



ORIGINAL ARTICLE

Top-down stepwise refinement identifies coding and noncoding RNA-associated epigenetic regulatory maps in malignant glioma

Yutao Huang¹  | Xiangyu Gao^{1,2} | Erwan Yang¹ | Kangyi Yue^{1,2} | Yuan Cao¹ | Boyan Zhao¹ | Haofuzi Zhang¹ | Shuhui Dai¹ | Lei Zhang¹ | Peng Luo¹  | Xiaofan Jiang¹

¹Department of Neurosurgery, Xijing Hospital, Fourth Military Medical University, Xi'an, China

²State Key Laboratory of Cancer Biology, Fourth Military Medical University, Xi'an, China

Correspondence

Peng Luo and Xiaofan Jiang, Department of Neurosurgery and Institute of Neurosurgery of PLA, PLA's Key Laboratory of Critical Care Medicine, Xijing Hospital, Fourth Military Medical University, 169 Changle Xi Road, Xi'an 710032, China.
Email: pengluo@fmmu.edu.cn (P. L.); jiangxf@fmmu.edu.cn (X. J.)

Funding information

This research was funded by the National Natural Science Foundation of China (Grant Nos. 81771322, 81671303 and 81871023) and the Youth Talent Lifting Project (Grant No. 17-JCJQ-QT-037).

Abstract

With the emergence of the molecular era and retreat of the histology epoch in malignant glioma, it is becoming increasingly necessary to research diagnostic/prognostic/therapeutic biomarkers and their related regulatory mechanisms. While accumulating studies have investigated coding gene-associated biomarkers in malignant glioma, research on comprehensive coding and noncoding RNA-associated biomarkers is lacking. Furthermore, few studies have illustrated the cross-talk signalling pathways among these biomarkers and mechanisms in detail. Here, we identified DEGs and ceRNA networks in malignant glioma and then constructed Cox/Lasso regression models to further identify the most valuable genes through stepwise refinement. Top-down comprehensive integrated analysis, including functional enrichment, SNV, immune infiltration, transcription factor binding site, and molecular docking analyses, further revealed the regulatory maps among these genes. The results revealed a novel and accurate model (AUC of 0.91 and C-index of 0.84 in the whole malignant gliomas, AUC of 0.90 and C-index of 0.86 in LGG, and AUC of 0.75 and C-index of 0.69 in GBM) that includes twelve ncRNAs, 1 miRNA and 6 coding genes. Stepwise logical reasoning based on top-down comprehensive integrated analysis and references revealed cross-talk signalling pathways among these genes that were correlated with the circadian rhythm, tumour immune microenvironment and cellular senescence pathways. In conclusion, our work reveals a novel model where the newly identified biomarkers may contribute to a precise diagnosis/prognosis and subclassification of malignant glioma, and the identified cross-talk signalling pathways would help to illustrate the noncoding RNA-associated epigenetic regulatory mechanisms of glioma tumorigenesis and aid in targeted therapy.

Yutao Huang, Xiangyu Gao, Erwan Yang and Kangyi Yue contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Journal of Cellular and Molecular Medicine* published by Foundation for Cellular and Molecular Medicine and John Wiley & Sons Ltd.

KEYWORDS

biomarker, Cox-Lasso regression, cross-talk signalling, epigenetic regulation, integrative analysis, malignant glioma, noncoding RNA

1 | INTRODUCTION

Malignant glioma, the most common primary malignant tumour of the central nervous system, usually occurs in the sixth through eighth decades of life,¹ has an annual incidence of 5.26 cases per 100,000 people and usually results in quick fatality.²⁻⁴ Malignant glioma basically consists of low-grade glioma (LGG, WHO grades II-III) and glioblastoma multiform (GBM, WHO grade IV).^{3,5,6} Nearly, all LGGs will ultimately progress to GBM.⁷ The current therapy for newly diagnosed malignant glioma is surgical removal of the maximum safe amount of tumour followed by adjuvant radiation therapy and temozolomide chemotherapy.⁸ This approach has a 2-year survival rate of 27% for newly diagnosed GBM, but the overall prognosis remains poor.^{1,2} Therefore, it is urgent to elucidate the mechanisms of malignant glioma and find new treatments. Once upon a time, malignant gliomas have ever been classified, diagnosed and therapized based on histological characteristics and resemblance with a supposed cell type of origin.⁹ In the past decade, however, the classification, diagnosis and therapy of glioma have dramatically changed. Deletions involving chromosomes 1p/19q, IDH mutations, and so on are closely related to prognosis.¹⁰⁻¹³ The latest guide incorporates molecular parameters, in addition to histology, into the classification, diagnosis and therapy of gliomas, contributing to a profound and in-depth molecular era.³

While accumulating studies have investigated coding gene-associated biomarkers in malignant glioma, research on comprehensive coding and noncoding RNA (ncRNA)-associated biomarkers is lacking. Noncoding RNAs, which are an emerging class of transcripts that are encoded by the genome but are mostly not translated into proteins, consist of housekeeping RNAs, small ncRNAs (sncRNAs, including microRNAs, miscRNAs, circular RNAs and piRNAs, etc.), and long ncRNAs (lncRNAs).^{14,15} sncRNAs and lncRNAs were once indicated as byproducts of the splicing procedure, until some researchers found that they may function in post-transcriptional modification, the organization of nuclear domains, and the regulation of proteins or RNA molecules, indicating that the biological functions of ncRNAs are more pervasive than previously suspected.¹⁶⁻²¹ Therefore, more research on these ncRNAs will lead to a greater understanding of cancer cell functions and may lead to novel clinical applications in oncology.¹⁵ The competing endogenous RNA (ceRNA) mechanism is one of the most important mechanisms of ncRNAs and indicates that endogenous coding and noncoding RNAs may share common microRNA-binding sites and thus indirectly regulate the expression of each other by competing for microRNA binding.²² Researchers have focused on lncRNA/miRNA/mRNA-associated ceRNA regulatory networks. However, ceRNA networks are not

limited to lncRNA-associated networks; they apply to all ncRNA-associated networks.²² Therefore, it is necessary and valuable to mine the comprehensive ncRNA-associated regulatory networks in glioma.

The emerging and accumulating applications of high-throughput sequencing make it easier to research and identify differentially expressed genes (DEGs) or biomarkers of various tumours. Based on bioinformatics methods, DEGs can be used to further mine critical signalling pathways or molecular mechanisms to provide guidance or illuminate tumour research questions. However, previous bioinformatics studies on glioma usually identified biomarkers or performed functional enrichment analyses. Few studies have illustrated the cross-talk signalling pathways among these biomarkers and mechanisms in detail. Illustration of the cross-talk signalling pathways among biomarkers and mechanisms is beneficial for understanding glioma tumorigenesis and malignant progression, as well as promoting glioma therapy.

'Stepwise refinement' was proposed by Swedish computer scientist Wirth in the 1970s and refers to compiling executable programs not in one step but in several steps and gradual refinement, where the program compiled in the first step is the most abstract, the program compiled in the second step is less abstract, and the program compiled in the last step is executable.²³ Here, we used two databases, The Cancer Genome Atlas (TCGA) database, whose samples were derived mainly from America (698 malignant glioma samples and 5 normal brain samples), and the Chinese Glioma Genome Atlas (CGGA) database, whose samples were derived mainly from Asia (1211 malignant glioma samples and 5 normal brain samples), to produce, validate and mine the ncRNA-associated networks. After the DEGs were identified, ncRNA-associated ceRNA networks were constructed. Subsequently, Cox and Lasso regression models were used to further identify the most valuable genes in glioma tumorigenesis and then validate them in LGG and GBM. 'Top-down' refers to dividing complex and large problems into small problems, determining the key points, and then qualitatively and quantitatively describing the problems with accurate thinking. Here, logical deduction based on references and top-down comprehensive integrative analysis, including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional enrichment, single nucleotide variation (SNV), immune infiltration, transcription factor binding site and molecular docking analyses, revealed that these genes were correlated with the circadian rhythm, tumour immune microenvironment, and cellular senescence pathways (Figure 1A,B). The epigenetic regulatory maps and signalling pathways among these genes were deduced and are discussed in the discussion section. This research aimed to determine the potential genes involved

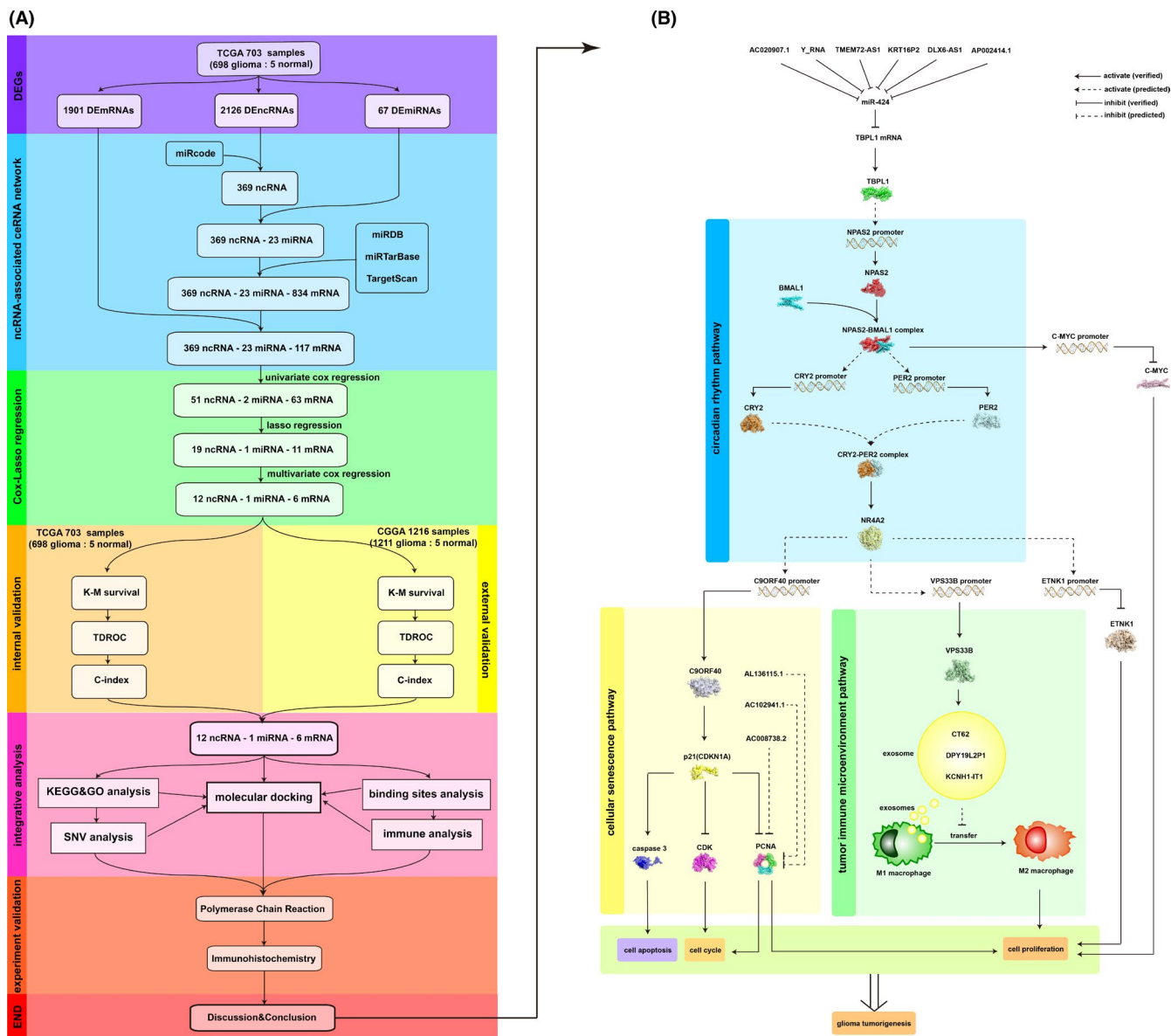


FIGURE 1 Algorithm tree and Cross-talk signalling pathways. (A) Algorithm tree of the top-down stepwise refinement method. TCGA, The Cancer Genome Atlas; CGGA, Chinese Glioma Genome Atlas; DEGs, differentially expressed genes; DE mRNAs, differentially expressed mRNAs; ncRNA, noncoding RNA; DE ncRNAs, differentially expressed ncRNAs; DE miRNAs, differentially expressed miRNAs; ceRNA, competing endogenous RNA; miRcode, miRcode database; miRDB, miRDB database; miRTarBase, miRTarBase database; TargetScan, TargetScan database; K-M survival, Kaplan–Meier survival; TROC, time-dependent receiver operating characteristic; C-index, concordance index; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology; SNV, single nucleotide variation. (B) Cross-talk signalling pathways and molecular mechanisms that involve the identified noncoding RNA-associated biomarkers

in glioma tumorigenesis and malignant progression and to illustrate the cross-talk signalling pathways among these genes and molecular mechanisms based on references and comprehensive integrative analysis. As a result, our work reveals a novel model in which the newly identified biomarkers would contribute to a precise diagnosis/prognosis and subclassification of malignant glioma, and the identified cross-talk signalling pathways would help to illustrate the noncoding RNA-associated epigenetic regulatory mechanisms of glioma tumorigenesis and malignant progression and aid in targeted therapy.

2 | MATERIALS AND METHODS

2.1 | Data acquisition

The mRNA sequencing data, miRNA sequencing data, SNV data and corresponding clinical data used to construct the critical gene network and Cox/Lasso model in malignant glioma were obtained from the TCGA in the National Institutes of Health - National Cancer Institute GDC data portal (<https://portal.gdc.cancer.gov/>, data release 26.0–8 September 2020). Data included

698 malignant glioma samples (529 GBM samples and 169 LGG samples) and 5 normal brain samples. The sequencing data were derived from the Illumina HiSeq platform (Illumina, Inc.) The research was conducted by the guidelines provided by the TCGA (<http://cancergenome.nih.gov/publications/publicationguidelines>).

The mRNA sequencing data, miRNA sequencing data and corresponding clinical data for validating the Cox and Lasso models were obtained from the CGGA (<http://www.cgga.org.cn/>, data release—14 June 2020). Data included 1211 malignant glioma samples (733 LGG samples and 478 GBM samples) and 5 normal brain samples. The mRNA sequencing data were derived from the Illumina HiSeq 2,000/2,500/4,000 Sequencing System, while the miRNA sequencing data were derived from the Agilent G2565BA Microarray Scanner System.

2.2 | Extraction of DEGs and construction of the ncRNA-miRNA-mRNA network

The sequenced genes were mapped to the human genome reference (Homo_sapiens.GRCh38.101.chr) from the Ensembl database (<https://asia.ensembl.org/index.html>). The mRNAs were recognized and extracted using the keyword 'protein_coding', while the ncRNAs (non-miRNA) were recognized and extracted using the following keywords (see supplementary material section I).

The differentially expressed mRNAs (DEmRNAs) and differentially expressed ncRNAs (DEncRNAs) (non-miRNA) were identified and depicted using the Linear Models for Microarray and RNA-Seq Data (LIMMA) package in R software (version 3.5.1).²⁴⁻²⁶ The differentially expressed miRNAs (DEmiRNAs) were identified and depicted using the Empirical Analysis of Digital Gene Expression Data (edgeR) package in R software.²⁷ The DEGs were identified using $|\log_2(\text{fold change})|$ ($|\log_2\text{FC}|$), which was set to 2, and the false discovery rate (FDR), which was set to 0.05.

The DEncRNAs (non-miRNA) and DEmiRNAs were mapped to the ncRNA-miRNA network in the miRcode database (<http://www.mircode.org/>) to identify the DEncRNAs (non-miRNA) and DEmiRNAs in the ceRNA network. The target mRNAs were identified through the miRDB database,²⁸ miRTarBase database²⁹ and TargetScan database.^{30,31} The ncRNA-miRNA-mRNA network was depicted using Cytoscape software³² (version 3.6.1).

2.3 | Cox and Lasso regression models

First, the DEGs in the ncRNA-miRNA-mRNA network were added to the univariate Cox proportional hazards model (Cox) using the 'survival' package in R software (version 3.5.1). The Cox model was constructed as follows:

$$h(t, X_i) = h_0(t) \times \exp(X_i \cdot \beta),$$

where $h_0(t)$ is the benchmark risk equation, which can be any non-negative equation for time; X_i is the eigenvector of instance i ; and β is the parameter vector, which is obtained by maximizing the Cox partial approximation.

Subsequently, the results were incorporated into the least absolute shrinkage and selection operator (Lasso) regression model using the 'glmnet' and 'survival' packages in R software (version 3.5.1). Given the objective function

$$\frac{1}{N} \sum_{i=1}^N f(x_i, y_i, \alpha, \beta),$$

the Lasso regression model would be constructed as follows:

$$\min_{\alpha, \beta} \frac{1}{N} \sum_{i=1}^N f(x_i, y_i, \alpha, \beta) \text{ subject to } \|\beta\|_1 \leq t,$$

where only β is penalized, while α is free to take any allowed value.

Finally, the results were incorporated into the multivariate Cox model using the 'survival' package in R software (version 3.5.1) to further filter the most valuable genes. The patients' risk scores were calculated and used to group patients into the high-risk group and low-risk group by the median using the 'predict' function of the 'survival' package in R software (version 3.5.1) based on gene expression in the model. Specifically speaking, if a gene is downregulated in malignant glioma and represents a worse prognosis, then this gene would be recognized as a useful gene in the multivariate Cox model (and obtains a regression coefficient; the same below), and the patients with this gene expression lower than the control would obtain a positive number score (where the concrete score is related to the gene expression level; the same below), while the patients with this gene expression higher than the control would obtain a negative number score. And if a gene is upregulated in malignant glioma and represents a worse prognosis, then this gene would be recognized as a useful gene in the model, and the patients with this gene expression higher than the control would obtain a positive number score, while the patients with this gene expression lower than the control would obtain a negative number score. Finally, after testing each gene expression, every patient would obtain a risk score which is calculated by the summation of the scores of each gene. Half of the patients (whose risk scores are higher than the median) would be categorized into the high-risk group, and the others would be categorized into the low-risk group.

The final model was evaluated through the internal (TCGA) and external (CGGA) datasets by Kaplan-Meier survival analysis, which was performed and depicted using the 'survival' and 'survminer' packages in R software (version 3.5.1). The TDROC curve was generated and depicted using the 'survivalROC' package in R software (version 3.5.1), and the C-index was determined using the 'survcomp' package in R software (version 3.5.1). The area under the TDROC curve (AUC) was calculated as follows:

$$\text{AUC}(f) = \frac{\sum_{t_0 \in D^0} \sum_{t_1 \in D^1} \mathbf{1}[f(t_0) < f(t_1)]}{|D^0| \cdot |D^1|},$$

where $1[f(t_0) < f(t_1)]$ denotes an indicator function that returns 1 if $f(t_0) < f(t_1)$ and 0 otherwise; D^0 is the set of negative examples; and D^1 is the set of positive examples.

The C-index was calculated as follows:

$$\frac{1}{M} \sum_{i: \delta_i = 1} \sum_{T_i < T_j} I[S(T_i, X_i) < S(T_j, X_j)],$$

where the function $I[C]$ refers to $I[C] = 1$ if C is true; otherwise, $I[C] = 0$. The first summation function $\sum_{i: \delta_i = 1}$ indicates that at least one instance of the pair has an event. The second summation function $\sum_{T_i < T_j}$ represents that the recording time T_j of another pair must be longer than the first instance event time. The two summation functions select all pair combinations that can be used for the comparison.

2.4 | SNV analysis

The median was used to divide patients into the high-risk group and low-risk group based on gene expression in the model through the 'predict' function of the 'survival' package in R software (version 3.5.1). SNV analysis was performed and depicted using the 'maftools' package (<https://bioconductor.org/packages/release/bioc/html/maftools.html>) in R software (version 3.5.1).

2.5 | Functional enrichment analysis

The functional enrichment analyses via the GO and KEGG databases were performed and depicted using the 'clusterProfiler', 'org.Hs.eg.db', 'enrichplot' and 'ggplot2' packages in R software (version 3.5.1) to identify the main biological functions of the genes in the model. The enrichment functions were identified using the adjusted p-value, which was set to 0.05.

2.6 | Analysis of infiltrating immune cells in malignant glioma

'The Estimation of Stromal and Immune Cells in Malignant Tumor Tissues Using Expression Data' (ESTIMATE, <https://r-forge.r-project.org/projects/estimate/>) package in R software (version 3.5.1) was used to predict tumour purity and the presence of infiltrating stromal/immune cells in glioma tissue. The correlation between the risk score and the immune filtration score was calculated and depicted using the 'ggplot2', 'openxlsx', 'ggpubr' and 'ggExtra' packages in R software (version 3.5.1). The proportion of each kind of immune cell in malignant glioma tissues based on various gene expression levels or risk scores was calculated and depicted using the 'ggplot2', 'ggthemes', 'ggpubr' and 'ggcorrplot' packages in R software (version 3.5.1).

2.7 | Molecular docking analysis

RNA sequences were obtained from the NIH genetic sequence database (GenBank),³³ while protein sequences were obtained from the UniProt database.³⁴ The minimum free energy (MFE) secondary structures of the RNAs were constructed using a loop-based energy model and the dynamic programming algorithm.³⁵ The three-dimensional structures of the RNAs were constructed based on the MFE secondary structure using the two-step procedure³⁶ and distance geometry (DG) method.³⁷ The three-dimensional structures of the RNAs were evaluated using all-heavy-atom knowledge-based statistical potential (3dRNAscore).³⁸ The three-dimensional structures of the proteins were obtained from the worldwide Protein Data Bank (PDB) database (<https://www.wwpdb.org/>) if they were known; otherwise, they were constructed using remote homology detection methods.³⁹ The interactions between RNAs and proteins were evaluated using the hybrid docking algorithm of template-based modelling and free docking (HDOCK).⁴⁰ The interactions between RNAs or proteins were evaluated using the docking energy score in HDOCK. The score used to evaluate the interaction between RNAs was set to -200 , which means that a docking energy less than -200 could be recognized as a stabilized interactive conformation. Similarly, the docking energy score used to evaluate the interaction between proteins or RNAs and proteins was set to -300 . The results were depicted using PyMOL software (The PyMOL Molecular Graphics System, Version 2.3.0 Open-Source, Schrodinger LLC).

2.8 | Transcription factor binding site analysis

The 2000 bp sequences upstream of the 5'-untranslated region (5'-UTR) are recognized as the potential sequences where promoters are located. The potential sequences where promoters are located were obtained from the National Center of Biotechnology Information (NCBI). Transcription factor binding sites were predicted using JASPAR software.⁴¹ The possibility of binding (relative score) was set to 0.9.

2.9 | Tissue samples collection

The glioma tissues and corresponding adjacent tissues (NC) were collected from the patients who underwent surgery at Xijing Hospital, Fourth Military Medical University. All the tissue samples that were obtained from the operation room were immediately frozen and then stored in liquid nitrogen until RNA and protein detection or staining. All the tissues were confirmed and graded by the department of pathology of Xijing Hospital, Fourth Military Medical University, according to the WHO stage guideline.

2.10 | Reverse-transcription and quantitative real-time polymerase chain reaction

The RNAiso plus (TAKARA) was used for isolating high purity of total RNA from tissues. RNA yields were determined with Nanodrop (Thermo Fisher Scientific). For microRNA reverse-transcription analysis, the PrimeScript™ RT reagent Kit (TAKARA) was used, and for other non-microRNA reverse-transcription analysis, the PrimeScript™ RT Master Mix (TAKARA, Kusatsu, Japan) was used. The stem-loop method was used to detect microRNA, and the procedure for microRNA reverse-transcription was (1) 42°C for 15 min, (2) 85°C for 5 s, (3) 4°C for +∞; and the process for non-microRNA reverse-transcription was (1) 37°C for 15 min, (2) 85°C for 5 s, (3) 4°C for +∞. For qRT-PCR, the SYBR Fast qPCR Mix (TAKARA) was used. The program for qRT-PCR was 1), Repts: 1, 95°C for 30 s; (2), Repts: 40, 95°C for 5 s, 60°C for 30 s; (3), Repts: 1, 95°C for 15 s, 60°C for 60 s. Relative expression of RNA was determined using the $2^{-\Delta\Delta CT}$ method. All Ct values were normalized to glyceraldehyde 3-phosphate dehydrogenase mRNA or vampyrus U6 spliceosomal RNA. The primer sequences were given in Supplementary Table 1.

2.11 | Immunohistochemistry (IHC)

The slides were retrieved by the method of heat-induced epitope retrieval and then blocked in 5% goat serum (LOT NO: SP-9001, ZSGB-bio) for 10 min. Then the primary antibody (anti-TBPL1 antibody, LOT NO: PA5-28088, 2 µg/ml, Invitrogen) was added and incubated with gentle agitation overnight at 4°C. After washing with PBS three times, the secondary antibody working solution (Goat Anti-Rabbit IgG H&L, HRP, LOT NO: SP-9001, ZSGB-bio) was subsequently added and incubated for 10 min at room temperature. The slides were washed with PBS three times, and then the Streptavidin/HRP working solution (LOT NO: SP-9001, ZSGB-bio) was added and incubated for 10 min at room temperature. After washing with PBS three times, the fresh DAB working solution (LOT NO: SP-9001, ZSGB-bio) was added and incubated for 5 min at room temperature. The slides were washed, haematoxylin stained, and finally observed under the inverted fluorescence microscopy (Olympus IX73).

2.12 | Statistical analysis

The interactions between RNAs or proteins were evaluated using the docking energy score in HDock. The score used to evaluate the interaction between RNAs was set to -200, which means that a docking energy less than -200 could be recognized as a stabilized interactive conformation. Similarly, the docking energy score used to evaluate the interaction between proteins or RNAs and proteins was set to -300. The DEGs were identified using $|\log_2(\text{fold change})|$ ($|\log_2FC|$), which was set to 2, and the FDR, which was set to 0.05. The enrichment functions were identified using the adjusted *p*-value, which was set to 0.05. Patients were grouped into the high-risk group and low-risk group using the median. All statistical analyses were performed using R software (version 3.5.1).

2.13 | Ethics approval

The glioma tissues and corresponding adjacent brain tissues were collected from the patients who underwent surgery at Xijing Hospital, Fourth Military Medical University. All the obtained tissues were approved by the Ethics Committee of Xijing Hospital, Fourth Military Medical University. The written informed consent was obtained from all patients or their relatives.

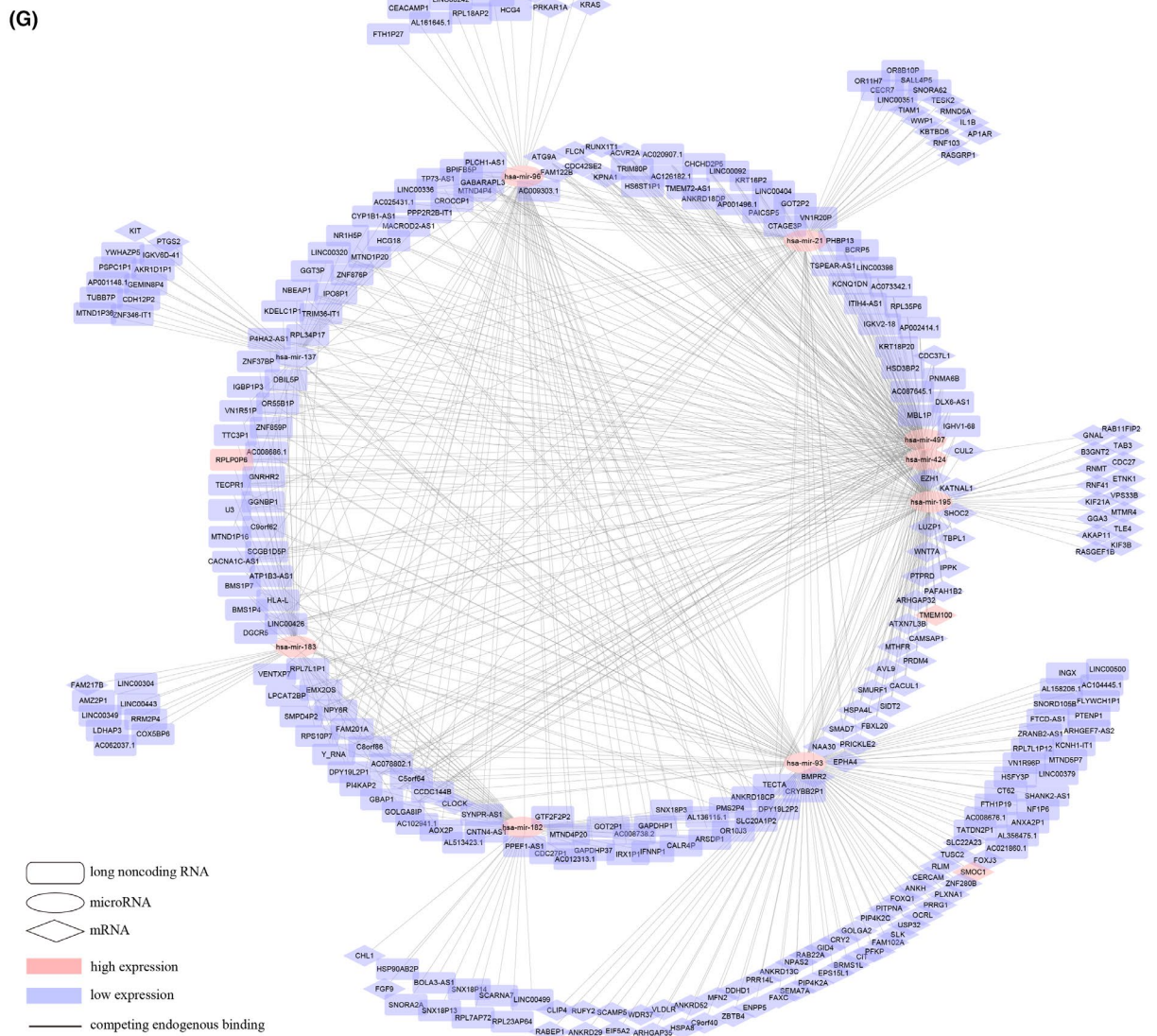
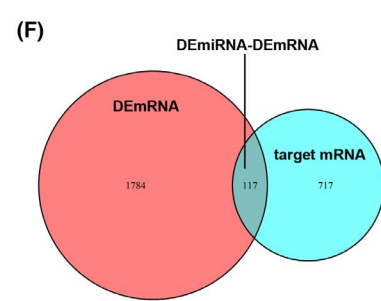
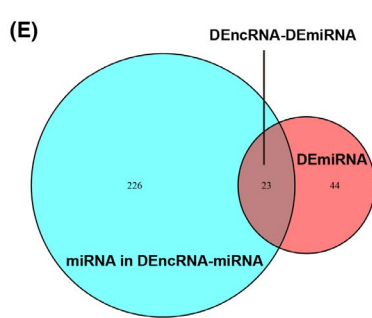
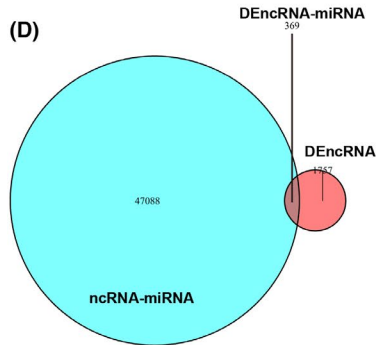
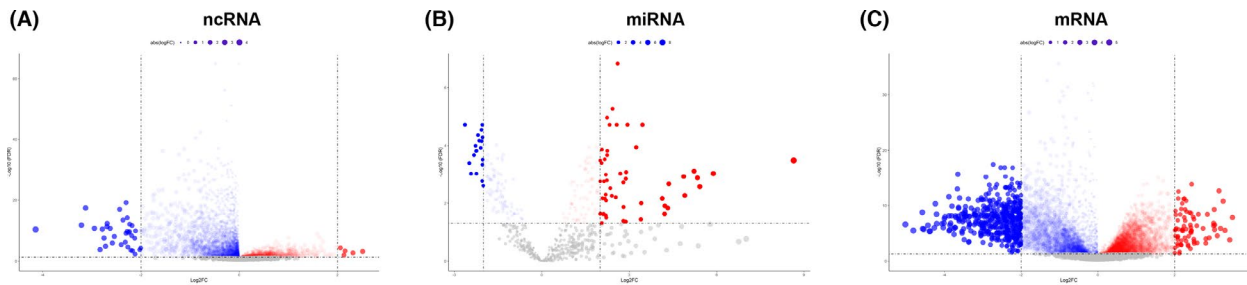
3 | RESULTS

3.1 | The DEGs and ceRNA network potentially involved in malignant glioma tumorigenesis

3.1.1 | DEGs in malignant glioma

To determine the genes related to malignant glioma tumorigenesis and progression, gene expression was analysed in 698 malignant glioma samples (169 GBM samples and 529 LGG samples) and compared to that in five normal brain samples. Ultimately, 4094 DEGs were identified: 2126 DEncRNAs (non-miRNA) (Figure 2A), 67 DEmiRNAs (Figure 2B), and 1901 DEmRNAs (Figure 2C). The DEncRNAs (non-miRNAs) accounted for almost half of the DEGs, which also highlights their importance in malignant glioma tumorigenesis and progression.

FIGURE 2 Potential DEGs and the ceRNA network in malignant glioma tumorigenesis and progression. Volcano plot of the differentially expressed (A) ncRNAs, (B) miRNAs and (C) mRNAs between malignant glioma and normal brain tissues; blue dots represent downregulated RNAs, and red dots represent upregulated RNAs; FC, fold change; FDR, false discovery rate. (D) Venn diagram of the differentially expressed ncRNAs involved in the ceRNA network; blue, ceRNA network; red, differentially expressed ncRNAs; intersection, a total of 369 differentially expressed ncRNAs involved in the ceRNA network were identified. (E) Venn diagram of the differentially expressed miRNAs involved in the ceRNA network; blue, miRNAs involved in the ceRNA network; red, differentially expressed miRNAs; intersection, a total of 23 differentially expressed miRNAs involved in the ceRNA network were identified. (F) Venn diagram of the differentially expressed mRNAs that serve as miRNA targets; blue, miRNA-targeted mRNAs; red, differentially expressed mRNAs; intersection, a total of 117 differentially expressed mRNAs that serve as miRNA targets were identified. (G) ceRNA network of the differentially expressed ncRNAs, miRNAs and mRNAs in glioma; rectangle, ncRNA; ellipse, miRNA; diamond, mRNA; blue, downregulated RNAs; red, upregulated RNAs



long noncoding RNA
 microRNA
 mRNA
 high expression
 low expression
 competing endogenous binding

3.1.2 | ncRNA-associated ceRNA networks among DEGs in malignant glioma

Considering that the ceRNA mechanism is one of the most common and critical mechanisms of ncRNAs, where an ncRNA competitively binds to a miRNA to modulate a regulatory complex and thus indirectly regulate mRNA expression,^{42,43} DEGs were screened through the ceRNA network. First, the 2126 DEncRNAs (non-miRNAs) were mapped to the ncRNA-miRNA network in the miRcode database, and 369 ncRNA-249 miRNA pairs remained (Figure 2D). Second, the 67 DEmiRNAs intersected with the 249 miRNAs that existed in the 369 ncRNA-249 miRNA pairs and 23 miRNAs remained (Figure 2E). Third, the target mRNAs were identified through the miRDB database, miRTarBase database, and TargetScan database based on the remaining 23 miRNAs. Only the mRNAs that were recognized by all three databases were considered candidate target mRNAs. Finally, the target mRNAs were intersected with the DE mRNAs, and 117 mRNAs remained (Figure 2F). The ncRNA-miRNA-mRNA network was constructed based on these 509 genes (369 ncRNAs, 23 miRNAs and 117 mRNAs) (Figure 2G).

3.2 | Potential genes for diagnosis, prognosis and therapy in malignant glioma

3.2.1 | Construction of the Cox and Lasso regression models

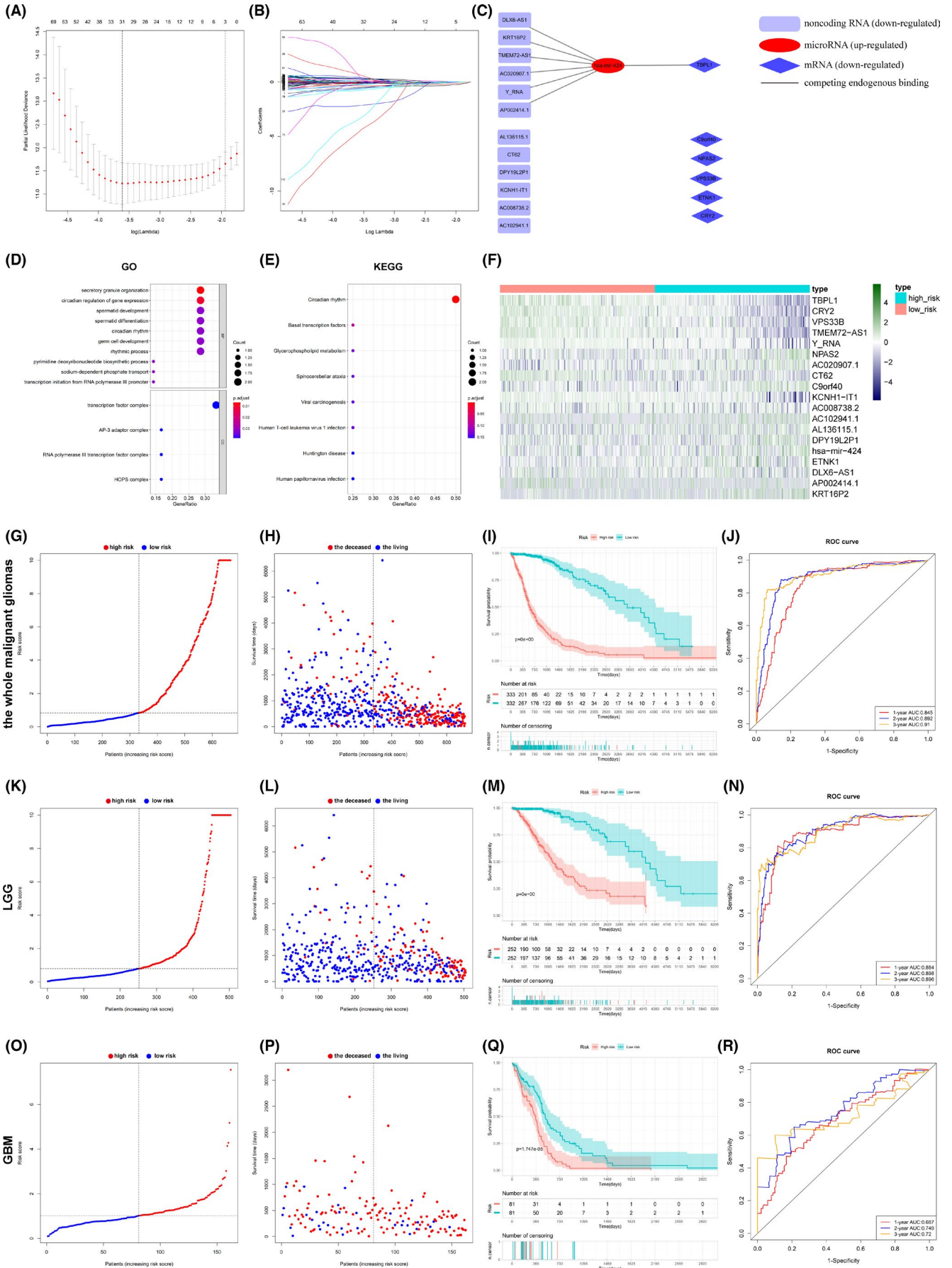
The 509 genes in the ceRNA network are obviously not suitable for evaluating malignant glioma patients and illustrate the signalling pathways in malignant glioma because of their quantity. Therefore, we needed to further screen the most valuable genes among these 509 genes. First, 509 genes (369 ncRNAs, 23 miRNAs and 117 mRNAs) in the ncRNA-miRNA-mRNA network (Figure 2G) were incorporated in the univariate Cox regression model, and ultimately, 116 genes (51 ncRNAs, 2 miRNAs and 63 mRNAs) remained. Second, 116 genes were incorporated in the Lasso regression model, and the results showed

that when $\log \lambda$ was close to -3.6 , the deviance was the lowest; thus, 31 genes remained (19 ncRNAs, 1 miRNA and 11 mRNAs) (Figure 3A,B). Subsequently, the 31 genes were incorporated in the multivariate Cox regression model, and 19 genes remained: 12 ncRNAs (AC020907.1, Y_RNA, TMEM72-AS1, KRT16P2, DLX6-AS1, AP002414.1, AC008738.2, AC102941.1, AL136115.1, CT62, DPY19L2P1 and KCN11-IT1), where Y_RNA was a miscRNA, KRT16P2 and DPY19L2P1 were transcribed unprocessed pseudogenes, AP002414.1 was a processed pseudogene, and the others were lncRNAs—1 miRNA (hsa-miR-424) and 6 coding genes (TBPL1, C9ORF40, CRY2, ETNK1, NPAS2 and VPS33B) (Figure 3C). These results demonstrate the importance of the ncRNA-associated network, not limited to the lncRNA-associated network. Among these 19 genes, hsa-miR-424 was upregulated in malignant glioma, and the others were downregulated. Moreover, DLX6-AS1, KRT16P2, TMEM72-AS1, AC020907.1, Y_RNA and AP002414.1 would competitively bind to hsa-miR-424, while hsa-miR-424 would bind to TBPL1 mRNA according to the ceRNA network (Figure 3C). Functional enrichment analyses via the GO and KEGG databases were performed to identify the biological functions of the 19 genes (Figure 3D,E). The results showed that the circadian rhythm and secretory granule organization would be the most important biological processes.

3.2.2 | Evaluation of the model in malignant glioma

The model was validated at the internal and external levels. According to the expressions of these 19 genes (Figure 3F) which were used to calculate the median of risk score, the whole malignant glioma patients (Figure 3G), LGG patients (Figure 3K), and GBM patients (Figure 3O) were grouped into two categories, a high-risk group and a low-risk group. The numbers of deceased patients among the whole malignant glioma patients (Figure 3H), LGG patients (Figure 3L), and GBM patients (Figure 3P) in the high-risk group were greater than that in the low-risk group. Kaplan-Meier survival analysis showed that the survival time of the whole malignant glioma patients (Figure 3I), LGG patients (Figure 3M), and GBM patients (Figure 3Q) in the high-risk group was significantly shorter than that in the low-risk group ($p < 10^{-20}$ in the

FIGURE 3 Critical genes for diagnosis, prognosis and therapy in glioma. (A) Cross-validation fitting into Lasso regression to select the best model. For each λ value, around the mean of the target parameter shown in the red dot, a confidence interval for the target parameter can be obtained. The two dashed lines indicate two particular λ values: λ_{\min} (dashed line on the left) refers to the minimum target parameter mean value in all λ values, while λ_{1se} (dashed line on the right) refers to the value of λ that gives the simplest model in a range of variances. (B) Tracing independent variable coefficients. Each curve represents the change in the trajectory of each independent variable coefficient. The ordinate is the coefficient value, and the lower abscissa is $\log(\lambda)$, while the upper abscissa is the number of nonzero coefficients in the model at this time. (C) Multivariate Cox regression yielded 19 genes to construct a model to predict and evaluate malignant glioma. Rectangle, ncRNA; ellipse, miRNAs; diamond, mRNAs; red, upregulated RNAs; blue, downregulated RNAs. (D) The risk group and the gene expression level of the 19 genes. (E) GO functional enrichment analysis: BP, biological process; CC, cell component. (F) KEGG functional enrichment analysis. The group of (G) the whole malignant glioma patients, (K) the LGG patients and (O) the GBM patients. Each dot represents a patient; red dot, high-risk group patients; blue, low-risk group patients. Survival state of (H) the whole glioma patients, (L) the LGG patients, and (P) the GBM patients; Each dot represents a patient; The dots on the left of the dashed line, low-risk group patients; the dots on the right of the dashed line, high-risk group patients; Red dots, deceased patients; blue dots, living patients. Kaplan-Meier survival analysis of (I) the whole glioma patients, (M) the LGG patients, and (Q) the GBM patients; red, patients in the high-risk group; blue, patients in the low-risk group; $p < 10^{-20}$ is shown as 0. TDROC curve of (J) the whole glioma patients, (N) the LGG patients, and (R) the GBM patients; red, 1-year curve; blue, 2-year curve; orange, 3-year curve



whole malignant glioma group and LGG group, $p = 1.747 \times 10^{-5}$ in GBM group). A time-dependent receiver operating characteristic (TDROC) curve was generated, and the area under the curve (AUC) of the whole malignant glioma patients was approximately 0.9 (Figure 3J), while the AUC of LGG patients was approximately 0.9 (Figure 3N), and the AUC of GBM patients was approximately 0.74 (Figure 3R). The concordance index (C-index) of the whole malignant glioma patients was 0.836, with a standard error of 0.012, a lower value of 0.814, a higher value of 0.859, and a p -value of 3.89×10^{-186} . The C-index of LGG patients was 0.859, with a standard error of 0.016, a lower value of 0.827, a higher value of 0.891, and a p -value of 2.46×10^{-106} . The C-index of GBM patients was 0.643, with a standard error of 0.026, a lower value of 0.591, a higher value of 0.694, and a p -value of 4.45×10^{-8} .

Subsequently, external patients from the CGGA database were used to validate the model, and the results were similar. The whole malignant glioma patients (Figure S1A), LGG patients (Figure S1B), and GBM patients (Figure S1C) from the CGGA database were grouped into two categories, a high-risk group and a low-risk group, according to the risk score, which is based on the expression of these 19 genes. The numbers of deceased patients among the whole malignant glioma patients (Figure S1D), LGG patients (Figure S1E), and GBM patients (Figure S1F) in the high-risk group were greater than that in the low-risk group. Kaplan-Meier survival analysis showed that the number of deceased patients among the whole malignant glioma patients (Figure S1G), LGG patients (Figure S1H), and GBM patients (Figure S1I) in the high-risk group was greater than that in the low-risk group, and the survival time of the patients in the high-risk group was significantly shorter than that in the low-risk group ($p < 10^{-20}$ in the whole malignant glioma group and LGG group, $p = 2.119 \times 10^{-2}$ in GBM group). The TDROC curve was generated, and the AUC of the whole malignant glioma patients was approximately 0.83 (Figure S1J), while the AUC of LGG patients was approximately 0.83 (Figure S1K), and the AUC of GBM patients was approximately 0.62 (Figure S1L). The C-index of the whole malignant glioma patients was 0.741, with a standard error of 0.009, a lower value of 0.722, a higher value of 0.759, and a p -value of 6.23×10^{-147} . The C-index of LGG patients was 0.759 with a standard error of 0.013, a lower value of 0.734, a higher value of 0.785, and a p -value of 3.87×10^{-86} . The C-index of GBM patients was 0.565, with a standard error of 0.018, a lower value of 0.529, a higher value of 0.601, and a p -value of 4.23×10^{-4} .

3.3 | Biological significance of the model in malignant glioma

3.3.1 | SNV analysis in malignant glioma

Before we could further determine and illustrate the molecular mechanisms among these genes, we must first examine the mutations in

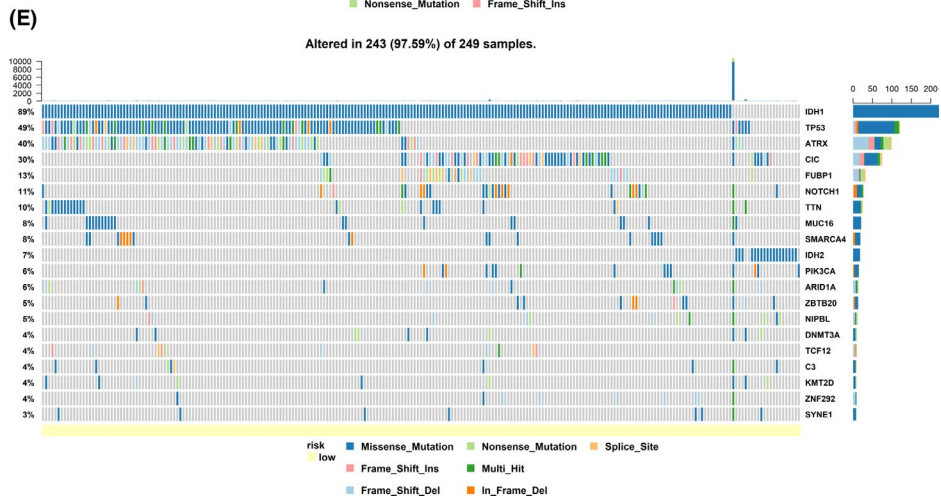
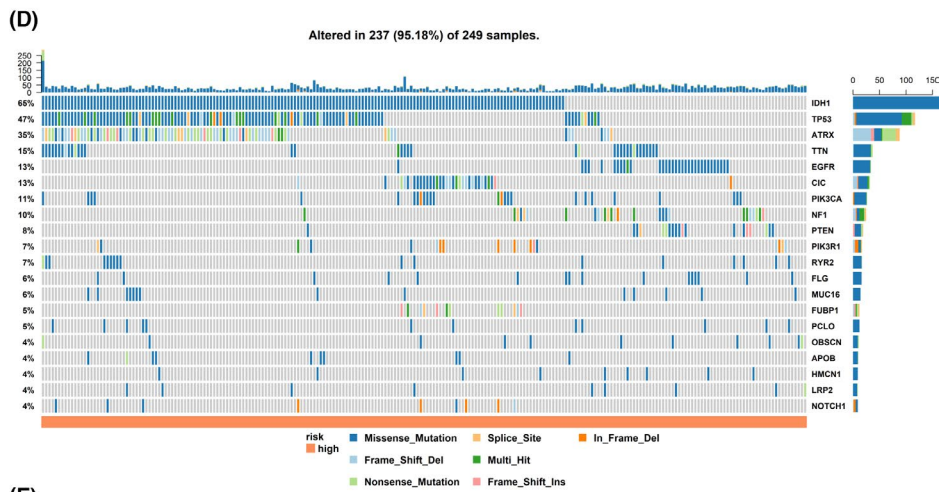
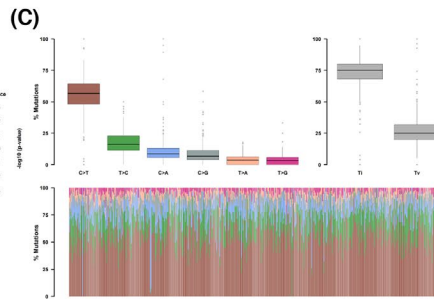
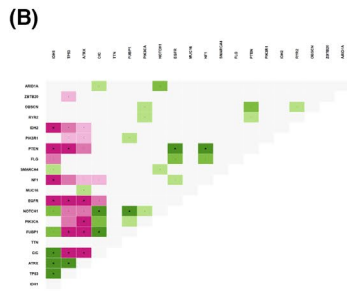
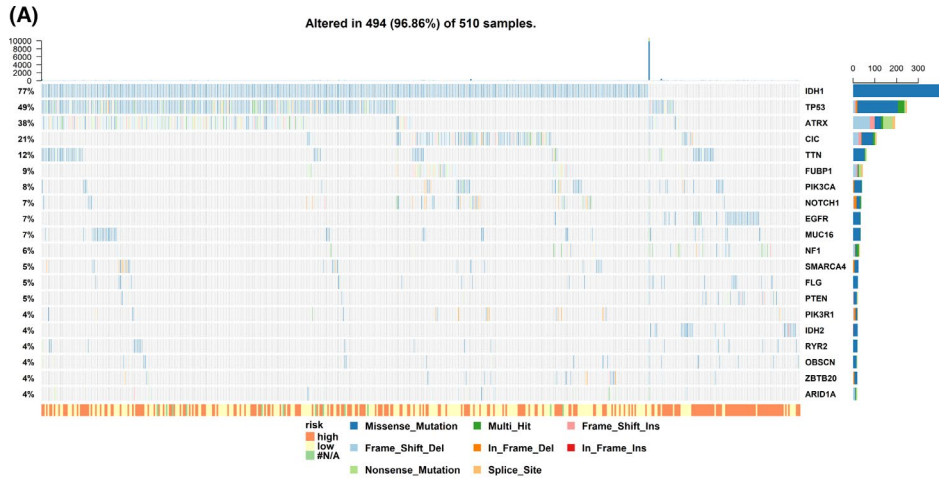
these 19 genes in glioma because if there were mutations in these 19 genes, the sequences and structures of their transcripts and proteins would change, thus hampering further analysis based on the transcripts and structures. The 510 malignant glioma samples were examined, and the top 20 mutated genes in malignant glioma were IDH1, TP53, ATRX, CIC, TTN, FUBP1, PIK3CA, NOTCH1, EGFR, MUC16, NF1, SMARCA4, FLG, PTEN, PIK3R1, IDH2, RYR2, OBSCN, ZBTB20 and ARID1A, where the most common type of mutation was missense (Figure 4A). The relations among these top 20 mutated genes were also examined (Figure 4B); for example, IDH1 mutations usually co-occurred with CIC, ARTX and TP53 mutations and exclusively with IDH2, PTEN, NF1 and EGFR mutations. Transition (Ti) was the major mutation type (mainly cytosine transitions to thymine or thymine transitions to cytosine), which accounted for approximately 75% of mutations, and transversion (Tv) accounted for approximately 25% of mutations (Figure 4C). Fortunately, mutations in the 19 genes were not found, suggesting that the effectiveness of the model would not be affected by gene mutations in malignant glioma.

Furthermore, the samples were grouped into two categories, a high-risk group and a low-risk group, based on the expression of the 19 genes to examine the incidence rate of mutations in various genes (Figure 4D,E). IDH1, TP53 and ATRX were the top 3 mutated genes in both the high-risk group and the low-risk group. The incidence rate of IDH1 mutation in the low-risk group (89%) was much higher than that in the high-risk group (65%), and simultaneously, the incidence rate of ATRX mutation in the low-risk group (40%) was much higher than that in the high-risk group (35%), while the incidence rate of TP53 mutation in the low-risk group (49%) was not significantly different from that in the high-risk group (47%). These phenomena are coincident with the consensus that mutations in IDH1 and ATRX are the initiating events for the development of many gliomas, and their presence dictates favourable clinical behaviour.⁴⁴⁻⁴⁶ This result confirms that our model based on the 19 genes would be reliable. Furthermore, as we know, IDH is the critical metabolic enzyme that catalyses the oxidative decarboxylation of isocitrate to α -ketoglutarate (α KG), NAD(P)H and CO_2 , while mutant enzymes produce (R)-2-hydroxyglutarate, which in turn inhibits α KG-dependent dioxygenase function, resulting in a global hypermethylation phenotype.⁴⁷ DNA methylation is one of the most important processes in gene expression regulation and usually suppresses gene expression. Interestingly, among these 19 genes, 18 were downregulated. Therefore, we conjecture that this hypermethylation may influence the expression of these 19 genes.

3.3.2 | Circadian rhythm in malignant glioma

(This section includes only the declarative results. Should you read this manuscript for the first time, we strongly suggest you read

FIGURE 4 SNV analysis. (A) SNV signatures in a total of 510 malignant glioma patients. (B) Correlations between the top 20 mutated genes. (C) Types and proportions of mutations. Ti, transition; Tv, transversion; A, adenine; T, thymine; C, cytosine; G, guanine. (D) SNV signatures in patients in the high-risk group based on the model. (E) SNV signatures in patients in the low-risk group based on the model



Section 4.1 at first.) Functional enrichment analyses via the GO and KEGG databases were performed to identify the biological functions of the 19 genes (Figure 3D,E). The results showed that the circadian rhythm and secretory granule organization would be the most important biological processes.

CRY2 and NPAS2 are the core components of the circadian clock.^{48,49} Moreover, NPAS2 interacts with BMAL1, and the compound then binds to the E-box of the CRY2 and PER2 promoters to facilitate expression.⁵⁰ Furthermore, CRY2 heterodimerizes with PER2 to bind to the transcription factors NR4A2, HNF4A, PPARA and NR1D1 to facilitate their activation.⁵⁰ Therefore, can TBPL1, a transcription factor,^{51,52} bind to the NPAS2 promoter to facilitate its expression, thus promoting CRY2 expression and further activating NR4A2, HNF4A, PPARA and NR1D1? To test whether TBPL1 could bind to the NPAS2 promoter to facilitate its expression, transcription factor binding site predictions were performed, and the results showed that the NPAS2 promoter has four potential binding sites, with the possibility of more than 0.9 to which TBPL1 would bind (Table S2). To test whether NPAS2 could interact with BMAL1 and the DNA promoter, molecular docking analysis was performed, and the results revealed that NPAS2 would interact with BMAL1 and bind to the E-box of the DNA promoter (Figure 5A), thus promoting PER2 and CRY2 expression. Moreover, CRY2 heterodimerizes with PER2 (Figure 5B) to bind to the transcription factors NR4A2, HNF4A, PPARA and NR1D1 to facilitate their activation.

As 6 coding genes (TBPL1, C9ORF40, CRY2, ETNK1, NPAS2 and VPS33B) were identified, we asked whether NR4A2, HNF4A, PPARA and NR1D1, as transcription factors, promote or suppress the expression of the other three genes (C9ORF40, ETNK1, or VPS33B). To examine whether NR4A2, HNF4A, PPARA and NR1D1 could promote or suppress the expression of C9ORF40, ETNK1, or VPS33B, similar transcription factor binding site predictions were performed, and the results showed that only the transcription factor NR4A2 would bind to the promoters of C9ORF40, ETNK1 and VPS33B to promote or suppress their expression (Table S2).

3.3.3 | Tumour immune microenvironment in malignant glioma

(This section includes only the declarative results. Should you read this manuscript for the first time, we strongly suggest you read Sections 4.1 and 4.2 at first.) Accumulating evidence indicates that VPS33B plays a critical role in vesicle-mediated transport and organization.⁵³⁻⁵⁶ It has been reported that the ncRNAs CT62, DPY19L2P1 and KCNH1-IT1 can be delivered by exosomes, while AC008738.2, AC102941.1 and

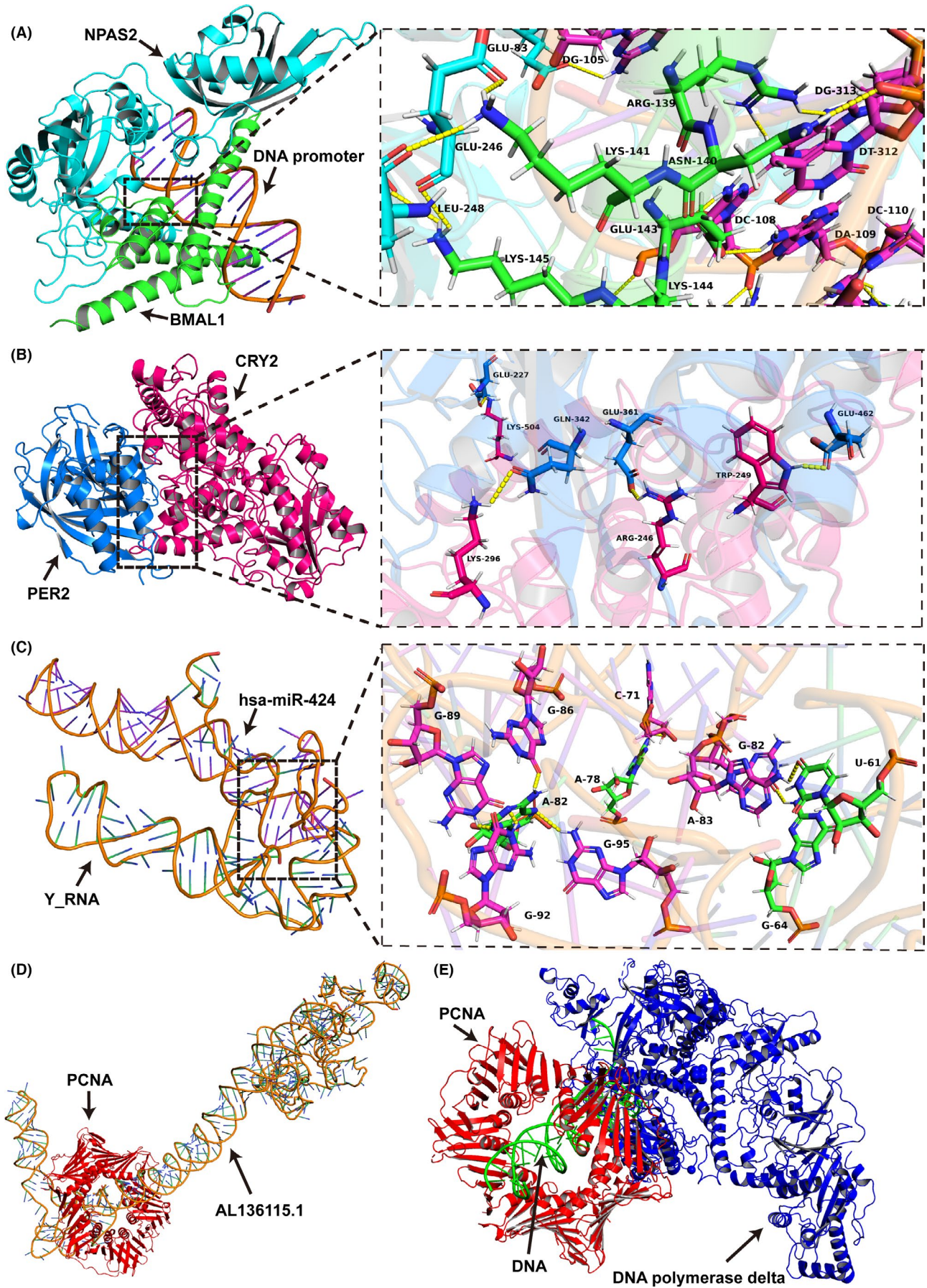
AL136115.1 have not been identified.^{57,58} Therefore, VPS33B would promote exosomes carrying CT62, DPY19L2P1 and KCNH1-IT1 to other cells, thus regulating the tumor microenvironment. Considering that immune cells are an important component of the tumor environment, glioma-derived exosomes would deliver CT62, DPY19L2P1 and KCNH1-IT1 to immune cells, regulating the tumour immune microenvironment.^{59,60} To test whether CT62, DPY19L2P1 and KCNH1-IT1 could regulate the tumour immune microenvironment, immune analyses were performed. The results revealed that as the risk score increased, the proportion of stromal cells and immune cells increased while tumour purity decreased (Figure 6A,B), indicating that immune cells play a critical role in malignant glioma tumorigenesis and progression. Moreover, M2 macrophages were the most abundant immune cells in the malignant glioma microenvironment (Figure 6C,D), suggesting that M2 macrophages would play an important role in malignant glioma tumorigenesis and progression. The proportions of M2 macrophages in the VPS33B (Figure 6E), DPY19L2P1 (Figure 6F), CT62 (Figure 6G) and KCNH1-IT1 (Figure 6H) low expression groups were higher than those in the high expression groups, while the proportions of M0 macrophages and M1 macrophages were lower than those in the high expression groups. These results suggest that VPS33B, CT62, DPY19L2P1 and KCNH1-IT1 would suppress M0 macrophage and M1 macrophage differentiation into M2 macrophages. What's more, considering TBPL1 and hsa-mir-424 that locate upstream of the pathways, the effects of TBPL1 and hsa-mir-424 have also been validated. The results show that the proportion of M2 macrophages in the TBPL1 (Figure 6I) low expression groups was higher than that in the high expression groups, while the proportion of M2 macrophages in the hsa-mir-424 (Figure 6J) low expression groups was lower than that in the high expression groups. This, again, verifies that our model is reliable.

3.3.4 | Cellular senescence in malignant glioma

(This section includes only the declarative results. Should you read this manuscript for the first time, we strongly suggest you read Sections 4.1-4.3 at first.) We have illustrated that Y_RNA, AC020907.1, TMEM72-AS1, KRT16P2, DLX6-AS1 and AP002414.1 would competitively bind to hsa-miR-424 and that hsa-miR-424 could bind to TBPL1 mRNA. To further verify these findings, molecular docking was performed. The results revealed that Y_RNA (Figure 5C), AC020907.1 (Figure S2A), TMEM72-AS1 (Figure S2B), KRT16P2 (Figure S2C), DLX6-AS1 (Figure S2D) and AP002414.1 (Figure S2E) could competitively bind to hsa-miR-424, while hsa-miR-424 could bind to TBPL1 mRNA (Figure S2F).

Although C9ORF40 has no correlation with either the circadian rhythm or secretory granule organization, it plays a critical

FIGURE 5 Molecular docking. (A) The interactions among NPAS2, BMAL1, and DNA promoters. (B) The interaction between CRY2 and PER2. (C) The interaction between hsa-miR-424 and Y_RNA. (D) The interaction between the PCNA trimer and AL136115.1. AL136115.1 competitively passes through the cavity of PCNA, which hinders the DNA from going through the same site. (E) The interactions among PCNA, DNA, and Pol δ . The PCNA trimer forms a hexagon-like structure with a cavity where DNA passes through PCNA and binds to Pol δ to replicate the lagging strandy.⁶⁶



role in cellular senescence by activating the p21 pathway.^{61,61} C9ORF40 activates p21 and thus inhibits proliferation by the CDK or proliferating cell nuclear antigen (PCNA) signalling pathway and simultaneously promotes apoptosis by the caspase-3 signalling pathway.⁶²⁻⁶⁵ To test whether the other three ncRNAs, AC008738.2, AC102941.1, and AL136115.1, could bind to p21, CDK, PCNA, or caspase-3, molecular docking analysis was performed. The results revealed that only PCNA would interact with the ncRNAs AL136115.1 (Figure 5D), AC008738.2 (Figure S2G), and AC102941.1 (Figure S2H). In eukaryotes, the PCNA trimer forms a hexagon-like structure with a cavity where DNA passes through PCNA and binds to DNA polymerase δ (Pol δ) to replicate the lagging strand (Figure 5E) and cooperates with flap endonuclease 1 (FEN1) to process the Okazaki fragments for ligation.⁶⁶ However, the DNA binding cavity for PCNA would be occupied by the ncRNAs AC008738.2, AC102941.1, and AL136115.1, hindering and impeding DNA replication and thus suppressing proliferation and other cell cycle processes in malignant glioma.

3.4 | Clinical significance of the circadian rhythm, tumour immune microenvironment, and cellular senescence pathways in malignant glioma

The eleven genes within the circadian rhythm pathway (AC020907.1, Y_RNA, TMEM72-AS1, KRT16P2, DLX6-AS1, AP002414.1, hsa-miR-424, TBPL1, NPAS2, CRY2, and ETNK1) were used to construct a model to evaluate the importance of these genes in malignant glioma. The patients were grouped into two groups based on the expression of these genes: a high-risk group and a low-risk group (Figure S3A). Kaplan–Meier survival analysis showed that the number of deceased patients in the high-risk group was greater than that in the low-risk group, and the survival time of the patients in the high-risk group was significantly shorter than that in the low-risk group ($p < 10^{-20}$) (Figure S3B,C). The TDROC curve was generated, and the AUC was approximately 0.9, with a 1-year AUC of 0.866, a 2-year AUC of 0.889, and a 3-year AUC of 0.887 (Figure S3D). The C-index was 0.841, with a standard error of 0.017, a lower value of 0.806, a higher value of 0.876, and a p -value of 2.60×10^{-82} . These results reveal that this model would also be accurate, highlighting that the circadian rhythm pathway and its related genes would be critical for malignant glioma and its evaluation.

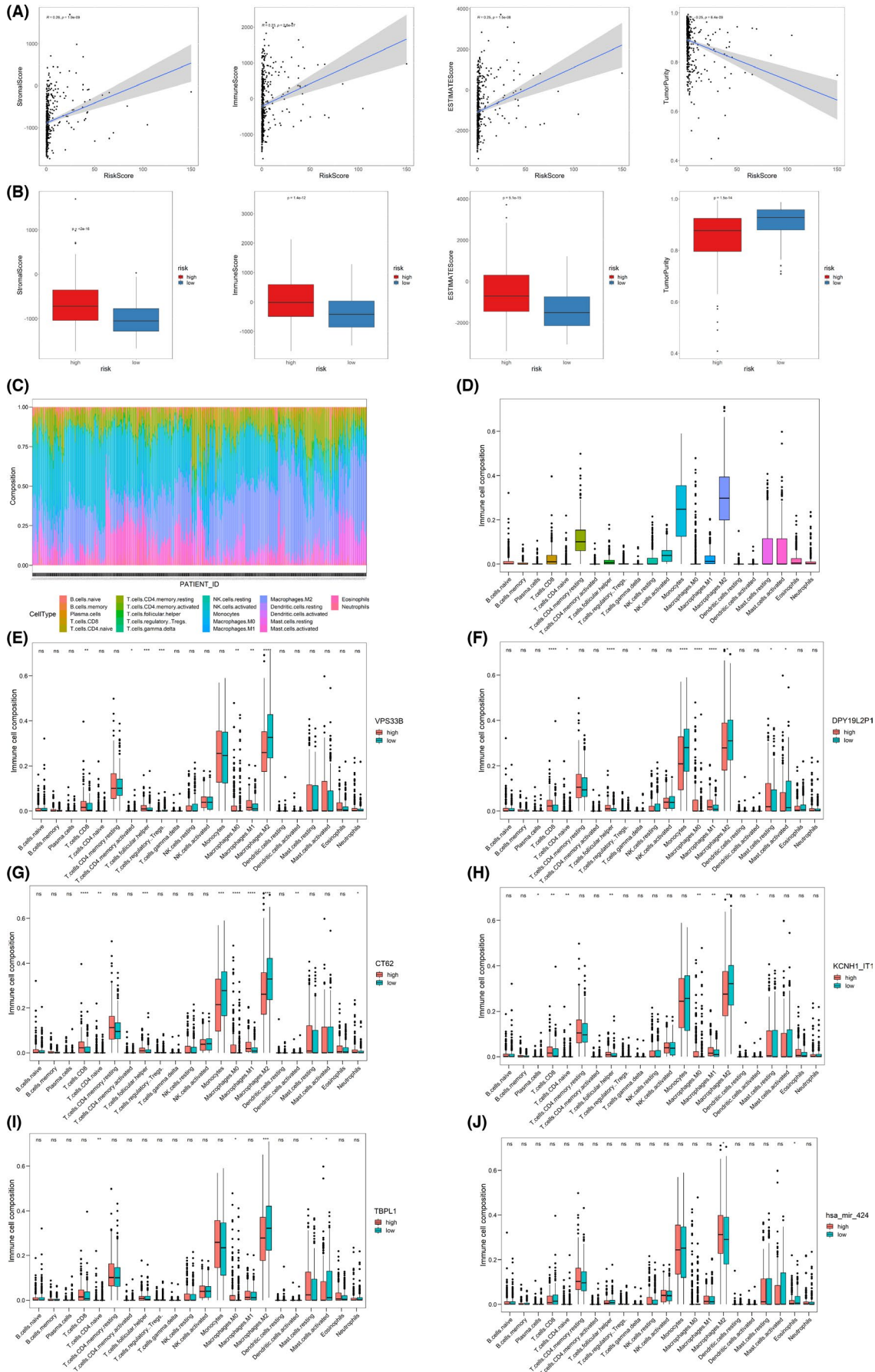
Fourteen genes within the cellular senescence pathway (AC020907.1, Y_RNA, TMEM72-AS1, KRT16P2, DLX6-AS1,

AP002414.1, hsa-miR-424, TBPL1, NPAS2, CRY2, C9ORF40, AL136115.1, AC102941.1, and AC008738.2) were used to construct a model to evaluate the importance of these genes in malignant glioma. The patients were grouped into two groups based on the expression of these genes: a high-risk group and a low-risk group (Figure S3E). Kaplan–Meier survival analysis showed that the number of deceased patients in the high-risk group was greater than that in the low-risk group, and the survival time of the patients in the high-risk group was significantly shorter than that in the low-risk group ($p < 10^{-20}$) (Figure S3F,G). The TDROC curve was generated, with a 1-year AUC of 0.866, a 2-year AUC of 0.900, and a 3-year AUC of 0.906 (Figure S3H). The C-index was 0.849, with a standard error of 0.016, a lower value of 0.817, a higher value of 0.881, and a p -value of 4.18×10^{-103} . Overall, this model would also be accurate, highlighting that the cellular senescence pathway and its related genes would be critical for malignant glioma and its evaluation.

Fourteen genes within the tumour immune microenvironment pathway (AC020907.1, Y_RNA, TMEM72-AS1, KRT16P2, DLX6-AS1, AP002414.1, hsa-miR-424, TBPL1, NPAS2, CRY2, VPS33B, CT62, DPY19L2P1, and KCNH1-IT1) were used to construct a model to evaluate the importance of these genes in malignant glioma. The patients were grouped into two groups based on the expression of these genes: a high-risk group and a low-risk group (Figure S3I). Kaplan–Meier survival analysis showed that the number of deceased patients in the high-risk group was greater than that in the low-risk group, and the survival time of the patients in the high-risk group was significantly shorter than that in the low-risk group ($p < 10^{-20}$) (Figure S3J,K). The TDROC curve was generated, with a 1-year AUC of 0.884, a 2-year AUC of 0.888, and a 3-year AUC of 0.878 (Figure S3L). The C-index was 0.845, with a standard error of 0.019, a lower value of 0.808, a higher value of 0.882, and a p -value of 1.62×10^{-73} . Overall, this model would also be accurate, highlighting that the tumour immune microenvironment pathway and its related genes would be critical for malignant glioma and its evaluation.

Since all three pathways act through AC020907.1, Y_RNA, TMEM72-AS1, KRT16P2, DLX6-AS1, AP002414.1, hsa-miR-424, TBPL1, NPAS2, and CRY2, these genes could serve as potential therapeutic targets for malignant glioma. Knockdown of hsa-miR-424 or overexpression of AC020907.1, Y_RNA, TMEM72-AS1, KRT16P2, DLX6-AS1, AP002414.1, and TBPL1 could regulate these three pathways to suppress malignant glioma tumorigenesis and progression.

FIGURE 6 Immune infiltration analysis. A) Linear correlations between the risk score and stromal score, immune score, ESTIMATE score, and tumour purity. ESTIMATE, Estimation of Stromal and Immune Cells in Malignant Tumour Tissues using Expression Data. B) Boxplot of the correlations among the risk score, stromal score, immune score, ESTIMATE score, and tumour purity. C) Landscape of infiltrating immune cells. D) Relative proportions of infiltrating immune cells. E) Relative proportions of infiltrating immune cells based on the expression of VPS33B (high or low). F) Relative proportions of infiltrating immune cells based on the expression of DPY19L2P1 (high or low). G) Relative proportions of infiltrating immune cells based on the expression of CT62 (high or low). H) Relative proportions of infiltrating immune cells based on the expression of KCNH1-IT1 (high or low). I) Relative proportions of infiltrating immune cells based on the expression of TBPL162 (high or low). J) Relative proportions of infiltrating immune cells based on the expression of hsa-mir-424 (high or low). *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.005$; ****, $p < 0.001$; ns, not significant



3.5 | Experiment validation

The PCR assays were performed to verify the 19-gene model preliminarily (Figure 7A–C). Nearly all the genes were consistent with our conclusions except CT62. The ncRNAs (DLX6-AS1, KRT16P2, TMEM72-AS1, AC020907.1, Y_RNA, AP002414.1, AL136115.1, DPY19L2P1, KCNH1-IT1, AC008738.2, and AC102941.1) and coding RNAs (TBPL1, C9ORF40, NPAS2, VPS33B, ETNK1, and CRY2) were low expressed in both LGG and GBM, while microRNA (hsa-miR-424) were high expressed in both LGG and GBM. In particular, DLX6-AS1, Y_RNA, KCNH1-IT1, hsa-miR-424, TBPL1, C9ORF40, NPAS2, and CRY2 have significance with $p < 0.001$, which indicates the priority of these genes in further experiments. Further, the IHC (Figure 7D,E) assays were performed to verify the expression of core protein TBPL1 which lies upstream of the three pathways. The results of IHC indicate that the expressions of TBPL1 in LGG and GBM were downregulated, and the expression of TBPL1 in GBM was much lower than LGG, which was consistent with our model.

4 | DISCUSSION

4.1 | Circadian rhythm pathway

After identifying DEGs, constructing ceRNA network, performing Cox/Lasso regression model, and evaluating the model, we obtained a novel 19-gene model (Figure 3C), including 12 ncRNAs (AC020907.1, Y_RNA, TMEM72-AS1, KRT16P2, DLX6-AS1, AP002414.1, AC008738.2, AC102941.1, AL136115.1, CT62, DPY19L2P1 and KCNH1-IT1), where Y_RNA is a miscRNA, KRT16P2 and DPY19L2P1 are transcribed unprocessed pseudogenes, AP002414.1 is a processed pseudogene, and the others are lncRNAs—1 miRNA (hsa-miR-424), and 6 coding genes (TBPL1, C9ORF40, CRY2, ETNK1, NPAS2, and VPS33B). But what's the biological function of these genes in malignant glioma? As mutations in the 19 genes were not found in malignant glioma (Figure 4), we could further determine the correlations between the 19 genes and malignant glioma. Functional enrichment analyses via the GO

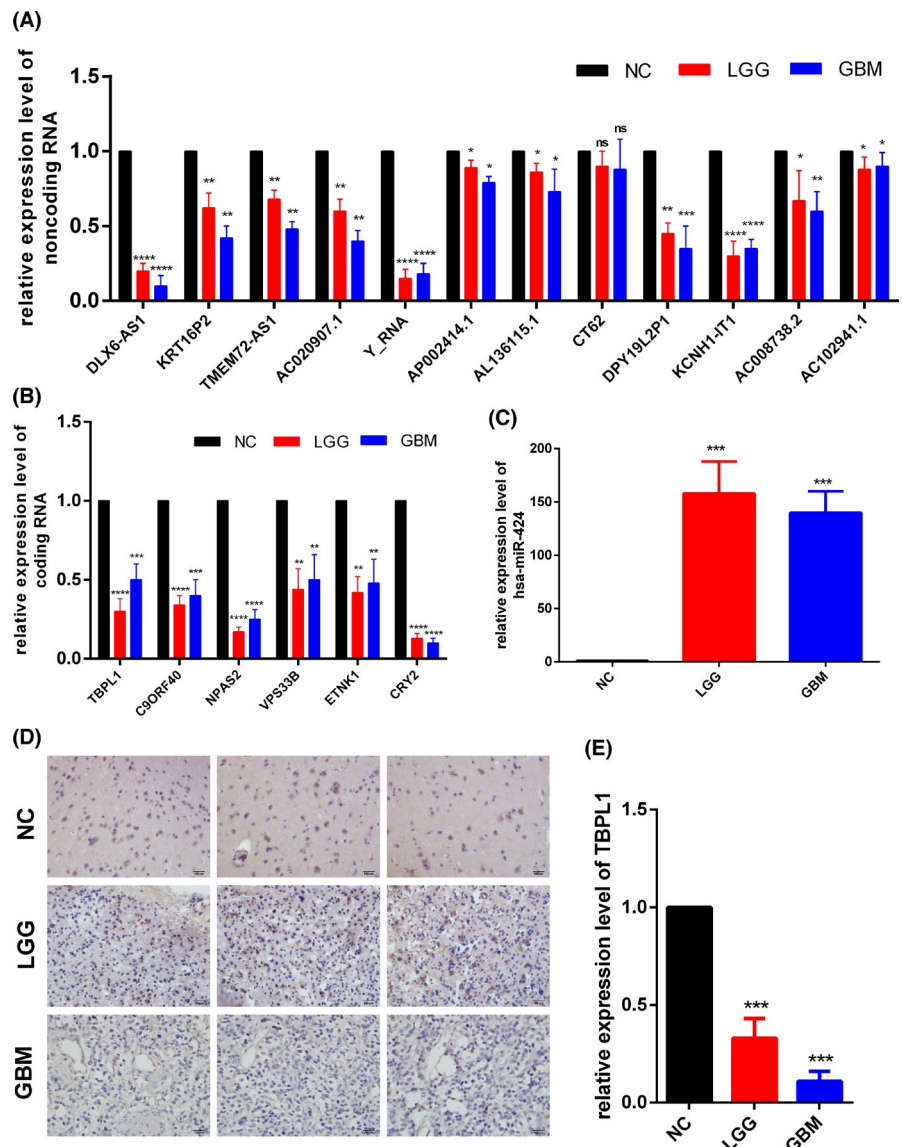


FIGURE 7 Experiment validation of the 19-gene model and core protein TBPL1. Relative expression level of (A) ncRNA, (B) coding RNA and (C) microRNA tested by PCR. (D,E) Relative expression level of TBPL1 protein detected by IHC. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.005$; ****, $p < 0.001$; ns, not significant

and KEGG databases were performed to identify the biological functions of the 19 genes (Figure 3D,E). The results showed that the circadian rhythm and secretory granule organization were the most important biological processes. First, we will discuss the circadian rhythm.

We previously found some relationships among these 19 genes—*Y_RNA* (Figure 5C), *AC020907.1* (Figure S2A), *TMEM72-AS1* (Figure S2B), *KRT16P2* (Figure S2C), *DLX6-AS1* (Figure S2D), and *AP002414.1* (Figure S2E)—which could competitively bind to *hsa-miR-424*, and *hsa-miR-424*, which could bind to *TBPL1* mRNA (Figure S2F). These correlations, however, are not enough to explain the biological functions of the 19 genes in malignant glioma tumorigenesis and progression. We observed that among the 6 coding genes (*TBPL1*, *C9ORF40*, *CRY2*, *ETNK1*, *NPAS2*, and *VPS33B*) (Figure 3C), *CRY2* and *NPAS2* were the core components of the circadian clock.^{48,49} Moreover, *NPAS2* interacts with *BMAL1*, and the compound then binds to the E-box of the *CRY2* and *PER2* promoters (Figure 5A) to facilitate their expression.⁵⁰ Furthermore, *CRY2* heterodimerizes with *PER2* (Figure 5B) to bind to the transcription factors *NR4A2*, *HNF4A*, *PPARA*, and *NR1D1* to facilitate their activation.⁵⁰ Therefore, we asked whether *TBPL1*, a transcription factor,^{51,52} can bind to the *NPAS2* promoter to facilitate its expression, thus promoting *CRY2* expression and further activating *NR4A2*, *HNF4A*, *PPARA*, and *NR1D1*. Transcription factor binding site predictions showed that *TBPL1* would bind to the *NPAS2* promoter (Table S2), indicating that *TBPL1* would promote activation of *NR4A2*, *HNF4A*, *PPARA*, and *NR1D1* indirectly through *NPAS2/CRY2*. Further, can *NR4A2*, *HNF4A*, *PPARA*, and *NR1D1*, as transcription factors, promote or suppress the expression of *C9ORF40*, *ETNK1*, or *VPS33B*? Similar transcription factor binding site predictions showed that only the transcription factor *NR4A2* would bind to the promoters of *C9ORF40*, *ETNK1*, and *VPS33B* to promote or suppress their expression (Table S2). Moreover, *VPS33B* expression is regulated by the circadian rhythm,⁶⁷ which is consistent with our conclusion. Furthermore, ethanolamine kinase 1 (*ETNK1*) is a rate-controlling enzyme in phosphatidylethanolamine biosynthesis, and its low expression would promote the proliferation and invasion of gastric cancer, larynx- or tongue-derived squamous cell carcinoma, among others,^{68–70} indicating that *TBPL1* would promote glioma tumorigenesis via the *NPAS2/CRY2/NR4A2/ETNK1* pathway. Moreover, accumulating evidence indicates that *PER2* plays important and critical roles in tumour suppression and the DNA damage response through the *NPAS2/BMAL1* complex, which regulates *TP53* and *C-MYC*.^{71,72} Overall, the noncoding RNAs *AC020907.1*, *Y_RNA*, *TMEM72-AS1*, *KRT16P2*, *DLX-AS1*, and *AP002414.1* competitively bind to *miR-424* to promote *TBPL1* expression, thus promoting *NPAS2* expression. Furthermore, *CRY2* and *PER2* would suppress glioma tumorigenesis by suppressing *ETNK1* expression through the transcription factor *NR4A2* or by suppressing *C-MYC* or *TP53* expression through the *NPAS2-BMAL1* complex (Figure 1B).

Interestingly, Ben Franklin's aphorism 'early to bed, early to rise, makes a man healthy, wealthy, and wise' also suggests the potential importance of the circadian rhythm in health. Regardless of bacteria or eukaryotes, circadian rhythms play an important role in controlling a variety of physiological processes, and disruptions to normal circadian biology can cause many diseases.⁷³ Many physiological processes, including hormone secretion, drug and xenobiotic metabolism, glucose homeostasis, cell cycle progression, and tumorigenesis, are regulated by the circadian rhythm,^{49,74,75}. Accumulating evidence indicates that circadian rhythm disorders cause many diseases.^{76–81} Therefore, understanding the correlation between the circadian rhythm and disease is critical and important for understanding the underlying mechanisms and potential therapies. Our work reveals that glioma tumorigenesis is significantly correlated with the circadian rhythm. We strongly suggest not staying up late if you can go to bed early.

4.2 | Tumour immune microenvironment

During the process of illustrating the circadian rhythm pathway, among the 6 coding genes, we exhibited the biological functions of *TBPL1*, *NPAS2*, *CRY2*, and *ETNK1*; therefore, what are the roles of *VPS33B* and *C9ORF40* in malignant glioma? As GO and KEGG functional enrichment analyses revealed that these genes are also related to secretory granule organization (Figure 3D,E), we are going to discuss secretory granule organization.

Accumulating evidence indicates that *VPS33B* plays a critical role in vesicle-mediated transport and organization.^{53–56} Therefore, given that exosomes, a kind of vesicle that mediates communication between cells by transferring ncRNAs, promote tumorigenesis,^{82–84} can *VPS33B* promote exosomes carrying ncRNAs to other cells in addition to malignant glioma and thus regulate the tumour microenvironment? First, the following 6 ncRNAs have not been discussed until now: *AC008738.2*, *AC102941.1*, *AL136115.1*, *CT62*, *DPY19L2P1*, and *KCNH1-IT1*. Therefore, do exosomes carry these 6 ncRNAs? Next, do these 6 ncRNAs regulate the tumour microenvironment? It has been reported that the ncRNAs *CT62*, *DPY19L2P1*, and *KCNH1-IT1* can be delivered by exosomes, while *AC008738.2*, *AC102941.1*, and *AL136115.1* have not been identified.^{57,58} Therefore, *VPS33B* would promote exosomes carrying *CT62*, *DPY19L2P1*, and *KCNH1-IT1* to other cells, thus regulating the tumour microenvironment. Considering that immune cells are an important component of the tumour environment, glioma-derived exosomes would deliver *CT62*, *DPY19L2P1*, and *KCNH1-IT1* to immune cells, regulating the tumour immune microenvironment.^{59,60} Immune analyses revealed that as the risk score increased, the proportions of stromal cells and immune cells increased, while tumour purity decreased (Figure 6A,B), indicating that immune cells play a critical role in malignant glioma tumorigenesis and progression. We noticed that M2 macrophages were the most abundant cell type in the malignant glioma microenvironment (Figure 6C,D), suggesting that M2 macrophages would play an important role in malignant

glioma tumorigenesis and progression. Moreover, the proportions of M2 macrophages in the VPS33B (Figure 6E), DPY19L2P1 (Figure 6F), CT62 (Figure 6G), and KCNH1-IT1 (Figure 6H) low expression groups were higher than those in the high expression groups, while the proportions of M0 macrophages and M1 macrophages in the low expression groups were lower than those in the high expression groups. These results suggest that VPS33B, CT62, DPY19L2P1, and KCNH1-IT1 would suppress M0 macrophage and M1 macrophage progression into M2 macrophages. Given that VPS33B plays a critical role in vesicle-mediated transport and organization⁵³⁻⁵⁶ and that the ncRNAs CT62, DPY19L2P1, and KCNH1-IT1 can be delivered by exosomes,^{57,58} VPS33B would promote the organization and release of exosomes that carry CT62, DPY19L2P1 and KCNH1-IT1 to immune cells, suppressing M0 macrophage and M1 macrophage progression into M2 macrophages. What's more, considering TBPL1 and hsa-mir-424 that locate upstream of the pathways, the effects of TBPL1 and hsa-mir-424 have also been validated. The results show that the proportion of M2 macrophages in the TBPL1 (Figure 6I) low expression groups was higher than that in the high expression groups, while the proportion of M2 macrophages in the hsa-mir-424 (Figure 6J) low expression groups was lower than that in the high expression groups. This, again, verifies our model is reliable. In M2 macrophages, arginine metabolism converts arginine into ornithine and polyamines, promoting proliferation and repair through polyamine and collagen synthesis, fibrosis and other tissue remodelling functions, thus promoting tumorigenesis.⁸⁵⁻⁸⁸ Therefore, the low expression of CT62, DPY19L2P1, KCNH1-IT1 and VPS33B in malignant glioma would promote the formation of M2 macrophages and thus promote malignant glioma tumorigenesis and progression (Figure 1B).

Interestingly, for a long time, the immune system was considered to play a protective role in tumour development.⁸⁹ Accumulating evidence, however, indicates that various types of immune and inflammatory cells are frequently present within tumours and are strongly correlated with tumours.⁹⁰ Exosomes, vesicles of endocytic origin released by cells, mediate communication between cells by transferring mRNAs, ncRNAs, liquids and proteins, thus promoting cell processes such as antigen presentation, tumorigenesis, neurodegeneration, immune responses, and intracellular trafficking.^{82-84,91,92} Overall, accumulating evidence highlights the important relationship between tumour cells and other cells mediated by exosomes. In our work, we reveal that the immune system plays an undeniably important role in the development of malignant glioma through exosomes. Thus, understanding the relationship among malignant glioma, the immune microenvironment, and exosomes is vital for illustrating the underlying mechanisms and improving malignant glioma therapy.

4.3 | Cellular senescence

After illustrating the circadian rhythm pathway and tumour immune microenvironment pathway, only one coding gene, C9ORF40, and three ncRNAs, AC008738.2, AC102941.1 and AL136115.1, remain

to be illustrated. Although C9ORF40 has no relationship with either the circadian rhythm or secretory granule organization, it plays a critical role in cellular senescence by activating the p21 pathway.⁶¹ p21, which is encoded by CDKN1A, is a key protein in cellular senescence and causes G1 phase arrest by interacting with CDK1 or CDK2 to disrupt the interaction between substrates and CDKs, thus inhibiting cell cycle progression.⁶²⁻⁶⁴ p21 can directly bind to PCNA and inhibit its activation, which inhibits DNA polymerase activity, transcription, and excision repair functions[65]. Moreover, p21 induces apoptosis by activating the caspase-3-mediated signalling pathway.^{62,63} Overall, C9ORF40 activates p21 and thus inhibits proliferation by the CDK or PCNA signalling pathway and simultaneously promotes apoptosis by the caspase-3 signalling pathway.

What are the roles of the ncRNAs AC008738.2, AC102941.1 and AL136115.1 in the cellular senescence pathway? One possible explanation is that the ncRNAs AC008738.2, AC102941.1 and AL136115.1 would interact with the p21, CDK, PCNA and caspase-3 proteins to regulate cellular senescence. Therefore, the question is whether p21, CDK, PCNA and caspase-3 are RNA-binding proteins. PCNA was identified as an RNA-binding protein[93], while the others were not. To test whether the ncRNAs AC008738.2, AC102941.1 and AL136115.1 could bind to p21, CDK, PCNA and/or caspase-3, molecular docking analysis was performed. The results revealed that only PCNA would interact with the ncRNAs AL136115.1 (Figure 5D), AC008738.2 (Figure S2G), and AC102941.1 (Figure S2H). In eukaryotes, the PCNA trimer forms a hexagon-like structure with a cavity where DNA passes through PCNA and binds to DNA polymerase δ (Pol δ) to replicate the lagging strand (Figure 5E) and cooperates with flap endonuclease 1 (FEN1) to process the Okazaki fragments for ligation[66]. However, the DNA binding cavity for PCNA would be occupied by the ncRNAs AC008738.2, AC102941.1 and AL136115.1, hindering and impeding DNA replication and thus suppressing proliferation and other cell cycle processes in malignant glioma.

In particular, the process of cellular senescence is considered a state of irreversible growth arrest. Cancer cells, however, do not enter this state and begin to proliferate infinitely.⁹⁴ Cellular senescence was found to be mediated by the two main tumour suppressor pathways of the cell: the ARF/p53 and INK4a/RB pathways.⁹⁴⁻⁹⁶ Overall, overwhelming evidence indicates that cellular senescence is a critical feature of mammalian cells to suppress tumorigenesis. In our work, we also revealed that cellular senescence is significantly correlated with glioma tumorigenesis (Figure 1B).

5 | CONCLUSIONS

- Nineteen genes (12 ncRNAs: AC020907.1, Y_RNA, TMEM72-AS1, KRT16P2, DLX6-AS1, AC008738.2, AC102941.1, AL136115.1, CT62, DPY19L2P1 and KCNH1-IT1; 1 miRNA: hsa-miR-424; and 6 coding genes: TBPL1, C9ORF40, CRY2, ETNK1, NPAS2 and VPS33B) were used to construct a model to evaluate malignant glioma patients (AUC of approximately 0.91 and C-index of

approximately 0.84). These 19 genes could be used as diagnostic/prognostic biomarkers for malignant glioma.

- The model is very suitable for LGG and is also suitable for GBM (AUC of approximately 0.90 and C-index of approximately 0.86 in LGG and AUC of approximately 0.75 and C-index of approximately 0.69 in GBM), which highlights the partial similarities and differences between LGG and GBM.
- The ncRNAs AC020907.1, Y_RNA, TMEM72-AS1, KRT16P2, DLX6-AS1 and AP002414.1 would competitively interact with hsa-miR-424, which would bind to TBPL1 mRNA, indirectly promoting TBPL1 expression.
- TBPL1 would bind to the NPAS2 promoter to promote its expression, and NPAS2 would interact with BMAL1 to promote CRY2 and PER2 expression and suppress C-MYC expression. CRY2, which complexes with PER2 would bind to the promoters of C9ORF40, ETNK1, and VPS33B to promote their expression. Low ETNK1 expression would directly promote malignant glioma proliferation.
- VPS33B would promote the organization and release of exosomes that deliver CT62, DPY19L2P1, and KCNH1-IT1 to immune cells, suppressing the differentiation of M1 macrophages into M2 macrophages.
- C9ORF40 would promote cellular senescence and thus inhibit tumorigenesis through p21/PCNA, p21/CDK, and p21/caspase-3.
- The ncRNAs AC008738.2, AC102941.1, and AL136115.1 would bind to PCNA and thus inhibit cell proliferation and promote cellular senescence.

ACKNOWLEDGEMENTS

We thank members of the Institute of Neurosurgery and the State Key Laboratory of Cancer Biology for their input and stimulating discussions and many other colleagues in the Department of Neurosurgery for helpful comments on this manuscript. We thank Research Assistant Xiaoyan Chen and Dan Liao for their assistance in IHC. We thank Jiaming Sun at TuSimple, Inc., and Lei Gan at NavInfo Co., Ltd., for their helpful suggestions on computer programming.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

Yutao Huang: Conceptualization (lead); Data curation (lead); Formal analysis (lead); Investigation (lead); Methodology (lead); Resources (lead); Software (lead); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review & editing (lead). **Xiangyu Gao:** Investigation (equal); Methodology (equal); Resources (equal); Validation (equal). **Erwan Yang:** Investigation (equal); Methodology (equal); Resources (equal); Validation (equal). **Kangyi Yue:** Investigation (equal); Methodology (equal); Resources (equal); Validation (equal). **Yuan Cao:** Investigation (equal); Resources (equal). **Boyan Zhao:** Investigation (equal); Resources (equal). **Haofuzi Zhang:** Investigation (equal); Resources (equal). **Shuhui Dai:** Investigation (equal); Resources (equal). **Lei Zhang:** Investigation (equal); Resources (equal). **Peng Luo:**

Conceptualization (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – review & editing (equal). **Xiaofan Jiang:** Funding acquisition (equal); Project administration (equal); Supervision (equal); Writing – review & editing (equal).

ORCID

Yutao Huang  <https://orcid.org/0000-0002-2433-8183>

Peng Luo  <https://orcid.org/0000-0003-0746-939X>

REFERENCES

1. Dolecek TA, Propp JM, Stroup NE, Kruchko C. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2005–2009. *Neuro-oncology*. 2012;14(Suppl 5):v1-49.
2. Omuro A, DeAngelis LM. Glioblastoma and other malignant gliomas: a clinical review. *JAMA*. 2013;310(17):1842-1850.
3. Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol*. 2016;131(6):803-820.
4. Wen PY, Kesari S. Malignant gliomas in adults. *N Engl J Med*. 2008;359(5):492-507.
5. Schiff D, Van den Bent M, Vogelbaum MA, et al. Recent developments and future directions in adult lower-grade gliomas: Society for Neuro-Oncology (SNO) and European Association of Neuro-Oncology (EANO) consensus. *Neuro-oncology*. 2019;21(7):837-853.
6. Phillips RE, Soshnev AA, Allis CD. Epigenomic reprogramming as a driver of malignant glioma. *Cancer Cell*. 2020;38(5):647-660.
7. Latha K, Yan J, Yang Y, et al. The role of fibrinogen-like protein 2 on immunosuppression and malignant progression in glioma. *J Natl Cancer Inst*. 2019;111(3):292-300.
8. Stupp R, Mason WP, van den Bent MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med*. 2005;352(10):987-996.
9. Louis DN, Ohgaki H, Wiestler OD, et al. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol*. 2007;114(2):97-109.
10. Su YT, Phan FP, Wu J. Perspectives on IDH mutation in diffuse Gliomas. *Trend Cancer*. 2018;4(9):605-607.
11. Kristensen BW, Priesterbach-Ackley LP, Petersen JK, Wesseling P. Molecular pathology of tumors of the central nervous system. *Ann Oncol*. 2019;30(8):1265-1278.
12. Gussyatiner O, Hegi ME. Glioma epigenetics: from subclassification to novel treatment options. *Semin Cancer Biol*. 2018;51:50-58.
13. Bourne TD, Schiff D. Update on molecular findings, management and outcome in low-grade gliomas. *Nat Rev Neurol*. 2010;6(12):695-701.
14. Bhan A, Soleimani M, Mandal SS. Long Noncoding RNA and Cancer: A New Paradigm. *Can Res*. 2017;77(15):3965-3981.
15. Prensner JR, Chinnaiyan AM. The Emergence of lncRNAs in Cancer Biology. *Cancer Discov*. 2011;1(5):391-407.
16. Hansen TB, Jensen TI, Clausen BH, et al. Natural RNA circles function as efficient microRNA sponges. *Nature*. 2013;495(7441):384-388.
17. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell*. 2013;154(1):26-46.
18. Lei B, Huang Y, Zhou Z, et al. Circular RNA hsa_circ_0076248 promotes oncogenesis of glioma by sponging miR-181a to modulate SIRT1 expression. *J Cell Biochem*. 2019;120(4):6698-6708.
19. Xin J, Du M, Jiang X, et al. Systematic evaluation of the effects of genetic variants on PIWI-interacting RNA expression across 33 cancer types. *Nucleic Acids Res*. 2020.

20. Slack FJ, Chinnaiyan AM. The role of non-coding RNAs in oncology. *Cell*. 2019;179(5):1033-1055.
21. Milligan MJ, Harvey E, Yu A, et al. Global intersection of long non-coding RNAs with processed and unprocessed pseudogenes in the human genome. *Front Genet*. 2016;7:26.
22. Wang Y, Hou J, He D, et al. The emerging function and mechanism of ceRNAs in cancer. *Trends Genet*. 2016;32(4):211-224.
23. Wirth N. Program development by stepwise refinement. *Commun ACM*. 1971;14(4):221-227.
24. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
25. Phipson B, Lee S, Majewski IJ, Alexander WS, Smyth GK. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann Appl Stat*. 2016;10(2):946-963.
26. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014;15(2):R29.
27. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*. 2010;26(1):139-140.
28. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res*. 2020;48(D1):D127-D131.
29. Chou CH, Shrestha S, Yang CD, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res*. 2018;46(D1):D296-D302.
30. Agarwal V, Bell GW, Nam Jin-Wu, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *eLife*. 2015;4:10.7554/eLife.05005
31. Garcia DM, Baek D, Shin C, Bell GW, Grimson A, Bartel DP. Weak seed-pairing stability and high target-site abundance decrease the proficiency of *Isy-6* and other microRNAs. *Nat Struct Mol Biol*. 2011;18(10):1139-1146.
32. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498-2504.
33. Benson DA, Cavanaugh M, Clark K, et al. *Nucleic Acids Res*. 2013;41(D1):D36-D42.
34. *Nucleic Acids Res*. UniProt: a worldwide hub of protein knowledge. 2019;47(D1):D506-d515.
35. Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res*. 1981;9(1):133-148.
36. Zhao Y, Huang Y, Gong Z, Wang Y, Man J, Xiao Y. Automated and fast building of three-dimensional RNA structures. *Sci Rep*. 2012;2:734.
37. Havel TF. Distance geometry: Theory, algorithms, and chemical applications. *Encyclopedia of Computational Chemistry*. 1998;120:723-742.
38. Wang J, Zhao Y, Zhu C, Xiao Y. 3dRNAscore: a distance and torsion angle dependent evaluation function of 3D RNA structures. *Nucleic Acids Res*. 2015;43(10):e63.
39. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015;10(6):845-858.
40. Yan Y, Tao H, He J, Huang SY. The HDock server for integrated protein-protein docking. *Nat Protoc*. 2020;15(5):1829-1852.
41. Mathelier A, Zhao X, Zhang AW, et al. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res* 2014, 42(Database issue):D142-147.
42. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. *Nature*. 2014;505(7483):344-352.
43. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*. 2011;146(3):353-358.
44. Turkalp Z, Karamchandani J, Das S. IDH mutation in glioma: new insights and promises for the future. *JAMA Neurol*. 2014;71(10):1319-1325.
45. Juratli TA, Kirsch M, Robel K, et al. IDH mutations as an early and consistent marker in low-grade astrocytomas WHO grade II and their consecutive secondary high-grade gliomas. *J Neurooncol*. 2012;108(3):403-410.
46. Reuss DE, Sahm F, Schrimpf D, et al. ATRX and IDH1-R132H immunohistochemistry with subsequent copy number analysis and IDH sequencing as a basis for an "integrated" diagnostic approach for adult astrocytoma, oligodendroglioma and glioblastoma. *Acta Neuropathol*. 2015;129(1):133-146.
47. Tommasini-Ghelfi S, Murnan K, Kouri FM, Mahajan AS, May JL, Stegh AH. Cancer-associated mutation and beyond: The emerging biology of isocitrate dehydrogenases in human disease. *Science Adv*. 2019;5(5):eaaw4543.
48. Bass J, Takahashi JS. Circadian integration of metabolism and energetics. *Science*. 2010;330(6009):1349-1354.
49. Takahashi JS, Hong HK, Ko CH, McDearmon EL. The genetics of mammalian circadian order and disorder: implications for physiology and disease. *Nat Rev Genet*. 2008;9(10):764-775.
50. Schmutz I, Ripperger JA, Baeriswyl-Aebischer S, Albrecht U. The mammalian clock component PERIOD2 coordinates circadian output by interaction with nuclear receptors. *Genes Dev*. 2010;24(4):345-357.
51. Chong JA, Moran MM, Teichmann M, Kaczmarek JS, Roeder R, Clapham DE. TATA-binding protein (TBP)-like factor (TLF) is a functional regulator of transcription: reciprocal regulation of the neurofibromatosis type 1 and c-fos genes by TLF/TRF2 and TBP. *Mol Cell Biol*. 2005;25(7):2632-2643.
52. Wang YL, Duttke SH, Chen K, et al. TRF2, but not TBP, mediates the transcription of ribosomal protein genes. *Genes Dev*. 2014;28(14):1550-1555.
53. Gissen P, Johnson CA, Gentle D, et al. Comparative evolutionary analysis of VPS33 homologues: genetic and functional insights. *Hum Mol Genet*. 2005;14(10):1261-1270.
54. Hershkovitz D, Mandel H, Ishida-Yamamoto A, et al. Defective lamellar granule secretion in arthrogyrosis, renal dysfunction, and cholestasis syndrome caused by a mutation in VPS33B. *Arch Dermatol*. 2008;144(3):334-340.
55. Pevsner J, Hsu SC, Hyde PS, Scheller RH. Mammalian homologues of yeast vacuolar protein sorting (*vps*) genes implicated in Golgi-to-lysosome trafficking. *Gene*. 1996;183(1-2):7-14.
56. Akbar MA, Ray S, Krämer H. The SM protein Car/Vps33A regulates SNARE-mediated trafficking to lysosomes and lysosome-related organelles. *Mol Biol Cell*. 2009;20(6):1705-1714.
57. Pefanis E, Wang JG, Rothschild G, et al. RNA exosome-regulated long non-coding rna transcription controls super-enhancer activity. *Cell*. 2015;161(4):774-789.
58. Li S, Li Y, Chen B, et al. exoRBase: a database of circRNA, lncRNA and mRNA in human blood exosomes. *Nucleic Acids Res*. 2018;46(D1):D106-d112.
59. Whiteside TL. The tumor microenvironment and its role in promoting tumor growth. *Oncogene*. 2008;27(45):5904-5912.
60. Hinshaw DC, Shevde LA. The tumor microenvironment innately modulates cancer progression. *Can Res*. 2019;79(18):4557-4566.
61. Avelar RA, Ortega JG, Tacutu R, et al. A multidimensional systems biology analysis of cellular senescence in aging and disease. *Genome Biol*. 2020;21(1):91.
62. Warfel NA, El-Deiry WS. p21WAF1 and tumorigenesis: 20 years after. *Curr Opin Oncol*. 2013;25(1):52-58.
63. Jin YH, Yoo KJ, Lee YH, Lee SK. Caspase 3-mediated cleavage of p21WAF1/CIP1 associated with the cyclin A-cyclin-dependent kinase 2 complex is a prerequisite for apoptosis in SK-HEP-1 cells. *J Biol Chem*. 2000;275(39):30256-30263.

64. Chen J, Saha P, Kornbluth S, Dynlacht BD, Dutta A. Cyclin-binding motifs are essential for the function of p21CIP1. *Mol Cell Biol.* 1996;16(9):4673–4682.
65. Luo Y, Hurwitz J, Massagué J. Cell-cycle inhibition by independent CDK and PCNA binding domains in p21Cip1. *Nature.* 1995;375(6527):159–161.
66. Lancey C, Tehseen M, Raducanu V-S, et al. Structure of the processive human Pol δ holoenzyme. *Nat Commun.* 2020;11(1):1109.
67. Chang J, Garva R, Pickard A, et al. Circadian control of the secretory pathway maintains collagen homeostasis. *Nat Cell Biol.* 2020;22(1):74–86.
68. Lykidis A, Wang J, Karim MA, Jackowski S. Overexpression of a mammalian ethanalamine-specific kinase accelerates the CDP-ethanalamine pathway. *J Biol Chem.* 2001;276(3):2174–2179.
69. Ratovitski EA. Delta Np63 alpha – Responsive microRNA modulate the expression of metabolic enzymes. *Curr Pharm Biotechnol.* 2015;16(9):832–850.
70. Chang H, Rha SY, Jeung HC, et al. Identification of genes related to a synergistic effect of taxane and suberoylanilide hydroxamic acid combination treatment in gastric cancer cells. *J Cancer Res Clin Oncol.* 2010;136(12):1901–1913.
71. Fu LN, Pelicano H, Liu JS, Huang P, Lee CC. The circadian gene Period2 plays an important role in tumor suppression and DNA damage response in vivo. *Cell.* 2002;111(1):41–50.
72. Fu LN, Lee CC. The circadian clock: Pacemaker and tumour suppressor. *Nat Rev Cancer.* 2003;3(5):350–361.
73. Bell-Pedersen D, Cassone VM, Earnest DJ, et al. Circadian rhythms from multiple oscillators: Lessons from diverse organisms. *Nat Rev Genet.* 2005;6(7):544–556.
74. Lowrey PL, Takahashi JS. Mammalian circadian biology: Elucidating genome-wide levels of temporal organization. *Annu Rev Genomics Hum Genet.* 2004;5:407–441.
75. Papagiannakopoulos T, Bauer MR, Davidson SM, et al. Circadian rhythm disruption promotes lung tumorigenesis. *Cell Metab.* 2016;24(2):324–331.
76. Videnovic A, Lazar AS, Barker RA, Overeem S. 'The clocks that time us'-circadian rhythms in neurodegenerative disorders. *Nat Rev Neurol.* 2014;10(12):683–693.
77. Hatfield CF, Herbert J, van Someren EJ, Hodges JR, Hastings MH. Disrupted daily activity/rest cycles in relation to daily cortisol rhythms of home-dwelling patients with early Alzheimer's dementia. *Brain.* 2004;127(Pt 5):1061–1074.
78. Gombert M, Carrasco-Luna J, Pin-Arboledas G, Codoñer-Franch P. The connection of circadian rhythm to inflammatory bowel disease. *Transl Res.* 2019;206:107–118.
79. Ji Y, Elkin K, Yip J, Guan L, Han W, Ding Y. From circadian clocks to non-alcoholic fatty liver disease. *Expert Rev Gastroenterol Hepatol.* 2019;13(11):1107–1112.
80. Rabinovich-Nikitin I, Lieberman B, Martino TA, Kirshenbaum LA. Circadian-regulated cell death in cardiovascular diseases. *Circulation.* 2019;139(7):965–980.
81. Masri S, Sassone-Corsi P. The emerging link between cancer, metabolism, and circadian rhythms. *Nat Med.* 2018;24(12):1795–1803.
82. Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ, Lotvall JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol.* 2007;9(6):654–U672.
83. Hoshino A, Costa-Silva B, Shen T-L, et al. Tumour exosome integrins determine organotropic metastasis. *Nature.* 2015;527(7578):329–335.
84. Bobrie A, Colombo M, Raposo G, Thery C. Exosome secretion: molecular mechanisms and roles in immune responses. *Traffic (Copenhagen, Denmark).* 2011;12(12):1659–1668.
85. Mills CD, Kincaid K, Alt JM, Heilman MJ, Hill AM. M-1/M-2 macrophages and the Th1/Th2 paradigm. *J Immunol (Baltimore, Md: 1950)* 2000;164(12):6166–6173.
86. Pesce JT, Ramalingam TR, Mentink-Kane MM, et al. Arginase-1-expressing macrophages suppress Th2 cytokine-driven inflammation and fibrosis. *PLoS Pathog.* 2009;5(4):e1000371.
87. MacMicking J, Xie QW, Nathan C. Nitric oxide and macrophage function. *Annu Rev Immunol.* 1997;15:323–350.
88. Italiani P, Boraschi D. From Monocytes to M1/M2 Macrophages: Phenotypical vs. Functional Differentiation. *Front Immunol.* 2014;5:514.
89. Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoeediting: from immunosurveillance to tumor escape. *Nat Immunol.* 2002;3(11):991–998.
90. Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. *Cell.* 2010;140(6):883–899.
91. Webber J, Steadman R, Mason MD, Tabi Z, Clayton A. Cancer exosomes trigger fibroblast to myofibroblast differentiation. *Can Res.* 2010;70(23):9621–9630.
92. Janowska-Wieczorek A, Wysoczynski M, Kijowski J, et al. Microvesicles derived from activated platelets induce metastasis and angiogenesis in lung cancer. *Int J Cancer.* 2005;113(5):752–760.
93. Baltz A, Munschauer M, Schwanhäusser B, et al. The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell.* 2012;46(5):674–690.
94. Collado M, Blasco MA, Serrano M. Cellular senescence in cancer and aging. *Cell.* 2007;130(2):223–233.
95. Gil J, Peters G. Regulation of the INK4b-ARF-INK4a tumour suppressor locus: all for one or one for all. *Nat Rev Mol Cell Biol.* 2006;7(9):667–677.
96. Kim WY, Sharpless NE. The regulation of INK4/ARF in cancer and aging. *Cell.* 2006;127(2):265–275.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Huang Y, Gao X, Yang E, et al. Top-down stepwise refinement identifies coding and noncoding RNA-associated epigenetic regulatory maps in malignant glioma. *J Cell Mol Med.* 2022;26:2230–2250. doi:[10.1111/jcmm.17244](https://doi.org/10.1111/jcmm.17244)