The Knee xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

The Knee

Candidate methylated genes in osteoarthritis explored by bioinformatics analysis

Jie Liu [a], Yangquan Hao [b,*], Yugui Wang [a], Shouye Hu [b], Ke Xu [b], Chao Lu [b]

[a] Shaanxi University of Chinese Medicine, Shiji Avenue, Xi'an-Xianyang New Economic Zone, Shaanxi 712046, PR China
[b] Department of Osteonecrosis and Joint Reconstruction, Honghui Hospital, Xi'an Jiao Tong University Health Science Center, 555 Youyi East Road, Xi'an, Shaanxi 710068, PR China

ARTICLE INFO

Article history:
Received 14 June 2016
Received in revised form 1 August 2016
Accepted 20 September 2016
Available online xxxx

Keywords:
Osteoarthritis
Methylation
Pathway
Protein–protein interaction

ABSTRACT

Background: This study aimed to explore potential novel genes correlated with osteoarthritis (OA).

Methods: The gene expression profile of GSE48422 was downloaded from the Gene Expression Omnibus (GEO) database. This dataset included five arthritic cartilage samples and five non-arthritic cartilage samples from five female OA patients. Differentially methylated genes (DMGs) between the two kinds of samples were identified, followed by their functional analysis and protein–protein interaction (PPI) analysis. Furthermore, the Comparative Toxicogenomics Database (CTD) was used to further identify OA-related genes among these DMGs.

Results: In total, 965 hypermethylated genes and 112 hypomethylated genes were identified in the arthritic cartilage samples. The hypermethylated genes (e.g., *ADCY4* and *ADCY6*) were significantly related to the calcium signaling pathway and gonadotropin-releasing hormone signaling pathway, while the hypomethylated genes were implicated in the mammalian target of rapamycin signaling pathway. In the PPI network, several genes had a higher degree, such as *ADCY4*, *ADCY6* and *GPR17*, and they interacted with each other. Additionally, 565 DMGs were predicted to be associated with OA, and five of them (e.g., *COMP* and *EDIL3*) were previously identified as OA markers.

Conclusions: The methylation of genes *ADCY4*, *ADCY6* and *GPR17*, as well as the gonadotropin-releasing hormone signaling pathway, was newly found to be potentially associated with OA. They may be novel OA markers.

1. Introduction

Osteoarthritis (OA), also known as degenerative arthritis, is a chronic disease of the joint characterized by a progressive degradation of subchondral bone [1]. Globally, approximately 250 million people (3.6% of the population) have osteoarthritis of the knee [2]. OA is believed to be caused by mechanical stress on the joint [3]. Because the pathogenesis of OA is complex and multifactorial with genetic, biological, and biomechanical components [4], its comprehensive pathogenesis is still unclear.

Articular cartilage plays an essential role in load transfer across the joint during the development of OA [5]. Due to changes in gene expression patterns within chondrocytes, OA joints are characterized by a dramatic increase in cartilage catabolism [6]. Cartilage damage in OA is probably the result of the aggregate effects of multiple genetic, environmental, mechanical, and cell

* Corresponding author.
E-mail address: haoyangqh@163.com (Y. Hao).

http://dx.doi.org/10.1016/j.knee.2016.09.020
0968-0160/

Please cite this article as: Liu J, et al, Candidate methylated genes in osteoarthritis explored by bioinformatics analysis, Knee (2016), http://dx.doi.org/10.1016/j.knee.2016.09.020

biological factors, which could cause changes in gene expressions [7]. Accumulating evidence suggests that epigenetics is closely associated with OA [8]. Genome-wide methylation screening has identified 1214 differentially methylated genes (DMGs) in cartilage between arthritic and non-arthritic knee cartilage, and these DMGs are predicted to be associated with the pathways of development and differentiation [6]. Another genome-wide DNA methylation study has identified a number of DMGs in OA patients, which have a great potential to be implicated in etiologic mechanisms of OA [9]. Furthermore, methylation of the promoter and 5′-untranslated regions (UTRs) of the OA-associated gene growth differentiation factor 5 (*GDF5*) was observed in joint tissues [10]. Despite the significant advances that have been achieved in the investigation of OA epigenetics, many epigenetic factors involved in OA remain undetected.

In the present study, a bioinformatics analysis was performed based on the gene expression profile of arthritic knee cartilage samples and non-arthritic knee cartilage samples deposited by Moazedi-Fuerst et al. [6]. The significant abnormal methylation genes were identified, and their related pathways and protein–protein interactions (PPIs) were investigated, followed by the key genes associated with OA exploration based on the Comparative Toxicogenomics Database (CTD). With these comprehensive analyses, we tried to explore more methylated genes that may be associated with OA, which may provide novel information for the understanding of the underlying mechanisms of OA progression and could be molecular targets for OA treatment.

2. Materials and methods

2.1. Data resources

Gene expression profile dataset GSE48422 [6] was downloaded from Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/). The data were produced by Agilent-014706 and Agilent-014707 Human Promoter ChIP-on-Chip Set 244K (Microarray combined G4489A) (Agilent Technologies, California, USA). The dataset included information from five paired arthritic knee cartilage samples (AC group) and non-arthritic knee cartilage samples (NAC group) from five different female OA patients. The NAC samples were taken from the healthy area between the femoral condyles, an area which is not affected by mechanical overloading in general; the AC samples were taken from an area of the same femur which was damaged, but where there was still enough cartilage present to collect a specimen [6].

2.2. Data preprocessing and differential expression analysis

The preprocessed gene expression profile data were downloaded. To explore the DMGs between the AC group and the NAC group, the *t*-test method in LIMMA (Linear Models for Microarray Data, http://www.bioconductor.org/packages/release/bioc/html/limma.html) package in R (v.3.0.0) (http://bioconductor.org/biocLite. R) [11] was utilized to calculate the significant *P*-value of DMGs. The *P*-value <0.05 and fold change (FC) ≥ 1.5 were selected as the cut-off criteria for DMG screening. Subsequently, the annotation files of methylation were downloaded, and the hypermethylated genes and hypomethylated genes were further identified based on the gene symbols (converted from probe IDs) of original gene expression profile data.

2.3. Gene ontology annotation and pathway analysis

The DAVID (Database for Annotation, Visualization and Integrated Discovery, http://david.abcc.ncifcrf.gov/) [12] is a gene functional classification tool that integrates various functional annotation tools for researchers to understand the biological meaning behind numerous genes. Gene Ontology (GO; http://www.Geneontology.org) function enrichment analysis and the Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.genome.jp/kegg/pathway.html) pathway enrichment analysis of hypermethylated genes and hypomethylated genes were performed based on DAVID [13]. GO functional categories included molecular function (MF), biological process (BP) and cellular component (CC). The *P*-value <0.05 and the count (the gene number) >2 were considered as cut-off values of significant GO and pathway terms.

2.4. PPI network construction

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING; http://string-db.org/) is a biological database and web resource of known and predicted PPIs [14]. According to information in the STRING database, proteins associated with DMGs (species: Homo) were selected with the criterion of combined score >0.7, and then the PPI network was constructed using Cytoscape software (http://cytoscape.org/) [15]. Subsequently, the modules of PPI network were further extracted using ClusterONE (http://www.paccanarolab.org/clusterone/) [16], followed by the pathway enrichment analysis. The $5E-4$ was considered as the threshold for significant module selection, and the ToppFun package in the ToppGene (https://toppgene.cchmc.org/) online database was used for the pathway enrichment analysis of genes in the network modules [17]. The Q-value calculated by the multiple correction method of <0.05 was considered as the cut-off criterion. In the network, a node represents a protein (gene), and lines represent the interactions of the proteins. The 'degree' of each node is equal to the number of nodes that interact with this node. The degree was used to evaluate the importance of the protein [18].

Please cite this article as: Liu J, et al, Candidate methylated genes in osteoarthritis explored by bioinformatics analysis, Knee (2016), http://dx.doi.org/10.1016/j.knee.2016.09.020