Chinese Medical Journal®

Identification of key genes underlying the effects of obesity on knee osteoarthritis

Siming Dai, Juan Zhang, Xiaoying Zhu, Yuxuan Lin, Ying Cui, Zhiyi Zhang, Zhiguo Lin

Department of Rheumatology, First Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang 150001, China.

To the Editor: Osteoarthritis (OA) is a disease associated with high morbidity and disability. Obesity, typically measured as a body mass index (BMI) ≥30 kg/m^2, is a well-known risk factor for the onset of OA and its related complications.[1] Interestingly, these two diseases may have common underlying mechanisms which contribute to the pathogenesis, such as genetic changes, and these mechanisms have not yet been fully explored. Gene expression profiling is a powerful tool used to systematically identify the molecular features underlying genetic variations associated with complex multifactorial diseases such as OA. In addition, it has been shown the significance of gene expression in understanding the mechanisms in the pathogenesis of obesity and OA.[2] This study hypothesized that hub genes could explain the relationship between obesity and OA. Consequently, identifying that these genes are involved could provide insights into the detailed regulatory mechanisms of obesity on OA.

The weighted gene co-expression network analysis (WGCNA) is a reliable method to screen significant modules highly related to clinical traits, and hub genes, which can provide more insights into the mechanism and progression of diseases.[3] In this research, we adopted the WGCNA, differential expression analysis, and the protein-protein interaction (PPI) approach to explore gene modules and critical hub genes which are highly associated with clinical traits of people with different BMI values using publicly available microarray datasets.

We first selected the GSE98460 dataset from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/), because it provides comprehensive and available baseline information on the BMI of participants enrolled.[4] The GSE98460 dataset contains 46 knee cartilage samples (23 each from the lateral and medial tibial plateaus, respectively) from 23 end-stage OA patients. Based on a variance analysis, the 7439 genes from 45 samples (top 50% of genes) were retained for further analysis [Supplementary Figure 1, http://links.lww.com/CM9/A705]. As shown in Figure 1A, GSM2596805, was an outlier and thus was excluded from subsequent analyses.

Next, a WGCNA co-expression network was constructed to extract significant modules that are highly related to clinical traits and hub genes. To ensure a scale-free network, an optimal soft-thresholding power of $\beta = 18$ (scale-free $R2 = 0.86$) was selected through the "pickSoftThreshold" function in the R package "WGCNA" [Figure 1B and 1C and Supplementary Figure 2, http://links.lww.com/CM9/A705]. Then, using the dynamic tree cutting and merged dynamic algorithm with a MEDissThres = 0.2, ten co-expression modules were detected from the 45 samples. The distinct modules are marked in different colors and are shown in the cluster dendrogram [Figure 1D]. The interaction relationships of the modules were analyzed by plotting a network heatmap [Supplementary Figure 3, http://links.lww.com/CM9/A705].

Next, the relationship between each module and BMI was assessed by correlating the eigengenes of each module with age, sex, and BMI [Supplementary Table 1, http://links.lww.com/CM9/A705]. It was found that three modules (light-green, salmon, and steel-blue) exhibited significant association with BMI values (absolute cor > 0.5, and $P < 0.05$) [Figure 1E]. This implied that 458 genes in these three modules are association between OA and obesity, therefore, these modules were identified as the key modules of interest for further analysis. To provide an interpretation of the biological functions and pathways underlying the impact of genes clustered in the three modules, Gene Ontology (GO) functional and Kyoto Encyclopedia for Genes and Genomes (KEGG) pathway enrichment analyses were performed on the 458 genes in these modules using the R package "clusterProfiler." In the GO analysis, the genes in these modules were significantly enriched for biological processes such as ossification, connective tissue

Access this article online

Quick Response Code:

Website: www.cmj.org

DOI: 10.1097/CM9.0000000000001670

Correspondence to: Zhiguo Lin, Department of Rheumatology, First Affiliated Hospital of Harbin Medical University, No. 23 Youzheng St., Harbin, Heilongjiang 150001, China
E-Mail: 1977linzhiguo@163.com

Chinese Medical Journal 2022;135(4)

Received: 04-03-2021; **Online:** 21-09-2021 **Edited by:** Lishao Guo

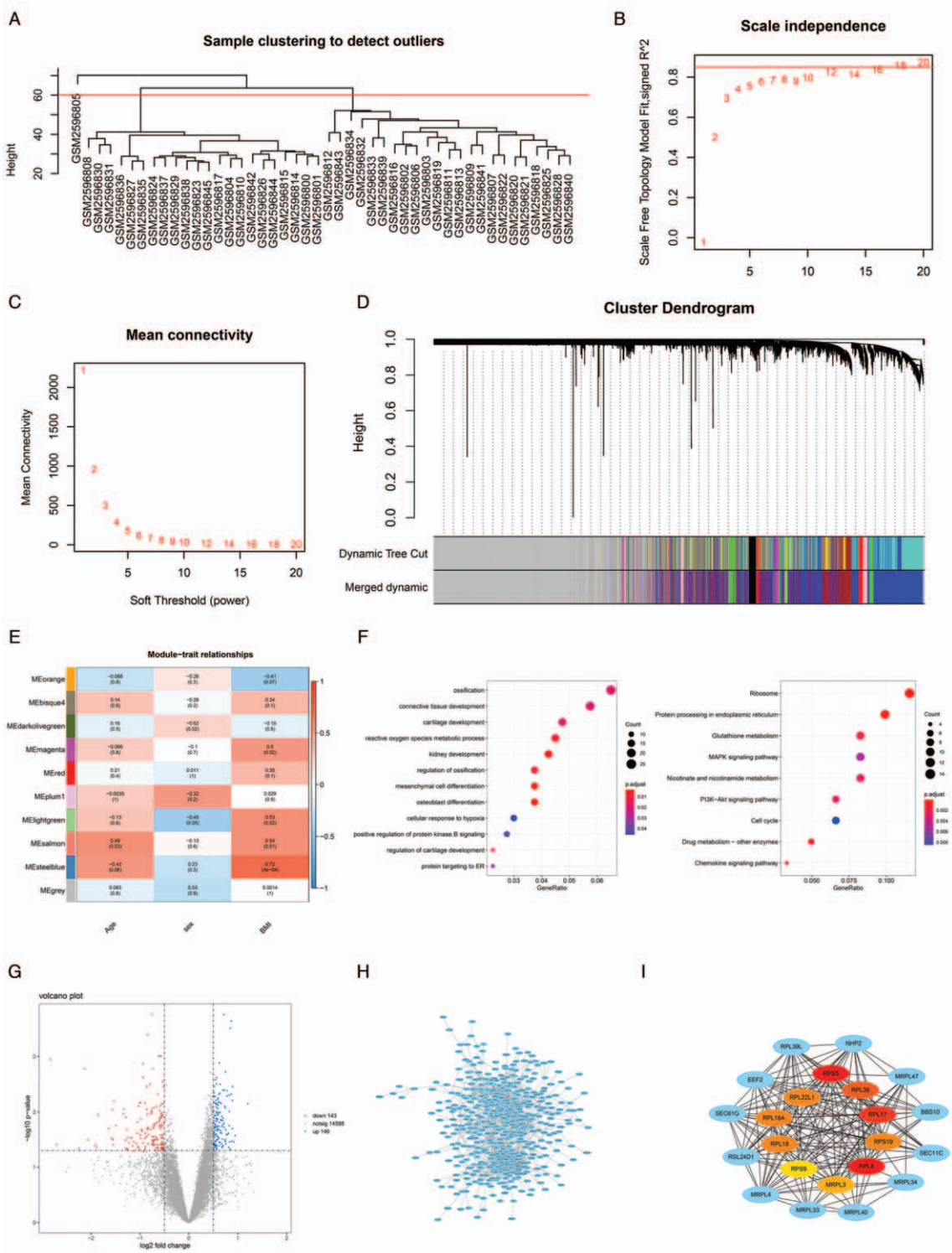

Figure 1: Identification of key genes underlying the effects of obesity on knee OA. (A) Clustering dendrogram of samples. The clustering was based on the expression data from the GSE98460 dataset. The top 50% of genes with the highest MAD values were used for WGCNA analysis. One outlier sample, GSM2596805 was excluded from subsequent analyses. (B) Analysis of the scale-free fit index for different soft-thresholding powers. (C) Analysis of the mean connectivity for different soft-threshold powers. (D) Cluster dendrogram and module assignment for modules from WGCNA. Genes were clustered based on dissimilarity measure (1-TOM). Bars below correspond to modules of genes with high interconnectivity. (E) Module-trait matrix was identified based on the correlation between clinical traits (age, sex, BMI) and ME. The number on the top of each cell illustrates the corresponding correlation coefficient and the number in brackets below, the P-value. High correlation coefficients are colored in red, and low in blue. (F) GO enrichment analysis (left panel) and KEGG pathway enrichment analysis (right panel) of 458 genes in the three hub modules. The color represents the adjusted *P* values (BH), and the size of the spots, the gene number. (G) Volcano plot of DEGs in the GSE98460 dataset. (H) PPI network of the DEGs and significant modules. The blue nodes represent the genes. Edges indicate interaction associations between nodes. (I) Identification of the key genes from the PPI network using the MCC algorithm. Edges represent the protein-protein associations, and red nodes, genes with a high MCC value. BMI: Body mass index; DEGs: Differentially expressed genes; GO: Gene Ontology; KEGG: Kyoto Encyclopedia for Genes and Genomes; OA: Osteoarthritis; PPI: Protein-protein interaction; WGCNA: Weighted gene co-expression network analysis.

development, cartilage development, reactive oxygen species metabolic process, and osteoblast differentiation [Figure 1F, left panel]. Whereas, as seen in the KEGG pathway analysis, the genes responsible for glutathione metabolism, MAPK signaling pathway, PI3K–Akt signaling pathway, and cell cycle were mostly enriched in the ribosome [Figure 1F, right panel].

Thereafter, the OA samples were subdivided into non-obese (BMI <30 kg/m^2) and obese (BMI ≥30 kg/m^2) groups. The R package "limma" was utilized to screen for the differentially expressed genes (DEGs) based on the cutoff criteria of false discovery rate (FDR) <0.05 and | logFC | ≥1. Finally, a total of 289 specific DEGs were detected, including 146 upregulated and 143 down-regulated genes [Figure 1G]. Also, the GO and KEGG enrichment analyses were utilized to understand the primary functions of the DEGs. As a result, GO terms analysis showed that the DEGs were categorized by their functions, mainly including negative regulation of cell migration, regulation of leukocyte migration, and negative regulation of cellular component movement [Supplementary Table 2 and Supplementary Figure 4A, http://links.lww.com/CM9/A705]. Additionally, the KEGG analysis revealed that the DEGs were involved in numerous pathways, including apoptosis, glutathione metabolism, and fatty acid metabolism [Supplementary Table 3 and Supplementary Figure 4B, http://links.lww.com/CM9/A705].

Lastly, for further identification of key genes, a total of 709 genes in the three significant modules and the DEGs described above were used to construct a PPI network [Figure 1H]. The network is based on the STRING online tool, which was designed for predicting PPI. Eventually, a PPI network consisting of 378 nodes (proteins) and 895 interactions (edges) was constructed, and 30 genes with a degree of >20 were selected as hub genes. Further, by analyzing the PPI network of hub genes using the MCC algorithm [Figure 1I and Supplementary Table 4, http://links.lww.com/CM9/A705], a total of ten genes were identified as key genes which might promote the pathogenesis of OA associated with obesity. These include *RPS5*, *RPL8*, *RPL17*, *RPL36*, *RPL18*, *RPL18A*, *RPL22L1*, *MRPL3*, *RPS19*, and *RPS9*. These genes code for ribosomal proteins relevant in ribosome biogenesis, a phenomenon which has not been reported in OA.

Notably, despite the underexplored role of the ribosome in OA pathology, the change in expression of ribosomal protein genes is highly suspected to be a compensatory response to stress stimuli such as obesity.[5] This study showed that the hub genes related to BMI are distributed individually among all three identified modules, and that the genes in these modules present significant enrichment of ossification, reactive oxygen species metabolic process, and osteoblast differentiation, all of which have been previously confirmed to promote the progression of OA.[6] Therefore, our findings suggest that the onset or progression of OA is partially due to obesity induced endochondral ossification, which leads to the degeneration of the cartilage. Additionally, we illustrate that this process is possibly regulated by genetic alterations of ribosomal protein.

In conclusion, we found that ribosomal protein genes may promote the onset and progression of OA induced by obesity. And we are currently carrying out several additional experiments to strengthen this hypothesis. So far, our findings provide a better understanding of the molecular processes which underlie obesity driven OA. Moreover, the key genes discussed may provide a novel research direction and potentially serve as a critical therapeutic strategy for OA.

Funding

This work was supported by grants from the National Natural Science Foundation of China (Nos. 81771749 and 81901638), Harbin Technology Bureau Research Fund (No. 2017RAQXJ196) and Scientific Research Innovation Foundation of the First Affiliated Hospital of Harbin Medical University (No. 2019B18).

Conflicts of interest

None.

References

1. Wang F, Shi L, Xue QY. Association of metabolic factors with symptomatic hand osteoarthritis in the Chinese Han population aged 40 years and above. Chin Med J 2016;129:2301–2307. doi: 10.4103/0366-6999.190660.
2. Ratneswaran A, Kapoor M. Osteoarthritis year in review: genetics, genomics, epigenetics. Osteoarthritis Cartilage 2021;29:151–160. doi: 10.1016/j.joca.2020.11.003.
3. Selvaraj G, Kaliamurthi S, Kaushik AC, Khan A, Wei YK, Cho WC, *et al*. Identification of target gene and prognostic evaluation for lung adenocarcinoma using gene expression meta-analysis, network analysis and neural network algorithms. J Biomed Inform 2018;86:120–134. doi: 10.1016/j.jbi.2018.09.004.
4. Rai MF, Sandell LJ, Barrack TN, Cai L, Tycksen ED, Tang SY, *et al*. A microarray study of articular cartilage in relation to obesity and severity of knee osteoarthritis. Cartilage 2020;11:458–472. doi: 10.1177/1947603518796122.
5. McDermott BT, Peffers MJ, McDonagh B, Tew SR. Translational regulation contributes to the secretory response of chondrocytic cells following exposure to interleukin-1β. J Biol Chem 2019;294:13027–13039. doi: 10.1074/jbc.RA118.006865.
6. Goldring SR, Goldring MB. Changes in the osteochondral unit during osteoarthritis: structure, function and cartilage-bone cross-talk. Nat Rev Rheumatol 2016;12:632–644. doi: 10.1038/nrrheum.2016.148.

How to cite this article: Dai S, Zhang J, Zhu X, Lin Y, Cui Y, Zhang Z, Lin Z. Identification of key genes underlying the effects of obesity on knee osteoarthritis. Chin Med J 2022;135:474–476. doi: 10.1097/CM9.0000000000001670