# RNA Velocity

Inferring Cell State Dynamics from Single-Cell RNA-seq Data

### 🧬 Spliced/Unspliced Ratio

Infer transcriptional dynamics from steady-state snapshots

### 📊 Velocity Estimation

Predict future cell states and trajectories

### ⚙️ Dynamic Models

Account for transcription, splicing, and degradation

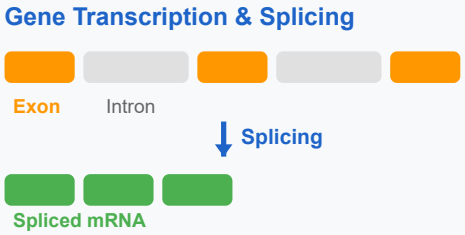
### 🚀 scVelo Improvements

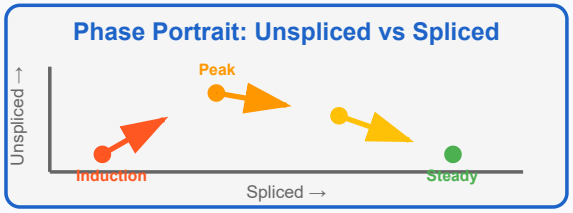
Dynamical model with latent time inference

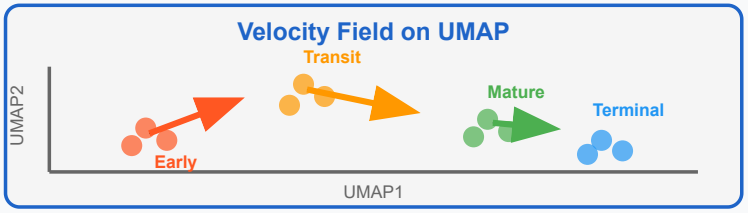
### 🎯 Interpretation

Direction and magnitude of cell state changes

**💡 RNA velocity adds a temporal dimension to static single-cell snapshots**



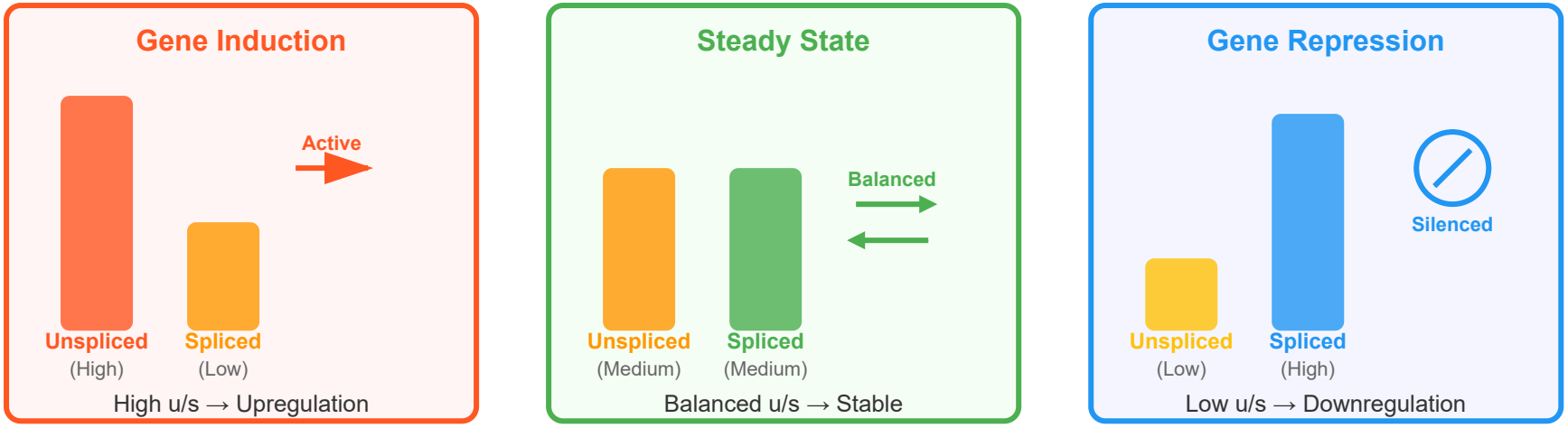



## 1 Spliced/Unspliced Ratio Analysis

RNA velocity leverages the natural biology of gene expression to infer cellular dynamics. During transcription, genes produce pre-mRNA containing both exons (coding sequences) and introns (non-coding sequences). Through splicing, introns are removed to create mature mRNA. By measuring both unspliced (pre-mRNA) and spliced (mature mRNA) counts in single cells, we can infer whether a gene is being actively induced, repressed, or in steady state.

### Key Concepts

- → **Unspliced mRNA:** Newly transcribed pre-mRNA containing introns, indicates recent transcriptional activity
- → **Spliced mRNA:** Mature mRNA after intron removal, represents the functional transcript pool
- → **Ratio dynamics:** High unspliced/spliced ratio suggests gene induction; low ratio suggests repression
- → **Temporal inference:** The relationship between unspliced and spliced abundance reveals transcriptional state
- → **Steady-state assumption:** Most scRNA-seq captures cells at equilibrium, but the ratio reveals direction of change




### Velocity Inference from Ratios

```
If u/s >> expected: positive velocity (gene turning ON)
    If u/s ≈ expected: zero velocity (steady state)
If u/s << expected: negative velocity (gene turning OFF)
```

## 2 Velocity Estimation Methods

RNA velocity estimation transforms the observed unspliced and spliced counts into predicted future states. The core idea is to fit a model that captures the relationship between unspliced and spliced abundances, then use deviations from steady-state to predict directional changes in gene expression. This velocity vector for each gene can then be projected into low-dimensional space (like UMAP or t-SNE) to visualize cell trajectories.

### Estimation Workflow

→ **Phase portrait construction:** Plot unspliced vs spliced counts for each gene across all cells

→ **Steady-state line fitting:** Identify the expected equilibrium relationship (typically linear)

→ **Deviation measurement:** Calculate how far each cell deviates from steady state

→ **Velocity vector computation:** Convert deviations into directional change predictions

→ **Dimensionality reduction projection:** Project gene-level velocities into embedding space

→ **Velocity field smoothing:** Average velocities across similar cells for robustness

### Mathematical Foundation

**Steady State Model:** At equilibrium, transcription rate $\alpha$ balances degradation: $\alpha = \beta s + \gamma u$

**Velocity Equation:** $v = du/dt - \gamma u$, where $du/dt$ represents change in unspliced abundance

**Linear Regression:** Fit $u = \gamma s$ to identify steady-state relationship from data

**Residual-based Velocity:** $v \propto u - \gamma s$
- positive = above line → gene increasing
- negative = below line → gene decreasing

**Projection:** Gene velocities projected to low-dimensional space: $v_{emb} = \Sigma\ (\partial x/\partial g) \times v_g$

## 3 Dynamic Models of RNA Kinetics

Dynamic models explicitly capture the kinetics of transcription, splicing, and degradation. Rather than assuming steady state, these models account for how unspliced and spliced abundances change over time through differential equations. This allows for more accurate velocity estimation, especially in systems with rapid transcriptional changes or where genes haven't reached equilibrium.

### Model Components

→ **Transcription rate (α):** Rate at which new unspliced mRNA is produced from the gene

→ **Splicing rate (β):** Rate at which unspliced mRNA is converted to spliced mRNA

→ **Degradation rates (γ_u, γ_s):** Rates at which unspliced and spliced mRNA are degraded

→ **Gene states:** Genes can be in induction, repression, or steady-state phases

→ **Time-dependent dynamics:** Models capture temporal evolution of mRNA populations

→ **Parameter estimation:** Infer kinetic parameters from observed u and s distributions

### Differential Equations for RNA Dynamics

**Unspliced mRNA:**

$$du/dt = \alpha(t) - \beta u - \gamma_u \cdot u$$

**Spliced mRNA:**

$$ds/dt = \beta u - \gamma_s \cdot s$$

**Gene States:**

- Induction: $\alpha$ switches from $0 \rightarrow \alpha_{max}$
- Steady: $\alpha$ constant, $du/dt \approx 0$, $ds/dt \approx 0$
- Repression: $\alpha$ switches from $\alpha_{max} \rightarrow 0$

### Steady-State Model (velocyto)

**Assumption:** Cells are at equilibrium
**Method:** Linear regression on u vs s
**Pros:** Simple, fast computation
**Cons:** Less accurate for transient states

### Dynamical Model (scVelo)

**Assumption:** Cells in different phases
**Method:** Fit full ODE system
**Pros:** Captures transitions accurately
**Cons:** More complex, computationally intensive

## 4 scVelo: Advanced Dynamic Modeling

scVelo represents a major advancement in RNA velocity analysis by implementing a fully dynamic model that accounts for gene-specific kinetics across all cells. Unlike the original velocyto approach which assumes steady state, scVelo models the complete transcription-splicing-degradation dynamics and infers a "latent time" variable for each cell, representing its progression through biological processes. This enables more accurate velocity estimates and better resolution of complex differentiation trajectories.

### Key Innovations

→ **Dynamical mode:** Jointly models induction, steady-state, and repression phases for each gene

→ **Latent time inference:** Estimates internal biological time for each cell along trajectories

→ **Gene-specific kinetics:** Learns individual α, β, γ parameters for each gene

→ **Likelihood-based fitting:** Uses EM algorithm to optimize parameters and latent time simultaneously

→ **Velocity confidence:** Provides uncertainty estimates for velocity predictions

→ **Improved projection:** Better velocity field smoothing and embedding projection methods

### scVelo Dynamical Model

**Likelihood-based Optimization:**
Maximize P(u, s | α, β, γ, t) across all cells

**For each gene:**
• Estimate switching time points (when α changes)
• Fit rate parameters (α, β, γ) that best explain observed u/s distributions
• Infer latent time t for each cell

**Velocity from Model:**
v = du/dt = α(t) - (β + γ_u)·u
Computed from fitted parameters and cell's position in latent time

## 5 Interpretation and Applications

Understanding RNA velocity results requires careful interpretation of both the direction and magnitude of velocity vectors in the context of biological processes. Velocity arrows indicate the predicted short-term future state of cells, enabling inference of differentiation trajectories, cell fate decisions, and regulatory dynamics. However, velocities should be validated against known biology and complemented with other analyses for robust conclusions.

### Interpretation Guidelines

→ **Arrow direction:** Points toward predicted future cell state along differentiation or cell cycle

→ **Arrow length:** Indicates magnitude of change; longer arrows suggest faster transitions

→ **Velocity consistency:** Coherent flow patterns indicate robust trajectories

→ **Terminal states:** Cells with short or inward-pointing arrows may represent endpoints

→ **Branching points:** Diverging velocities reveal cell fate bifurcations

→ **Confidence assessment:** Check velocity confidence scores and gene-level contributions

→ **Biological validation:** Verify predictions with known markers and experimental perturbations

### 🧬 Developmental Trajectories

Map cell fate decisions during embryogenesis and organogenesis. Velocity analysis reveals the directionality of differentiation pathways and identifies progenitor populations.

### 🔄 Cell Cycle Analysis

Distinguish cycling from non-cycling cells and determine cell cycle phase progression. Velocity vectors show circular patterns for proliferating cells.

### 🦠 Disease Progression

Study cancer evolution, immune response dynamics, and disease state transitions. Identify aberrant differentiation patterns in pathological conditions.

### 🧪 Drug Response

Assess how perturbations alter cell trajectories. Compare velocity fields before and after treatment to identify mechanistic effects.

### 🧠 Neuronal Differentiation

Trace neural stem cell differentiation into diverse neuronal subtypes. Resolve complex branching patterns in neurogenesis.

### ☆ Best Practices

✓ **Quality control:** Filter low-quality cells and genes before velocity analysis

✓ **Multiple approaches:** Compare steady-state and dynamical models

✓ **Confidence metrics:** Examine velocity confidence and gene contributions

✓ **Biological context:** Interpret results in light of known markers and biology

✓ **Validation:** Cross-validate with trajectory inference methods (Monocle, Slingshot)

✓ **Visualization:** Use multiple embeddings (UMAP, PCA, force-directed) to assess robustness

### ⚠ Common Pitfalls to Avoid

✗ **Over-interpretation:** RNA velocity predicts SHORT-TERM futures, not long-term fates

✗ **Batch effects:** Technical variation can create spurious velocity patterns

✗ **Low coverage:** Sparse data reduces velocity reliability, especially for rare cell types

✗ **Assuming causality:** Velocity shows correlation with trajectories, not causal mechanisms

✗ **Ignoring biology:** Computational predictions must align with experimental evidence

✗ **Single-gene focus:** Analyze velocity across many genes, not individual genes in isolation

**Summary:** RNA velocity is a powerful computational method that adds temporal resolution to single-cell RNA-seq data. By leveraging the natural dynamics of RNA splicing, it enables prediction of cell state transitions, identification of differentiation

trajectories, and discovery of regulatory relationships. When combined with careful quality control, appropriate modeling choices (steady-state vs. dynamical), and biological validation, RNA velocity provides unique insights into cellular dynamics that are invisible in static gene expression measurements.

## Key References


- **velocyto:** La Manno et al. (2018). "RNA velocity of single cells." Nature.
- **scVelo:** Bergen et al. (2020). "Generalizing RNA velocity to transient cell states through dynamical modeling." Nature Biotechnology.
- **RNA velocity analysis:** Lange et al. (2022). "CellRank for directed single-cell fate mapping." Nature Methods.
- **Velocity interpretation:** Gorin et al. (2022). "RNA velocity unraveled." PLoS Computational Biology.