AR(2) Eigenvalue Modulus as a Measure of Temporal Persistence in Gene Expression: Circadian Hierarchy Emerges from Two Coefficients

Michael Whiteside[1,*]

[1]Independent Researcher, United Kingdom

[*]Corresponding author: mickwh@msn.com
ORCID: 0009-0000-0643-5791

February 2026

Abstract

Background: The mammalian circadian clock organizes gene expression into a hierarchical architecture, but no single quantitative metric has previously captured this hierarchy from expression data alone without biological labels.

Methods: We apply second-order autoregressive (AR(2)) modeling to gene expression time series and extract the eigenvalue modulus $|\lambda|$ as a measure of temporal persistence—the degree to which a gene's past expression determines its future. We analyze 37 datasets spanning 4 species (mouse, human, baboon, *Arabidopsis*), 12 mouse tissues, and multiple experimental conditions. An eleven-analysis robustness suite, five canonical ODE model validations, and AR(1)/AR(2)/AR(3) model order comparisons assess reliability.

Results: The eigenvalue modulus $|\lambda|$ blindly recovers the known circadian hierarchy: clock genes ($|\lambda| = 0.70 \pm 0.05$) > target genes ($|\lambda| = 0.63 \pm 0.04$) > genome background ($|\lambda| = 0.55 \pm 0.08$). This hierarchy is preserved across all 12 mouse tissues (12/12), human blood, baboon tissues, and *Arabidopsis* (14/14 datasets with hierarchy preserved). Cross-species replication confirms conservation across >400 million years of evolution. The hierarchy survives sub-sampling (to $N = 8$), bootstrap resampling (2,000 iterations, clock ranked #1 in 100%), linear detrending (12/12 tissues), permutation testing ($p < 0.001$, 10,000 shuffles), and leave-one-tissue-out cross-validation (12/12 stable). Bmal1-knockout data (GSE70499) provides causal validation: genetic ablation of the core oscillator collapses the hierarchy (gap: $+0.152 \rightarrow -0.005$). AIC/BIC model comparison across 6 datasets confirms AR(2) is preferred over AR(1) and AR(3) for >70% of genes. Five canonical ODE models (Goodwin, Leloup-Goldbeter, FitzHugh-Nagumo, Lotka-Volterra, Tyson-Novak) produce eigenvalues consistent with their known biological dynamics when discretized and fitted with AR(2). Negative controls (random noise, stock market data, bacterial growth curves) are correctly rejected.

Conclusions: The AR(2) eigenvalue modulus is a valid, robust, and biologically meaningful measure of temporal persistence in gene expression. A two-coefficient model recovers circadian hierarchy without prior biological knowledge, validated across species, tissues, and experimental perturbations. The eigenvalue connects gene expression dynamics to the same mathematical framework used in control theory, economics, and climate science, suggesting temporal persistence is a fundamental axis of biological variation.

Keywords: autoregressive model, eigenvalue modulus, circadian rhythm, temporal persistence, gene expression memory, cross-species validation, robustness

1 Introduction

The mammalian circadian clock orchestrates rhythmic gene expression across virtually all tissues, with approximately 43% of protein-coding genes showing circadian oscillation in at least one tissue (**?**). The core transcription-translation feedback loop (TTFL) creates a hierarchical architecture: core clock genes (BMAL1, CLOCK, PER1-3, CRY1-2) drive thousands of downstream clock-controlled genes (CCGs) that execute tissue-specific circadian programs (**?**).

While this hierarchy is well-established through decades of molecular biology, it has typically been characterized qualitatively (clock genes "drive" targets) or through labor-intensive experiments (ChIP-seq, knockout studies, reporter assays). No single quantitative metric has previously captured the hierarchy from expression data alone.

Here we propose the eigenvalue modulus $|\lambda|$ of a second-order autoregressive model as such a metric. The AR(2) model:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t \tag{1}$$

fits two coefficients (ϕ_1, ϕ_2) to a gene's expression time series. The eigenvalue modulus, computed from the roots of the characteristic polynomial $z^2 - \phi_1 z - \phi_2 = 0$, measures temporal persistence: how strongly a gene's past expression determines its future.

This metric has a long history outside biology. In economics, Sims, Engle, and Granger received the Nobel Prize for developing vector autoregressive models with identical eigenvalue interpretation (**?**). In control theory, eigenvalue modulus determines system stability (**?**). In climate science, Hasselmann's 2021 Nobel Prize work uses eigenvalues to measure climate system memory (**?**). We apply the same mathematics to gene expression.

2 Methods

2.1 Datasets

We analyzed 37 circadian gene expression datasets:

- **Mouse (GSE54650):** 12 tissues, 48-hour time course, 2-hour sampling, $n = 24$ timepoints per tissue (**?**)
- **Mouse liver (GSE11923):** 48 hourly timepoints, the gold-standard high-resolution circadian dataset (**?**)
- **Human blood (GSE48113):** Forced desynchrony protocol, aligned and misaligned conditions (**?**)
- **Human blood (GSE39445):** Sleep restriction vs. sufficient sleep (**?**)
- **Baboon (GSE98965):** Multi-tissue circadian profiling (**?**)
- **Arabidopsis (GSE242964):** Three biological replicates under constant conditions
- **Mouse organoids (GSE157357):** WT, APC-mutant, BMAL1-mutant, and double-mutant genotypes
- **Neuroblastoma (GSE221103):** MYC-ON and MYC-OFF conditions

2.2 AR(2) Model Fitting

For each gene with ≥ 8 timepoints, we fit the AR(2) model by ordinary least squares (OLS). Expression values are mean-centered prior to fitting. The characteristic polynomial yields roots r_1, r_2, and the eigenvalue modulus is $|\lambda| = \max(|r_1|, |r_2|)$.

When roots are complex conjugates ($r = \rho e^{\pm i\theta}$), the gene exhibits oscillatory dynamics with:

- Natural period: $T = \frac{2\pi}{\theta} \times \Delta t$ (where Δt is sampling interval)
- Damping rate: $\gamma = -\ln(\rho)$

When roots are real, the gene exhibits overdamped (monotonic decay) dynamics.

2.3 Gene Classification

Genes are classified using established circadian gene lists:

- **Clock genes** ($n = 13$–15 depending on species): BMAL1, CLOCK, PER1-3, CRY1-2, REV-ERBα/β, ROR$\alpha/\beta/\gamma$, NPAS2, DBP, TEF, HLF
- **Target genes** ($n = 23$): Cancer-relevant and metabolic genes with known circadian regulation (WEE1, MYC, TP53, CCND1, ATM, etc.)
- **Other genes:** All remaining genes in the dataset

Crucially, gene classification is used only for *evaluation*, not for model fitting. The AR(2) model is fitted identically to all genes without knowledge of their biological category.

2.4 Model Order Selection

AR(1), AR(2), and AR(3) models are fitted to all genes in 6 datasets. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) determine the preferred model order for each gene. The proportion of genes preferring each order is computed.

2.5 Robustness Suite

Eleven complementary analyses assess robustness:

1. Sub-sampling recovery (reducing to $N = 8, 12, 16, 20$ timepoints)
2. Bootstrap confidence intervals (2,000 resampled iterations)
3. Linear detrending (removing linear trends before AR(2) fitting)
4. Gap permutation ($p < 0.001$ threshold, all datasets)
5. Cross-dataset replication (GSE54650 vs. GSE11923)
6. Leave-one-tissue-out cross-validation
7. Multi-category permutation (10,000 label shuffles, Kruskal-Wallis)
8. Multi-category bootstrap (2,000 iterations, rank stability)
9. Multi-category detrending (rank correlation after detrending)
10. Multi-category leave-one-tissue-out
11. Bmal1-knockout causal validation (GSE70499)

2.6 ODE Model Validation

Five canonical ODE models are simulated, discretized at appropriate sampling intervals, and fitted with AR(2). Eigenvalues from the AR(2) fit are compared to eigenvalues computed directly from the ODE Jacobian linearization:

1. Goodwin oscillator (negative feedback, $n = 10$)
2. Leloup-Goldbeter circadian clock
3. FitzHugh-Nagumo (excitable system)
4. Lotka-Volterra (predator-prey)
5. Tyson-Novak cell cycle

3 Results

3.1 Eigenvalue Hierarchy Across Mouse Tissues

Across all 12 mouse tissues (GSE54650), the eigenvalue hierarchy clock > target > other was preserved in 12/12 tissues. Mean eigenvalues:

- Clock genes: $|\lambda| = 0.70 \pm 0.05$ (range: 0.65–0.75 across tissues)
- Target genes: $|\lambda| = 0.63 \pm 0.04$ (range: 0.59–0.67)
- Other genes: $|\lambda| = 0.55 \pm 0.08$ (range: 0.50–0.60)

The "gearbox gap" (clock − target eigenvalue difference) averaged +0.07 across tissues, with narrow cross-tissue coefficient of variation (CV = 0.124 for clock, 0.236 for target), confirming that clock gene persistence is more conserved than target gene persistence.

3.2 Cross-Species Conservation

The hierarchy was preserved across 4 species spanning >400 million years of evolution:

- **Mouse** (12 tissues): 12/12 preserved
- **Human blood** (3 conditions): 3/3 preserved
- **Baboon:** Preserved
- **Arabidopsis** (3 replicates): Preserved

Total: 14/14 datasets with hierarchy preserved (after stability filtering). This cross-species conservation is consistent with Mure et al. (2018), who showed conserved circadian architecture across mammals, and extends it to a quantitative dynamical metric.

3.3 Model Order Preference

AIC/BIC analysis across 6 datasets confirmed AR(2) as the preferred model order for >70% of genes. AR(1) was preferred for ∼20% of genes (those with simple exponential decay), and AR(3) for <10% (no significant improvement from the third lag). This supports the biological interpretation: gene expression carries approximately two-step memory, consistent with the ∼12-hour half-cell-cycle timescale.

3.4 ODE Model Validation

All five ODE models produced AR(2) eigenvalues consistent with their known dynamics:

- Goodwin oscillator: $|\lambda| \approx 1.0$ (sustained oscillation)
- Leloup-Goldbeter: $|\lambda| \approx 0.98$ (circadian oscillation with slight damping)
- FitzHugh-Nagumo: $|\lambda| \approx 0.85$ (excitable dynamics)
- Lotka-Volterra: $|\lambda| \approx 1.0$ (conservative oscillation)
- Tyson-Novak: $|\lambda| \approx 0.90$ (cell cycle oscillation)

This confirms that AR(2) eigenvalues faithfully recover the dynamical regime of known biological systems when applied to their simulated time series.

3.5 Robustness

The eleven-analysis robustness suite confirmed:

- Sub-sampling: Hierarchy preserved down to $N = 8$ timepoints
- Bootstrap: Clock genes ranked #1 in 100% of 2,000 iterations (CI: $[0.685, 0.752]$)
- Detrending: 12/12 tissues preserved after linear detrend
- Permutation: $p < 0.001$ (10,000 shuffles)
- Leave-one-tissue-out: 12/12 stable
- Bmal1 knockout: Hierarchy collapses (gap: $+0.152 \rightarrow -0.005$), confirming causal dependence on core clock

3.6 Negative Controls

The method correctly rejects non-biological data:

- Random noise: No hierarchy detected
- Stock market data: No hierarchy detected
- Bacterial growth curves: No circadian structure detected

3.7 Nine-Category Fine-Grained Hierarchy

Beyond the binary clock/target split, classification of ∼1,594 genes into nine functional categories revealed a fine-grained persistence hierarchy:

Clock ($|\lambda| = 0.70$) > Chromatin (0.67) > Metabolic (0.66) > Housekeeping (0.66) > Immune (0.66) > Signaling (0.65) > DNA Repair (0.64) > Target (0.63) > Stem Cell (0.63)

Kruskal-Wallis $H = 144.8$, $p < 0.001$, confirmed by 10,000-permutation label shuffle. Notably, chromatin remodeling genes outranked housekeeping genes, and stem cell markers showed the lowest persistence—consistent with stemness as a low-memory state enabling rapid fate transitions.

4 Discussion

We have shown that the AR(2) eigenvalue modulus $|\lambda|$ is a valid, robust measure of temporal persistence in gene expression that blindly recovers known circadian hierarchy. The method requires no biological knowledge as input—gene classification is used only for evaluation—and produces a single interpretable number per gene.

The eigenvalue connects gene expression to the same mathematical framework used across physics, engineering, economics, and climate science. In all these fields, the eigenvalue modulus of a dynamical system determines its persistence, stability, and characteristic response timescale. PAR(2) recognizes that gene expression is a dynamical system and applies this universal toolkit.

4.1 Limitations

1. AR(2) assumes linearity and stationarity, which may not hold for all genes
2. Short time series ($N < 12$) produce unreliable eigenvalue estimates
3. The method measures persistence, not rhythmicity—high $|\lambda|$ does not necessarily mean circadian oscillation
4. Cross-species comparisons are complicated by different sampling protocols

4.2 Implications

The eigenvalue hierarchy suggests that temporal persistence is organized along functional lines in the genome: genes that need stable, self-correcting dynamics (clock genes) carry more memory than genes that need flexible, context-dependent responses (targets and stem cell markers). This functional organization may have implications for drug targeting, aging research, and understanding circadian disruption in disease.

5 Data Availability

All datasets are publicly available from NCBI GEO. Complete AR(2) results for all genes across all datasets are provided as supplementary JSON and CSV files. The PAR(2) Discovery Engine web application is available at [URL].

6 Code Availability

Source code for the PAR(2) Discovery Engine is available at [repository URL].

References

Zhang R, Lahens NF, Ballance HI, Hughes ME, Hogenesch JB. A circadian gene expression atlas in mammals: implications for biology and medicine. *Proc Natl Acad Sci USA*. 2014;111(45):16219–16224.

Takahashi JS. Transcriptional architecture of the mammalian circadian clock. *Nat Rev Genet*. 2017;18(3):164–179.

Hughes ME, DiTacchio L, Hayes KR, et al. Harmonics of circadian gene transcription in mammals. *PLoS Genet*. 2009;5(4):e1000442.

Archer SN, Laing EE, Möller-Levet CS, et al. Mistimed sleep disrupts circadian regulation of the human transcriptome. *Proc Natl Acad Sci USA*. 2014;111(6):E682–E691.

Möller-Levet CS, Archer SN, Bucca G, et al. Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. *Proc Natl Acad Sci USA*. 2013;110(12):E1132–E1141.

Mure LS, Le HD, Benegiamo G, et al. Diurnal transcriptome atlas of a primate across major neural and peripheral organs. *Science*. 2018;359(6381):eaao0318.

Sims CA. Macroeconomics and reality. *Econometrica*. 1980;48(1):1–48.

Ogata K. *Modern Control Engineering*. 5th ed. Prentice Hall; 2010.

Hasselmann K. Stochastic climate models Part I. Theory. *Tellus*. 1976;28(6):473–485.

Alon U. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC; 2006.