scientific reports

Check for updates

OPEN

Prediction of lung adenocarcinoma prognosis and diagnosis with a novel model anchored in circadian clock-related genes

Qihang Sun[1,2], Shubin Zheng[1,2], Wei Tang[1], Xiaoyu Wang[1], Qi Wang[1], Ruijie Zhang[1], Ni Zhang[1]✉ & Wei Ping[1]✉

Lung adenocarcinoma is the most common primary lung cancer seen in the world, and identifying genetic markers is essential for predicting the prognosis of lung adenocarcinoma and improving treatment outcomes. It is well known that alterations in circadian rhythms are associated with a higher risk of cancer. Moreover, circadian rhythms play a regulatory role in the human body. Therefore, studying the changes in circadian rhythms in cancer patients is crucial for optimizing treatment. The gene expression data and clinical data were sourced from TCGA database, and we identified the circadian clock-related genes. We used the obtained TCGA-LUAD data set to build the model, and the other 647 lung adenocarcinoma patients' data were collected from two GEO data sets for external verification. A risk score model for circadian clock-related genes was constructed, based on the identification of 8 genetically significant genes. Based on ROC analyses, the risk model demonstrated a high level of accuracy in predicting the overall survival times of lung adenocarcinoma patients in training folds, as well as external data sets. This study has successfully constructed a risk model for lung adenocarcinoma prognosis, utilizing circadian rhythm as its foundation. This model demonstrates a dependable capacity to forecast the outcome of the disease, which can further guide the relevant mechanism of lung adenocarcinoma and combine behavioral therapy with treatment to optimize treatment decision-making.

Keywords Lung adenocarcinoma, Circadian rhythm, Bioinformatics, Prognosis, Immune infiltration

While cancer remains an enduring global health challenge of paramount concern. According to the Global Cancer Burden data from the IARC, the year 2020 alone witnessed a staggering 19.29 million new cancer cases worldwide, with lung cancer accounting for a formidable 2.27 million cases, and lung cancer stands as the second most frequently diagnosed malignancy and the leading cause of cancer-related mortality[1,2].The American Cancer Society estimates that in 2022, the United States will witness approximately 350 new lung cancer-related fatalities each day[3]. Smoking continues to be the predominant risk factor for lung cancer. Beyond smoking, other risk factors further heighten the susceptibility to lung cancer, encompassing environmental and occupational exposures, chronic respiratory conditions, and lifestyle factors[4–6].

Lung adenocarcinoma, which is the most prevalent histological subtype of non-small cell lung cancer (NSCLC), poses a significant global health challenge, accounting for a substantial share of cancer-related morbidity and mortality[7]. This cancer originate from the bronchial mucosal epithelium, and occasionally from the mucous glands of the larger bronchi[8]. Surgery is currently the most effective treatment for early-stage LUAD. However, non-surgical radiotherapy and chemotherapy have become the primary modalities for managing advanced-stage lung cancer patients, owing to its insidious nature and lack of specificity, often presents with advanced-stage diagnoses[9]. Despite considerable progress in deciphering its genetic and molecular intricacies, many facets of this intricate disease remain elusive. The clinical prognosis of lung adenocarcinoma, characterized by inherent heterogeneity and diverse outcomes, continues to be a complex puzzle[4,10,11]. Thus, it is vital to establish an effective prognostic model.

[1]Department of Thoracic Surgery, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China. [2]These authors contributed equally: Qihang Sun and Shubin Zheng. ✉email: nizhang@tjh.tjmu.edu.cn; 247046170@qq.com

The circadian rhythm represents a fundamental phenomenon in the organism's life processes, encompassing physiological, biochemical, and behavioral aspects, cyclically driven by clock genes and clock-controlled genes with a periodicity of approximately 24 hours[12]. Circadian rhythm plays a pivotal role in maintaining internal equilibrium. In recent years, the World Health Organization (WHO) has identified circadian rhythm disruption as a plausible carcinogen based on population and laboratory research findings. Perturbations in circadian rhythms have been linked not only to a heightened cancer risk but also to compromised treatment outcomes and early mortality in cancer patients, including breast, colon, liver, prostate, pancreatic, ovarian, and lung cancers[13–17]. Besides, disruption of the circadian rhythm plays a key role in tumorigenesis and facilitates the establishment of cancer hallmarks[18].

Moreover, emerging evidence suggests the involvement of circadian rhythms in the tumor immune microenvironment[19–21]. Dysregulation of circadian rhythm genes can impact critical pathways in cancer development and progression, including metabolic regulation, cell cycle control, apoptosis, and DNA damage response[22]. Given the substantial influence of the tumor immune microenvironment on the efficacy of cancer immunotherapy, circadian rhythms may potentially influence the sensitivity to immunotherapeutic interventions by modulating the immune milieu[23]. Furthermore, it can lead to increased cellular heterogeneity, affecting gene transcription and modification, which in turn accelerates tumor growth[24–26]. Therefore, it is very important to consider circadian clock-associated genes for cancer treatment optimization. Nevertheless, the impact of circadian clock-related genes on the prognosis of lung adenocarcinoma patients, as well as their potential as prognostic biomarkers and therapeutic targets, presents an intriguing area for further exploration.

This paper embarks on a comprehensive exploration of the complex relationship between circadian clock-related genes and the outcome of lung adenocarcinoma. Leveraging advanced computational techniques, including monogram and Lasso regression models, we delve into the molecular intricacies of these genes and their multifaceted roles in cancer biology. Although we do not present experimental evidence, our research provides significant findings into the potential of circadian clock-related genes as key determinants of clinical outcomes in lung adenocarcinoma. Through a rigorous analysis of clinical data and a sophisticated predictive model, our endeavor is to unearth previously uncharted territory at the nexus of circadian biology and cancer prognosis. We anticipate that the findings from this investigation will provide a solid foundation for future advancements in precision oncology, offering innovative avenues for personalized treatment strategies and enhancing the overall management of lung adenocarcinoma patients.

Materials and methods

Data set acquisition and data processing

The NCI Genomic Data Commons (https://portal.gdc.cancer.gov/repository) was utilized to obtain the clinical characteristics and RNA sequencing data (HTSeq-FPKM) of 515 patients diagnosed with LUAD. From this cohort, 535 samples were collected from LUAD tumor tissue, while 59 samples were obtained from adjacent normal tissue. 469 of 500 individuals have both complete sequencing data and comprehensive clinical information available.

297 genes related to the circadian rhythm were extracted from 13-gene sets in MSigDB[27,28], and they are presented in Table S1. The circadian clock-related genes (CCRGs) expression validation data sets GSE31210[29] and GSE68465[30] were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The data containing normalized counts was downloaded, with a cut-off date of February 2, 2023. Patients without follow-up data or information on circadian gene expression were excluded from the analysis.

The DESeq2 algorithm was used for gene expression data processing[31]. Data retrieved from the TCGA and GEO databases is a public resource, so ethical clearance was not deemed necessary for the conduct of this research, which adhered to their data access and publication policies.

Investigation of differentially expressed gene and pinpointing the circadian clock-related genes

Primary objective of this step was identifying DEGs within the tumor and normal samples using the "DESeq2" package, with a p adjust < 0.05, and fold change > 2 or < 0.5. We analyzed the DEGs obtained from the above steps, along with 297 circadian clock-related genes by Venn diagram and identified 76 genes related to circadian rhythm. The gene expression data from the GSE31210 and GSE68465 datasets were obtained from the GEO repository. The "ComplexHeatmap" package for R software (version 4.2.1) was utilized to create a gene expression heatmap representing the 76 circadian clock-related genes[32]. Functional enrichment analysis and visualization were conducted on a set of 89 genes associated with immune infiltration using the software packages "clusterProfiler," "org.Hs.eg.db," and "GOplot." [33–35].

Establishment and validation of a risk scoring system

Preliminary univariate Cox regression analyses were carried out among the 76 genes associated with circadian rhythm. Subsequently, 16 circadian clock-related genes were chosen for additional examination due to having a p value < 0.05. The 16 genes underwent LASSO tenfold cross-validation using the "glmnet" and "survival" packages[36,37]. Through the analysis, a selection of the 8 most significant predictive genes and the risk-score models was made. Following this, the GENEMANIA software was employed to construct protein–protein interaction (PPI) networks for these identified genes[38]. After that, the genes were assimilated into risk features, resulting in the creation of a risk evaluation mechanism that was founded upon the standardized values of gene expression and their respective coefficients. The risk evaluation mechanism was devised according to the subsequent mathematical expression: Risk score=$\sum_{i=1}^{n} \boldsymbol{expr}_{genei} \times \boldsymbol{coeffieicent}_{genei}$ [39]. The TMM algorithm from the "edgeR" package was utilized to compute standardized gene expression level[40]. The risk-factor plot was created with the

"ggplot2", and ROC curves were generated utilizing the "timeROC" software package[41,42]. Patients were divided into high risk group and low risk group depending on the median risk score. Survival curves were generated using the "survminer" package[43], and the association between risk score and clinical features was examined through dot plots produced by the "ggplot2" software package.

Building and evaluation of the nomogram

To ascertain the viability of the risk scoring system as the stand-alone predictor, the eight gene expression parameters (*ADRB1*, *LGR4*, *BMAL2*, *RORA*, *TYMS*, *NPAS2*, *PTGDS*, and *SFTPC*) underwent both univariate and multivariate Cox regression analyses using the "survival" package. And by using the "rms" package, a nomogram, which helped predict overall survival (OS) probabilities for 1, 3, 5 years, was constructed based on the genes. ROC and calibration analyses were carried out to assess the discriminatory power of the nomogram. We integrate the data of GSE31210 and GSE68465 as the external verification group of our clinical prediction model to test the effectiveness of the model.

GSEA analysis

The "clusterProfiler" package was employed for Gene Set Enrichment Analysis (GSEA), while the "ggplot2" package was utilized for data visualization.

Immune cell infiltration level and correlation analysis

Single-sample gene enrichment analysis (ssGSEA) was introduced to quantify the correlation of 28 immune cell types in the circadian clock-related genes[44–46].

Statistical analysis

R software (version 4.2.1) was utilized to carry out all statistical analyses in this study. Kaplan–Meier survival analysis was performed using the log-rank test, while hazard ratios (HRs) and 95% confidence intervals (CIs) were calculated during the regression analysis. For the comparison of two groups of continuous variables, the statistical significance of normally distributed variables was estimated using the independent Student's *t* test, while the differences between non-normally distributed variables were analyzed using the Mann–Whitney U test (also known as the Wilcoxon rank-sum test). Chi-squared tests or Fisher's exact tests were employed to compare and analyze the statistical significance between two groups of categorical variables. Univariate and multivariate Cox analyses were conducted to identify independent prognostic factors. All statistical P-values were two-sided, with a p value of < 0.05 considered statistically significant.

Results

Identification of circadian clock-related genes in LUAD patients

Utilizing the DESeq2 algorithm, a total of 5382 DEGs were identified between 535 TCGA-LUDA samples and 59 normal lung samples, in accordance with the DEGs criteria. Venn diagrams were analyzed for these 5382 DEGs and 276 circadian clock-related genes retrieved from the MsigDB database. The analysis conducted above led to the identification of 76 genes related to circadian rhythm (Fig. 1A). We then validated the expression of 76 circadian clock-related genes in the TCGA-LUAD as well as the GSE31210 and GSE68465 datasets obtained from the GEO database (Fig. 1B, C and Figure S1). Figure 1D and E. display the Top10 results of each ontology enrichment analysis. From Fig. 1D, we can observe that in the GO analysis, these genes show significant enrichment in areas such as regulation of rhythmic behavior, response to radiation and light, gene expression, gene transcription, biological metabolism, and signal transduction. Figure 1E presents the KEGG results, revealing notable enrichment in embryonic neural development, signal transduction, osteoclast differentiation, amphetamine addiction, and renin secretion. The comprehensive findings were presented in Table S2.

Establishment and evaluation of risk scoring system

First, a univariate Cox regression analysis was conducted to examine the expression of 76 circadian clock-related genes in relation to the OS time of patients in the TCGA-LUAD cohort. By applying a Cox *P-value* threshold of less than 0.05, a total of 16 potential predictive genes associated with OS were identified and documented in Table S3. Subsequently, the gene sets were further refined using LASSO regression analysis, as depicted in Fig. 2A and B. From this refinement process, eight genes were determined to be the most significant predictive genes. We used RhythmicDB[47] to investigate whether these 8 genes have been experimentally validated as oscillating genes. The results showed that all 8 genes have been identified as oscillating genes in multiple species. In Homo sapiens, *PTGDS*, *ADRB1*, and *TYMS* have been confirmed as oscillating genes. Based on the formula, a risk scoring system was established and documented in Table 1. The results of Kaplan–Meier analysis are presented in Figure S2. A gene–gene interaction network was developed for the eight genes in order to investigate their functions utilizing the GeneMANIA database. The central hub node, which symbolized these genes, was encircled by 20 nodes that represented genes showing significant correlations with them. Additionally, the network highlighted the top seven related functions associated with these genes (Fig. 2C).

To examine the manifestation of the selected genes between LUAD and normal tissues, additional investigation was carried out utilizing immunohistochemical information sourced from the Human Protein Atlas (HPA) database[48]. The findings are depicted in Fig. 3. Regrettably, the immunohistochemical data for certain genes were not accessible at the time from the HPA database.

Subsequently, we assessed the correlation between eight specific genes and various clinical indicators (Fig. 4). Our analysis revealed distinct variations in the expression of "*ADRB1*", "*BMAL2*", and "*RORA*" across different

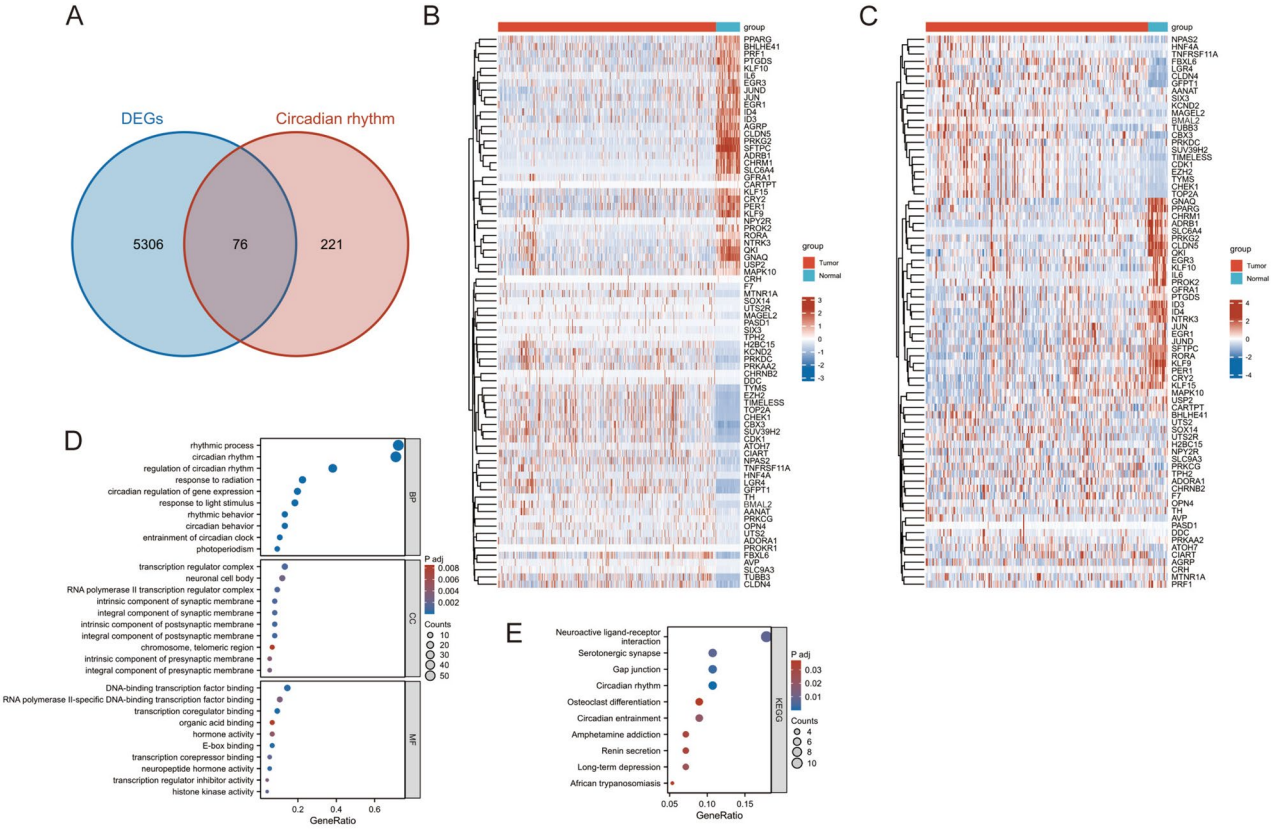

Figure 1. Identification and functional enrichment analysis of circadian clock-related genes between the TCGA-LUAD cohort and normal lung samples. (**A**) Venn diagram of the intersection between circadian clock-related genes and DEGs identified by the DESeq2 algorithm. (**B**) Heat map of 76 DEGs related to circadian rhythm in the data set TCGA-LUAD. (**C**) Heat map of 76 DEGs related to circadian rhythm in the data set GSE31210. Terms of Gene Ontology (GO) enrichment analysis (**D**) and KEGG pathways (**E**) related to the 76 circadian clock-related genes.

T stages. Similarly, "*BMAL2*", "*RORA*", and "*PTGDS*" showed differential expression across various N stages. Notably, "*TYMS*" exhibited significant differences in expression between different M stages. These five genes demonstrated substantial variability in expression across different pathological stages. We also observed significant differences in the expression levels of "*RORA*" between smokers and non-smokers. In individuals smoking ≥ 40 packs per year, the expression level of "*ADRB1*", "*LGR4*" and "*RORA*" was significantly less compared to those smoking < 40 packs annually, and the expression of "*TYMS*" demonstrated a contrasting trend. Additionally, we conducted a correlation analysis of these eight genes with different tumor locations. Besides, analysis demonstrated increased levels of "*NPAS2*" expression in tumors located on the left side, and "*BMAL2*" showed higher expression in central-type lung cancers. Besides, significant variations in the expression of these eight genes were also observed across different biological sexes, ages, and races. "*BMAL2*" and "*TYMS*" exhibited higher expression levels in males, while "*PTGDS*" showed significantly increased expression in females. The expression of "*ADRB1*" and "*PTGDS*" was more pronounced in individuals over 65 years of age. Additionally, "*LGR4*" and "*BMAL2*" demonstrated more significant expression in white people.

Through an examination of the expression levels and regression coefficients of 8 circadian clock-related genes, the risk score for each patient was calculated. The distribution of risk scores and the expression levels of the 8 circadian clock-related genes are visually represented in Fig. 5A. Subsequently, patients were categorized into high- and low-risk groups based on median risk scores.

Furthermore, the analysis of the survival time distribution revealed a clear association between risk score and prognosis, with higher risk scores indicating a worse forecast ($P < 0.001$; Fig. 5B). Additionally, our findings demonstrate a negative correlation between the expression levels of four genes (*BMAL2*, *LGR4*, *NPAS2*, and *TYMS*) and overall survival (OS), while the expression levels of four other genes (*ADRB1*, *PTGDS*, *RORA*, and *SFTPC*) show a positive correlation with OS (Figure S2). Figure 5C presents the evaluation of the risk scoring system's performance using time-ROC curves for 1-, 3- and 5-year prognoses. The AUCs for 1-, 3-, and 5-year OS times were found to be 0.726, 0.679, and 0.654, respectively.

Association between risk scores and genes

Additionally, we examined correlation ship between risk scores and circadian clock-related genes, and there are a noteworthy association among them (Fig. 6A). The high-risk group exhibit a higher level in *LGR4* ($P < 0.001$),

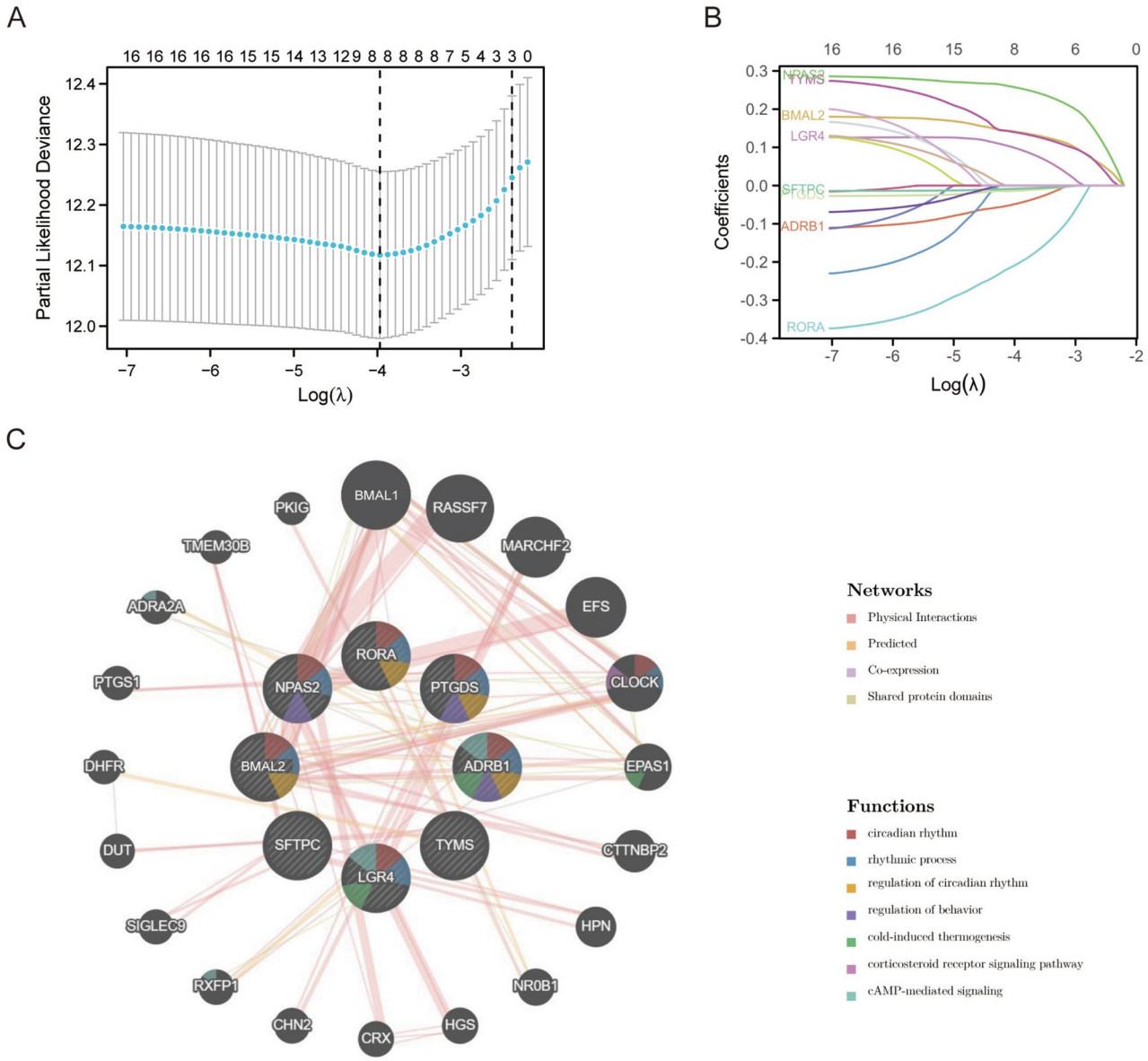

Figure 2. Demonstration of DEGs with univariate Cox regression *P value* < 0.05. (**A**) The LASSO regression model of the 16 circadian clock-related genes performed by Lasso-ten-fold cross-validation. (**B**) The coefficient distribution in the LASSO regression mode. (**C**) The gene–gene interaction network for the eight potential genes.

Gene Symbol	Description	logFC	P.Value	adj.P
ADRB1	Beta-1 adrenergic receptor	2.488927511	2.91089E − 36	8.78832E − 35
LGR4	Leucine-rich repeat-containing G-protein coupled receptor 4	− 3.147768189	1.13084E − 42	4.74852E − 41
BMAL2	Basic helix-loop-helix ARNT-like protein 2	− 1.021687521	4.75207E − 13	3.06504E − 12
RORA	Nuclear receptor ROR-alpha	− 1.72205315	1.38125E − 21	1.68858E − 20
TYMS	Thymidylate synthase	− 4.729705542	4.24937E − 37	1.33646E − 35
NPAS2	Neuronal PAS domain-containing protein 2	1.22031332	4.02077E − 16	3.29097E − 15
PTGDS	Prostaglandin-H2 D-isomerase	1.875724194	5.19283E − 48	2.88576E − 46
SFTPC	Pulmonary surfactant-associated protein C	2.720691626	1.03421E − 90	4.13116E − 88

Table 1. 8 circadian clock-related genes identified by LASSO regression analysis. adj.P, adjusted *P* value.

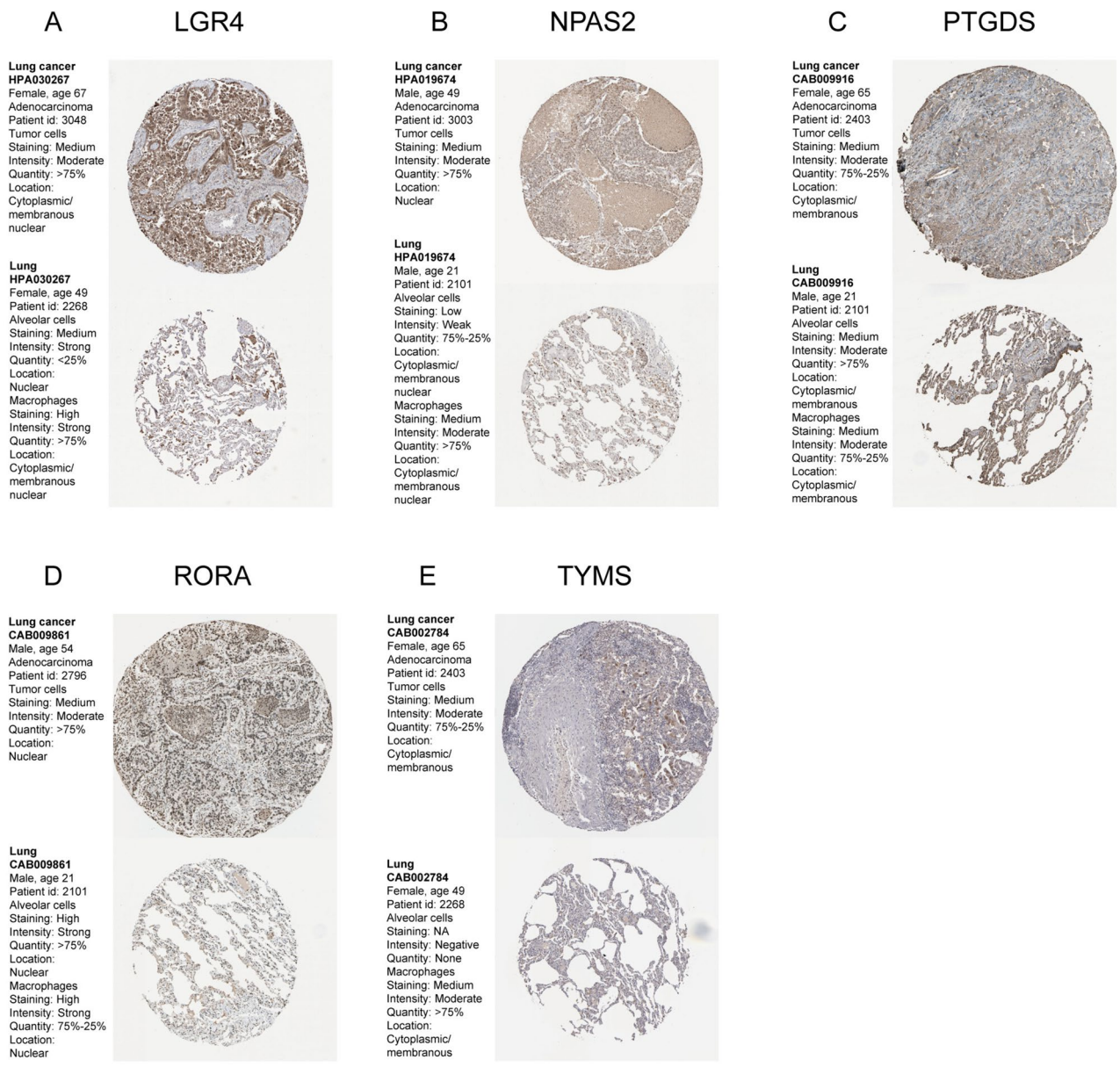

Figure 3. Immunohistochemical analysis of LUAD and normal lung tissue determined by HPA database. (**A**) *LGR4*; (**B**) *NPAS2*; (**C**) *PTGDS*; (**D**) *RORA*; (**E**) *TYMS*.

BMAL2 ($P < 0.001$), *TYMS* ($P < 0.001$) and *NPAS2* ($P < 0.001$), the low-risk group has a higher score in *ADRB1* ($P < 0.001$), *RORA* ($P < 0.001$), *PTGDS* ($P < 0.001$), *SFTPC* ($P < 0.001$). Besides, the scores had a negative correlation with *ADRB1*, *PTGDS*, *RORA* and *SFTPC*, while positively correlated with *BMAL2*, *LGR4*, *NPAS2* and *TYMS* (Fig. 6B-I).

Development and validation of nomogram

Initially, univariate and multivariate Cox regression analyses were conducted on candidate circadian clock-associated genes, including *ADRB1*, *LGR4*, *BMAL2*, *RORA*, *TYMS*, *NPAS2*, *PTGDS*, and *SFTPC*, to investigate their impact on the prognosis of patients with lung adenocarcinoma (Table 2).

The findings of the study indicated that several genes, namely *LGR4*, *BMAL2*, *RORA*, *TYMS*, *NPAS2*, and *SFTPC*, were identified as autonomous predictors of overall survival (OS) among individuals diagnosed with lung adenocarcinoma. These 8 genes were then incorporated into a nomogram model, as depicted in Fig. 7A. The nomogram model we developed exhibited a C-index of 0.692 (0.664–0.720), indicating its moderate predictive accuracy. Subsequently, we assessed its predictive ability through ROC analysis. In the TCGA-LUAD cohort, the area under the AUC curve values of the nomogram for the 1-, 3-, and 5-year OS were 0.710, 0.667, and 0.648, respectively, as shown in Fig. 7B.

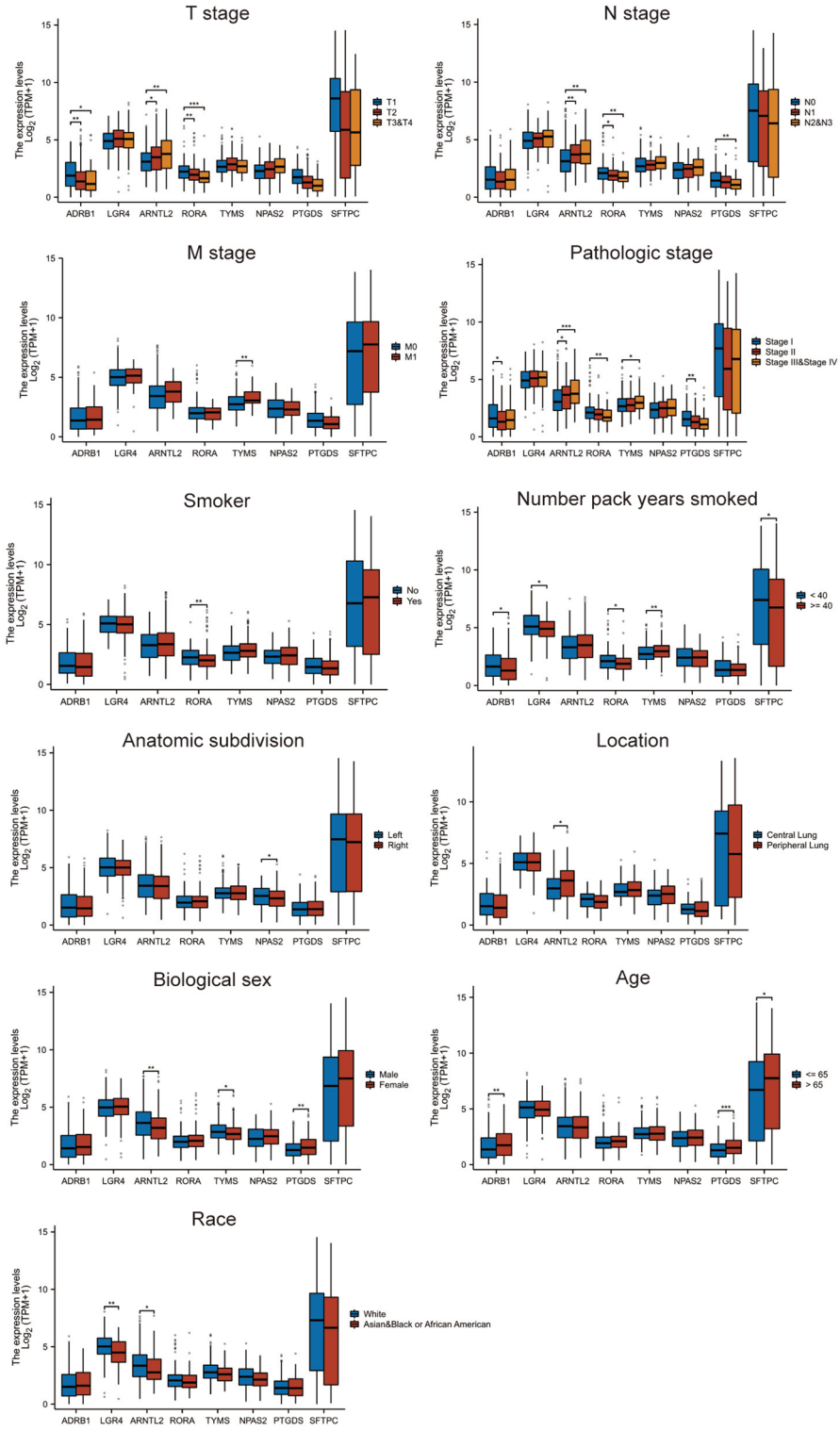

Figure 4. The correlation between eight specific genes and various clinical indicators.

Furthermore, we employed calibration curves to assess the agreement of the nomogram. The calibration plots of the model demonstrated favorable coherence with the ideal model for the 1-, 3-, and 5-year OS rates, as depicted in Fig. 7C.

We integrate the data of GSE31210 and GSE68465 as an external verification group to test the effectiveness of the model. The C-index of the external validation set was 0.686(0.669–0.703). The AUCs for the 1-, 3-, and 5-year were 0.685, 0.659, 0.645 (Fig. 7D), and we also used the calibration curve to evaluate the effectiveness (Fig. 7E).

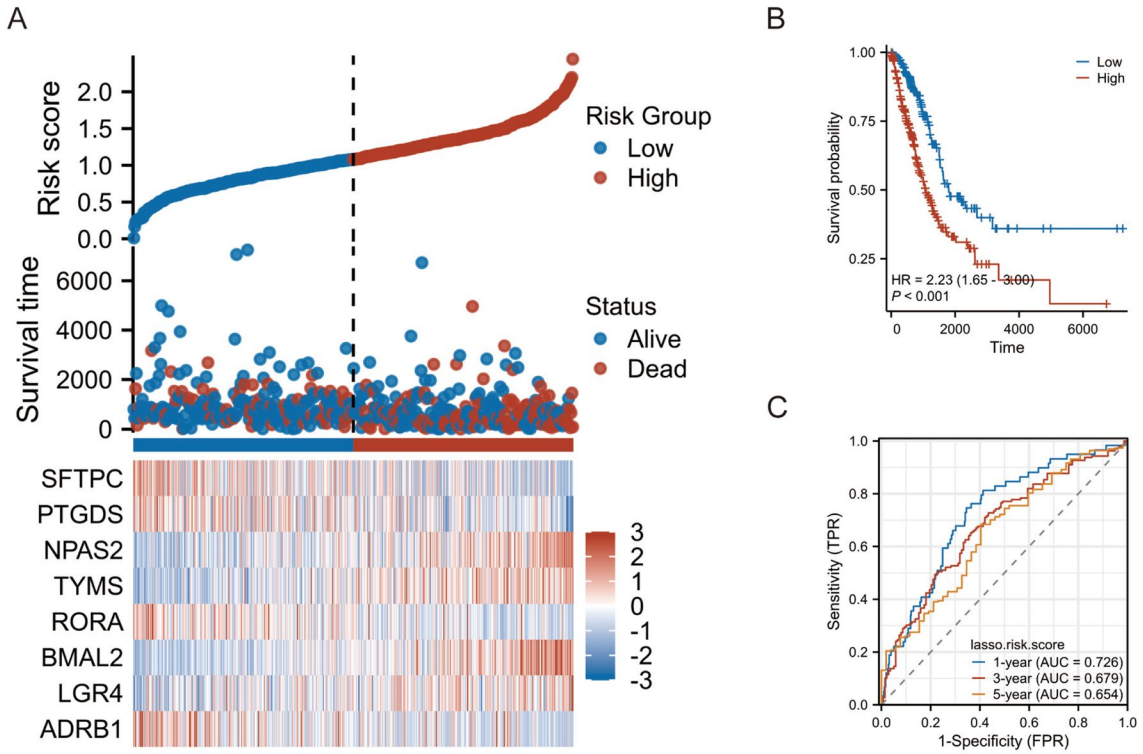

Figure 5. The risk score analysis, prognostic performance and survival analysis of the risk scoring model based on the differential expression of the 8 circadian clock-related genes in TCGA-LUAD patients. (**A**) The risk score, survival time distributions and gene expression heat map of circadian clock-related genes in the TCGA-LUAD cohort. (**B**) The ROC curves of the risk scoring model predicting OS of 1-year, 3-year, and 5-year in the TCGA-LUAD cohort. (**C**) Kaplan–Meier survival analysis of the OS between the risk groups in the TCGA-LUAD cohort.

GSEA

To assess the potential influence of the genes on incidence and progression of LUAD, GSEA was performed. And results demonstrated that DEGs between high-risk group and low-risk group were predominantly enriched in various pathways. These pathways include CD22 mediated BCR regulation, FcγR activation, the role of phospholipids in phagocytosis, the role of LAT2/NTAL/LAB in calcium mobilization, antigen activation of BCR leading to the generation of second messengers, FcεRI mediated Ca2 + mobilization, FcεRI mediated MAPK activation, FCGR3A mediated IL10 synthesis, assembly of the ORC at the origin of replication, and HDACs deacetylate histones (Fig. 8). The results indicate that genes associated with circadian rhythm could potentially influence the regulation of the immune system, lipid metabolism, calcium mobilization, cell signaling pathways, and cell cycle in LUAD.

Immune cell infiltration level analysis

Correlation analysis of the circadian clock-related genes and the 28 different immune cell types (immunome) showed that *TYMS* had a positive correlation with 27 of 28 the different immune cell types ($P < 0.01$), and no correlation with Eosinophil. *ADRB1* had a negative correlation with all 28 different immune cell types, with significance of $P < 0.01$ for 24 pathways, and 2 pathways with significant $P < 0.05$, and the remaining 2 pathways were not significant. *BMAL2* had a positive correlation with Type 2 T helper cell and neutrophil, and a negative correlation with the other 26 of 28 pathways. *NPAS2* had no correlation with the 28 immune cell type. There was no clear trend or significance between *LGR4*, *RORA*, *PTGDS* and *SFTPC* and these immune pathways (Fig. 9A).

The association between the forecasting model and the infiltration of immune cells were also computed. There is a reduce in immune cell infiltration among individuals classified as high-risk, such as activated B cell ($P < 0.001$), eosinophil ($P < 0.001$), immature B cell ($P < 0.001$), immature dendritic cell ($P < 0.001$), and mast cell ($P < 0.001$). Conversely, the high-risk group displayed elevated levels of activated CD4 T cell ($P < 0.001$), central memory CD8 T cell ($P < 0.01$), gamma delta T cell ($P < 0.01$), natural killer T cell ($P < 0.01$), neutrophil ($P < 0.01$), and type 2 T helper cell ($P < 0.001$) (Fig. 9B). Furthermore, there was a negative correlation observed between the risk score and the infiltration of immune cells, including activated B cell, eosinophil, immature B cell, immature dendritic cell, and mast cell (Fig. 9C-M).

Finally, in order to investigate the potential therapeutic drugs, we conducted drug prediction using the DGIdb database (https://dgidb.org/). A total of 137 drugs that target 5 specific genes were identified. To establish a prognostic genes-drug network, we utilized 5 prognostic genes (*ADRB1*, *LGR4*, *PTGDS*, *RORA*, *TYMS*) and 137

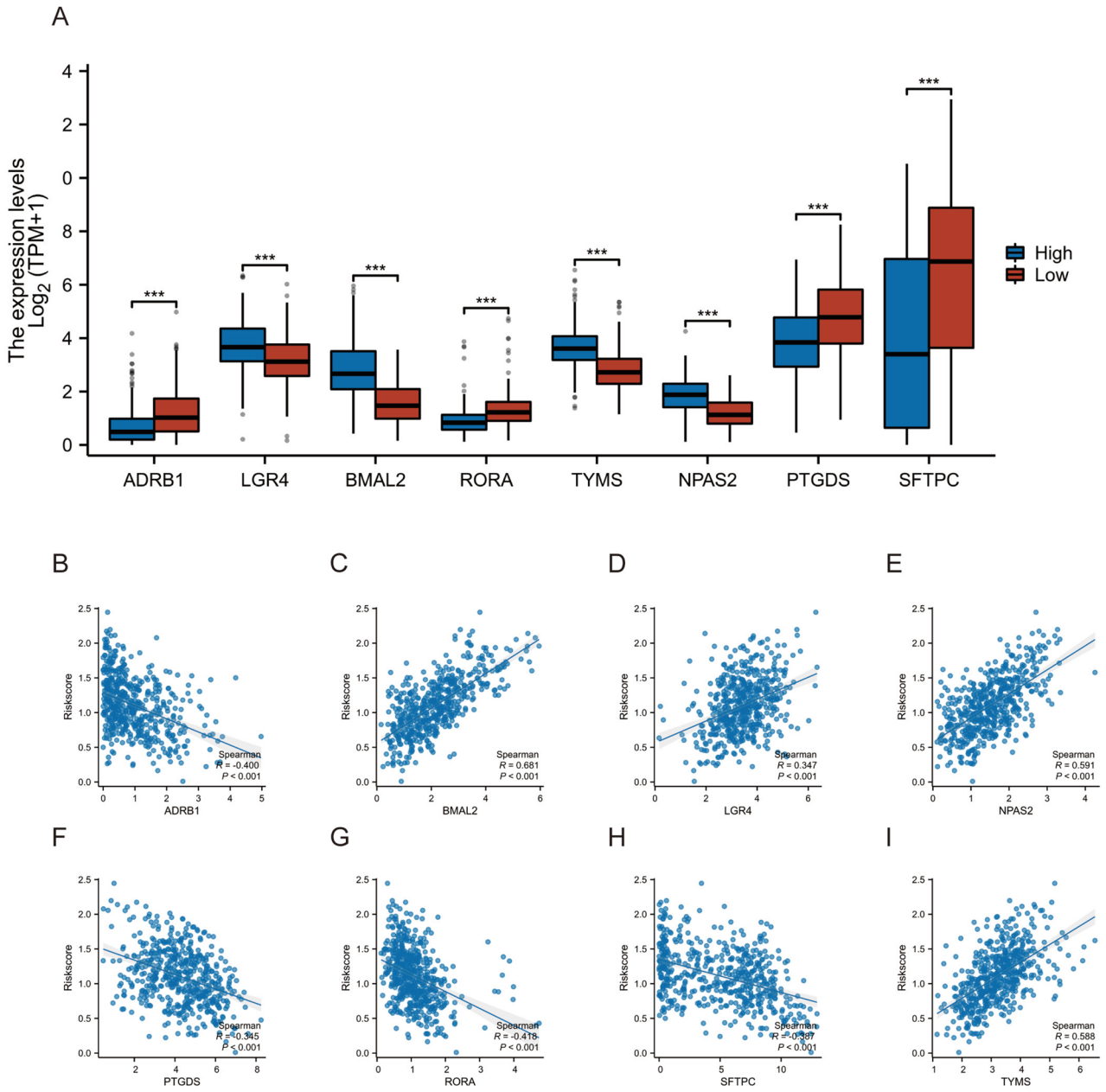

Figure 6. Correlation between the risk score and the circadian clock-related genes in the TCGA-LUAD data set. (**A**) The box plot showed the expression levels of the circadian clock-related genes between the high-risk group and low-risk group in LUAD patients. Scatter plots of correlation between the expression and risk score (**B**) *ADRB1*; (**C**) *BMAL2*; (**D**) *LGR4*; (**E**) *NPAS2*; (**F**) *PTGDS*; (**G**) *RORA*; (**H**) *SFTPC*; (**I**) *TYMS*). $^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$.

drugs (including ROSMANTUZUMAB, AMINOCAMPTOTHECIN, and INTERFERON BETA-1A) (Fig. 9N). This network consisted of 142 nodes and 137 edges.

Discussion

Lung adenocarcinoma, which poses a significant global health challenge, accounting for a substantial share of cancer-related morbidity and mortality, is the most prevalent subtype of NSCLC, and has a highly heterogeneous morphological characteristic and is remarkably variable in prognosis[7,10,11,49,50]. In recent years, due to the rapid advancements in high-throughput technologies and bioinformatics methods, increasing attention has been focused on the crucial role of gene markers based on specific correlations in predicting the prognosis of LUAD[51–53].

There is clear evidence of differential expression of CCRGs in various diseases, including cancer[54]. Besides, in recent years, the World Health Organization (WHO) has identified "night shift work involving circadian disruption" as "probably carcinogenic to humans" (Group 2A) based on population and laboratory research

Characteristics	Total(N)	Univariate analysis		Multivariate analysis	
		HR (95% CI)	*P* value	HR (95% CI)	*P* value
ADRB1	530		0.089		
Low	264	Reference		Reference	
High	266	0.779 (0.584–1.040)	0.090	0.984 (0.715–1.354)	0.921
LGR4	530		0.003		
Low	265	Reference		Reference	
High	265	1.540 (1.154–2.055)	0.003	1.351 (1.005–1.818)	0.047
BMAL2	530		0.001		
Low	265	Reference		Reference	
High	265	1.616 (1.206–2.165)	0.001	1.391 (1.008–1.919)	0.044
RORA	530		0.004		
Low	263	Reference		Reference	
High	267	0.647 (0.481–0.871)	0.004	0.700 (0.514–0.953)	0.024
TYMS	530		0.003		
Low	264	Reference		Reference	
High	266	1.547 (1.158–2.067)	0.003	1.278 (0.937–1.744)	0.121
NPAS2	530		< 0.001		
Low	266	Reference		Reference	
High	264	1.675 (1.255–2.236)	< 0.001	1.530 (1.137–2.057)	0.005
PTGDS	530		0.171		
Low	264	Reference		Reference	
High	266	0.819 (0.615–1.091)	0.172	0.983 (0.727–1.331)	0.913
SFTPC	530		0.042		
Low	265	Reference		Reference	
High	265	0.742 (0.556–0.990)	0.042	0.898 (0.657–1.227)	0.498

Table 2. Univariate and multivariate Cox regression analysis between the genes' expression level and OS in the TCGA-LUAD cohort. *HR* Hazard ratio, 95%*CI* 95% confidence interval.

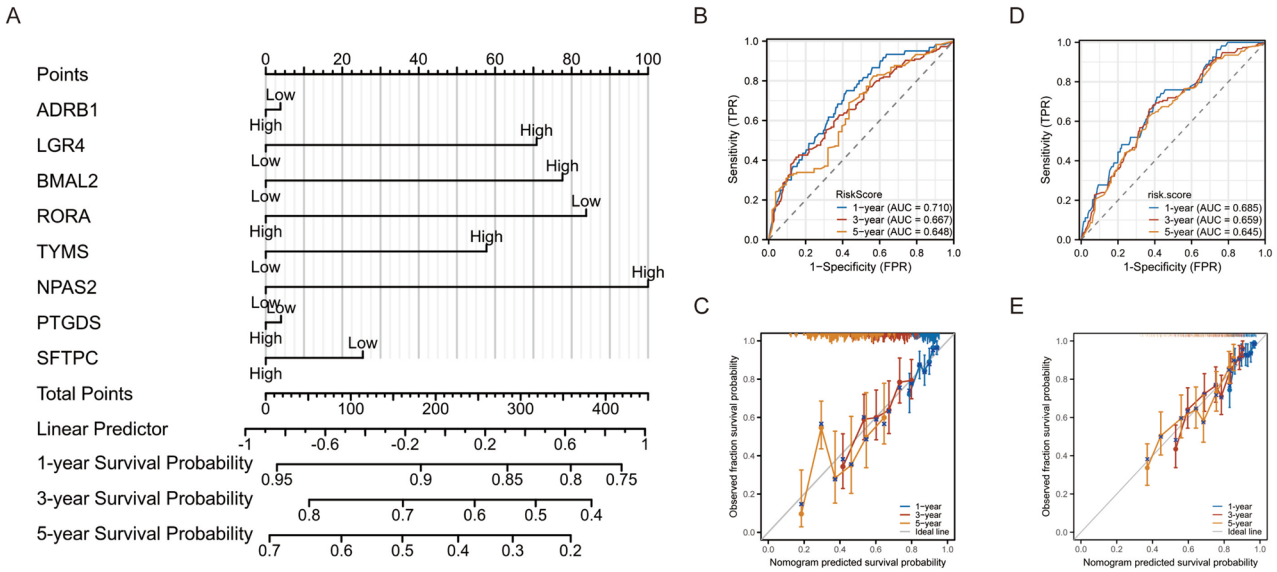

Figure 7. Prognostic nomogram for the 1-year, 3-year, and 5-year OS of LUAD patients. (**A**) The independent risk factors that affect the OS of LUAD patients screened by multiple Cox regression were incorporated into the nomogram model. (**B**) The ROC curves for predicting the nomogram of 1-year, 3-year, and 5-year OS in the TCGA-LUAD cohort. (**C**) The nomogram calibration curves for predicting the 1-year, 3-year, and 5-year OS of the TCGA-LUAD cohort. (**D**) The ROC curves in the external verification group. (**E**) The nomogram calibration curves of the external verification group.

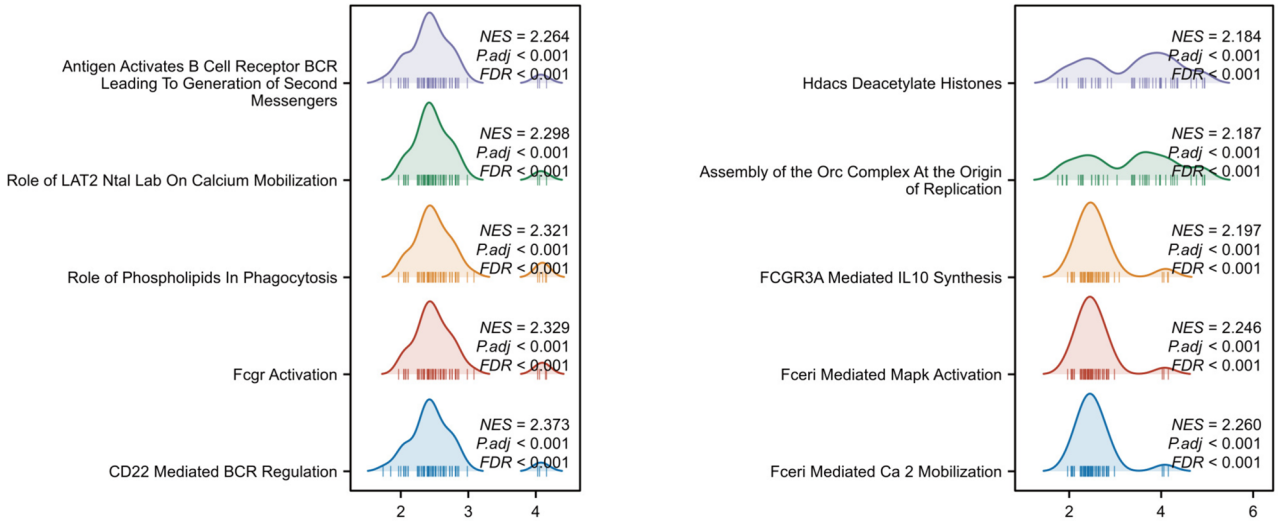

Figure 8. GSEA analysis of the DEGs between the high-risk group and the low-risk group in the TCGA-LUAD cohort. Top 10 terms of GSEA analysis (CD22 mediated BCR regulation, FcγR activation, role of phospholipids in phagocytosis, role of LAT2/NTAL/LAB on calcium mobilization, antigen activates BCR leading to generation of second messengers, FcεRI mediated Ca2+ mobilization, FcεRI mediated MAPK activation, FCGR3A mediated IL10 synthesis, assembly of the ORC at the origin of rep-lication and HDACs deacetylate histones).

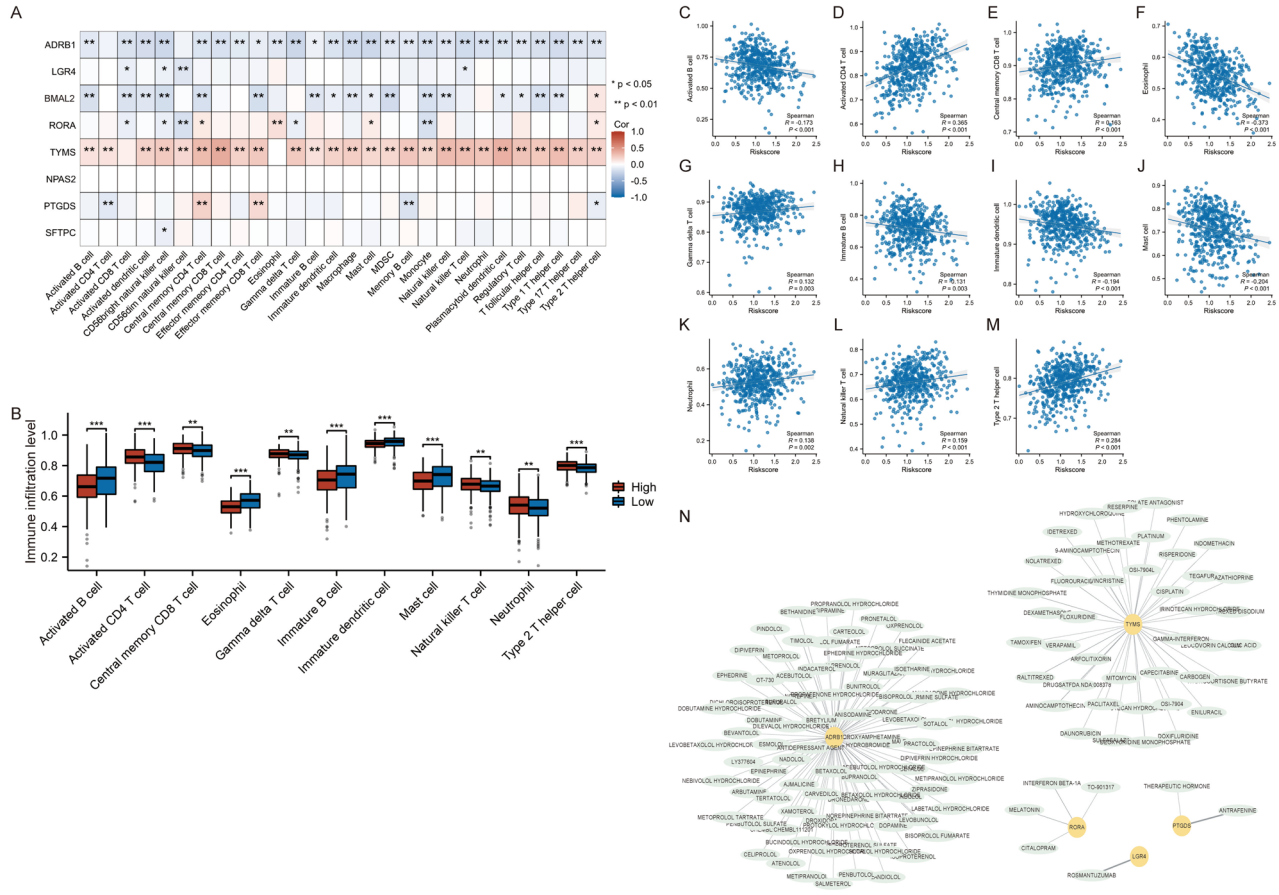

Figure 9. Analysis of immune cell infiltration in TCGA-LUAD cohort. (**A**) The heatmap showed the correlation between the circadian clock-related genes and the 28 different immune cell types. (**B**) The box plot showed the levels of immune cell infiltration between the high-risk group and low-risk group in LUAD patients. Scatter plots of correlation between immune cell infiltrations and risk score (**C**) Activated B cell; (**D**) Activated CD4 T cell; (**E**) Central memory CD8 T cell; (**F**) Eosinophil; (**G**) Gamma delta T cell; (**H**) Immature B cell, (**I**) Immature dendritic cell; (**J**) Mast cell; (**K**) Neutrophil; (**L**) Natural killer T cell; (**M**) Type 2 T helper cell). (**N**) The network of drugs and prognostic genes. $^{*}P<0.05$, $^{**}P<0.01$, $^{***}P<0.001$.

findings[55]. Existing research results indicate that CCRGs play a significant role in the regulation of epigenetic control mechanisms during tumor initiation and progression. Disruption of circadian rhythms can promote the expression of cancer biomarkers, increase cellular heterogeneity, affect gene transcription and modification, and accelerate tumor growth[24–26]. This has been reported in various studies on different types of tumors, highlighting the increasing importance of circadian rhythm biology in enhancing our understanding of molecular mechanisms in cancer cells[13–15,56].

By analyzing, we identified 5382 DEGs. Comparing these differentially expressed genes with CCRGs, we obtained a total of 76 DEGs associated with circadian rhythm. Among them, 16 genes which significantly correlated with overall survival were identified as potential prognostic markers. After performing LASSO regression and multivariate Cox regression analysis, we identified 8 potential prognostic markers (*ADRB1*, *BMAL2*, *LGR4*, *NPAS2*, *PTGDS*, *RORA*, *SFTPC* and *TYMS*). According to the results from RhythmicDB, *PTGDS*, *ADRB1*, and *TYMS* have been identified as oscillating genes in Homo sapiens. Additionally, the other genes have been confirmed as oscillating genes in different species. This warrants further research to explore the specific effects of these genes. We constructed a prognostic model using these 8 CCRGs. Among them, the expression levels of 4 genes (*BMAL2*, *LGR4*, *NPAS2*, and *TYMS*) were negatively correlated with OS, while the expression levels of 4 genes (*ADRB1*, *PTGDS*, *RORA*, and *SFTPC*) were positively correlated with OS (Figure S2).

ADRB1, also recognized as β-1 adrenergic receptor (AR), is part of the G-protein coupled receptor family. Anna-Maria Globig and her colleagues have pointed out that in situations where there is long-term exposure to antigens, the signaling of adrenergic receptors on T cells through *ADRB1* is crucial in the final differentiation of T cells into an exhausted state[57].

BMAL2 plays a crucial role in the regulation of biological rhythms and exerts a substantial influence on the clinical characteristics and prognosis of individuals diagnosed with lung adenocarcinoma. Researchers have identified a significant correlation between elevated *BMAL2* expression and lymph node metastasis in individuals diagnosed with lung adenocarcinoma. Moreover, the expression level of *BMAL2* in lung adenocarcinoma is associated with tumor immune infiltration and immune checkpoint activity[58,59].

LGR4 is a transmembrane receptor member of the GPCRs superfamily[60] and is highly expressed in lung adenocarcinoma. Yang and colleagues demonstrated that miR-449b targets *LGR4*, with decreased expression levels observed in LUAD in comparison to normal tissues. The upregulation of miR-39b results in decreased proliferation and invasion abilities of lung cancer cell lines through the downregulation of *LGR4* expression[61], thus, highlighting the role of *LGR4* in migration and invasion processes.

The absence of normal *NPAS2* may lead to several defects in the aspects of the circadian rhythm system, such as sleep and behavioral patterns[62]. Previous reports have indicated that *NPAS2* is involved in tumorigenesis by regulating PER2, which can act as a tumor suppressor[63], and by inhibiting transcription of the oncogene c-Myc[64] and there is a close correlation between genetic variations in *NPAS2* and clinical outcomes in NSCLC patients[65].

PTGDS is the key enzyme responsible for regulating the synthesis of PGD2. Currently, research indicates that *PTGDS* is downregulated in lung adenocarcinoma and is associated with the prognosis[66]. The invasive ability of tumor cells is closely related to *PTGDS*[67,68].

RORα belongs to the superfamily of orphan nuclear receptors (NRs). The fundamental capacity of RORα to control the expression of target genes plays a crucial role in the activation of tumor-suppressive genes[69].

Surfactant protein C (SFTP-C) is the key element of pulmonary surfactant that helps to prevent the collapse of alveoli[70]. A previous research report observed the absence of *SFTPC* in NSCLC tissues[71], and Bin Li et al. found that *SFTPC* was downregulated in lung adenocarcinoma tissues and cell lines (H1650, H441, and A549), and low expression of *SFTPC* was associated with poor prognosis[72].

TYMS, a rate-limiting enzyme in the process of DNA synthesis[73]. High expression levels of *TYMS* have been associated with adverse reactions to 5-FU, shorter survival time, and other unfavorable clinical outcomes in various solid tumors[74,75].

The correlations identified in this study between various circadian rhythm genes and clinical indicators suggest that these genes might have a multifaceted impact on the initiation and advancement of tumors through multiple pathways[76,77]. *BMAL2* is highly expressed in individuals with higher T or N stages, consistent with previous findings that elevated *BMAL2* levels affect lymph node metastasis[59]. In contrast, *RORA* and *PTGDS* exhibit an opposite trend to *BMAL2* across different T and N stages. According to previous research, high *PTGDS* expression reduces tumor metastasis, which aligns with our findings[66,67]. It is well known that smoking is a significant risk factor for lung cancer. In our subgroup analysis of smokers, we found that *RORA* expression is elevated in individuals with no smoking history and those who smoke fewer than 40 packs per year. This suggests that tobacco may influence tumor suppression through the *RORA* pathway, leading to the development of lung cancer[69]. Besides, the observed differences in circadian rhythm gene expression across biological sexes, ages, ethnicities, and tumor locations were unexpected, underscoring the critical role these genes play in biological growth and development. These results indicate that these genes possess commendable sensitivity and specificity.

The GSEA indicated that these CCRGs could play a potential part in CD22 mediated BCR regulation, FcγR activation, role of phospholipids in phagocytosis, role of LAT2/NTAL/LAB on calcium mobilization, antigen activates BCR leading to generation of second messengers, FcεRI mediated Ca2+ mobilization, FcεRI mediated MAPK activation, FCGR3A mediated IL10 synthesis, assembly of the ORC at the origin of replication and HDACs deacetylate histones. Tuscano et al. determined through experiments that CD22 antigen is broadly expressed on lung cancer cells and suggested its potential as a therapeutic target for lung cancer[78]. Multiple studies have indicated that the utilization of FcγR binding may potentially enhance anti-tumor activity, offering a promising strategy for the treatment of lung cancer[79]. Yingfang Lu et al. discovered that the FcεRI signaling pathway may be closely associated with the inhibition of lung cancer and the regulation of gut microbiota balance[80]. Geetha Shanmugam et al. further elucidated the significance of HDAC inhibitors in the treatment of lung cancer, highlighting the potential of HDAC as a targeted therapeutic approach[81]. The above-mentioned

study has already demonstrated the dependability of our result; however, extra research is needed to explore the mechanisms of different pathways.

Our analysis demonstrated the engagement of these genes in diverse biological activities, encompassing cell cycle management, metabolic processes, immune system regulation, inflammatory reactions, cytoskeleton restructuring, chromatin modification, DNA damage-induced apoptosis, and protein synthesis and trafficking. Therefore, these processes in tumor cells may be influenced by circadian rhythms. Some studies have suggested that significant epigenetic changes in circadian clock-related genes promote the occurrence and progression of lung cancer, leading to decreased survival rates[77,82].

Examination of immune cell infiltration reveals that an appreciable surge in the levels of infiltration was noted for key cells, such as activated CD4 T cells and central memory CD8 T cells, while immature immune cell infiltration decreases, in specific anti-tumor immune processes in high-risk patients. The current viewpoint suggests that the CCRGs are widely expressed in immune cells and exhibit a fixed circadian rhythm[83]. Differential expressions of the CCRGs gene are important in the development and specificity of immune cell lineages. Evidence suggests that disruptions in circadian rhythm caused by differential gene expression in cancer cells may lead to immune dysregulation, which could be attributed to mutations in clock genes, environmental disturbances, or age and the tumor itself. Disruption of normal circadian rhythm may result in differential expression of CCRGs and metabolic rhythm, potentially contributing to both host immune support and increased susceptibility to tissue damage and catastrophic vulnerability[84].

Moreover, the drug prediction outcomes indicated that various small-molecule drugs specifically target *ADRB1*, *LGR4*, *PTGDS*, *RORA*, *TYMS*. Among them, ROSMANTUZUMAB targeted *LGR4*. ROSMANTUZUMAB, a humanized monoclonal antibody designed for the treatment of cancer, has been shown to modulate the Wnt pathway for tumor cell survival[85]. ROSMANTUZUMAB demonstrated promising antitumor activity in some preclinical models of colon cancer[86]. We hypothesize ROSMANTUZUMAB could exert its effects by targeting *LGR4*. AMINOCAMPTOTHECIN has been found to exhibit inhibitory biological activity against repair of single-strand DNA breakages[87]. Thus, AMINOCAMPTOTHECIN might exert inhibitory effects on LUAD by targeting *TYMS*. Numerous other pharmaceuticals that target prognostic genes have also shown potential in treating cancer. Moving ahead, our group aim to validate the effectiveness of these drugs through clinical data and investigate their mechanisms of action. All of the findings imply this work could pave the way for new approaches to treating LUAD.

Conclusions

In summary, we have innovatively constructed and rigorously a novel risk stratification system founded upon circadian clock-associated genes, meticulously derived from the TCGA dataset. This pioneering approach facilitates the precise prognostic assessment of patients afflicted with LUAD. Furthermore, we built a nomogram model, meticulously designed for the prediction of OS, showcasing remarkable predictive precision. And through our meticulous investigation, we have successfully unearthed eight differentially expressed genes closely linked to the circadian rhythm in the context of LUAD. These groundbreaking findings are poised to serve as the bedrock for the future advancement of precision oncology. They open innovative avenues for tailoring personalized treatment strategies and, in turn, promise to elevate the overall standard of care and management for individuals grappling with lung adenocarcinoma.

Data availability

The data sets analyzed during the current study are all available in public databases. The gene expression data and clinical data of 535 LUAD samples and 59 control samples were obtained from the TCGA database (https://portal.gdc.cancer.gov/). The immune infiltration-related gene expression validation data sets GSE31210 and GSE68465 were obtained from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). The immune infiltration-related gene data were downloaded from MSigDB (https://www.gsea-msigdb.org/gsea/). The immunohistochemical data of immune infiltration-related genes in HCC and normal liver tissues were obtained from the Human Protein Atlas (HPA) database (https://www.proteinatlas.org/).

Received: 27 March 2024; Accepted: 22 July 2024
Published online: 06 August 2024

References

1. Herbst, R. S., Morgensztern, D. & Boshoff, C. The biology and management of non-small cell lung cancer. *Nature* **553**, 446–454. https://doi.org/10.1038/nature25183 (2018).
2. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **71**, 209–249. https://doi.org/10.3322/caac.21660 (2021).
3. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics 2022. *CA Cancer J. Clin.* **72**, 7–33. https://doi.org/10.3322/caac.21708 (2022).
4. Zhang, J. *et al.* T cell-related prognostic risk model and tumor immune environment modulation in lung adenocarcinoma based on single-cell and bulk RNA sequencing. *Comput. Biol. Med.* **152**, 106460. https://doi.org/10.1016/j.compbiomed.2022.106460 (2023).
5. Torre, L. A., Siegel, R. L. & Jemal, A. in *Lung Cancer and Personalized Medicine: Current Knowledge and Therapies* (eds Aamir Ahmad & Shirish Gadgeel) 1–19 (Springer International Publishing, 2016).
6. Chen, W. *et al.* Cancer statistics in China. *CA Cancer J. Clin.* **66**, 115–132. https://doi.org/10.3322/caac.21338 (2016).
7. Hirsch, F. R. *et al.* Lung cancer: Current therapies and new targeted treatments. *The Lancet* **389**, 299–311. https://doi.org/10.1016/S0140-6736(16)30958-8 (2017).
8. Saito, M. *et al.* Gene aberrations for precision medicine against lung adenocarcinoma. *Cancer Sci.* **107**, 713–720. https://doi.org/10.1111/cas.12941 (2016).

9. Gavin, S. J. & David, R. B. Recent advances in the management of lung cancer. *Clin. Med.* **18**, s41. https://doi.org/10.7861/clinmedicine.18-2-s41 (2018).
10. Yang, L. *et al.* Identification and validation of a novel six-lncRNA-based prognostic model for lung adenocarcinoma. *Front. Oncol.* **11**, 775583. https://doi.org/10.3389/fonc.2021.775583 (2022).
11. Jiang, X. *et al.* Systematic analysis and validation of the prognosis, immunological role and biology function of the ferroptosis-related lncRNA GSEC/miRNA-101–3p/CISD1 axis in lung adenocarcinoma. *Front. Mol. Biosci.* **8**, 793732. https://doi.org/10.3389/fmolb.2021.793732 (2022).
12. Koronowski, K. B. & Sassone-Corsi, P. Communicating clocks shape circadian homeostasis. *Science* **371**, eabd0951. https://doi.org/10.1126/science.abd0951 (2021).
13. Sancar, A. *et al.* Circadian clock, cancer, and chemotherapy. *Biochemistry* **54**, 110–123. https://doi.org/10.1021/bi5007354 (2015).
14. Sancar, A. Mechanisms of DNA repair by photolyase and excision nuclease (nobel lecture). *Angew Chem. Int. Ed. Engl.* **55**, 8502–8527. https://doi.org/10.1002/anie.201601524 (2016).
15. Lunn, R. M. *et al.* Health consequences of electric lighting practices in the modern world : A report on the National Toxicology Program's workshop on shift work at night, artificial light at night, and circadian disruption. *Sci. Total Environ.* **607–608**, 1073–1084. https://doi.org/10.1016/j.scitotenv.2017.07.056 (2017).
16. Stevens, R. G. *et al.* Considerations of circadian impact for defining "shift work" in cancer studies: IARC working group report. *Occup. Environ. Med.* **68**, 154–162. https://doi.org/10.1136/oem.2009.053512 (2009).
17. Straif, K. *et al.* Carcinogenicity of shift-work, painting, and fire-fighting. *Lancet Oncol.* **8**, 1065–1066. https://doi.org/10.1016/S1470-2045(07)70373-X (2007).
18. Sulli, G., Lam, M. T. Y. & Panda, S. Interplay between circadian clock and cancer: New frontiers for cancer treatment. *Trends Cancer* **5**, 475–494. https://doi.org/10.1016/j.trecan.2019.07.002 (2019).
19. Xuan, W. *et al.* Circadian regulation of cancer cell and tumor microenvironment crossta lk. *Trends Cell Biol.* **31**, 940–950. https://doi.org/10.1016/j.tcb.2021.06.008 (2021).
20. Li, M. *et al.* Circadian rhythm-associated clinical relevance and tumor microenvironment of non-small cell lung cancer. *J. Cancer* **12**, 2582–2597. https://doi.org/10.7150/jca.52454 (2021).
21. Aiello, I. *et al.* Circadian disruption promotes tumor-immune microenvironment remodeling favoring tumor cell proliferation. *Sci. Adv.* **6**, eaaz4530. https://doi.org/10.1126/sciadv.aaz4530 (2020).
22. Shafi, A. A. & Knudsen, K. E. Cancer and the circadian clock. *Cancer Res.* **79**, 3806–3814. https://doi.org/10.1158/0008-5472.CAN-19-0566 (2019).
23. Lesicka, M., Nedoszytko, B. & Reszka, E. Disruptions of circadian genes in cutaneous melanoma—An in silico analysis of transcriptome databases. *Int. J. Mol. Sci.* **24**, 10140 (2023).
24. Janich, P. *et al.* The circadian molecular clock creates epidermal stem cell heterogeneity. *Nature* **480**, 209–214. https://doi.org/10.1038/nature10649 (2011).
25. Malhan, D., Basti, A. & Relógio, A. Transcriptome analysis of clock disrupted cancer cells reveals differential alternative splicing of cancer hallmarks genes. *NPJ Syst. Biol. Appl.* **8**, 17. https://doi.org/10.1038/s41540-022-00225-w (2022).
26. Schwartz, P. B. *et al.* The circadian clock is disrupted in pancreatic cancer. *PLOS Genet.* **19**, e1010770. https://doi.org/10.1371/journal.pgen.1010770 (2023).
27. Mootha, V. K. *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273. https://doi.org/10.1038/ng1180 (2003).
28. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Nat. Acad. Sci.* **102**, 15545–15550. https://doi.org/10.1073/pnas.0506580102 (2005).
29. Okayama, H. *et al.* Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res.* **72**, 100–111. https://doi.org/10.1158/0008-5472.CAN-11-1403 (2012).
30. Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma. *et al.* Gene expression-based survival prediction in lung adenocarcinoma: A multi-site, blinded validation study. *Nat. Med.* **14**, 822–827. https://doi.org/10.1038/nm.1790 (2008).
31. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. https://doi.org/10.1186/s13059-014-0550-8 (2014).
32. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849. https://doi.org/10.1093/bioinformatics/btw313 (2016).
33. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. Clusterprofiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287. https://doi.org/10.1089/omi.2011.0118 (2012).
34. Walter, W., Sánchez-Cabo, F. & Ricote, M. GOplot: An R package for visually combining expression data with functional analysis. *Bioinformatics* **31**, 2912–2914. https://doi.org/10.1093/bioinformatics/btv300 (2015).
35. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* https://doi.org/10.1016/j.xinn.2021.100141 (2021).
36. Friedman, J. H., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22. https://doi.org/10.18637/jss.v033.i01 (2010).
37. Therneau, T.M., Grambsch, P.M. *Modeling survival data: Extending the {C}ox Model.* (Springer, 2000).
38. Warde-Farley, D. *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **38**, W214-220. https://doi.org/10.1093/nar/gkq537 (2010).
39. Xu, S., Wang, Z., Ye, J., Mei, S. & Zhang, J. Identification of iron metabolism-related genes as prognostic indicators for lower-grade glioma. *Front. Oncol.* **11**, 729103. https://doi.org/10.3389/fonc.2021.729103 (2021).
40. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297. https://doi.org/10.1093/nar/gks042 (2012).
41. Blanche, P., Dartigues, J. F. & Jacqmin-Gadda, H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat. Med.* **32**, 5381–5397 (2013).
42. Fagotti, A. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).
43. Biecek, A. K. a. M. K. a. P. https://cran.r-project.org/web/packages/survminer/index.html (2021).
44. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **14**, 7. https://doi.org/10.1186/1471-2105-14-7 (2013).
45. Charoentong, P. *et al.* Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* **18**, 248–262. https://doi.org/10.1016/j.celrep.2016.12.019 (2017).
46. Ru, B. *et al.* TISIDB: An integrated repository portal for tumor-immune system intera ctions. *Bioinformatics* **35**, 4200–4202. https://doi.org/10.1093/bioinformatics/btz210 (2019).
47. Castellana, S. *et al.* RhythmicDB: A database of predicted multi-frequency rhythmic transcripts. *Front. Genet.* https://doi.org/10.3389/fgene.2022.882044 (2022).
48. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419. https://doi.org/10.1126/science.1260419 (2015).
49. Matsuda, T. & Machii, R. Morphological distribution of lung cancer from cancer incidence in five continents Vol X. *Jpn. J. Clin. Oncol.* **45**, 404. https://doi.org/10.1093/jjco/hyv041 (2015).

50. The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550. https://doi.org/10.1038/nature13385 (2014).
51. Wang, Z. *et al.* Establishment and validation of a prognostic signature for lung adenocarcinoma based on metabolism-related genes. *Cancer Cell Int.* **21**, 219. https://doi.org/10.1186/s12935-021-01915-x (2021).
52. Wang, W., Ren, S., Wang, Z., Zhang, C. & Huang, J. Increased expression of TTC21A in lung adenocarcinoma infers favorable prognosis and high immune infiltrating level. *Int. Immunopharmacol.* **78**, 106077. https://doi.org/10.1016/j.intimp.2019.106077 (2020).
53. Yu, L. *et al.* Prognostic value and immune infiltration of a novel stromal/immune score-related P2RY12 in lung adenocarcinoma microenvironment. *Int. Immunopharmacol.* **98**, 107734. https://doi.org/10.1016/j.intimp.2021.107734 (2021).
54. Kelleher, F. C., Rao, A. & Maguire, A. Circadian molecular clocks and cancer. *Cancer Lett.* **342**, 9–18. https://doi.org/10.1016/j.canlet.2013.09.040 (2014).
55. Ward, E. M. *et al.* Carcinogenicity of night shift work. *Lancet Oncol.* **20**, 1058–1059. https://doi.org/10.1016/S1470-2045(19)30455-3 (2019).
56. Masri, S. & Sassone-Corsi, P. The emerging link between cancer, metabolism, and circadian rhythms. *Nat. Med.* **24**, 1795–1803. https://doi.org/10.1038/s41591-018-0271-8 (2018).
57. Globig, A.-M. *et al.* The β1-adrenergic receptor links sympathetic nerves to T cell exhaustion. *Nature* https://doi.org/10.1038/s41586-023-06568-6 (2023).
58. Mazzoccoli, G. *et al.* ARNTL2 and SERPINE1: Potential biomarkers for tumor aggressiveness in colorectal cancer. *J. Cancer Res. Clin. Oncol.* **138**, 501–511. https://doi.org/10.1007/s00432-011-1126-6 (2011).
59. Zhang, H. *et al.* ARNTL2 is an indicator of poor prognosis, promotes epithelial-to-mesen chymal transition and inhibits ferroptosis in lung adenocarcinoma. *Transl. Oncol.* **26**, 101562. https://doi.org/10.1016/j.tranon.2022.101562 (2022).
60. Van Loy, T. *et al.* Comparative genomics of leucine-rich repeats containing G protein-coup led receptors and their ligands. *Gen. Comp. Endocrinol.* **155**, 14–21. https://doi.org/10.1016/j.ygcen.2007.06.022 (2008).
61. Yang, D., Li, J.-S., Xu, Q.-Y., Xia, T. & Xia, J.-H. Inhibitory effect of MiR-449b on cancer cell growth and invasion through LGR4 in non-small-cell lung carcinoma. *Curr. Med. Sci.* **38**, 582–589. https://doi.org/10.1007/s11596-018-1917-y (2018).
62. Dudley, C. A. *et al.* Altered patterns of sleep and behavioral adaptability in *NPAS2*-deficient mice. *Science* **301**, 379–383. https://doi.org/10.1126/science.1082795 (2003).
63. Fu, L. & Lee, C. C. The circadian clock: Pacemaker and tumour suppressor. *Nat. Rev. Cancer* **3**, 350–361. https://doi.org/10.1038/nrc1072 (2003).
64. Fu, L., Pelicano, H., Liu, J., Huang, P. & Lee, C. The circadian gene Period2 plays an important role in tumor suppression and DNA damage response in vivo. *Cell* **111**, 41–50. https://doi.org/10.1016/s0092-8674(02)00961-3 (2002).
65. He, Y. *et al.* Genetic variants in NPAS2 gene and clinical outcomes of resectable non-small-cell lung cancer. *Future Oncol.* **17**, 795–805. https://doi.org/10.2217/fon-2020-0211 (2021).
66. He, L. P., Chen, Y. F. & Yang, J. Investigation on the role and mechanism of prostagland in D2 synthase in non-small cell lung cancer. *Zhonghua Yi Xue Za Zhi* **97**, 3022–3027. https://doi.org/10.3760/cma.j.issn.0376-2491.2017.38.016 (2017).
67. Hsia, T.-C. *et al.* Cantharidin impairs cell migration and invasion of human lung cancer NCI-H460 cells via UPA and MAPK signaling pathways. *Anticancer Res.* **36**, 5989–5997 (2016).
68. Cao, Q., Mao, Z. D., Shi, Y. J., Chen, Y. & Peng, L. P. J. O. MicroRNA-7 inhibits cell proliferation, migration and invasion in human non-small cell lung cancer cells by targeting FAK through ERK/MAPK signaling pathway. *Oncotarget* **7**, 77468 (2016).
69. Lee, J. M., Kim, H. & Baek, S. H. Unraveling the physiological roles of retinoic acid receptor-related orphan receptor α. *Exp. Mol. Med.* **53**, 1278–1286. https://doi.org/10.1038/s12276-021-00679-8 (2021).
70. Whitsett, J. A. & Weaver, T. E. Hydrophobic surfactant proteins in lung function and disease. *N Engl. J. Med.* **347**, 2141–2148. https://doi.org/10.1056/NEJMra022387 (2002).
71. Li, R. *et al.* Genetic deletions in sputum as diagnostic markers for early detection of stage I non-small cell lung cancer. *Clin. Cancer Res.* **13**, 482–487. https://doi.org/10.1158/1078-0432.CCR-06-1593 (2007).
72. Li, B. *et al.* MiR-629–3p-induced downregulation of SFTPC promotes cell proliferation and predicts poor survival in lung adenocarcinoma. *Artif. Cells Nanomed. Biotechnol.* **47**, 3286–3296. https://doi.org/10.1080/21691401.2019.1648283 (2019).
73. Gangjee, A. *et al.* Design, synthesis, and X-ray crystal structure of a potent dual inhibitor of thymidylate synthase and dihydrofolate reductase as an antitumor agent. *J. Med Chem.* **43**, 3837–3851. https://doi.org/10.1021/jm000200l (2000).
74. Formentini, A., Henne-Bruns, D. & Kornmann, M. Thymidylate synthase expression and prognosis of patients with gastrointestinal cancers receiving adjuvant chemotherapy: A review. *Langenbecks Arch. Surg.* **389**, 405–413. https://doi.org/10.1007/s00423-004-0510-y (2004).
75. Lee, S.-W. *et al.* Overexpression of thymidylate synthetase confers an independent prognostic indicator in nasopharyngeal carcinoma. *Exp. Mol. Pathol.* **95**, 83–90. https://doi.org/10.1016/j.yexmp.2013.05.006 (2013).
76. Zhang, H., Liu, R., Zhang, B., Huo, H. & Song, Z. Advances in the study of circadian genes in non-small cell lung cancer. *Integr. Cancer Therap.* **21**, 15347354221096080. https://doi.org/10.1177/15347354221096080 (2022).
77. Papagiannakopoulos, T. *et al.* Circadian rhythm disruption promotes lung tumorigenesis. *Cell Metab.* **24**, 324–331. https://doi.org/10.1016/j.cmet.2016.07.001 (2016).
78. Tuscano, J. M. *et al.* CD22 antigen is broadly expressed on lung cancer cells and is a target for antibody-based therapy. *Cancer Res.* **72**, 5556–5565. https://doi.org/10.1158/0008-5472.CAN-12-0173 (2012).
79. Lutterbuese, P. *et al.* Exchanging human Fcgamma1 with murine Fcgamma2a highly potentiates anti-tumor activity of anti-EpCAM antibody adecatumumab in a syngeneic mouse lung metastasis model. *Cancer Immunol. Immunother.* **56**, 459–468. https://doi.org/10.1007/s00262-006-0218-7 (2006).
80. Lu, Y. *et al.* Spirulina polysaccharide induces the metabolic shifts and gut microbiota change of lung cancer in mice. *Curr. Res. Food Sci.* **5**, 1313–1319. https://doi.org/10.1016/j.crfs.2022.08.010 (2022).
81. Shanmugam, G., Rakshit, S. & Sarkar, K. HDAC inhibitors: Targets for tumor therapy, immune modulation and lung diseases. *Transl. Oncol.* **16**, 101312. https://doi.org/10.1016/j.tranon.2021.101312 (2022).
82. Kettner, N. M., Katchy, C. A. & Fu, L. Circadian gene variants in cancer. *Ann. Med.* **46**, 208–220. https://doi.org/10.3109/07853890.2014.914808 (2014).
83. Wen, P. *et al.* Identifying hub circadian rhythm biomarkers and immune cell infiltration in rheumatoid arthritis. *Front. Immunol.* https://doi.org/10.3389/fimmu.2022.1004883 (2022).
84. Haspel, J. A. *et al.* Circadian rhythm reprogramming during lung inflammation. *Nat. Commun.* **5**, 4753. https://doi.org/10.1038/ncomms5753 (2014).
85. Fischer, M. M. *et al.* RSPO3 antagonism inhibits growth and tumorigenicity in colorectal tumors harboring common Wnt pathway mutations. *Sci. Rep.* **7**, 15270. https://doi.org/10.1038/s41598-017-15704-y (2017).
86. Storm, E. E. *et al.* Targeting PTPRK-RSPO3 colon tumours promotes differentiation and loss of stem-cell function. *Nature* **529**, 97–100. https://doi.org/10.1038/nature16466 (2016).
87. Takimoto, C. H. & Thomas, R. The clinical development of 9-aminocamptothecin. *Ann. New York Acad. Sci.* **922**, 224–236. https://doi.org/10.1111/j.1749-6632.2000.tb07041.x (2000).

Acknowledgements
We acknowledge TCGA, HPA and GEO database for providing their platforms and contributors for uploading their meaningful datasets.

Author contributions
Q.S., W.P. and N.Z. developed the idea and Q.S. and S.Z. selected the corresponding methodological content; Q.S., W.T. and X.W. performed the preliminary data analysis; Q.S., S.Z., W.T. and X.W. wrote the manuscript; N.Z., W.P., Q.W. and R.Z. reviewed and edited the article. N.Z., W.P., Q.W. and R.Z. reviewed and edited the article; W.P. and N.Z. managed the project. All authors have read and agreed to the published version of the manuscript.

Funding
Supported by National Natural Science Foundation of China (81902347).

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-68256-3.

Correspondence and requests for materials should be addressed to N.Z. or W.P.