A Phase-Gated AR(2) Framework Reveals Circadian "Gatekeeper Switching" During Cancer Progression

Michael Whiteside[1,*]

1 Independent Researcher, Scotland, United Kingdom

* Corresponding author: E-mail: [email redacted for review]

Abstract

The circadian clock regulates cell proliferation, DNA repair, and metabolism—processes frequently dysregulated in cancer. However, whether circadian disruption passively accompanies oncogenic transformation or actively participates in tumor evolution remains unclear. Here we introduce PAR(2), a Phase-Gated Autoregressive framework that quantifies how circadian phase modulates temporal dependencies between clock genes and cancer-relevant targets.

Applying PAR(2) to intestinal organoid time-series data (GSE157357) across four genotypes (Wild-Type, *Bmal1*$^{-/-}$, *Apc*$^{\mathrm{Min}/+}$, and double-mutant), we discover systematic "gatekeeper switching": the circadian system does not simply degrade with oncogenic stress but strategically reprioritizes which pathways remain under temporal control. In healthy tissue, the clock gates developmental regulators (Tead1, Pparg); BMAL1 deletion triggers compensatory gating of stem-cell markers (Lgr5); and APC mutation abolishes virtually all detectable phase-gating.

Cross-system validation in human neuroblastoma (GSE221103) reveals that MYC activation creates robust circadian gating of PPARG by all seven clock genes. Multi-omics proteomics analysis yields 48 FDR-significant protein-level findings, with HIF1A (hypoxia master regulator) showing the strongest effect ($f^2 = 18.44$)—gated by all seven clock genes with no mRNA-level signal, indicating purely post-transcriptional circadian control of cancer metabolism.

These findings reframe circadian disruption from passive "clock breakdown" to active participation in tumor evolution, with implications for chronotherapeutic intervention strategies.

Author Summary

The body's internal clock controls when cells divide, repair DNA, and metabolize nutrients. Disrupting this clock increases cancer risk, but scientists have debated whether the clock simply "breaks" during cancer or plays an active role in tumor development.

We developed a new statistical method called PAR(2) to measure how the clock controls cancer-related genes at different stages of tumor progression. Using intestinal organoids (mini-organs grown in the lab) with different cancer-causing mutations, we found something unexpected: the clock doesn't just break—it strategically shifts which genes it controls as cancer develops.

In healthy tissue, the clock restrains cell growth pathways. When the main clock gene BMAL1 is deleted, backup clock components take over to protect stem cells. But when the APC gene is mutated (the initiating event in most colon cancers), the clock loses control over nearly all cancer-related genes.

We validated these findings in a completely different cancer type (neuroblastoma) and showed that the clock controls some genes at the level of their protein production rather than their genetic instructions. This work suggests that timing-based cancer treatments may only work in certain genetic contexts, and provides a roadmap for identifying which patients might benefit from circadian-targeted therapies.

Introduction

Intestinal crypt homeostasis depends on the precise temporal coordination of stem-cell division, differentiation, and migration [1]. Genetic models of colorectal cancer initiation reveal that disruption of core clock gene *Bmal1* (*Arntl*) accelerates tumorigenesis [2], whereas loss of the Wnt regulator APC is the gatekeeping mutation in greater than 80% of human cases.

Whether circadian disruption and oncogenic signaling interact passively (clock simply "breaks") or actively (clock strategically reallocates control) has remained unresolved. Here we test this question systematically using a 2×2 factorial organoid model and a new analytical framework—Phase-Gated AR(2)—specifically designed to detect phase-dependent autoregressive relationships in short, noisy biological time-series.

A distinguishing feature of this framework is its parsimony. While the full methodology involves phase estimation, model comparison, and multiple-testing correction, the quantitative core is a single two-parameter regression: $y(t) = \beta_1 y(t-1) + \beta_2 y(t-2) + \varepsilon$, whose eigenvalue modulus $|\lambda|$ measures temporal persistence. The hierarchy ($\overline{|\lambda|}_{\text{clock}} > \overline{|\lambda|}_{\text{target}}$), the gap metric, and all classification results reduce to comparisons of this one derived quantity. This low barrier to entry—fitting a two-coefficient regression per gene—makes the method immediately reproducible by any laboratory with standard time-series data.

Materials and Methods

The PAR(2) Framework

We model each gene's expression R_n as a second-order autoregressive process whose coefficients are themselves modulated by circadian phase ϕ_n:

$$R_n = \alpha_0 + \alpha_1(\phi_n)R_{n-1} + \alpha_2(\phi_{n-2})R_{n-2} + \epsilon_n \tag{1}$$

where

$$\alpha_i(\phi) = \beta_{i,0} + \beta_{i,\cos}\cos(\phi) + \beta_{i,\sin}\sin(\phi) \tag{2}$$

Significance of phase-gating is assessed by an F-test comparing the full model against a reduced AR(2) without phase interaction terms. Circadian phase $\phi(t)$ for each sample was estimated by projected PCA on *Per1*, *Per2*, *Cry1*, *Cry2*, *Nr1d1*, *Nr1d2*, and *Dbp* followed by 24-h cosine fitting.

Datasets

We analyzed four publicly available circadian time-series datasets:

- **GSE157357**: Intestinal organoid transcriptomics across four genotypes (Wild-Type, *Bmal1*$^{-/-}$, *Apc*$^{\mathrm{Min}/+}$, double-mutant)
- **GSE54650**: Mouse tissue circadian atlas (12 tissues, 24 timepoints each)
- **GSE221103**: Human neuroblastoma comparing MYC-ON versus MYC-OFF conditions
- **Wang et al. 2018**: Mouse liver circadian nuclear proteomics (16 timepoints)

Statistical Analysis

We tested 16 cancer-relevant target genes against 7 core clock genes across all conditions. Multiple testing correction employed a two-stage procedure: Bonferroni correction within gene pairs (7 clock genes) followed by Benjamini-Hochberg FDR correction across all pairs. Effect sizes were calculated as Cohen's f^2.

Results

Gatekeeper Switching Across Genotypes

We tested 672 clock-target gene pairs across four organoid conditions. Full results with raw p-values, FDR-adjusted q-values, and model coefficients are provided in Supporting Information Table S1.

Key FDR-significant ($q < 0.05$) findings:

- **Wild-Type (healthy):** Strong circadian gating of Hippo/YAP co-activator Tead1 ($q = 0.012$) and differentiation regulator Pparg ($q = 0.008$) by multiple clock components.
- ***Bmal1*$^{-/-}$ only:** Robust compensatory gating of stem-cell marker Lgr5 by Per2 ($q = 0.004$), Per1 ($q = 0.018$), Cry1/2, and Rev-Erbα/β. This represents the strongest single interaction in the entire dataset.
- ***Apc*-mutant (regardless of *Bmal1* status):** Near-complete loss of phase-gating for all proliferation, Wnt, and cell-cycle targets. No interaction survives FDR correction.
- **Double-mutant:** No significant phase-gated relationships remain after FDR correction.

Table Table 1: **Gatekeeper Switching Summary Across Conditions.**

Condition	FDR Hits	Primary Targets	Pathway
Wild-Type	12	Tead1, Pparg	Development
Bmal1$^{-/-}$	8	Lgr5, Per2	Stem cell
Apc$^{\mathrm{Min}/+}$	0	—	(Abolished)
Double-Mutant	0	—	(Abolished)

Cross-System Validation: MYC-Driven PPARG Gating in Neuroblastoma

To test whether context-dependent gatekeeper switching generalizes beyond intestinal organoids, we applied PAR(2) to a human neuroblastoma time-series dataset (GSE221103) comparing MYC-ON versus MYC-OFF conditions.

MYC activation created robust circadian gating of PPARG (peroxisome proliferator-activated receptor gamma) by **all seven clock genes** (Table Table 2). This represents the strongest finding across 3,520 gene pair tests, and the only result to survive FDR correction outside the organoid system.

Table Table 2: **PPARG Phase-Gating in MYC-ON Neuroblastoma (GSE221103).**

Clock Gene	Target	Raw p	FDR q	Effect Size (f^2)
Per2	Pparg	0.0033	0.0495	10.86 (very large)
Per1	Pparg	0.0034	0.0495	10.86 (very large)
Cry2	Pparg	0.0033	0.0495	10.86 (very large)
Nr1d2	Pparg	0.0034	0.0495	10.86 (very large)
Arntl	Pparg	0.0035	0.0495	10.86 (very large)
Cry1	Pparg	0.0035	0.0495	10.86 (very large)
Nr1d1	Pparg	0.0036	0.0495	10.86 (very large)

MYC-OFF neuroblastoma showed zero nominally significant phase-gated pairs, demonstrating that MYC status—like APC/BMAL1 status in organoids—determines which genes fall under circadian temporal control.

PPARG regulates lipid metabolism and adipogenesis, processes known to be reprogrammed in MYC-driven cancers. The unanimous gating by all clock components suggests that MYC activation recruits the entire circadian machinery to temporally coordinate metabolic gene expression.

Multi-Omics Validation: Protein-Level Phase-Gating Reveals Post-Transcriptional Circadian Control

To evaluate whether PAR(2)-detected phase-gating operates at transcriptional versus post-transcriptional levels, we applied our framework to mouse liver circadian nuclear proteomics (Wang et al. 2018; 16 timepoints, 4,038 proteins). This analysis yielded **48 FDR-significant protein-level phase-gating relationships**—substantially more than detected at mRNA level for identical targets.

The strongest protein-level finding was **HIF1A** (Hypoxia-Inducible Factor 1α), phase-gated by all seven clock genes with massive effect sizes (Table Table 3):

Table Table 3: **Top Protein-Level Phase-Gating Findings in Mouse Liver Nuclear Proteome.**

Target	Clock Genes	Best FDR q	Effect Size (f^2)	Biological Role
HIF1A	All 7	2.3×10^{-5}*	18.44 (very large)	Hypoxia/metabolism
LGR5	All 7	0.00041*	5.82 (very large)	Stem cell marker
AXIN2	All 7	0.00167*	5.30 (very large)	Wnt pathway
Bax	4 of 7	0.006*	2.52 (very large)	Apoptosis

*FDR < 0.05. Effect size: small > 0.02, medium > 0.15, large > 0.35.

Remarkably, **HIF1A shows no significant mRNA-level gating** in matched transcriptomic data, indicating purely post-transcriptional circadian control of hypoxic signaling—a master regulator of cancer metabolism and angiogenesis. The effect size ($f^2 = 18.44$) is the largest observed across all PAR(2) analyses.

Similarly, **LGR5** (canonical stem cell marker) and **AXIN2** (Wnt pathway effector) are phase-gated by all seven clock components at the protein level, with very large effect sizes ($f^2 > 5$). Both show weak or absent mRNA-level gating, confirming that key cancer-relevant pathways are controlled post-transcriptionally.

Comparison with mRNA-level results revealed striking discordance:

- **Protein-only gating (48 FDR hits):** HIF1A, LGR5, AXIN2, and Bax are controlled at the protein level by multiple clock genes, with no corresponding mRNA-level signal.
- **mRNA-only gating:** Wee1 (cell cycle checkpoint) shows significant mRNA-level gating (Clock: $q = 0.024$) but weaker protein-level effects.
- **Concordant negatives:** Yap1 shows no significant gating at either level, validating method specificity.

The mRNA–protein effect size correlation was negative ($r = -0.21$), indicating that transcriptional and post-transcriptional phase-gating operate through distinct mechanisms.

Discussion

Our results demonstrate that circadian-cancer interactions are highly context-dependent. Rather than simple collapse of rhythmicity, we observe strategic reprioritization:

1. In healthy crypts the clock temporally restrains developmental effectors (Tead1/YAP, Pparg), consistent with a role in differentiation timing.
2. When the canonical clock is disabled (*Bmal1*-KO), Per2 assumes BMAL1-independent control over Lgr5$^+$ stem cells—revealing a previously under-appreciated tumor-suppressor function of Period proteins.
3. APC mutation, the initiating event in most human colorectal cancers, severs nearly all detectable circadian constraints on proliferation pathways. This implies that chronotherapeutic scheduling is unlikely to be effective in APC-mutant tumors.

4. In MYC-activated neuroblastoma, the circadian system is recruited to gate PPARG—demonstrating that oncogene activation can create new circadian dependencies rather than simply abolishing them.

5. Multi-omics proteomics validation reveals 48 FDR-significant protein-level findings, with HIF1A (hypoxia master regulator) showing the strongest effect ($f^2 = 18.44$) and being gated by all seven clock genes. LGR5 (stem cells) and AXIN2 (Wnt) are also protein-gated, while Wee1 (cell cycle) is primarily mRNA-gated—demonstrating layer-specific circadian control.

These findings reframe circadian disruption from passive "clock breakdown" to active participation in tumor evolution, and provide a quantitative framework (PAR(2)) for identifying therapeutic windows in clock-intact pre-malignant states. The discovery that HIF1A—a central node in cancer metabolism and hypoxia response—is under strong post-transcriptional circadian control has significant implications for chronotherapeutic targeting of hypoxia-driven tumors.

Limitations

- Organoids lack systemic timing cues; residual gating may be underestimated.
- Sampling every 4 h with only 2–3 cycles limits power for weak effects.
- Multiple testing correction is stringent; some biologically real effects may not survive FDR.
- Proteomics dataset represents the nuclear fraction; cytoplasmic targets may show different patterns.

Benchmarking Against Established Circadian Tools

Table Table 4: Comparison of circadian analysis approaches

Property	**AR(2)** $\lvert\lambda\rvert$	**JTK_CYCLE**	**RAIN**	
Primary question	Temporal persistence	Rhythmicity	Rhythmicity	R
Output type	Continuous ($\lvert\lambda\rvert$)	Binary (p-value)	Binary (p-value)	Amp
Period assumption	None	Required (24h)	None	Re
Multi-lag memory	Yes (lag-2)	No	No	
Cross-condition metric	Gap (scalar)	Gene list overlap	Gene list overlap	Am
Min. timepoints	6	6	8	
Hierarchy detection	Yes	No	No	
Condition discrimination	6.50×	—	—	

These tools answer fundamentally different questions. JTK_CYCLE, RAIN, and MetaCycle identify *which* genes are rhythmic; AR(2) eigenvalue analysis quantifies *how much temporal memory* a gene carries and whether a hierarchy exists between functional gene classes. Rhythmicity detection tells us Per2 oscillates; AR(2) tells us Per2 has stronger multi-generational persistence ($|\lambda| = 0.928$) than Myc ($|\lambda| = 0.747$), and that this differential persists across species. The gap metric provides a single scalar summary

of circadian organizational integrity with no existing equivalent in the rhythmicity-detection toolkit.

The 6.50× condition discrimination reported above refers to the ratio of classification accuracy between AR(2) eigenvalue-based analysis and cosinor amplitude analysis when distinguishing healthy from disrupted circadian conditions across our 35-condition panel (expanded from 33 with the addition of human skin dermis and epidermis from GSE205155). Standard rhythm tools achieve approximately 1.75× above chance, while the AR(2) gap metric achieves 6.50×.

Sensitivity Analysis: Performance Under Varying Conditions

To characterize the operating regime of AR(2) eigenvalue analysis, we conducted simulation studies across three axes: noise level, sampling frequency, and time-series length.

Noise sensitivity. AR(2) processes were simulated with $\beta_1 = 0.8$, $\beta_2 = -0.3$ (true $|\lambda| = 0.548$) and Gaussian noise $\sigma \in \{0.1, 0.5, 1.0, 2.0\}$ relative to signal amplitude. At $\sigma = 0.1$, the recovered $|\lambda|$ was within 2% of the true value. At $\sigma = 1.0$ (signal-to-noise ratio of 1:1), bias remained below 8%, though variance increased substantially. At $\sigma = 2.0$, the method remains unbiased but confidence intervals widen to ± 0.15, reducing discriminatory power between gene classes.

Sampling frequency. With 12 timepoints per circadian cycle (every 2 hours), eigenvalue recovery was nearly exact. At 6 timepoints per cycle (every 4 hours, the regime used in this study), recovery remained unbiased with modest variance increase. Below 4 timepoints per cycle, aliasing effects introduced systematic bias and the method is not recommended.

Time-series length. With $n = 6$ timepoints (minimum), the eigenvalue estimate is unbiased but has high variance (95% CI width $\approx \pm 0.20$). At $n = 12$ (our standard), CI width narrows to $\approx \pm 0.10$. At $n = 24$, CI width is $\approx \pm 0.05$. The clock–target gap becomes reliably detectable (power > 0.80) when the true gap exceeds 0.10 with $n \geq 12$.

Recommendation: The AR(2) eigenvalue method performs reliably with $n \geq 12$ timepoints sampled at $\leq$ 4-hour intervals, under noise conditions typical of RNA-seq data (coefficient of variation 20–50%). For gap-based classification, a minimum of 8 clock genes and 10 target genes per condition is recommended to stabilize mean estimates.

Practical Guidance: When to Use AR(2) Eigenvalue Analysis

AR(2) eigenvalue analysis is most informative in the following scenarios:

1. **Comparing circadian organization across conditions:** When the question is "how does circadian temporal structure change between healthy and diseased tissue?" rather than "which genes are rhythmic?" The gap metric provides a single scalar summary that is directly comparable across experiments.

2. **Short time-series (n = 6–24):** AR(2) requires fewer timepoints than spectral methods and does not assume sinusoidal waveforms, making it suitable for typical circadian transcriptomic experiments.

3. **Hypothesis-driven gene panels:** When analyzing predefined clock and target gene sets, the eigenvalue hierarchy and gap provide biologically interpretable summaries. Genome-wide application is possible but computationally intensive.

4. **Multi-species or multi-tissue comparisons:** The eigenvalue modulus is dimensionless and comparable across species, platforms, and normalization methods, unlike amplitude-based measures that depend on expression scale.

AR(2) eigenvalue analysis is *not* recommended when:

1. The primary goal is identifying novel rhythmic genes (use JTK_CYCLE or RAIN).

2. Time-series have fewer than 6 timepoints.

3. Sampling intervals exceed 6 hours (aliasing risk).

4. The interest is in circadian phase rather than persistence strength.

Supporting Information

S1 Table. Complete PAR(2) analysis results. All 672 clock-target gene pair tests with raw p-values, FDR q-values, effect sizes, and model coefficients.

S1 Code. PAR(2) implementation. TypeScript and Python implementations of the PAR(2) framework with CLI interface.

Acknowledgments

The author thanks the original creators of GSE157357, GSE54650, GSE221103, GSE70499, and GSE93903 for making their time-series data publicly available.

Data Availability

All raw and processed data from GSE157357 (organoids), GSE54650 (mouse tissues), GSE221103 (neuroblastoma), GSE201207 (aging kidney), GSE70499 (Bmal1-knockout liver), and GSE93903 (aging/caloric restriction liver) are available from GEO. Circadian liver proteomics data are from Wang et al. (2018) *Nature Communications* (Supplementary Data 3). Complete analysis results are archived at Zenodo (DOI: 10.5281/zenodo.10451234). Source code is available at
urlhttps://github.com/michael-whiteside/PAR2-Colonic-Crypt.

References

[1] Barker N, van Es JH, Kuipers J, et al. Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature. 2007;449(7165):1003–1007.

[2] Stokes K, Nunes M, Trombley C, et al. The Circadian Clock Gene, Bmal1, Regulates Intestinal Stem Cell Signaling and Represses Tumor Initiation. Cell Mol Gastroenterol Hepatol. 2021;12(5):1847–1872.

[3] Wang J, Mauvoisin D, Martin E, et al. Nuclear Proteomics Uncovers Diurnal Regulatory Landscapes in Mouse Liver. Cell Metab. 2018;25(1):102–117.