Research Paper

A two-gene random forest model to diagnose osteoarthritis based on RNA-binding protein-related genes in knee cartilage tissue

Wenhua Yin[1,*], Ying Lei[2,*], Xuan Yang[1], Jiawei Zou[1]

[1]Department of Orthopaedics, Yuebei People's Hospital Affiliated to Medical College of Shantou University, Shaoguan, Guangdong 512026, China
[2]Department of Audit, Yuebei People's Hospital Affiliated to Medical College of Shantou University, Shaoguan, Guangdong 512026, China
[*]Equal contribution

Correspondence to: Wenhua Yin; **email:** yinwenhuasohu@sohu.com, https://orcid.org/0000-0003-1868-1310
Keywords: osteoarthritis, RNA-binding protein, diagnosis, machine learning, random forest
Received: October 4, 2022 **Accepted:** December 20, 2022 **Published:** January 5, 2023

ABSTRACT

Osteoarthritis (OA) is one of the most common diseases in the orthopedic clinic, characterized by progressive cartilage degradation. RNA-binding proteins (RBPs) are capable of binding to RNAs at transcription and translation levels, playing an important role in the pathogenesis of OA. This study aims to investigate the diagnosis values of RBP-related genes in OA. The RBPs were collected from previous studies, and the GSE114007 dataset (control = 18, OA = 20) was downloaded from the Gene Expression Omnibus (GEO) as the training cohort. Through various bioinformatical and machine learning methods, including genomic difference detection, protein-protein interaction network analyses, Lasso regression, univariate logistic regression, Boruta algorithm, and SVM-RFE, RNMT and RBM24 were identified and then included into the random forest (RF) diagnosis model. GSE117999 dataset (control = 10, OA = 10) and clinical samples collected from local hospital (control = 10, OA = 11) were used for external validation. The RF model was a promising tool to diagnose OA in the training dataset (area under curve [AUC] = 1.000, 95% confidence interval [CI] = 1.000-1.000), the GSE117999 cohort (AUC = 0.900, 95% CI = 0.769–1.000), and local samples (AUC = 0.759, 95% CI = 0.568–0.951). Besides, qPCR and Western Blotting experiments showed that RNMT ($P < 0.05$) and RBM24 ($P < 0.01$) were both down-regulated in CHON-001 cells with IL-1β treatment. In all, an RF model to diagnose OA based on RNMT and RBM24 in cartilage tissue was constructed, providing a promising clinical tool and possible cut-in points in molecular mechanism clarification.

INTRODUCTION

As one of the most common diseases worldwide, osteoarthritis (OA) severely affects the life quality of patients and brings a tremendous socioeconomic burden [1]. The pathogenic factors of OA mainly include the alternation of genetics, metabolism, inflammation, and biomechanics, contributing to cartilage destruction and bone fragmentation [2]. However, given the fact that there is no cure for OA with the existing medical treatment except for artificial joint replacement [3], investigating the latent molecular mechanisms to better clarify the etiology and pathogenesis and to develop novel therapeutic agents is urgently demanded.

RNA-binding protein (RBP) is defined as a group of proteins that are able to distinguish and bind to the target RNAs and thus influent these RNAs' transcription, editing, splicing, and other biological processes [4]. RBPs are vital to maintaining body homeostasis, and the disorders of RBPs lead to a series of diseases, such as inflammatory diseases [5],

metabolism-related diseases [6], malignant cancer [7], and genetic diseases [8]. Similar to these situations, RBPs are also involved in the pathological processes of OA. For instance, the RBP PUM1 could alleviate the cartilage destruction caused by OA through down-regulating TLR4 [9]; CPEB1, an RBP binding to the 3′-UTR-cytoplasmic polyadenylation element of the target RNAs, was up-regulated in the cartilage samples of OA patients, and was associated with the severity of the disease [10]; the lack of the RBP Samd4 could lead to the chondrogenesis defects in mice [11]. This evidence proves that RBPs exert non-negligible functions in OA. Nevertheless, our understandings of the roles of RBPs in the initiation and development of OA are far from enough, and the associated researches are at the initial stage for the moment. Seeking more RBPs serving as biomarkers in OA is important and meaningful.

Herein, the present study downloaded the transcriptome sequencing data of the knee cartilage tissue extracted from OA patients from the Gene Expression Omnibus (GEO) since knee OA is the most common form of arthritis. Diverse bioinformatical methods and machine learning algorithms, including genomic difference detection, protein-protein interaction (PPI) network analyses, Lasso regression, univariate logistic regression, Boruta algorithm, and SVM-RFE, were used for feature selection. Then, a random forest (RF) model was developed to diagnose OA. 21 clinical samples from the local hospital and the GSE117999 dataset from GEO were used for external validation. At last, we established an *in vitro* OA model in CHON-001 cells using Interleukin-1β (IL-1β) and measured the expression levels of the screened genes via real-time quantitative PCR (RT-qPCR) and Western Blotting.

MATERIALS AND METHODS

Data collection and processing

1525 RBPs identified by experimental research and high-throughput screening were collected from Stefanie and his colleagues' report [12]. GSE114007 [13], which contained the RNA sequencing (RNA-seq) data of the knee cartilage tissue isolated from 18 healthy control and 20 OA subjects, was downloaded from GEO (https://www.ncbi.nlm.nih.gov/geo/) as the training dataset. The GSE117999 dataset, which was also obtained from GEO and included the transcriptome data of the knee cartilage tissue extracted from 10 control and 10 OA cases, was set as the external validation dataset. To reduce the divergence as possible, we transformed the RNA-seq data into Transcripts-per-Million (TPM) format and then used the sva package in R software (version 3.6.3) to correct the batch effects. The detailed information of these public datasets is displayed in Table 1.

Ethics and clinical specimens

The protocol of the present study has been revised and approved by the Medical Ethics Committee of Yuebei People's Hospital Affiliated to Medical College of Shantou University. All the participants have signed the informed consent. The knee cartilage samples were collected from 11 OA patients (4 female, 7 male, age range 49–73 years) undergoing total knee replacement and 10 subjects (6 female, 4 male, age range 52–71 years) receiving traumatic amputation in the absence of rheumatoid arthritis or OA in Yuebei People's Hospital Affiliated to Medical College of Shantou University between January 12, 2022 and July 28, 2022. The cartilage tissue was then immediately stored in liquid nitrogen for RNA extraction. The diagnosis of OA depends on the criteria recommended by the American College of Rheumatology.

Genomic difference analyses

The limma package in R software was used for genomic divergence detection. The false discovery rate (FDR) < 0.05 and |logFC| > 1 were set as the filtering thresholds.

Gene functional annotation

The gene functional annotation was conducted via the Metascape database [14] (https://metascape.org/) or the clusterProfile package in R. The terms with $P < 0.05$ were considered to be significant.

PPI network construction

The STRING database (https://cn.string-db.org/) was utilized to construct the PPI network with a confidence level of 0.4. Subsequently, the Cytoscape software (version 3.8.0) was used to visualize the network. The importance of the genes in the network was measured by the cytoHubba plug-in [15], which is a widely-used tool to calculate node scores according to various algorithms including MCC, DMNC, Degree, EPC, Bottle neck, EcCentricity, closeness, radiality, betweenness, stress, and clustering coefficient. Here, we chose "Degree" algorithm to measure the importance of the genes, and the Top 20 genes with the highest degree were included in further analysis.

Feature selection via machine learning

To identify the genes significantly associated with OA, we implemented various machine learning algorithms, as previously reported [16]. Lasso regression with

Table 1. The detailed information of the public datasets obtained from GEO.

GEO series	Platform	Experiment type	Tissue	Control/OA	Region
GSE114007	GPL11154, GPL18573	RNA-seq	Human knee cartilage	18/20	USA
GSE117999	GPL20844	Microarray	Human knee cartilage	10/10	USA

Abbreviations: GEO: Gene Expression Omnibus; OA: Osteoarthritis.

Table 2. The primer sequence used in this study.

Gene	Sequence (5′-3′)
RBM24	F: GAACCTGGCATACTTAGGAGCA
	R: AGGTCTTTGTATAAGGGCTGGA
RNMT	F: ATAGCACTTGAGGATGTTCCTGA
	R: ACTACGCTTCTCCAAACCAAC
GAPDH	F: GGAGCGAGATCCCTCCAAAAT
	R: GGCTGTTGTCATACTTCTCATGG

10-fold cross-validation was conducted via the glmnet package. The Boruta algorithm was conducted by the Boruta package, and the features labeled with "Confirmed" were selected. The caret package was used to conduct the SVM-RFE. Last, univariate logistic regression was performed with the rms package, where $P < 0.01$ was statistically significant.

RF model construction

Compared with the conventional modeling method, RF attracted increasing attention due to its high precision and accuracy [17, 18]. Here, we attempted to construct the diagnosis model via RF. The randomForest package in R software was used to develop the RF diagnosis model with ntree = 500 and mtry = 3. The importance of the genes in the RF model was quantified by mean decrease accuracy and mean decrease Gini [16].

Functional co-expression analyses

The Top 20 functionally co-expressed genes of the target gene were queried in the GeneMANIA database (http://genemania.org/) with the Max resultant genes = 20 and the Max resultant attributes = 10. The interaction types included physical interactions, co-expression, prediction, co-localization, genetic interactions, pathways, and shared protein domains. Subsequently, the functional enrichment of these genes was investigated via the clusterProfile package, where Gene Ontology (GO) datasets were selected as the reference.

Cell culture and treatment

CHON-001 cell line was obtained from the American Type Culture Collection (Manassas, VA, USA) and cultured in DMEM (Gibco; Thermo Fisher Scientific, USA) supplemented with 10% FBS (Thermo Fisher Scientific, USA) and 1% penicillin-streptomycin at 37°C with 5% CO_2. CHON-001 cells were incubated with 10 ng/mL IL-1β (Sigma-Aldrich, China) for 48 hours to establish the OA cell model [19, 20].

RT-qPCR

The total RNA of the clinical samples and cells was extracted with the TRIzol reagent (Invitrogen, USA). The cDNA was synthesized by PrimeScript RT Reagent Kit (Takara, China), and the qPCR experiments were conducted via SYBR Premix ExTaq kit (Takara, China). GAPDH was chosen as the internal reference gene, and the $2^{-\Delta\Delta Ct}$ method was used to normalize the gene expression levels. The primer sequence of this study is shown in Table 2.

Western blotting

The cell and tissue samples were lysed with RIPA lysis buffer (Thermo Fisher Scientific, USA) with protease inhibitor (Sigma-Aldrich, USA) on ice. The proteins were transferred onto nitrocellulose membranes (Millipore, USA) after the separation on SDS-PAGE. The membranes were washed using TBS 5 times and then blocked via 5% skimmed milk for 1 hour. Subsequently, the membranes were incubated with primary antibodies overnight at 4°C. Antibodies used in this study: GAPDH (dilution: 1:1000, AC001, ABclonal, China), RNMT (dilution: 1:500, PA5-41778, Thermo Fisher Scientific, USA), and RBM24 (dilution: 1:500, PA5-66881, Thermo Fisher Scientific, USA). Subsequently, the membranes were incubated with secondary anti-rabbit IgG antibodies for 1 hour. MiniChmei Chemiluminescence imager (Sagecreation, China) was used to quantify the protein levels.

Statistical analyses

The statistical analyses of the present study were based on R software (version 3.6.3) and GraphPad Prism 8 (version 8.4.3). The receiver operating characteristic (ROC) curves were plotted and the areas under the curve (AUCs) were calculated via the pROC package. The gene expression levels detected via RT-qPCR were compared with the student's *t*-test. Unless otherwise specified, $P < 0.05$ was considered to be statistically significant. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

RESULTS

Identification of the differentially-expressed RBPs

The workflow of this study is illustrated in Figure 1, and the R codes used in this study were shown in Supplementary File 1. After the genomic difference analysis, a sum of 62 differentially-expressed RBPs were identified between the control and OA samples of the GSE114007 cohort, including 38 up-regulated and 24 down-regulated genes (Supplementary Table 1). The heatmap and the volcano plot used to visualize the genomic difference detection results were displayed in Figure 2A, 2B, respectively. Functional annotation from Metascape revealed that these differentially-expressed RBPs were involved in multiple critical cellular processes, including ncRNA metabolic process (GO term, $P < 0.001$), mRNA metabolic process (GO term, $P < 0.001$), nucleic acid phosphodiester bond hydrolysis (GO term, $P < 0.001$), mRNA surveillance pathway (KEGG term, $P < 0.001$), regulation of mRNA metabolic process (GO term, $P < 0.001$), metabolism of RNA (Reactome term, $P < 0.001$), DNA methylation or demethylation (GO term, $P < 0.001$), antiviral mechanism by IFN-stimulated genes (Reactome term, $P < 0.001$), negative regulation of viral process (GO term, $P < 0.001$), cellular response to oxidative stress (GO term, $P < 0.001$), mRNA processing (WikiPathways term, $P < 0.001$), RNA 3′-end processing (GO term, $P < 0.01$), diseases of programmed cell death

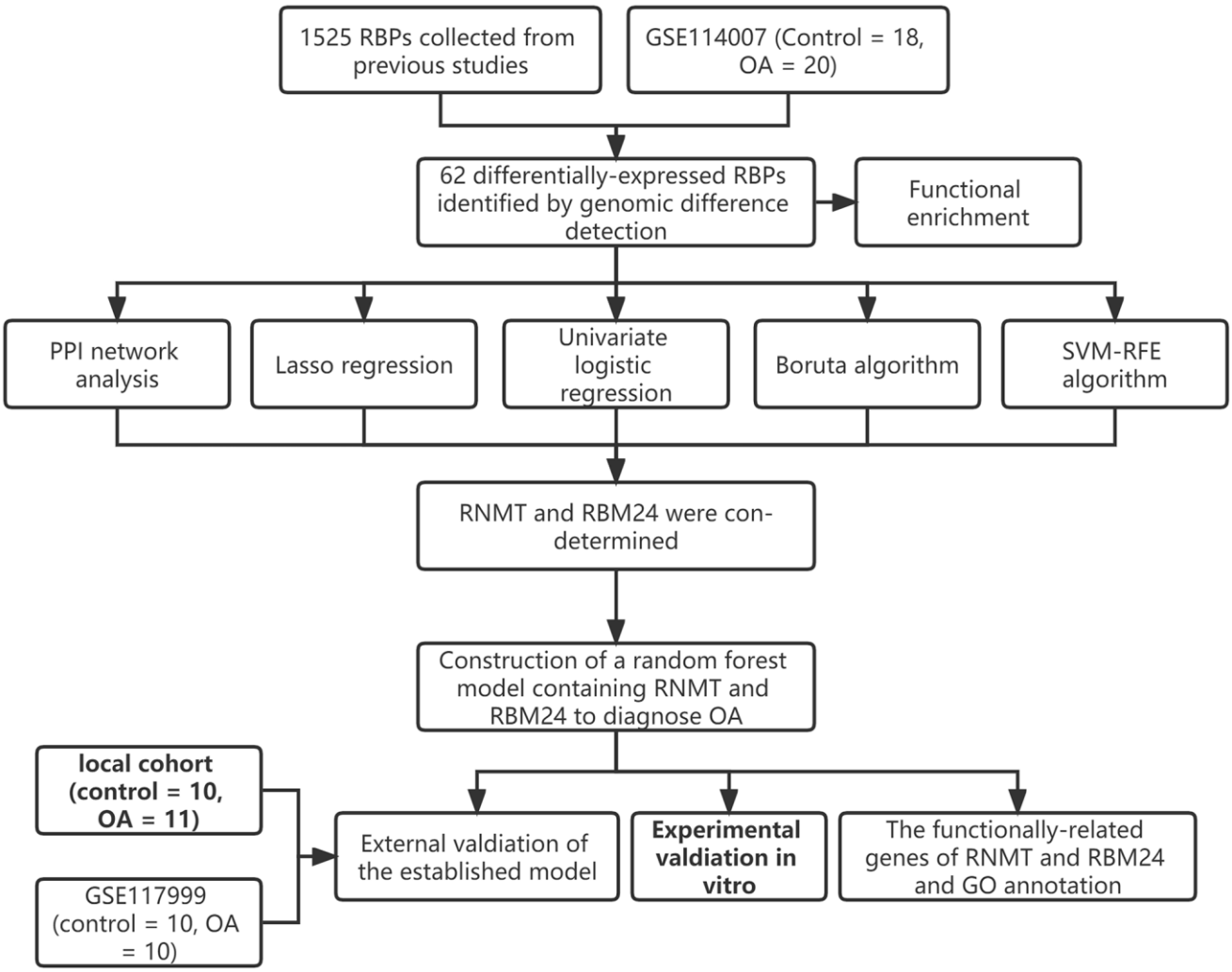


Figure 1. The workflow of the present study.

(Reactome term, $P < 0.01$), blastocyst development (GO term, $P < 0.01$), cellular response to DNA damage stimulus (GO term, $P < 0.01$), skeletal system development (GO term, $P < 0.01$), protein autophosphorylation (GO term, $P < 0.01$), and translation (GO term, $P < 0.01$), uncovering the underlying biological functions of these RBPs in the pathogenesis of OA (Figure 2C).

PPI network analyses

To detect the potential interactions of the 62 genes, we uploaded these genes into the STRING database to construct the PPI network (Figure 3A). Next, the importance of these genes in the network was quantified using the cytoHubba app in Cytoscape (Supplementary Table 2). The Top 20 genes showing the highest degree were chosen for further study, including EIF4A3, DDX28, KHDRBS3, AEN, CLK3, EIF4E2, RBM24, NIP7, ZFP36, NOVA1, CELF2, PABPC4L, CLP1, MYEF2, RNMT, PUS1, XAB2, EXO1, FDXACB, and RPS27 (Figure 3B).

RBM24 and RNMT were identified as potential diagnostic biomarkers to OA

15 of 62 RBPs were determined as significant features to evaluate the possibility of OA development via Lasso regression (Figure 4A). The parameters of these 15 features in the Lasso regression model are shown in Figure 4B. To render the predictive model more concise, we implemented other feature selection methods at the same time. SVM-RFE algorithm identified 37 variables significantly associated with the outcomes (Figure 4C). 47 genes were determined via univariate logistic regression with $P < 0.01$ filtering (Supplementary Table 3), and the Boruta algorithm helped to identify 37 genes (Figure 4D). Ultimately, RNA Guanine-7 Methyltransferase (RNMT) and RNA Binding Motif Protein 24 (RBM24) were

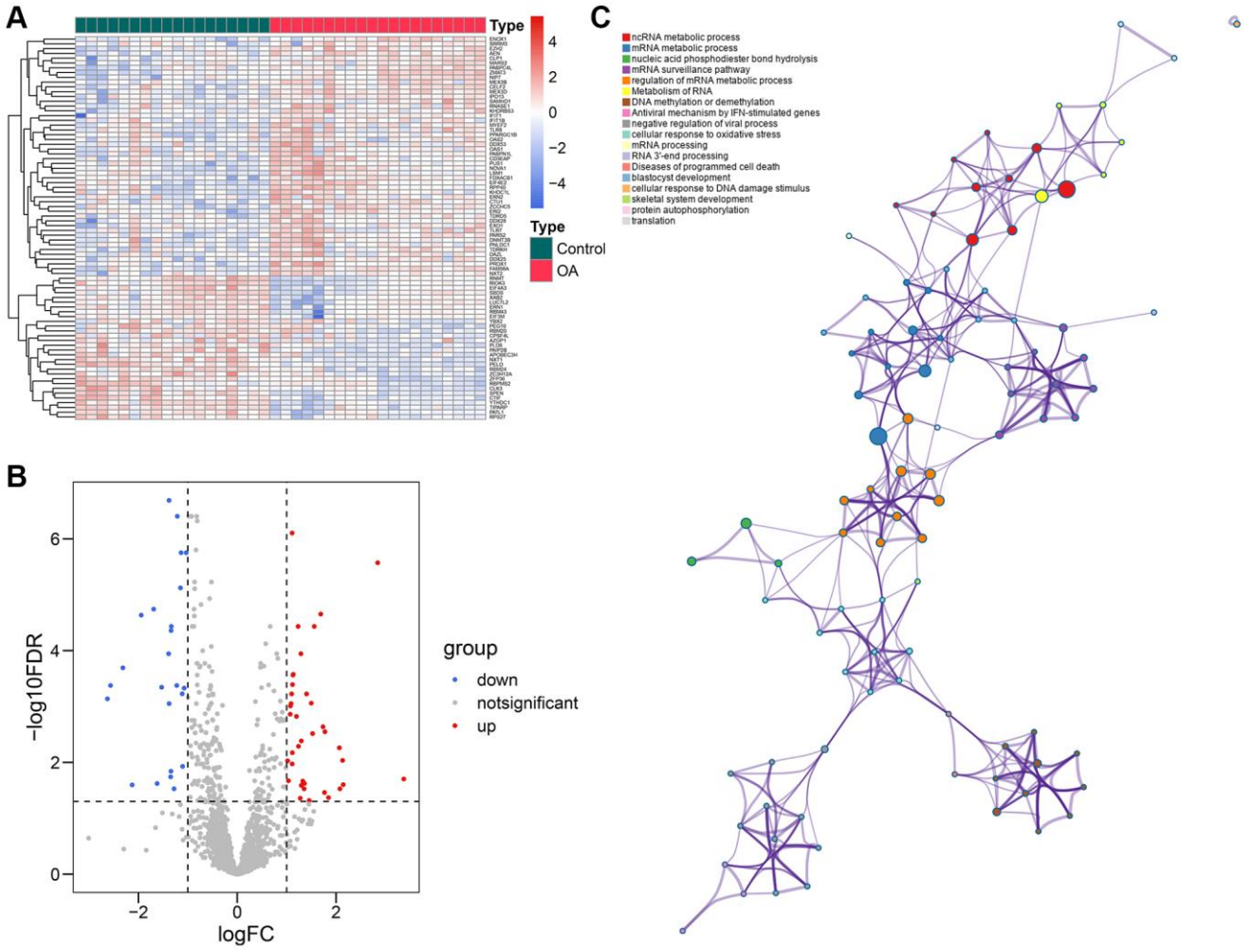


Figure 2. Identification of the differentially-expressed RBPs. (**A**, **B**) The volcano plot (**A**) and the heatmap (**B**) showed that a total of 62 differentially-expressed RBPs were detected between the control and OA samples. (**C**) The functional annotation of the 62 genes via the Metascape database. Abbreviations: RBP: RNA-binding protein; OA: osteoarthritis.

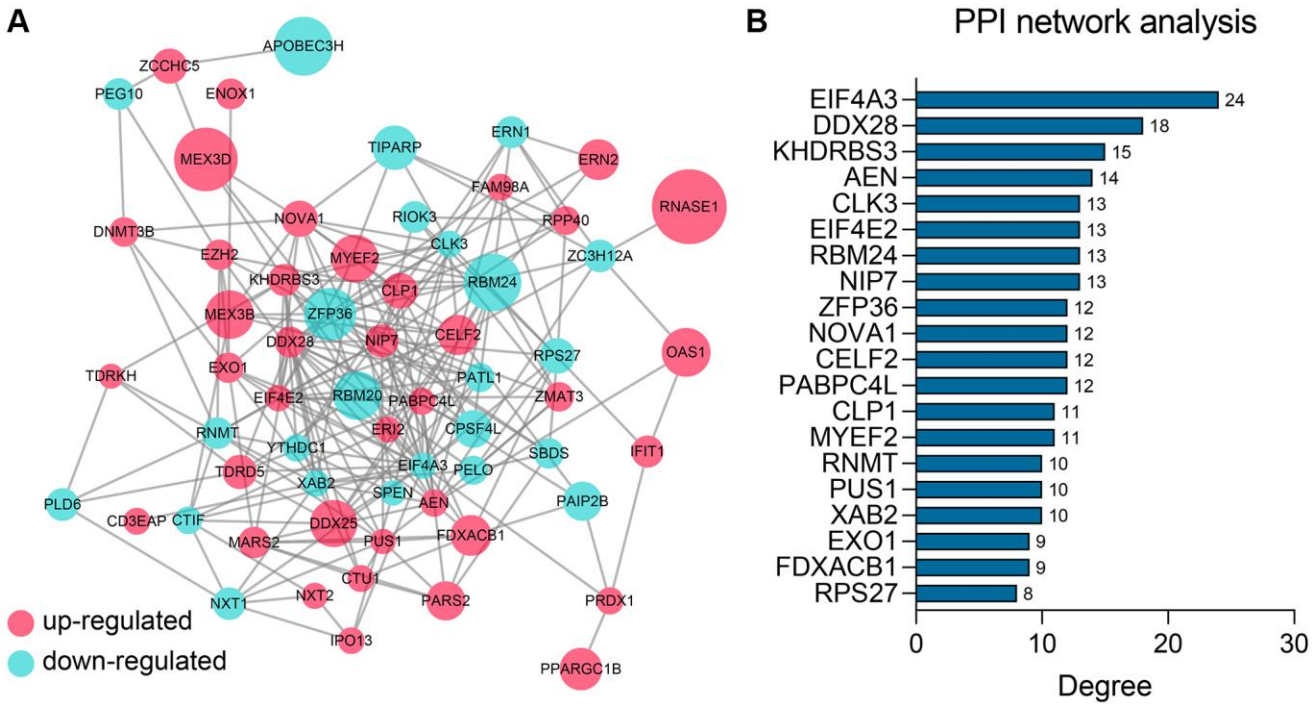


Figure 3. PPI network construction and analysis. (**A**) The PPI network of the 62 differentially-expressed RBPs. (**B**) The Top 20 genes ranked by degree in the network. Abbreviation: PPI: protein-protein interaction.

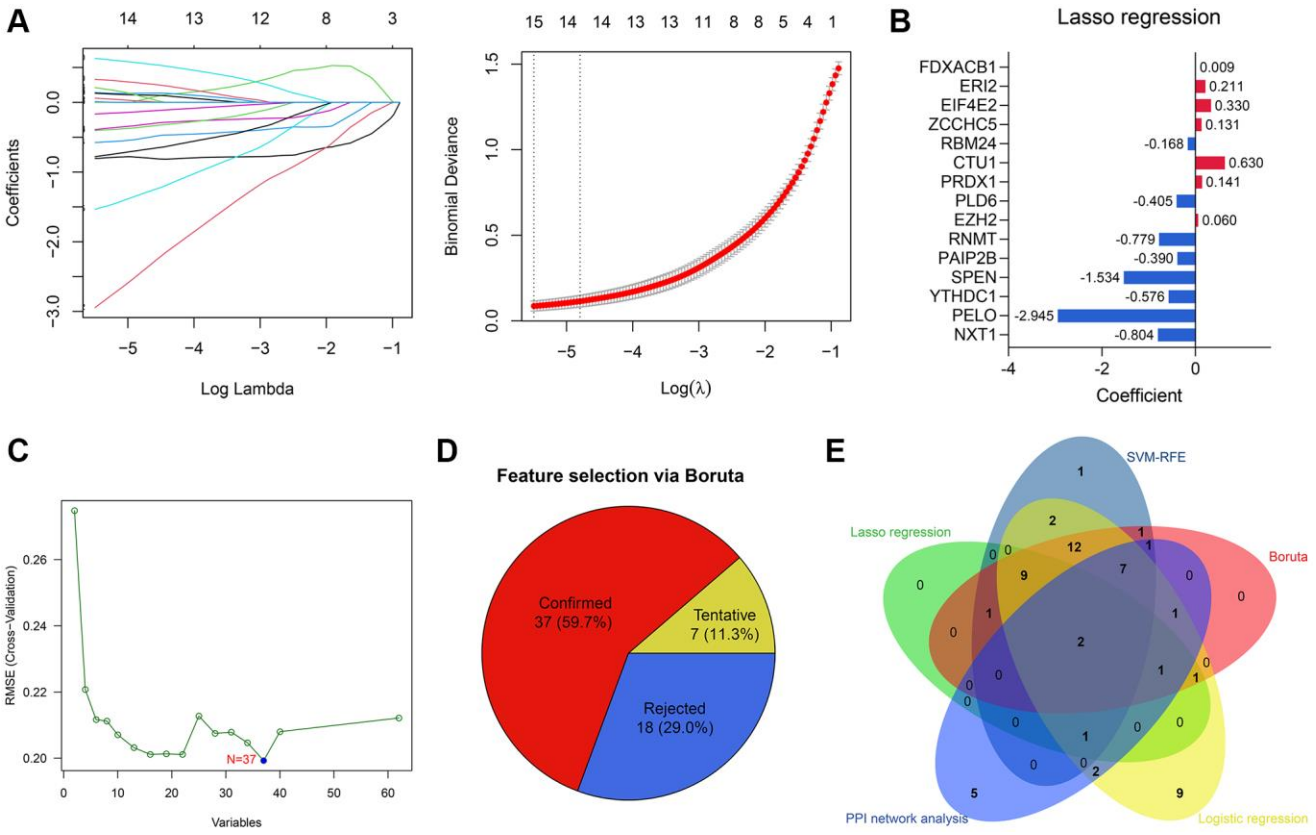


Figure 4. Feature selection via machine learning algorithms. (**A**) 15 genes were determined by Lasso regression. (**B**) The parameters of the variables in the Lasso regression model. (**C**) 37 genes were identified by SVM-RFE algorithm. (**D**) Boruta algorithm helped to select 37 genes. (**E**) RNMT and RBM24 were con-determined by the PPI network analysis and machine learning algorithms.

con-determined by the feature selection algorithms and PPI network analysis (Figure 4E) and then included in the diagnosis model. To clarify the specificity and sensitivity of RNMT and RBM24, we then performed the ROC analyses in the training cohort. The optimal cut-off value for RBM24 was 0.500, where its specificity and sensitivity were 0.944 and 0.800, respectively (Supplementary Figure 1A). The best cut-off for RNMT was 2.917, and its specificity and sensitivity were 0.778 and 0.900, respectively (Supplementary Figure 1B).

Construction and external validation of the RF model

Based on the expressions of RBM24 and RNMT, an RF diagnosis model was developed, and the modeling parameters are stated above. The RF model exhibited high efficacy to distinguish the PA samples from the control cases, both in the training dataset (AUC = 1.000, 95% CI = 1.000–1.000, Figure 5A) and the GSE117999 dataset (AUC = 0.900, 95% CI = 0.769–1.000, Figure 5B). The confusion matrices in the training dataset and the GSE117999 dataset were displayed in Figure 5C and Figure 5D, respectively.

Given the fact that the GSE117999 dataset has not been publicly published, we collected 10 control and 11 OA samples from the local hospital to re-confirm the reliability of the model. The baseline clinicopathological parameters of the training cohort and the local cohort were displayed in Table 3. Since the GSE117999 dataset has not been published, the

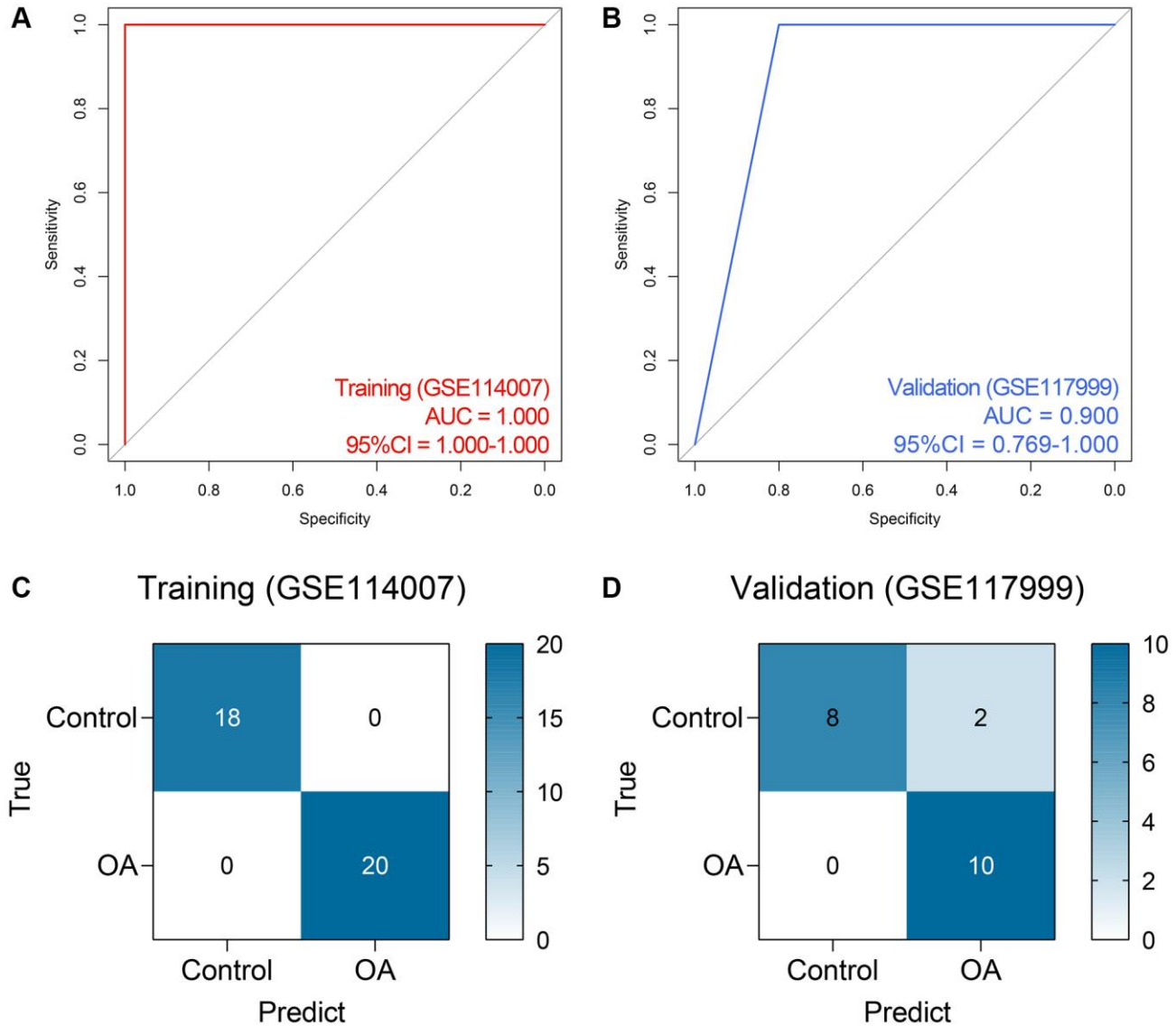


Figure 5. The performance of the random forest model in the public datasets. (**A**, **B**) The ROC analysis of the diagnosis model in the training (**A**) and external validation (**B**) datasets. (**C**, **D**) The confusion matrices of the model in the training (**C**) and external validation (**D**) datasets. Abbreviations: ROC: receiver operating characteristic.

Table 3. The baseline clinical traits of the training cohort and the local cohort.

Parameters	GSE114007		Local cohort	
	Control (*n* = 18)	OA (*n* = 20)	Control (*n* = 10)	OA (*n* = 11)
Age	–	–	56 ± 8.2	62 ± 7.4
Gender				
Male	13 (72.2%)	8 (40.0%)	4 (40.0%)	7 (63.6%)
Female	5 (27.8%)	12 (60.0%)	6 (60.0%)	4 (36.4%)
BMI	32.4 ± 8.0	30.7 ± 8.1	29.6 ± 9.1	33.5 ± 7.4
Kellgren-Lawrence Grade				
I	–	–	–	0 (0.0%)
II	–	–	–	0 (0.0%)
III	–	–	–	2 (18.2%)
IV	–	–	–	9 (81.8%)
Varus Deformity	–	–	–	10 (90.9%)
Valgus Deformity	–	–	–	1 (9.1%)

Abbreviation: BMI: body mass index. The data is presented as *n* (%) or mean ± standard deviation.

clinical information of the GSE117999 cohort was unavailable. The gene expression levels of RBM24 and RNMT were measured via RT-qPCR since the predictive model was based on the RNA expression values. The raw CT values of the genes in the clinical samples were exhibited in Supplementary Table 4. Compared with the control samples, RBM24 ($P < 0.05$, Figure 6A) and RNMT ($P < 0.05$, Figure 6B) were both down-regulated in the OA samples. The RF model could also diagnose OA in the local cohort to some extent (AUC = 0.759, 95% CI = 0.568–0.951, Figure 6C, 6D showed the corresponding confusion matrix. The RF model exhibited moderate diagnosis performance in the local cohort compared with that in the public datasets, and the different gene expression detection platforms (RT-qPCR vs. high-throughput sequencing) and the relatively small sample size might account for this divergence. Other assessment indexes, including accuracy, precision, recall, F-measure, sensitivity, specificity, positive predictive value, and negative predictive value, of the RF model in each cohort were shown in Table 4.

RBM24 and RNMT were associated with genesis of OA

The ROC analyses revealed the predictive ability of RBM24 and RNMT to OA in the training dataset (RNMT, AUC = 0.906; RBM24, AUC = 0.889; Figure 7A), the GSE117999 dataset (RNMT, AUC = 0.840; RBM24, AUC = 0.590; Figure 7B), and the local cohort (RNMT, AUC = 0.855; RBM24, AUC = 0.736; Figure 7C). We observed that except for RBM24 in the GSE117999 cohort, the variables in these cohorts all exhibited high diagnosis values (AUC > 0.7). At the same time, the mean decrease accuracy and mean decrease Gini of RNMT were superior to those of RBM24 (Figure 7D), suggesting that RNMT was a relatively more reliable biomarker than RBM24. At last, we constructed the OA model *in vitro* and detected the RNA and protein levels of RNMT and RBM24. RT-qPCR and Western Blotting indicated that RNMT ($P < 0.05$) and RBM24 ($P < 0.01$) were both down-regulated in the OA model group, as shown in Figure 7E, 7F respectively. This evidence demonstrated that RBM24 and RNMT both serve as critical biomarkers for OA and are associated with OA development.

The functional annotation of RBM24 and RNMT

The Top 20 genes most relevant to RBM24 and RNMT were illustrated in Figure 8A and Figure 8B, respectively. Gene Ontology (GO) analysis indicated that RBM24 and its associated genes were mainly enriched in muscle development and differentiation and RNA metabolism, splicing, and processing (Figure 8C), and RNMT and its correlated genes mainly participated in RNA synthesis and modification, ATPase activity, and serine/threonine kinase activity (Figure 8D). These results might provide clues to elucidate the roles RBM24 and RNMT play in OA.

DISCUSSION

Currently, the diagnosis of OA mainly depends on the clinical presentations and imaging examination, often

Table 4. The predictive performance of the RF model in each cohort.

Cohort	Accuracy	Precision	Recall	F-measure	Sensitivity	Specificity	Positive predictive value	Negative predictive value
GSE114007	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GSE117999	0.900	0.833	1.000	0.909	1.000	0.800	0.833	1.000
Local Cohort	0.762	0.750	0.818	0.783	0.818	0.700	0.750	0.778

leading to delayed diagnosis and missing the optimal time to intervene [21]. Therefore, many efforts have been devoted to seeking reliable diagnosis biomarkers in recent years, and some novel molecules as OA diagnosis markers have been reported, such as ATF3 [22], Apolipoprotein D [23], and CXCL13 [24]. The rapid development of genomic sequencing technology and big-data analysis methods represented by machine learning generates new opportunities to disclose novel biomarkers and to develop novel diagnostic tools, and

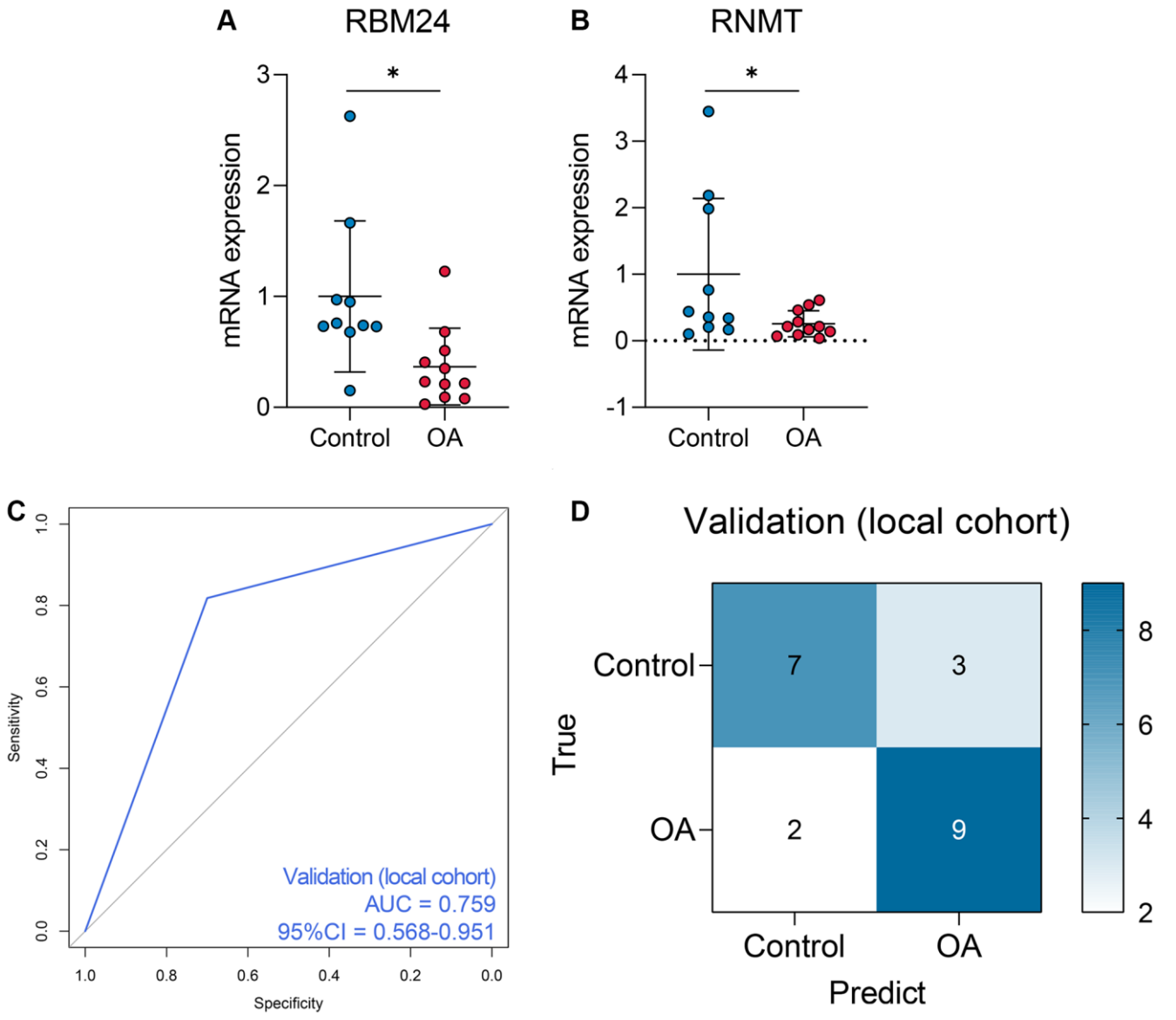


Figure 6. The performance of the random forest model in the local cohort. (**A**, **B**) The qPCR experiments indicated that RNMT and RBM24 were both down-regulated in the knee cartilage tissue extracted from OA patients. (**C**, **D**) The ROC analysis (**C**) and the confusion matrix (**D**) of the random forest model in the local cohort.

many achievements have been obtained in OA [25–27]. Nevertheless, the reports about OA diagnosis models based on machine algorithms are relatively rare at the moment.

As important posttranscriptional regulators, RBPs are capable of modulating RNA metabolism and thus affect the expression levels of proteins [28]. The huge effects of RBPs on gene expression enlighten the investigators to study the roles RBPs play in OA, and some RBPs, such as GNL3 [29], SND1 [30], and ZFP36L1 [31], have been verified as critical regulators in the pathogenesis of OA. However, the number of RBP-related research in OA is still limited for the moment.

Herein, the present study collected 1525 RBPs from previous reports and detected their expression levels in the knee cartilage tissue isolated from control and OA subjects. A sum of 62 differentially-expressed RBPs were identified, RNMT and RBM24 of which were determined as the significant biomarkers for OA through multiple machine learning algorithms and bioinformatical analyses. To achieve better predictive performance, we constructed an RF model using the expression values of RNMT and RBM24. Importantly, the RF model has been externally validated in another public dataset and the clinical samples collected from the local hospital. Besides, the *in vitro* OA model was constructed in CHON-001 cells, which were treated with IL-1β to mimic OA. The RNA and protein levels of RNMT and RBM24 significantly decreased in the OA model group, suggesting RNMT and RBM24 might be important regulators in the development of OA.

We firstly reported that RNMT and RBM24 acted as potential biomarkers for OA and conducted the preliminary verification in the OA cell model and clinical samples. RNMT, a regulatory factor for 7-methylguanosine mRNA capping, exerts inhibitory effects against 5′-exonucleases and promotes RNAs' export and translation [32]. It is worth mentioning that the polarity of RNMT ligands is low, rendering the compounds targeting this molecule easier to cross the

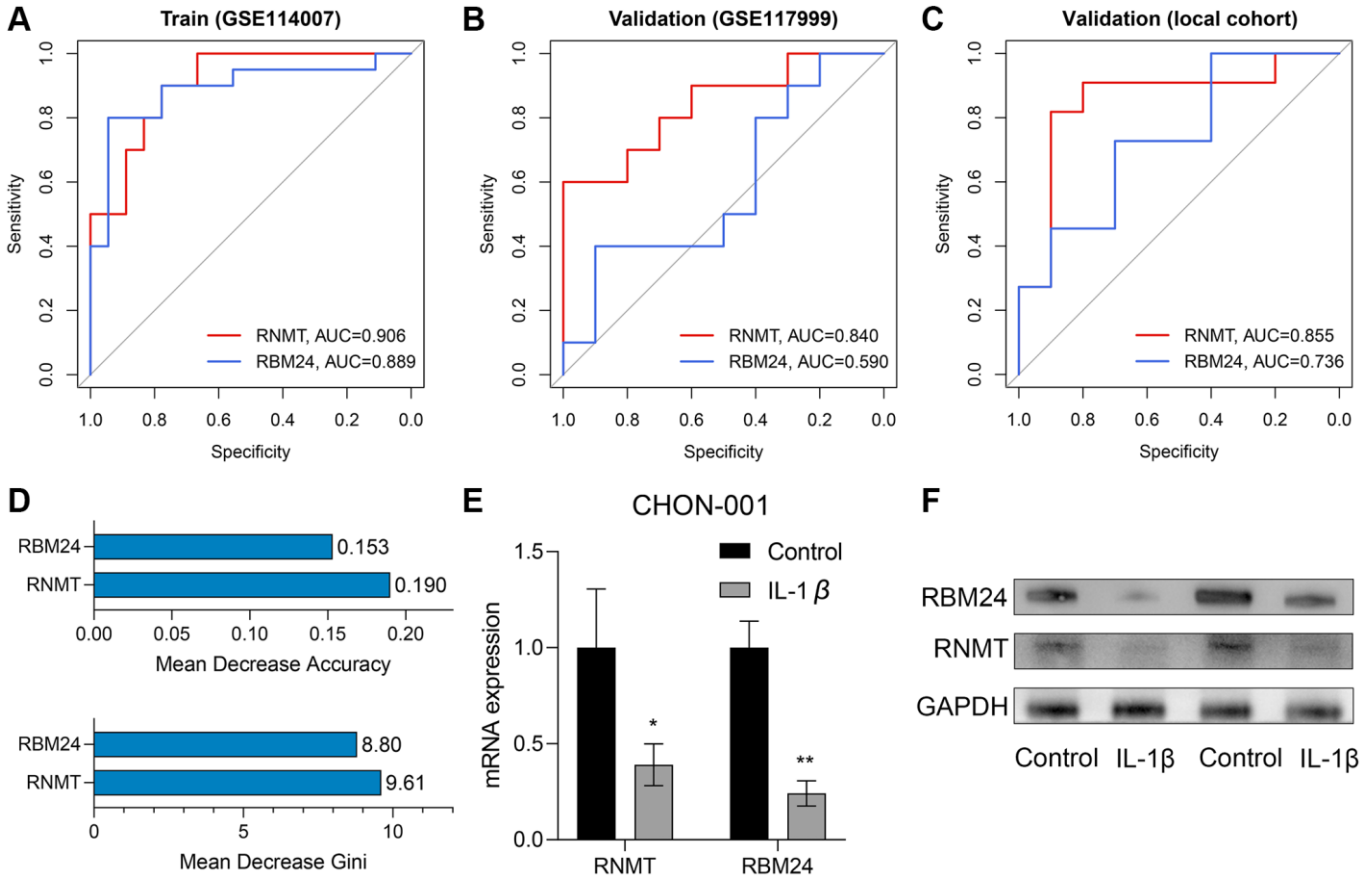


Figure 7. RNMT and RBM24 were associated with genesis of OA. (**A**–**C**) The diagnosis value of RNMT and RBM24 in the training cohort (**A**), the GSE117999 cohort (**B**), and the local cohort (**C**). (**D**) The mean decrease accuracy (up) and the mean decrease Gini (bottom) of RNMT and RBM24 in the random forest model. (**E**, **F**) The qPCR experiments (**E**) and Western Blotting (**F**) displayed that RNMT and RBM24 were both down-regulated in the CHON-001 cells treated with 10 ng/mL IL-1β.

plasma membrane [33]. Alison et al., reported that RNMT was an essential mediator for T cell activation [34]; meanwhile, many studies have demonstrated that T cells exhibit high numbers in OA samples compared with the control samples [35], but their correlations need to be further clarified. RBM24, a highly conserved RBP, is an important regulator for cell differentiation and cellular homeostasis [36]. RBM24 was reported to be recruited in stress granules, which were formed in cells under stress, and to protect its target RNAs [37], which might be the underlying biological function of RBM24 in OA development. However, more direct experimental evidence supporting these assumptions in OA is demanded in future studies.

The shortcomings of this study should be stated. First, although the RF model was validated in the clinical samples from the local hospital, a prospective, large-scale, and multi-center clinical trial would be helpful to better clarify its usefulness. Additionally, the roles of RNMT, RBM24, and the RF model in OA's early detection or prognosis prediction are unknown, which is the research direction in the future. Second, more experimental exploration should be performed to elucidate the biological functions and underlying mechanisms of RNMT and RBM24 in OA. Third, although CHON-001 cells treated with IL-1β were widely used for *in vitro* OA model construction, it should be emphasized that CHON-001 cells were isolated from the cartilage tissue of an embryo (age 18 weeks) and exhibited a fibroblast-like morphology, leading to different characteristics compared with primary chondrocytes. For instance, adult chondrocytes rarely divide throughout life, but CHON-001 cells are proliferative cells. This is the limitation of the methodology in this study.

In conclusion, an RF model based on RNMT and RBM24 was established to diagnose OA, which was externally validated in public datasets, local clinical samples, and *in vitro* cell experiments.

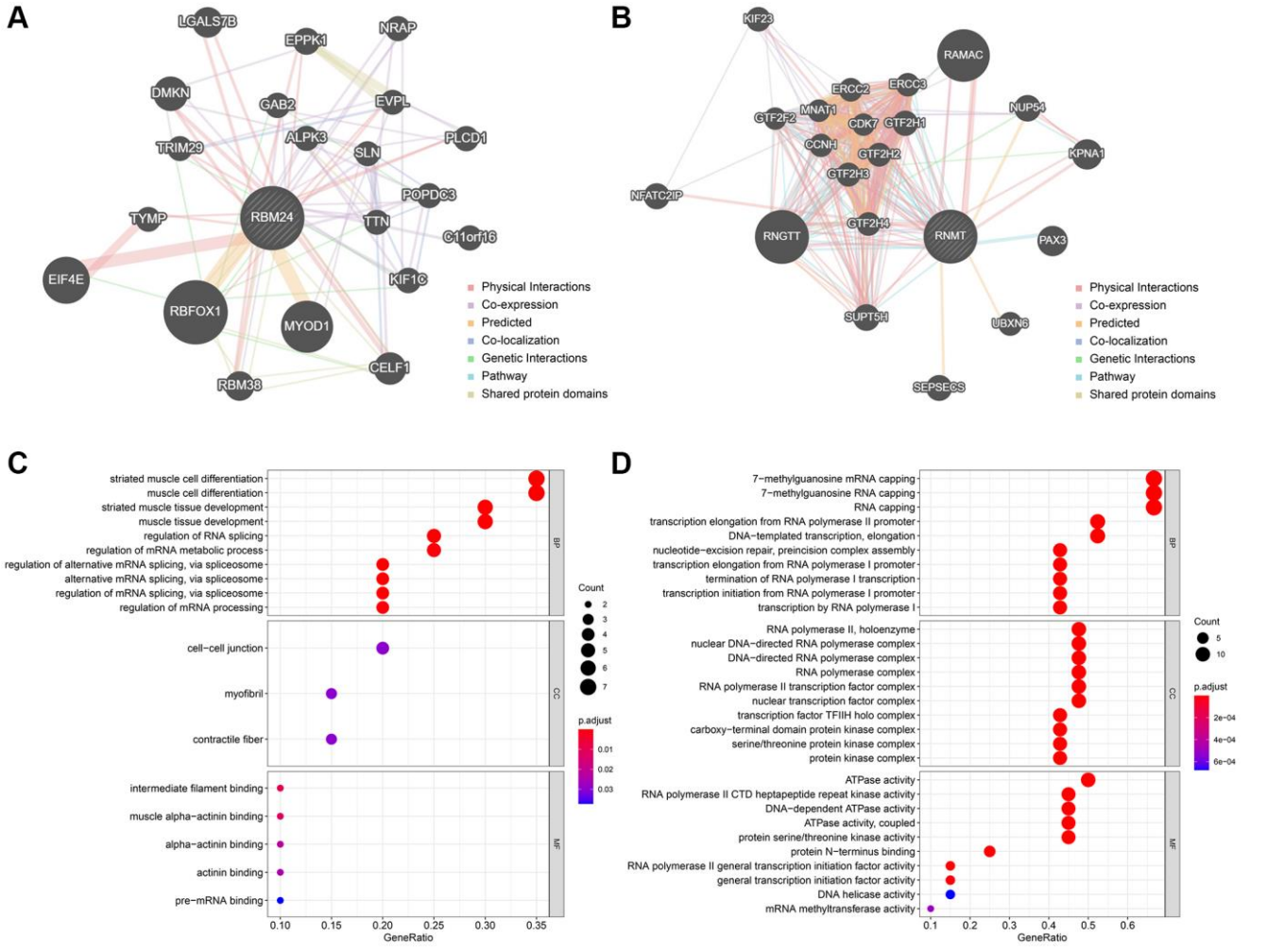


Figure 8. The functionally-associated genes and GO enrichment. (**A**, **B**) The Top 20 genes associated with RBM24 (**A**) and RNMT (**B**). GO functional annotation of the associated genes of RBM24 (**C**) and RNMT (**D**). Abbreviation: GO: gene ontology.

AUTHOR CONTRIBUTIONS

WHY designed the whole study, provided financial support, and conducted the experiments. YL developed the algorithm and conducted the experiments. XY wrote the original draft. JWZ helped to collect the clinical samples and edited the original manuscript. All authors reviewed the manuscript and approved the submission version.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

ETHICAL STATEMENT AND CONSENT

The protocol of the present study has been revised and approved by the Medical Ethics Committee of Yuebei People's Hospital Affiliated to Medical College of Shantou University. All the participants have signed the informed consent.

REFERENCES

1. Katz JN, Arant KR, Loeser RF. Diagnosis and Treatment of Hip and Knee Osteoarthritis: A Review. JAMA. 2021; 325:568–78. https://doi.org/10.1001/jama.2020.22171 PMID:33560326

2. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. Lancet. 2019; 393:1745–59. https://doi.org/10.1016/S0140-6736(19)30417-9 PMID:31034380

3. Martel-Pelletier J, Barr AJ, Cicuttini FM, Conaghan PG, Cooper C, Goldring MB, Goldring SR, Jones G, Teichtahl AJ, Pelletier JP. Osteoarthritis. Nat Rev Dis Primers. 2016; 2:16072. https://doi.org/10.1038/nrdp.2016.72 PMID:27734845

4. Hentze MW, Castello A, Schwarzl T, Preiss T. A brave new world of RNA-binding proteins. Nat Rev Mol Cell Biol. 2018; 19:327–41. https://doi.org/10.1038/nrm.2017.130 PMID:29339797

5. Turner M, Díaz-Muñoz MD. RNA-binding proteins control gene expression and cell fate in the immune system. Nat Immunol. 2018; 19:120–9. https://doi.org/10.1038/s41590-017-0028-4 PMID:29348497

6. de Bruin RG, Rabelink TJ, van Zonneveld AJ, van der Veer EP. Emerging roles for RNA-binding proteins as effectors and regulators of cardiovascular disease. Eur Heart J. 2017; 38:1380–8. https://doi.org/10.1093/eurheartj/ehw567 PMID:28064149

7. Mir C, Garcia-Mayea Y, LLeonart ME. Targeting the "undruggable": RNA-binding proteins in the spotlight in cancer therapy. Semin Cancer Biol. 2022; 86:69–83. https://doi.org/10.1016/j.semcancer.2022.06.008 PMID:35772609

8. Gebauer F, Schwarzl T, Valcárcel J, Hentze MW. RNA-binding proteins in human genetic disease. Nat Rev Genet. 2021; 22:185–98. https://doi.org/10.1038/s41576-020-00302-y PMID:33235359

9. Yoon DS, Lee KM, Choi Y, Ko EA, Lee NH, Cho S, Park KH, Lee JH, Kim HW, Lee JW. TLR4 downregulation by the RNA-binding protein PUM1 alleviates cellular aging and osteoarthritis. Cell Death Differ. 2022; 29:1364–78. https://doi.org/10.1038/s41418-021-00925-6 PMID:35034101

10. Li L, Lan J, Ye Y, Yang B, Yang X, Cai Z. CPEB1 Expression Correlates with Severity of Posttraumatic Ankle Osteoarthritis and Aggravates Catabolic Effect of IL-1β on Chondrocytes. Inflammation. 2019; 42:628–36. https://doi.org/10.1007/s10753-018-0920-6 PMID:30411210

11. Niu N, Xiang JF, Yang Q, Wang L, Wei Z, Chen LL, Yang L, Zou W. RNA-binding protein SAMD4 regulates skeleton development through translational inhibition of Mig6 expression. Cell Discov. 2017; 3:16050. https://doi.org/10.1038/celldisc.2016.50 PMID:28163927

12. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. Nat Rev Genet. 2014; 15:829–45. https://doi.org/10.1038/nrg3813 PMID:25365966

13. Fisch KM, Gamini R, Alvarez-Garcia O, Akagi R, Saito M, Muramatsu Y, Sasho T, Koziol JA, Su AI, Lotz MK. Identification of transcription factors responsible for dysregulated networks in human osteoarthritis cartilage by global gene expression analysis. Osteoarthritis Cartilage. 2018; 26:1531–8. https://doi.org/10.1016/j.joca.2018.07.012 PMID:30081074

14. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019; 10:1523. https://doi.org/10.1038/s41467-019-09234-6 PMID:30944313

15. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014 (Suppl 4); 8:S11. https://doi.org/10.1186/1752-0509-8-S4-S11 PMID:25521941

16. Zhou R, Liang J, Chen Q, Tian H, Yang C, Liu C. A 3-Gene Random Forest Model to Diagnose Non-obstructive Azoospermia Based on Transcription Factor-Related Henes. Reprod Sci. 2022. [Epub ahead of print]. https://doi.org/10.1007/s43032-022-01008-8 PMID:35715550

17. Everson TM, Lyons G, Zhang H, Soto-Ramírez N, Lockett GA, Patil VK, Merid SK, Söderhäll C, Melén E, Holloway JW, Arshad SH, Karmaus W. DNA methylation loci associated with atopy and high serum IgE: a genome-wide application of recursive Random Forest feature selection. Genome Med. 2015; 7:89. https://doi.org/10.1186/s13073-015-0213-8 PMID:26292806

18. Song JK, Zhang Y, Fei XY, Chen YR, Luo Y, Jiang JS, Ru Y, Xiang YW, Li B, Luo Y, Kuai L. Classification and biomarker gene selection of pyroptosis-related gene expression in psoriasis using a random forest algorithm. Front Genet. 2022; 13:850108. https://doi.org/10.3389/fgene.2022.850108 PMID:36110207

19. Ou D, Liu S, Tong C, Tan H, Yang Y, He C. LIM mineralization protein-1 inhibits IL-1β-induced human chondrocytes injury by altering the NF-κB and MAPK/JNK pathways. Exp Ther Med. 2022; 23:61. https://doi.org/10.3892/etm.2021.10983 PMID:34934432

20. Wang X, Fan J, Ding X, Sun Y, Cui Z, Liu W. Tanshinone I Inhibits IL-1β-Induced Apoptosis, Inflammation And Extracellular Matrix Degradation In Chondrocytes CHON-001 Cells And Attenuates Murine Osteoarthritis. Drug Des Devel Ther. 2019; 13:3559–68. https://doi.org/10.2147/DDDT.S216596 PMID:31686786

21. Xiao S, Chen L. The emerging landscape of nanotheranostic-based diagnosis and therapy for osteoarthritis. J Control Release. 2020; 328:817–33. https://doi.org/10.1016/j.jconrel.2020.11.007 PMID:33176171

22. Yang J, Fan Y, Liu S. ATF3 as a potential diagnostic marker of early-stage osteoarthritis and its correlation with immune infiltration through bioinformatics analysis. Bone Joint Res. 2022; 11:679–89. https://doi.org/10.1302/2046-3758.119.BJR-2022-0075.R1 PMID:36082523

23. Li B, Zhan H, Luo J, Wang X, Cao T, Wei B. A novel serological biomarker are associated with disease severity in patients with osteoarthritis. J Bone Miner Metab. 2022; 40:1007–13. https://doi.org/10.1007/s00774-022-01364-0 PMID:36036835

24. Hu X, Ni S, Zhao K, Qian J, Duan Y. Bioinformatics-Led Discovery of Osteoarthritis Biomarkers and Inflammatory Infiltrates. Front Immunol. 2022; 13:871008. https://doi.org/10.3389/fimmu.2022.871008 PMID:35734177

25. You R, Liu S, Tan J. Screening and identification of osteoarthritis related differential genes and construction of a risk prognosis model based on bioinformatics analysis. Ann Transl Med. 2022; 10:444. https://doi.org/10.21037/atm-22-1135 PMID:35571384

26. Han Y, Wu J, Gong Z, Zhou Y, Li H, Wang B, Qian Q. Identification and development of a novel 5-gene diagnostic model based on immune infiltration analysis of osteoarthritis. J Transl Med. 2021; 19:522. https://doi.org/10.1186/s12967-021-03183-9 PMID:34949204

27. Jiang L, Zhou Y, Shen J, Chen Y, Ma Z, Yu Y, Chu M, Qian Q, Zhuang X, Xia S. RNA Sequencing Reveals LINC00167 as a Potential Diagnosis Biomarker for Primary Osteoarthritis: A Multi-Stage Study. Front Genet. 2021; 11:539489. https://doi.org/10.3389/fgene.2020.539489 PMID:33519887

28. Li W, Deng X, Chen J. RNA-binding proteins in regulating mRNA stability and translation: roles and mechanisms in cancer. Semin Cancer Biol. 2022; 86:664–77. https://doi.org/10.1016/j.semcancer.2022.03.025 PMID:35381329

29. Zhu Z, Xie J, Manandhar U, Yao X, Bian Y, Zhang B. RNA binding protein GNL3 up-regulates IL24 and PTN to promote the development of osteoarthritis. Life Sci. 2021; 267:118926. https://doi.org/10.1016/j.lfs.2020.118926 PMID:33358901

30. Lv M, Cai Y, Hou W, Peng K, Xu K, Lu C, Yu W, Zhang W, Liu L. The RNA-binding protein SND1 promotes the degradation of GPX4 by destabilizing the HSPA5 mRNA and suppressing HSPA5 expression, promoting ferroptosis in osteoarthritis chondrocytes. Inflamm Res. 2022; 71:461–72. https://doi.org/10.1007/s00011-022-01547-5 PMID:35320827

31. Son YO, Kim HE, Choi WS, Chun CH, Chun JS. RNA-binding protein ZFP36L1 regulates osteoarthritis by modulating members of the heat shock protein 70 family. Nat Commun. 2019; 10:77. https://doi.org/10.1038/s41467-018-08035-7 PMID:30622281

32. Bueren-Calabuig JA, G Bage M, Cowling VH, Pisliakov AV. Mechanism of allosteric activation of human mRNA cap methyltransferase (RNMT) by RAM: insights from accelerated molecular dynamics simulations. Nucleic Acids Res. 2019; 47:8675–92. https://doi.org/10.1093/nar/gkz613 PMID:31329932

33. Varshney D, Petit AP, Bueren-Calabuig JA, Jansen C, Fletcher DA, Peggie M, Weidlich S, Scullion P, Pisliakov AV, Cowling VH. Molecular basis of RNA guanine-7 methyltransferase (RNMT) activation by RAM. Nucleic Acids Res. 2016; 44:10423–36. https://doi.org/10.1093/nar/gkw637 PMID:27422871

34. Galloway A, Kaskar A, Ditsova D, Atrih A, Yoshikawa H, Gomez-Moreira C, Suska O, Warminski M, Grzela R, Lamond AI, Darzynkiewicz E, Jemielity J, Cowling VH. Upregulation of RNA cap methyltransferase RNMT drives ribosome biogenesis during T cell activation. Nucleic Acids Res. 2021; 49:6722–38. https://doi.org/10.1093/nar/gkab465 PMID:34125914

35. Woodell-May JE, Sommerfeld SD. Role of Inflammation and the Immune System in the Progression of Osteoarthritis. J Orthop Res. 2020; 38:253–7. https://doi.org/10.1002/jor.24457 PMID:31469192

36. Grifone R, Shao M, Saquet A, Shi DL. RNA-Binding Protein Rbm24 as a Multifaceted Post-Transcriptional Regulator of Embryonic Lineage Differentiation and Cellular Homeostasis. Cells. 2020; 9:1891. https://doi.org/10.3390/cells9081891 PMID:32806768

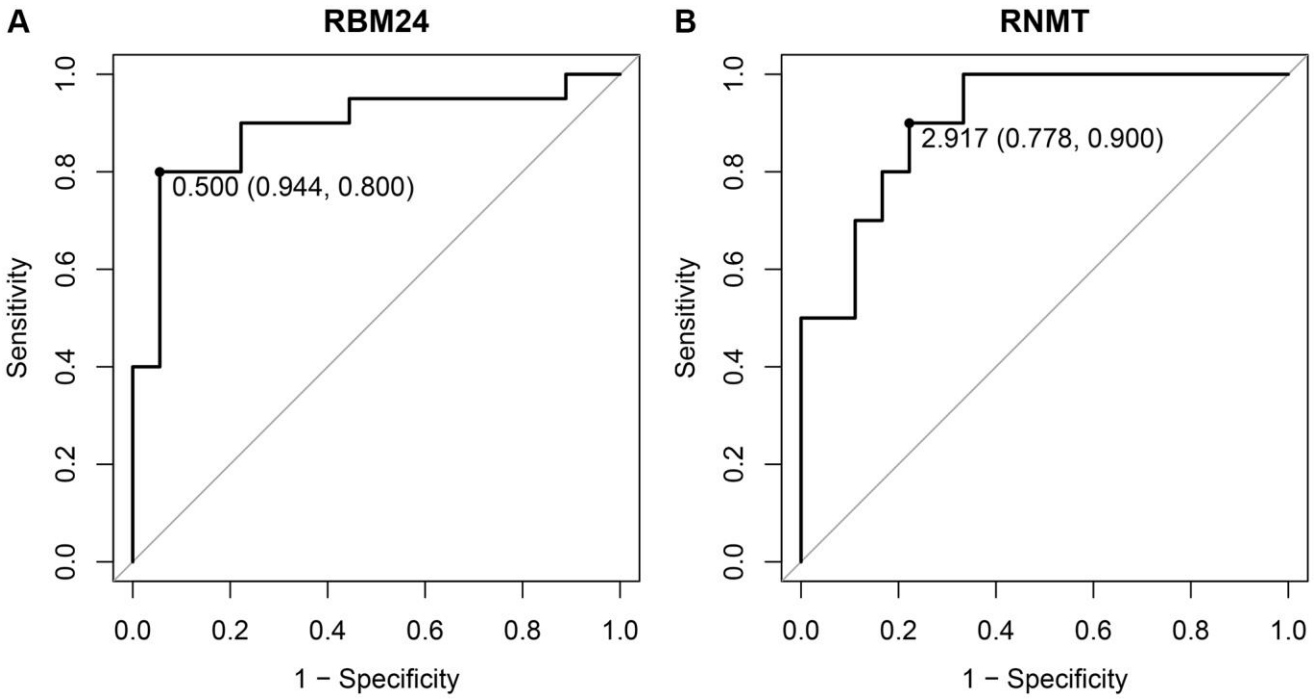
37. Wang Y, Li W, Zhang C, Peng W, Xu Z. RBM24 is localized to stress granules in cells under various stress conditions. Biochem Biophys Res Commun. 2022; 608:96–101. https://doi.org/10.1016/j.bbrc.2022.03.160 PMID:35395551

SUPPLEMENTARY MATERIALS

Please browse Full Text version to see the data of Supplementary File 1.

Supplementary File 1. The R codes used in this study.

Supplementary Figure



Supplementary Figure 1. The optimal cut-off values and their corresponding specificity and sensitivity of RBM24 (**A**) and RNMT (**B**).

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Table 1.

Supplementary Table 1. The genomic difference of the RBPs between control and OA samples.

Supplementary Table 2. The importance of the genes in the PPI network.

node_name	MCC	DMNC	MNC	Degree	EPC	Bottle Neck	Ec Centricity	Closeness	Radiality	Betweenness	Stress	Clustering Coefficient
RNASE1	1	0	1	1	4.847	1	0.2	19.71667	3.77049	0	0	0
PPARGC1B	1	0	1	1	3.284	1	0.16667	18.68333	3.60656	0	0	0
PLD6	5	0.30898	3	4	14.959	1	0.2	24.65	4.29508	8.83413	34	0.33333
RIOK3	7	0.46346	3	4	15.012	1	0.2	24.61667	4.31148	4.0303	10	0.5
SBDS	6	0.30898	3	5	17.239	1	0.25	27.5	4.62295	32.16725	132	0.2
TDRKH	5	0.30898	3	4	15.913	2	0.25	26.08333	4.5082	14.6171	68	0.33333
MEX3D	4	0.30779	2	4	15.241	3	0.25	26.58333	4.57377	188.86849	586	0.16667
ENOX1	1	0	1	1	5.764	1	0.2	20.95	4	0	0	0
RPS27	22	0.29157	8	8	21.191	4	0.2	30.28333	4.78689	57.34252	228	0.35714
IPO13	5	0.30898	3	4	14.444	3	0.2	27.2	4.62295	34.32918	166	0.33333
PRDX1	4	0	1	4	11.589	4	0.2	26.95	4.59016	147.75238	372	0
PEG10	3	0.30779	2	3	8.566	1	0.2	22.11667	4.01639	45.47489	140	0.33333
DNMT3B	7	0.2842	4	5	16.202	2	0.25	27.83333	4.67213	57.32172	260	0.3
PELO	5	0.30779	2	5	18.033	1	0.2	27.7	4.62295	18.74944	84	0.2
FAM98A	12	0.47366	4	4	17.613	1	0.2	27.45	4.63934	0.7	4	0.83333
TDRD5	28	0.45378	5	7	21.877	3	0.2	29.45	4.7377	51.78556	218	0.38095
PATL1	52	0.40246	7	7	24.155	1	0.25	30.75	4.90164	20.30137	132	0.52381
PARS2	26	0.56839	4	6	19.344	1	0.2	26.86667	4.4918	14.7224	54	0.4
CTU1	32	0.42794	6	6	21.549	1	0.2	28.2	4.63934	7.53108	52	0.6
RNMT	12	0.2842	4	10	25.452	4	0.25	32.25	4.93443	165.03031	636	0.11111
NXT1	12	0.32413	5	7	20.419	1	0.2	29.78333	4.77049	80.40835	338	0.2381
NXT2	4	0.30898	3	3	10.828	1	0.16667	22.15	4	1.65476	10	0.66667
CTIF	9	0.2842	4	7	19.976	2	0.2	29.53333	4.7541	80.61883	390	0.14286
RPP40	11	0.32413	5	6	19.265	2	0.2	27.45	4.55738	33.51416	124	0.33333
CPSF4L	19	0.45378	5	6	23.226	1	0.25	30.58333	4.91803	9.97532	78	0.46667
EZH2	10	0.2842	4	8	21.669	2	0.25	30.75	4.86885	155.14999	572	0.14286
CLP1	34	0.31026	9	11	28.098	3	0.25	33.58333	5.04918	119.79573	576	0.23636
ERN1	9	0.25931	5	6	20.189	1	0.2	27.53333	4.55738	31.67548	112	0.26667
TIPARP	7	0.2842	4	5	18.085	2	0.25	28	4.65574	27.03146	102	0.3
SPEN	6	0.30898	3	5	17.382	2	0.25	27.5	4.62295	29.50759	94	0.2
IFIT1	3	0	1	3	9.69	1	0.2	24.36667	4.32787	26.23653	62	0
YTHDC1	10	0.25931	5	7	23.735	1	0.25	31.33333	4.95082	44.47497	218	0.2381
MYEF2	42	0.27149	11	11	27.945	1	0.25	32.91667	4.96721	76.29101	368	0.29091
ERN2	4	0.30898	3	3	14.205	1	0.2	24.95	4.39344	0.58333	4	0.66667
ZFP36	23	0.23866	9	12	27.219	9	0.25	33.66667	5.01639	240.56739	718	0.15152

NOVA1	95	0.3733	11	12	28.321	2	0.33333	34.16667	5.08197	166.62022	676	0.33333
CLK3	31	0.25938	10	13	28.181	2	0.25	34.58333	5.08197	236.89318	858	0.17949
RBM20	50	0.47549	6	6	22.614	1	0.25	29.91667	4.85246	11.90253	72	0.66667
KHDRBS3	149	0.38288	14	15	30.382	2	0.33333	36.33333	5.19672	255.50036	1018	0.32381
MEX3B	13	0.28529	6	7	22.503	1	0.25	29.5	4.77049	39.68178	198	0.28571
DDX25	51	0.47549	6	7	23.388	1	0.2	30.36667	4.83607	18.70413	110	0.47619
EIF4E2	88	0.29378	13	13	30.084	1	0.25	34.91667	5.11475	116.73785	632	0.29487
PAIP2B	4	0	1	4	15.943	1	0.2	26.61667	4.54098	23.74821	88	0
CELF2	54	0.33919	10	12	27.883	4	0.25	33.41667	5	133.31217	520	0.25758
CD3EAP	1	0	1	1	5.89	1	0.2	21.28333	4.04918	0	0	0
ZCCHC5	3	0	1	3	5.698	2	0.2	20.56667	3.77049	122.75362	364	0
APOBEC3H	1	0	1	1	2.304	1	0.16667	15.35	2.78689	0	0	0
ZC3H12A	8	0.30898	3	7	18.743	3	0.25	29.5	4.7541	164.1295	490	0.14286
ERI2	6	0.30779	2	6	21.132	1	0.2	29.86667	4.81967	38.27614	164	0.13333
DDX28	168	0.27914	18	18	31.26	7	0.25	38.08333	5.2623	260.52881	1046	0.24837
EXO1	15	0.28529	6	9	24.219	4	0.25	32.25	4.98361	202.10351	666	0.19444
OAS1	3	0.30779	2	3	11.445	2	0.2	24.95	4.40984	16.05	48	0.33333
PABPC4L	128	0.45891	10	12	28.914	1	0.25	34	5.04918	108.82034	496	0.34848
RBM24	90	0.30655	13	13	29.633	2	0.25	34.75	5.09836	136.56635	586	0.30769
PUS1	59	0.38186	9	10	25.75	3	0.2	32.03333	4.90164	79.96712	330	0.35556
XAB2	28	0.32073	8	10	26.788	4	0.25	33.08333	5.03279	192.58545	738	0.24444
ZMAT3	6	0.30779	2	6	21.248	1	0.25	29.41667	4.78689	48.69737	194	0.13333
MARS2	44	0.43905	7	7	21.926	1	0.2	29.53333	4.7541	26.50462	116	0.57143
EIF4A3	231	0.26149	23	24	32.539	8	0.25	41.75	5.42623	753.02289	2586	0.19565
FDXACB1	49	0.4082	8	9	24.676	1	0.2	31.03333	4.83607	62.82318	266	0.38889
NIP7	68	0.30655	13	13	28.032	2	0.2	34.61667	5.06557	168.88007	710	0.30769
AEN	68	0.32197	12	14	29.865	2	0.2	35.61667	5.13115	276.14754	1026	0.25275

Supplementary Table 3. The univariate logistic regression of the 62 differentially-expressed RBPs.

Gene	OR (95% CI)	*P* value
NXT1	0.002 (0–0.052)	0.006
PELO	0 (0–0.028)	0.019
FAM98A	134.891 (12.223–7503.703)	0.002
YTHDC1	0.009 (0–0.084)	0.001
SPEN	0.005 (0–0.066)	0.002
MEX3D	15.761 (3.807–267.657)	0.006
CTIF	0.001 (0–0.041)	0.012
PAIP2B	0.036 (0.002–0.209)	0.005
PARS2	13.52 (3.664–114.511)	0.002
TIPARP	0.121 (0.025–0.338)	0.001
RNMT	0.035 (0.001–0.21)	0.006
ZMAT3	27.999 (4.929–515.098)	0.003

NOVA1	24.972 (4.49–481.767)	0.005
RIOK3	0.057 (0.005–0.249)	0.003
EZH2	15.864 (3.791–141.354)	0.002
PLD6	0.076 (0.011–0.279)	0.001
ZFP36	0.238 (0.067–0.519)	0.005
PRDX1	292.35 (12.659–47754.741)	0.006
AEN	12.52 (3.305–88.927)	0.002
CTU1	17.521 (3.864–151.39)	0.002
RBM24	0.296 (0.108–0.571)	0.003
PATL1	0.057 (0.006–0.258)	0.002
RPS27	0.138 (0.029–0.397)	0.002
EIF4A3	0.036 (0.002–0.224)	0.004
CLK3	0.096 (0.015–0.343)	0.003
NIP7	12.205 (3.244–83.971)	0.002
IPO13	18.338 (3.702–210.704)	0.004
APOBEC3H	0.466 (0.268–0.702)	0.001
ZCCHC5	4.805 (2.009–17.163)	0.003
EIF4E2	7.687 (2.502–39.611)	0.003
ZC3H12A	0.16 (0.037–0.439)	0.003
NXT2	16.012 (3.44–161.178)	0.004
ERI2	11.737 (3.025–76.228)	0.002
RPP40	5.531 (2.115–21.847)	0.003
CELF2	2.978 (1.585–7.319)	0.004
FDXACB1	5.437 (2.051–22.03)	0.005
CLP1	8.75 (2.711–47.714)	0.002
ENOX1	3.793 (1.712–11.693)	0.005
DNMT3B	3.709 (1.684–10.989)	0.005
DDX25	2.652 (1.454–6.348)	0.009
PABPC4L	3.705 (1.654–11.053)	0.006
MEX3B	1.771 (1.235–2.781)	0.005
TDRKH	3.744 (1.628–11.302)	0.006
CD3EAP	3.398 (1.557–9.86)	0.008
XAB2	0.167 (0.028–0.53)	0.016
PEG10	0.403 (0.184–0.732)	0.009
ERN1	0.212 (0.051–0.576)	0.011
RNASE1	1.36 (1.105–1.754)	0.008
PUS1	3.407 (1.496–10.302)	0.011
MARS2	2.67 (1.377–6.502)	0.012
KHDRBS3	2.311 (1.303–4.897)	0.012
CPSF4L	0.562 (0.354–0.838)	0.008
OAS1	1.567 (1.147–2.292)	0.009
RBM20	0.632 (0.426–0.869)	0.01
DDX28	4.343 (1.635–19.791)	0.018

IFIT1	17.749 (3.721–199.691)	0.003
SBDS	0.453 (0.225–0.789)	0.012
MYEF2	1.534 (1.131–2.201)	0.01
ERN2	1.643 (1.148–2.514)	0.012
PPARGC1B	1.601 (1.129–2.542)	0.019
EXO1	1.997 (1.195–3.84)	0.018
TDRD5	1.81 (1.157–3.322)	0.023

Supplementary Table 4. The CT values of the genes in the clinical samples.

Subject ID	Group	CT value		
		GAPDH	RBM24	RNMT
2022003	Control	14.74525716	18.07840008	20.03488368
2022012	Control	15.84250541	19.18721133	20.05887191
2022008	Control	16.64254397	19.61031053	21.15128554
2022023	Control	15.23590747	20.84740272	19.82626215
2022009	Control	15.01622463	18.46975354	20.61515179
2022002	Control	17.17888197	19.33891721	19.07583418
2022014	Control	14.80743867	17.74364263	21.10228253
2022016	Control	15.99036566	19.33979752	18.0255821
2022018	Control	16.13291468	17.63390788	17.37191675
2022028	Control	15.83299278	19.12467119	19.24682853
OA2022001	OA	17.26160472	21.12238607	21.00300764
OA2022002	OA	16.02803631	20.22529893	22.95781598
OA2022005	OA	15.80973734	20.8112529	21.41373541
OA2022006	OA	15.76713681	18.36677247	19.90915417
OA2022007	OA	15.84871631	21.0072574	22.34982956
OA2022008	OA	14.30135821	20.64377909	22.01430636
OA2022013	OA	15.28747625	21.83092286	20.52466488
OA2022014	OA	16.54314623	21.64015343	21.78957776
OA2022015	OA	15.3648146	23.39388081	21.22732165
OA2022017	OA	16.39616359	20.8044486	21.23231182
OA2022018	OA	17.36021479	20.80605607	21.2718392