

Dynamic Proportionality: A Unified Compositional Framework for Genomic Phase Transitions

DyProp Development Team

December 12, 2025

Abstract

The analysis of genomic dynamics—whether transcriptomic differentiation, mutational evolution, or microbial succession—relies on reconstructing continuous trajectories from snapshot data. However, current statistical methods are fundamentally limited by the constraints of Compositional Data Analysis (CoDa). They largely rely on univariate regression or correlation-based networks, both of which yield spurious associations when applied to sequencing counts constrained by library depth. Here, we propose **Dynamic Proportionality (DyProp)**, a unified statistical engine that integrates CoDa with Dynamical Systems Theory. By modeling the continuous evolution of gene-gene log-ratios on a quasi-potential landscape, DyProp identifies specific “Boundary Functions”—the mathematical descriptions of rapid network rewiring events. Using a high-performance vectorized template-matching engine, this framework moves beyond static comparisons to distinguish between controlled stoichiometric switches and chaotic network decoupling events, offering a rigorous new standard for analyzing phase transitions in scRNA-seq, scDNA-seq (CNV), and longitudinal metagenomics.

1 Introduction: The Static Limit

1.1 The Compositional Constraint

High-throughput sequencing data is inherently compositional^[Ref:131]. The total number of reads in a cell is an artifact of sequencer capacity, not a biological property. Consequently, an increase in one feature (e.g., an amplified oncogene) mathematically forces a decrease in the relative abundance of all other features^[Ref:132]. This constraint confines data to a simplex geometry. Standard statistical methods that assume data exists in real Euclidean space (\mathbb{R}^p)—including Pearson correlation and Euclidean distance—are flawed in this context^[Ref:133]. As demonstrated by Aitchison [1], applying these methods to compositional data yields spurious negative associations and false positive correlations driven solely by sequencing depth variance^[Ref:134].

1.2 The Dynamic Gap

While static solutions like `propr` have resolved these issues for group-wise comparisons, biology is dynamic^[Ref:136]. Current trajectory inference tools (e.g., `tradeSeq`, `scVelo`) model gene expression along pseudotime^[Ref:137]. However, they generally operate under one of two limitations:

1. **Univariate Analysis:** They track individual genes in isolation, ignoring the regulatory couplings that define cell states^[Ref:139].
2. **Correlation-Based Inference:** They infer dynamic networks using correlation metrics susceptible to compositional artifacts (the “Pearson Trap”)^[Ref:140].

There is currently no statistical framework that models **how the proportionality (stoichiometry) between features evolves continuously** during a cell-state transition.

2 The Unified Vision: A General Theory of Genomic Dynamics

We propose that genomic transitions—regardless of modality—can be mathematically described as **Singular Perturbations of Stoichiometry** [Ref:144]. A biological system exists in a stable homeostatic state (Outer Solution), passes through a rapid, unstable transition (Boundary Layer), and settles into a new stable state [Ref:145].

Dynamic Proportionality (DyProp) is a platform-agnostic engine designed to characterize this process. By abstracting the input data into a generic Compositional Matrix \mathbf{X} , DyProp provides a single mathematical language to describe stability and change across diverse fields [Ref:147].

Table 1: Platform-Agnostic Applications of Dynamic Proportionality

Modality	Compositional Unit	The “Tipping Point” Event
Transcriptome	mRNA Counts	Network Rewiring: A regulatory loop breaks (Decoupling) or forms (Coupling), such as an oncogene escaping homeostatic control.
Genome (CNV)	Read Depth	Dosage Crisis: A sudden shift in chromosomal copy number ratios, pinpointing the exact pseudotime of chromothripsis.
Microbiome	OTU/ASV Counts	Guild Breakdown: The metabolic partnership between species collapses (Dysbiosis) or a pathogen invades.
Epigenome	Peak Counts	State Commitment: The physical locking of an enhancer-promoter loop prior to gene expression.

3 Theoretical Framework

3.1 Integration of CoDa and Dynamics

DyProp integrates two distinct mathematical fields:

- **Compositional Data Analysis (CoDa):** We utilize log-ratios as the fundamental unit of measurement to ensure rigorous independence from sequencing depth [Ref:153].
- **Singular Perturbation Theory:** We model the transition between states not as a simple gradient, but as a dynamic system governed by a “Boundary Function.” This allows us to quantify the *sharpness* (ϵ) of a biological transition [Ref:154].

3.2 The Dual Metrics

To fully characterize a genomic trajectory, we must distinguish between the stability of a ratio and the strength of the association. We define two continuous metrics:

Dynamic Instability ($\Phi(t)$) – The Kinetic Metric: Measures the **Variance of the Log-Ratio**. A spike in $\Phi(t)$ indicates the system is in a “Boundary Layer” or phase transition. It serves as a seismometer for genomic stress [Ref:160].

Dynamic Coupling ($\rho(t)$) – The Structural Metric: Measures **Regulatory Integrity**. High $\rho(t)$ indicates that two features form a tight regulatory module. A collapse in $\rho(t)$ signifies network decoupling^[Ref:162].

4 Discussion

The impact of this framework is to shift the analytical question from “What is different?” to “When does the difference arise, and how?”^[Ref:165].

4.1 From Driver Genes to Driver Events

Traditional analysis yields lists of differentially expressed genes. DyProp yields a catalog of **Network Events**. We can now identify:

- The exact moment a tumor suppressor decouples from its downstream effectors^[Ref:169].
- The specific “Crisis Point” in a tumor’s history where copy number stability collapsed.
- The “Tipping Point” where a stem cell commits to a lineage.

4.2 Reconstructing the Rewired Network

Beyond detection, DyProp enables **Differential Topology Reconstruction**. By slicing the trajectory based on the detected Tipping Point τ , we can calculate the Rewiring Matrix $\Delta P = P_{post} - P_{pre}$. This allows us to map:

- **Rewiring:** Edge weights shifting from Target A to Target B.
- **Haywire Hubs:** Regulators (e.g., TP53) that lose connectivity with all targets simultaneously (massive degree loss).

5 The Physics of Regulation: Landscape Dynamics

5.1 Landscape Curvature and Stiffness

We reframe the gene regulatory network (GRN) not as a thermal system, but as a dynamical system traversing a quasi-potential landscape (e.g., the Waddington Epigenetic Landscape) [2].

In homeostatic states, the GRN minimizes transcriptional noise via strong restoring forces, corresponding to a steep “basin of attraction.” Here, the log-ratio variance is tightly constrained ($\Phi \approx 0$). To facilitate a phenotypic transition, these constraints must be relaxed. The system traverses a “saddle point” or unstable manifold where the restoring force diminishes [4].

5.2 Dynamic Instability as Regulatory Stiffness

Consequently, we reinterpret the **Dynamic Instability** metric ($\Phi(t)$) as the inverse of **Regulatory Stiffness** (or Landscape Curvature).

- **Homeostasis ($\Phi \approx 0$):** High stiffness. Strong repression minimizes transcriptional noise.
- **Transition ($\Phi \gg 0$):** Loss of stiffness. The landscape flattens, allowing the system to overcome energy barriers and rewire its topology [3].

5.3 The Energetic Cost of Fidelity

We posit that maintaining high regulatory stiffness is metabolically costly (e.g., ATP-dependent chromatin remodeling). A spike in Φ represents a momentary release of this cost to permit network reorganization. This framework distinguishes between:

1. **Coherent Switching (Type 1):** Controlled traversal to a new attractor.
2. **Incoherent Decoupling (Type 2):** Loss of attractor stability (Network Melting).

6 Methodology: The DyProp Engine

DyProp employs a three-stage hybrid architecture designed for scale ($> 10^8$ gene pairs) and statistical rigor:

1. **Stage I: Vectorized Template Matching (Discovery).** To avoid the convergence failures of iterative non-linear solvers (GAMLSS), we employ a dense linear algebra approach. We convolve the observed log-ratio trajectories against a pre-computed basis set of Singular Perturbation boundary functions via matrix multiplication. This performs a global search for topological tipping points (τ) and transition sharpness (ϵ) in $O(1)$ time per pair.
2. **Stage II: Multi-Evidence Classification (Inference).** Candidate events are classified using a decision tree that evaluates three dimensions: Kinetic Instability (Φ), Structural Coupling (ρ), and Transition Coherence (Shape Fit). This strictly separates controlled stoichiometric switches from chaotic decoupling events.
3. **Stage III: Hierarchical Validation (GLMM).** For validated candidates, we implement a post-hoc Generalized Linear Mixed Model (GLMM) module. This accounts for random effects (e.g., patient or batch variability) to ensure that detected dynamics are biological drivers rather than technical artifacts.

7 Future Horizons

This framework establishes a foundation for a new class of “Genomic Physics” tools:

- **Spatial Proportionality ($\nabla\Phi$):** Defining the invasive margin of tumors as a boundary of maximal instability.
- **Causal Inference:** Using multi-omic lags to prove that chromatin decoupling precedes transcriptional decoupling.
- **Resilience Analysis:** Measuring the depth of the attractor basin to predict the efficacy of combination therapies.

References

- [1] Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman and Hall.
- [2] Waddington, C. H. (1957). *The Strategy of the Genes*. Allen & Unwin.
- [3] Scheffer, M., et al. (2009). Early-warning signals for critical transitions. *Nature*, 461(7260), 53-59.
- [4] Ferrell, J. E. (2012). Bistability, bifurcations, and Waddington’s epigenetic landscape. *Current Biology*, 22(11), R458-R466.