Screening of Osteoarthritis-Related Genes and Function Determination Based on Bioinformatics Tools

Zheng Fu
The First Affiliated Hospital of Chongqing Medical University
Weiqian Jiang
The First Affiliated Hospital of Chongqing Medical University
Wenlong Yan
The First Affiliated Hospital of Chongqing Medical University
Fei Xie
The First Affiliated Hospital of Chongqing Medical University
Yu Chen
The First Affiliated Hospital of Chongqing Medical University
Jian Zhang (zhangjiancqmu@sohu.com)
The First Affiliated Hospital of Chongqing Medical University https://orcid.org/0000-0003-2375-2507

Research

Keywords: Osteoarthritis, bioinformatics, differentially expressed gene, GO and KEGG pathway enrichment analysis, protein-protein interaction network, miRNA-mRNA regulatory network

Posted Date: December 6th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1095207/v1

Abstract

Background: Osteoarthritis(OA), commonly seen in the middle-aged and elderly population, imposes a heavy burden on patients from the clinical, humanistic and economic aspects. Our work aims at the discovery of early diagnostic and therapeutic targets for OA and new candidate biomarkers for experimental studies on OA via bioinformatics analysis.

Methods: The dataset GSE114007 was downloaded from GEO to identify differentially expressed genes(DEGs) in R using 3 different algorithms. Overlapping DEGs were subject to GO and KEGG pathway enrichment analysis and functional annotation. Following the identification of DEGs, a protein-protein interaction(PPI) network was established and imported into Cytoscape to screen for hubgenes. The expression of each hubgene was verified in two other datasets and create miRNA-mRNA regulatory networks.

Results:174 upregulated genes and 117 downregulated genes were identified among the overlapping DEGs. According to the results of GO enrichment analysis,MF enrichment was basically found in ECM degradation and collagen breakdown; enrichment was also present in the development, ossification, and differentiation of cells. The KEGG pathway enrichment analysis suggested significant enrichment in such pathways as PI3K-AKT, P53, TNF, and FoxO. 23 hubgenes were obtained from the PPI network, and 11 genes were identified as DEGs through verification. 8 genes were used for the establishment of miRNA-mRNA regulatory networks.

Conclusion:OA-related genes, proteins, pathways and miRNAs that were identified through bioinformatics analysis may provide a reference for the discovery of early diagnostic and therapeutic targets for OA, as well as candidate biomarkers for experimental studies on OA.

Background

Osteoarthritis (OA) is a chronic disease characterized by cartilage degradation, bone remodeling, osteophyte formation, joint inflammation and loss of normal joint function.[1] OA often occurs in middle-aged and elderly patients, especially in women above 55 years old and men over the age of 65, with the prevalence of OA in women substantially higher when compared with men.[2] Between 1990 and 2019, the global burden of OA had increased by 48%, and over 500 million of the world's population are currently affected by the disease.[3] As a major cause of disability among the elderly, OA entails a great clinical, humanistic and economic burden, which is complicated by the continuous increase in the aging and obese populations.[3,4]

Despite the extensive use of imaging methods, the diagnosis of OA fundamentally depends on clinical techniques.[5] A step-by-step treatment strategy has been developed to treat OA patients in different conditions.[6] However, conservative treatment demonstrates little efficacy in OA patients; for those with advanced OA, surgical treatment is considered only when the minimally invasive approaches yield

unsatisfactory outcomes.[5,7] All this unravels the profound significance and excellent prospect of research and studies on the prevention and early diagnosis and treatment of OA.[5] The extraordinary growth of bioinformatics enables big data analysis of genes and proteins to facilitate the identification of potential diagnostic and therapeutic targets in oncology and other research fields.[8] However, bioinformatics tools are seldom used in OA studies.

The Gene Expression Omnibus (GEO) supported by the National Center for Biotechnology Information (NCBI) is a data repository that archives gene expression datasets, original series and platform records. In the recently updated high-throughput sequencing mRNA expression profile dataset GSE114007,[9] 18 normal articular cartilage samples and 20 OA articular cartilage samples were subject to RNA sequencing. Based on this high-throughput sequencing dataset, different R packages were used for the identification of differentially expressed genes (DEGs) and overlapping DEGs in R. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed on the DEGs to elucidate their potential biological functions and relevant signaling pathways. Subsequently, a protein-protein interaction (PPI) network was set up to visualize the interactions between proteins and identify hubgenes using MCODE, a plugin for Cytoscape. Following that, the latest microarray datasets GSE55235[10]and GSE169077 were downloaded from the GEO database to verify each key DEG. Finally, target genes obtained from the verification process were utilized to establish miRNA-mRNA regulatory networks. In this study, the aim was to identify early diagnostic and therapeutic targets for OA and new candidate biomarkers for experimental studies on OA by exploring updated biological information associated with the development of OA using bioinformatics tools.

Materials And Methods

Data Downloading

The dataset GSE114007 was selected and downloaded from the NCBI GEO database (https://www.ncbi.nlm.nih.gov/gds/) to obtain data of knee-joint articular cartilage samples, including 18 normal cartilage samples and 20 OA cartilage samples from two high-throughput sequencing platforms, namely GPL11154 Illumina HiSeq 2000 (Homo sapiens) and GPL18573 Illumina NextSeq 500 (Homo sapiens).

Two other datasets, including GSE55235 and GSE169077 based on the GPL96 [HG-U133A] Affymetrix Human Genome U133A Array, were selected and downloaded from the GEO database. Data in GSE55235 were obtained from the synovial tissue, including 10 normal samples and 10 OA samples. Data in GSE169077 were derived from OA patients who underwent knee joint replacement, including 5 normal cartilage samples and 6 OA cartilage samples.

Data analysis and visualization were conducted in R (R ×64 4.0.3, https://www.r-project.org/).

Identification of DEGs

In R, DESeq2,[11] EdgeR, and Limma packages are usually used for high-throughput sequencing data analysis. In this study, these three packages were loaded respectively to identify DEGs between the normal and OA samples in GSE114007. DESeq2: $\log_2$FC =2, adjusted P value =0.01; EdgeR: logFC =2, P value =0.01; Limma: $\log_2$FC =2, adjusted P value =0.01. The pheatmap and ggplot2 R packages were utilized to draw three heatmaps and three volcano plots of DEGs, respectively. Following that, the DESeq2, EdgeR, and Limma methods were used to obtain overlapping DEGs from the upregulated and downregulated DEGs and depict Venn diagrams of these DEGs.

GO and KEGG Enrichment

The GO database is set up by the Gene Onotology Consortium to define and describe the functions of genes and proteins in terms of three ontologies including biological process (BP), cell component (CC) and molecular function (MF).[12] KEGG is an extensive database resource intended for understanding high-level functions and utilities of biological systems such as cells, organisms and ecosystems from information at the molecular level, particularly large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies.[13] R developers have developed a number of R packages for GO and KEGG enrichment analysis. In R, the org.Hs.eg.db package was used for ID conversion of the overlapping DEGs, the clusterProfiler package for GO and KEGG enrichment analysis[14] and the ggplot2 for visualization of the analytical results. A p-value less than 0.05 was defined as the threshold for statistical significance.

PPI Network Integration and Identification of Hubgene

The STRING database (https://string-db.org/cgi/input.pl) supports collection, scoring and integration of all public sources of PPI information. Moreover, it is able to complement such sources with computational predictions to clarify the relationships between proteins, including direct (physical) and indirect (functional) interactions.[15] DEGs were imported into the STRING database to generate a PPI network with a confidence cutoff score of 0.7 (high confidence).

Cytoscape 3.8.2 (https://cytoscape.org/) is a free, open-source bioinformatics application program designed for the visualization of molecular interaction networks integrated with gene expression profiles, phenotypes, and other molecular state data.[16] The PPI network data were imported into the software to depict the PPI network. MCODE is a Cytoscape plugin that detects highly interconnected regions in a PPI network, analyzes interactions between DEG-coded proteins and identifies hubgenes depending on given cluster-finding parameters. In this study,[17] MCODE was used for the discovery of highly interconnected modules in the PPI network within the following parameters: MCODE scores >5, degree cut-off =2, node score cut-off =0.2, k-score =2 and Max depth =100.

The hubgenes were verified by the microarray datasets GSE55235 and GSE169077. Gene ID conversion was conducted in R to obtain the expression profile data of the hubgenes and determine the corresponding $\log_2$ values. Further, the expression profile data of normal and OA samples in these

datasets were analyzed and demonstrated to meet the criteria for the Shapiro-Wilk normality test and the Levene's test. Therefore, such data were examined by the independent-samples t-test, and results were considered significant if $p < 0.05$. The ggplot2 R package was used to visualize the results.

Prediction of miRNAs

The miRWalk database[18] (http://mirwalk.umm.uni-heidelberg.de/) stores massive predicted data of miRNA-target interactions obtained with a machine learning algorithm, including miRNA-target interactions that are experimentally verified and available in miRTarBase, TargetScan, and miRDB. All verified hubgenes were imported into the miRWalk database as targets for miRNA prediction. For each GeneSymbol, the parameters were set as follows: min-p-value =0.95, position =3UTR, and experimentally verified evidence from at least one database (e.g., TargetScan, miRDB, miRTarBase). The selected miRNAs were defined as source nodes, and the corresponding target genes as target nodes. The miRNA-mRNA regulatory networks were visualized on Cytoscape.

Results

Identification of DEGs

A total of 479 DEGs, including 172 downregulated genes and 307 upregulated genes, were identified using the DESeq2 algorithm. See the volcano plots and heatmaps of DEG expression (**Fig.1**). A total of 600 DEGs, including 200 downregulated genes and 400 upregulated genes, were identified with the EdgeR method (**Fig.2**). A total of 366 DEGs, including 156 downregulated genes and 210 upregulated genes, were identified using the Limma algorithm (**Fig.3**). The DESeq2, EdgeR, and Limma methods yielded 174 upregulated and 117 downregulated overlapping genes. See the heatmap of overlapping DEG expression in normal and OA samples (Fig. 4)

GO and KEGG Enrichment

The GO and KEGG enrichment analysis suggested that the upregulated genes were mainly involved in the biological processes including: extracellular matrix organization (GO:0030198), ossification (GO:0001503), osteoblast differentiation (GO:0001649), cartilage development (GO:0051216), extracellular matrix disassembly (GO:0022617), collagen catabolic process (GO:0030574), replacement ossification (GO:0036075), and endochondral ossification (GO:0001958); In terms of KEGG pathways, the upregulated genes are predominantly enriched in the PI3K-Akt signaling pathway (hsa04151), focal adhesion (hsa04510), ECM-receptor interaction (hsa04512), the relaxin signaling pathway (hsa04926), leukocyte transendothelial migration (hsa04670), complement and coagulation cascades (hsa04610) and the P53 signaling pathway (hsa04115) (**Fig.5**). The relationships between the upregulated DEGs and different GO or KEGG terms are depicted by the string diagrams (**Fig.6**).

The GO and KEGG enrichment analysis suggested that the downregulated genes were primarily associated with such biological processes as responses to nutrient levels (GO:0031667), steroid hormone

(GO:0048545) and oxygen levels (GO:0070482), positive regulation of cellular catabolic process (GO:0031331), cellular response to external stimulus (GO:0071496), regulation of T cell activation (GO:0050863), response to glucocorticoid (GO:0051384), and cellular response to starvation (GO:0009267); in addition, the downregulated genes were largely involved in the following KEGG pathways: TNF signaling pathway (hsa04668), FoxO signaling pathway (hsa04068), osteoclast differentiation (hsa04380), HIF-1 signaling pathway (hsa04066), PPAR signaling pathway (hsa03320), prolactin signaling pathway (hsa04917) and fructose and mannose metabolism (hsa00051) (**Fig.7**). The relationships between the downregulated DEGs and different GO or KEGG terms are shown in the string diagrams (**Fig.8**).

PPI Network Integration and Identification of Hubgene

The PPI network consists of 263 nodes and 494 edges, with an average node degree of 3.76 and an average cluster coefficient of 0.379 (**Fig.9**); 23 hubgenes were identified via MCODE; KIF20A is the only seed node, and the rest including BUB1B, RRM2, BUB1, CDK1, CEP55, KIF23, TPX2, PBK, CENPF, ASPM, TOP2A, MELK, TTK, ZWINT, NUF2, ECT2, BIRC5, MKI67, DEPDC1, ANLN, HJURP, and KIF18B are all cluster nodes (**Fig.10**). MCODE scores of the hubgenes are visualized by a histogram (**Fig.11**).

The expression of hubgenes was verified using the GSE55235 and GSE169077 datasets. After ID conversion, the expression profiles of 21 genes were obtained from the GSE55235 and GSE169077 datasets, and the verification results suggested statistically significant differences in the expression of 11 out of these 23 genes (**Fig.12**)

miRNA-mRNA regulatory networks

With these 11 hubgenes as targets, 8 were enriched in miRNAs with potential regulatory properties using a miRWalk algorithm, which was supported by the validated data from at least one of the TargetScan, miRDB, and miRTarBase databases. Particularly, RRM2 and ZWINT were enriched in more miRNAs (up to 19 miRNAs) than other hubgenes; miR-6838-5p might have regulatory effects on CEP55 and CDK1 simultaneously; miRNA-mRNA interactions are depicted by the following networks (**Fig.13**).

Discussion

Osteoarthritis (OA) is a condition that involves movable joints and features cellular stress and ECM degradation attributed to micro- and macro-injury, which activates maladaptive repair responses including pro-inflammatory pathways of innate immunity. OA first occurs as a molecular derangement (metabolic abnormality in joint tissue) and subsequently as anatomic and/or physiologic derangements (characterized by cartilage degradation, bone remodeling, osteophyte formation, joint inflammation, and loss of normal joint function) and eventually results in illness.[1] Traditionally, it is perceived that cartilage is first afflicted by early OA. Under physiological stress, chondrocytes are responsible for maintaining the balance between cartilage matrix degradation and synthesis. When chondrocytes are overloaded by injury or stress, the rate of matrix degradation exceeds the rate of synthesis, which induces joint tissue

degradation and culminates in OA.[19] Researchers presume that OA is initiated by changes in the subchondral bone.[20] In OA, alterations in the subchondral bone mainly include increases in subchondral bone volume fraction, trabecular thickness and connectivity density of the subchondral bone due to loss of cartilage, which promotes bone remodeling and results in osteophyte formation.[21] OA also involves the innervated synovial tissue, a potential source of peripheral pain in OA; knee joints with painful OA are shown to largely enriched with gene-encoding neuronal proteins that promote neuronal survival under cellular stress, participate in calcium-dependent exocytosis in synaptic terminals and regulate GABA (γ-aminobutyric acid)-ergic activity.[22] Epidemiological studies prove that OA is an intricate disorder involving numerous factors.[23] In terms of OA diagnosis, clinical approaches remain the mainstay in spite of the extensive use of imaging techniques; as to OA treatment, conservative interventions are demonstrated to have limited efficacy; patients with advanced OA usually have no other options besides surgery. All this unravels the profound significance and excellent prospect of research and studies on the prevention and early diagnosis and treatment of OA.[5] The extraordinary growth of bioinformatics enables big data analysis of genes and proteins to facilitate the identification of potential diagnostic and therapeutic targets in oncology and other research fields.[8] However, bioinformatics tools are seldom used in OA studies. This study aimed to explore the latest biological information that emerges with the onset of OA and identify new candidate biomarkers by using bioinformatics tools for big data analysis of OA-related molecular expression, thereby offering reference to the discovery of potential targets for the early diagnosis and treatment of OA, as well as experimental molecules for OA studies.

Based on the high-throughput sequencing dataset obtained from the GEO database, three R packages, including DESeq2, EdgeR, and Limma, were utilized to identify DEGs, respectively. Eventually, 117 downregulated and 174 upregulated overlapping DEGs were identified. The GO analysis revealed that the DEGs were mostly enriched in two categories of biological processes: (1) ECM degradation and collagen breakdown; (2) development, ossification and differentiation of cells (chondrocytes) and cellular responses to hormones, nutrient levels, oxygen levels, external stimulus, and starvation.

ECM status changes, damage and degradation have been accepted as basic pathological changes in OA.[24] Cartilage does not contain blood vessels but a dense ECM with a sparse distribution of chondrocytes.[25] Accounting for 1-5% of the total cartilage tissue volume, chondrocytes provide powerful support for the synthesis of ECM proteins such as collagen, hyaluronic acid or glycoprotein and proteoglycan, with the collagen level contributing to approximately 60% of the dry weight of cartilage.[26,27] ECM degrading enzymes (metalloproteinases, i.e., ADAMTSs) and matrix metallo proteinases (MMPs) can induce ECM degradation, which is considered as a characteristic manifestation of OA.[28] Hyaline cartilage can be precursory or permanent. Healthy cartilage that exists in articular joints is defined as permanent cartilage, which consists of resting chondrocytes; under normal conditions, resting chondrocytes have a low proliferation rate and do not undergo terminal differentiation or endochondral ossification (EO).[29] However, in the presence of certain diseases, some articular chondrocytes are likely to lose their normal phenotypes and EO-like proliferation.[30] EO occurs when chondrocytes undergo active

proliferation and produce a cascade of cells, with some being enlarged and others growing into hypertrophic chondrocytes through hypertrophic changes. As these cells undergo a dramatic increase in their volumes, the surrounding is mineralized to develop bone tissue.[31] In the meantime, cartilage is hardened due to alteration in its elastic nature through calcification, making it increasingly difficult for chondrocytes to absorb nutrients. This leads to apoptosis of most chondrocytes and formation of small cavities within the tissue, leaving sufficient room for blood vessel invasion in the hardened structure and resulting in a shift from cartilage towards trabecular bone. Numerous studies have shown that the highlight events that occur in EO are also observed in OA, such as chondrocyte proliferation, hypertrophic differentiation of chondrocytes, cell death, calcification or mineralization, blood vessel invasion, and chondrocyte apoptosis.[28,29,30] Particularly, hypertrophic differentiation of chondrocytes is considered as one of the major pathological changes in OA.[32] Hypertrophy generally refers to an increase in the size and volume of cells. Hypertrophic chondrocytes start to express osteogenic differentiation-related genes and produce mineralized ECM proteins.[33,34] Besides, hypertrophic cells undergo decreases in hyaline cartilage markers, such as aggrecan, collagen type II, and SOX9. Despite the essential role of hypertrophic changes in chondrocytes in bone growth and development, this mechanism is a double-edged sword in the presence of disease.[35] Although chondrocyte clarification is known as the final stage of chondrocyte hypertrophy, it remains unclear whether hypertrophic chondrocytes are able to fully differentiate into bone cells. Despite all that, apoptosis of hypertrophic chondrocytes and lacunar emptying within OA cartilage are reported to give rise to the loss of articular cartilage and contribute to the generation of osteophytes by subchondral bone, that is, cartilage tissue is gradually replaced by bone tissue during the EO-like process in OA.[28, 29, 36] Additionally, chondrocyte senescence is suggested to play a crucial role in the pathological process of OA.[37] Cellular senescence may result from external stimuli or stress, oxidative stress, DNA damage, telomere shortening, oncogene activation and other factors.[38] It is found that mechanical stress stimuli increase oxidative stress to accelerate the progression of chondrocyte senescence in vitro.[39] Cellular responses to oxygen levels and external stimuli as shown in the GO analysis are consistent with these opinions.

The results of our KEGG enrichment analysis also support these opinions. The PI3K/AKT/mTOR signaling pathway has a complex structure known to involve over 150 proteins[40] and play an essential role in many cellular processes to maintain homeostasis, such as cell cycles and cell survival, inflammation, metabolism and apoptosis.[41] Many studies have shown that the PI3K/AKT/mTOR signaling pathway is strongly associated with cartilage degradation, synovial inflammation, and subchondral bone sclerosis in OA.[42] Chondrocytes are suggested to exhibit two different mechanisms of senescence: (1) replicative senescence, and (2) stress-induced premature senescence. Notably, the P53 signaling pathway that was shown to be enriched with upregulated genes in the KEGG enrichment analysis is thought to play a role in the replicative senescence and apoptosis of chondrocytes.[43] In addition to chondrocytes, synovioblasts are also suggested to accelerate the progression of OA through senescence. When exposed to H□O□ or TNF-α, synovioblasts exhibit increased senescence and promote mRNA expression and protein secretion in IL6, CXCL8, CCL2 and MMP3,[44] with the TNF signaling

pathway being included in the KEGG enrichment results. Likewise, the FoxO signaling pathway present in the KEGG enrichment results is considered as a critical signaling pathway in OA. FoxO1, FoxO3, FoxO4, and FoxO6 belong to the family of FoxO transcription factors that play a crucial role in the development and aging processes.[45] It is found in a study that in OA chondrocytes, inflammatory mediators and cartilage-degrading enzymes are reduced due to overexpression of FoxO1, which demonstrates the critical role of FoxOs in cartilage development, maturation and homeostasis as well as the protective properties of FoxOs against OA-associated cartilage damage.[46]

Following the analysis of DEGs based on the PPI network, MCODE was used for the identification of hubgenes, which yielded a total of 23 highly expressed hubgenes. These genes were ranked by MCODE score and verified by the other two datasets. Finally, 11 out of the 23 genes were found to exhibit differential expression in those datasets. The two datasets used for verification have a limited number of registered samples, which may bring a bias to this study and fail to reflect gene expression accurately. Particularly, because GSE169077 contains only 5 normal samples and 6 OA samples, the differences in gene expression were shown to lack statistical significance. Despite all that, we still spotted its value for analysis. KIF20A, also known as RAB6KIFL, is a kinesin family member involved in such cellular processes as mitosis, migration and intracellular transport and essential to cell division.[47] Currently, most KIF20A studies focus on oncology. Overexpression of KIF20A is found in melanoma, bladder cancer, and breast cancer;[48,49,50] migration and invasion of pancreatic cancer cells are inhibited by downregulating the expression of KIF20A.[51] Ribonucleotide reductase (RNR) is an enzyme that participates in cell cycles, which contains two subunits, including the regulatory subunit RRM1 and the catalytic subunit RRM2 involved in DNA replication and repair.[52] Aberrantly upregulated expression of RRM2 promotes rapid cell division through an increased accumulation of dNTPs,[53] which is closely associated with many types of cancer. However, the role of KIF20A or RRM2 has not been reported. Since chondrocytes have a relatively low proliferation rate, it is puzzling that KIF20A and RRM2, which have a strong association with cell division, are highly expressed in OA chondrocytes; further studies are needed to clarify the mechanisms of action of these genes in OA chondrocytes and investigate whether they are potential therapeutic targets for OA treatment. CDK1 is a member of the cyclin-dependent kinase (CDK) family. Increased CDK activity induced by the alteration of DNA damage and mitotic checkpoints is demonstrated to drive cell cycles.[54] If the activation of CDKs is out of control, this can lead to unexpected cell proliferation, as well as chromosomal and genomic instability.[55] It is currently believed that CDK1 can serve as a potential therapeutic target for cancer treatment, which achieves therapeutic goals by gaining control over cell proliferation in lung or breast cancer.[56,57] CEP55 (centrosomal protein 55 kDa) is dispensable for the whole cell cycle, especially in the stage of cytokinesis.[58] Moreover, the CEP55 gene is found to induce breast cancer cell death in vitro during mitosis and sensitize breast cancer cells to antimitotic agents; CEP55-dependent antimitotic treatment leads to early activation of CDK1/Cyclin B and thereby initiates cell death from the G2/M phase.[59] Meanwhile, through observation of the miRNA-mRNA regulatory networks, it was noted that miR-6838-5p might produce regulatory effects on CEP55 and CDK1 simultaneously. There are few published works in relation to miR-6838-5p and most are oncology studies,

in which miR-6838-5p is found to inhibit the Wnt pathway by targeting WNT3A to suppress cell proliferation and migration in triple-negative breast cancer.[60] Another similar report pointed out that by targeting GPRIN3 via the Wnt/β-Catenin signaling pathway, miR-6838-5p inhibited the malignant behaviors of gastric cancer cells, including cell growth, migration and invasion.[61] Since the effects of CDK1, CEP55 and miR-6838-5p on the cellular processes of chondrocytes such as cell cycle, mitosis, and death are unclear, it is worthwhile to conduct further studies and explore their value as therapeutic targets.

This study has some limitations: (1) the analysis of a single dataset may produce biased results in spite of the superior data quality and application of different algorithms; (2) The verification of some molecules is based on two other datasets with a smaller sample size, especially GSE169077, which only contains 5 normal samples and 6 OA samples. Therefore, it is clear that the verification results will bring a bias to the present study. (3) Because this study is carried out on the basis of an online database, the results should be verified by experiments to draw more accurate conclusions.

Conclusion

To sum up, this study used bioinformatics tools to analyze DEGs in OA and obtained 174 upregulated and 117 downregulated overlapping DEGs based on different algorithms. These DEGs were predominantly involved in such biological processes as ECM degradation, collagen breakdown, and cell development, ossification and differentiation. Hubgenes such as KIF20A, RRM2, CDK1 and CEP55 and regulatory miRNAs like miR-6838-5p were identified after an in-depth big data analysis. The valuable biological information regarding OA-related genes, proteins, pathways and regulatory miRNAs derived from the bioinformatics analysis may provide a reference for the discovery of early diagnostic and therapeutic targets for OA, as well as candidate biomarkers for experimental studies on OA.

Declarations

Ethics approval and consent to participate

This study was carried out abiding by the principles established by the Declaration of Helsinki and approved by the ethics committee of Chongqing Medical University.

Consent for publication

Not applicable.

Availability of data and material

The gene expression profifiling data are available at the Gene Expression Omnibus (GEO) database. The other data generated and analyzed during this study are included in the manuscript and the additional materials.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by Chongqing Technology Innovation and Application Development Special Project (Project No: CQYC202010).

Authors' contributions

ZF: Conceptualization, methodology, data curation, writing-original draft; WQJ: Conceptualization, Investigation, methodology; WLY, FX & YC: Data curation, resources; JZ: Project administration, validation, writing-review & editing. All authors read and approved the final manuscript.

Acknowledgements

We thank the School of Life Science, Chongqing Medical University.

Authors' information

1 Department of Orthopedics, the First affiliated Hospital of Chongqing Medical University, Chongqing 400016, China

2 Department of Orthopedics, Binzhou peoples' hospital, Shandong, China

3 Orthopedic laboratory of Chongqing Medical University, Chongqing, China

References

1. Kraus VB, Blanco FJ, Englund M, Karsdal MA, Lohmander LS. Call for standardized definitions of osteoarthritis and risk stratification for clinical trials and clinical use. Osteoarthritis and Cartilage. 2015;23(8).
2. Marita C, Emma S, Damian H, Sandra N, Ilana A, Marlene F, et al. The global burden of hip and knee osteoarthritis: estimates from the global burden of disease 2010 study. Annals of the rheumatic diseases. 2014;73(7).
3. J HD, Lyn M, Mabel C. Osteoarthritis in 2020 and beyond: a Lancet Commission. The Lancet. 2020;396(10264).
4. Zhao X, Shah D, Gandhi K, Wei W, Dwibedi N, Webster L, et al. Clinical, humanistic, and economic burden of osteoarthritis among noninstitutionalized adults in the United States. Osteoarthritis and Cartilage. 2019;27(11).
5. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. The Lancet. 2019;393(10182).

6. Emery CA, Whittaker JL, Mahmoudian A, Lohmander LS, Roos EM, Bennell KL, et al. Establishing outcome measures in early knee osteoarthritis. Nature Reviews Rheumatology. 2019;15(Suppl. 10).
7. Pinto D, Robertson MC, Hansen P, Abbott JH. Cost-Effectiveness of Nonpharmacologic, Nonsurgical Interventions for Hip and/or Knee Osteoarthritis: Systematic Review. Value in Health. 2012;15(1).
8. Zhanyu Y, Jiangdong N, Letian K, Yongquan G, Shibin T. Identification of genes and pathways associated with subchondral bone in osteoarthritis via bioinformatic analysis. Medicine. 2020;99(37).
9. Fisch KM, Gamini R, Alvarez-Garcia O, Akagi R, Saito M, Muramatsu Y, et al. Identification of transcription factors responsible for dysregulated networks in human osteoarthritis cartilage by global gene expression analysis. Osteoarthritis and Cartilage. 2018.
10. Dirk W, Rene H, Peter K, Dirk P, Michael P, Dominik D, et al. Identification of rheumatoid arthritis and osteoarthritis patients by transcriptome-based rule set generation. Arthritis research & therapy. 2014;16(2).
11. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15(12).
12. The Gene Ontology project in 2008. Narnia. 2008;36(suppl1).
13. M K, S G. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000;28(1).
14. Guangchuang Y, Li-Gen W, Yanyan H, Qing-Yu H. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics : a journal of integrative biology. 2012;16(5).
15. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Research. 2019;47(D1).
16. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Cold Spring Harbor Laboratory Press. 2003;13(11).
17. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BioMed Central. 2003;4(1).
18. Dweep H, Sticht C, Pandey P, Gretz N. miRWalk – Database: Prediction of possible miRNA binding sites by “walking” the genes of three genomes. Journal of Biomedical Informatics. 2011;44(5).
19. Cheng C, T TD, Linhong D, Liu Y. Biomechanical properties and mechanobiology of the articular chondrocyte. American journal of physiology Cell physiology. 2013;305(12).
20. Mansell JP, Collins C, Bailey AJ. Bone, not cartilage, should be the major focus in osteoarthritis. Nature Clinical Practice Rheumatology. 2007;3(6).
21. Chiba K, Uetani M, Kido Y, Ito M, Okazaki N, Taguchi K, et al. Osteoporotic changes of subchondral trabecular bone in osteoarthritis of the knee: a 3-T MRI study. Osteoporosis International. 2012;23(2).
22. Anna B-N, Francesc C-G, Mojca F-B, Sirisha A, F FS, Ralph S, et al. Pain-Associated Transcriptome Changes in Synovium of Knee Osteoarthritis Patients. Genes. 2018;9(7).

23. Neogi T, Zhang Y. Epidemiology of Osteoarthritis. Rheumatic Disease Clinics of North America. 2013;39(1).
24. Bruckner P. Suprastructures of extracellular matrices: paradigms of functions controlled by aggregates rather than molecules. Cell and Tissue Research. 2010;339(1).
25. Rim YA, Nam Y, Ju JH. Application of Cord Blood and Cord Blood-Derived Induced Pluripotent Stem Cells for Cartilage Regeneration. Cell Transplantation. 2019;28(5).
26. Goldring MB. Chondrogenesis, chondrocyte differentiation, and articular cartilage metabolism in health and osteoarthritis. Therapeutic Advances in Musculoskeletal Disease. 2012;4(4).
27. Akkiraju H, Nohe A. Role of Chondrocytes in Cartilage Formation, Progression of Osteoarthritis and Cartilage Regeneration. JDB. 2015;3(4).
28. Alice RY, Yoojun N, Hyeon JJ. The Role of Chondrocyte Hypertrophy and Senescence in Osteoarthritis Initiation and Progression. International journal of molecular sciences. 2020;21(7).
29. Kraan PMvd, Berg WBvd. Chondrocyte hypertrophy and osteoarthritis: role in initiation and progression of cartilage degeneration? Osteoarthritis and Cartilage. 2012;20(3).
30. Dreier R. Hypertrophic differentiation of chondrocytes in osteoarthritis: the developmental aspect of degenerative joint disorders. Arthritis Research & Therapy. 2010;12(5).
31. J ME, A AY, L T, K-S C, M M. Endochondral ossification: how cartilage is converted into bone in the developing skeleton. The international journal of biochemistry & cell biology. 2008;40(1).
32. Chengjie L, Xudong W, Xianjian Q, Zizhao W, Bo G, Lei L, et al. Collagen type II suppresses articular chondrocyte hypertrophy and osteoarthritis progression by promoting integrin β1-SMAD1 interaction. Bone research. 2019;7.
33. Findlay DM, Atkins GJ. Osteoblast-Chondrocyte Interactions in Osteoarthritis. Current Osteoporosis Reports. 2014;12(1).
34. Jianmei L, Shiwu D. The Signaling Pathways Involved in Chondrocyte Differentiation and Hypertrophic Differentiation. Stem cells international. 2016;2016.
35. Donkelaar CC, Wilson W. Mechanics of chondrocyte hypertrophy. Biomechanics and Modeling in Mechanobiology. 2012;11(5).
36. Edith C, Biserka R, Céline D, Olivier M, Sophie N, Julie C, et al. Insights on Molecular Mechanisms of Chondrocytes Death in Osteoarthritis. International journal of molecular sciences. 2016;17(12).
37. Kendal M, J LG, Singh RT. Cellular senescence in osteoarthritis pathology. Aging cell. 2017;16(2).
38. Claire V, Eduardo D, Jerome G, Beatriz C. Role of the Inflammation-Autophagy-Senescence Integrative Network in Osteoarthritis. Frontiers in physiology. 2018;9.
39. A MJ, Thomas B, Anneliese H, A BJ. Post-traumatic osteoarthritis: the role of accelerated chondrocyte senescence. Biorheology. 2004;41(3-4).
40. Jafari M, Ghadami E, Dadkhah T, Akhavan-Niaki H. PI3k/AKT signaling pathway: Erythropoiesis and beyond. Journal of Cellular Physiology. 2019;234(3).

41. Fengyuan T, Yuhua W, A HB, Curzio R, Gongda X. PKB/Akt-dependent regulation of inflammation in cancer. Seminars in cancer biology. 2018;48.
42. Sun K, Luo J, Guo J, Yao X, Jing X, Guo F. The PI3K/AKT/mTOR signaling pathway in osteoarthritis: a narrative review. Osteoarthritis Cartilage. 2020;28(4):400-9.
43. S A, B-H C, J-S K, J A, I H, H P, et al. Regulation of senescence associated signaling mechanisms in chondrocytes for cartilage tissue regeneration. Osteoarthritis and cartilage. 2016;24(2).
44. J DRM, Álvaro V, Alicia U, Sandra E, Eduardo M, Cristina M, et al. Senescent synovial fibroblasts accumulate prematurely in rheumatoid arthritis tissues and display an enhanced inflammatory phenotype. Immunity & ageing : I & A. 2019;16.
45. Kahn AJ. FOXO3 and related transcription factors in development, aging, and exceptional longevity. J Gerontol A Biol Sci Med Sci. 2015;70(4):421-5.
46. Matsuzaki T, Alvarez-Garcia O, Mokuda S, Nagira K, Olmer M, Gamini R, et al. FoxO transcription factors modulate autophagy and proteoglycan 4 in cartilage homeostasis and osteoar thritis. Sci Transl Med. 2018;10(428).
47. Neef R, Gruneberg U, Barr FA. Assay and functional properties of Rabkinesin-6/Rab6-KIFL/MKlp2 in cytokinesis. Methods Enzymol. 2005;403:618-28.
48. Junji Y, Satoshi F, Masatoshi J, Noritoshi H, Katunari M, Keisuke S, et al. Kinesin family member 20A is a novel melanoma-associated antigen. Acta dermato-venereologica. 2012;92(6).
49. Yan L, Pengyuan L, Weidong W, J GC, R TR, P MJ, et al. Cross-species comparison of orthologous gene expression in human bladder cancer and carcinogen-induced rodent models. American journal of translational research. 2010;3(1).
50. Khongkow P, Gomes AR, Gong C, Man EPS, Tsang JW-H, Zhao F, et al. Paclitaxel targets FOXM1 to regulate KIF20A in mitotic catastrophe and breast cancer paclitaxel resistance. Oncogene. 2016;35(8).
51. Daniela S, Mert E, Malte B, Thomas G, Christoph M, Susanne R, et al. Kif20a inhibition reduces migration and invasion of pancreatic cancer cells. The Journal of surgical research. 2015;197(1).
52. Pär N, Peter R. RIBONUCLEOTIDE REDUCTASES. Annual Review of Biochemistry. 2006;75.
53. Sun H, Yang B, Zhang H, Song J, Zhang Y, Xing J, et al. RRM2 is a potential prognostic biomarker with functional significance in glioma. Int J Biol Sci. 2019;15(3):533-43.
54. Lei D, Jiaqi C, Wen L, Hongjian C, Xianhui X, Hongshun A, et al. The Roles of Cyclin-Dependent Kinases in Cell-Cycle Progression and Therapeutic Strategies in Human Breast Cancer. International journal of molecular sciences. 2020;21(6).
55. Anke M, Eline M, De VK, Ken M, Karin VE, Elke DB. The therapeutic potential of cell cycle targeting in multiple myeloma. Oncotarget. 2017;8(52).
56. Li M, He F, Zhang Z, Xiang Z, Hu D. CDK1 serves as a potential prognostic biomarker and target for lung cancer. Journal of International Medical Research. 2020;48(2).

57. Nivrutti KH, YuChi W, Kumar YV, Prateeti S, Bashir L, Ntlotlang M, et al. In-Silico Evaluation of Genetic Alterations in Ovarian Carcinoma and Therapeutic Efficacy of NSC777201, as a Novel Multi-Target Agent for TTK, NEK2, and CDK1. International journal of molecular sciences. 2021;22(11).
58. Lee HH, Elia N, Ghirlando R, Lippincott-Schwartz J, Hurley JH. Midbody Targeting of the ESCRT Machinery by a Noncanonical Coiled Coil in CEP55. Science. 2008;322(5901).
59. Kalimutho M, Sinha D, Jeffery J, Nones K, Srihari S, Fernando WC, et al. CEP 55 is a determinant of cell fate during perturbed mitosis in breast cancer. EMBO Molecular Medicine. 2018;10(9).
60. Liu G, Wang P, Zhang H. MiR-6838-5p suppresses cell metastasis and the EMT process in triple-negative breast cancer by target ing WNT3A to inhibit the Wnt pathway. J Gene Med. 2019;21(12):e3129.
61. Zhou W, Ding X, Jin P, Li P. miR-6838-5p Affects Cell Growth, Migration, and Invasion by Targeting GPRIN3 via the Wnt/¦Â-Catenin Si gnaling Pathway in Gastric Cancer. Pathobiology. 2020;87(6):327-37.

Figures

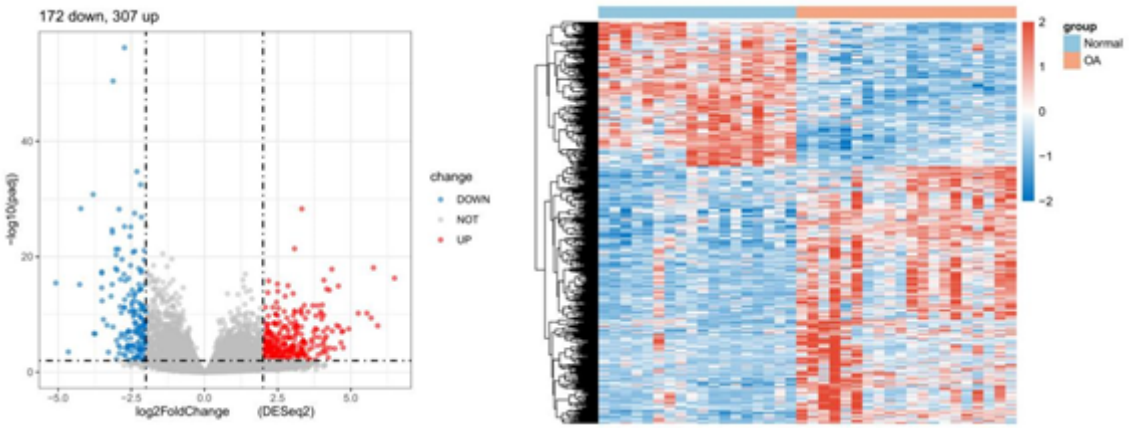

Figure 1

DESeq2: The volcano plot has log2FC on the x-axis and -log10(adjusted P value) on the y-axis; the heatmap has sample name on the x-axis and gene name on the y-axis; downregulated genes are indicated in blue and upregulated genes in red.

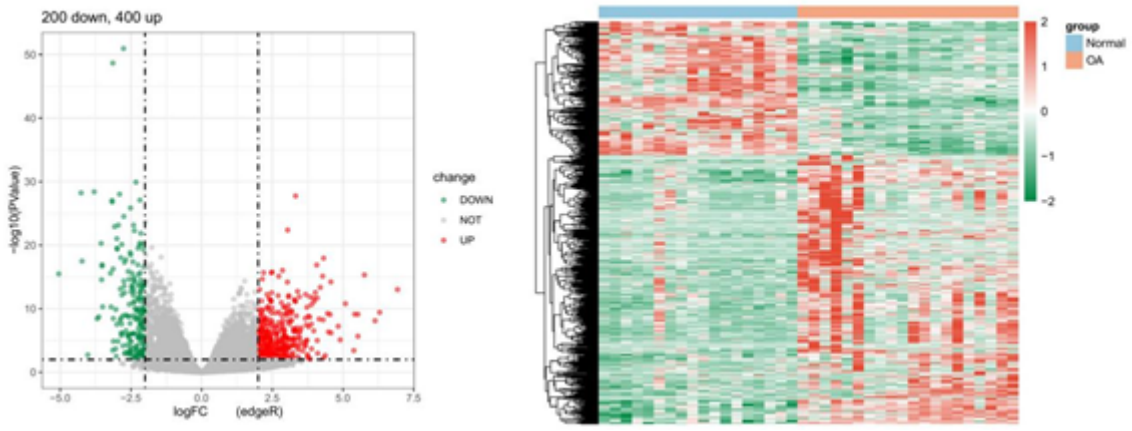

Figure 2

EdgeR: The volcano plot has log2FC on the x-axis and -log10(P value) on the y-axis; the heatmap has sample name on the x-axis and gene name on the y-axis; downregulated genes are indicated in green and upregulated genes in red.

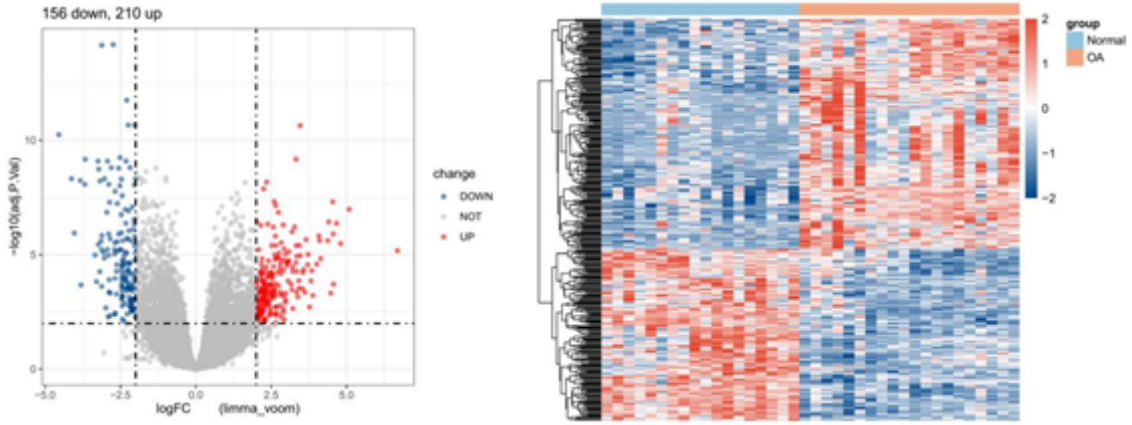

Figure 3

Limma: The volcano plot has log2FC on the x-axis and -log10(adjusted P value) on the y-axis; the heatmap has sample name on the x-axis and gene name on the y-axis; downregulated genes are indicated in dark blue and upregulated genes in red.

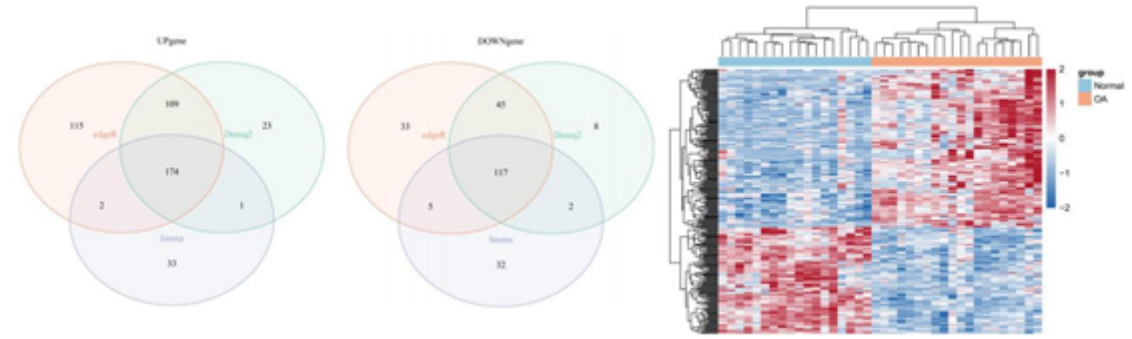

Figure 4

Venn diagrams: Overlapping DEGs yielded by the DESeq2, EdgeR, and Limma algorithms, including 174 upregulated genes and 117 downregulated genes. Heatmap of overlapping DEGs: Sample name is plotted on the x-axis, and gene name on the y-axis; downregulated genes are indicated in blue and upregulate genes in red.

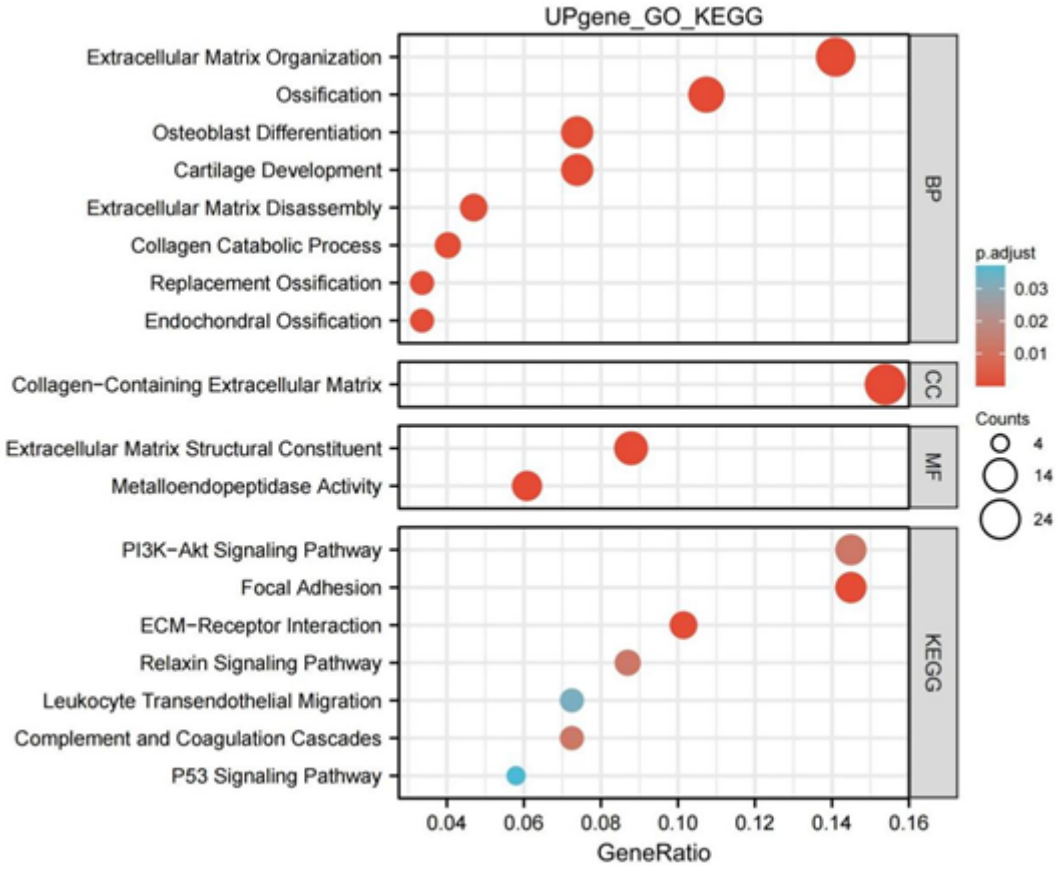

Figure 5

Bubble plot: GO and KEGG enrichment analysis of upregulated DEGs. The intensity of gradient color stands for the adjusted P value, and the size of bubble indicates the number of genes assigned to the term.

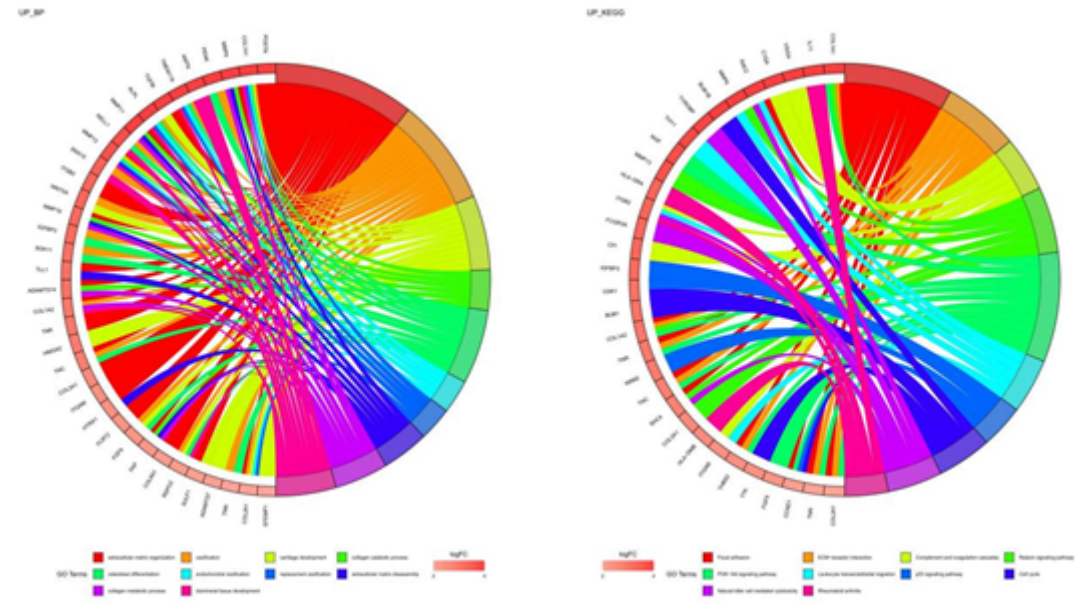

Figure 6

String diagrams: Top 10 enriched GO terms in the GO and KEGG enrichment analysis of upregulated DEGs, with biological processes on the left and KEGG pathways on the right. On the GO enrichment string diagram (left), the log2FC value of each gene is indicated in gradient color. On the KEGG enrichment string diagram (right), each color stands for a GO term.

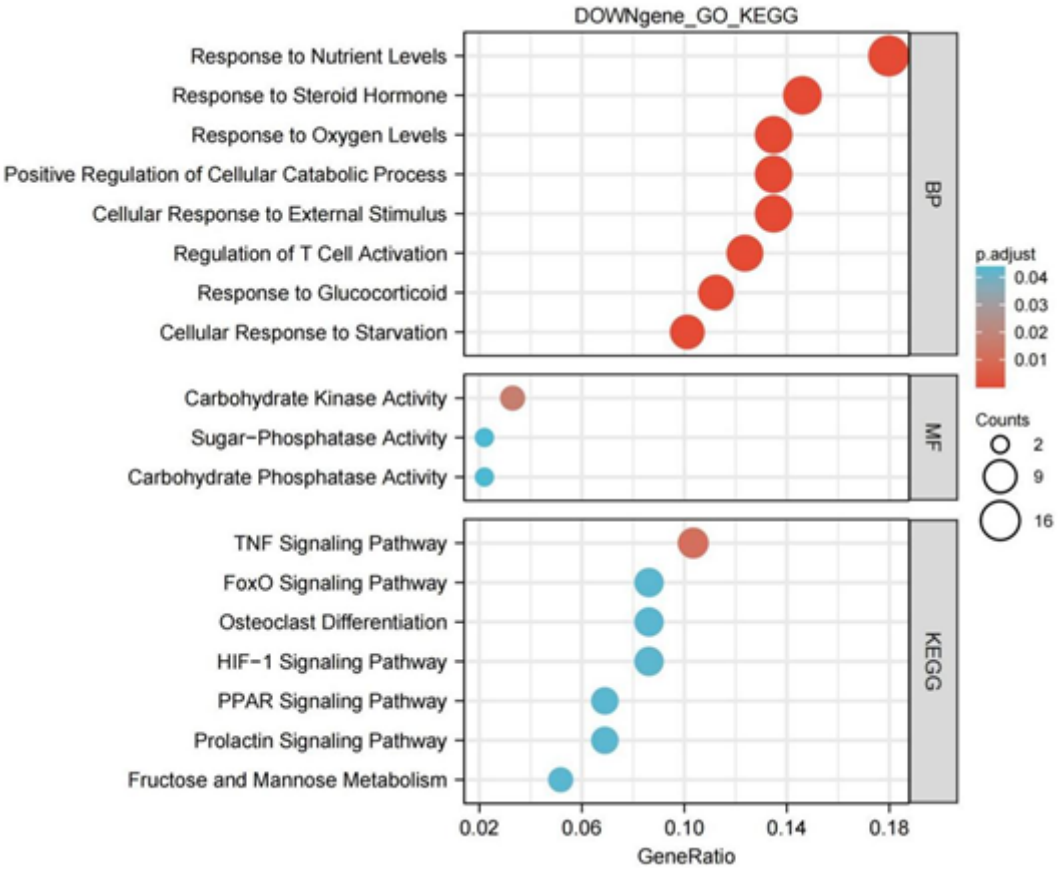

Figure 7

Bubble plot: GO and KEGG enrichment analysis of downregulated DEGs. The intensity of gradient color stands for the adjusted P value, and the size of bubble indicates the number of genes assigned to the term.

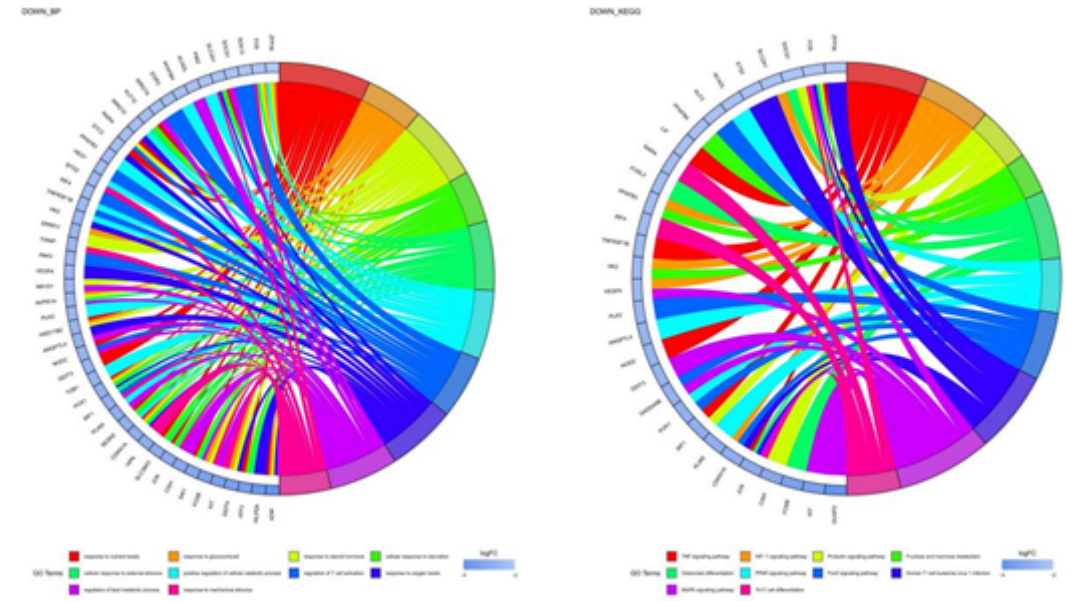

Figure 8

String diagrams: Top 10 enriched GO terms in the GO and KEGG enrichment analysis of downregulated DEGs, with biological processes on the left and KEGG pathways on the right. On the GO enrichment string diagram (left), the log2FC value of each gene is indicated in gradient color. On the KEGG enrichment string diagram (right), each color stands for a GO term.

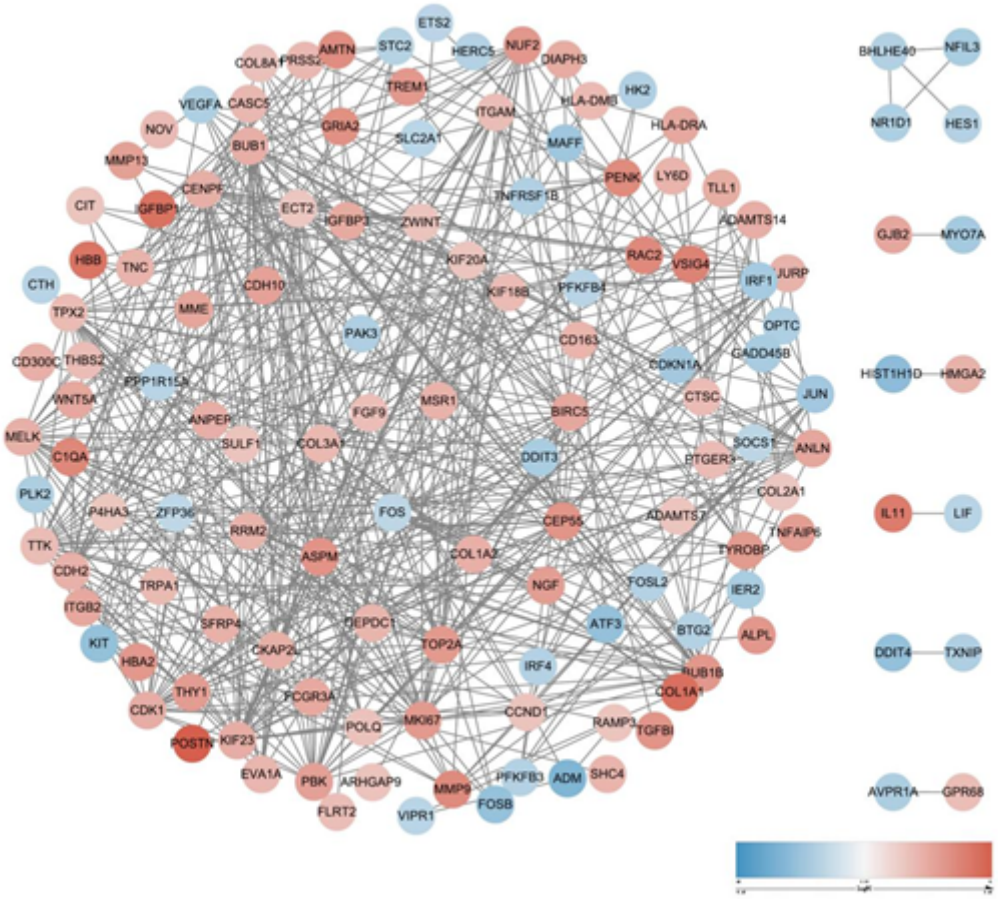

Figure 9

PPI network: The log2FC value (yielded by the DESeq2 algorithm) of each gene is indicated in gradient color.

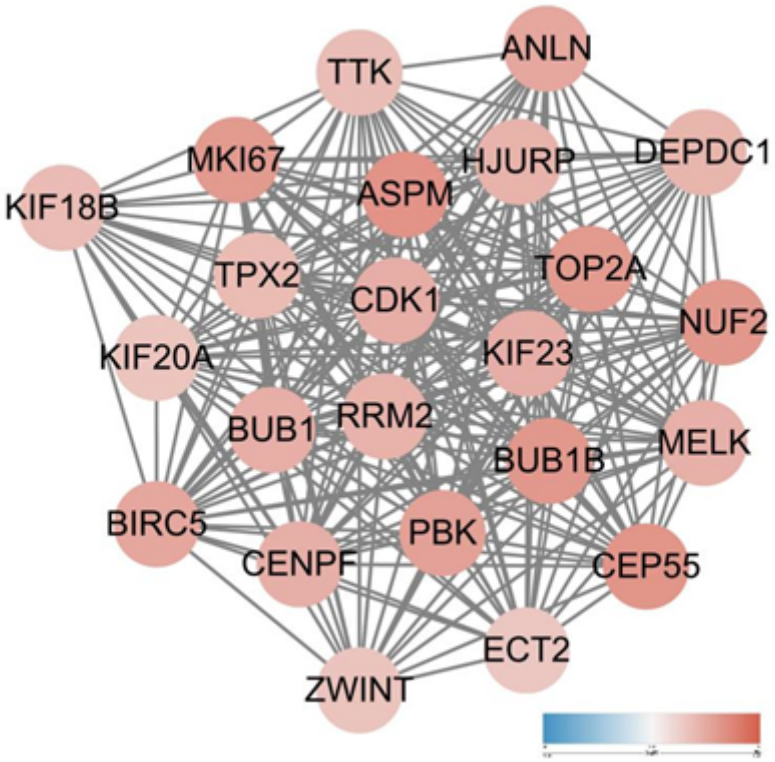

Figure 10

Hubgene network: The log2FC value (yielded by the DESeq2 algorithm) of each gene is indicated in gradient color.

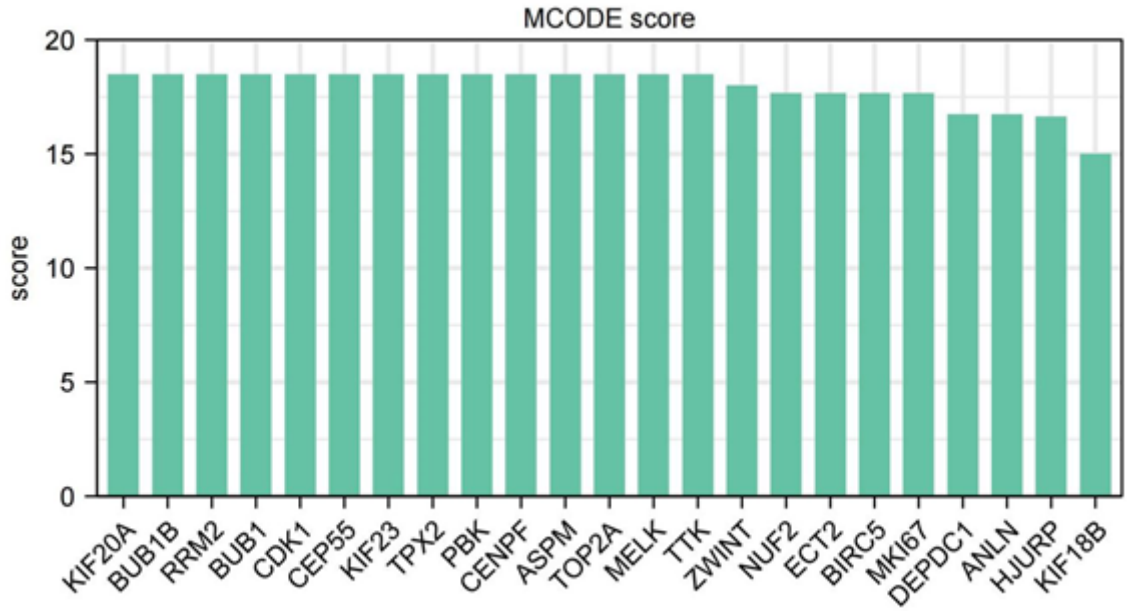

Figure 11

MCODE scores of hubgenes, with gene name on the x-axis and score on the y-axis.

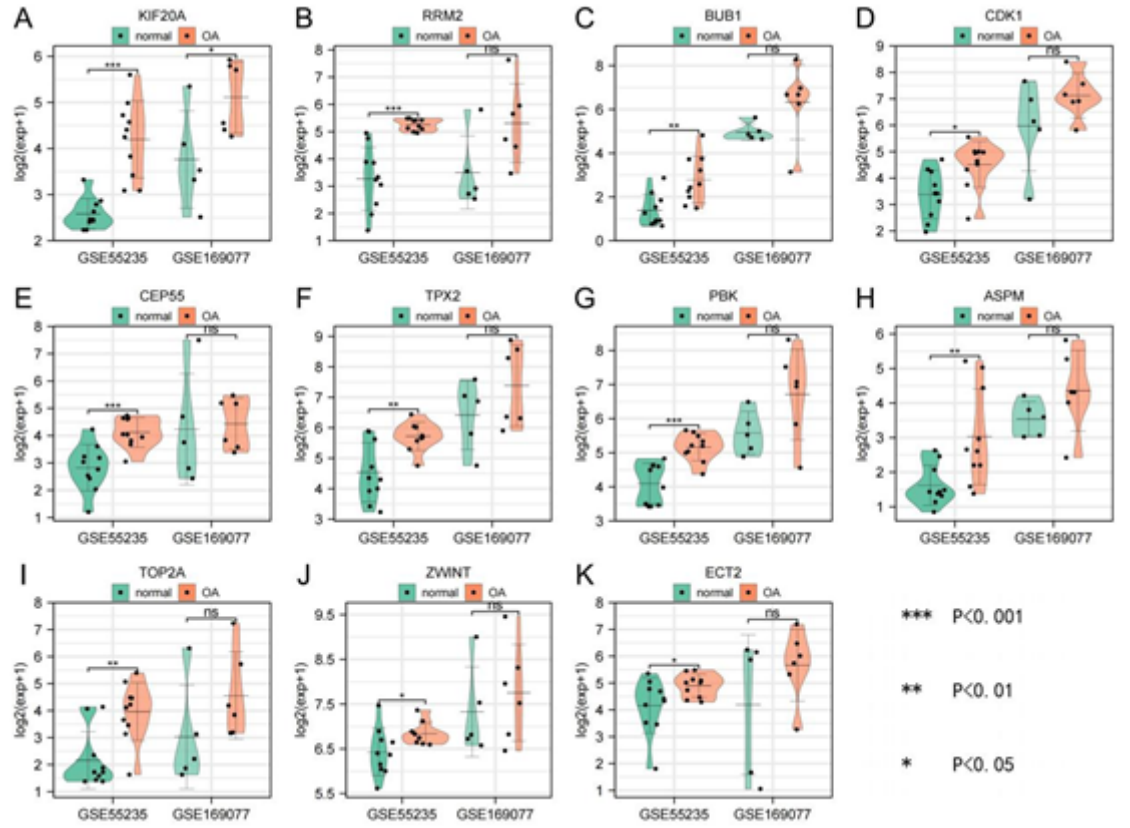

Figure 12

Violin plots: Hubgenes showing significant differences in expression levels in the GSE55235 and GSE169077 datasets. Normal samples are indicated in green and OA samples in orange.

Figure 13

miRNA-mRNA network: blue diamonds represent miRNAs, circles represent targets; the log2FC value (yielded by the DESeq2 algorithm) of each gene is indicated in gradient color; the size of circle standards for the value of MCODE score.