Identification of Ion Channel-Related Genes as Diagnostic Markers and Potential Therapeutic Targets for Osteoarthritis

Yongming Liu  
Shuguang Hospital, Shanghai University of Traditional Chinese Medicine

Xiong Yizhe  
Shuguang Hospital, Shanghai University of Traditional Chinese Medicine

Qian Zhikai  
The Second Affiliated Hospital of Soochow University

Wang Yupeng  
Shuguang Hospital, Shanghai University of Traditional Chinese Medicine

Wang Xiang  
Shuguang Hospital, Shanghai University of Traditional Chinese Medicine

Yin Mengyuan  
Tongji University School of Medicine

Du Guoqing (✉ 530427540@qq.com)  
Shuguang Hospital, Shanghai University of Traditional Chinese Medicine

Zhan Hongsheng  
Shuguang Hospital, Shanghai University of Traditional Chinese Medicine

Research Article

Keywords: Osteoarthritis, Transient receptor potential, Ion channel, Immune cell infiltration, Diagnostic Markers, bioinformatics

Posted Date: September 28th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3275709/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.
Abstract

Osteoarthritis (OA) is a debilitating joint disorder characterized by the progressive degeneration of articular cartilage. Although the role of ion channels in OA pathogenesis is increasingly recognized, diagnostic markers and targeted therapies remain limited. In this study, we analyzed the GSE48556 dataset to identify differentially expressed ion channel-related genes (DEGs) in OA and normal controls. We identified a total of 47 DEGs, with the majority involved in transient receptor potential (TRP) pathways. To select potential diagnostic markers, we employed machine learning algorithms, LASSO and SVM-RFE, and identified seven genes (CHRNA4, GABRE, HTR3B, KCNG2, KCNJ2, LRRC8C, and TRPM5) as the best characteristic genes for distinguishing OA from healthy samples. The differential expression of these seven marker genes was validated, and gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA) were performed to explore their involvement in biological pathways. We performed clustering analysis and identified two distinct subtypes of OA, C1 and C2, with differential gene expression and immune cell infiltration profiles. Using weighted gene co-expression network analysis (WGCNA), we identified three key genes (PPP1R3D, ZNF101, and LOC651309) associated with OA. We constructed a prediction model using these genes and validated it using the GSE46750 dataset, demonstrating reasonable accuracy and specificity. Our findings provide novel insights into the role of ion channel-related genes in OA pathogenesis and offer potential diagnostic markers and therapeutic targets for the treatment of OA.

1. Introduction

Osteoarthritis (OA) is a widespread degenerative joint disease characterized by cartilage degeneration, abnormal bone remodeling, osteophyte formation and joint inflammation[1]. That affects various structures of the joint, including the cartilage, bone, ligaments, and synovium, which often result in mobility loss, stiffness, and pain[2]. While OA commonly affects joints located in the hands, knees, hips, lower back, and neck, its symptoms are more common in people aged over 50 years. However, younger individuals, particularly those with prior joint injuries, can also develop this condition. Globally, OA affects approximately 1/10 of men and 1/50 of women over the age of 60, and it significantly impairs personal health, leading to a reduced quality of life and disability. Moreover, it imposes a substantial economic burden on society, with estimates ranging between 1.0% and 2.5% of the gross domestic product for Western countries. The present therapy protocol, which includes self-education and cyclooxygenase-2 inhibitors, mostly alleviates symptoms, and no disease-modifying OA medicine is available due to a lack of understanding of OA etiology[3–6]. The pathophysiology of OA involves the interaction of synovial tissue, bone, ligaments, menisci and synovial fluid around the articular cartilage, where sclerosis of the subchondral bone and synovial lesions occur[7–9]. However, the precise role of synovial lesions in OA pathogenesis remains poorly understood. Thus, this study aims to investigate the potential mechanisms and diagnostic markers of OA by examining synovial histopathology.

Pain is a prevalent and debilitating symptom of osteoarthritis (OA) that significantly impacts patients’ quality of life. The underlying causes of OA-associated pain can be broadly classified into two categories:
immunologically related factors and nociceptive sensitization[10]. Immunologically related factors are triggered by the degradation of joint tissues such as cartilage, synovium, and ligaments, leading to the release of pro-inflammatory cytokines and consequent activation of an inflammatory response, culminating in pain. Conversely, nociceptive sensitization occurs when biomechanical changes in the joint stimulate nociceptive receptors, triggering pain. In particular, ion channels such as transient receptor potential (TRP) play a pivotal role in nociceptive sensitization. TRP channels are non-selective cation channels located in the cell membrane that convert mechanical and temperature stimuli into electrical signals. In diseased joint tissue, chronic pain stimulation leads to TRP channel sensitization, eventually resulting in nociceptive sensitization.

The emergence of next generation sequencing (NGS) technology has opened up new avenues for exploring the molecular mechanisms of disease occurrence, progression, and regression. In the field of orthopedics, NGS has been employed in various studies to identify osteoarthritis (OA) biomarkers using bioinformatics methods[11, 12]. However, small sample sizes and single data analysis methods limit the effectiveness of such studies. Machine learning, a subset of artificial intelligence, has become increasingly popular in the medical field, particularly in cardiovascular research, where it has been applied to explore disease biomarkers, pathogenesis, therapeutic targets, predicting survival outcomes, and healthcare[13–15]. In recent years, integrated learning methods such as Random Forest (RF) and Support Vector Machines (SVM) have gained considerable attention from researchers due to their high accuracy and generalization capabilities[16]. RF is increasingly being used in cancer survivability prediction studies, exhibiting higher prediction accuracy than artificial neural networks, logistic regression, and other models[17, 18]. On the other hand, SVM creates a decision boundary between two categories based on margin calculation principles, minimizing classification error. The classification function of SVM has been extended to cancer genomics, enabling the discovery of new biomarkers, drug targets, and deeper insights into cancer-inducing genes in cancer genome classification or typing[19–21]. In this study, we compare RF and SVM models to select the best model for predicting the probability of multiple indicators or disease onset or progression based on individual patient characteristics. The nomogram model, based on multivariate analysis, integrates multiple predictors, demonstrating high accuracy and precision, and holds significant potential for joint diagnosis and disease prediction in clinical practice.

This study aimed to identify potential biomarkers and effective molecular targets for the diagnosis and treatment of OA. To this end, we retrieved eligible OA-related microarrays from the Gene Expression Omnibus (GEO) database and quantified ion channel-related genes based on the processed gene expression matrix. We then established a prediction model through RF, SVM and nomogram modeling, based on three candidate ion channel regulators (PPP1R3D, ZNF101, and LOC651309) to predict OA susceptibility, which effectively determined the prognosis of OA patients[22]. Additionally, we employed the consensus clustering algorithm and principal component analysis (PCA) to reveal two different patterns of TRP regulation. Subsequently, we performed single-sample gene-set enrichment analysis (ssGSEA) and immune infiltration analysis on the TRP genotyping results to investigate the pathogenesis of OA and explore potential biomarkers for clinical decision-making. The ssGSEA and immune infiltration analysis not only confirmed the pivotal role of ion channels in OA pathogenesis but also provided new.
insights into tissue engineering techniques for the treatment of OA. Our findings have significant implications for improving the diagnosis and treatment of OA and could pave the way for the development of novel therapeutic interventions.

2. Methods

The overall process is shown in Fig. 1.

2.1 Data source

The study employed publicly available osteoarthritis expression data from the GEO database, the GSE48556 and GSE46750 datasets. The former dataset comprised 106 peripheral blood samples collected from osteoarthritis patients and 33 samples from healthy controls, while the latter dataset contained 12 samples of synovial inflammation in osteoarthritis patients and 12 samples from unaffected areas. A total of 360 genes associated with ion channels were identified through a comprehensive literature review and then crossed with 330 genes from the HUGO Gene Nomenclature Committee (HGNC) database (https://www.genenames.org/data/genegroup/#! /group/177), which resulted in the identification of 323 distinct genes associated with ion channels.

2.2 Pre-processing and differentially expressed genes (DEGs) analysis

The "limma" (3.54.2) R package was used to perform background correction and normalization on each dataset and further convert them into gene symbols by referencing the probe name in the probe annotation file in each dataset. Next, we used the GSE48556 dataset as the training set and the GSE46750 dataset as the validation set. Specifically, we extracted the expression matrix of ion channel-related genes from the GSE48556 dataset and conducted differential expression analysis using the "limma" R package. DEGs were selected based on a threshold of P < 0.05, which allowed for the identification of the most robust and reliable genes associated with ion channels in osteoarthritis.

2.3 Correlation analysis

The "corrplot" (0.92) R package was used to investigate the relationship between 47 differentially expressed genes (DEGs) associated with ion channels in osteoarthritis. We utilized the "cor.mtest" function to conduct correlation analysis of the DEGs and employed the "corrplot" R package to perform visual enrichment analysis.

2.4 DEGs gene loci location

The "RCircos" (1.2.2) R package was used to locate the sites of differentially expressed ion channel related genes within 23 chromosomes.

2.5 KEGG and GO analysis
The "Cluster Profiler"(4.6.2) R package was used to perform KEGG (Kyoto Encyclopedia of Genes and Genomes) and GO (Gene Ontology) analysis on differentially expressed ion channel related genes, and "ggplot2" (3.4.1) and "RCircos" (1.2.2) R packages were used to visualize the results.

2.6 Screen the most relevant ion channel genes for osteoarthritis

The "glmnet" (4.1-7) R package was used to employ the LASSO (least absolute shrinkage and selection operator) algorithm to reduce data dimensions. Using DEGs as the selection feature, the LASSO algorithm was used to identify the most relevant ion channel genes in osteoarthritis. At the same time, the study used the SVM algorithm to build a support vector machine recursive feature estimation (SVM-RFE) model, and compare the average false positive rate with a 10-fold cross validation. The genes obtained through the two algorithms were considered to be the most relevant ion channel genes for osteoarthritis. The receiver operating characteristic (ROC) curve was used to determine the area under the curve (AUC) value, sensitivity, specificity, and accuracy to assess whether the genetic markers we selected have diagnostic value.

2.7 Single sample gene set enrichment analysis

The "GSEA" (4.1.0) R package was used to perform SsGSEA (Single sample gene set enrichment analysis). In order to explore the rich pathways of these seven marker genes, the enrichment results of marker genes were analyzed based on the GSE48556 dataset. In addition, this work treats the KEGG pathway set as a predefined set for detecting gene concentration enrichment levels. The specific enrichment results of each marker gene are integrated into a table (GSEA. result. txt).

2.8 Single sample gene set variation analysis

The "GSVA" (1.46.0) R package was used for single gene GSVA (Single sample gene set variation analysis). The study used the KEGG pathway set as the background gene set for different marker genes. Meanwhile, different GSVA scores of marker genes in high and low expression groups and different subtype groups were analyzed using the "limma" R package at the significance level of $|t|$. When $|t|>2$ and $p < 0.05$, the pathway was considered to be activated.

2.9 Immunologic Infiltration Analysis

The "GSVA" R package was used to calculate the enrichment fraction of each immune-related cell (22 species) through SsGSEA. The "ggplot2" package was used to draw a box graph to visualize the differences in immune cell infiltration between arthritis and healthy samples, as well as between different subtypes. Pearson correlation analysis observed the relationship between 7 characteristic genes and infiltrating immune cells. The "ggplot2" R package was used to visualize the analysis results.

2.10 Subcomponent typing

The "Conensus ClusterPlus" (1.62.0) R package was used to perform cluster analysis (1000 iterations and 80% resampling rate) on GSE48556 patients with osteoarthritis to evaluate the gene expression patterns
between each subtype and subdivide patients into two subgroups (subgroup 1 and subgroup 2) by using the optimal k-means clustering method. The "ggplot2" R package was used to visualize the differences in the expression pattern of 7 ion channel genes among different genotypes.

2.11 Principal Component Analysis (PCA)

The "limma" and "ggplot2" R packages was used to perform PCA analysis to explore the distribution of patients with different subtypes.

2.12 WGCNA network construction and module identification

The "WGCNA" (1.72-1) R package was used to construct a co-expression network between the osteoarthritis group and the healthy group, as well as the C1 and C2 subtype groups. First, cluster the samples to assess whether there are any significant outliers. Secondly, using the automatic network construction function to construct a co-expression network, The "pickSoftThreshold" function was used to calculate the soft threshold power β, and coexpression similarity was extracted to calculate adjacency relationships. Thirdly, hierarchical clustering and a dynamic tree cutting function detection module were used; Fourthly, calculate gene significance (GS) and module membership (MM) to correlate modules with clinical traits. Finally, extract the corresponding module gene information for further analysis and visualize the characteristic gene network. The "venn" (1.11) R package was used to visualize the key modules and obtain the intersection genes.

2.13 Construction and evaluation of RF, XGB, GLM, and SVM models

Create a random forest model (RF), extreme gradient enhancement model (XGB), support vector machine model (SVM), and a generalized linear model (GLM) based on the key genes obtained from the WGCNA intersection. Then, the "DALEX" (2.4.3) R package was used to analyze the interpretation characteristics of the above four models, draw residual distribution diagrams, and then obtain the best model based on the test set. Finally, we analyzed the significance of the variables and selected the most important explanatory variables for further research.

2.14 Construction and validation of a nomograph model for the diagnosis of OA.

The "rms" (6.5-0) R package was used to establish a nomograph model to predict the occurrence of osteoarthritis. "Score" refers to the score of the corresponding factor, and "Total Score" refers to the total score of all the above factors. The calibration curve was then used to evaluate the predictive power of the nomograph model. Finally, decision curve analysis and the clinical impact curve were used to evaluate the clinical value of the model. Finally, the impact of the risk model constructed from key genes on the prognosis of the dataset GSE46750 was validated using time dependent receiver operating characteristic curve.
2.15 Statistical Analysis

The R program (4.2.2) was used for statistical analysis in this study. Unless otherwise specified, a two-tailed p < 0.05 was considered statistically significant.

3. Results

3.1 Expression of ion channel-related genes and selection of DEGs

A total of 47 DEGs were identified in the GSE48556 dataset, including 16 upregulated and 31 downregulated genes (Figure 2A). and gene associations and locations were visualized in Figures 2B and 2C, respectively. To elucidate the biological functions and pathways of ion channel-related genes, we performed GO and KEGG functional annotation and reactome pathway analyses. Notably, most of the DEGs were found to be involved in ion channel (TRP) related pathways (Figures 2D and 2E).

3.2 Seven DEGs as genes for diagnosis of OA

To identify candidate diagnostic markers for OA, we employed two machine learning algorithms, LASSO and SVM-RFE, to screen for DEGs that significantly differentiated OA from healthy controls, based on the GSE48556 dataset. Penalty parameters were adjusted by 10-fold cross-validation in LASSO logistic regression, which selected 20 features associated with OA-related features (Figures 3A, B), and SVM-RFE was then used to filter 8 DEGs to identify the best combination of characteristic genes. Finally, this work identified 7 genes (minimum RMSE = 0.173, maximum accuracy = 0.827) as the best characterized genes (Figure 3C). Ultimately, seven genes (CHRNA4, GABRE, HTR3B, KCNG2, KCNJ2, LRRC8C, and TRPM5) were chosen as the best characteristic genes for distinguishing OA from healthy samples, with an AUC value of 0.930 based on the logistic regression model constructed by the "glm" R package (Figures 3D, E). Furthermore, we mapped the ROC curves of all seven marker genes to investigate whether a single gene can distinguish OA from healthy controls, and found that all seven genes had an AUC value of >0.6. These findings indicate that our logistic regression model is more accurate and specific in distinguishing OA from healthy samples than individual marker genes.

3.3 Differential expression of seven marker genes

We next evaluated the differential expression of the seven marker genes in the GSE48556 dataset. Our results revealed that CHRNA4 was upregulated (p = 0.00046), whereas GABRE (p = 0.021), HTR3B (p = 0.023), and TRPM5 (p = 0.054) were downregulated in the OA group. Additionally, KCNG2 (p = 0.0057), KCNJ2 (p = 1.4e-05), and LRRC8C (p = 0.026) were upregulated in the OA group (Figure 4).

3.4 GSEA and GSVA analysis of characteristic genes

In an endeavor to elucidate the underlying signal pathways associated with the seven key genes, gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA) were employed. Our findings revealed that HTR3B and KCNJ2 predominantly participate in the lysosomal and ribosomal synthesis pathways. In
contrast, CHRNA4 is mainly implicated in the extracellular matrix (ECM) receptor interaction pathway. GABRE is primarily engaged in the antigen processing and presentation pathway, whereas TRPM5 is chiefly involved in glycosaminoglycan biosynthesis, particularly heparan sulfate. The nod-like receptor signaling pathway is predominantly regulated by KCNG2, and LRRC8C is principally associated with the DNA replication pathway. These insights provide a foundation for further investigation into the molecular mechanisms governing these genes’ roles in their respective pathways (Figure. 5).

3.5 Correlation between characteristic genes and immune cells

Under the GSVA algorithm, Tfh cells, Th17 cells, and others were significantly increased in the OA group, while dendritic cells were significantly decreased (Figure. 6A). Although Figure-6B revealed no significant correlation between the 7 characteristic genes and immune cells.

3.6 Cluster analysis and PCA

To determine the clustering relationship of ion channel genes related to OA, we conducted cluster analysis using "Consensus ClusterPlus" based on the 7 characteristic ion channel genes in patients with OA from the GSE48556 dataset. Our analysis revealed 2 distinct subtypes of OA, C1 and C2, as visualized in Figures. 7A and 7B. The distribution of C1 and C2 samples is depicted in Figure. 7C. And Figure. 7D displays the expression of 7 characteristic genes in the 2 phenotypes. However, only 2 characteristic genes, KCNJ2 and LRRC8C, were significantly differentially expressed in the two subtypes C1 and C2.

3.7 Construction and validation of risk scores

To obtain the key modules most relevant to the ion channel genes of OA, we conducted WGCNA analysis of the healthy and OA groups, as well as the C1 and C2 groups from the GSE48556 dataset (Figure. 8A, B). Our analysis revealed that the correlation between the MEyellow module in different subtype groups and the MEred module in the disease group was the highest (Figures. 8C, D). We then selected key genes from important modules, resulting in 125 key genes from the healthy and disease groups and 61 key genes from different genotyping groups. By taking the intersection, we ultimately obtained 3 key genes (PPP1R3D, ZNF101, and LOC651309) (Figure. 8E).

3.8 Machine learning

We further evaluated the 3 key genes using 4 machine learning methods (RF, XGB, GLM, and SVM), of which mean squared deviations (Figure. 9A). Our analysis revealed that ZNF101 was the most important gene across all 4 machine learning methods (Figure. 9B). The ROC curve analysis in Figure. 9C indicated that the RF model had the largest ROC value (0.901) and was the best model.

3.9 Model validation

We validated the 3 gene model using the RF machine learning model. The nomogram in Figure. 10A depicts the score of the 3 genes in terms of their expression, and the calibration curve in Figure. 10B
demonstrates the predictive power of the nomogram model. The DCA curve in Figure 10C assesses the clinical value of the nomogram model. Finally, the ROC curve in Figure 10D, based on the validation set GSE46750, yielded an AUC of 0.667, indicating that the 3-gene model has reasonable accuracy and specificity.

4. Discussion

TRP(TRP) channels have gained attention as possible drug targets for the management of OA, rheumatoid arthritis, and gout due to accumulating clinical and preclinical evidence[23–25]. TRP channels are ligand-gated ion channels. Some members of this family of ion channels can detect external harmful stimuli such as heat, cold and chemicals. They respond to these stimuli by opening the pores and allowing calcium to enter the cells[26]. Some TRP channels are highly expressed in sensory neurons, when they are activated and followed by calcium influx, depolarization and action potentials can be triggered, causing acute or persistent pain[27]. In OA, TRP channels have been implicated in the nociceptive and inflammatory processes that contribute to the development and progression of the disease[28]. Among the TRP channel family members, TRPV1, TRPV4, and TRPA1 have garnered significant interest[23, 29]. TRPV1, for instance, is activated by various noxious stimuli and has been shown to play a role in pain transmission and sensitization. Additionally, TRPV4 is involved in mechanotransduction and has been linked to the regulation of chondrocyte function and cartilage integrity. Meanwhile, TRPA1 is associated with the detection of noxious cold and chemical irritants and has been implicated in inflammation and pain in OA[30]. Current OA research endeavors to elucidate the precise mechanisms underlying the involvement of TRP channels in disease progression and pain perception. Investigations also focus on the development of novel pharmacological agents targeting TRP channels, with the goal of improving therapeutic outcomes for OA patients. Such approaches aim to reduce pain, inflammation, and joint deterioration while minimizing side effects commonly associated with existing treatments.

In this study, we endeavor to elucidate potential TRP signaling pathways implicated in OA through the application of bioinformatics methodologies. By meticulously curating a robust collection of gene microarrays and employing various gene expression and microarray data sets, we effectively minimize error rates and bolster the reliability of our findings. Consequently, this study furnishes crucial insights for the clinical management and prevention of OA, thereby advancing our understanding of the interplay between TRP pathways and the immune microenvironment within the context of this debilitating disease. We identified seven key ion channel genes, specifically CHRNA4, GABRE, HTR3B, KCNG2, KCNJ2, LRRC8C, and TRPM5, that are implicated in OA. Utilizing a logistic regression model based on these seven genes, we achieved an AUC value of 0.930, highlighting the accuracy and specificity of the model in differentiating between OA and healthy samples. In contrast, the AUC values for each individual marker gene were approximately 0.6. Following a comprehensive functional enrichment analysis of these genes, CHRNA4 and KCNJ2 emerged as the most significant contributors, with the highest AUC values.
The CHRNA4 gene is responsible for encoding the α4 subunit of the acetylcholine receptor, known as the CHRNA4 protein, which plays a critical role in the rapid transmission of synaptic signals[31], and acetylcholine receptors can be a therapeutic target for the chronic pain control[32]. In contrast, KCNJ2 codes for the inward rectifier potassium channel, which facilitates the entry of potassium ions into the cell more efficiently than their extracellular movement[33]. This channel is primarily expressed in skeletal muscle and cardiac myocytes, and it is physiologically involved in regulating the resting membrane potential and modulating cellular excitability[33]. Hence, KCNJ2 is a vital determinant of the action potential duration, the resting membrane potential, and the pacemaker activity of the heart, among other functions.

GABRE encodes the gamma-aminobutyric acid (GABA) A receptor which plays a crucial role in signal transduction and can function as a neurotransmitter[34]. Migraine and OA have a clinical symptom of pain at the same time, Through the study of A case-control study, it proves that NRXN2, GABRE and CASK the synergetic effect between those genes and its relation with migraine susceptibility[35]. In the future, we can revolve around GABRE to research OA pain relief and treatment. Catastrophizing and knee pain were reported in 346 (10%) and 917 (24%) of the 3813 participants (mean 64.9 years, 58% female) at baseline in Data from the Osteoarthritis Initiative[36]. HTR3B was found to be associated with pain catastrophizing scores, future studies will be of great interest to HTR3B descending pain modulation pathways[37]. Inflammation in the synovial membrane response exhibits features of a T cell immune response, LRRC8C is an essential component of the VRAC channel in T cells and negative regulator of T cell function[38].

KCNG2 encodes one of voltage-gated potassium channels (Kv), Kv are important regulator of neuronal excitability, which regulates resting membrane potential and repolarization, Kv activity usually inhibits excitability of sensory neurons[39]. In fact, reduced Kv activity appears to be a marker of hyperexcitability seen in many pain syndromes ranging from traumatic injury and painful diabetic neuropathy to autoimmune diseases[40].

TRPM5 is a Ca2+-activated cation channel that mediates signaling in taste and other chemosensory cells, it can be activated physiologically by inositol 1,4,5-trisphosphate-producing receptor agonists, and it may therefore couple intracellular Ca2+ release to electrical activity and subsequent cellular responses[41]. TRPM5, TRPM7, and TRPV1 displayed an enhanced gene expression in freshly isolated chondrocytes[42].

Pain in OA has been associated with alterations in ion channels and the onset of inflammation. To analyze immune cell infiltration differences between OA patients and healthy controls, we employed the CIBERSORT and GSVA algorithms. Our findings revealed a significant increase in T.follicular.helper.cell and Type.17.T.helper.cell in the OA group, while plasmacytoid dendritic cells decreased considerably. Correlation analysis between the seven TRP genes and these immune cells did not yield significant results, suggesting that the TRP genes may not induce pain in OA patients via immune-mediated effects. To discern the clustering relationship of OA-associated ion channel genes, we conducted a consensus
clustering analysis on GSE48556 OA patients using the seven identified ion channel genes. This approach yielded two distinct subtypes, C1 and C2, with KCNJ2 exhibiting the most significant differential expression between them. We applied WGCNA to identify for co-expressed genes based on these subtypes, ultimately identifying three co-expressed key genes: PPP1R3D, ZNF101, and LOC651309. Lastly, we employed a RF machine learning approach to validate the three-gene model. The results confirmed the accuracy of the model, indicating its potential utility as a diagnostic marker for OA.

5. Conclusions

This study provide novel insights into the role of ion channel-related genes in OA pathogenesis and offer potential diagnostic markers and therapeutic targets for the treatment of OA.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets presented in this study can be found in online repositories, further inquiries can be directed to the corresponding author.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author’s Contributions

All authors contributed to the study conception and design. Conceptualization, Liu Yongming and Xiong Yizhe.; methodology, Liu Yongming.; software, Liu Yongming; validation, Xiong Yizhe, Wang Yupeng and Qian Zhikai; formal analysis, Xiong Yizhe.; investigation, Yin Mengyua.; resources, Qian Zhikai; data curation, Yin Mengyua.; writing—original draft preparation, Liu Yongming; writing—review and editing, Wang Xiang; visualization, Du Guoqing; supervision, Zhan Hongsheng.; project administration, Wang Xiang; funding acquisition, Du Guoqing, and Zhan Hongsheng. All authors have read and agreed to the published version of the manuscript.

Funding
This research was funded by the National Natural Science Foundation of China (82174403 82074466); the Summit Plateau Team Project in Traumatology of Shanghai University of TCM, Shanghai Chronic Musculoskeletal Disease Clinical Medical Research Center (20mc1920600); Shanghai clinical specialty “traditional Chinese medicine orthopaedic traumatology” (shslczd zk03901); the second-round construction project of the National Traditional Chinese Medicine Academic School Inheritance Studio "Shi’s Traumatology Department"; the innovation team of "Research and Transformation of Chronic Musculoskeletal Diseases" in Shanghai’s high-level local universities (Shanghai Education Commission [2022] No. 3); "Shanghai School of Traditional Chinese Medicine Inheritance" Extension Plan" (ZY(2021-2023)-0209-02), Zhan Hongsheng National Famous Veteran Traditional Chinese Medicine Expert Inheritance Studio Construction Project (2022-75).

References


Figures
Figure 1

Overall workflow chart of this study
Figure 2

Differential expression of ion channel genes (DEGs) between osteoarthritis and normal samples. (A) heat map of DEGs; (B) Correlation of DEGs; (C) Chromosome loci of DEGs; (D) GO analysis bar chart; (E) KEGG analysis bar chart.
Figure 3

Identify candidate diagnostic markers for OA. (A, B) Screening of diagnostic markers by the LASSO logistic regression algorithm. (C) Screening of diagnostic markers by SVM-RFE algorithm. (D) Venn diagram of genes identified by LASSO and SVM-RFE algorithms. (E) AUC of logistic regression model for identifying disease samples. (F) ROC curves of 7 marker genes.
Figure 4

Validation of the expression of seven ion channel related genes in OA versus normal controls. (A) CHRNA4; (B) GABRE; (C) HTR3B; (D) KCNG2; (E) KCNJ2; (F) LRRC8C; (G) TRPM5.

Figure 5

Functional enrichment analysis of ion channel related genes.
Figure 6

Profile of immune cell infiltration in OA. (A) Box plot of the proportion of 22 immune cell infiltrates calculated by GSVA. Red in the boxes represents the OA group and blue represents the normal group. (*\(P < 0.05\); **\(P < 0.01\); ns, not significant) (B) Heat map of the correlation between 22 immune cells and 7 ion channel genes calculated by the "GSVA" R package. Positive and negative correlations are indicated in blue and red, respectively. The darker the color, the stronger the correlation.
Figure 7

Cluster analysis of OA-associated ion channel genes. (A, B) Cluster analysis was used to classify OA into subgroup 1 and subgroup 2 based on the expression of ion channel genes. (C) PCA of C1 and C2. (D) Expression of seven ion channel genes in C1 and C2.
Figure 8

Identify important genes between different subtypes, between OA and healthy controls based on WGCNA. (A, B) WGCNA based on gene expression data identified gene modules with highly synergistic changes. (C) Heat map of normal versus osteoarthritic module-trait relationships. (D) Heat map of module-trait relationships for different subtypes. (E) Wayne diagram of important gene intersections between normal and OA, and between different subtypes.
Figure 9

Construction and evaluation of the 3-gene model by RF, XGB, GLM, and SVM. (A) Box plot of sample residuals. Red dots represent the mean square of residuals. (B) Significance of variables in RF, GLM, and SVM models. (C) ROC curves for different models.
Figure 10

Validation of the ion channel related 3-gene model. (A) Nomogram for predicting the occurrence of GDM. (B) Calibration curves to assess the predictive power of the nomogram model. (C) DCA curves to assess the clinical value of the nomogram model. (D) ROC curves for the validation set GSE46750 based on the 3-gene model (AUC=0.667).