectively. The DEGs with

down-regulated expression were dramatically enriched in autophagy and muscle structure development. Autophagy is a conservative decomposition process and has been extensively studied in human degenerative diseases[18]. Autophagy mediates the energy cycle of damaged proteins and dysfunctional organelles to maintain the metabolic homeostasis of cells under certain pressure[19]. Research by Xiaodong Li et al. concludes that curcumin inhibits cell apoptosis and inflammatory signals through its effect on chondrocyte autophagy induced by ERK 1/2. Chondrocyte apoptosis exhibits a close relationship to OA progression[20]. Ayumi Tsukada et al. point out that muscle weakness is related to the pathology of OA, and the occurrence of OA can cause changes in muscle development[21].

By constructing the PPI network through Metascape software, we obtained genes related to OA: TAS2R40, PDYN, KALRN, PRRC2C and SIN3B. Among them, the expression of TAS2R40 gene was also related to the proliferation of HASMC cells. TAS2R is the member of the bitter receptor family belonging to the G protein coupled receptor superfamily, and mainly expressed in taste receptor cells of tongue and palate epithelium[22]. TAS2R encodes a seven transmembrane receptor protein that functions as a bitter taste receptor. TAS2R is flocked together with 8 other taste receptor genes on chromosome 7 and Chloramphenicol is a TAS2R agonist[23]. PDYN (Prodynorphin)-related diseases contain Spinocerebellar Ataxia 23[24] and Cocaine Dependence[25]. GO annotations related to this gene include opioid peptide activity[26]. KALRN (Kalirin RhoGEF Kinase)-associated diseases consist of KALRN include Coronary Heart Disease 5 and Schizophrenia 1[27]. The main KALRN-related pathways are p75 NTR receptor-mediated signaling and GPCR signaling. TRIO is homologous to KALRN. PRRC2C (Proline Rich Coiled-Coil 2C) associated diseases contained Spastic Paraplegia 8, Autosomal Dominant and Encephalopathy, Ethylmalonic. PRRC2A is one homologous gene of PRRC2C[28]. SIN3A is one homologous gene of SIN3B. SIN3B (SIN3 Transcription Regulator Family Member B)-associated pathways include PEDF Induced Signaling and HDAC Class I mediated signaling events. Chromatin binding is the main pathway of SIN3B

by GO analysis[29].

GSEA analysis data demonstrate key pathways related to all OA genes were 16p11.2 distal deletion, biomarkers for pyrimidine metabolism disorders, pyrimidine metabolism and related diseases and base excision repair. Ao Wang et al. find that the X-ray severity of OA is positively correlated with insulin resistance, and the results indicate that OA may be related to impaired glucose metabolism[30].

Conclusion

Collectively, we analyze gene expression profile GSE117999 to acquire 1,318 DEGs associated with OA, including 893 up-regulated and 425 down-regulated genes. We further assess OA-induced impact on gene transcripts and biological processes in the articular cartilage, and obtain the hub genes and signaling pathways related to OA. Among them, the low-expressed TAS2R40 gene is found to affect the proliferation process of HASMC cells through experiments. The above results are of great significance for exploring the pathogenesis and biomarkers of OA.

Acknowledgements

Not applicable.

Funding

None

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors contributions

Development of methodology: Weizheng Mao;

Sample collection: Zhendi Song;

Analysis and interpretation of data: Chen Chao;

Writing, review, and/or revision of the manuscript: Hong Ren.

Ethics approval and consent to participate

This research has received approval from the Ethics Committee of the Qingdao Municipal Not applicable.

Hospital and all the patients were informed and consent to this study.

Conflict of interest

The authors declare that they have no conflict of interest.

Figure 1. The Heatmap diagram showed 893 DEGs with up-regulation and 425 DEGs with down-regulation. Yellow and Purple separately indicated DEGs with up-regulation and down-regulation, respectively. The expression intensity value was obtained from the gene expression level after R software analysis.

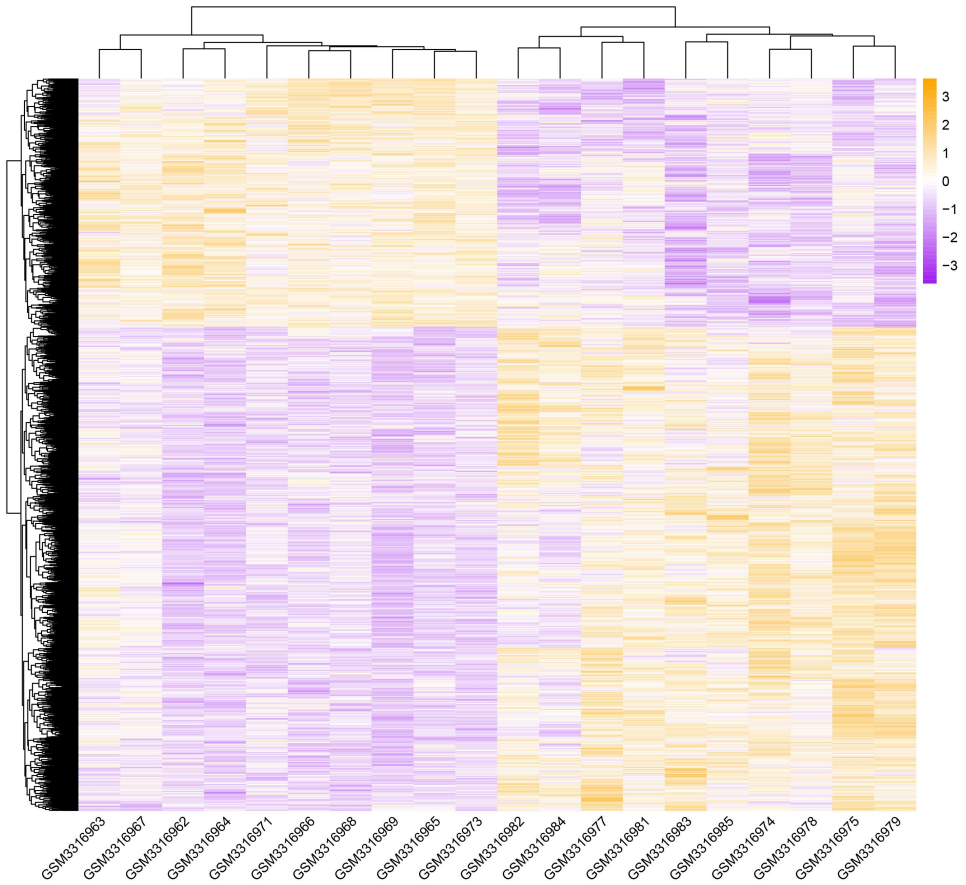
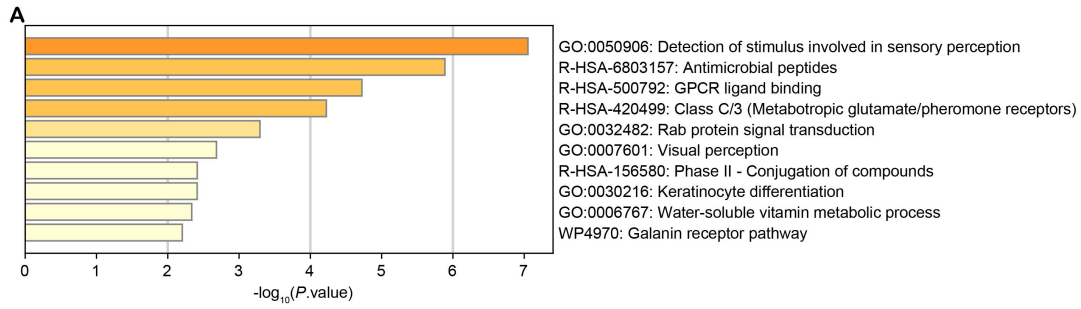


Figure 2. Functional analysis of 893 DEGs with up-regulation.

A. Metascape analysis of the top 10 significantly enriched clusters amid up-regulated DEGs.

B. Metascape analysis of the relationship between the top enriched clusters.



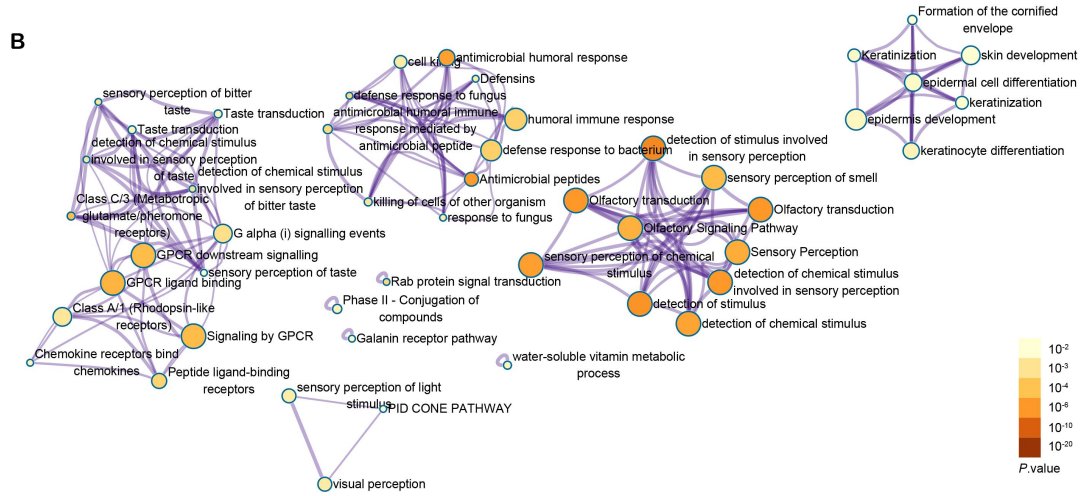


Figure 3. Functional analysis of 425 DEGs with down-regulation.

A. Metascape analysis of the top 19 significantly enriched clusters amid down-regulated DEGs.

B. Metascape analysis of the relationship between the top enriched clusters.

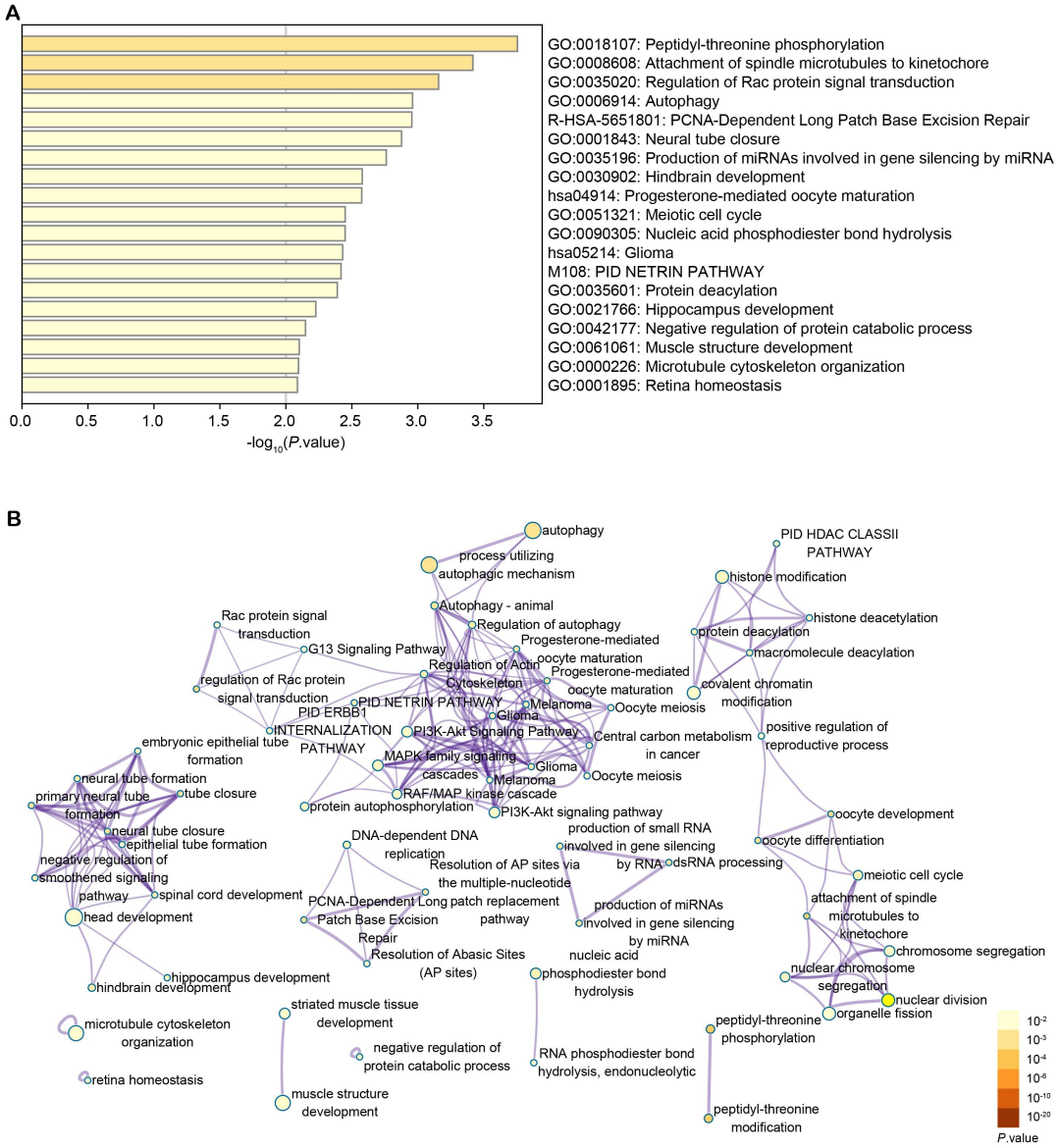


Figure 4. GSEA analysis of pathways related to OA based on GSE117999.

A. The gene sets of 116p11.2 distal deletion syndrome were largely enriched in patients with OA samples.

B. The gene sets of biomarkers for pyrimidine metabolism disorders were largely enriched in patients with OA samples.

C. The gene sets of pyrimidine metabolism and related diseases were greatly enriched in patients with OA samples.

D. The gene sets of base excision repair were greatly enriched in patients with OA samples.

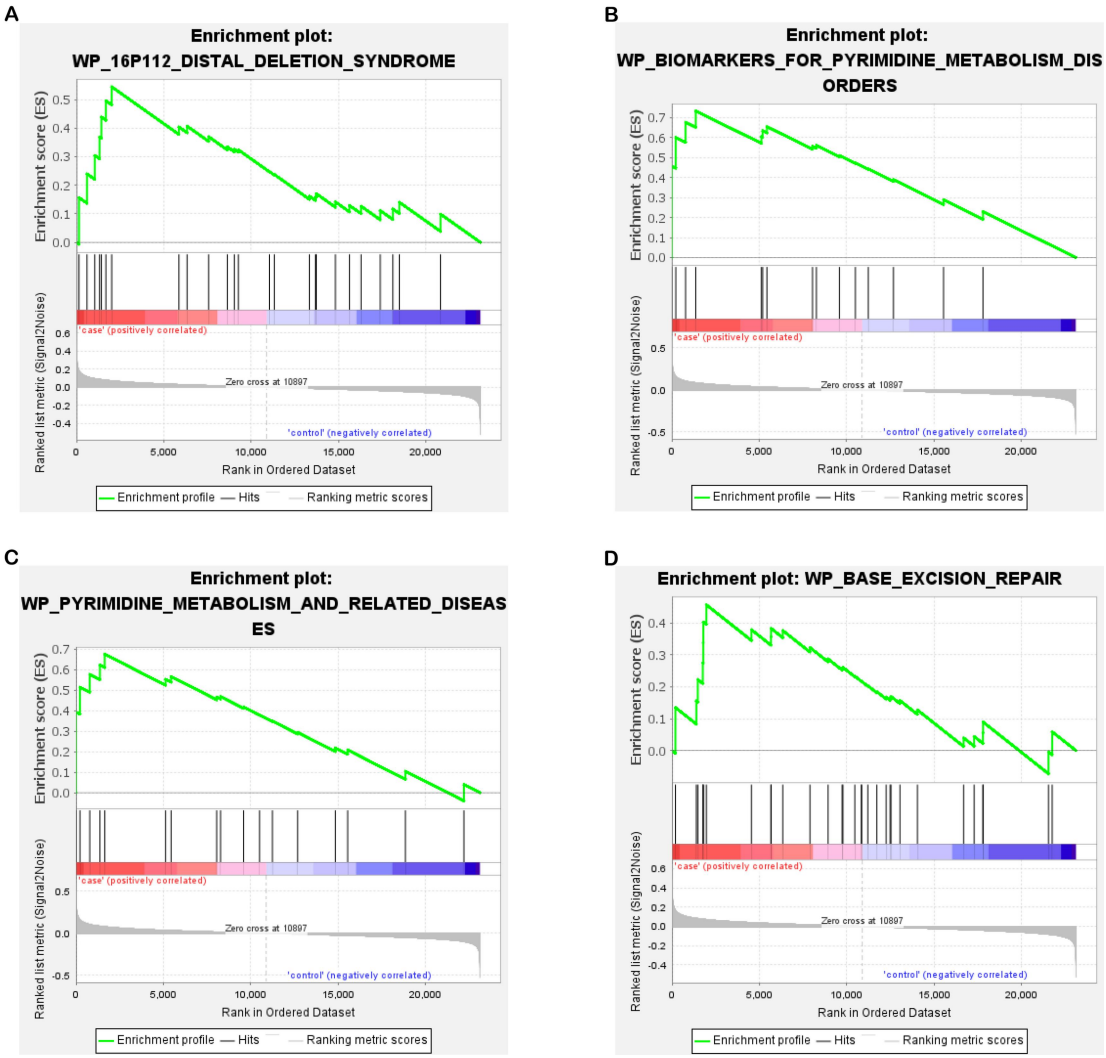


Figure 5. PPI network and MCODE analysis of the up-regulated DEGs. (A) PPI network analysis was conducted by Metascape. (B) MCODE method was used to select key protein modules from PPI. Colors indicated different MCODE clusters.

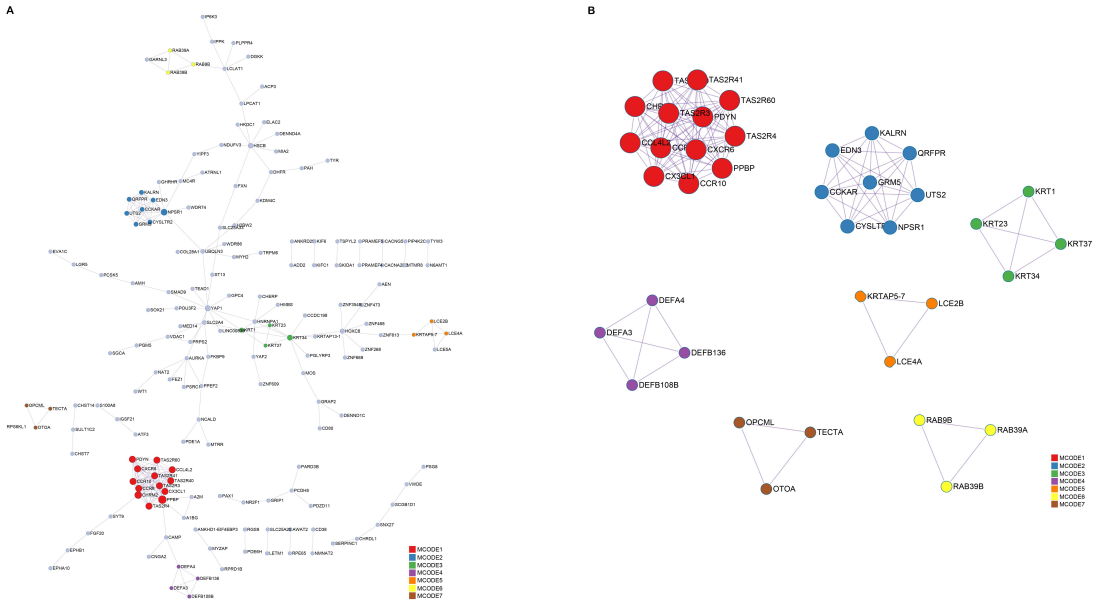


Figure 6. PPI network and MCODE analysis of the down-regulated DEGs. (A) PPI network analysis was determined using Metascape. (B) Key protein modules were selected from the PPI using the MCODE method. Colors represent the different MCODE clusters.

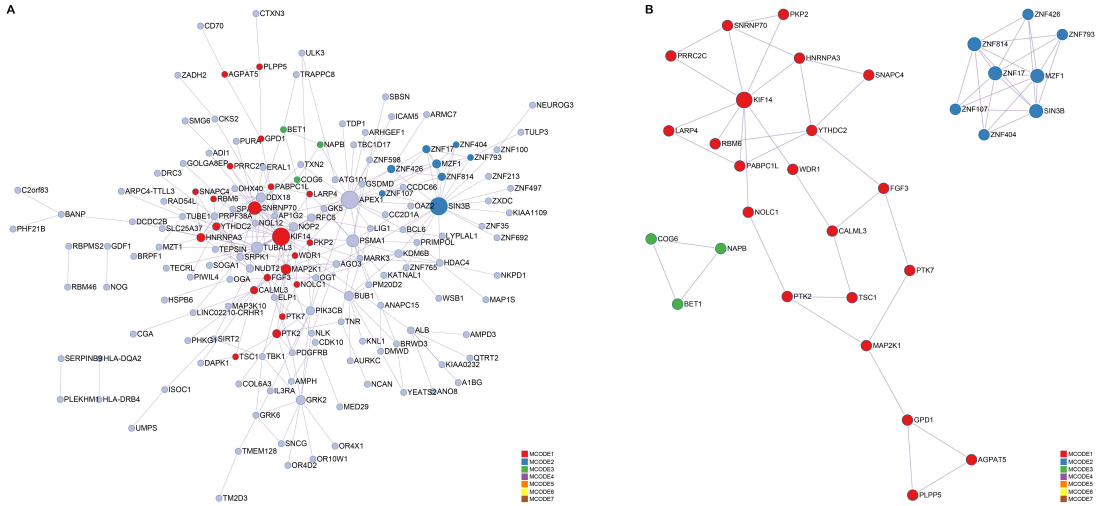
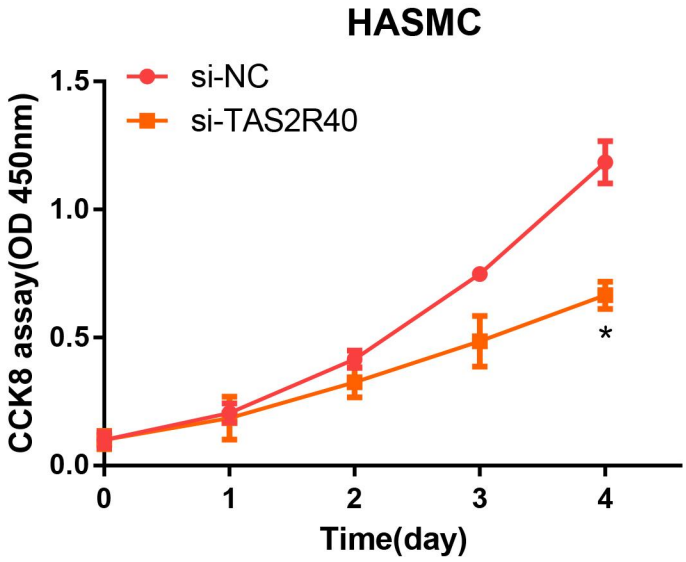


Figure 7. CCK-8 experimental results. The abscissa represents time, and the ordinate represents the OD value at 450nm wavelength.



Reference

1. Arden, N. and M.C. Nevitt, *Osteoarthritis: epidemiology.* Best Pract Res Clin Rheumatol, 2006. **20**(1): p. 3-25.
2. Spector, T.D. and A.J. MacGregor, *Risk factors for osteoarthritis: genetics.* Osteoarthritis Cartilage, 2004. **12 Suppl A**: p. S39-44.
3. Loeser, R.F., *Age-related changes in the musculoskeletal system and the development of osteoarthritis.* Clin Geriatr Med, 2010. **26**(3): p. 371-86.
4. Calce, S.E., et al., *The relationship of age, activity, and body size on osteoarthritis in weight-bearing skeletal regions.* Int J Paleopathol, 2018. **22**: p. 45-53.
5. Hurley, M.V., et al., *Sensorimotor changes and functional performance in patients with knee osteoarthritis.* Ann Rheum Dis, 1997. **56**(11): p. 641-8.
6. Tiulpin, A., et al., *Multimodal Machine Learning-based Knee Osteoarthritis Progression Prediction from Plain Radiographs and Clinical Data.* Sci Rep, 2019. **9**(1): p. 20038.
7. Felson, D.T., *Epidemiology of hip and knee osteoarthritis.* Epidemiol Rev, 1988. **10**: p. 1-28.
8. Georgiev, T. and A.K. Angelov, *Modifiable risk factors in knee osteoarthritis: treatment implications.* Rheumatol Int, 2019. **39**(7): p. 1145-1157.
9. Wang, Z., A. Lachmann, and A. Ma'ayan, *Mining data and metadata from the gene expression omnibus.* Biophys Rev, 2019. **11**(1): p. 103-110.
10. The Gene Ontology, C., *The Gene Ontology Resource: 20 years and still GOing strong.* Nucleic Acids Res, 2019. **47**(D1): p. D330-D338.

11. Zito, A., et al., *Gene Set Enrichment Analysis of Interaction Networks Weighted by Node Centrality.* Front Genet, 2021. **12**: p. 577623.

12. Parry, E.L., M.J. Thomas, and G. Peat, *Defining acute flares in knee osteoarthritis: a systematic review.* BMJ Open, 2018. **8**(7): p. e019804.

13. Gato-Calvo, L., et al., *Platelet-rich plasma in osteoarthritis treatment: review of current evidence.* Ther Adv Chronic Dis, 2019. **10**: p. 2040622319825567.

14. Chen, D., et al., *Osteoarthritis: toward a comprehensive understanding of pathological mechanism.* Bone Res, 2017. **5**: p. 16044.

15. Liu, M., et al., *Combining chemoinformatics with bioinformatics: in silico prediction of bacterial flavor-forming pathways by a chemical systems biology approach "reverse pathway engineering".* PLoS One, 2014. **9**(1): p. e84769.

16. Shenoy, N., *HIF1alpha is not a target of 14q deletion in clear cell renal cancer.* Sci Rep, 2020. **10**(1): p. 17642.

17. Zhu, L., et al., *Pan-Cancer Analysis of the Mitophagy-Related Protein PINK1 as a Biomarker for the Immunological and Prognostic Role.* Front Oncol, 2020. **10**: p. 569887.

18. Condello, M., et al., *Targeting Autophagy to Overcome Human Diseases.* Int J Mol Sci, 2019. **20**(3).

19. Miller, D.R. and A. Thorburn, *Autophagy and organelle homeostasis in cancer.* Dev Cell, 2021. **56**(7): p. 906-918.

20. Li, X., et al., *Curcumin Inhibits Apoptosis of Chondrocytes through Activation ERK1/2 Signaling Pathways Induced Autophagy.* Nutrients, 2017. **9**(4).

21. Tsukada, A., et al., *Unilateral-dominant reduction in muscle volume in female knee osteoarthritis patients: computed tomography-based analysis of bilateral sides.* J Orthop Surg Res, 2020. **15**(1): p. 543.

22. Di Pizio, A., M. Behrens, and D. Krautwurst, *Beyond the Flavour: The Potential Druggability of Chemosensory G Protein-Coupled Receptors.* Int J Mol Sci, 2019. **20**(6).

23. Di Bona, D., et al., *Taste receptor polymorphisms and longevity: a systematic review and meta-analysis.* Aging Clin Exp Res, 2020.

24. Wu, F., et al., *Spinocerebellar ataxia type 23 (SCA23): a review.* J Neurol, 2020.

25. Nagaya, D., et al., *An analysis of genetic association in opioid dependence susceptibility.* J Clin

Pharm Ther, 2018. **43**(1): p. 80-86.

26. Henriksson, R., et al., *PDYN, a gene implicated in brain/mental disorders, is targeted by REST in the adult human brain.* Biochim Biophys Acta, 2014. **1839**(11): p. 1226-32.

27. Li, M., et al., *KALRN mutations promote antitumor immunity and immunotherapy response in cancer.* J Immunother Cancer, 2020. **8**(2).

28. Jacobo-Baca, G., et al., *Proteomic profile of preeclampsia in the first trimester of pregnancy.* J Matern Fetal Neonatal Med, 2020: p. 1-7.

29. Latypova, X., et al., *Haploinsufficiency of the Sin3/HDAC corepressor complex member SIN3B causes a syndromic intellectual disability/autism spectrum disorder.* Am J Hum Genet, 2021. **108**(5): p. 929-941.

30. Wang, A., et al., *Nutraceuticals and osteoarthritis pain.* Pharmacol Ther, 2018. **187**: p. 167-179.