# Dynamic Proportionality: A Unified Compositional Framework for Genomic Phase Transitions

DyProp Development Team

December 11, 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 for analyzing these trajectories are fundamentally limited. They largely rely on univariate regression or correlation-based network inference, both of which are mathematically invalid when applied to compositional data (sequencing counts constrained by arbitrary library depth). Here, we propose **Dynamic Proportionality (DyProp)**, a unified statistical engine that integrates Compositional Data Analysis (CoDa) with Generalized Additive Models for Location, Scale, and Shape (GAMLSS). By modeling the continuous evolution of gene-gene log-ratios, DyProp moves beyond static comparisons to identify the specific "Boundary Functions"—the mathematical descriptions of rapid network rewiring events—that drive biological phase transitions. This framework is platform-agnostic, offering a rigorous new standard for analyzing dynamics 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. The total number of reads in a cell is an artifact of the 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.

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, Euclidean distance, and standard differential expression—are flawed in this context. As demonstrated by Aitchison, applying these methods to compositional data yields spurious negative associations and false positive correlations driven solely by sequencing depth variance.

## 1.2 The Dynamic Gap

While static solutions like `propr` have resolved these issues for group-wise comparisons, biology is dynamic. Current trajectory inference tools (e.g., `tradeSeq`, `scVelo`) model gene expression along pseudotime. 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.
2. **Correlation-Based Inference:** They infer dynamic networks using correlation metrics that are susceptible to compositional artifacts (the "Pearson Trap").

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**. 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.

**Dynamic Proportionality (DyProp)** is a platform-agnostic engine designed to characterize this process across the central dogma. 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.

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. We employ a **Bayesian Imputation Strategy** for zero-handling, ensuring that "On/Off" switches (biological zeros) are captured as quantifiable, high-magnitude shifts in the log-ratio, rather than being masked as missing data.
- **Singular Perturbations:** 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* ($\epsilon$) of a biological transition.

### 3.2 The Two-Stage Strategy (Anti-Blur)

To reconcile the trade-off between computational speed and boundary precision, DyProp employs a split architecture:

1. **Stage I (Exploratory Scan):** Uses kernel-weighted covariance to rapidly identify regions of instability across the genome. While kernels introduce smoothing, they are essential for robust discovery.
2. **Stage II (Mechanistic Quantification):** Uses **Cell-Level Imputed Ratios** (without temporal smoothing) to fit the boundary function. By fitting GAMLSS to the raw scatter of points, we recover the precise sharpness ($\epsilon$) of the transition, preventing the "blurring" effect of kernels.

### 3.3 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 Phi ($\Phi(t)$) – The Kinetic Metric:** Measures **Instability**. A spike in $\Phi(t)$ indicates the system is in a "Boundary Layer" or phase transition. It serves as a seismometer for genomic stress.

**Dynamic Rho ($P(t)$) – The Structural Metric:** Measures **Coupling**. High $P(t)$ indicates that two features form a tight regulatory module. A drop in $P(t)$ signifies network decoupling.

## 4 From Events to Mechanisms: Network Reconstruction

Beyond detecting events, DyProp enables the explicit reconstruction of the rewired network topology.

### 4.1 Temporal Slicing

Using the statistically derived Tipping Point $\tau$ and Sharpness $\epsilon$, we partition the trajectory into stable regimes: Pre-Transition ($t < \tau - 2\epsilon$) and Post-Transition ($t > \tau + 2\epsilon$).

### 4.2 Differential Topology ($\Delta P$)

We 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 ($\sum \Delta P_{ij} \ll 0$). This explicitly resolves whether a process is being moderated (rewired) or completely switched off (decoupled).

## 5 Discussion

The impact of this framework is to shift the analytical question from "What is different?" to "**When does the difference arise, and how?**"

### 5.1 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.
- 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.

### 5.2 Clinical Implications

This framework enables the discovery of **Process-Based Biomarkers**. Instead of targeting the final state (the cancer cell), therapies could target the transition mechanism itself—stabilizing the network before the "Boundary Layer" is crossed.