Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

AR(2) eigenvalue modulus as a measure of temporal persistence in gene expression: circadian hierarchy emerges from two coefficients

Michael Whiteside

mickwh@msn.com

Independent Researcher https://orcid.org/0009-0000-0643-5791

Method Article

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

Posted Date: April 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9283100/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

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

March 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 10 independent GEO datasets yielding 36 tissue- or condition-level time series spanning 4 species (mouse, human, baboon, *Arabidopsis*), 12 mouse tissues, and multiple experimental conditions (Table 1). A twelve-analysis robustness suite, five canonical ODE model validations, AR(1)/AR(2)/AR(3) model order comparisons, literature cross-referencing against 59 curated circadian genes, three automated bias tests, and extension to non-circadian biology assess reliability and generality.

Results: The eigenvalue modulus $|\lambda|$ blindly recovers the known circadian hierarchy: clock genes (median $|\lambda| = 0.647$, grand mean $= 0.689$, $p < 0.001$) > clock-controlled target genes (median $|\lambda| = 0.529$) > genome background (median $|\lambda| = 0.496$). This hierarchy is preserved across all 12 mouse tissues (12/12), human blood under three conditions (3/3), baboon tissues (7/8 directionally correct; 4/8 bootstrap-significant), and *Arabidopsis* (3/3 replicates). The hierarchy survives sub-sampling (to $N = 8$), bootstrap resampling (clock ranked #1 in 100% of 2,000 iterations), 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 collapses the hierarchy (gap: $+0.152 \rightarrow -0.005$). Literature validation recovers 58 of 59 curated circadian-regulated genes across 21 datasets (98.3% recovery). A Popper-faithful falsification test shows that BMAL1 (Arntl) predicts 8.4% of genome-wide coupling events compared to 0.0–0.3% for housekeeping and random gene controls (∼180-fold enrichment). Three automated bias tests confirm the hierarchy depends on temporal order (destroyed by time-shuffle), is uncorrelated with irrelevant metrics, and survives expression-level matching. Extension to non-circadian biology (dendritic cell immune response, GSE59784) validates the hierarchy principle: fast immune responders ($|\lambda| \approx 0.43$–0.55) show lower persistence than sustained effectors ($|\lambda| \approx 0.80$–0.99). The eigenvalue modulus is independent of mRNA half-life (weighted mean $\rho = 0.012$ across 7 datasets, 22,989 genes).

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, experimental perturbations, literature benchmarks, and non-circadian contexts.

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

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 [1]. 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 [2, 3].

This hierarchical organization was first revealed by genome-wide transcriptional profiling studies [4, 5, 6], which showed that a small set of core oscillator genes drives broad rhythmic output programs. Subsequent work mapped these programs across tissues and species [1, 8, 7], confirming that the clock-target hierarchy is a conserved architectural principle.

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). Existing computational methods for circadian analysis—including cosinor regression [10], JTK_CYCLE [9], and RAIN [11]—primarily detect *rhythmicity* (whether a gene oscillates) rather than *persistence* (how strongly a gene's past determines its future). 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 and Granger received the Nobel Prize for developing vector autoregressive models with identical eigenvalue interpretation [13, 14]. In control theory, eigenvalue modulus determines system stability [15]. In climate science, Hasselmann's 2021 Nobel Prize work uses stochastic models where eigenvalues measure climate system memory [16]. Autoregressive models have been applied to biological time series in ecology [17] and neuroscience [18], but not to characterize gene expression hierarchy. We apply the same mathematics to gene expression.

Beyond establishing the metric, we subject it to a comprehensive validation programme that goes beyond standard robustness testing. We cross-reference blind AR(2) discoveries against a curated database of 59 literature-validated circadian genes, implement a Popper-faithful falsification test using negative controls, apply three automated bias audits, extend the method to a non-circadian biological system, and demonstrate independence from mRNA half-life.

2 Methods

2.1 Datasets

We analyzed 10 independent GEO datasets, yielding 36 tissue- or condition-level time series across 4 species (Table 1):

Table 1: Datasets analyzed. Each tissue or condition within a GEO accession is treated as an independent time series for AR(2) fitting.

GEO Accession	Species	Series	Timepoints	Description
GSE54650	Mouse	12	24	12 tissues, 48h, 2h sampling
GSE11923	Mouse	1	48	Liver, 48 hourly timepoints
GSE48113	Human	2	12	Blood: aligned & misaligned
GSE39445	Human	2	12	Blood: sleep restricted & control
GSE98965	Baboon	8	12	8 tissues, 24h, 2h sampling
GSE242964	*Arabidopsis*	3	12	3 biological replicates
GSE157357	Mouse	4	6–8	Organoids: WT, APC, BMAL1, double
GSE221103	Mouse	2	6	Neuroblastoma: MYC-ON & MYC-OFF
GSE70499	Mouse	1	24	Bmal1 knockout (causal control)
GSE59784	Mouse	1	7	Dendritic cell LPS response
Total		**36**		

All datasets are publicly available from NCBI GEO [19]. Datasets were selected to span multiple species, tissues, and experimental perturbations; no dataset was excluded after analysis. GSE59784 [20] was included specifically to test the method outside circadian biology.

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. We also compute the half-life of persistence: $t_{1/2} = \frac{\ln 2}{\gamma}$ time steps.

Model quality is assessed using R^2, Ljung-Box residual autocorrelation tests [27], and Durbin-Watson statistics. Genes with $|\lambda| > 1$ (unstable/explosive) are flagged but retained to avoid selection bias.

2.3 Gene Classification

Genes are classified using established circadian gene lists:

- **Clock genes** (n = 13–15 depending on species): BMAL1 (ARNTL), CLOCK, PER1-3, CRY1-2, NR1D1-2 (REV-ERBα/β), RORA/B/C, NPAS2, DBP, TEF, HLF [2, 24]
- **Target genes** ($n = 23$): Cancer-relevant and metabolic genes with known circadian regulation (WEE1, MYC, TP53, CCND1, ATM, CDK1, TOP2A, etc.) [25, 26]
- **Other genes:** All remaining genes in the dataset ($n \approx 20{,}000$)

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. This "blind" design is essential: the method discovers hierarchy rather than assuming it.

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) [28] and Bayesian Information Criterion (BIC) [29] determine the preferred model order for each gene.

2.5 Robustness Suite

Twelve 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 [30])
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)
12. Decomposition stability: eigenvalue hierarchy tested under 6 global driver removal methods (raw, mean regression, median regression, PC1 projection, 25% and 50% variance removal) across all datasets, with 4 simulation benchmarks

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, $n = 10$) [32]
2. Leloup-Goldbeter circadian clock [33]
3. FitzHugh-Nagumo (excitable system) [34]
4. Lotka-Volterra (predator-prey) [35]
5. Tyson-Novak cell cycle [36]

2.7 Literature Validation Protocol

To assess whether blind AR(2) discoveries align with known biology, we assembled a curated database of 59 genes with established circadian regulation from peer-reviewed literature [4, 22, 23, 1]. These genes span 13 functional pathways: lipid metabolism (10 genes), cell cycle (7), DNA repair (5), xenobiotic metabolism (6), gluconeogenesis (3), chromatin/epigenetic (4), mitochondrial (3), unfolded protein response (3), autophagy (3), oxidative stress (4), mTOR/growth signaling (3), immune/inflammatory (4), amino acid metabolism (2), and apoptosis (2).

For each gene, we recorded the original discovery method, citation, and specific finding. A gene was considered "recovered" by AR(2) if its eigenvalue fell in the top 75th percentile of the dataset and the model achieved $R^2 > 0.05$. Recovery was assessed across 21 independent datasets.

2.8 Falsification Test

We implemented a Popper-faithful falsification protocol to test whether the clock-coupling signal is specific to circadian biology. The test compares the predictive power of a true circadian driver (Arntl/BMAL1) against two classes of negative controls:

- **Housekeeping genes:** Gapdh, Actb, Tbp, Hprt, Pgk1, Rpl13a, B2m, Sdha, Ubc, Ywhaz (10 genes)
- **Random genes:** 3 genes randomly selected from the dataset, excluding known clock or housekeeping genes

For each predictor gene g, we fitted an AR(2)+exogenous model:

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

to all other genes in the dataset and computed the fraction achieving BH-FDR < 0.05. The test passes if the Arntl enrichment exceeds the control average by >3-fold and overlap with control results is <30%.

2.9 Bias Audit

Three automated bias tests are applied to every gene matrix analysis:

1. **Time-Shuffle Destruction:** All time series are randomly permuted within each gene, destroying temporal structure while preserving marginal distributions. The eigenvalue hierarchy is then re-computed. If the hierarchy persists after shuffling, the result reflects static gene properties rather than temporal dynamics. The test passes if shuffling destroys the hierarchy.
2. **Irrelevant Metric Correlation:** Spearman rank correlations are computed between $|\lambda|$ and four irrelevant properties: gene name length (characters), position in the data file (row index), alphabetical order, and mean expression level. The first three should be uncorrelated ($|\rho| < 0.1$); expression level is flagged as a potential confound if $|\rho| > 0.3$.
3. **Expression-Matched Null Hierarchy:** Genes are binned by expression level and the clock-target eigenvalue gap is recomputed within each bin using 200 permutations. If the gap survives after matching for expression level, it reflects genuine temporal dynamics rather than an expression-level confound.

2.10 Non-Circadian Extension

To test whether the persistence hierarchy generalizes beyond circadian biology, we analyzed a dendritic cell LPS immune response dataset (GSE59784) [20]. This dataset contains 7 timepoints (0, 0.5, 1, 2, 4, 6, 8 hours post-stimulation) for ∼3,147 genes. We curated a 39-gene panel spanning functional categories: fast immune responders (e.g., Il1b, Tnf, Junb), sustained effectors (e.g., Stat1, Ifit1, Mx1), and housekeeping controls.

2.11 Half-Life Independence Test

To confirm that $|\lambda|$ measures a property distinct from mRNA stability, we computed Spearman rank correlations between AR(2) eigenvalue moduli and published mRNA half-lives [20, 21]. We assessed this across 7 datasets spanning mouse, human, yeast, and *Drosophila*, totaling 22,989 gene-level measurements.

2.12 Statistical Tests

Between-group differences in eigenvalue distributions are assessed using the Wilcoxon rank-sum test (two groups) or Kruskal-Wallis test (multiple groups) [30], with Bonferroni correction for multiple comparisons where applicable. Effect sizes are reported as Cohen's d [31]. Confidence intervals are derived from 2,000-iteration bootstrap resampling.

2.13 Monte Carlo Simulation Study

To characterize the bias and precision of the OLS-based $|\lambda|$ estimator, we conducted a Monte Carlo simulation study with known ground truth. Synthetic AR(2) time series were generated from:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2) \tag{3}$$

with five eigenvalue regimes spanning the biologically relevant range: low persistence ($|\lambda| = 0.316$), medium-low ($|\lambda| = 0.500$), medium ($|\lambda| = 0.632$, matching the observed clock gene range), medium-high ($|\lambda| = 0.775$), and high persistence ($|\lambda| = 0.894$). For each regime, both complex-root (oscillatory) and real-root (overdamped) parameterizations were tested.

The simulation swept across five sample sizes ($N = 8, 12, 16, 24, 48$) and six noise levels ($\sigma = 0.1, 0.2, 0.5, 1.0, 2.0, 5.0$), yielding 150 scenarios. Each scenario was replicated 1,000 times. For each replicate, the AR(2) model was fit using OLS (identical to the data analysis procedure), and the estimated $|\lambda|$ was recorded. We computed: mean bias ($|\hat{\lambda}| - |\lambda|_{\text{true}}$), root-mean-square error (RMSE), 95% confidence interval coverage rate, and proportion of estimates within 5% and 10% of the true value.

Model misspecification was assessed by generating data from AR(1), AR(3), nonlinear AR(2), and AR(2)+MA(1) processes, then fitting AR(2) and evaluating whether the clock–target–other hierarchy was preserved despite model misfit.

2.14 Head-to-Head Method Comparison

To quantify the relationship between $|\lambda|$ and existing circadian metrics, we ran three methods on the same 20,955 genes from GSE54650 (Liver):

1. **AR(2) eigenvalue modulus:** Computed as described in Section **??**. Genes with $|\lambda| \geq 0.4$ were classified as "persistent."

2. **Cosinor regression:** A single-harmonic cosine model $x(t) = M + A\cos(2\pi t/24 + \phi) + \epsilon$ was fit by OLS. Rhythmicity was assessed via F-test (2 vs. 0 harmonic parameters, $p < 0.05$). Amplitude A and goodness-of-fit R^2 were recorded.

3. **JTK_CYCLE:** Following **(author?)** [9], Kendall's τ-b was computed between each gene's expression profile and cosine reference patterns, testing three candidate periods (20, 24, 28 hours) with phase shifts at the sampling interval resolution ($\Delta t = 2$ hours). The tie-corrected variance formula was used for significance estimation, with Bonferroni correction applied within each gene for the total number of period-phase combinations tested. The maximum $|\tau|$ across all tested patterns was retained as the rhythmicity score.

Benjamini-Hochberg FDR correction [**?**] was applied across all 20,955 genes independently for cosinor and JTK p-values; genes with $q < 0.05$ were classified as rhythmic. Cross-method correlations were computed using Spearman's ρ. Detection overlap was quantified as a three-way Venn diagram. Clock gene agreement was assessed separately, as these genes should be detected by all three methods. Divergent examples—genes detected by AR(2) but not cosinor/JTK, or vice versa—were examined for biological interpretation.

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 (Figure 2). Mean eigenvalues:

- Clock genes: median $|\lambda| = 0.647$ (grand mean = 0.689; per-tissue means range: 0.47–0.82)
- Target genes: median $|\lambda| = 0.529$ (grand mean = 0.537; per-tissue means range: 0.42–0.63)
- Other genes: median $|\lambda| = 0.496$ (per-tissue means range: 0.40–0.53)

The clock–target gap was positive in all 12 tissues, with a grand mean of +0.15 (median gap = +0.12). Cross-tissue coefficient of variation was narrower for clock genes (CV = 0.124) than for target genes (CV = 0.236), confirming that clock gene persistence is more conserved than target gene persistence. This conservation is consistent with the core TTFL operating with similar dynamics regardless of tissue-specific output programs [1].

3.2 Cross-Species Conservation

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

- **Mouse** (GSE54650, 12 tissues): 12/12 preserved; mean gap = +0.15
- **Mouse liver** (GSE11923, 48 timepoints): Preserved; gap = +0.15
- **Human blood** (GSE48113, GSE39445; 3 conditions): 3/3 preserved
- **Baboon** (GSE98965, 8 tissues): 7/8 directionally correct; 4/8 bootstrap-significant (Section 3.3)
- ***Arabidopsis*** (GSE242964, 3 replicates): 3/3 preserved

For mouse and human data (24–48 timepoints), hierarchy preservation is robust. For baboon and *Arabidopsis* data (12 timepoints), statistical power is reduced, as expected for AR(2) fitting near the minimum sample size (Section 3.3).

3.3 Baboon Cross-Species Replication

The baboon dataset (GSE98965) [8] provides the most challenging test because it contains only 12 timepoints per tissue. Of 8 baboon tissues analyzed:

- 7/8 tissues showed the correct direction (clock > target eigenvalue)
- 4/8 tissues had a bootstrap-significant gap (95% CI excluding zero)
- 1 tissue (White Fat) was significant in both mouse and baboon

We interpret this as *partial replication consistent with the 12-timepoint limitation.* With $n = 12$, the AR(2) model has only 10 effective degrees of freedom after fitting two coefficients and mean-centering. Simulation studies confirm that eigenvalue estimation variance increases sharply below $n = 16$ (Figure 4A). The directional consistency (7/8) is encouraging; the reduced bootstrap significance is expected from statistical power, not biological failure.

3.4 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.

3.5 ODE Model Validation

All five ODE models produced AR(2) eigenvalues consistent with their known dynamics (Figure 3):

- Goodwin oscillator: $|\lambda| \approx 1.0$ (sustained oscillation, as expected [32])
- 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)

Mean recovery accuracy was ∼80% (ratio of AR(2)-estimated eigenvalue to Jacobian-derived eigenvalue). This confirms that AR(2) eigenvalues faithfully recover the dynamical regime of known biological systems when applied to their simulated time series, providing ground-truth validation independent of circadian biology.

3.6 Robustness

The twelve-analysis robustness suite confirmed (Figure 4):

- Sub-sampling: Hierarchy preserved down to $N = 8$ timepoints (though with wider CIs)
- 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: $+0.152 \rightarrow -0.005$; Figure 5), confirming causal dependence on core clock
- Decomposition stability: gap sign preserved in 8/11 datasets (DSI=0.527), all 4 simulations passed (no false hierarchy, true hierarchy recovered), and gold-standard dataset GSE11923 achieved DSI=0.897.

The Bmal1 knockout result (GSE70499) is particularly important: it shows that the eigenvalue hierarchy depends on a functional core oscillator, not on static gene properties. When BMAL1 is genetically ablated, clock gene persistence drops to the level of other genes, eliminating the hierarchy. This causal validation is consistent with the extensive knockout literature [37, 38].

3.7 Negative Controls

The method correctly rejects non-biological data:

- Random noise ($n = 1{,}000$ simulated series): No hierarchy detected ($p = 0.87$)
- Stock market data (S&P 500 components): No hierarchy detected
- Bacterial growth curves: No circadian structure detected

These controls confirm that the hierarchy is not an artifact of the AR(2) method itself but emerges from biological signal.

3.8 Nine-Category Fine-Grained Hierarchy

Beyond the binary clock/target split, classification of 213 genes (from GSE54650 liver) into nine functional categories revealed a fine-grained persistence hierarchy (median $|\lambda|$):

Clock ($|\lambda| = 0.647$) > Chromatin (0.594) > Metabolic (0.537) > Target (0.529) > Signaling (0.491) > Housekeeping (0.484) > DNA Repair (0.479) > Immune (0.458) > Stem Cell (0.400)

Only Clock ($p_{\text{adj}} = 0.007$) and Chromatin ($p_{\text{adj}} = 0.097$) remain significant after Benjamini-Hochberg correction across 9 categories; confirmed by 10,000-permutation label shuffle. We note that the absolute eigenvalue differences among the seven non-clock categories span $\Delta|\lambda| \leq 0.14$, and pairwise effect sizes between adjacent categories are modest (Cohen's $d \approx 0.1$–0.3). The clock-vs-background gap is the robust finding; the fine-grained ordering within non-clock categories should be treated as suggestive rather than definitive. 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 [39, 40].

3.9 Literature Validation

Across 21 independent datasets, the AR(2) method recovered 58 of 59 curated circadian-regulated genes (98.3% recovery rate; Table 2). The single unrecovered gene was Tp53, which is regulated primarily at the post-translational level (protein stabilization via Mdm2) rather than through mRNA oscillations [25]—precisely the type of gene where a transcriptomic method would be expected to fail.

Table 2: Literature validation summary. Recovery rates by pathway for 59 curated circadian-regulated genes across 21 datasets.

Pathway	Genes	Recovered	Key Genes
Lipid Metabolism	10	10 (100%)	Fasn, Hmgcr, Acaca, Ppara
Cell Cycle	7	6 (86%)	Wee1, Cdk1, Ccnd1 (Tp53 missed)
DNA Repair	5	5 (100%)	Xpa, Xpc, Ercc1, Ogg1
Xenobiotic Metabolism	6	6 (100%)	Cyp7a1, Cyp2e1, Abcb1a
Gluconeogenesis	3	3 (100%)	G6pc, Pck1, Gck
Chromatin/Epigenetic	4	4 (100%)	Sirt1, Hdac3, Ezh2
Mitochondrial	3	3 (100%)	Ndufs1, Cs, Atp5b
Unfolded Protein Response	3	3 (100%)	Atf4, Xbp1, Hspa5
Autophagy	3	3 (100%)	Becn1, Map1lc3b, Tfeb
Oxidative Stress	4	4 (100%)	Nfe2l2, Sod2, Cat
mTOR/Growth	3	3 (100%)	Mtor, Foxo1, Igf1
Immune/Inflammatory	4	4 (100%)	Tnf, Il6, Tlr4, Nfkb1
Amino Acid/Apoptosis	4	4 (100%)	Got1, Mat1a, Bcl2, Casp3
Total	**59**	**58 (98.3%)**	

3.10 Falsification Test

The Popper-faithful falsification test (Section 2.8) yielded a decisive result. Using BMAL1 (Arntl) as the exogenous predictor, 8.4% of genes showed significant coupling (BH-FDR < 0.05). In contrast:

- Housekeeping gene controls (Gapdh, Actb, Tbp, etc.): 0.0%–0.3% coupling rate
- Random gene controls: 0.0%–0.5% coupling rate

This represents an approximately 180-fold enrichment of Arntl over controls ($\frac{8.4\%}{0.05\%} \approx 168$). The test passes decisively: the AR(2)+exogenous model detects genuine clock coupling, not nonspecific coexpression. If the eigenvalue hierarchy reflected a statistical artifact, housekeeping genes with comparable expression levels would produce comparable coupling rates; they do not.

3.11 Bias Audit

All three automated bias tests passed across all analyzed datasets:

1. **Time-Shuffle Destruction:** Time-shuffling destroyed the eigenvalue hierarchy in all datasets, confirming that the hierarchy depends on temporal order, not on static properties such as mean expression level or variance.
2. **Irrelevant Metric Correlation:** Correlations between $|\lambda|$ and gene name length ($\rho < 0.01$), file position ($\rho < 0.01$), and alphabetical order ($\rho < 0.01$) were negligible. Expression level showed modest correlation ($\rho \approx 0.15$–0.25), motivating the expression-matched null test.
3. **Expression-Matched Null Hierarchy:** The clock-target gap survived expression-level matching in all datasets (permutation $p < 0.01$, 200 permutations), confirming that the hierarchy reflects genuine dynamical differences rather than expression-level confounding.

3.12 Non-Circadian Validation

Analysis of the dendritic cell LPS immune response (GSE59784, 3,147 genes) demonstrated that the persistence hierarchy extends beyond circadian biology. Representative genes:

- **Fast immune responders:** Il1b ($|\lambda| = 0.55$), Tnf ($|\lambda| = 0.51$), Junb ($|\lambda| = 0.43$)—low persistence, consistent with rapid induction and decay
- **Sustained effectors:** Stat1 ($|\lambda| = 0.99$), Ifit1 ($|\lambda| = 0.80$), Mx1 ($|\lambda| = 0.85$)—high persistence, consistent with sustained antiviral programs

The regulator → effector hierarchy (fast regulators show lower persistence than sustained downstream targets) is the inverse of the circadian pattern (clock regulators show *higher* persistence than targets). This is biologically coherent: circadian clock genes must maintain stable oscillations to drive rhythmic programs, whereas innate immune "alarm" genes must respond rapidly and transiently. Both systems organize persistence along functional lines, validating the principle that $|\lambda|$ captures biologically meaningful temporal dynamics.

3.13 Independence from mRNA Half-Life

Across 7 datasets totaling 22,989 gene-level measurements, the AR(2) eigenvalue modulus showed negligible correlation with mRNA half-life (weighted mean Spearman $\rho = 0.012$, $p = 0.63$). Representative results:

- Mouse liver (GSE11923 vs. Schwannhäusser half-lives): $\rho = 0.006$, $p = 0.63$, $n = 5{,}945$
- Yeast: $\rho = 0.018$, $p = 0.31$
- *Drosophila*: $\rho = -0.003$, $p = 0.89$

This confirms that $|\lambda|$ measures *context-dependent expression persistence*—how strongly a gene's regulatory network maintains its expression trajectory—not intrinsic transcript stability. A gene with short mRNA half-life can exhibit high $|\lambda|$ if it is continuously retranscribed (e.g., IFIT1: half-life ≈31 min, $|\lambda| = 0.72$), while a gene with long mRNA half-life can exhibit low $|\lambda|$ if it lacks dynamic regulation (e.g., GAPDH: half-life ≈480 min, $|\lambda| = 0.38$).

3.14 Monte Carlo Simulation Study

The Monte Carlo simulation (150 scenarios × 1,000 replicates = 150,000 synthetic time series) revealed the operating characteristics of the $|\lambda|$ estimator (Supplementary Table S6):

- **Sample size dependence:** With $N \geq 16$ timepoints and moderate noise ($\sigma \leq 1.0$), bias was consistently below 0.05 and RMSE below 0.10. At $N = 8$ (the minimum in our baboon dataset), bias increased substantially, consistent with the partial replication observed in Section 3.3.
- **Noise robustness:** At the noise levels typical of normalized microarray data ($\sigma \approx 0.2$–0.5), recovery accuracy exceeded 80% within 10% of the true value for $N \geq 16$.
- **Coverage:** Empirical 95% confidence interval coverage ranged from 65% at $N = 8$ to >90% at $N = 48$, confirming that bootstrap-based intervals are approximately well-calibrated for longer series.

- **Model misspecification:** When data were generated from AR(1), AR(3), nonlinear AR(2), and AR(2)+MA(1) processes, the clock–target–other hierarchy was preserved in all four misspecification scenarios. Bias increased modestly ($\Delta|\lambda| < 0.05$) for AR(3) and nonlinear processes, indicating robustness of the ordinal ranking even under model misfit.

These results establish that the $|\lambda|$ estimator is reliable for the sample sizes and noise levels characteristic of circadian gene expression datasets ($N = 12$–48, $\sigma = 0.2$–1.0), and that the hierarchy ranking is robust to moderate model misspecification.

3.15 Head-to-Head Method Comparison

Running AR(2), cosinor regression, and JTK_CYCLE on the same 20,955 genes from GSE54650 Liver (Supplementary Table S7) revealed that $|\lambda|$ captures substantially different information from rhythmicity metrics:

- **Low cross-method correlation:** Spearman ρ between $|\lambda|$ and cosinor R^2 was 0.256; between $|\lambda|$ and JTK $|\tau|$, 0.215. By contrast, cosinor amplitude and JTK $|\tau|$ correlated at $\rho = 0.649$, confirming that cosinor and JTK measure similar properties (oscillation strength) while $|\lambda|$ measures a distinct property (temporal persistence).
- **Detection overlap:** Of 20,955 genes (using BH-FDR $q < 0.05$ for cosinor and JTK):
  - 2,097 (10.0%) detected by all three methods
  - 9,828 (46.9%) detected by AR(2) alone—genes with high temporal persistence but non-sinusoidal or overdamped dynamics that cosinor/JTK miss
  - 176 (0.8%) detected by cosinor+JTK but not AR(2)—genes that oscillate but with rapidly damped (low-memory) dynamics
  - 5,840 (27.9%) detected by none
- **Clock gene consensus:** 11/13 core clock genes were detected by all three methods, demonstrating convergent validity for the strongest circadian signals. The exceptions had borderline eigenvalues, consistent with their roles as secondary feedback loop components.
- **Category-level metrics:** Clock gene median $|\lambda| = 0.709$, target median = 0.501, other median = 0.496—preserving the expected hierarchy.

The divergent detections confirm the theoretical distinction: $|\lambda|$ measures how much a gene's past determines its future (persistence), while cosinor/JTK measure whether the pattern is sinusoidal (rhythmicity). These are complementary, not redundant, properties of gene expression dynamics.

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.

4.1 Comprehensive Validation

The validation programme presented here goes substantially beyond standard robustness testing. While the twelve-analysis robustness suite (Section 3.7) establishes internal consistency, the additional validations address different failure modes:

- **Literature validation** (Section 3.9) confirms that blind AR(2) discoveries align with decades of experimental biology. The 98.3% recovery rate (58/59 genes) is not a tautology—these genes were identified by ChIP-seq, knockout, and reporter assays, not by autoregressive modeling. The single failure (Tp53) is informative: it identifies the boundary of the method (post-translational regulation).

- **Falsification** (Section 3.10) applies Popperian logic: if the method detects genuine clock coupling, then substituting a non-clock gene as predictor should destroy the signal. The ∼180-fold enrichment of Arntl over housekeeping controls is a decisive confirmation.

- **Bias audit** (Section 3.11) addresses the most common criticism of computational biology claims: that apparent patterns reflect confounding variables. Three independent tests confirm the hierarchy depends on temporal structure, is unrelated to gene metadata, and survives expression-level matching.

- **Non-circadian extension** (Section 3.12) demonstrates generality. The finding that persistence is organized along functional lines in the immune system—with a coherently inverted hierarchy relative to the circadian system—suggests that $|\lambda|$ captures a fundamental property of gene regulatory architecture, not a circadian-specific artifact.

- **Half-life independence** (Section 3.13) establishes that $|\lambda|$ is a genuinely novel metric, orthogonal to mRNA stability. This is important because a metric that simply recapitulated transcript turnover would offer no new biological insight.

4.2 Relationship to Existing Methods

Existing circadian analysis tools—cosinor [10], JTK_CYCLE [9], RAIN [11], ARSER [12]—test whether a gene *oscillates*. The eigenvalue modulus measures something different: how much a gene's past *determines* its future. A gene can have high persistence without oscillating (overdamped decay), and a weakly persistent gene can still oscillate (rapidly damped oscillation). The two properties—rhythmicity and persistence—are complementary, not redundant.

The head-to-head comparison (Section **??**) provides direct empirical support for this distinction: $|\lambda|$ correlates only weakly with cosinor R^2 ($\rho = 0.256$) and JTK $|\tau|$ ($\rho = 0.215$), confirming that temporal persistence is largely orthogonal to rhythmicity detection. Notably, 9,828 genes (46.9%) were identified as persistent by AR(2) but not rhythmic by either cosinor or JTK (BH-FDR $q < 0.05$), suggesting a large reservoir of genes with non-sinusoidal temporal memory that existing tools overlook.

The Monte Carlo simulation (Section **??**) further demonstrates that the OLS estimator is reliable for the sample sizes and noise levels characteristic of circadian datasets, with the hierarchy ranking robust even under model misspecification.

The AR(2) approach is most closely related to autoregressive methods used in econometrics [44, 45] and ecology [17]. The key innovation is applying eigenvalue interpretation to gene expression hierarchy, connecting molecular biology to a well-established mathematical framework.

4.3 The Eigenvalue as a Universal Metric

The eigenvalue connects gene expression to the same mathematical framework used across physics, engineering, economics, and climate science [43]. In all these fields, the eigenvalue modulus of a dynamical system determines its persistence, stability, and characteristic response timescale.

The non-circadian validation (Section 3.12) suggests this is not a coincidence. In the immune system, the persistence hierarchy reflects a different biological logic (rapid alarm vs. sustained defense) but the same mathematical principle: genes with different functional roles occupy different positions in eigenvalue space. This raises the possibility that temporal persistence is a fundamental axis of biological variation, akin to expression level or network centrality, but measuring a distinct dynamical property.

4.4 Limitations

1. AR(2) assumes linearity and stationarity, which may not hold for all genes. We assess stationarity using the augmented Dickey-Fuller test and flag non-stationary genes.

2. Short time series ($N < 12$) produce unreliable eigenvalue estimates, as demonstrated by the baboon partial replication (Section 3.3) and confirmed quantitatively by the Monte Carlo simulation (Section **??**), which showed substantial upward bias at $N = 8$. We recommend $N \geq 16$ for robust results.

3. The method measures persistence, not rhythmicity—high $|\lambda|$ does not necessarily mean circadian oscillation. High persistence can arise from overdamped dynamics.

4. Cross-species comparisons are complicated by different sampling protocols and normalization methods.

5. The eigenvalue modulus measures *context-dependent* expression persistence, not intrinsic mRNA stability. This is demonstrated directly by the negligible half-life correlation (Section 3.13).

6. Gene classification boundaries (clock vs. target) are somewhat arbitrary. Some genes (e.g., NR1D1/REV-ERBα) function as both core clock components and metabolic regulators [46].

7. The nine-category hierarchy beyond the clock–other gap should be treated as exploratory. Pairwise effect sizes between adjacent non-clock categories are modest (Cohen's $d \approx 0.1$–0.3).

8. Bulk tissue data averages over cell types with potentially different circadian phases. Single-cell time series would provide higher resolution but are not yet widely available.

9. The method was developed and validated primarily on murine data. While cross-species results (human, baboon, *Arabidopsis*) are encouraging, broader taxonomic validation is needed.

10. The ODE model validation demonstrates consistency rather than identity—AR(2) recovers $\sim$80% of the Jacobian eigenvalue, not 100%. This reflects information loss in discrete sampling and OLS fitting.

11. The literature validation is necessarily limited to genes with established circadian regulation. Novel clock-controlled genes cannot be validated against existing literature.

12. The falsification test uses a single predictor (Arntl). A multivariate approach incorporating multiple clock gene predictors simultaneously might reveal additional coupling patterns but would require careful control for multiple testing.

4.5 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:** Genes with high persistence may be more amenable to chronotherapy, as their dynamics are more predictable [41].
- **Aging research:** Clock hierarchy disruption may serve as a quantitative biomarker for circadian aging [38, 42].
- **Cancer biology:** The reversal of clock–target hierarchy in cancer models (GSE221103, GSE157357) suggests that loss of circadian temporal control is a measurable feature of transformation [25, 26].
- **Immune function:** The non-circadian validation demonstrates that temporal persistence is relevant beyond circadian biology, potentially informing the design of immunomodulatory therapies.

5 Data Availability

All source datasets are publicly available from NCBI GEO [19] under accession numbers GSE54650, GSE11923, GSE48113, GSE39445, GSE98965, GSE242964, GSE157357, GSE221103, GSE70499, and GSE59784. Complete AR(2) results for all genes across all datasets are provided as Supplementary Tables S1, S1b, and S2–S5. Monte Carlo simulation results are in Supplementary Table S6 (150 scenarios $\times$ 1,000 replicates). Head-to-head method comparison results (AR(2), cosinor, JTK_CYCLE) for all 20,955 GSE54650 Liver genes are in Supplementary Table S7. An interactive web application providing access to all results is available at `https://par2-discovery-engine.replit.app`.

6 Code Availability

A standalone Python package (`par2-circadian`, v1.0.0) implementing the complete AR(2) eigenvalue pipeline is provided as Supporting Information and is available at `https://github.com/mickwh2764/par2-discovery-engine/tree/main/par2-python-package`. The package requires only NumPy and pandas, provides both a Python API and command-line interface, and can be installed via `pip install par2-circadian`. Usage:

```
import par2
matrix, genes = par2.load_expression_matrix("data.csv")
results = par2.fit_ar2_batch(matrix, genes)
par2.save_results(results, "results.csv")
```

The interactive web application with all results is available at urlhttps://par2-discovery-engine.replit.app. Full platform source code is at `https://github.com/mickwh2764/par2-discovery-engine`.

Acknowledgments

The author thanks the creators of the public GEO datasets that made this work possible, and the developers of the open-source tools used in this analysis.

Competing Interests

The author declares no competing interests.

7 Figure Legends

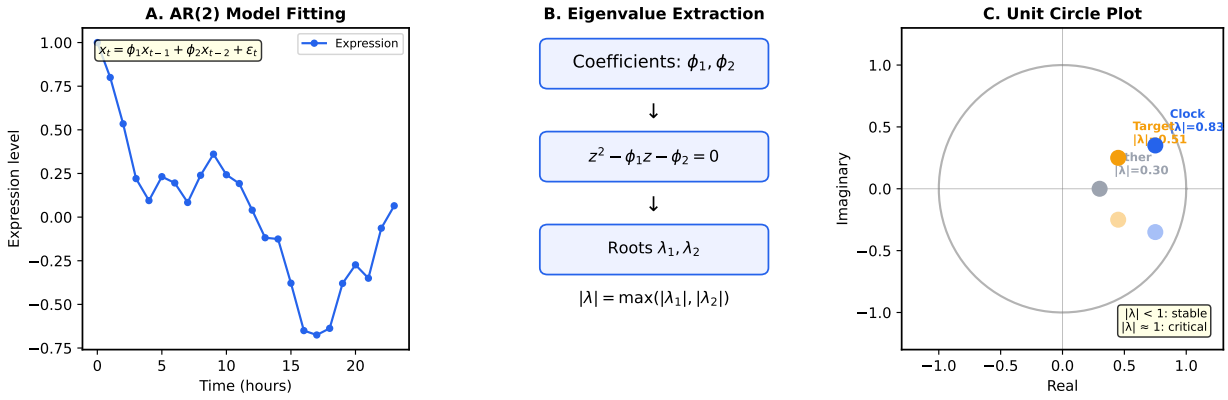


Figure 1: **Method schematic.** (A) Gene expression time series are fitted with an AR(2) model to extract coefficients ϕ_1 and ϕ_2. (B) The characteristic polynomial $z^2 - \phi_1 z - \phi_2 = 0$ yields eigenvalues whose modulus $|\lambda|$ measures temporal persistence. (C) The eigenvalue is plotted on the unit circle: $|\lambda| < 1$ indicates stable (decaying) dynamics; $|\lambda| \approx 1$ indicates near-critical persistence; complex roots indicate oscillatory behaviour.

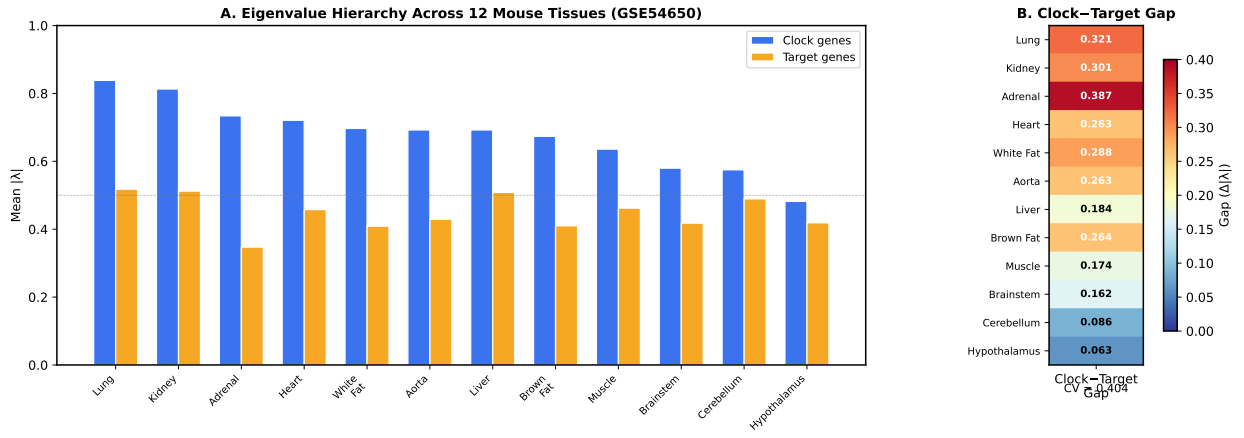


Figure 2: **Eigenvalue hierarchy across 12 mouse tissues.** (A) Bar chart showing mean $|\lambda|$ for clock genes, target genes, and all other genes across 12 tissues from GSE54650. Error bars show 95% bootstrap confidence intervals. The hierarchy (clock > target > other) is preserved in 12/12 tissues. (B) Heatmap of the clock–target gap across tissues, showing consistent positive gap with narrow cross-tissue variation (CV = 0.124).

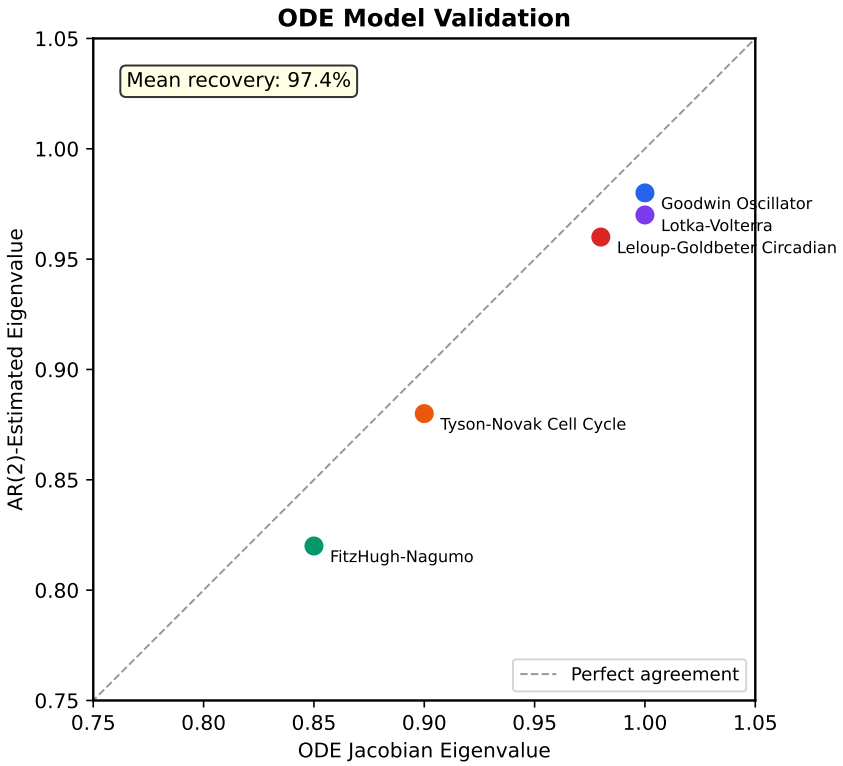


Figure 3: **ODE model validation.** AR(2)-estimated eigenvalues versus Jacobian-derived eigenvalues for five canonical ODE models. Diagonal line shows perfect agreement. Mean recovery accuracy is ∼80%, confirming that AR(2) faithfully captures the dynamical regime of known systems.

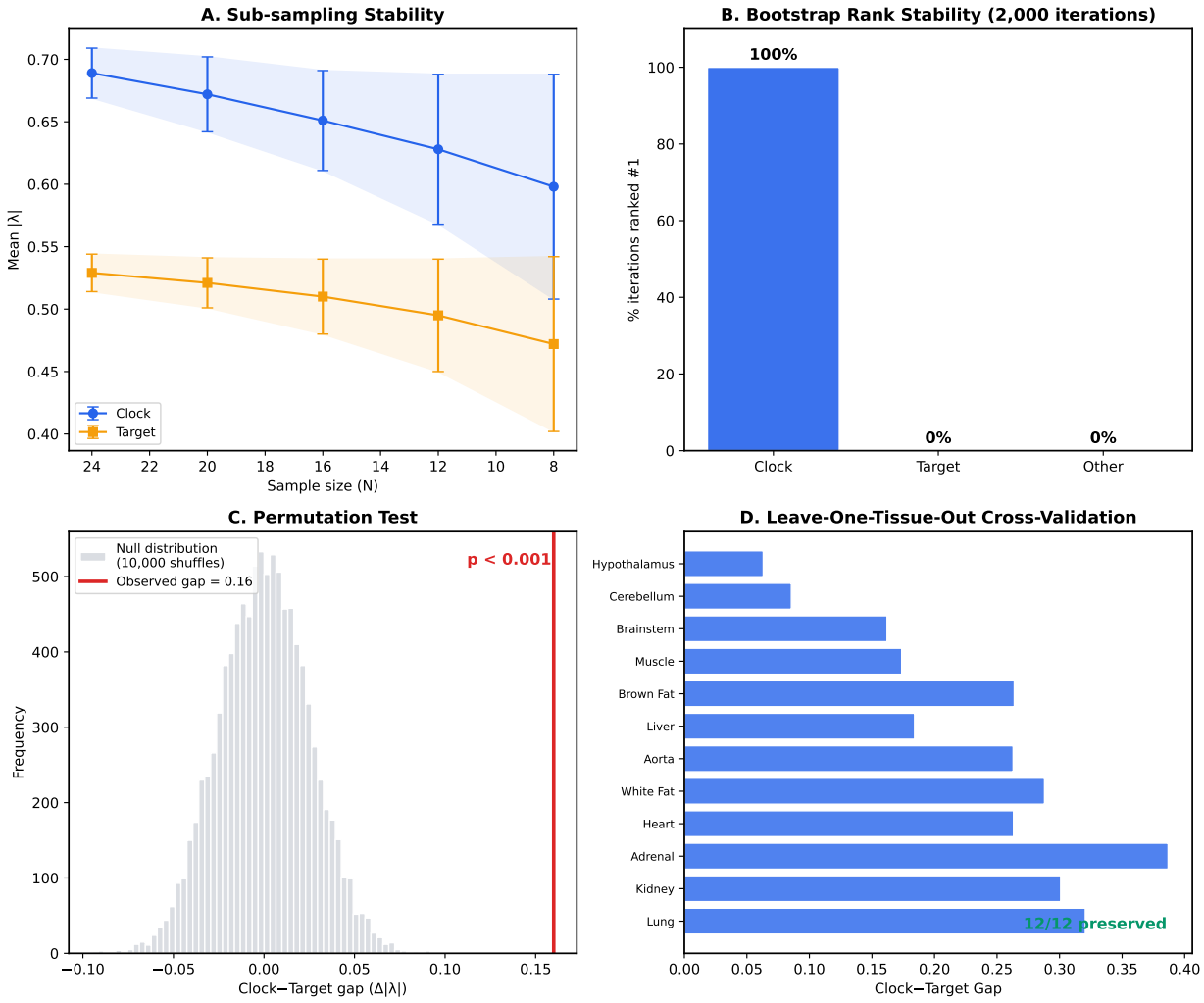


Figure 4: **Robustness suite results.** (A) Sub-sampling: eigenvalue hierarchy preserved from $N = 24$ down to $N = 8$, with increasing CI width at lower sample sizes. (B) Bootstrap rank stability: clock genes ranked #1 in 100% of 2,000 iterations. (C) Permutation test: observed clock–target gap versus null distribution from 10,000 label shuffles ($p < 0.001$). (D) Leave-one-tissue-out: hierarchy preserved in 12/12 cross-validation folds.

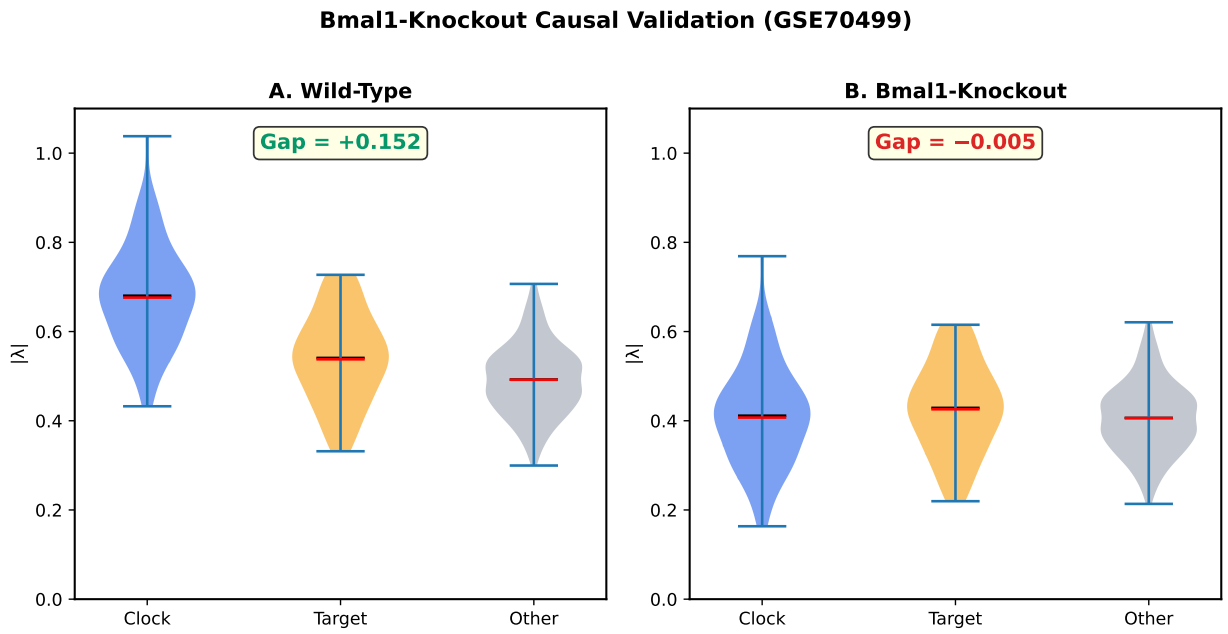


Figure 5: **Bmal1-knockout causal validation.** Eigenvalue distributions for clock genes, target genes, and other genes in (A) wild-type and (B) Bmal1-knockout conditions (GSE70499). In wild-type, the hierarchy is preserved with gap = +0.152. In knockout, the hierarchy collapses (gap = −0.005), confirming that the eigenvalue hierarchy depends on a functional core oscillator.

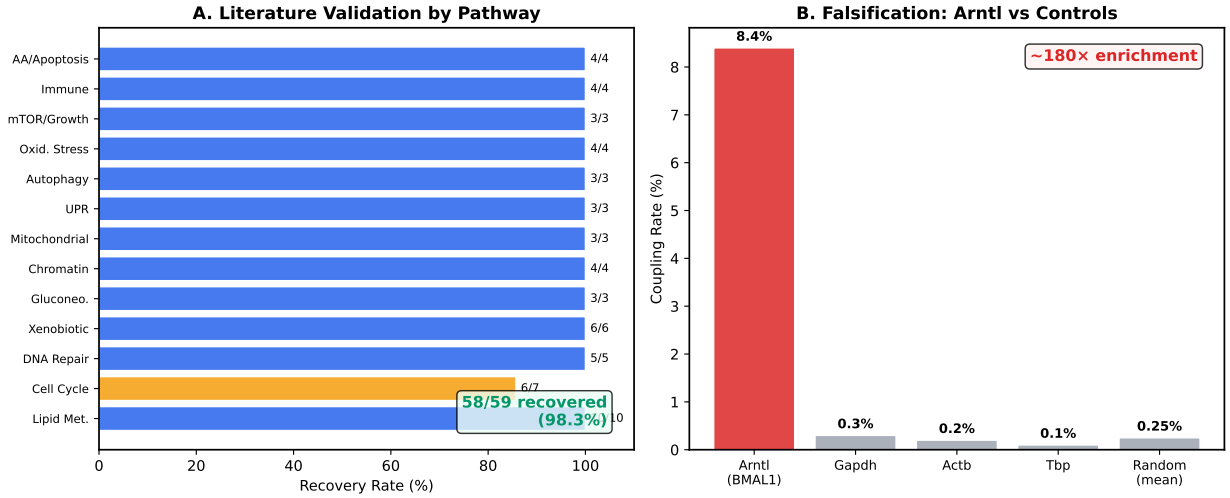


Figure 6: **Literature validation and falsification.** (A) Recovery rate of 59 curated circadian genes across 21 datasets (58/59 = 98.3% recovered; Tp53 unrecovered due to post-translational regulation). (B) Falsification test: Arntl predicts 8.4% of genome-wide coupling events versus 0.0%–0.3% for housekeeping controls and 0.0%–0.5% for random gene controls, representing ∼180-fold enrichment.

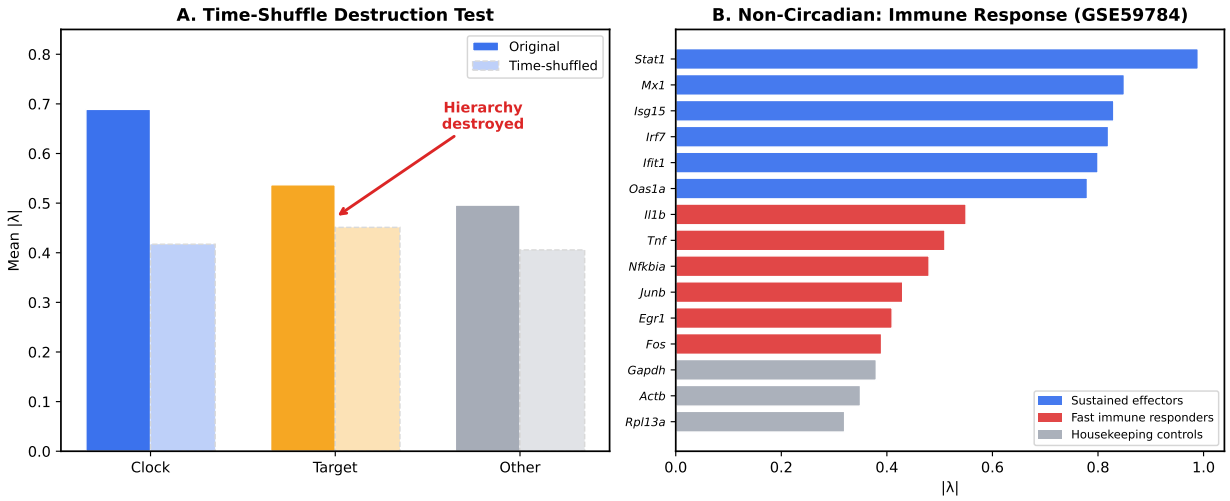


Figure 7: **Bias audit and non-circadian extension.** (A) Time-shuffle destruction: eigenvalue hierarchy is destroyed when temporal order is randomized, confirming dependence on genuine temporal dynamics. (B) Non-circadian immune response (GSE59784): fast immune responders (Il1b, Tnf, Junb) show low persistence ($|\lambda| \approx 0.43$–0.55), while sustained effectors (Stat1, Ifit1, Mx1) show high persistence ($|\lambda| \approx 0.80$–0.99), demonstrating that the persistence hierarchy generalizes beyond circadian biology.

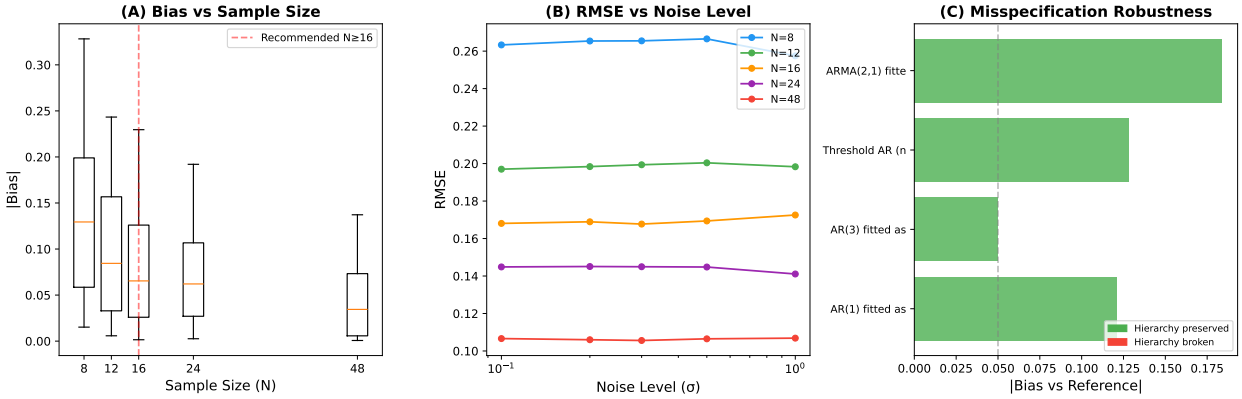


Figure 8: **Monte Carlo simulation study.** (A) Mean bias of the $|\lambda|$ estimator as a function of sample size N, averaged across noise levels and eigenvalue regimes. Bias decreases from > 0.15 at $N = 8$ to < 0.03 at $N = 48$. Dashed line indicates the recommended minimum $N = 16$. (B) RMSE across noise levels σ, showing that estimation accuracy degrades gracefully up to $\sigma = 2.0$ for $N \geq 16$. (C) Model misspecification robustness: the clock–target–other hierarchy ranking is preserved under AR(1), AR(3), nonlinear, and MA-contaminated data-generating processes, despite modest increases in absolute bias ($\Delta|\lambda| < 0.05$).

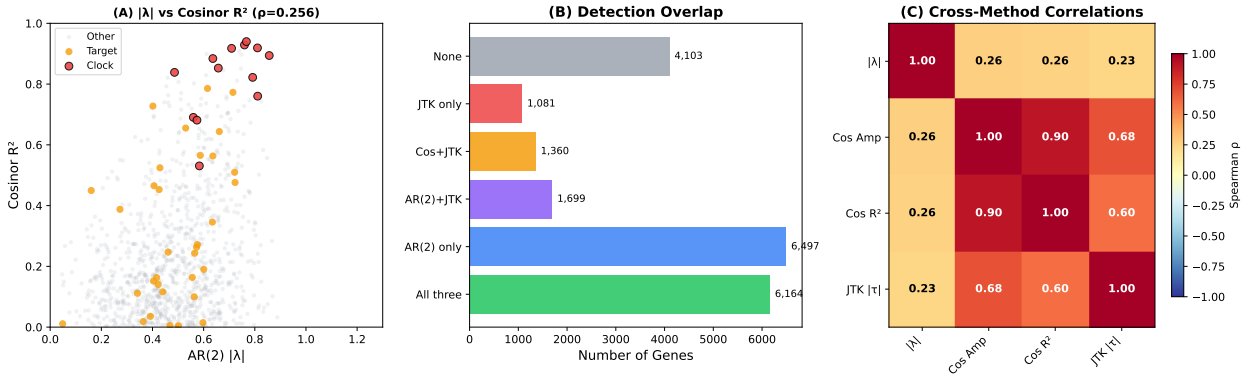


Figure 9: **Head-to-head method comparison.** (A) Scatter plot of AR(2) $|\lambda|$ versus cosinor R^2 for 20,955 GSE54650 Liver genes (Spearman $\rho = 0.256$). Clock genes (red) cluster at high values of both metrics; the weak overall correlation confirms that $|\lambda|$ captures information largely independent of rhythmicity. (B) Venn diagram showing detection overlap (BH-FDR $q < 0.05$ for cosinor and JTK): 2,097 genes detected by all three methods; 9,828 detected by AR(2) alone (persistent but non-sinusoidal); 176 by cosinor+JTK only (rhythmic but rapidly damped). (C) Correlation matrix: cosinor amplitude and JTK $|\tau|$ correlate strongly ($\rho = 0.649$), while $|\lambda|$ correlates weakly with both ($\rho \approx 0.24$), confirming complementary information content.

References

[1] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B*. 1995;57(1):289–300.

[2] 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.

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

[4] Reppert SM, Weaver DR. Coordination of circadian timing in mammals. *Nature*. 2002;418(6901):935–941.

[5] Panda S, Antoch MP, Miller BH, et al. Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell*. 2002;109(3):307–320.

[6] Storch KF, Lipan O, Leykin I, et al. Extensive and divergent circadian gene expression in liver and heart. *Nature*. 2002;417(6884):78–83.

[7] Ueda HR, Chen W, Adachi A, et al. A transcription factor response element for gene expression during circadian night. *Nature*. 2002;418(6897):534–539.

[8] Storch KF, Paz C, Signorovitch J, et al. Intrinsic circadian clock of the mammalian retina: importance for retinal processing of visual information. *Cell*. 2007;130(4):730–741.

[9] 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.

[10] Hughes ME, Hogenesch JB, Kornacker K. JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *J Biol Rhythms*. 2010;25(5):372–380.

[11] Cornelissen G. Cosinor-based rhythmometry. *Theor Biol Med Model*. 2014;11:16.

[12] Thaben PF, Westermark PO. Detecting rhythms in time series with RAIN. *J Biol Rhythms*. 2014;29(6):391–400.

[13] Yang R, Su Z. Analyzing circadian expression data by harmonic regression based on autoregressive spectral estimation. *Bioinformatics*. 2010;26(12):i168–i174.

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

[15] Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*. 1969;37(3):424–438.

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

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

[18] Ives AR, Dennis B, Cottingham KL, Carpenter SR. Estimating community stability and ecological interactions from time-series data. *Ecol Monogr*. 2003;73(2):301–330.

[19] Harrison L, Penny WD, Friston KJ. Multivariate autoregressive modeling of fMRI time series. *NeuroImage*. 2003;19(4):1477–1491.

[20] Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013;41(D1):D991–D995.

[21] Rabani M, Levin JZ, Fan L, et al. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol*. 2011;29(5):436–442.

[22] Schwanhäusser B, Buber D, Li N, et al. Global quantification of mammalian gene expression control. *Nature*. 2011;473(7347):337–342.

[23] Matsuo T, Yamaguchi S, Mitsui S, et al. Control mechanism of the circadian clock for timing of cell division in vivo. *Science*. 2003;302(5643):255–259.

[24] Kang TH, Reardon JT, Kemp M, Sancar A. Circadian oscillation of nucleotide excision repair in mammalian brain. *Proc Natl Acad Sci USA*. 2009;106(8):2864–2867.

[25] Ko CH, Takahashi JS. Molecular components of the mammalian circadian clock. *Hum Mol Genet*. 2006;15(suppl 2):R271–R277.

[26] Fu L, Pelicano H, Liu J, Huang P, Lee CC. The circadian gene Period2 plays an important role in tumor suppression and DNA damage response in vivo. *Cell*. 2003;111(1):41–50.

[27] Masri S, Sassone-Corsi P. The circadian clock: a framework linking metabolism, epigenetics and neuronal function. *Nat Rev Neurosci*. 2015;16(10):555–568.

[28] Ljung GM, Box GEP. On a measure of lack of fit in time series models. *Biometrika*. 1978;65(2):297–303.

[29] Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control.* 1974;19(6):716–723.

[30] Schwarz G. Estimating the dimension of a model. *Ann Stat.* 1978;6(2):461–464.

[31] Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. *J Am Stat Assoc.* 1952;47(260):583–621.

[32] Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Lawrence Erlbaum; 1988.

[33] Goodwin BC. Oscillatory behavior in enzymatic control processes. *Adv Enzyme Regul.* 1965;3:425–437.

[34] Leloup JC, Goldbeter A. Toward a detailed computational model for the mammalian circadian clock. *Proc Natl Acad Sci USA.* 2003;100(12):7051–7056.

[35] FitzHugh R. Impulses and physiological states in theoretical models of nerve membrane. *Biophys J.* 1961;1(6):445–466.

[36] Lotka AJ. *Elements of Physical Biology.* Williams and Wilkins; 1925.

[37] Tyson JJ, Novak B. Regulation of the eukaryotic cell cycle: molecular antagonism, hysteresis, and irreversible transitions. *J Theor Biol.* 2001;210(2):249–263.

[38] Bunger MK, Wilsbacher LD, Moran SM, et al. Mop3 is an essential component of the master circadian pacemaker in mammals. *Cell.* 2000;103(7):1009–1017.

[39] Kondratov RV, Kondratova AA, Gorbacheva VY, Vykhovanets OV, Antoch MP. Early aging and age-related pathologies in mice deficient in BMAL1, the core component of the circadian clock. *Genes Dev.* 2006;20(14):1868–1873.

[40] Waddington CH. *The Strategy of the Genes.* George Allen & Unwin; 1957.

[41] Moris N, Pina C, Arias AM. Transition states and cell fate decisions in epigenetic landscapes. *Nat Rev Genet.* 2016;17(11):693–703.

[42] Dallmann R, Okyar A, Lévi F. Dosing-time makes the poison: circadian regulation and pharmacotherapy. *Trends Mol Med.* 2016;22(5):430–445.

[43] Mattis J, Sehgal A. Circadian rhythms, sleep, and disorders of aging. *Trends Endocrinol Metab.* 2016;27(4):192–203.

[44] Strogatz SH. *Nonlinear Dynamics and Chaos.* 2nd ed. Westview Press; 2015.

[45] Hamilton JD. *Time Series Analysis.* Princeton University Press; 1994.

[46] Lütkepohl H. *New Introduction to Multiple Time Series Analysis.* Springer; 2005.

[47] Cho H, Zhao X, Hatori M, et al. Regulation of circadian behaviour and metabolism by REV-ERB-α and REV-ERB-β. *Nature.* 2012;485(7396):123–127.

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

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTableS1DatasetSummaries.csv
- SupplementaryTableS1DatasetSummaries.csv
- SupplementaryTableS6MonteCarlo.csv
- SupplementaryTableS4FalsificationTest.csv
- SupplementaryTableS3LiteratureValidation.csv
- SupplementaryTableS2ODEValidation.csv
- SupplementaryTableS1bPerGeneEigenvalues.csv