Gene signature based on glycolysis is closely related to immune infiltration of patients with osteoarthritis

Ziyi Chen
Department of Sports Medicine, Huashan Hospital, Fudan University

Wenjuan Wang
Department of Sports Medicine, Huashan Hospital, Fudan University

Yinghui Hua (✉ hua_cosm@aliyun.com)
Department of Sports Medicine, Huashan Hospital, Fudan University

Research Article

Keywords: metabolism, glycolysis, inflammation, immune infiltration, osteoarthritis

Posted Date: October 10th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2132594/v1

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

**Background:** Osteoarthritis (OA) is a degenerative joint disease characterized by low-grade inflammation and high levels of clinical heterogeneity. Aberrant metabolism such as shifting from oxidative phosphorylation to glycolysis is a response to changes in the inflammatory micro-environment and may play a key role in cartilage degeneration and OA progression. Therefore, there is a pressing need to identify glycolysis regulators in the diagnosis of OA, determination of individualized risk, discovery of therapeutic targets, and improve understanding of pathogenesis.

**Methods:** We systematically studied glycolysis patterns mediated by 141 glycolysis regulators in 74 samples and discussed the characteristics of the immune microenvironment modified by glycolysis. The random forest was applied to screen candidate glycolysis regulators to predict the occurrence of OA. RT-qPCR was performed to validate these glycolysis regulators. Then two distinct glycolysis patterns were identified and systematic correlation between these glycolysis patterns and immune cell infiltration was analyzed. The glycolysis score was constructed to quantify glycolysis patterns together with immune infiltration of individual OA patient.

**Results:** 56 differentially expressed genes (DEGs) of glycolysis were identified between OA and normal samples. STC1, VEGFA, KDELR3, DDIT4 and PGAM1 were selected as candidate genes to predict the risk of OA using the random forest (RF) method. Two glycolysis patterns in OA were identified and glycolysis scoring system was constructed to show distinct individual immune characteristics. Glycolysis cluster A and higher glycolysis score was revealed to be related to an inflamed phenotype.

**Conclusions:** Taken together, these results established a genetic signature for OA based on glycolysis, which has reference significance for the in-depth study of the metabolic mechanism of OA and the exploration of new clinical treatment strategies.

Introduction

Osteoarthritis (OA), the most ubiquitous degenerative disease affecting the entire joint, is characterized by cartilage degradation and synovial inflammation[1]. Currently, OA is the leading cause of disability in older adults, and affects more than 500 million people worldwide, and represents a substantive burden to health care systems and socioeconomic costs[2–4]. Diagnosis of OA is based on clinical manifestations and physical examinations, however, because of its high clinical heterogeneity it is hard to confirm until reversible joint destruction happens[3]. Additionally, treatment for OA consists of pain medication and joint replacement surgery with a mass of limitations such as complications and lifespan of prostheses[3, 5, 6]. Therefore, an effective model is demanded to predict the risk of OA.

Although the pathogenesis of OA remains poorly understood, synovial inflammation is known to play an important role in OA development[1, 7]. Synovial tissue inflammation, which is known to contribute to OA development, exists in all stages of OA, even in the early stages[8]. Inflammatory regulators and matrix degradation enzymes produced by OA synoviocytes contribute to the progression of OA[9, 10]. For
example, the production of adhesion molecules such as intercellular adhesion molecule-1 (ICAM-1), vascular cell adhesion molecule-1 (VCAM-1) and chemokines causes synovial histological changes in the OA synovium, accompanied by infiltrated mononuclear cells (such as monocytes and macrophages) and lymphocytes (activated T cells and B cells)[11, 12]. These regulators are positively associated with the clinical symptoms of OA, such as inflammatory pain, joint swelling and disease development[1]. However, studies on OA pathophysiology have focused more on cartilage degeneration, rather than on the inflamed synovium[1].

Apart from mechanical and inflammatory factors, metabolic elements are also involved in the complex pathogenesis of OA[13]. Clinical evidence also indicates that OA often co-exists with metabolic diseases and comorbidities such as diabetes mellitus, and cardiovascular diseases[14]. In the environment of a joint with OA, chondrocytes experience a pathological shift of metabolic homeostasis and cartilage remodeling characterized by enhancement of the glycolytic pathway, disturbed mitochondrial function and so on[15]. Similarly, in synovial tissues of OA, metabolic alternations also occur. Glycolysis, for example, as one of six major metabolic pathways are involved in immunometabolism, plays an important role in the metabolic reprogramming of synoviocytes[15]. Under hypoxia and inflammatory conditions in OA joints, cells undergo enhanced anaerobic glycolysis with upregulation of glucose transporter type I (GLUT1), hexokinase (HK), pyruvate kinase (PK) and lactate dehydrogenase (LDH)[13]. Damerau et al. also showed that by inhibition of pathologically overexpressed pyruvate dehydrogenase kinases in OA synovial fibroblasts metabolically reprogrammed[16]. However, researches about distinct glycolysis patterns and their correlation with inflammatory environment in OA synovial tissues are still scarce in this field.

Therefore, in our study, 141 glycolysis regulator genes were compared in 74 OA and normal samples and five genes (STC1, VEGFA, KDELR3, DDIT4 and PGAM1) were selected to construct a prediction model of OA. Then we distinguished two distinct glycolysis patterns and analyzed its connection with inflammatory environment in OA. Finally, we established a set of scoring system to quantify the glycolysis pattern in individual patients.

**Methods And Materials**

**Data Acquisition and Difference Analysis**

The GSE1919, GSE41038, GSE55235, GSE82107 and GSE55457 dataset containing 38 OA patients and 36 normal controls were obtained from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/). To adjust the microarray, we employed the SVA method to merge the GSE1919, GSE41038, GSE55235, GSE82107 and GSE55457 dataset[17]. 200 cell glycolysis regulator genes obtained from the Gene Set Enrichment Analysis (GSEA) (https://software.broadinstitute.org/gsea/index.jsp/). Finally, 141 glycolysis regulators were extracted to identify distinct glycolysis patterns. Data normalization and probe annotation were performed on the data of the GSE1919, GSE41038, GSE55235, GSE82107 and GSE55457 dataset using the ‘limma’ and
‘GEOquery’ packages of R software (version 4.0.1) with DEG screening criteria of adjusted P-value < 0.05 and |log fold change (FC)| > 1.[18]

**Construction of prediction model**

The support vector machine (SVM) and the random forest (RF) methods were adopted to construct a training model to predict the risk of OA. “Boxplots of residual”, “Reverse cumulative distribution of residual”, and a receiver operating characteristic (ROC) curve were used to assess the accuracy of the model. The RF method was applied to predict the occurrence of OA using the R library “randomForest”[19]. In this study, mtry and ntree were set to 3 and 500, respectively. The optimal ntree was selected according to minimum cross-validation error in 10-fold cross-validation. The significance of differentially-expressed glycolysis regulators was evaluated with the optimal ntree. Nomogram construction was performed using the “rms” package[20]. The consistency between the actual observed values and the predicted values was evaluated using calibration curves, and a clinical impact curve and decision curve analysis were conducted to evaluate the clinical benefit of our model.

**Identification of glycolysis clusters**

Glycolysis patterns were identified using consensus clustering based on Glycolysis differentially expressed genes (DEGs) and the “ConsensusClusterPlus” R package was used[21]. The principal component analysis (PCA) was used to correlate the principal component with senescence clusters[22].

**Single-sample gene-set enrichment analysis (ssGSEA)**

The relative infiltration levels of 23 immune cells in the GSE1919, GSE41038, GSE55235, GSE82107 and GSE55457 dataset were quantified using the ssGSEA algorithm[23]. Spearman correlations were calculated for 23 immune infiltrating cells with 56 glycolysis regulators, followed by visualization using the ‘ggplot2’ package[24].

**Functional enrichment analysis**

DEGs between glycolysis clusters were screened using the “limma” package in R, and the criterion for screening differential genes was P < 0.05[18]. Meanwhile, GO enrichment analyses (www.geneontology.org/) and KEGG pathway (www.genome.jp/kegg/pathway.html/) were performed using the R package, including “clusterProfiler”, “org.Hs.eg.db”, “enrichplot”, “ggplot2”, “RColorBrewer”, “dplyr”, and “ComplexHeatmap”[19, 20, 25–32].

**Generation of glycolysis signature**

To quantify the glycolysis patterns of individual patients, we constructed a set of scoring system – the glycolysis signature based on DEGs between OA and control using PCA. We then conducted PCA to construct m6A relevant gene signature. Both principal component 1 and 2 were selected to act as signature scores.

\[ \text{m6A score} = \sum PC1i \]
where $i$ is the expression of m6A phenotype-related genes.

**Sample collection**

Synovial tissue from 3 patients of meniscus injury and OA were collected from Huashan hospital. All subjects read and signed the informed consent form. The study was in conformance with the guidelines of the 1975 Declaration of Helsinki and was approved by the ethics committee of Huashan Hospital with informed written consent (KY2020-060).

**Reverse Transcription-Quantitative Polymerase Chain Reaction (RT-qPCR)**

The total RNA was isolated from synovial tissue samples by using Trizol (Thermo, California, USA). After uniform quality between groups, total RNA was reverse transcribed to complementary DNA (cDNA) using PrimeScript™ RT Master Mix (TaKaRa, Tokyo, Japan). The qRT-PCR was performed using the SYBR Green Master Mix (Thermo, California, USA) using cDNA according to the manufacturer. The primer sequence of genes was shown in Table 1. Genes were normalized to GAPDH. Relative expression of mRNA was calculated through the $2^{-\Delta\Delta CT}$ method. Each biological sample tests were technically replicated performed in triplicate.

**Statistical Analysis**

All data calculations and statistical analyses were performed using R software (https://www.r-project.org/, version 4.1.1/). Comparison between two groups of continuous normally distributed variables were realized by performing the independent Student’s t-test, and the difference between non-normally distributed variables was analyzed using the Mann-Whitney U test (Wilcoxon rank-sum test). All statistical p values were bilateral, and statistical significance was set at $p < 0.05$.

**Results**

**Analysis of expression characteristics of glycolysis regulators in OA synovial tissues**

A total of 56 glycolysis regulators were finally identified in this study. The differences in glycolysis gene expression between groups were visualized in heatmap (Fig. 1A). The location of glycolysis regulator genes in chromosome was displayed in Fig. 1B. Specifically, qRT-PCR showed that KDELR3 and PGAM1 were expressed at a high level in OA, while STC1, VEGFA and DDIT4 in normal (Fig. 1C).

**Construction of the OA predictive model using the SVM and RF methods**

The prediction performance between the SVM and RF methods were compared. “Boxplots of residual” (Fig. 2A), “reverse cumulative distribution of residual” (Fig. 2B), and a ROC curve (Fig. 2C) showed that RF
has substantially high prediction accuracy. The RF model is optimal as all of the samples in the model had minimum residuals. In this study, mtry and ntree were set to 3 and 500, respectively. The optimal ntree was selected according to the minimum cross-validation error in 10-fold cross-validation (Fig. 2D). After ranking the genes according to their importance, we identified the 30 glycolysis regulators (Fig. 2E). The calibration curves (Fig. 2G), clinical impact plots (Fig. 2H) and decision curve analysis (DCA) (Fig. 2I) showed that the nomogram model may be an ideal predictive model for OA. The nomogram evaluation model was constructed based on five glycolysis regulators to predict the probability of OA, showing remarkable predictive power (Fig. 2F).

**Identification of two distinct glycolysis clusters**

Two glycolysis clusters were identified (glycolysis cluster A and B) based on 56 glycolysis DEGs between OA and normal synovial samples using consensus clustering (Fig. 3A-D). The bloxplot and heatmap showed the 56 senescence genes’ expression in the two distinct glycolysis groups (Fig. 3E, F). PCA was applied to verify the two distinct glycolysis clusters divided by consensus clustering of the 56 glycolysis regulators (Fig. 3G).

**Immune cell infiltration analysis in the two senescence phenotypes**

We analyzed the difference degree of immune infiltration between the two distinct glycolysis clusters using ssGSEA method. A more significant infiltration was found in activated dendritic cell (P < 0.001), immature dendritic cell (P < 0.01), MDSC (P < 0.001), macrophage (P < 0.001), natural killer (NK) T cell (P < 0.05), NK cell (P < 0.01), plasmacytoid dendritic cell (P < 0.05), regulatory T cell (P < 0.001) and type 1 helper T cell (P < 0.01) (Fig. 4A, B). GNPDA1, AKR1A1, ENO1, SDC3 and ME2 were positively related to immune cell infiltration, while VEGFA, CTH and STC1 were negatively connected (Fig. 4C-J).

**Function analysis in the two senescence phenotypes**

GO annotation and KEGG pathway analyses were conducted on DEGs between the two glycolysis clusters to perform the gene functional enrichment analysis. The biological process (BP) of GO term result emphasized positive regulation of cell activation, leukocyte migration, myeloid leukocyte activation, leukocyte proliferation and muscle contraction; for the molecular function (MF) of GO term, DEGs were significantly enriched in membrane raft, membrane microdomain, actin cytoskeleton, secretory granule membrane and endocytic vesicle; the cellular components (CC) terms highlighted that DEGs were mainly concentrated on actin binding, amide binding, peptide binding, immune receptor activity and amyloid-beta binding (Fig. 4K). The KEGG result manifested that DEGs were found to be mostly related to myeloid leukocyte activation, myeloid cell activation involved in immune response, cell activation involved in immune response, leukocyte proliferation antigen processing and presentation of exogenous, peptide antigen via MHC class I, macrophage activation, positive regulation of cell activation, leukocyte activation involved in immune response, muscle contraction and leukocyte migration (Fig. 4L).
Generation of senescence signatures

To further evaluate the glycolysis pattern and immune infiltration of an individual patient with OA, we constructed the senescence score using PCA based on DEGs between OA and normal. We found that glycolysis scores correlated with glycolysis clusters (P = 0.00024), to illustrate, glycolysis cluster A had higher scores. The boxplots showed differential gene expression of mitogen-activated protein kinases (MAPK), toll-like receptors (TLR), necroptosis and transforming growth factor-β (TGF-β) pathways which is most related to OA (Fig. 5B-E).

Discussions

OA is the one of most common arthritis leading to pain, joint destruction and disability characterized by cartilage degradation and synovial inflammation\[1\]. Metabolism is important for cartilage and synovial joint function\[13\]. In the past few years, several studies have demonstrated that metabolism has a key role in inflammatory joint diseases\[15\]. Evidence suggests that patients with OA fall into multiple phenotypic subgroups defined on the basis of the main driver of disease, included a metabolic phenotype and a synovitis-driven inflammatory phenotype, although all OA phenotypes probably involve metabolic alterations\[15\]. Glucose is an important metabolic fuel being vital for extracellular matrix synthesis and degradation, and it is metabolized via glycolysis\[15\]\[33\]. Warburg effect, exhibiting enhanced glycolysis even under aerobic conditions, with the purpose of providing sufficient energy to support rapid biosynthesis, also exists in the pathogenesis of OA\[13, 34\]. Systematic analysis of the interaction between glycolysis and OA patient prognosis and its potential associations with immune infiltration is lacking but urgently needed. In this study, we aimed to distinguish distinct metabolic phenotypes related to glycolysis and figured out the correlation between metabolic patterns with inflammatory microenvironment in synovial tissues of OA.

Firstly, we observed a significant difference in the expression of 56 glycolysis regulator factors between OA patients and normal controls. Next, we established a glycolysis nomogram for predicting the risk of OA from the perspective of metabolism. Different scores were assigned to factors such as STC1, VEGFA, KDELR3, DDIT4 and PGAM1. Consistent results were obtained via RT-qPCR, which validated our findings. Moreover, we investigated the association between glycolysis regulatory factors and the immune properties of OA, including the gene set for immune cell infiltration and inflammatory response. Unsupervised clustering of OA samples using glycolysis regulator expression profiles led to two subtypes with distinctive glycolysis patterns. Furthermore, 56 glycolysis DEGs was used to generate glycolysis scores for every patient by PCA. Most patients from glycolysis cluster A were further classified into high glycolysis score group and an inflamed phenotype; whereas patients from glycolysis cluster B were classified into low glycolysis score group and a non-inflamed phenotype. Our results also in the two distinct glycolysis patterns there is a significant difference of gene expression of MAPK, TLR, necroptosis and TGF-β pathways which all significantly involved in OA\[35\]\[36\].
Among the five genes, STC1 (Stanniocalcin 1) is a paracrine factor associated with inflammation and carcinogenesis that can help mesenchymal cells to protect cancer cells from apoptosis and enhance the Warburg effect[37]. DDIT4 is a mammalian target of rapamycin (mTOR) inhibitor[38]. Eddie et al showed that IL-10 inhibits lipopolysaccharide-induced glucose uptake and glycolysis and promotes oxidative phosphorylation through the induction of an mTOR inhibitor, DDIT4[38]. PGAM1, phosphoglycerate mutase 1, plays a key role in glycolysis. Shen et al. demonstrated that S1P/S1PR3 axis promotes aerobic glycolysis by YAP/c-MYC/PGAM1 axis in osteosarcoma[38]. These suggested that these glycolysis regulators may also involve in OA and served as biomarkers of diagnosis.

OA patients often exhibit inflammatory infiltration of synovial membranes by macrophages, T cells, mast cells, B cells, plasma cells, natural killer cells, dendritic cells, granulocytes, etc[39]. Glycolysis also have a neglectable role in immune infiltration as well. For example, glycolysis plays an important role in T cells, which are assumed to be associated with the pathogenesis of OA, as significant abnormalities in the T-cell profile have been found in the synovial membranes of OA patients[39]. Both CD8 and CD4 T cell-dependent immune responses and the helper T cell-dependent inflammation were attenuated by Pgam1 deficiency[40]. Macrophages are among the most abundant cell type present in the cellular infiltrates found in the inflamed synovium in OA[41]. Moreover, Cai et al. showed that the glycolysis inhibitor 2-deoxyglucose ameliorates adjuvant-induced arthritis by regulating macrophage polarization in an AMPK-dependent manner[42]. All these studies had implications in our results; however, further research is urgently acquired to further demonstrate the correlation between glycolysis and inflammatory environment in OA.

Taken together, glycolysis pathway plays a significant role in OA. Firstly, we constructed a glycolysis OA nomogram to assess the risk of OA, thus providing a reference for the clinical diagnosis of OA. Then two distinct glycolysis patterns and a strong correlation between these patterns and inflammatory environment were identified. Meanwhile, a novel scoring system to quantify glycolysis pattern in individual patients was built. Our findings provided novel ideas for promoting personalized immunotherapy, and demonstrating a new horizon for the investigations of the pathogenesis of OA in the future.

Declarations

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. All data used in this work were obtained from the GEO (https://www.ncbi.nlm.nih.gov/geo/).

AUTHOR CONTRIBUTIONS

Chen, Wang and Hua developed the research question. Chen, Wang and Hua wrote the first draft of the manuscript. All authors contributed to the development of the review protocol, data analysis, and refining
of the manuscript, and approved the final manuscript. Chen, Wang and Hua critically read and revised the manuscript before submission.

**FUNDING**

This work was supported by the China National Key R&D Program SQ2020YFA080008.

**ACKNOWLEDGMENTS**

Not applicable.

**References**


27. Hadley Wickham, Romain François, Lionel Henry, Kirill Müller. A grammar of data manipulation [R package dplyr version 1.0.0]. Media. 2020;


Figures
Figure 1

Differentially expressed glycolysis genes in OA. (A) Heatmap showing the significantly different expression of senescence glycolysis genes in OA and normal synovial tissues. (B) The location of these genes on the chromosome. (C) The mRNA expression in the normal and OA group. (All figures * represents $p < 0.05$, ** represents $p < 0.01$, *** represents $p < 0.001$) OA, osteoarthritis.
Figure 2

Construction of OA risk predictive model. Boxplot of the residual distribution (A) and reverse cumulative distribution of residual (B) as a function of the values and ROC curves (C) showing the observed sensitivity between RF and SVM. (D) RF: prediction error curves based on 10-fold cross-validation. (E) The importance of the 30 glycolysis regulators based on the RF model. (F) Nomogram graph of the predictive model based on five glycolysis regulators. The calibration curves (G), clinical impact plot (H) and DCA (I)
Figure 3

Identification of two glycolysis clusters. (A-D) Clustering of OA synovial samples based on glycolysis regulators. (E) Boxplot demonstrating the gene expression of glycolysis regulators. (F) Heatmap showing
the DEGs of the two glycolysis clusters. (G) PCA was utilized to verify the two glycolysis clusters. (All figures * represents p <0.05, ** represents p <0.01, *** represents p <0.001) OA, osteoarthritis; DEGs, differentially expressed genes; PCA, principal component analysis.

Figure 4
Immune cell infiltration and function enrichment analyses. (A) Heatmap showing the correlation between the expression of the 56 glycolysis regulators and immune cells infiltration using the ssGSEA method. (B) Boxplot showing the infiltrating immune cells in the two glycolysis clusters. (C-J) Connection between key genes (GNPDA1, AKR1A1, VEGFA, EN01, CTH, STC1, SDC3 and ME2) and immune cell infiltration. (K) The bubble diagram showing the top 10 terms of GO categories of BP, MF and CC. (L) Barplot diagram showing the KEGG enrichment analysis. (All figures * represents p <0.05, ** represents p <0.01, *** represents p <0.001) ssGSEA, single-sample gene-set enrichment analysis; GO, gene ontology; KEGG, Kyoto Encyclopedia of Gene and Genome; BP, biological process; MF, molecular function; CC, cellular component.

Figure 5

Generation of glycolysis signature. (A) The glycolysis score in the two glycolysis clusters. (B) Boxplot showing the gene expression of MAPK pathway in the two glycolysis clusters. (C) Boxplot showing the gene expression of TLR signaling pathway in the two glycolysis gene clusters. (D) Boxplot showing the gene expression of necroptosis signaling pathway in the two glycolysis gene clusters. (E) Boxplot showing the gene expression of TGF-β signaling pathway in the two glycolysis gene clusters. (All figures * represents p <0.05, ** represents p <0.01, *** represents p <0.001) MAPK, mitogen-activated protein kinases; TLR, toll-like receptors; TGF-β, transforming growth factor-β.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarytable.docx