



System biology analysis reveals circadian rhythm disorder associated with development and progression in colorectal cancer



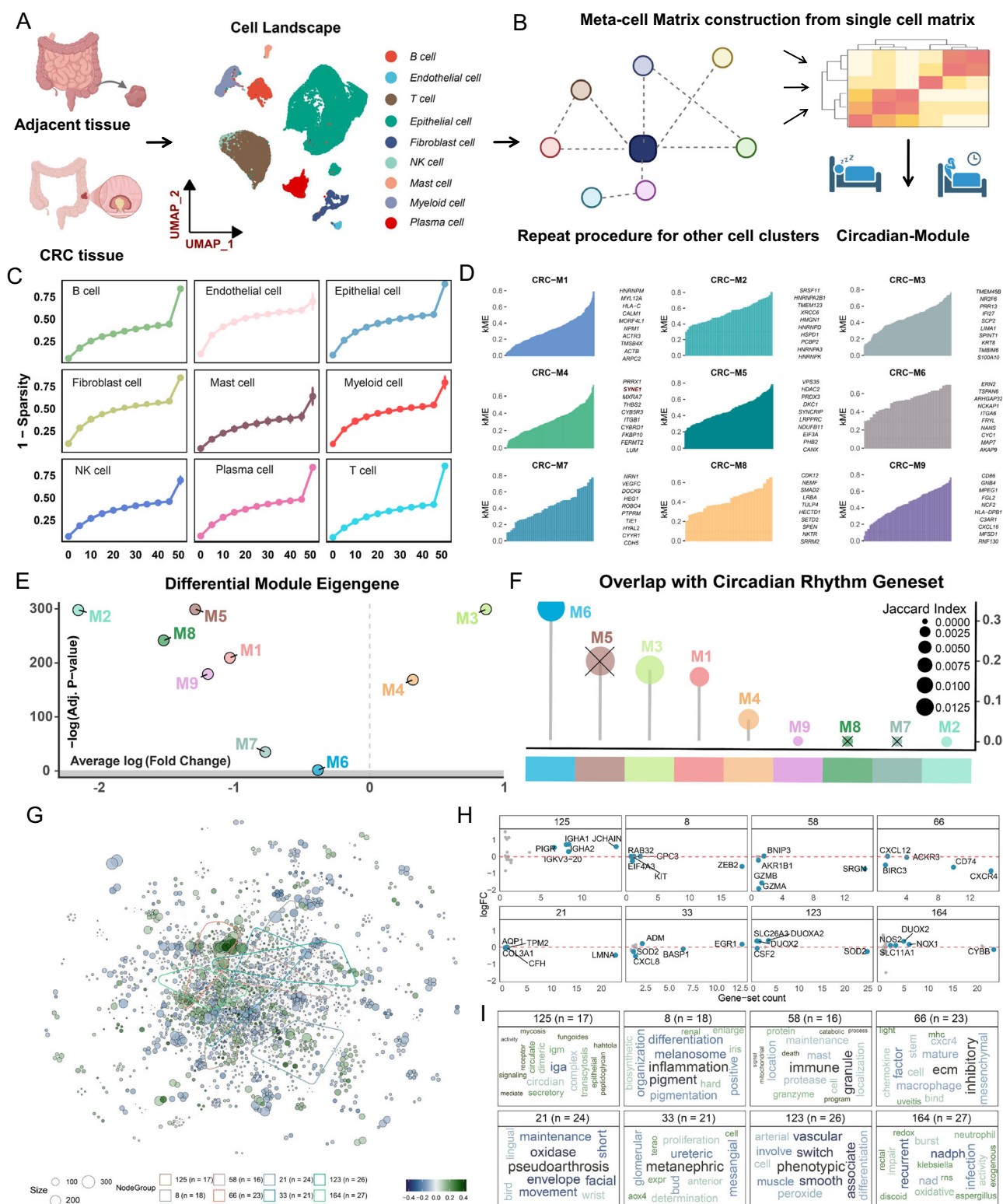
Shiqian Zhang^{1,2,13}, Shanshan Cai^{3,4,12,13}, Naijing Hou^{5,13}, Qian Guo⁶, Pengpeng Zhang⁷, Zhijie Zhao⁸, Song-Bin Guo⁹, Xufeng Huang¹⁰, Huaqing Wang^{1,2}, Haonan Zhang⁶, Chaoyang Yu⁶, Ruhao Wu⁶, Chunze Zhang^{1,2}✉, Shing Yau Tam¹¹✉ & Ge Zhang⁶✉

Circadian rhythm disorders represent an abstract concept lacking standardized quantitative metrics. Existing circadian indicators, including traditional rhythm parameters and a limited set of clock gene or physiological biomarkers, are insufficient to robustly capture steady-state endogenous circadian homeostasis in complex disease contexts, thereby constraining quantitative assessment of circadian disruption and limiting its translational applicability. Chronic circadian rhythm disruption is associated with various diseases, including metabolic disorders and malignancies. However, the mechanisms by which circadian disruption influences tumor microenvironment formation and colorectal cancer progression remain incompletely understood. This study employs systems biology analysis to decipher the molecular characteristics of circadian rhythm disruption in colorectal cancer progression. We analyzed single-cell RNA sequencing data from 13 CRC tissue samples and 12 normal mucosal samples, combined with 3733 samples from 34 public batch RNA, microarray, and single-cell RNA sequencing cohorts. We developed and validated the ClockProCRC system, which detects and quantifies intrinsic circadian misalignment in CRC. The ClockProCRC score elucidates how circadian misalignment drives CRC progression trajectories, shapes clinical phenotypes, regulates disease manifestations, and reshapes the tumor microenvironment. SYNE1 gene was identified as a key mediator of circadian misalignment, promoting tumorigenesis by driving epithelial-like phenotypic conversion and demonstrating therapeutic potential in colorectal cancer management. This study establishes a foundation for integrating rhythmic information into clinical practice and advances circadian biology research in the field of CRC.

In contemporary society, human lifestyles have undergone dramatic changes, which have profoundly impacted the body's circadian rhythm and its regulatory mechanisms. As a highly conserved evolutionary endogenous timing system, the circadian rhythm plays an indispensable role in maintaining normal physiological functions and homeostatic balance^{1–3}. However, various modern factors, such as frequent trans-meridian travel, irregular shift work, and excessive exposure to artificial light sources, have led to an increasing prevalence of Circadian rhythm disorder (CRD)^{4,5}. Numerous clinical and basic studies have demonstrated a strong link

between CRD and the development of various diseases, including cardiovascular disorders, metabolic diseases, neurological conditions, and multiple forms of cancer^{5–8}.

CRC is the third most commonly diagnosed cancer worldwide and the second leading cause of cancer-related death globally. In 2020, China accounted for 28.8% of all newly diagnosed CRC cases and 30.6% of CRC-related deaths worldwide^{9–11}. Increasing evidence suggests that disruptions in the circadian rhythm are closely associated with the onset of various cancers, including breast, prostate, colorectal, pancreatic, liver, and lung



cancers^{12–15}. These findings highlight the critical regulatory role of the intrinsic biological clock in cancer development.

Disruption of circadian rhythms can impact tumor cell biological behavior through multiple mechanisms. On one hand, circadian rhythm disturbances may interfere with the normal regulation of the cell cycle and weaken immune system function, resulting in an imbalance between cell proliferation and apoptosis, thereby creating favorable conditions for malignant tumor cell proliferation^{16,17}. On the other hand, circadian

disruption may also alter the ecological balance and metabolic function of the gut microbiota. The gut microbiota plays a crucial role in maintaining intestinal immune homeostasis and metabolic health, and changes in circadian rhythms may lead to alterations in microbiota composition and metabolic products, thereby triggering chronic inflammation and creating a microenvironment conducive to tumorigenesis^{18,19}. Additionally, the results of a study point to a direct effect of disruption of circadian core clock genes (ARNTL, PER2, or NR1D1) during selective splicing events associated with

Fig. 1 | High-dimensional network analysis of the circadian rhythm system in colorectal cancer development. **A** Schematic representation of a high-dimensional network analysis procedure. UMAP plot projecting individual cells in normal mucosa and colorectal cancer tissue from diseased patients (GSE166555 dataset). **B** Schematic overview of the high-resolution network analysis procedure. To address the sparsity inherent in scRNA-seq data, highly similar cells were aggregated into ‘Meta-cells’, representing transcriptionally similar cells within the circadian profile and capturing distinct cell states. **C** Depiction of expression matrix density (1, sparsity) for single-cell (sc), pseudo-bulk (PB), and meta-cell matrices, considering varying values of K in each cell type. **D** Module eigengene-based connectivity (kME) of genes within the nine identified co-expression modules. For each module, genes are ranked according to their kME values, which represent the correlation between individual gene expression and the corresponding module eigengene. Higher kME values indicate stronger intramodular connectivity and suggest that the gene is more

representative of the module. **E** Overlap analysis comparing sets of genes associated with circadian rhythms and identified modules using Fisher’s exact test. The similarity was assessed by the Jaccard similarity coefficient (Jaccard index, where X indicates non-significant overlap (FDR > 0.05)). **F** Differential module signature gene (DME) analysis of the circadian co-expression network revealed upregulated or downregulated modules between normal mucosal tissues and colorectal cancer tissues. **G** Identification, summarization, and visualization of higher-order functional themes underlying CTD were performed using vissE analysis. The overlap network was generated based on the gene-set similarity data and annotated using the MSigDB categories. **H** Genes with high logFC and high frequencies were considered key players in the biological processes represented by the communities. **I** vissE analysis automatically annotated each community by performing text mining analysis of either gene-set names or descriptions.

cancer markers, which may support the onset or progression of cancer²⁰. Although current research is increasingly refining the mechanisms linking biological clock disruption to CRC development²¹. However, unraveling these mechanisms is critical to developing targeted prevention and treatment strategies to address the impact of circadian rhythm disruption in colorectal cancer disease.

To systematically elucidate the potential molecular mechanisms underlying CRD in the initiation and progression of CRC and its impact on clinical disease progression, this study employs a systems biology approach, integrating high-throughput data with external cellular and molecular functional experiments. Through this comprehensive approach, we investigate the potential molecular mechanisms of CRD in CRC and its clinical significance. By developing the quantitative scoring system ‘ClockProCRC’, we are able to assess the endogenous circadian rhythm dysregulation expression profile in CRC patients, providing a comprehensive evaluation from both clinical and biological perspectives. Additionally, our study identifies that dysregulation of the novel biomarker SYNE1 gene in epithelial cells of CRC patient tissues may be associated with the progression of CRC.

Results

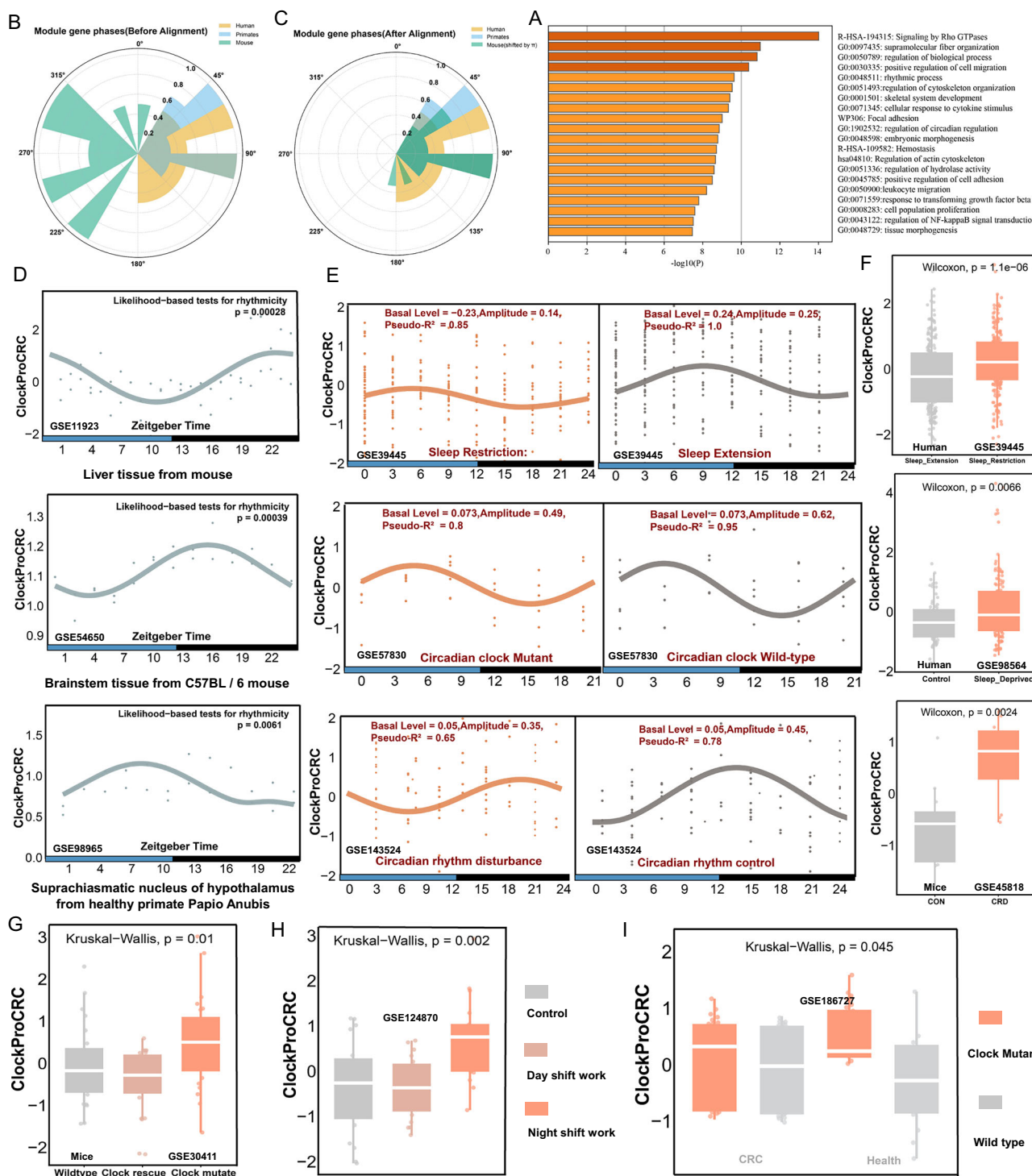
Single-cell transcriptomic profiling identifies CRC cell populations and maps CRD gene expression patterns

To systematically investigate the role of circadian rhythm disruption (CRD) in colorectal cancer (CRC), we established an integrative analytical framework and technical roadmap (Supplemental Fig. 1). Briefly, CRD-related genes were first defined from circadian perturbation datasets and subsequently projected onto bulk and single-cell CRC cohorts. To characterize the cellular landscape of colorectal cancer and establish the single-cell framework for subsequent circadian analyses, we first analyzed the GSE166555 scRNA-seq cohort after standard quality control and batch correction. After screening and filtering, a total of 94683 single cells were retained for subsequent analysis. The filtered cells underwent principal component analysis for dimensionality reduction, then uniform manifold approximation and projection (UMAP) clustering analysis revealed 19 distinct cell clusters based on similar gene expression patterns (Supplemental Fig. 2A). Subsequently, we identified nine major cell types: B cells, endothelial cells, T cells, epithelial cells, fibroblasts, natural killer (NK) cells, mast cells, stromal cells, and plasma cells (Supplemental Fig. 2B). We next analyzed the single-cell expression patterns of predefined CRD-related genes in the colorectal cancer scRNA-seq cohort, following cell clustering and annotation (Fig. 1A). In the GSE10644 dataset, 906 expressed genes were identified as the CRD gene set following a previously published screening strategy, and were subsequently used as the predefined CRD signature in our downstream analyses²². GO enrichment analysis revealed significant enrichment of CRD genes in processes related to circadian rhythm disorder, rhythmic processes, regulation of protein ubiquitination, and the circadian sleep-wake cycle (Supplemental Fig. 2C). The defined CRD gene set was mapped to the

GSE166555 single cell expression matrix, and a subset of CRD genes expressible in the scRNA-seq data set was screened out. Based on this, subsequent single cell expression distribution, cell type-specific enrichment and module score analysis were carried out. This ensured feature compatibility for downstream single-cell analyses, including cell-type-specific expression mapping and CRD module scoring.

Deciphering the transcriptome drivers of the circadian rhythm system during the development of colorectal cancer through high-dimensional network analysis

To address the inherent sparsity of high-dimensional data, we further aggregated highly similar cells into meta-cells. Using the meta-cell circadian expression matrix, we constructed networks for each cell type, adjusting the number of cells folded into a single meta-cell using KNN clustering (Fig. 1B). We observed a significant reduction in data sparsity in the meta-cell matrix compared to the single-cell matrix, with the sparsity reaching an optimal reduction at K = 50 (Fig. 1C). To reduce the complexity of the network, we merged modules with a similarity greater than 0.8, resulting in hierarchical clustering of nine circadian rhythm co-expression modules and their key genes (Fig. 1D, Supplemental Fig. 2D). Using UMAP for dimensionality reduction, we visualized key genes within the co-expression network and found that these circadian rhythm modules exhibited distinct spatial expression patterns and cell-type-specific expression profiles (Supplemental Fig. 2E). Bar plot displaying module membership values for the top 10 hub genes within the M4 module, where the SYNE1 gene exhibits a high value (Supplemental Fig. 2F). We used scatter plots to visualize the expression profiles of the nine CRC module genes in normal and colorectal cancer tissues. The results show that module genes 1, 3, 4, 6, and 9 are upregulated in tumor samples (Supplemental Fig. 2G). Modules 5 and 6 showed significant overlap with colorectal cancer cells, indicating a strong correlation between them (Supplemental Fig. 2H). To investigate the functional relevance of CRD-associated modules in the development of colorectal cancer, we compared gene expression in these modules between normal colorectal mucosa and CRC tissues, and further assessed their correlation with tumor-associated transcriptional features. Modules that were significantly upregulated in CRC tissue and correlated with malignant features were considered likely to be involved in the progression of CRC. At the cellular level, overlap analysis, further confirmed by Fisher’s exact test and Jaccard index, showed that four modules (M1, M3, M4, and M6) were significantly enriched in circadian-related features from multiple biological databases (BIOCARTA, KEGG, GO, REACTOME, and WIKIPATHWAYS), suggesting that the rhythm genes they contain may play a crucial role in colorectal cancer progression (Fig. 1E, Supplemental Fig. 2I). Differential module feature gene (DME) analysis further highlighted their distinct co-expression patterns during the transition from normal tissue to colorectal cancer (Fig. 1F). Thus, this 559-clock gene signature composed of four modules may represent the circadian rhythm transcriptional drivers (CTDs) promoting CRC progression (Supplemental Table 3).



CTD gene clusters reveal biological processes associated with CRC progression

To characterize the biological functions represented by CTD-associated genes, we applied *vissE* and GO enrichment analyses, which identified eight prioritized gene clusters corresponding to distinct biological themes. *vissE* analysis leverages the correlations between gene sets and the inherent hierarchy present in pathway databases and Gene Ontology. Initially, we calculated the similarity network between the CTD signature and all MSigDB gene sets based on overlap coefficients to identify gene sets that share similarities with the CTD signature. By performing graph clustering on the similarity

network, we identified eight prioritized gene clusters (Fig. 1G). To elucidate the biological relevance of each gene cluster, we mapped the gene-level statistics (fold change, Log FC) from differential expression analysis between normal mucosal and colorectal cancer tissues to each cluster (Fig. 1H). Additionally, GO enrichment analysis was performed on the CTD genes to explore the biological pathways they are involved in or regulate. The results of the *vissE* analysis revealed eight prominent biological functional themes within these gene clusters, which are primarily associated with circadian rhythm signaling, immune responses, extracellular matrix organization, oxidative stress, and rhythmic processes (Figs. 1I and 2A).

Fig. 2 | The robustness of ClockProCRC in circadian rhythm misalignment quantification. **A** Functional enrichment analysis of four colorectal cancer circadian disruption-related modules (M1, M3, M4, and M6). **B** Distribution of module gene phases before alignment across species. **C** Distribution of module gene phases after alignment (mouse shifted by 12 h). **D** ClockProCRC demonstrated accurate detection of circadian rhythms in different samples, including liver tissue from mice (GSE11923), heart tissue from C57BL/6 gene mutation mice (GSE54650), and the inferior supraoptic nucleus of *Papio Anubis* (GSE98965). **E** Comparison of ClockProCRC temporal profiles under circadian-perturbed and corresponding control conditions, including sleep restriction versus sleep extension, circadian clock-mutant versus wild-type mice, and circadian rhythm disturbance versus control conditions. Individual points represent observations at the indicated Zeitgeber times, and solid curves represent fitted 24 h rhythmic profiles. Basal level and amplitude characterize the estimated baseline and magnitude of oscillation, respectively, whereas pseudo- R^2 reflects the goodness-of-fit of the rhythmic model. These parameters demonstrate condition-dependent alterations in ClockProCRC temporal organization rather than a uniform directional change in baseline levels across all datasets. **F** Distribution of ClockProCRC scores across three independent human transcriptomic datasets (GSE39445: Sleep Extension ($n = 222$), Sleep

Restriction ($n = 216$); GSE98564: Well-rested Control ($n = 71$), Sleep Deprived ($n = 122$); and GSE45818: Wildtype ($n = 12$), Clock mutant ($n = 12$)), illustrating the association between ClockProCRC and circadian rhythm perturbation under different physiological and experimental conditions. ClockProCRC scores were calculated individually for each sample based on normalized bulk expression matrices using the predefined circadian-related gene set and the expression-matched random background correction strategy described in the methods. Each dot represents one independent biological sample. **G** Distribution of ClockProCRC scores in mice harboring core clock gene mutations (GSE30411: Wildtype ($n = 24$), Clock rescue ($n = 24$), Clock mutant ($n = 24$)). Scores were computed per sample using the same standardized pipeline as in (C). Each data point corresponds to one sample. **H** Distribution of ClockProCRC values across three distinct work-shift patterns (Control ($n = 18$), Day shift ($n = 18$), Night shift ($n = 18$)) derived from GSE124870. Individual sample-level scores are shown as jittered points overlaid on boxplots. **I** ClockProCRC scores in circadian clock mutant and wild-type mice (GSE186727: CRC ($n = 76$), Health ($n = 23$)), further stratified into colorectal cancer and normal mucosa subgroups. Each dot represents one independent sample. Higher ClockProCRC scores indicate a greater degree of endogenous circadian rhythm disruption.

ClockProCRC score detection and quantification of endogenous circadian disorder states

To overcome the limitations of using a single clock gene or a small number of molecular markers to infer overall circadian status, we developed a quantitative scoring system, ClockProCRC, to measure endogenous circadian dysregulation in colorectal cancer. This system aims to identify and quantify the molecular mechanisms driving circadian rhythm disruption and activity in these patients. To validate the performance and robustness of ClockProCRC, we employed 16 independent public datasets encompassing various circadian rhythm patterns. To enable cross-species comparisons, module eigenvalues were first z-score normalized within homologous gene sets. In addition, given the nocturnal behavior of mice compared with the diurnal patterns of humans and primates, mouse Zeitgeber time was shifted by 12 hours to align circadian phases across species (Fig. 2B, C). Notably, using likelihood-based rhythmicity tests, we observed significant temporal rhythmicity of ClockProCRC in both rodent and primate models under standard or circadian-perturbed conditions. These rhythmic patterns were detected in the hypothalamic suprachiasmatic nucleus of healthy primates (*Papio anubis*) and in the brainstem and liver tissues of C57BL/6 mice (Fig. 2D). Furthermore, ClockProCRC distinguished temporal patterns between circadian-perturbed and corresponding control conditions across independent datasets, including human subjects exposed to sleep restriction versus sleep extension, circadian clock-mutant versus wild-type mice, and other experimentally perturbed circadian conditions. Importantly, these perturbations produced condition-dependent alterations in baseline levels and oscillatory characteristics rather than a uniform directional change in baseline ClockProCRC across all datasets. The relatively circadian-robust conditions generally exhibited more pronounced oscillatory profiles, as reflected by the corresponding rhythmic model parameters (Fig. 2E). At the group level, several independent validation cohorts showed significantly different ClockProCRC distributions between circadian-perturbed and corresponding control conditions, with higher scores observed in the disrupted groups in the cohorts displayed in Fig. 2F and Supplemental Fig. 3A. Furthermore, ClockProCRC effectively discriminated genetically defined circadian states, with significant differences among wild-type, clock-mutant, and clock-recovery conditions, with the highest levels observed in circadian gene mutant mice compared to wild-type mice, and progressively lower scores in the gene recovery mutant mice (Fig. 2G).

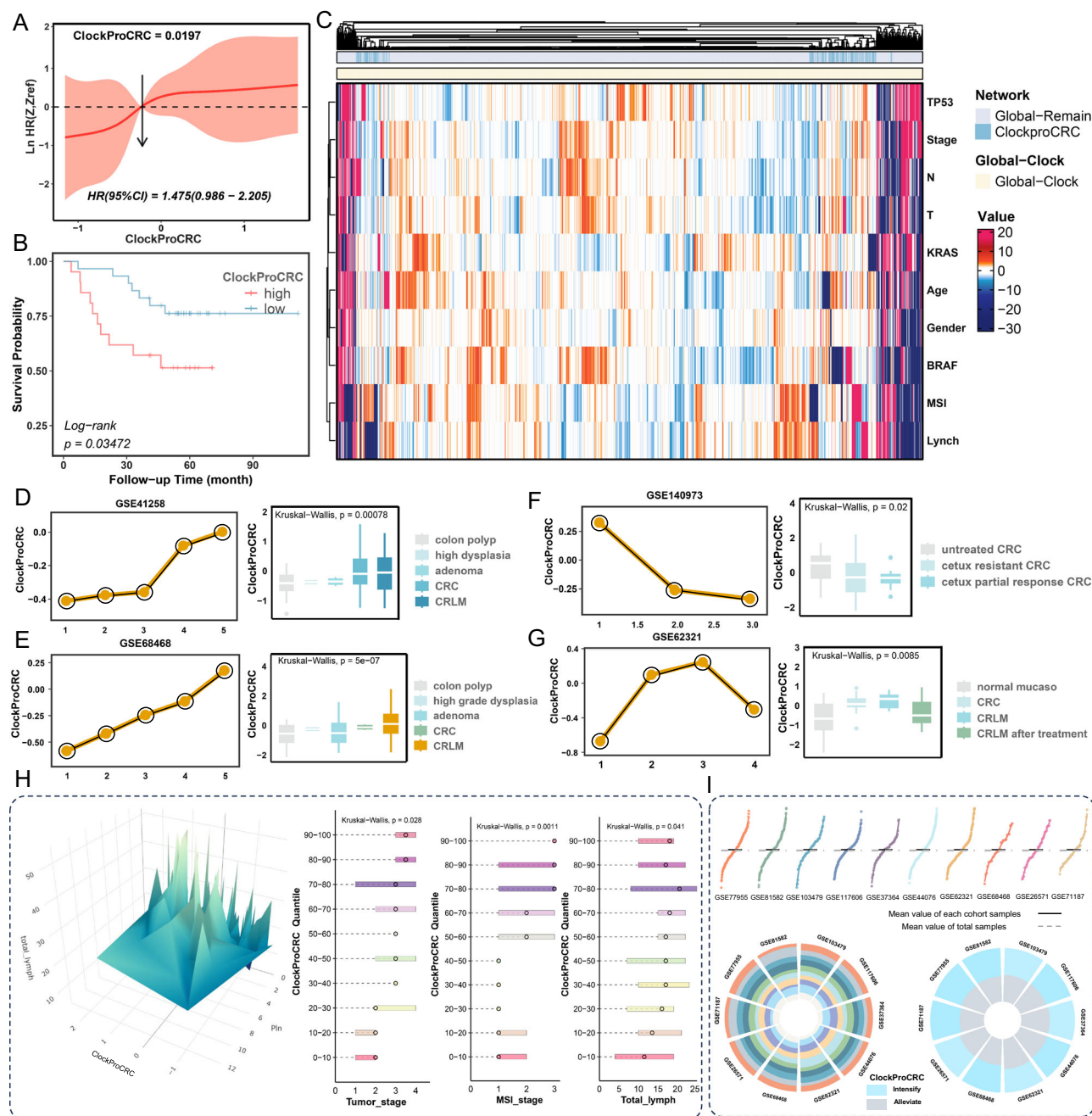
Shift work is a major disruptor of circadian rhythms, as the internal circadian clock becomes misaligned with the external light-dark cycle, which maybe has been associated with an increased risk of colorectal cancer. To assess the ability of ClockProCRC to detect circadian rhythm disruptions induced by shift work, we employed a shift-work mouse model. Notably, we observed significant differences in ClockProCRC scores across various shift work patterns, with the highest levels observed in night shift mice, followed

by day shift mice, and the non-shift control group (Fig. 2H). Furthermore, ClockProCRC scores were higher in circadian rhythm mutant mice compared to wild-type mice, and also higher in CRC samples compared to healthy controls (Fig. 2I). These findings demonstrate that ClockProCRC can effectively detect circadian rhythm status under different conditions.

Analysis of clinical significance of ClockProCRC score

Our previous analysis highlighted the correlation between ClockProCRC scores and the biology and rhythm of CRC. To further explore the clinical manifestations associated with CRD, we conducted an initial evaluation focusing on the predictive accuracy of ClockProCRC scores for clinical outcomes. Restricted cubic spline analysis revealed a significant linear prognostic relationship between ClockProCRC scores and colorectal cancer survival, corresponding HR = 1 near the reference point, and the overall trend increased risk with increasing ClockProCRC (Fig. 3A). Kaplan-Meier survival curves demonstrated that an increase in ClockProCRC scores was associated with a gradual decline in survival probability in CRC patients (log-rank $P < 0.01$, Fig. 3B). Additionally, we assessed the correlation between ClockProCRC scores and various clinical variables, including clinical presentation, comorbidities, tumor staging, and microsatellite instability (MSI) status in the GSE103479 and GSE39084 datasets. Hierarchical clustering based on Spearman's correlation revealed a strong association between ClockProCRC scores and tumor stage, as well as the mutation status of genes such as TP53 and KRAS (Fig. 3C, Supplemental Fig. 3B). ClockProCRC scores progressively increased with the progression of colorectal cancer (normal colorectal mucosa, adenoma, severe dysplasia, colorectal cancer, and colorectal cancer with liver metastasis (CRLM), Fig. 3D). This trend was further confirmed in the GSE68468 dataset, using the same grading criteria ($P < 0.0001$, Fig. 4E). These results suggest that ClockProCRC scores accurately reflect disease activity and progressively increase with the severity of the disease.

Moreover, we observed that ClockProCRC scores fluctuated with disease activity and remission. Specifically, ClockProCRC levels were higher during the active phase of the disease and decreased during remission. Notably, ClockProCRC scores were significantly elevated during the acute phase of the disease, particularly in untreated CRC patients, followed by treatment-resistant CRC patients and treatment-sensitive patients ($P < 0.05$, Fig. 3F). In a separate cohort of CRLM, ClockProCRC levels were highest in liver metastases, followed by primary CRC lesions, treated CRLM, and normal colorectal mucosa ($P < 0.05$, Fig. 3G). Further analysis revealed that as the ClockProCRC decile scores increased, tumor clinical stage ($P < 0.05$), MSI index ($P < 0.01$), and the total number of harvested lymph nodes ($P < 0.05$) also increased (Fig. 3H). Analysis of more than 1400 colorectal cancer patients across 10 independent cohorts showed a broadly consistent



distribution pattern of ClockProCRC scores. Using the cohort-specific median as the cutoff, samples were divided into ClockProCRC-Intensify and ClockProCRC-Alleviate groups with approximately balanced proportions, supporting the generalizability and robustness of the scoring system across independent (Fig. 3I).

Dissecting the circadian rhythm disorder of colorectal cancer from single-cell resolution

Using data from two CRC single-cell cohorts (GSE200977 and GSE196964), we investigated the mechanisms underlying microenvironmental changes induced by circadian rhythm misalignment in CRC progression. Notably, compared to adjacent normal mucosal tissue, CRC cells exhibited significantly higher ClockProCRC scores (Fig. 4A). To assess the tissue enrichment of different cell populations, we performed *Ro/e* analysis. We found that epithelial/plasma cell populations preferentially occupied the core regions of CRC with exacerbated CRD, while fibroblasts and

endothelial cells were preferentially enriched in the regions with stable circadian rhythm. In regions with intensified circadian rhythm disruption, we observed a reduction in endothelial and fibrotic cell expression, accompanied by an increase in pro-inflammatory epithelial and plasma cell numbers. *Ro/e* values greater than 1 indicate enrichment of the corresponding cell type relative to expectation, whereas values lower than 1 indicate depletion (Fig. 4B, C).

We further transformed the traditional CRC cellular map into a single-cell signaling map to explore changes in the perceived microenvironment induced by alterations in cellular circadian rhythm states during dynamic progression (Supplemental Fig. 3C). We observed significant spatial overlap between epithelial cell-centered communication networks and disrupted circadian rhythm states. In contrast, fibroblast-driven interactions primarily occurred in the spatial regions with a stable circadian rhythm (Fig. 4D). Overall, exacerbation of CRD enhanced intercellular crosstalk (Fig. 4E, Supplemental Fig. 3D). Specifically, under exacerbated CRD, epithelial/

Fig. 3 | Macroscopic clinical interpretability of circadian rhythm disorders.

A Restricted cubic spline analysis of the association between ClockProCRC and overall survival. The red line represents the estimated log hazard ratio [ln(HR)], and the shaded region indicates the 95% confidence interval, based on a Cox proportional hazards model. The reference ClockProCRC value was set at 0.0197, where $\ln(\text{HR}) = 0$ (equivalent to $\text{HR} = 1$). The dashed horizontal line denotes the reference risk level. **B** Kaplan-Meier curves of prognostic correlations with overall survival in patients with colorectal cancer in the GSE17537 cohort ($n = 55$), with p values derived from the log-rank test. **C** Heatmap of associations between circadian gene profiles and clinical variables in cohort GSE39084, where genes are grouped by marker (ClockProCRC, CTD). “Global-Clock” represents the comprehensive trend of overall expression changes of circadian rhythm-related genes in the cohort. The color scale indicates the direction and magnitude of ClockProCRC-associated circadian signal changes across cohorts (red: intensification; blue: alleviation). Color intensity represents the standardized effect size of the trend. **D–G** Fold and box plots depicting the mean and distribution of ClockProCRC scores across the four cohorts (GSE41258: colon polyps ($n = 24$), high dysplasia ($n = 10$), adenoma ($n = 6$), CRC samples ($n = 185$), CRLM ($n = 67$); GSE68468: colon polyps ($n = 48$), high dysplasia ($n = 7$), adenoma ($n = 8$), CRC samples ($n = 193$), CRLM ($n = 47$); GSE140973: CRC

samples ($n = 10$), Cetuximab partial response to CRC ($n = 11$), Cetuximab resistant to CRC ($n = 26$); GSE62321: normal mucosal tissues ($n = 17$), CRC samples ($n = 20$), CRLM samples ($n = 10$), CRLM after treatment ($n = 10$)), highlighting associations with colorectal cancer course, progression, and treatment remission timeline.

H Three-dimensional plots depicting the interactions between ClockProCRC, total lymph nodes obtained, and positive lymph nodes, as well as Gantt plots of tumor stage, microsatellite stability, and total lymph nodes harvested in the ClockProCRC quartiles. **I** Assessment of the robustness and generalizability of ClockProCRC. Top: Robustness and generalizability assessment of ClockProCRC in 10 independent colorectal cancer cohorts ($n > 1400$). The ClockProCRC score is calculated for each sample, and the gray dashed line represents the median ClockProCRC score of all samples. Left: The circular bar chart illustrates the distribution of ClockProCRC deciles among ten cohorts. Each bar represents the proportion of the number of samples in different ClockProCRC quantile intervals in a cohort, reflecting the stability of the ClockProCRC score distribution in the data set. Right: Donut charts show the proportion of samples in each cohort classified as ClockProCRC-Intensify or ClockProCRC-Alleviate using the cohort-specific median ClockProCRC as the cutoff value. Sky blue is the proportion of the Intensify group in different cohorts, and gray is the proportion of the Alleviate group in different cohorts.

plasma cells showed stronger interaction weights, while fibroblasts, crucial for epithelial-mesenchymal transition and tumor cell suppression, displayed reduced communication activity (Fig. 4F, Supplemental Fig. 3E). Furthermore, progression from the alleviate to the intensify CRD state was associated with decreased apoptotic pathway activity and increased pro-inflammatory cell-cell communication signaling, larger Euclidean distances indicate stronger differences in intercellular communication patterns between circadian disruption states (Fig. 4G). These findings highlight the potential role of epithelial cells, fibroblasts, and their related pathways in CRC progression under circadian rhythm misalignment.

Key circadian and night regulatory genes screening and their effects on circadian rhythm disorder in colorectal cancer

To further investigate the regulatory mechanisms of circadian rhythm disorder genes in CRC progression, we constructed a regulatory network based on CTD genes derived from high-dimensional co-expression network analysis and identified key dysregulated genes (Fig. 5A). During the construction of regulatory networks, the strength of regulatory interactions between gene pairs was calculated using transcription factor binding sites and transcription factor motif data. After 100 successive generations, 61 key axes of regulatory interactions emerged (Fig. 5B). By integrating differential gene expression and regulatory strength, we quantified gene dysregulation scores. Results indicate that NFATC2's regulatory strength over SYNE1 was significantly elevated in colorectal cancer tissues (0.607, 95% CI 0.412–0.802), whereas it was markedly weakened in normal mucosal tissues (–0.102, 95% CI –0.058–0.262), suggesting this regulatory relationship is markedly activated in the tumor state (Fig. 5C). Furthermore, during the iterative scoring and comparison of candidate regulatory axes, SYNE1 consistently demonstrated significant effect size and network robustness in positively regulating colorectal cancer, outperforming other high-frequency regulatory combinations. Consequently, the NFATC2-SYNE1 axis was identified as a key dysregulated process accelerating CRC progression, with SYNE1 receiving particular attention as a critical gene involved in circadian rhythm dysregulation (Supplemental Fig. 3F).

Notably, SYNE1 exhibited a time-dependent expression trend across the sampled Zeitgeber times, and this trend differed between control and circadian disruption conditions. In the GSE43073 and GSE35026 datasets, fitted curves suggested an oscillatory expression pattern of SYNE1 over the cycle relative to ZT0 (lights on), whereas this pattern became weaker or temporally shifted under circadian misalignment (Fig. 5D). This rhythmic change may play a potential role in CRC initiation and progression. In human experiments involving normal controls, sleep recovery groups, and sleep deprivation groups, SYNE1 gene expression levels increased in the sleep deprivation group while showing a decreasing trend in the sleep

recovery group (GSE56931). SYNE1 gene expression was significantly higher in both the Clock-rescued and Clock-mutant groups compared to wild-type mice, though the Clock-rescued group exhibited lower expression than the Clock-mutant group (GSE30411). These findings indicate that circadian misalignment disrupts the rhythmicity of SYNE1 gene, potentially influencing colorectal cancer development (Fig. 5E).

Notably, SYNE1 exhibited a modest time-dependent oscillatory expression profile across the sampled Zeitgeber times, although its rhythmicity was substantially weaker than that of canonical clock genes (Fig. 5D and Supplemental Fig. 4). In GSE35026, the circadian disruption condition showed a lower pseudo- R^2 (0.028 vs 0.081) and amplitude (0.031 vs 0.046) than the corresponding control condition. A similar pattern was observed in GSE43073, where Clock-mutant samples exhibited a lower pseudo- R^2 (0.119 vs 0.305) and amplitude (0.15 vs 0.21) than wild-type samples. These findings suggest that circadian perturbation is associated with attenuation or remodeling of the temporal expression profile of SYNE1 rather than a complete loss of rhythmicity. Consistent with this observation, SYNE1 expression also varied across independent circadian perturbation models. In GSE56931, SYNE1 gene expression was increased in the sleep-deprivation group and showed a decreasing trend following sleep recovery. In GSE30411, SYNE1 gene expression was higher in both Clock-mutant and Clock-rescued mice than in wild-type mice, while the Clock-rescued group showed lower expression than the Clock-mutant group (Fig. 5E). Collectively, these results indicate that SYNE1 expression is sensitive to circadian perturbation, supporting its potential involvement as a downstream circadian disruption-associated factor rather than a canonical clock oscillator.

SYNE1 gene interference will promote the development of CRC

To investigate the role of the SYNE1 gene in CRC progression, we first examined its expression levels in normal mucosal and CRC tissues. The results indicated that SYNE1 gene expression was significantly higher in CRC tissues compared to adjacent normal mucosal tissues (Supplemental Fig. 5A, B). Similarly, in the GSE41258 dataset, Kaplan-Meier analysis revealed that lower SYNE1 expression was associated with improved survival prognosis (Supplemental Fig. 5C). Functional enrichment analyses were performed to investigate the biological roles associated with SYNE1. GO analysis revealed that SYNE1-related genes were significantly enriched in processes such as regulation of cell migration and inflammatory response (Fig. 5F). Consistently, GSEA based on Hallmark pathways indicated enrichment in epithelial-mesenchymal transition, inflammatory signaling, and adherens junction pathways (Fig. 5G).

In metastasis-related gene expression profiling, we found that SYNE1 gene expression levels were significantly higher in CRLM than in primary

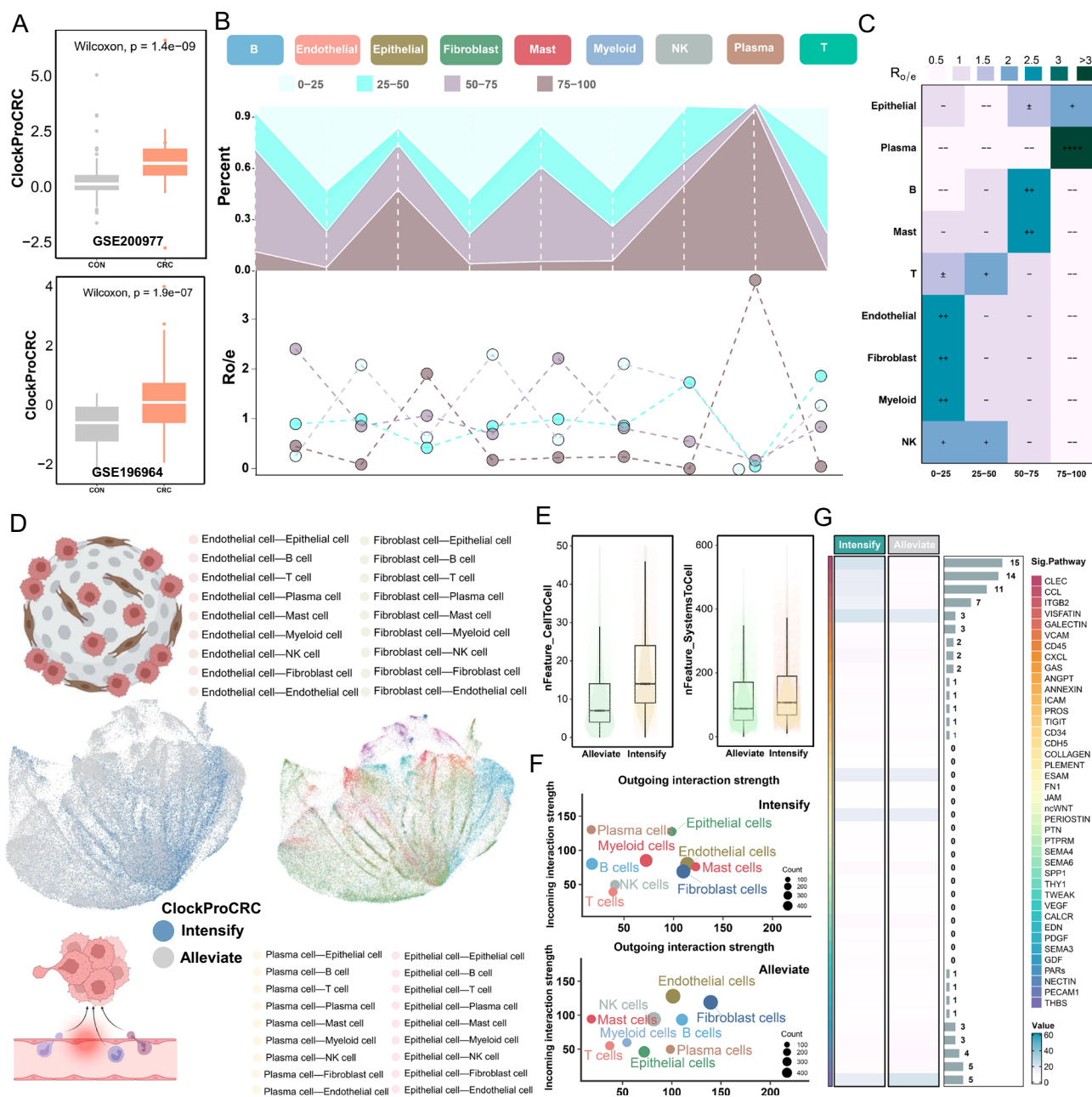
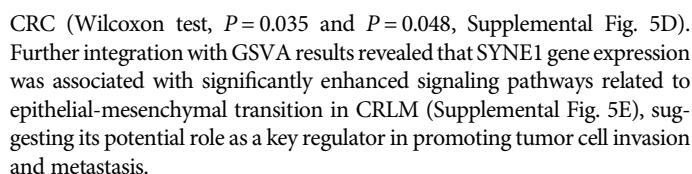


Fig. 4 | Biological effects of circadian dysregulation on colorectal carcinogenesis.

A Comparison of ClockProCRC distribution in different colorectal cancer single-cell datasets: colorectal cancer tissues and adjacent normal mucosal tissues (GSE200977, GSE196964). **B** Distribution of major cell types across different tumor proportion intervals. The stacked area plot shows the relative percentage of each cell type across tumor proportion groups (0–25%, 25–50%, 50–75%, and 75–100%). The lower panel presents the Ro/e (observed/expected) ratio for each cell type, indicating enrichment or depletion across tumor progression stages. **C** Tissue preference analysis of different cell types during colorectal cancer progression. The heatmap displays the ratio of observed to expected cell numbers (Ro/e) for each cell type across tumor proportion intervals. Rows represent cell types and columns represent tumor proportion intervals. Higher Ro/e values indicate enrichment of the corresponding cell type in the specified tumor proportion group. Ro/e values > 1 indicate enrichment, whereas values < 1 indicate depletion of the corresponding cell type. **D** Cell-cell signaling mapping within colorectal cancer cells allows observation of relationships occupying spatial regions of their own mechanisms (circadian rhythm

disruption) as well as relationships overlapping with other cell-cell relationships. **E** Comparison of System to Cell and Cell to Cell signaling activity in colorectal cancer cells in states of remission and exacerbation of circadian rhythm disruption according to ClockProCRC scores. **F** Identification of cell populations that undergo significant changes in sending or receiving signals between states of remission and exacerbation of circadian disruption. Two-dimensional spatial visualization of the comparison of efferent and afferent interaction strengths. **G** Signaling pathways with differential intercellular communication between the ClockProCRC intensify and alleviate states. The left heatmap-like panels represent the relative activity of individual signaling pathways in the two circadian disruption states (Intensify and Alleviate). The middle bars indicate the extent of pathway-specific differences, measured as Euclidean distance, and the right panel lists the significantly altered signaling pathways. The color scale (Value) represents relative pathway activity strength. Larger Euclidean distances indicate greater differences in communication network organization between the two states.



To further explore the functional significance of SYNE1 gene, we utilized siRNA targeting SYNE1. As shown in Fig. 6A, siRNA transfection effectively downregulated SYNE1 expression in both HCT116 and SW480 CRC cell lines, and these cell lines were subsequently chosen for further experiments.

Fig. 5 | Construction of gene regulatory network and biological significance of SYNE1 gene. **A** GRNs were constructed by predicting binding sites for TFs within promoter regions using TF motif data. Gene dysregulation was identified by integrating three criteria, including differential regulation, differential expression of targets, and concordance between differential regulation and differential expression. **B** Identifying dysregulations that are strongly associated with CRC progression. In each iteration of the genetic algorithm, the frequency of each disorder in the first ten individuals is colored. **C** Key gene dysregulation axis NFATC2-SYNE1. Regression lines were calculated by univariate regression to visualize differences in gene regulation under different conditions. **D** Temporal expression profiles and 24-h rhythmicity analysis of SYNE1 under normal and circadian-disrupted conditions in the GSE35026 and GSE43073 datasets. Each dot represents one individual sample measured at the indicated Zeitgeber time. The solid curve indicates the fitted rhythmic trend generated using a fixed 24 h sinusoidal model, which was applied to

illustrate the overall oscillatory pattern rather than to interpolate each individual data point. Rhythmicity P indicates statistical evidence for a 24 h oscillatory component, pseudo- R^2 represents the goodness-of-fit of the rhythmic model, and amplitude reflects the magnitude of the fitted oscillation. **E** Distribution of SYNE1 expression across independent datasets under different circadian perturbation conditions. Boxplots show SYNE1 expression levels in human and mouse datasets comparing control versus circadian disruption-related conditions, including sleep deprivation (GSE98565), sleep recovery and sleep deprivation (GSE56931), and wild-type, clock rescue, and clock mutant mice (GSE30411). **F** GO enrichment analysis of SYNE1-associated genes in the colorectal cancer transcriptomic dataset GSE71187. According to the enrichment degree (number of enriched genes), different categories of GO items are displayed, and their ontologies are marked with different colors. **G** GSEA of Hallmark pathways based on genes ranked according to their correlation with SYNE1 expression in the GSE71187 dataset.

To assess the contribution of the SYNE1 gene to CRC cell phenotypes, we designed and synthesized two specific siRNAs targeting SYNE1, si-SYNE1-01 and si-SYNE1-02. Transwell assays demonstrated that knockdown of SYNE1 gene significantly reduced proliferation and invasion in both cell lines (Fig. 6B, C). To further explore potential mechanisms, we performed transcriptomic analysis of colorectal cancer cells under different conditions, including SYNE1 silencing (si-SYNE1), CLOCK silencing (si-CLOCK), and melatonin treatment (si-CLOCK + melatonin). RNA-seq analysis revealed significantly increased circadian dysregulation in clock-silenced cells compared to controls (higher ClockProCRC levels), which was partially alleviated by melatonin treatment (Fig. 6D). These findings indicate that melatonin treatment can partially mitigate circadian dysregulation in colorectal cancer. Consistent with the macroscopic finding that circadian dysregulation leads to SYNE1 upregulation and impaired rhythmic expression, qRT-PCR revealed a clear circadian rhythm in SYNE1 gene expression under control conditions (si-NC, $P = 2.5 \times 10^{-4}$), disrupted by clock silencing (si-CLOCK, $P = 0.0704$), and restored following melatonin treatment (si-CLOCK + melatonin, $P = 3.4 \times 10^{-3}$, Fig. 6E). In the si-CLOCK group, basal SYNE1 gene expression was elevated compared to the control group, and both the circadian oscillation and phase of the SYNE1 gene were altered. Results indicate that the expression inflection point of the SYNE1 gene occurs during the transition between day and night (approximately ZT10–ZT12h) (Fig. 6E). These findings further confirm the positive influence of circadian rhythm disruption on SYNE1 regulation at the molecular level of gene control, a phenomenon we validated in colorectal cancer cells (Fig. 6F). ROC analysis demonstrated strong predictive value for endogenous circadian rhythm disruption of SYNE1 (AUC = 0.915, Fig. 6G).

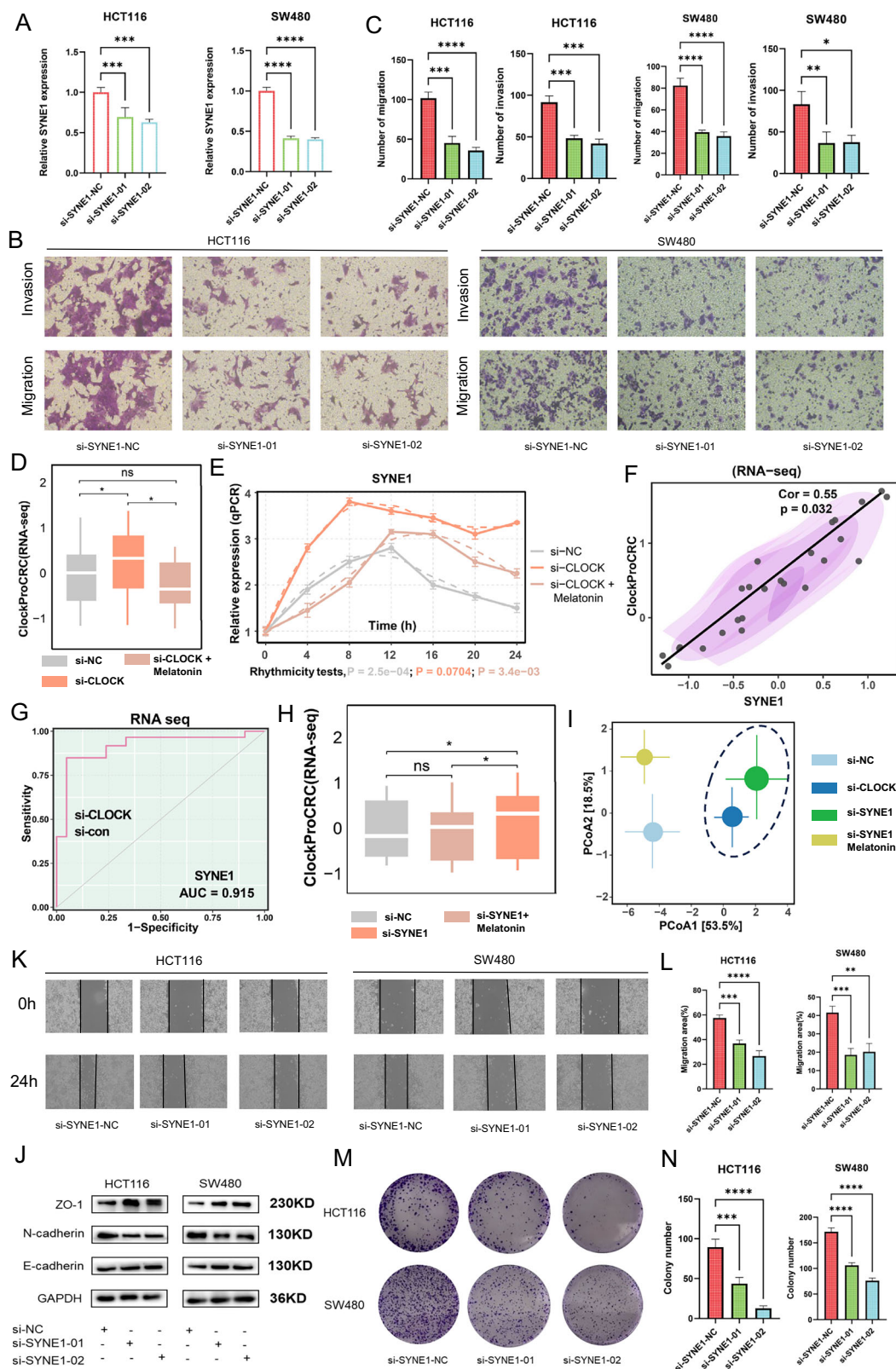
RNA sequencing analysis revealed significant circadian rhythm disruption in the si-SYNE1 group compared to the si-NC, which was markedly alleviated by melatonin treatment (Fig. 6H). Dimension reduction analysis demonstrated strong transcriptional similarity between the si-CLOCK and si-SYNE1 two groups, indicating these perturbations share common regulatory effects (Fig. 6I). Western blot analysis further revealed that SYNE1 gene knockout significantly alters the expression pattern of epithelial markers. Specifically, in the si-SYNE1 group, E-cadherin and ZO-1 levels were markedly upregulated, while expression of the mesenchymal marker N-cadherin was significantly downregulated (Fig. 6J, Supplemental Figs. 6, 7). This suggests that SYNE1 gene may accelerate distant organ metastasis by promoting epithelial-mesenchymal transition, thereby mediating phenotypic reprogramming and immune evasion in colorectal cancer cells. Similarly, wound healing assays showed that compared to si-NC cells, si-SYNE1 cells exhibited a reduced migration area after 48 hours of culture (Fig. 6K, L). To confirm the impact of SYNE1 gene on CRC cell proliferation, colony formation assays were performed on HCT116 and SW480 cells. SYNE1 gene knockdown significantly decreased the number of colonies formed by both cell lines, indicating that SYNE1 gene may play a critical role in regulating CRC cell proliferation (Fig. 6M, N).

Discussion

The circadian clock functions as an internal time-keeping system and an evolutionarily conserved molecular mechanism that generates circadian rhythms and coordinates diverse physiological processes, including energy metabolism and homeostasis^{23,24}. Increasing evidence suggests that disruptions in circadian rhythms play a crucial role in tumorigenesis, promoting cancer development and progression^{13,25–27}. Moreover, circadian rhythm disruption has been reported to interfere with multiple signaling pathways, and altered expression of PER family genes has been associated with tumor growth, invasiveness, and metastatic potential in several cancer contexts^{28,29}. CRY1, a member of the cryptochrome family, is upregulated in CRC and is significantly associated with tumor progression and poor prognosis^{30,31}. Importantly, core circadian components directly or indirectly regulate the expression of hundreds of peripheral circadian genes, many of which appear to be linked to cancer invasiveness³². Overall, accumulating evidence suggests that dysregulation of circadian rhythm genes is associated with increased tumor heterogeneity, unfavorable prognosis, and resistance to anticancer therapies, although the mechanistic links among these processes remain to be fully defined³³.

In recent years, CRD has increasingly been recognized as a significant risk factor for CRC. This disruption may promote tumor initiation and progression through multiple mechanisms. First, circadian rhythm disturbances may interfere with the normal regulation of the cell cycle and impair immune system function, leading to an imbalance between cell proliferation and apoptosis^{16,17}. This imbalance creates a favorable environment for the malignant proliferation of tumor cells. Studies have shown that the aberrant expression of core circadian genes, such as BMAL1 and PER2, is closely linked to the proliferation rate of tumor cells. The underlying mechanisms likely involve the regulation of cyclin-dependent kinases (CDKs) and cyclins^{34–36}. Furthermore, circadian rhythm disruption may also affect the ecological balance and metabolic functions of the gut microbiota. The gut microbiota plays a critical role in maintaining intestinal immune homeostasis and metabolic health, and circadian alterations can lead to changes in its composition and metabolic products. These changes may trigger chronic inflammatory responses, creating a microenvironment conducive to tumorigenesis^{18,19}. Research has demonstrated that in CRC patients, the diversity of the gut microbiota is significantly reduced, with an increase in the relative abundance of pro-inflammatory bacterial genera, accompanied by a decrease in beneficial metabolic products such as short-chain fatty acids^{37,38}.

Clinical studies have demonstrated a significant association between the expression levels of circadian rhythm genes in tumor tissues of CRC patients and both tumor metastatic potential and patient prognosis. For example, studies have shown that the disruption of the PER1 gene is associated with an increased risk of distant metastasis and poorer survival rates in CRC patients³⁹. Although these studies provide valuable insights into the role of circadian rhythm disruption in CRC, they predominantly focus on single-gene analyses, lacking comprehensive and systemic multi-gene co-expression analyses⁴⁰. Given the complexity of circadian rhythm disruption,



it is likely to exert its biological effects and mechanisms in CRC development through the modulation of multiple gene modules and signaling pathways.

To systematically elucidate the potential molecular mechanisms of CRD in CRC development and its impact on clinical disease progression, this study integrates high-throughput data with external cellular and molecular functional validation experiments through high-dimensional

artificial intelligence algorithms^{11,41–45}. We developed the quantitative scoring system ClockProCRC, which enables the assessment of endogenous circadian rhythm dysregulation expression profiles in CRC patients, providing a comprehensive evaluation from both clinical and biological perspectives. After analyzing multiple external public datasets related to CRC and circadian rhythm, our findings demonstrate that this scoring system

Fig. 6 | Functional validation of SYNE1 gene in circadian disruption-associated colorectal cancer progression. **A** The efficiency of SYNE1 gene knockdown by qRT-PCR in HCT116 and SW480 cell lines. **B** The transwell assay shows that after the knockdown of SYNE1, both invasion and migration are reduced compared to NC. **C** The statistical graph of invasion and migration in HCT116, SW480 cell lines transwell statistic graph. The mean $\pm$ SD of the bar graph represents the mean $\pm$ standard deviation from 3 independent biological replicates, where the value for each experiment is the average of 4 random field counts in that experiment. **D** ClockProCRC score significantly increased after CLOCK knockdown (si-CLOCK) compared to control, and was partially rescued by melatonin supplementation. **E** qPCR time-course analysis showing the rhythmic expression pattern of SYNE1 in si-NC, si-CLOCK, and si-CLOCK + melatonin conditions; rhythm restoration observed with melatonin treatment. Rhythmicity was assessed using a constrained 24 h fitting model. **F** Correlation analysis between SYNE1 gene expression and ClockProCRC score based on RNA-seq data. **G** ROC curve indicating high discriminatory power of SYNE1 expression in differentiating CLOCK

knockdown samples. **H** ClockProCRC score increased significantly after SYNE1 interference, and melatonin could reverse this change. RNA-seq data showed that the ClockProCRC of si-SYNE1 group was significantly higher than that of si-NC group ($P < 0.05$), and melatonin combined intervention (si-SYNE1+Melatonin) could partially restore the rhythm stability ($P < 0.05$). **I** PCoA analysis based on transcriptome revealed that the global expression profiles of si-CLOCK and si-SYNE1 treatment groups were significantly different from those of the si-NC and si-NC+ melatonin group, and showed similar distribution trends. **J** Expression of E-cadherin, ZO-1, and N-cadherin in HCT116 and SW480 cells transfected with si-SYNE1 was detected by Western blotting. **K** HCT116 and SW480 cell lines showed reduced cell migration due to si-SYNE1 compared to the si-NC group in 0 h and 24 h wound healing assays. **L** Migration statistics of HCT116 and SW480 cell lines in wound healing assays. **M** Cell cloning experiments of HCT116 and SW480 cell lines, in which SYNE1 gene knockdown resulted in reduced cell proliferation compared to the NC group. **N** Proliferation statistics of HCT116 and SW480 cell cloning experiments.

exhibits strong robustness and universality in external validation datasets. It accurately quantifies the degree of circadian rhythm disruption and rhythm variations in CRC samples. Notably, in samples exposed to circadian rhythm disruption or under standard control conditions, the scoring system reveals rhythmic changes over time. Similarly, in experimental groups with circadian rhythm disruption, the scores were significantly higher than those in control groups.

More importantly, our findings demonstrate that the ClockProCRC score is significantly associated with clinical variables in CRC patients, such as survival prognosis, tumor staging, and treatment sensitivity. Specifically, higher ClockProCRC scores were found to correlate with increased tumor malignancy and poorer prognosis. Additionally, the ClockProCRC score exhibited a notable trend of change at various stages of CRC progression, from normal colorectal mucosa to adenoma, advanced atypical hyperplasia, and ultimately liver metastasis, suggesting that this scoring system can reflect the dynamic progression of CRC to some extent. A randomized clinical trial has shown that, compared to continuous infusion, the FOLFIRI regimen (which includes irinotecan, oxaliplatin, and 5-fluorouracil) with circadian rhythm modulation reduces tumor burden and drug toxicity in metastatic CRC patients, while also improving objective response rate, overall survival, and progression-free survival⁴⁶. These findings underscore the potential for circadian rhythm-based chronotherapy to gain greater attention in clinical treatment strategies. The results of a study analyzed by multi-omics data revealed the significance of biological clock genes (CLOCK, CRY1, PER3) in metastasis, prognosis, and especially in regulating the immune micro-environment through EMT in metastatic CRC, and will provide preliminary evidence for the use of cancer time therapy combined with immunotherapy in immunotherapy regimens for metastatic CRC⁴⁷.

The ClockProCRC scoring system was conceived not merely as a simple stratification tool, but to reflect the degree of circadian rhythm disruption and its molecular characteristics within patients. By integrating single-cell, multi-cohort transcriptomics and functional experiments, it reveals to some extent whether a patient's circadian rhythm state is disrupted, providing a potential basis for individualized implementation of chronotherapy. Based on ClockProCRC scores, patients are categorized into high-score groups (indicating significant circadian disruption) and low-score groups (with relatively preserved rhythms). High-score patients may exhibit greater sensitivity to rhythm-optimized therapies (e.g., chemotherapy administered during morning/evening windows), as their disrupted endogenous rhythms could enhance time-dependent drug metabolism and DNA repair processes⁴⁸. Additionally, the ClockProCRC score can serve as a companion biomarker to help determine optimal dosing windows (e.g., morning vs. evening) to maximize efficacy and minimize toxicity.

Furthermore, our study suggests that circadian dysregulation of the novel biomarker SYNE1 in epithelial cells may be associated with colorectal cancer progression and may contribute to metastasis-related phenotypes. SYNE1 expression not only exhibits significant circadian rhythmic changes under CRD conditions, but its levels are also significantly upregulated in

CRC tissues and the circadian rhythm disruption intervention group, correlating closely with poor prognosis in patients. More importantly, in vitro experiments, including transwell and scratch assays, demonstrated that knockdown of SYNE1 significantly inhibited the invasive and migratory capabilities of CRC cells. Similarly, colony formation assays showed that knockdown of SYNE1 reduced CRC cell proliferation. These external cell experiments further confirmed the critical role of SYNE1 in CRC progression, suggesting that it may serve as a potential therapeutic molecular target to inhibit CRC advancement.

Although this study provides new insights into the relationship between CRD and CRC, several issues remain to be addressed in future research. Firstly, our study primarily relies on transcriptomic data from public databases, lacking validation from large-scale prospective clinical cohorts. There is an urgent need for prospective, multicenter, large-sample external validation in clinical cohorts to further assess the clinical predictive accuracy and intervention feasibility of the ClockProCRC score system. Additionally, the specific application of the ClockProCRC scoring system in clinical practice and its integration with existing clinical risk assessment models require further validation and optimization. Secondly, while our in vitro cell experiments have functionally verified the role of the SYNE1 gene in CRC cell migration, invasion, and proliferation, further animal studies will help fully elucidate the bridging role of the SYNE1 gene between physiological circadian rhythms and the tumor microenvironment.

In conclusion, this study provides new insights into the molecular mechanisms of CRC by revealing the critical role of CRD in tumorigenesis and progression. We propose a circadian-based precision oncology framework and identify SYNE1 as a critical circadian-associated regulator in CRC, laying the groundwork for chronotherapy interventions aimed at improving clinical outcomes and quality of life for colorectal cancer patients.

Methods

Dataset source and collection

The public datasets used in this study, which include data from humans, primates, and mice, were sourced from the Gene Expression Omnibus (GEO) database hosted by the National Center for Biotechnology Information (NCBI). The acquisition and application of these datasets adhered to relevant guidelines and policies. A total of 3758 samples from 35 public datasets were included and re-analyzed (Supplemental Table 1). The publicly available CRC single-cell RNA sequencing datasets included GSE166555, GSE200997, and GSE196964. The Bulk RNA-seq and microarray chip datasets related to CRC included GSE41258, GSE44076, GSE68468, GSE62321, GSE140973, GSE17537, GSE39084, GSE103479, GSE77955, GSE81582, GSE117606, GSE37364, GSE26571, GSE186727, GSE71187, GSE49355, and GSE51244. The Bulk RNA-seq and microarray chip datasets related to circadian rhythm included GSE124870, GSE98965, GSE54650, GSE10644, GSE57830, GSE143524, GSE56931, GSE98564, GSE30411, GSE39445, GSE45818, GSE11923, GSE98565, GSE43073, and GSE35026. Single-cell RNA sequencing dataset GSE166555 was used to

study biological mechanisms and scoring systems related to circadian rhythms in CRC development. The criteria for inclusion of data sets are as follows: (a) Publicly accessible and reproducible: All datasets used in this study were sourced from the GEO database, featuring downloadable raw expression matrices or normalized matrices, accompanied by comprehensive platform annotation files to ensure reproducibility of data processing. (b) Sample type clarity: In the included dataset, sample types must be clearly labeled as primary CRC tissue and its adjacent normal mucosal tissue. Similarly, the criteria for exclusion of data sets are as follows: (a) Technical Incomparability: Data will be excluded if significant batch effects are present in bulk RNA-seq or microarray cohorts, and systematic bias persists after correction by methods such as sva, or if platform annotation information is missing, preventing gene annotation reproducibility. (b) Non-compliant rhythmic design: Time-series data that fail to meet study specifications, lack rhythmic condition annotations or time point records, or cannot be used for rhythmic validation will not be employed in the rhythmic performance assessment of the ClockProCRC score.

Single-cell RNA sequencing, bulk RNA sequencing, and microarray chip cohort processing

We performed quality control and data normalization on the single-cell dataset GSE166555 using the Seurat package^{49,50}. First, cells with more than 5% mitochondrial gene content, fewer than 500 genes, or more than 5500 genes were excluded to eliminate potential outliers and reduce interference from cytoplasmic effects in subsequent analyses. Next, the 'FindIntegrationAnchors' and 'IntegrateData' functions were applied to integrate cells from different samples. The vst function was used to identify the top 2000 highly variable genes, which were then normalized and used as input features for dimensionality reduction via principal component analysis. Subcluster-specific marker genes were identified using the 'FindAllMarkers' function, and Uniform Manifold Approximation and Projection (UMAP) analysis was performed using the 'RunUMAP' function. Finally, cell clustering and annotation were performed using the SingleR package^{51,52}, along with marker genes obtained from the literature for specific cell types, thereby providing a clear presentation of cell types and states across all clusters. For the processing of bulk RNA-seq and microarray data, each probe set was annotated using the official gene symbols from the Ensembl database along with the corresponding platform annotation information. Data normalization and batch effect correction were performed using the limma and sva packages, respectively^{53,54}. The normalization process involved scaling the gene expression data, enabling comparisons of gene expression levels across different samples on a uniform scale. Batch effect correction was conducted using a classical Bayesian algorithm, effectively removing systematic errors arising from experimental batch variations (e.g., differences in time or handling by different operators). This approach ensured that the gene expression data accurately reflected the biological differences between samples in different groups, providing reliable data support for subsequent analyses.

Expression of the circadian gene set in the colorectum

To establish a robust reference gene set for circadian rhythm disorders, this study referenced HOGERWERF et al.'s systematic transcriptomic analysis of mouse distal colon tissue under a 24-hour light-dark cycle. In that study, experimental mice underwent a two-week adaptation period to a 12-hour light/dark (LD) cycle, followed by sampling every 4 hours during the formal experiment (6 time points in total)²². Subsequently, researchers employed the maSigPro (Microarray Significant Profiles) algorithm for temporal modeling analysis of all genes, followed by False Discovery Rate (FDR) control using the Benjamini-Hochberg (B-H) method⁵⁵. Rhythmic genes exhibiting significant temporal trends and high reliability were selected using thresholds of FDR < 0.01 and goodness-of-fit $R^2 \geq 0.7$. The final set of 906 genes was categorized as circadian rhythm expression genes associated with multiple pathophysiological functions, including colonic motility, inflammatory regulation, and colorectal cancer development, forming the gene set for subsequent studies (Supplemental Table 2).

High-dimensional co-expression network analysis

Single-cell transcriptome aggregation produces a 'meta-cell' expression matrix. After cell annotation in the GSE166555 dataset, the resulting single-cell expression matrix was intersected with the circadian rhythm disruption gene set identified in previous analyses. This intersection allowed for the selection of circadian rhythm disruption genes expressed in colorectal cancer cells. Subsequently, a K-nearest neighbors (KNN) algorithm was applied to generate a meta-cell expression matrix. Meta-cells are defined as small clusters of transcriptionally similar cells that represent distinct cell states. The algorithm groups cells with similar gene expression patterns into meta-cells, ensuring that key sample information, such as disease status, is preserved for downstream analysis⁵⁶. Given the sparsity of single-cell gene expression matrices, which often contain a large proportion of zero-value entries, the sparsity is quantified as the proportion of zero entries relative to the total number of entries in a matrix containing N_g genes and N_c cells⁵⁷. A matrix is considered sparse if its sparsity exceeds 0.5.

$$sparsity = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_c} \begin{cases} 1 & \text{if } X_{ij} = 0 \\ 0 & \text{else} \end{cases}}{N_g \times N_c} \quad (1)$$

The inherent sparsity and noise of single-cell data often lead to spurious gene-gene correlations, complicating network analysis. Additionally, the correlation structures of single-cell data vary significantly across different subsets of cells, such as distinct cell types and states⁵⁸. To address these limitations, we first cluster highly similar cells into 'meta-cells' to reduce sparsity while preserving cellular heterogeneity.

We defined two sets: G for circadian-related genes (Supplemental Table 2) and C for unique cell barcodes, with sizes $|G| = N_g$ and $|C| = N_c$. To identify cells with similar circadian gene expression profiles, we applied the KNN algorithm to a reduced-dimensional representation D of the expression matrix X , generating N_c sets of k -nearest cells. Since overlap between these sets was expected, a parameter m was introduced to limit the extent of overlap. Cells were randomly selected from C , and their gene expression data in X were aggregated (via summation or averaging) with their k closest neighbors. Cells that had excessive overlap with previously selected ones were excluded to avoid redundancy in the meta-cell matrix. This process continued until no cells met the overlap condition, yielding a final meta-cell expression matrix (M). To ensure the process did not increase the sparsity of the data, we calculated the sparsity of both X and M at each step. The vector for the i -th row of M is represented as M_{i*} , and the vector for the i -th column is M_{*i} .

High-dimensional co-expression network analysis and construction of circadian rhythm module. Subsequently, we constructed a high-dimensional co-expression network based on the CRC meta-cell expression matrix. The gene-gene adjacency matrix was determined by calculating the pairwise correlations between genes in the meta-cell expression matrix (M). For any pair of genes (i, j), their gene expression vectors in M are denoted as $x_i = M_{i*}$, $x_j = M_{j*}$, respectively. The correlation between gene pairs is then defined as follows:

$$correlation_{i,j} = \frac{1 + cor(x_i, x_j)}{2} \quad (2)$$

The correlations were then weighted by setting a soft threshold, while the lowest soft power threshold with scale-free topological model fit ≥ 0.8 was selected to construct the topological overlap matrix (TOM), and the genes were grouped into modules using a dynamic tree-cutting algorithm based on the TOM network representation⁵⁹. This method hierarchically clusters genes based on their dissimilarity in the TOM, thus identifying co-expression modules composed of genes with high topological similarity⁶⁰.

Calculate the module characteristic gene. The module gene expression profile (ME) serves as a convenient metric summarizing the gene expression profile of an entire co-expression module. ME is computed by performing principal component analysis on subsets of the meta-cell expression matrix

corresponding to each module. The first principal component obtained from each principal component analysis is defined as the ME of the module, encapsulating the overall expression profile of that module for subsequent analysis. This approach allows for a precise focus on the circadian rhythm disruption genes within key modules that are closely involved in the progression of colorectal cancer, which will be prioritized for further investigation. Finally, the 'Eigengenes' function was used to quantitatively assess the correlation between co-expressed gene modules and circadian disruption, selecting modules significantly associated with circadian disruption for further analysis.

Development of ClockProCRC scoring system

To assess the level of endogenous CRD in individual cells or large sample cohorts, the CRDscore package was utilized, integrating genes identified as associated with circadian rhythm in colon tissues²². A ClockProCRC scoring system was developed, which can finely adjust for discrepancies based on the expression profiles of cells and genes⁶¹. The calculation process is as follows:

For a given normalized meta-cellular expression matrix C , the average expression E_i of gene i in N samples or cells is defined as:

$$E_i = \log_2 \left(\frac{\sum_j 10 \times (2^{C_{ij}} - 1)}{N} + 1 \right), j = 1, \dots, N; \text{scRNAseq} \quad (3)$$

Next, a random sampling strategy was employed to divide genes into 100 expression bins based on their average expression levels (E_i). The frequency of circadian rhythm disorder genes (CRGs) identified within each bin was calculated and designated as B_{CRGs} . Randomly selected feature genes were chosen from each bin according to the frequency of B_{CRGs} , ensuring the total number of selected feature genes (K) matched the number of CRGs. This random procedure was repeated 1000 times.

Finally, the circadian rhythm center gene expression matrix can be interpreted as data without excessive migration signals, which means:

$$Z_{ij} = C_{ij} - \frac{\sum_j C_{ij}}{N}, j = 1, \dots, N \quad (4)$$

where Z_{ij} is the central expression of gene i in cell or sample j . To quantify the abundance of CRGs, we created a random score S_{random} as the average of the $K \times 1000$ random features sampled above, and created the score S_{CRGs} as the average of the K CRGs using the central expression data for each sample or cell.

The characteristics of the ClockProCRC score are normal distributions or mixtures of normal distributions, and the calculation method is as follows:

$$\text{ClockProCRC} = S_{random} - S_{CRGs} \quad (5)$$

The 75% cutoff value based on the quartile was used as the threshold for CRD levels in single-cell RNA sequencing cohorts, while the median value was commonly used in Bulk RNA-seq cohorts.

Analysis of biological functional enrichment

We performed Gene Ontology (GO) enrichment analysis on the circadian rhythm disorder gene set defined in the GSE10644 dataset and the key module genes identified from the CRC single-cell cohort using the clusterProfiler package⁶². The analysis employs Fisher's exact test. P -values are corrected using the B-H method (FDR correction). This approach explores the functional characteristics and biological activities of these genes from a biological process perspective^{63–65}. visE analysis is a tool for visualizing gene set or functional pathway enrichment results, typically used in conjunction with functional enrichment analysis^{66,67}. It takes as input BH-corrected pathway results with $FDR < 0.05$ and constructs a network-like structure based on overlap rates to present findings from gene expression or pathway analysis. We employed gene set variation analysis (GSVA) to investigate

expression activity differences and biological significance of epithelial-mesenchymal transition (EMT) related pathways between primary CRC lesions and liver metastases. Using the classic EMT signature gene sets from the MSigDB database as the analysis target, we calculated enrichment scores for each EMT pathway in every sample via the GSVA algorithm to quantify their activation levels. GSVA was performed using the GSVA software package⁶⁸. Intergroup differences in EMT pathway enrichment scores between primary tumors and liver metastases were compared using the Wilcoxon signed-rank test, with P values corrected using the B-H method ($FDR < 0.05$ considered significant).

Detection of the biological clock rhythm of the ClockProCRC score

To evaluate the quantification scoring system we developed based on circadian rhythm disorder genes significantly expressed in colorectal cancer cells, we conducted assessments of its rhythmicity over time in multiple cohorts with circadian rhythm fluctuations. This monitoring was based on a likelihood-based method for detecting circadian rhythmicity in gene expression data and comparing patterns. The method includes: (I) detecting the circadian rhythmicity of genes; (II) comparing differences in the circadian rhythmic patterns of genes under two experimental conditions. The approach assumes that the relationship between gene expression levels and circadian time follows a sinusoidal curve model, quantifying the strength and dynamic changes of circadian rhythm characteristics through curve-fitting parameters such as amplitude, phase, and period⁶⁹. The corresponding calculation formula is as follows:

$$y_i = A \sin(\omega(t_i + \varphi)) + C + \varepsilon_i \quad (6)$$

Where y_i denotes the observed expression value of a gene at time t_i ; t_i represents the circadian time (e.g., hours since the onset of the subjective day); A is the amplitude of oscillation, reflecting the peak-to-trough difference; $\omega = 2\pi/T$ is the angular frequency, with $T = 24$ hours; φ indicates the phase, representing the timing of peak expression; C is the baseline expression level, corresponding to the average expression; and ε_i represents random noise, modeled as a Gaussian distribution.

Circadian rhythmicity detection: To assess whether a feature exhibited rhythmic behavior, we employed a likelihood ratio test (LRT) to compare two models. The null hypothesis (H_0): Expression remains constant over time ($A = 0$), indicating no circadian rhythm. The alternative hypothesis (H_A): Expression follows a sinusoidal pattern ($A \neq 0$), suggesting circadian rhythmicity. The likelihoods for the null ($L(H_0)$) and alternative ($L(H_A)$) models were estimated via nonlinear regression, and the LRT statistic was computed as follows:

$$\text{LRT} = -2(\ln \mathcal{L}(H_0) - \ln \mathcal{L}(H_A)) \quad (7)$$

Features with significant LRT values were categorized as rhythmic. The resulting rhythmicity P value therefore represents statistical evidence for 24 h rhythmicity and pseudo- R^2 was used as a goodness-of-fit measure reflecting the proportion of observed expression variability captured by the fitted sinusoidal model. Differences in basal level, amplitude and phase between experimental conditions were evaluated separately using differential circadian pattern analysis.

Assessment of cross-species rhythmic module concordance

To assess the conservation of ClockProCRC rhythmicity across species, we integrated data from mice, humans, and primates. Given the distinct diurnal activity patterns among species (e.g., nocturnal mice versus diurnal humans and primates), we employed a unified circadian time alignment strategy. First, we performed independent module construction and rhythm analysis within each species. Subsequently, we shifted the mouse Zeitgeber Time by 12 hours to align with the circadian rhythm phase of humans⁷⁰. Using the expression peaks of core module genes as anchors, we performed phase fine-tuning via cross-correlation methods⁷¹.

Analysis of the evaluation and clinical correlation of circadian rhythm disorder scores

To evaluate the clinical significance of ClockProCRC, we assessed its association with survival outcome using a restricted cubic spline (RCS) Cox regression model. ClockProCRC was modeled as a continuous variable, and the resulting spline function was visualized on the log hazard ratio [ln(HR)] scale relative to the reference value. Kaplan-Meier survival curves were further employed to validate their correlation with survival probability. Additionally, we performed correlation analyses between the ClockProCRC score and various clinical variables, including patient age, sex, tumor stage, genetic mutations, and comorbidities, using data from the GSE39084 and GSE103479 cohorts. At the molecular level, we explored the correlation of ClockProCRC scores with different clinical variables, investigating how the score varies across patient subgroups with distinct clinical characteristics. Furthermore, we validated the dynamic changes of the ClockProCRC score across different stages of CRC progression and treatment in four clinical cohorts, further examining its relationship with disease progression and treatment sensitivity.

Single-cell RNA data analysis

We explored the molecular phenotype of the ClockProCRC score using the Singscore algorithm⁶⁴. The core computation of the algorithm can be summarized in two steps: (1) calculating the gene ranks within each sample; (2) using these ranks to calculate a genomic score. The results yield two values: the score and the dispersion. The score represents the median expression of all genes in the genome relative to all measured genes. Dispersion measures the coordination of gene expression within the feature. We calculated the ratio of observed to expected cell counts (*Ro/e*) for each cell type in different groups to quantify the preferences of each cell type.

The Nonlinear Interactions and Communication Heterogeneity in Extracellular Signals (NICHEs) framework was applied to the annotated single-cell transcriptomic data to infer ligand-receptor signaling at single-cell resolution⁷². Briefly, normalized expression matrices and cell-type annotations were used as input to construct sender-receiver signaling maps across major cell populations. NICHEs-transformed signaling matrices were then used to compare intercellular communication patterns between ClockProCRC Intensify and Alleviate states, thereby enabling downstream visualization of signaling heterogeneity and network differences. Default ligand-receptor reference settings were used, and signaling outputs were summarized at the cell-type level for downstream heatmap and network analyses.

Construction of the circadian rhythm gene regulation network (GRN)

To comprehensively understand the key regulatory role of CRD genes in CRC initiation and progression, we constructed a regulatory network analysis based on CRD genes using the Dysregis software package⁷³. We utilized transcription factor binding site prediction tools and transcription factor motif data to assess the regulatory strength of each gene pair from the key modules of CRD genes identified in a CRC single-cell cohort. A gene regulatory network was then constructed. The network analysis algorithm was employed to evaluate the centrality of each gene within the network and quantify the regulatory strength between genes, selecting significantly dysregulated genes with a threshold of $P < 0.05$. Using a genetic algorithm, we evaluated the frequency of occurrence of each potential regulatory axis in individuals after 100 iterations as a quantitative indicator of robustness to identify key dysregulated axes promoting CRC development. This approach allowed for an in-depth exploration of the pathological processes and regulatory mechanisms of circadian rhythm genes in CRC progression. For the screened key gene dysregulation axis, GO enrichment analysis was conducted using the clusterProfiler package, significantly enriched items were corrected for multiple testing using the B-H method, and adjusted p value < 0.05 was used as the significance threshold. Gene Set Enrichment Analysis (GSEA) was performed using Hallmark gene sets

from the MSigDB database and is calculated using the fgsea algorithm⁷⁴. All genes were sorted according to SYNE1-related expression changes, and the enrichment score (ES) and its statistical significance were calculated. FDR-adjusted p value < 0.05 was also used as the screening criterion for significant pathways.

Cell lines

HCT116 cells and SW480 cells were both from the Shanghai Cell Bank of the Chinese Academy of Sciences (Shanghai, China). HCT116 and SW480 cells were cultured in DMEM/high sugar (Gibco, Carlsbad, CA, USA). All cells were cultured at 37 °C and 5% CO₂ supplemented with 10% fetal bovine serum (Clark Bioscience, Richmond, Virginia, USA).

Cell transfection with siRNA and plasmids

The siRNAs were mixed with Lipofectamine 3000 (Thermo Fisher Scientific, Carlsbad, CA, USA), incubated in DMEM, and then transduced into HCT116 and SW480 cells, and the solution was changed after 6–8 h. siRNA-NC (siN0000001–1–5) targeting CLOCK (si-CLOCK, stB0000881–1–2) and siRNAs-SYNE1 (stB0008815A–1–5 and stB0008815B–1–5) with scrambled siRNAs as negative controls were purchased from Guangzhou Ruibo Biological Co. To assess CLOCK and SYNE1 gene silencing under circadian restoration, cells were treated with 0.1 mM melatonin (Cat# HY-B0075, MedChemExpress) dissolved in DMSO (final DMSO $< 0.1\%$) for 24 h before RNA collection. Total RNA from all experimental conditions was submitted to Novogene (Beijing, China) for RNA sequencing. Gene silencing was achieved using short interfering RNAs (siRNAs) targeting CLOCK (si-CLOCK) or SYNE1 (si-SYNE1), with scrambled siRNAs as negative controls (si-NC), all synthesized by RiboBio (Guangzhou, China).

Bulk RNA-seq

Total RNA was isolated from CRC cell lines using the TRIzol reagent (Cat#15596026, Thermo Fisher, USA) following the manufacturer's protocol. The quality of RNA was evaluated with the Bioanalyzer 2100 system (Agilent Technologies, USA), and its purity was determined by measuring absorbance using a NanoDrop 2000 spectrophotometer (Thermo Fisher, USA). Only RNA samples with an OD260/280 ratio ≥ 1.9 and an RNA Integrity Number (RIN) ≥ 7.0 were included for downstream procedures. For mRNA enrichment, oligo(dT) magnetic beads were employed, followed by synthesis of double-stranded complementary DNA (cDNA) using Superscript II reverse transcriptase (Invitrogen, USA). Fragmentation of cDNA and subsequent library preparation were carried out using the NEBNext Ultra RNA Library Prep Kit (New England Biolabs, USA). Paired-end sequencing (150 bp) was performed on the Illumina NovaSeq 6000 platform by Novogene Co., Ltd (Beijing, China).

Total RNA isolation, reverse transcription, and quantitative real-time PCR (qRT-PCR)

Total RNA was isolated from cells and tissues using RNAiso Plus (Takara, Dalian, China). 1 µg of total RNA was reverse transcribed to cDNA using a reverse transcription kit manufactured by Takara, after the concentration and purity of the total RNA were determined using NanoDrop2000 (Thermo Fisher Scientific, Carlsbad, CA, USA). After identifying the concentration and purity of total RNA, 1 µg of total RNA was reverse transcribed to cDNA using the Reverse Transcription Kit from Takara. qRT-PCR was performed using the Hieff qPCR SYBR Green Master Mix kit (Yeasen, Shanghai, China) according to the manufacturer's instructions. Each assay was repeated three times under the following reaction conditions:

(1) Pre-denaturation: 95 °C for 5 min of continuous heating; (2) PCR reaction: 95 °C for 10 s, 60 °C for 30 s; (3) Drawing of lysis curves, 95 °C for 15 s, 60 °C for 60 s, 95 °C for 15 s. The test was terminated after 40 cycles. At the end of the reaction, the PCR products were collected and stored temporarily in a refrigerator at -20 °C for subsequent nucleic acid electrophoresis. The primer sequences were:

Gene	Forward (5' to 3'-)	Reverse (5' to 3'-)
SYNE1CLOCK	ACCTCCAATGGTGGTG- GACCGACGAGAACUUGG- CAUUGAATTH	CGTGCCAATGTTAGC- CACA UUCAAUGCCAAGUUCUC- GUCGTT
GAPDH	AACGGATTGTGGTCG- TATTGG	TTGATTTTGGAGG- GATCTCG

Relative quantification of RNA for the SYNE1 and CLOCK genes was calculated by the $2^{-\Delta\Delta Ct}$ method, and all data were analyzed and normalized to GAPDH.

Transwell assay and scratch assay

Transwell assays were performed to evaluate the migratory and invasive capacities of colorectal cancer cells. For migration assays, 5×10^5 modified HCT116 and SW480 cells were seeded into Transwell inserts with 8- μ m pore membranes (Corning, NY, USA). For invasion assays, inserts were pre-coated with Matrigel (Corning, NY, USA) before cell seeding. After 36–72 h incubation, non-migrated/non-invaded cells on the upper membrane surface were gently removed, and cells on the lower surface were fixed and stained with Giemsa stain (Solarbio, Beijing, China). For quantification, four random microscopic fields per insert were photographed and counted. Each experiment was performed in three independent biological replicates. Quantitative results are presented as mean $\pm$ SD of the averaged cell counts from the independent experiments.

For the scratch assay, after knocking down the expression of the SYNE1 gene, cells were digested with trypsin and counted. 1×10^5 cells/well were planted into 12-well plates, and after being adhered to the wall, serum-free medium was changed, and a 'ten' was scratched out with a white lance tip, and photographed and recorded as 0 h. The cells were photographed again after 24 h of incubation and observed for cell healing.

Western blot analysis

To verify whether the SYNE1 gene further promotes distant metastasis through impaired epithelial mucosal proteins and barrier function, given that epithelial barrier integrity and epithelial-mesenchymal transition represent the most direct phenotypes of this process, we designated N-cadherin, E-cadherin, and ZO-1 as protein validation markers. Total protein was obtained by lysing cells and centrifuging tissue samples on ice using RIPA buffer containing protease and phosphatase inhibitors, and homogenized using the BCA kit (Beyotime, Shanghai, China). Denaturation by heating. SDS-PAGE gels were assembled at appropriate percentages according to the expected protein size range. The gels were run at a constant voltage of 100 V until the proteins were adequately separated. The separated proteins were transferred from the gel to a PVDF membrane (Millipore, MA, USA) using a wet transfer system. After blocking with a fast blocking solution, the membrane was blocked with ZO-1 (1:5000) (Proteintech, Wuhan, China), E-cadherin (1:20000) (Proteintech, Wuhan, China), N-cadherin (1:5000) (Proteintech, Wuhan, China), with SYNE1 (1: 3000) (Abcam, Cambridge, UK) and GAPDH primary antibody (1:5000) (Proteintech, Wuhan, China) were hybridized overnight, washed with TBST, and then hybridized with the secondary antibody for 1 h at 37 °C, washed again with TBST and exposed. The membrane was incubated with the secondary antibody attached to HRP for 1 h in blocking buffer. Enhanced chemiluminescence reagent (Millipore, MA, USA) was used to visualize protein bands. Protein banding signals were captured using a chemiluminescent imaging system.

Cell cloning formation experiment

Cell clone formation assay to detect the proliferative ability of the SYNE1 gene on colorectal tumor cells in vitro. After cell collection, the cells were diluted into a low concentration of cell suspension by medium according to the number of cells, and an appropriate amount of cell suspension (1000 cells/per well) was aspirated and transferred to the culture wells of a 12-well plate, and fresh medium was added to reach a volume of 1.5 mL of medium

in each well. The plate was then placed in an incubator for 7–10 days, and the medium was aspirated as much as possible, washed with PBS solution, and fixed with 4% paraformaldehyde for 30 min. The cells were fixed with 4% paraformaldehyde for 30 min, then washed with PBS, and 500 μ L of crystal violet (concentration: 1%) was added to each well and stained at room temperature for 30 min. The cells were then aspirated as much as possible, washed twice with PBS, and inverted onto filter paper to dry. Finally, the 12-well plate was placed on top of a white A4 paper to capture images.

Statistical analysis

All data processing, statistical analysis, and graphing were performed in R software (version 4.3.3). Comparisons between groups of categorical variables were performed using the chi-square test. Comparisons between groups of continuous variables were conducted using the Wilcoxon rank sum test or t-test. For correlation analysis, Spearman's correlation analysis was used to assess the relationship between two categorical variables or between a categorical variable and a continuous variable. Pearson's correlation analysis was used to assess the relationship between two continuous variables. The Kruskal-Wallis test was employed for comparing multiple groups of continuous variables. All analyses involving multiple significance tests were FDR-corrected using the B-H method, as precious study. All statistical tests were two-sided, and $P < 0.05$ indicated statistically significant differences.

Data availability

All the data come from the GEO database hosted by the National Biotechnology Information Center (<http://www.ncbi.nlm.nih.gov/geo/>), and are processed and analyzed in strict accordance with relevant guidelines to ensure that the privacy and rights of any participant are not infringed.

Code availability

Essential original scripts for implementing the ClockProCRC system in this work are available through the GitHub website (<https://github.com/DrZoggy/ClockProCRC>).

Received: 2 May 2025; Accepted: 1 September 2026;

Published online: 16 September 2026

References

1. Bass, J. Interorgan rhythmicity as a feature of healthful metabolism. *Cell Metab.* **36**, 655–669 (2024).
2. Patke, A., Young, M. W. & Axelrod, S. Molecular mechanisms and physiological importance of circadian rhythms. *Nat. Rev. Mol. cell Biol.* **21**, 67–84 (2020).
3. Zhang, T. et al. Circadian immunity and vascular inflammation in cardiovascular disease chronotherapy windows: time for a re-assessment. **2**, 236–279 (2026).
4. Straif, K. et al. Carcinogenicity of shift-work, painting, and fire-fighting. *Lancet Oncol.* **8**, 1065–1066 (2007).
5. 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 (2011).
6. Richards, J. & Gumz, M. L. Advances in understanding the peripheral circadian clocks. *FASEB J.* **26**, 3602–3613 (2012).
7. Durgan, D. J. & Young, M. E. The cardiomyocyte circadian clock: emerging roles in health and disease. *Circ. Res.* **106**, 647–658 (2010).
8. Hou, Y. -c et al. Biodegradable Mg alloy modified with bioactive exosomes for cardiovascular stent application. *J. Magnes. Alloy.* **12**, 4988–5004 (2024).
9. Orlich, M. J. et al. Vegetarian dietary patterns and the risk of colorectal cancers. *JAMA Intern. Med.* **175**, 767–776 (2015).
10. 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 (2021).

11. Zhang, S. et al. Artificial intelligence hybrid survival assessment system for robot-assisted proctectomy: a retrospective cohort study. *JCO Precis. Oncol.* e2400089 (2024).
12. Lin, H. H. & Farkas, M. E. Altered circadian rhythms and breast cancer: from the human to the molecular level. *Front. Endocrinol.* **9**, 219 (2018).
13. Wendeu-Foyet, M. G. & Menegaux, F. Circadian disruption and prostate cancer risk: an updated review of epidemiological evidences. *Cancer Epidemiol. Biomarkers Prev.* **26**, 985–991 (2017).
14. Tahara, Y. & Shibata, S. Circadian rhythms of liver physiology and disease: experimental and clinical evidence. *Nat. Rev. Gastroenterol. Hepatol.* **13**, 217–226 (2016).
15. Papagiannakopoulos, T. et al. Circadian rhythm disruption promotes lung tumorigenesis. *Cell Metab.* **24**, 324–331 (2016).
16. Gallego, M. & Virshup, D. M. Post-translational modifications regulate the ticking of the circadian clock. *Nat. Rev. Mol. cell Biol.* **8**, 139–148 (2007).
17. de Assis, L. V. M. et al. Expression of the circadian clock gene BMAL1 positively correlates with antitumor immunity and patient survival in metastatic melanoma. *Front. Oncol.* **8**, 185 (2018).
18. Soliz-Rueda, J. R. et al. Gut microbiota and eating behaviour in circadian syndrome. *Trends Endocrinol. Metab.* **36**, 15–28 (2025).
19. Niu, Y. et al. Targeting the intestinal circadian clock by meal timing ameliorates gastrointestinal inflammation. *Cell. Mol. Immunol.* **21**, 842–855 (2024).
20. 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 (2022).
21. Innominato, P. F. et al. Circadian rhythm in rest and activity: a biological correlate of quality of life and a predictor of survival in patients with metastatic colorectal cancer. *Cancer Res.* **69**, 4700–4707 (2009).
22. Hoogerwerf, W. A. et al. Transcriptional profiling of mRNA expression in the mouse distal colon. *Gastroenterology* **135**, 2019–2029 (2008).
23. Huang, W., Ramsey, K. M., Marcheva, B. & Bass, J. Circadian rhythms, sleep, and metabolism. *J. Clin. Investig.* **121**, 2133–2141 (2011).
24. Li, X.-J. et al. Integrated-omics analysis defines subtypes of hepatocellular carcinoma based on circadian rhythm. *J. Integr. Med.* **23**, 445–456 (2025).
25. Dickerman, B. A. et al. Sleep disruption, chronotype, shift work, and prostate cancer risk and mortality: a 30-year prospective cohort study of Finnish twins. *Cancer Causes Control* **27**, 1361–1370 (2016).
26. Markt, S. C. et al. Sleep duration and disruption and prostate cancer risk: a 23-year prospective study. *Cancer Epidemiol. Biomarkers Prev.* **25**, 302–308 (2016).
27. Zhang, S. et al. An assessment system for clinical and biological interpretability in ulcerative colitis. *Aging* **16**, 3856–3879 (2024).
28. Deng, F. & Yang, K. Current status of research on the period family of clock genes in the occurrence and development of cancer. *J. Cancer* **10**, 1117–1123 (2019).
29. Xuan, W. et al. Circadian regulation of cancer cell and tumor microenvironment crosstalk. *Trends Cell Biol.* **31**, 940–950 (2021).
30. Pogue-Geile, K. L., Lyons-Weiler, J. & Whitcomb, D. C. Molecular overlap of fly circadian rhythms and human pancreatic cancer. *Cancer Lett.* **243**, 55–57 (2006).
31. Gery, S. et al. Transcription profiling of C/EBP targets identifies Per2 as a gene implicated in myeloid leukemia. *Blood* **106**, 2827–2836 (2005).
32. Lane, J. M. et al. Genetics of circadian rhythms and sleep in human health and disease. *Nat. Rev. Genet.* **24**, 4–20 (2023).
33. Diamantopoulou, Z., Gvozdenovic, A. & Aceto, N. A new time dimension in the fight against metastasis. *Trends Cell Biol.* **33**, 736–748 (2023).
34. Kondratov, R. V., Shamanna, R. K., Kondratova, A. A., Gorbacheva, V. Y. & Antoch, M. P. Dual role of the CLOCK/BMAL1 circadian complex in transcriptional regulation. *FASEB J.* **20**, 530–532 (2006).
35. Kwak, Y. et al. Cyclin-dependent kinase 5 (Cdk5) regulates the function of CLOCK protein by direct phosphorylation. *J. Biol. Chem.* **288**, 36878–36889 (2013).
36. Zhao, X. et al. Circadian amplitude regulation via FBXW7-targeted REV-ERBa degradation. *Cell* **165**, 1644–1657 (2016).
37. Sun, Y. et al. Lactobacillus intestinalis facilitates tumor-derived CCL5 to recruit dendritic cell and suppress colorectal tumorigenesis. *Gut Microbes* **17**, 2449111 (2025).
38. Guodong, W. et al. Fecal occult blood affects intestinal microbial community structure in colorectal cancer. *BMC Microbiol.* **25**, 34 (2025).
39. Gery, S. et al. The circadian gene per1 plays an important role in cell growth and DNA damage control in human cancer cells. *Mol. Cell* **22**, 375–382 (2006).
40. Wang, Y. et al. Upregulation of circadian gene ‘hClock’ contribution to metastasis of colorectal cancer. *Int. J. Oncol.* **50**, 2191–2199 (2017).
41. Zhang, G. et al. AI hybrid survival assessment for advanced heart failure patients with renal dysfunction. *Nat. Commun.* **15**, 6756 (2024).
42. Zhang, G., Zhang, S., Zhang, H., Wu, R. & Zheng, H. Reply to: Addressing bias in feature importance: a hybrid approach for risk prediction in prognostic survival models. *JCO Precis. Oncol.* e2400875 (2025).
43. Xu, S. et al. Plasma olink proteomics reveals novel biomarkers for prediction and diagnosis in dilated cardiomyopathy with heart failure. *J. Proteome Res.* **23**, 4139–4150 (2024).
44. Li, Z. et al. Advancing predictive markers in lung adenocarcinoma: a machine learning-based immunotherapy prognostic prediction signature. *Environ. Toxicol.* **39**, 4581–4593 (2024).
45. Ubaid, S., Kushwaha, R., Kashif, M. & Singh, V. Comprehensive analysis of oncogenic determinants across tumor types via multi-omics integration. *Cancer Genet.* **298–299**, 44–62 (2025).
46. Jiang, M. et al. EMT and cancer stem cells: drivers of therapy resistance and promising therapeutic targets. *Drug Resist. Updat.* **83**, 101276 (2025).
47. Zhang, L. et al. The genomic and transcriptomic landscapes of clock genes reveal the significance of circadian rhythm in the progression and immune microenvironment of metastatic colorectal cancer. *Clin. Transl. Med.* **12**, e755 (2022).
48. Mormont, M. C. et al. Marked 24-h rest/activity rhythms are associated with better quality of life, better response, and longer survival in patients with metastatic colorectal cancer and good performance status. *Clin. Cancer Res.* **6**, 3038–3045 (2000).
49. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e3529 (2021).
50. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
51. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
52. Zheng, X. et al. Single-cell transcriptomic profiling unravels the adenoma-initiation role of protein tyrosine kinases during colorectal tumorigenesis. *Signal Transduct. Target. Ther.* **7**, 60 (2022).
53. Xie, X. et al. Effects of COL1A1 and SYTL2 on inflammatory cell infiltration and poor extracellular matrix remodeling of the vascular wall in thoracic aortic aneurysm. *Chin. Med. J.* **137**, 1105–1114 (2024).
54. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
55. Conesa, A., Nueda, M. J., Ferrer, A. & Talón, M. maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics* **22**, 1096–1102 (2006).

56. Komarov, I., Dashti, A. & D'Souza, R. M. Fast k-NNG construction with GPU-based quick multi-select. *PLoS ONE* **9**, e92409 (2014).
 57. Morabito, S. et al. Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat. Genet.* **53**, 1143–1155 (2021).
 58. Jiang, R., Sun, T., Song, D. & Li, J. J. Statistics or biology: the zero-inflation controversy about scRNA-seq data. *Genome Biol.* **23**, 31 (2022).
 59. Horvath, S. & Dong, J. Geometric interpretation of gene coexpression network analysis. *PLoS Comput Biol.* **4**, e1000117 (2008).
 60. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720 (2008).
 61. Jerby-Arnon, L. et al. A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. *Cell* **175**, 984–997.e924 (2018).
 62. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
 63. Zhang, G. et al. Smooth muscle cell fate decisions decipher a high-resolution heterogeneity within atherosclerosis molecular subtypes. *J. Transl. Med.* **20**, 568 (2022).
 64. Zhang, H. et al. Optimized dynamic network biomarker deciphers a high-resolution heterogeneity within thyroid cancer molecular subtypes. **1**, 10–31 (2025).
 65. Zhang, G. et al. Uncovering the genetic links of SARS-CoV-2 infections on heart failure co-morbidity by a systems biology approach. *ESC Heart Fail.* **9**, 2937–2954 (2022).
 66. Bhuva, D. D. et al. vissE: a versatile tool to identify and visualise higher-order molecular phenotypes from functional enrichment analysis. *BMC Bioinform.* **25**, 64 (2024).
 67. Zhang, G. et al. Atherosclerotic plaque vulnerability quantification system for clinical and biological interpretability. *iScience* **26**, 107587 (2023).
 68. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* **14**, 7 (2013).
 69. Ding, H. et al. Likelihood-based tests for detecting circadian rhythmicity and differential circadian patterns in transcriptomic applications. *Brief. Bioinform.* **22**, (2021).
 70. Petersen, K. A. et al. Comparative rhythmic transcriptome profiling of human and mouse striatal subregions. *Neuropsychopharmacology* **49**, 796–805 (2024).
 71. Möller-Levet, C. S., Laing, E. E., Archer, S. N. & Dijk, D. J. Diurnal and circadian rhythmicity of the human blood transcriptome overlaps with organ- and tissue-specific expression of a non-human primate. *BMC Biol.* **20**, 63 (2022).
 72. Raredon, M. S. B. et al. Comprehensive visualization of cell-cell interactions in single-cell and spatial transcriptomics with NICHES. *Bioinformatics* **39** (2023).
 73. Li, Q. et al. DysRegSig: an R package for identifying gene dysregulations and building mechanistic signatures in cancer. *Bioinformatics* **37**, 429–430 (2021).
 74. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA.* **102**, 15545–15550 (2005).
- Construction Project (Approval No. TJYXZDXK-044A), Hospital Management Research Project of Tianjin Hospital Association (Approval No. 2019ZZ07) and Tianjin Health Technology Project (Grant number: TJWJ2023QN040), Beijing-Tianjin-Hebei Basic Research Cooperation Project (Grant numbers: 23JCZXJC00390 and J230038) and Tianjin Health Technology Project (Grant number: TJWJ2021QN021), Zhengzhou University Young Student Basic Research Projects for PhD students (ZDBJ2026009). This work was also supported by the UGC Research Matching Grant (Grant number: RMGS240019) (S.Y.T.). We are grateful to the editors at NPJ Precision Oncology for their constructive comments and suggestions. Special thanks to every member of our team for their unwavering support, continuous companionship, and excellent work.

Author contributions

Study concept and design: S.Z., S.C. and N.H. Draft of the manuscript: S.Z., P.Z., Z.Z. and S.G. Acquisition and analysis of the data: Q.G., X.H., S.C. and H.W. Draw figures and visualization: H.Z., C.Y. and R.W. Writing—revision, supervision, resources, and funding acquisition: S.C., C.Z., S.Y.T. and G.Z. All the authors read and approved the final version of the manuscript.

Competing interests

All authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. S.C. is the founder and a shareholder of PyraGeo Biosciences Co., Ltd. The company had no role in the funding, study design, data collection, data analysis, interpretation of the results, or preparation of the manuscript. The authors declare no other competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41698-026-01699-1>.

Correspondence and requests for materials should be addressed to Chunze Zhang, Shing Yau Tam or Ge Zhang.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.

Acknowledgements

This research was supported by the Key research Project of Tianjin Science and Technology Support Program (Approval No. 19YFZCSY00420), Tianjin Natural Science Foundation (Grant numbers: 21JCZDJC00060, 21JCYBJC00180 and 21JCYBJC00340), Tianjin Key Medical Discipline

¹School of Medicine, Nankai University, Tianjin, China. ²Department of Colorectal Surgery, Tianjin Union Medical Center and the First Affiliated Hospital of Nankai University, Tianjin, China. ³Department of Thoracic Surgery, Peking University Shenzhen Hospital, Shenzhen, China. ⁴Cancer Research Laboratory, Department of Research and Development, PyraGeo Biosciences Co., Ltd., Changsha, China. ⁵The Second Medical Center of Chinese PLA General Hospital, Beijing, China. ⁶Zhengzhou University, Zhengzhou, China. ⁷Department of Lung Cancer, Tianjin Lung Cancer Center, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China. ⁸Department of Plastic and Reconstructive Surgery, Shanghai Ninth People's Hospital, Shanghai JiaoTong University School of Medicine, Shanghai, China. ⁹Department of Medical Oncology, Sun Yat-sen University Cancer Center, Guangzhou, China. ¹⁰Faculty of Dentistry, University of Debrecen, Debrecen, Hungary. ¹¹School of Medical and Health Sciences, Tung Wah College, Hong Kong, China. ¹²Present address: Division of Biomedical and Life Sciences, Faculty of Health and Medicine, Lancaster University, Lancaster, UK. ¹³These authors contributed equally: Shiqian Zhang, Shanshan Cai, Naijing Hou.

✉ e-mail: chunze.zhang@nankai.edu.cn; marcotam@twc.edu.hk; zg381492242@163.com