scientific reports

Check for updates

OPEN

A multidimensional approach reveals the function of lactylation related genes in osteoarthritis

Shanjie Luan[1]✉ & Jian Luan[2]

As a complex joint disease, osteoarthritis (OA) increasingly affects the elderly. Currently, existing drugs cannot cure OA. There is an urgent need for new targets. Lactylation is closely related to inflammation and is an emerging target in treatment. However, the potential of lactylation-related genes (LRGs) in OA is poorly understood. This study identified differentially expressed lactylation-related genes (DELRGs) through bioinformatics analysis, constructed a model through a combination of various machine learning methods, and performed immune infiltration analysis, single-cell analysis and molecular docking to predict drugs. Mendelian randomization was used to study the causal relationships between eQTLs and the three types of osteoarthritis. Finally, we used RT-qPCR and CCK-8 assays to validate the results of the bioinformatics analysis. We generated a model with good diagnostic efficacy and seven hub genes, which revealed that osteoarthritis is associated with the infiltration of immune cells such as dendritic cells and macrophages, as well as with the cell communication between fibroblasts and macrophages. Azacitidine, with significant docking results, was obtained through seven hub genes. The results of RT-qPCR verified the expression of LRGs and CCK-8 assay indicated that azacitidine can significantly inhibit the proliferation of OA cells. Overall, we established a lactylation-based diagnostic model and obtained novel biomarkers, which are expected to lead to the development of new strategies for the diagnosis and treatment of OA.

Keywords Osteoarthritis, Lactylation, Diagnostic model, Machine learning

Osteoarthritis (OA) is a degenerative joint disease that affects elderly people the most severely[1]. The epidemiology of the disease is complex and multifactorial, in-cluding genetic, biological, and biomechanical components[2]. Currently, joint replacement surgery is an effective way to treat osteoarthritis[3], while drug treatment can only relieve the symptoms of OA associated with inflammation and pain[4]. However, the etiology and pathogenesis of the disease are still unknown[5]. Therefore, there is an urgent need for more efficient and more appropriate biomarkers to play a role in all stages of OA diagnosis, prognosis, drug development and treatment[6].

With the development of high-throughput sequencing technology, people are committed to analyzing the epigenome of human cells and have discovered an increasing number of novel epigenetic codes, including lactylation[7]. Histone lysine lactylation was first reported in 2019, indicating that lactic acid-derived lactylation of histone lysine residues acts as an epigenetic modification that directly stimulates gene transcription on chromatin[8]. Histone lactylation is an indicator of lactic acid levels and glycolysis and is intrinsically related to cell metabolism[9], which is involved in biological processes such as tumor proliferation, nerve excitation, and inflammation[10].

A number of studies have shown that histone lactylation and inflammation are closely related. For example, the TLR signaling adapter BCAP regulates the transition from inflammation to repair in macrophages by promoting histone lactylation[11]. A study on the lactylation of macrophages in acute neonatal brain injury was performed to elucidate the potential mechanisms of inflammation[12]. Small molecules targeting PKM2 provided the molecular basis for the inhibition of rheumatoid arthritis by lactylation-dependent fibroblast-like synovial cell proliferation[13]. These findings illustrate the effectiveness of using lactylation as a therapeutic target for the regulation of inflammation and may provide a new perspective for the treatment of inflammation-related diseases[14]. Although lactylation has received widespread attention, related articles are still limited. In particular, there are few articles on the modification of lactylation in OA.

In this study, potential lactylation targets were obtained from the published articles, and the data of OA patients were integrated to obtain differentially expressed genes. The hub lactylation-related genes (LRGs) were

[1]School of Basic Medical Sciences, Shandong University, 44 Wenhua Xi Road, Jinan 250012, Shandong, China. [2]Department of Spine Surgery, Qingdao Municipal Hospital, No. 5, Middle Dong Hai Road, Qingdao 266000, Shandong, China. ✉email: Luan20050519@163.com

selected by using a combination of 14 machine learning methods. Cluster analysis, immune analysis, single-cell analysis, drug prediction and Mendelian randomization were subsequently performed. It is worth noting that, to date, no LRGs have been reported to effectively predict the prognosis of OA, and further analysis is needed. This paper fills the gap in this field and is expected to provide an effective basis for OA treatment.

Materials and methods

Data source

Eight gene expression array datasets, GSE98918, GSE169077, GSE55235, GSE55457, GSE51588, GSE82107, GSE10575, and GSE206848, were obtained from the public GEO database (http://www.ncbi.nlm.nih.gov/geo) and included data of control groups and OA patients. The three datasets GSE169077, GSE55235, and GSE55457 were merged and named as GEO-Merge No. (1) Three other datasets GSE82107, GSE10575, and GSE206848 were merged and named as GEO-Merge No. (2) Missing values in the data were imputed using the "impute" package, and the intergroup differences were removed using the "ComBat" function of the "sva" package.

Identification of differentially expressed lactylation-related genes (DELRGs)

The "limma" package was used to carry out differential analysis on the GEO-Merge No. 1 dataset, and the results of differential analysis were visualized with a volcano map. Four hundred LRGs were obtained from published literature[15]. The intersection of the differentially expressed genes and the LRGs was drawn in a Venn diagram. For gene set functional enrichment analysis, we first downloaded the subset of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) from the Molecular Signatures Database[16] and used it as a background to map the genes to the background set. The "ClusterProfiler" package was used to perform enrichment analysis on the DELRGs. The results were drawn as a chord diagram, and pathways with p values less than 0.05 and gene ratios greater than 3 were drawn.

Integrate various types of machine learning

Since its proposal in 2022, the medical application of more than ten machine learning-based ensemble programs has received widespread attention and promotion[17]. Because it was difficult to obtain the specific survival time of OA patients, we used a binary classification method to analyze. We integrated 14 machine learning algorithms: neural networks, logistic regression, linear discriminant analysis, quadratic discriminant analysis, KNN nearest neighbors, decision tree classification, random forest classification, XGBoost, ridge regression, lasso regression, elastic net regression, support vector machines, gradient boosting machines, stepwise logistic regression models, naive Bayes, and obtained 279 algorithm combinations. A binary classification model was developed using the GEO-Merge No. 1 dataset as the training set and the GEO-Merge No. 2, GSE51588, and GSE98918 datasets as the test set.

The diagnostic accuracy of each model on the four datasets was calculated, and the consistency index (C-index) for model selection was also calculated. We defined the model with the highest average C-index in the test set as the optimal model. We chose the top 120 models in terms of accuracy to demonstrate. The number and average ranking of the 22 DELRGs included in the model are shown in the heatmap. We ranked the genes by calculating the average importance ranking and occurrence frequency of all models. When filtering genes, priority is given to those that appear frequently in multiple models and those with a higher average importance ranking. This ranking principle ensures that both the stability of genes across multiple models and their importance in each model are taken into account. Ultimately, we selected seven LRGs.

Based on the machine learning results, we selected the model with the highest accuracy to obtain the model score and seven hub genes. The receiver operating characteristic (ROC) curve of the optimal model was plotted, and the area under the curve (AUC) was calculated.

Consensus clustering and immune infiltration analysis

We used the "ConsensusClusterPlus" software package[18] to cluster the 51 samples in the GEO-Merge No. 1 dataset according to the DELRGs, classified all the individuals into two subgroups, and used a Sankey diagram to compare the sample information with the grouping status. ImmuCellAI[19] was used to assess the abundance of immune cell types in the samples. This method collected characteristic genes for each type of immune cell through literature and other analytical methods (such as CIBERSORT, xCell) to establish a specific gene set. Using the single-sample gene set enrichment analysis (ssGSEA) algorithm, it calculated the deviation from the reference expression profile, thereby computing enrichment scores for each type of immune cell. After adjustment with a compensation matrix and least squares regression, it output the abundance data of immune cell types for each sample, and the results were expressed in pie charts and stacked histograms. We also evaluated the correlation between immune cells in the two subgroups, and the results were presented as a heatmap.

Single-cell analysis

We obtained scRNA-seq data of 12 OA patients' cartilage and meniscus from GSE220243 and used "Seurat" package to process the data[20]. After normalization, data integration and principal component analysis (PCA) were performed. Cell clusters were created using the "FindClusters" and "FindNeighbors" functions. The "CELLiD" was used for automatic cell annotation, and regularization was performed manually. We used the "SCP" package to illustrate the expression of LRGs in different cell types, and we calculated the lactylation score for each cell type based on the gene expression data, with the results displayed in a violin plot. We also used "CellChat" to analyze the cell communication between different cell types[21].

Drug prediction

Interactive visualization of gene expression characteristics of over 16,000 drugs and small molecules for up- and downregulated genes induced in cell lines was recorded in the L1000FWD database[22]. We used the seven hub genes to screen and obtain the drugs and small molecules that were negatively correlated, and finally, we obtained azacitidine. The crystal structures of the proteins encoded by the hub genes were obtained from the RCSB Protein Data Bank (PDB) website (www.rcsb.org/pdb/home/home.do). In addition, drug structures were retrieved from PubChem (https://www.ncbi.nlm.nih.gov/pccompound).

CB-Dock2 was used to perform molecular docking of seven hub genes and azacitidine. CB-Dock2 is an improved protein-ligand blind docking tool. It combines structure-based cavity detection with template fitting for blind docking. In template fitting, the system predicts binding sites and docking poses by querying similar protein-ligand complex structures using template information; if no similar templates are available, it resorts to structure-based docking alone. Benchmark tests show that this tool has a binding site prediction success rate of about 85% (RMSD < 2.0 Å), significantly outperforming other commonly used blind docking tools such as SwissDock and COACH-D. Compared to existing blind docking methods, CB-Dock2 has improved the success rate of docking pose predictions by 16–30%, demonstrating its advantages in accuracy and efficiency[23].

Mendelian randomization

IEU Open GWAS (https://gwas.mrcieu.ac.uk) data was used for Mendelian randomization (MR). Importantly, in the relevant original research, every participant in the IEU Open GWAS database gave informed consent.

In order to derive unconfounded estimates of the causal effect of the exposures of interest on the outcome variables, genetic variation was included as an instrumental variable (IV) in multiple regression analysis. First, throughout the entire genome, we found single nucleotide polymorphisms (SNPs) connected to exposure that were statistically significant ($p < 5^{*}10^{-8}$). We employed the aggregation technique to exclude SNPs in significant linkage disequilibrium (LD) with the clumping size of 10,000 kb and R^2 value less than 0.001.

Mendelian randomization analysis was performed using "TwoSampleMR" package to verify the causal relationship between the exposures and the results. This study used two MR methods, the inverse variance weighting (IVW) method and the Mendelian randomization-Egger method (MR-Egger), to ascertain the causal relationship be-tween exposure and outcomes. The IVW method was chosen because of its superior statistical validity compared with existing methods and its ability to consistently estimate the causal impact of exposure on outcomes.

Based on nSNPs ≥ 2, we obtained eQTL data for five genes, BRD4, CEBPZ, CRABP2, HLTF, MNDA, and performed analysis with knee osteoporosis (ebi-a-GCST007090), hip osteoporosis (ebi-a-GCST007091), knee or hip osteoporosis (ebi-a-GCST007092).

Cell culture

Human chondrocytes were cultured using immortalized cell culture medium at 37 °C with 5% CO2. Cells were treated during the logarithmic growth phase. Upon thawing, frozen cells were melted in a 37 °C water bath, then resuspended in culture medium and incubated in a culture chamber. For passaging, cells were washed with PBS and digested with trypsin. The treated cells were resuspended by centrifugation and inoculated at a predetermined density. Based on previous studies, we chose to induce chondrocytes with 10ng/ml of IL-1β to obtain osteoarthritis cells.

Real-time quantitative polymerase chain reaction (RT-qPCR)

Total RNA was extracted from cells using the RNAprep FastPure kit, and the concentration and purity of RNA were measured using a spectrophotometer. The extracted RNA was reverse transcribed into cDNA using SynScript®III RT SuperMix, with reaction temperature conditions set at 50 °C for 15 min. PCR reactions were performed using SYBR Green dye and specific primers, amplified with the ABI ViiA 7 PCR instrument. The reaction program included an initial denaturation at 95 °C, using β-actin as an internal control. Detailed primer information can be found in the supplementary material (Supplementary Table S8).

Cell counting Kit-8 (CCK-8)

5-Azacytidine and IL-1β were dissolved to prepare stock solutions, which were then diluted to different concentrations for treatment. After 24 h, the CCK-8 kit was used to assess cell viability. 10μL of CCK-8 solution was added to each well, and after incubating for 3 h, absorbance was measured at 450 nm using a microplate reader to evaluate cell activity.

Results

Lactylation differential gene analysis

We screened 1,514 DELRGs from the GEO Merge No. 1 dataset based on |LogFC|>0.58 and p value < 0.05 (Fig. 1A). These genes were intersected with the LRGs to obtain 22 DELRGs (Fig. 1B). GO and KEGG enrichment analyses were performed on 22 genes. MF terms such as RNA binding and protein domain specific binding; BP terms such as response to starvation; CC terms such as ribonucleoprotein complex and nuclear protein containing complex; and KEGG pathways such as the spliceosome were identified (Fig. 1C, D, Supplementary Table S2).

Integrate various types of machine learning

Among the 279 machine learning combinations, the "ENR – CV:10-fold (cutoff: 0.5, alpha: 0.5)" model had the highest accuracy (Fig. 2A). We scored each dataset based on the coefficient obtained from this model and obtained four ROC curves, whose AUC values all exhibited good predictive ability (Fig. 2B-E). In addition,

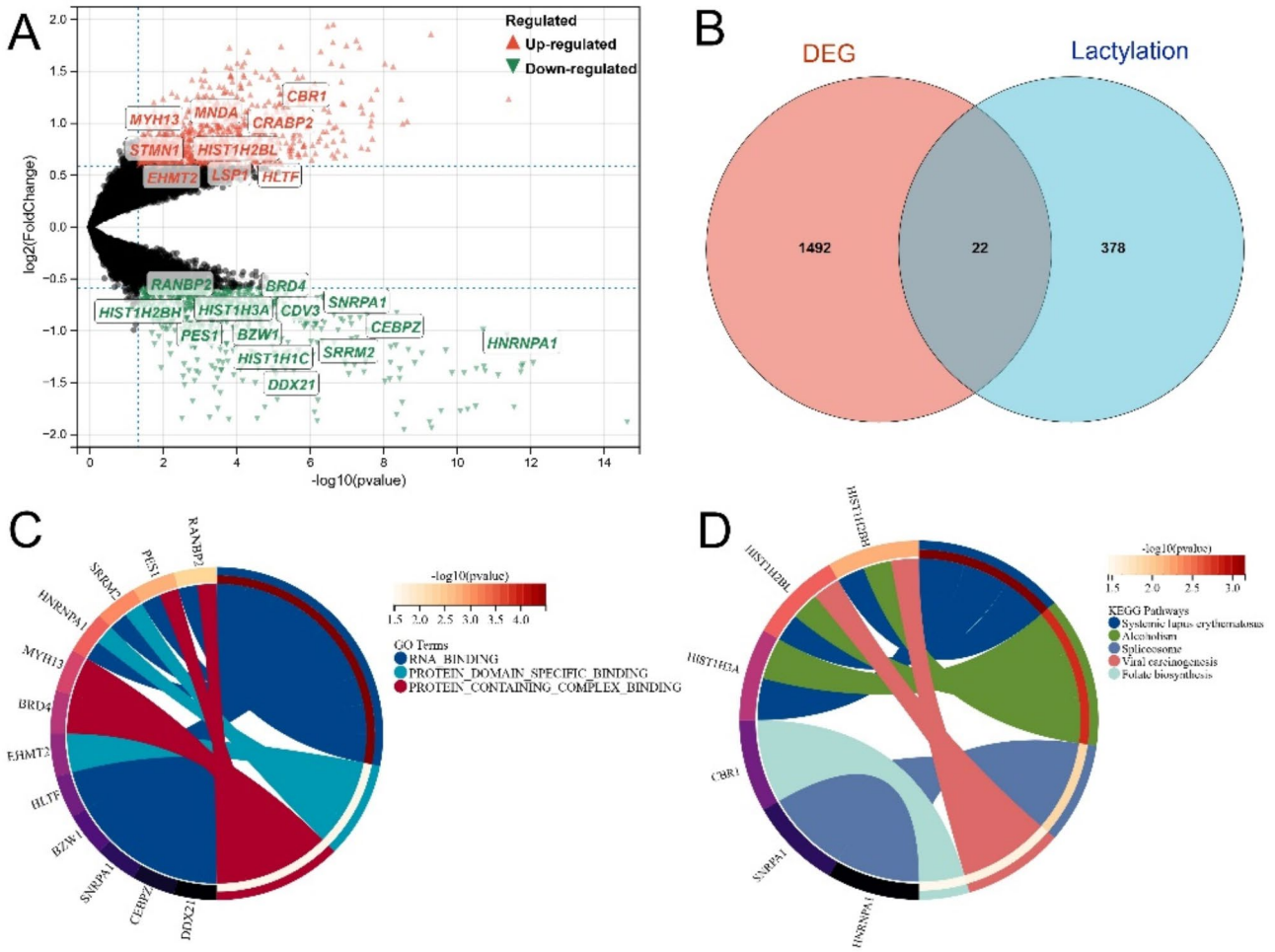

Fig. 1. (**A**) Volcano plot of gene expression in the GEO-Merge No. 1 dataset. (**B**) Venn diagram of the intersection of DEGs and LRGs. (**C**) GO: MF enrichment results of the DELRGs. (**D**) KEGG enrichment results for the DELRGs.

after synthesizing all the models obtained by machine learning, we found that the seven genes, BRD4, CEBPZ, CRABP2, HIST1H2BL, HLTF, HNRNPA1, and MNDA not only had high average rankings but were also included in most of the models (Fig. 2F). We obtained ROC curves for these seven genes in four datasets, and according to their AUC values, these seven genes all had good predictive performance (Fig. 2G-J). They had the potential to become therapeutic targets.

Consensus clustering and immune analysis

After the samples were divided into two subgroups, C1 and C2 (Fig. 3A), we found that most diseased individuals were in the C2 subgroup (Fig. 3B). Compared with those in the C1 subgroup, the number of dendritic cells (DCs) in the C2 subgroup was signifi-cantly greater, while the number of natural killer T cells (NKT) and CD8 + naive cells was significantly lower (Fig. 3C, D). Pie charts and stacked histograms visualize the differences in DCs, NKT, CD8 + naive cells and other immune cells between the two subgroups and individuals (Fig. 3E, F, G). In the C1 and C2 subgroups, CD4 + T cells and Tr1, Tfh, Th17 cells exhibited strong positive correlations. Strong negative correlations between macrophages and Tfh cells, NKT cells and CD8 naive cells were also observed in the C2 subgroup (Fig. 3H, I).

Single-cell analysis

After quality control and removal of batch effects, these cells were classified and then annotated into ten cell types (Fig. 4A). The lactylation score significantly increases except red blood cells (Fig. 4B). BRD4, CEBPZ, HNRNPA1 are highly expressed. In addition, the expression levels of various genes in fibroblasts are significantly higher than in other cells. (Fig. 4C). The results of cell communication showed that most interactions occurred between different types of fibroblasts and macrophages (Fig. 4D). In summary, fibroblasts may be the key cells involved in the pathogenesis of osteoarthritis and are related to lactylation.

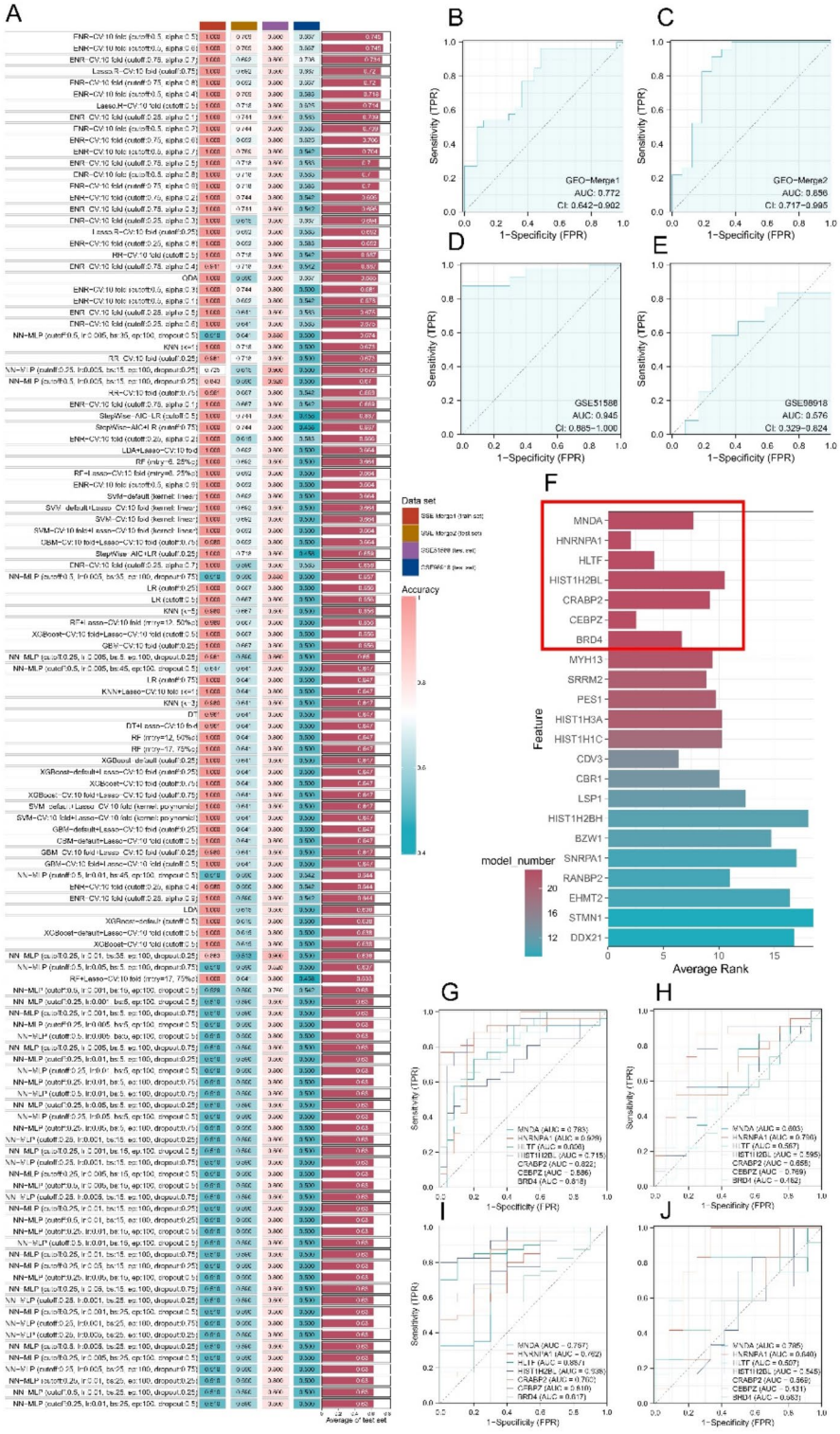

Fig. 2. (**A**) Accuracy of various machine learning combinations. (**B**-**E**) ROC curves of the mod-el score in the GEO-Merge NO. 1, GEO-Merge NO. 2, GSE-51,588, and GSE-98,918 datasets. (**F**) The number of DELRGs included in the model and their average ranking. (**G**-**J**) ROC curves of the seven hub genes in the GEO-Merge NO. 1, GEO-Merge NO. 2, GSE-51,588, and GSE-98,918 datasets.

Drug prediction

The molecular structure of azacitidine has been demonstrated (Fig. 5A). The vina score of MNDA was − 6.1 kcal/mol (Fig. 5B), the vina score of HNRNPA1 was − 5.9 kcal/mol (Fig. 5C), the vina score of HLTF was − 6.3 kcal/mol (Fig. 5D), the vina score of HIST1H2BL was − 5.1 kcal/mol (Fig. 5E), the vina score of CRABP2 was

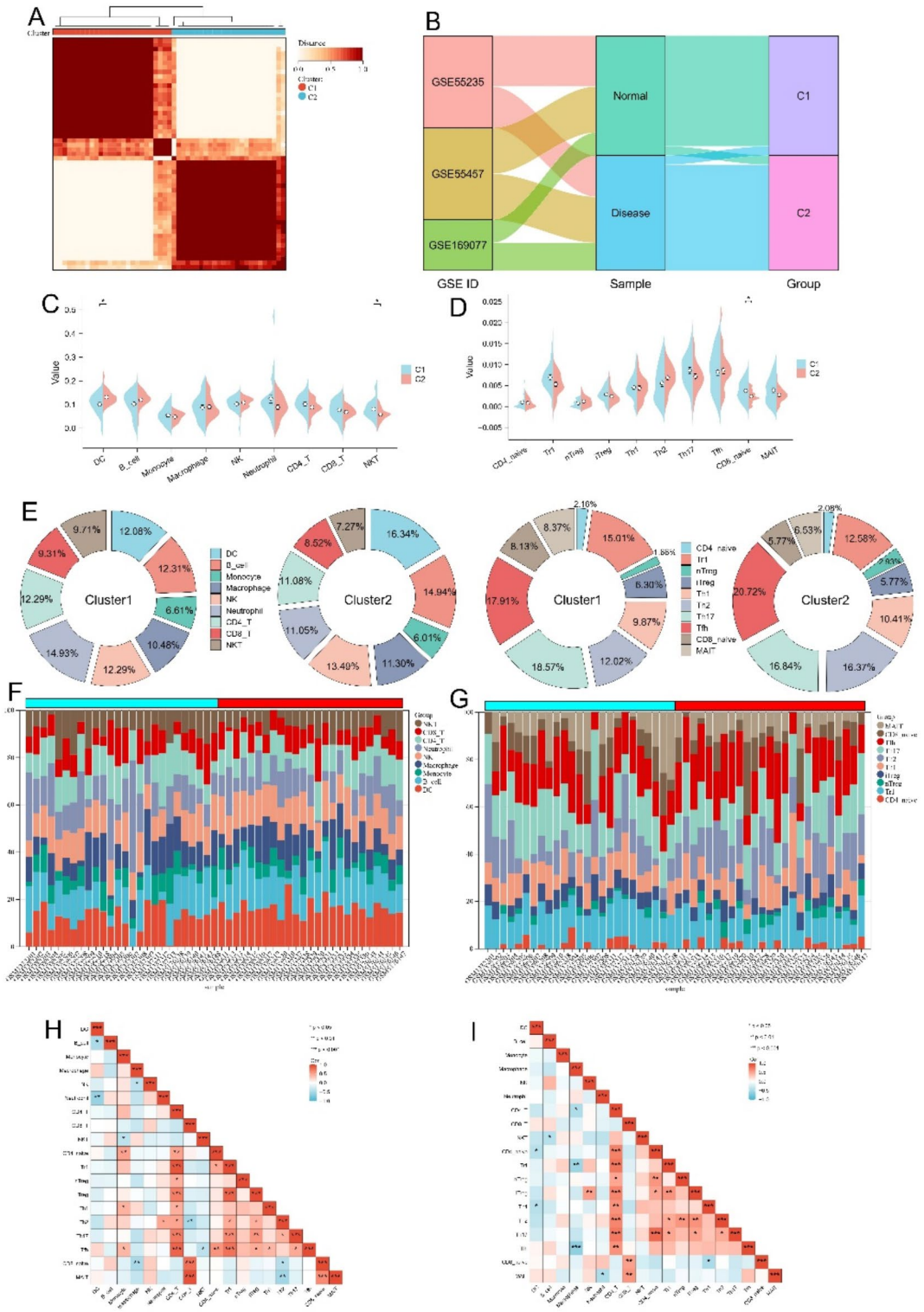

Fig. 3. (**A**) Consensus clustering heatmap. (**B**) Sankey diagram of the clustering re-sults. (**C**,**D**) Comparison of 19 types of immune cells. (**E**) The proportion of different immune cells in the two clusters. (**F**,**G**) Stacked histograms of 19 types of immune cells. (**H**,**I**) Immune cell correlation heatmaps in the two subgroups.

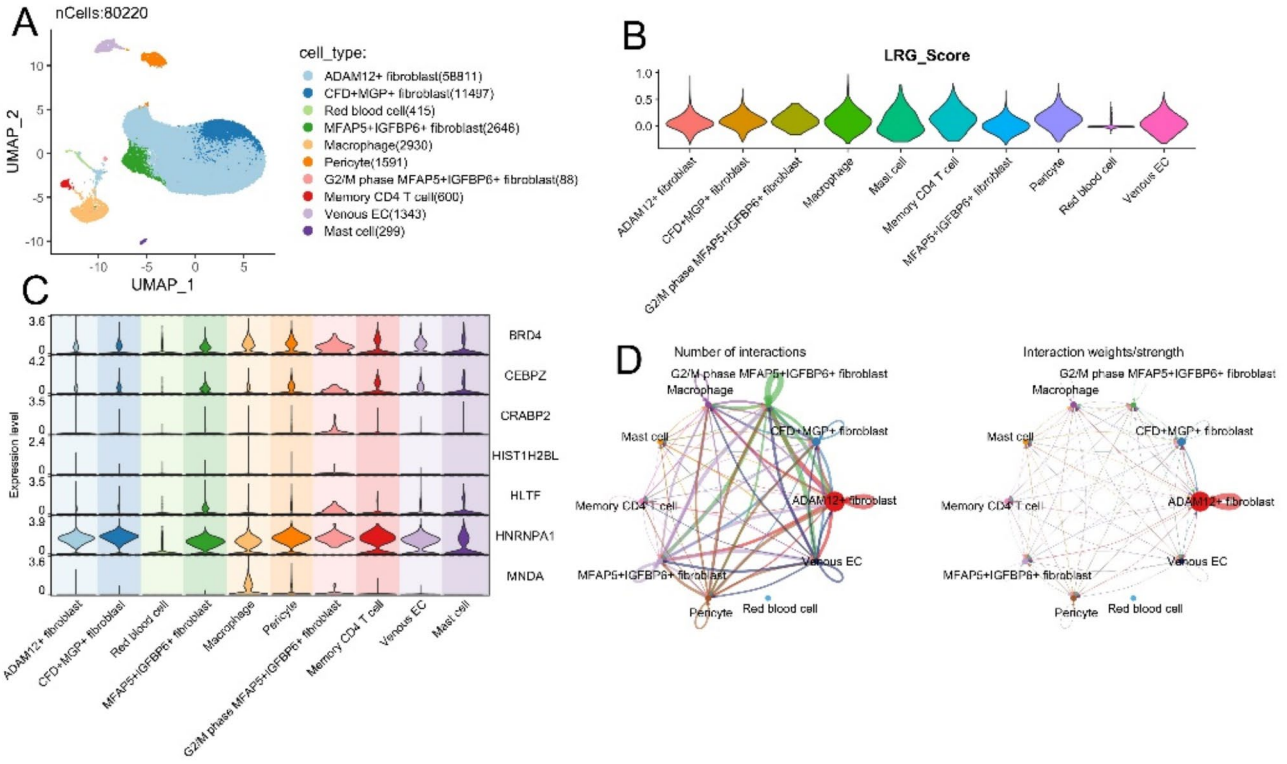

Fig. 4. (**A**) Results of cell clustering and annotation. (**B**) The score of LRG in different cell types. (**C**) Violin plot of gene expression in different cells. (**D**) Number and strength of cell communication results.

−5.9 kcal/mol (Fig. 5F), the vina score of CEBPZ was −6.7 kcal/mol (Fig. 5G), and the vina score of BRD4 was −6.7 kcal/mol (Fig. 5H). All molecular docking scores are less than −5 kcal/mol. These findings indicate that azacitidine has good affinity for LRGs and may be used for the treatment of osteoarthritis.

Mendelian randomization

MR results showed the association between the eQTLs of the five genes and OA. The results showed that HLTF, BRD4, CRABP2 and MNDA are risk factors of OA, while CEBPZ is a protective factor of OA (Fig. 6).

Experimental results

From the RT-qPCR results, compared to normal chondrocytes, the relative expression levels of CRABP2, HIST1H2BL, HLTF, and MNDA were significantly increased in osteoarthritis cells (Fig. 7A-D), which is consistent with our bioinformatics analysis results.

In addition, the azacitidine treatment group showed a significant inhibitory effect on cell viability, especially at higher concentrations (such as 10μM and 20μM), where the inhibitory effect of the drug was very pronounced. Furthermore, the inhibitory effect of azacitidine on IL-1β induced osteoarthritis cells was significantly enhanced compared to normal chondrocytes, suggesting that the drug has a stronger effect in inhibiting the proliferation of osteoarthritis cells (Fig. 7E).

Discussion

The development of artificial intelligence (AI), especially machine learning (ML), in the health care field has led to important improvements and discoveries, especially in rheumatology and osteoarthritis (OA)[24,25]. The trends and applications of ma-chine learning are constantly increasing, and machine learning is still an emerging field with incredible potential[26]. In recent years, many articles have used machine learning to identify and diagnose OA-associated genes and have demonstrated good diagnostic performance[27,28]. However, due to database limitations, the size of the included samples generally did not exceed 100, and comprehensive evaluation and analysis of the obtained models are seldom performed. A variety of machine learning algorithm combinations were included in this study, and an optimal model and seven hub genes were obtained, followed by immune infiltration analysis, single-cell analysis, drug prediction, Mendelian randomization and experiments. It is expected to have certain value for the diagnosis and treatment of OA.

Among the seven hub genes, Nanotherapy effectively reduces osteoarthritis by targeting BRD4-regulated synovial macrophage polarization[29]. CEBPZ, a C/EBP transcription factor, is involved in osteoarthritis cell activation and interacts with various signaling pathways[30]. The level of the retinoic acid-binding protein CRABP2 was increased in an OA mouse model[31]. β-Hydroxybutyrate relieves cartilage aging during the development of OA through HNRNPA1-mediated upregulation of PTEN[32]. Enrichment analysis revealed that

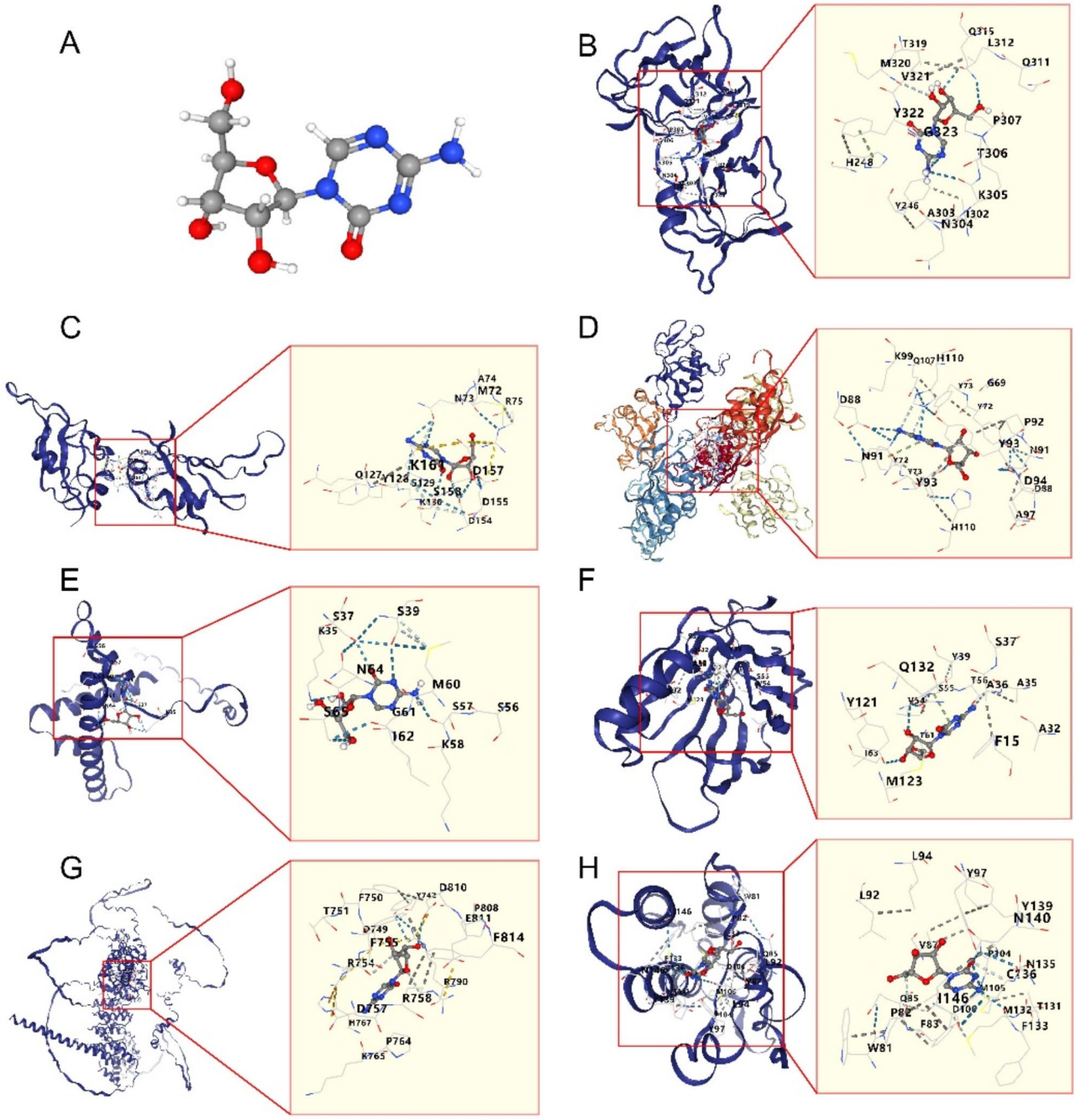

Fig. 5. (**A**) Molecular structure of azacitidine. (**B**) Molecular docking results of MNDA. (**C**) Molecular docking results of HNRNPA1. (**D**) Molecular docking results of HLTF. (**E**) Molecular docking results of HIST1H2BL. (**F**) Molecular docking results of CRABP2. (**G**) Molecular docking results of CEBPZ. (**H**) Molecular docking results of BRD4.

the response to hunger, nucleoprotein complexes, and RNA binding pathways may affect the cell's adaptability to environmental changes by regulating the function of RNA binding proteins (RBPs), which in turn influences RNA metabolism and post-transcriptional modifications[33]. Dysfunction in these processes can lead to cellular aging and osteoarthritis[34]. KEGG enrichment showed that the involved spliceosome pathway was also enriched under normoxia and hypoxia[35]. This finding confirmed that the selected genes were significantly involved in osteoarthritis and lactylation process.

Immune infiltration and single-cell analysis can reveal the molecular programs and lineage progression models of human OA pathogenesis[36]. A significantly in-creased number of dendritic cells undergo metabolic reprogramming from oxidative phosphorylation to glycolysis under the stimuli of hypoxia, nutrient deprivation and inflammation[37]. Fibroblasts, which play a significant role in cell communication, have been shown to

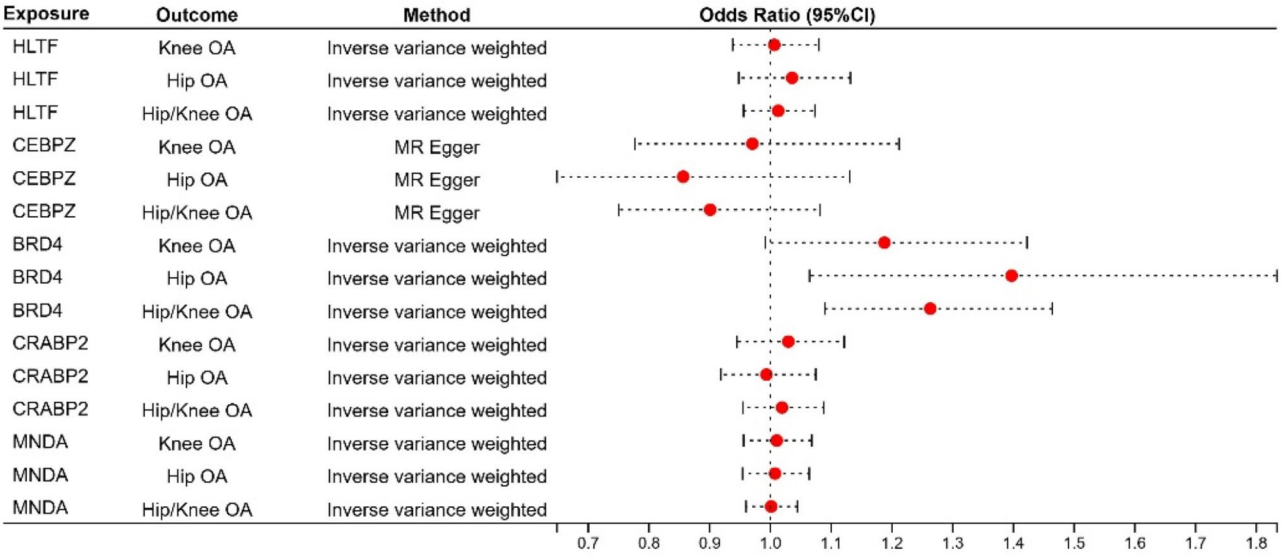

Fig. 6. Results of Mendelian randomization.

promote the profibrotic activity of macrophages through lactic ac-id-induced histone lactylation[38], and targeting YAP1-regulated glycolysis in fibro-blast-like synovial cells can also impair macrophage infiltration to improve the progression of osteoarthritis[39]. Therefore, we believe that fibroblasts and macrophages are directly or indirectly involved in the regulation of immune processes by lactylation, thus affecting the development and progression of OA.

Molecular docking is a mature computer structure-based method that is extensively used in drug discovery[40]. Azacitidine is a DNA methylation inhibitor that was approved by the US Food and Drug Administration (FDA) in May 2004 for the treatment of myelodysplastic syndrome[41]. In recent years, it has also been used to treat acute lymphocytic leukemia and X-linked sideroblastic anemia[42,43]. In 2024, a patient with auto-inflammatory syndrome who presented with sacroiliitis had a molecular response to azacitidine[44]. Our experimental results also indicate that azacitidine can significantly inhibit the proliferation of osteoarthritis cells. Therefore, we speculated that this compound might be a potential drug for the treatment of OA.

Mendelian randomization (MR) uses genetic variation as an instrumental variable to infer whether risk factors causally affect health outcomes[45]. It is also extensively used in osteoarthritis and can be used to identify biomarkers[46]. Among the identified risk factors, HLTF, CRABP2, and MNDA were upregulated in OA patients, while the expression of the protective factor CEBPZ was downregulated in OA patients, which also verified the accuracy of the results of the present study.

However, this study still has many limitations. First, although more than 100 samples were included, there were only 24 individuals in the validation set GSE98918, causing errors to occur, and the ROC curve results were not very accurate or obvious. Second, we also noted that the mechanisms related to lactylation in OA are still unclear. Therefore, further studies are needed to reveal the influences of related genes on pathways such as immune cell infiltration in OA. Finally, since this paper used public databases for analysis, the accuracy of the model in the prediction and diagnosis of OA still needs to be verified by large-scale clinical trials.

Conclusion

In summary, we identified an OA diagnostic model through a combination of various machine learning methods and identified seven hub genes associated with lactylation that had high diagnostic specificity for OA. In addition, immune infiltration and single-cell analyses revealed that lactylation-related genes may affect the development of OA by regulating the immune response, and azacitidine was identified as a possible therapeutic drug.

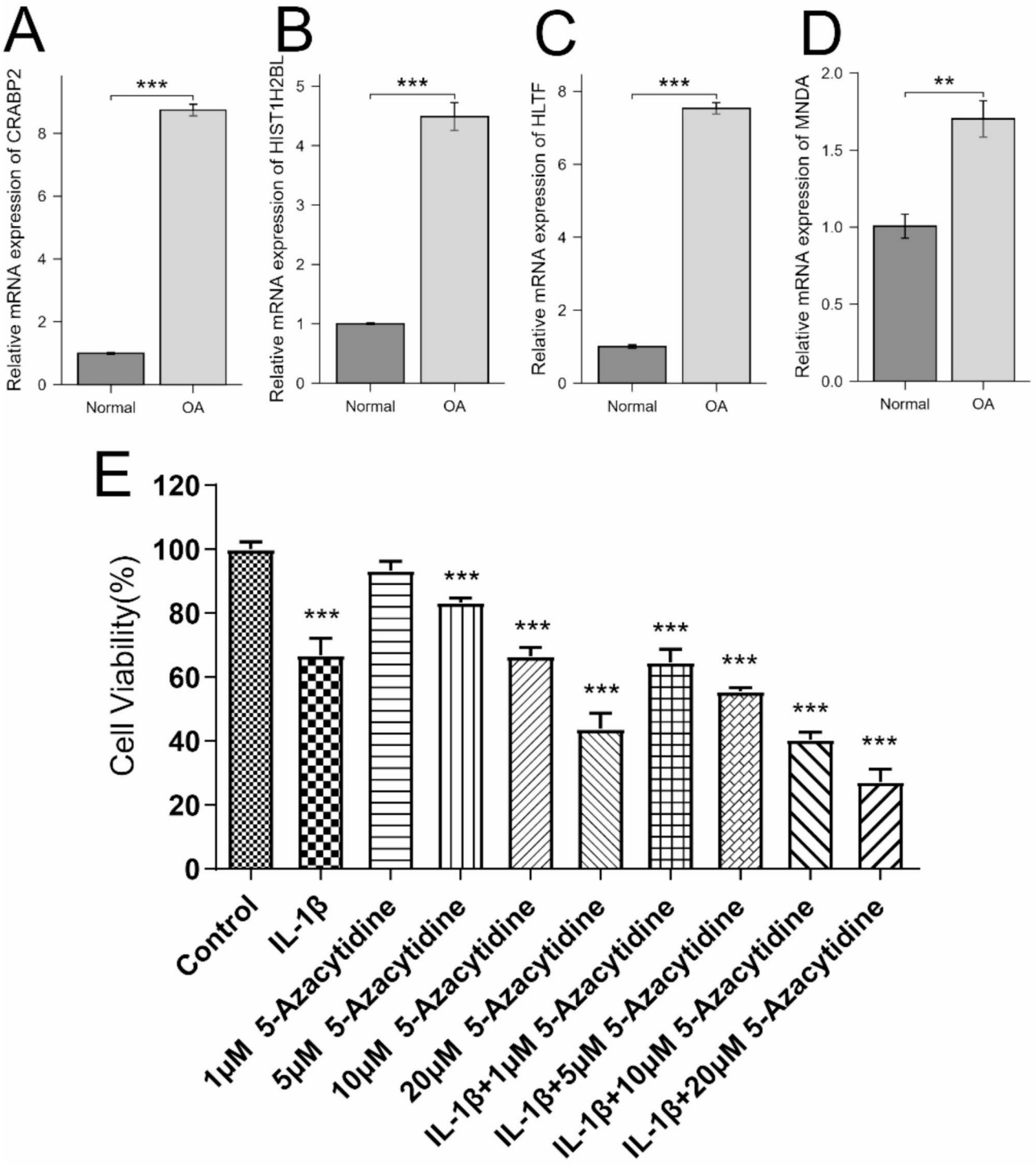

Fig. 7. Relative expression level of CRABP2 (**A**), HIST1H2BL (**B**), HLTF (**C**), MNDA (**D**). CCK-8 results of chondrocytes induced by IL-1β and treated with azacitidine (**E**). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Data availability

The data available in this article were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/), the RCSB Protein Data Bank (PDB) database (www.rcsb.org/pdb/home/home.do), the L1000FWD database (https://maayanlab.cloud/l1000fwd/), the PubChem database (https://www.ncbi.nlm.nih.gov/pccompound) and the IEU Open GWAS database (https://gwas.mrcieu.ac.uk). Other specific experimental data can be obtain ed from the corresponding author.

Received: 22 June 2024; Accepted: 3 February 2025

Published online: 05 March 2025

References

1. Mi, B. et al. Icariin activates autophagy via down-regulation of the NF-κB signaling-mediated apoptosis in chondrocytes. *Front. Pharmacol.* **9**, 605. https://doi.org/10.3389/fphar.2018.00605 (2018).
2. Glyn-Jones, S. et al. *Lancet* **386**, 376–387, doi:https://doi.org/10.1016/S0140-6736(14)60802-3. (2015).
3. Xia, B., Di Zhang, C., Hu, J., Jin, S. & Tong, H. Osteoarthritis pathogenesis: A review of molecular mechanisms. *Calcif Tissue Int.* **95**, 495–505. https://doi.org/10.1007/s00223-014-9917-9 (2014).
4. Cho, Y. et al. Disease-modifying therapeutic strategies in osteoarthritis: Current status and future directions. *Exp. Mol. Med.* **53**, 1689–1696. https://doi.org/10.1038/s12276-021-00710-y (2021).
5. Hu, X., Ni, S., Zhao, K., Qian, J. & Duan, Y. Bioinformatics-led discovery of osteoarthritis biomarkers and inflammatory infiltrates. *Front. Immunol.* **13**, 871008. https://doi.org/10.3389/fimmu.2022.871008 (2022).
6. Kraus, V. B. & Karsdal, M. A. Clinical monitoring in osteoarthritis: Biomarkers. *Osteoarthr. Cartil.* **30**, 1159–1173. https://doi.org/10.1016/j.joca.2021.04.019 (2022).
7. Lv, X., Lv, Y., Dai, X. & Lactate histone lactylation and cancer hallmarks. *Expert Rev. Mol. Med.* **25**, e7. https://doi.org/10.1017/erm.2022.42 (2023).
8. Zhang, D. et al. Metabolic regulation of gene expression by histone lactylation. *Nature* **574**, 575–580. https://doi.org/10.1038/s41586-019-1678-1 (2019).
9. Chen, A. N. et al. Lactylation, a Novel Metabolic Reprogramming Code: Current status and prospects. *Front. Immunol.* **12**, 688910. https://doi.org/10.3389/fimmu.2021.688910 (2021).
10. Li, X. et al. Correction: Lactate metabolism in human health and disease. *Sig Transduct. Target. Ther.* **7**, 372. https://doi.org/10.1038/s41392-022-01206-5 (2022).
11. Irizarry-Caro, R. A. et al. TLR signaling adapter BCAP regulates inflammatory to reparatory macrophage transition by promoting histone lactylation. *Proc. Natl. Acad. Sci. U S A.* **117**, 30628–30638. https://doi.org/10.1073/pnas.2009778117 (2020).
12. Zhou, Y., Yang, L., Liu, X. & Wang, H. Lactylation may be a novel posttranslational modification in inflammation in neonatal hypoxic-ischemic encephalopathy. *Front. Pharmacol.* **13**, 926802. https://doi.org/10.3389/fphar.2022.926802 (2022).
13. Wang, Y. H. et al. Small-molecule targeting PKM2 provides a molecular basis of lactylation-dependent fibroblast-like synoviocytes proliferation inhibition against rheumatoid arthritis. *Eur. J. Pharmacol.* **972**, 176551. https://doi.org/10.1016/j.ejphar.2024.176551 (2024).
14. Fang, Y. et al. Emerging roles of lactate in acute and chronic inflammation. *Cell. Commun. Signal.* **22**, 276. https://doi.org/10.1186/s12964-024-01624-8 (2024).
15. Pan, J., Zhang, J., Lin, J., Cai, Y. & Zhao, Z. Constructing lactylation-related genes prognostic model to effectively predict the disease-free survival and treatment responsiveness in prostate cancer based on machine learning. *Front. Genet.* **15**, 1343140. https://doi.org/10.3389/fgene.2024.1343140 (2024).
16. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740. https://doi.org/10.1093/bioinformatics/btr260 (2011).
17. Liu, Z. et al. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. *Nat. Commun.* **13**, 816. https://doi.org/10.1038/s41467-022-28421-6 (2022).
18. Wilkerson, M. D., Hayes, D. N. & ConsensusClusterPlus. A class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573. https://doi.org/10.1093/bioinformatics/btq170 (2010).
19. Miao, Y. et al. ImmuCellAI: A unique method for comprehensive T-cell subsets abundance prediction and its application in cancer immunotherapy. *Adv. Sci.* **7**, 1902880. https://doi.org/10.1002/advs.201902880 (2020).
20. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304. https://doi.org/10.1038/s41587-023-01767-y (2024).
21. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088. https://doi.org/10.1038/s41467-021-21246-9 (2021).
22. Wang, Z., Lachmann, A., Keenan, A. B. & Ma'ayan, A. L1000FWD: Fireworks visualization of drug-induced transcriptomic signatures. *Bioinformatics* **34**, 2150–2152. https://doi.org/10.1093/bioinformatics/bty060 (2018).
23. Liu, Y. et al. CB-Dock2: Improved protein–ligand blind docking by Integrating Cavity Detection, docking and homologous template fitting. *Nucleic Acids Res.* **50**, W159–W164. https://doi.org/10.1093/nar/gkac394 (2022).
24. Pandit, A. & Radstake, T. R. D. J. Machine learning in Rheumatology approaches the clinic. *Nat. Rev. Rheumatol.* **16**, 69–70. https://doi.org/10.1038/s41584-019-0361-0 (2020).
25. Kokkotis, C., Moustakidis, S., Papageorgiou, E., Giakas, G. & Tsaopoulos, D. E. Machine learning in knee osteoarthritis: A review. *Osteoarthr. Cartil. Open.* **2**, 100069. https://doi.org/10.1016/j.ocarto.2020.100069 (2020).
26. Binvignat, M. et al. Use of machine learning in osteoarthritis research: A systematic literature review. *RMD Open.* **8**, e001998. https://doi.org/10.1136/rmdopen-2021-001998 (2022).
27. Zhou, J. et al. Identification of aging-related biomarkers and Immune infiltration characteristics in Osteoarthritis based on bioinformatics analysis and machine learning. *Front. Immunol.* **14**, 1168780. https://doi.org/10.3389/fimmu.2023.1168780 (2023).
28. Li, J. et al. Identification of immune-associated genes in diagnosing osteoarthritis with metabolic syndrome by integrated bioinformatics analysis and machine learning. *Front. Immunol.* **14**, 1134412. https://doi.org/10.3389/fimmu.2023.1134412 (2023).
29. Xu, Y. D. et al. Apoptotic body-inspired nanotherapeutics efficiently attenuate osteoarthritis by targeting BRD4-regulated synovial macrophage polarization. *Biomaterials* **122483**. https://doi.org/10.1016/j.biomaterials.2024.122483 (2024).
30. Berenbaum, F. Signaling transduction: Target in osteoarthritis. *Curr. Opin. Rheumatol.* **16**, 616–622. https://doi.org/10.1097/01.bor.0000133663.37352.4a (2004).
31. Welch, I. D., Cowan, M. F., Beier, F. & Underhill, T. M. The retinoic acid binding protein CRABP2 is increased in murine models of degenerative joint disease. *Arthritis Res. Ther.* **11**, R14. https://doi.org/10.1186/ar2604 (2009).
32. Xia, G. et al. β-Hydroxybutyrate alleviates cartilage senescence through hnRNP A1-mediated up-regulation of PTEN. *Exp. Gerontol.* **175**, 112140. https://doi.org/10.1016/j.exger.2023.112140 (2023).
33. Zigdon, I. et al. Beyond RNA-binding domains: Determinants of Protein-RNA binding. *RNA 2024*, **30**, 1620–1633, https://doi.org/10.1261/rna.080026.124
34. Yoon, D. S. et al. TLR4 downregulation by the RNA-binding protein PUM1 alleviates cellular aging and osteoarthritis. *Cell. Death Differ.* **29**, 1364–1378. https://doi.org/10.1038/s41418-021-00925-6 (2022).
35. Song, F. et al. Lactylome analyses suggest systematic lysine-lactylated substrates in oral squamous cell carcinoma under Normoxia and Hypoxia. *Cell. Signal.* **120**, 111228. https://doi.org/10.1016/j.cellsig.2024.111228 (2024).
36. Ji, Q. et al. Single-cell RNA-Seq analysis reveals the progression of human osteoarthritis. *Ann. Rheum. Dis.* **78**, 100–110. https://doi.org/10.1136/annrheumdis-2017-212863 (2019).
37. Gan, P., Wu, H., Zhu, Y., Shu, Y. & Wei, Y. Glycolysis, a driving force of rheumatoid arthritis. *Int. Immunopharmacol.* **132**, 111913. https://doi.org/10.1016/j.intimp.2024.111913 (2024).
38. Cui, H. et al. Lung myofibroblasts promote macrophage profibrotic activity through lactate-induced histone lactylation. *Am. J. Respir Cell. Mol. Biol.* **64**, 115–125. https://doi.org/10.1165/rcmb.2020-0360OC (2021).

39. Yang, J. et al. Targeting YAP1-regulated glycolysis in fibroblast-like synoviocytes impairs macrophage infiltration to ameliorate diabetic osteoarthritis progression. *Adv. Sci.* **11**, 2304617. https://doi.org/10.1002/advs.202304617 (2024).
40. Pinzi, L. & Rastelli, G. Molecular docking: Shifting paradigms in drug discovery. *IJMS* **20**, 4331. https://doi.org/10.3390/ijms20184331 (2019).
41. Issa, J. P. J., Kantarjian, H. M., Kirkpatrick, P. & Azacitidine *Nat. Rev. Drug Discov* **4**, 275–276, doi:https://doi.org/10.1038/nrd1698. (2005).
42. Cheung, L. C. et al. Preclinical efficacy of Azacitidine and Venetoclax for Infant KMT2A-Rearranged Acute Lymphoblastic Leukemia reveals a New Therapeutic Strategy. *Leukemia* **37**, 61–71. https://doi.org/10.1038/s41375-022-01746-3 (2023).
43. Morimoto, Y. et al. Azacitidine is a potential therapeutic drug for pyridoxine-refractory female X-Linked sideroblastic Anemia. *Blood Adv.* **6**, 1100–1114. https://doi.org/10.1182/bloodadvances.2021005664 (2022).
44. Costa, P. D. et al. Case Report: VEXAS Syndrome: An atypical indolent presentation as Sacroiliitis with Molecular Response to Azacitidine. *Front. Immunol.* **15**, 1403808. https://doi.org/10.3389/fimmu.2024.1403808 (2024).
45. Bowden, J. & Holmes, M. V. Meta-analysis and mendelian randomization: A review. *Res. Synthesis Methods*. **10**, 486–496. https://doi.org/10.1002/jrsm.1346 (2019).
46. Gu, Y. et al. Causality of genetically determined metabolites and metabolic pathways on osteoarthritis: A two-sample mendelian randomization study. *J. Transl Med.* **21**, 357. https://doi.org/10.1186/s12967-023-04165-9 (2023).

Acknowledgements
We are very grateful for all the databases that provided data sources in our article, as well as the platform for technical exchange provided by the Shandong University Bioinformatics Association.

Author contributions
Shanjie Luan completed all the research, writing, and proofreading portions of this manuscript in its entirety. Jian Luan proposed and guided the direction of this study.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-89072-3.

Correspondence and requests for materials should be addressed to S.L.