

Inferenza di dinamica del trascrittoma  
da dati di sequenziamento

Guido Sanguinetti

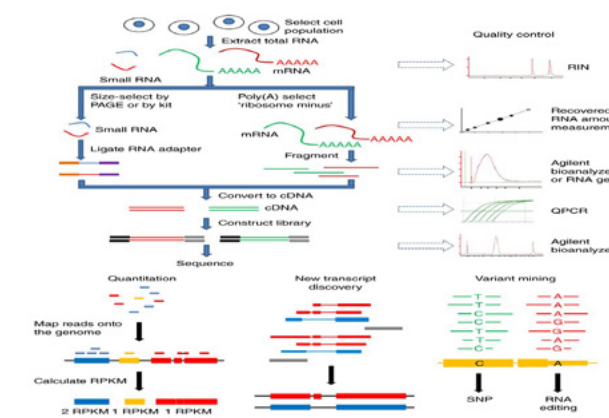
Theoretical and Scientific Data Science  
SISSA, Trieste

Sanita' 2030, Codroipo 06/24

### Roadmap of today

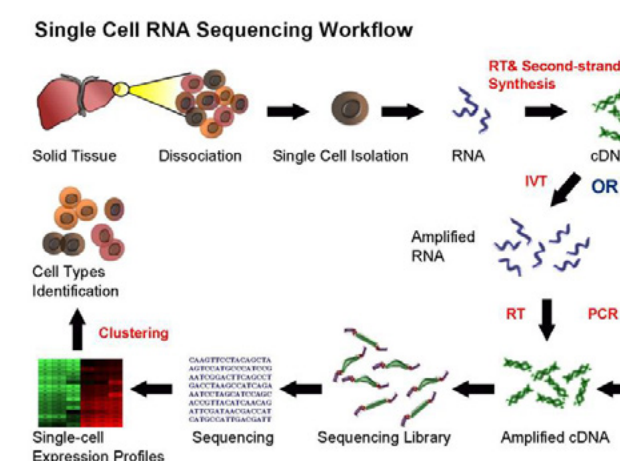
- 1 Single-cell RNA-seq
- 2 Dynamics from static data
- 3 Neural Differential equations
- 4 NeuroVelo (Idris Kouadri-Boudjelthia)

## RNA-seq



Output (after pre-processing): fragments of DNA (reads) aligned to a standard genome (.bam file). Then intersect with genomic coordinates (.bed file, e.g. exon coordinates) to obtain data matrix.

## Going single-cell



Same output as RNA-seq, but a) many cells and b) far fewer reads per cell

### Different technologies

- RNA-seq can be enriched/ depleted for specific transcript types (e.g. polyA, ribo-minus)

### Different technologies

- RNA-seq can be enriched/ depleted for specific transcript types (e.g. polyA, ribo-minus)
- scRNA-seq has various flavours: plate-based (one cell per well) or droplet-based (in principle unlimited number of cells)

## Different technologies

- RNA-seq can be enriched/ depleted for specific transcript types (e.g. polyA, ribo-minus)
- scRNA-seq has various flavours: plate-based (one cell per well) or droplet-based (in principle unlimited number of cells)
- Sequencing can be whole transcript (reads distributed across all transcript) or polyA (reads enriched at the end of transcripts)

## Different technologies

- RNA-seq can be enriched/ depleted for specific transcript types (e.g. polyA, ribo-minus)
- scRNA-seq has various flavours: plate-based (one cell per well) or droplet-based (in principle unlimited number of cells)
- Sequencing can be whole transcript (reads distributed across all transcript) or polyA (reads enriched at the end of transcripts)
- Unique Molecular Identifiers (UMIs) can be added to reduce amplification biases (in theory, each read comes from a unique transcript)



### What the data looks like

- Bulk RNA seq typically quantifies all genes/ transcripts ( $\approx$  100M reads) in triplicates, largest studies 100s of samples, time series with 10 time points.

## What the data looks like

- Bulk RNA seq typically quantifies all genes/ transcripts ( $\approx$  100M reads) in triplicates, largest studies 100s of samples, time series with 10 time points.
- Plate-based, full transcript scRNA-seq protocols measure 3-4K "genes" in 100s of cells, typical number of reads 100K per cell

## What the data looks like

- Bulk RNA seq typically quantifies all genes/ transcripts ( $\approx$  100M reads) in triplicates, largest studies 100s of samples, time series with 10 time points.
- Plate-based, full transcript scRNA-seq protocols measure 3-4K "genes" in 100s of cells, typical number of reads 100K per cell
- Droplet-based (10X) measure 1-2K "genes" in 10-100K cells (only polyA, typical number of reads 3-4K per cell)

## What the data looks like

- Bulk RNA seq typically quantifies all genes/ transcripts ( $\approx$  100M reads) in triplicates, largest studies 100s of samples, time series with 10 time points.
- Plate-based, full transcript scRNA-seq protocols measure 3-4K "genes" in 100s of cells, typical number of reads 100K per cell
- Droplet-based (10X) measure 1-2K "genes" in 10-100K cells (only polyA, typical number of reads 3-4K per cell)
- A word on dynamics

### Roadmap of today

- 1 Single-cell RNA-seq
- 2 Dynamics from static data
- 3 Neural Differential equations
- 4 NeuroVelo (Idris Kouadri-Boudjelthia)

### Uncovering dynamics: RNA velocity

- scRNA-seq is destructive → static snapshots from a dynamic process

## Uncovering dynamics: RNA velocity

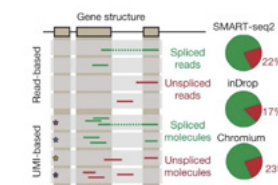
- scRNA-seq is destructive → static snapshots from a dynamic process
- **IDEA** (La Manno et al, 2018): use spliced/ unspliced reads to derive *rate of change* of RNA levels

$$\frac{dx_u}{dt} = \alpha - \beta x_u \quad \frac{dx_s}{dt} = \beta x_u - \gamma x_s$$

## Uncovering dynamics: RNA velocity

- scRNA-seq is destructive → static snapshots from a dynamic process
- **IDEA** (La Manno et al, 2018): use spliced/ unspliced reads to derive *rate of change* of RNA levels

$$\frac{dx_u}{dt} = \alpha - \beta x_u \quad \frac{dx_s}{dt} = \beta x_u - \gamma x_s$$

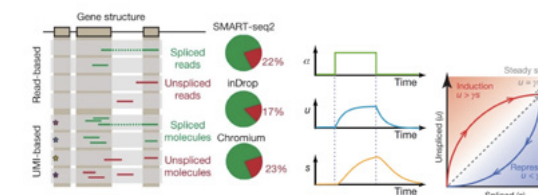




### Uncovering dynamics: RNA velocity

- scRNA-seq is destructive → static snapshots from a dynamic process
- **IDEA** (La Manno et al, 2018): use spliced/ unspliced reads to derive *rate of change* of RNA levels

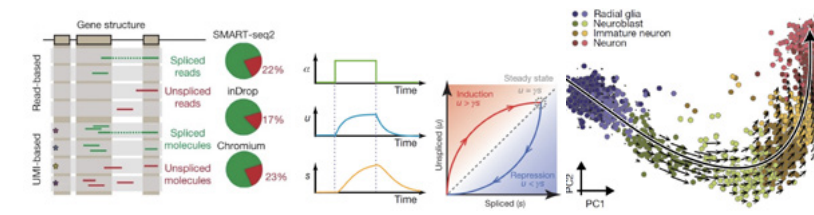
$$\frac{dx_u}{dt} = \alpha - \beta x_u \quad \frac{dx_s}{dt} = \beta x_u - \gamma x_s$$



### Uncovering dynamics: RNA velocity

- scRNA-seq is destructive → static snapshots from a dynamic process
- **IDEA** (La Manno et al, 2018): use spliced/ unspliced reads to derive *rate of change* of RNA levels

$$\frac{dx_u}{dt} = \alpha - \beta x_u \quad \frac{dx_s}{dt} = \beta x_u - \gamma x_s$$



### Problems and solutions

- Splicing signal is **very noisy** in single cells
- No reason why timescale of splicing should be the relevant one

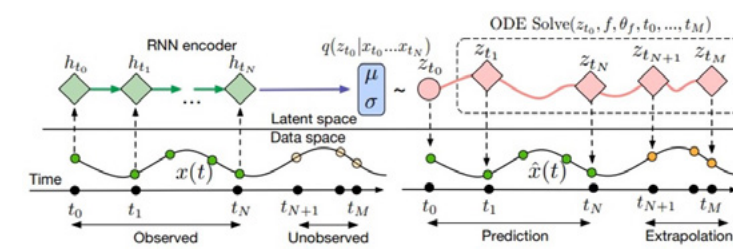
### Problems and solutions

- Splicing signal is **very noisy** in single cells
- No reason why timescale of splicing should be the relevant one
- **IDEA**: Underlying (low dimensional) nonlinear dynamical system should govern long-term evolution of cells' transcriptomes
- Spliced/ unspliced ratio gives a noisy measurement of *instantaneous* rate of change

## Problems and solutions

- Splicing signal is **very noisy** in single cells
- No reason why timescale of splicing should be the relevant one
- **IDEA**: Underlying (low dimensional) nonlinear dynamical system should govern long-term evolution of cells' transcriptomes
- Spliced/ unspliced ratio gives a noisy measurement of *instantaneous* rate of change
- Couple the two components in the spirit of *physics informed machine learning*

Neural ODEs



Autoencoding structure in time. ODE in latent space with drift parametrised by a NN. Efficient evaluation of gradients by Pontryagin adjoint.

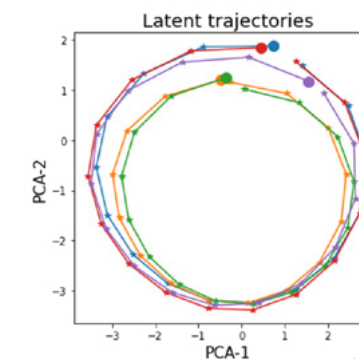


A (somewhat contrived) example



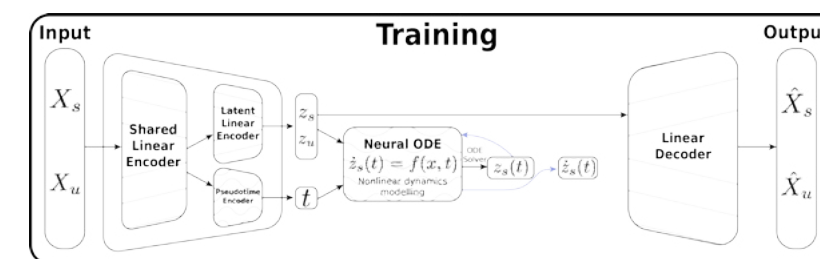


A (somewhat contrived) example





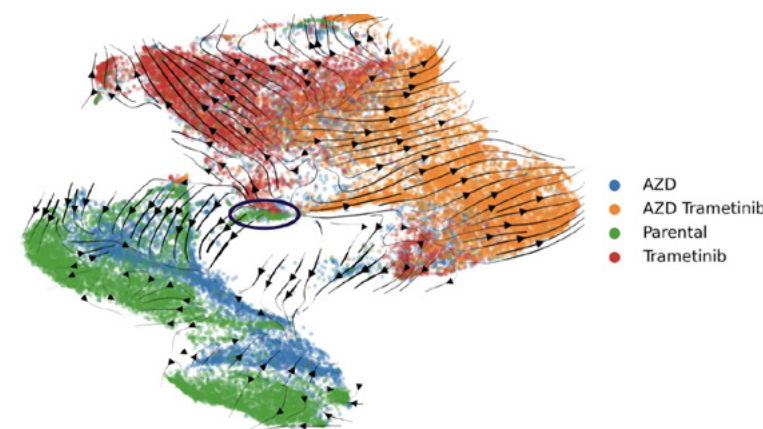
NeuroVelo (Idris Kouadri Boudjelthia)



$$\mathcal{L} = \text{MSE}(X, \hat{X}) + \text{MSE}(\dot{z}_s, \beta z_u - \gamma z_s)$$

Because the encoding/ decoding is linear, the RNA velocity equations apply also in latent space. Notice we need no assumptions on the transcription rate function.

NeuroVelo on CRC

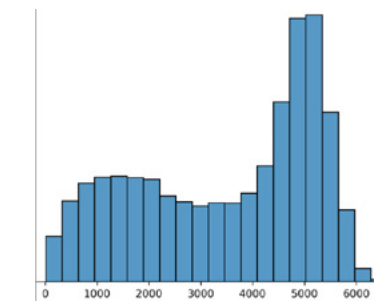


### Interpreting Neurovelo

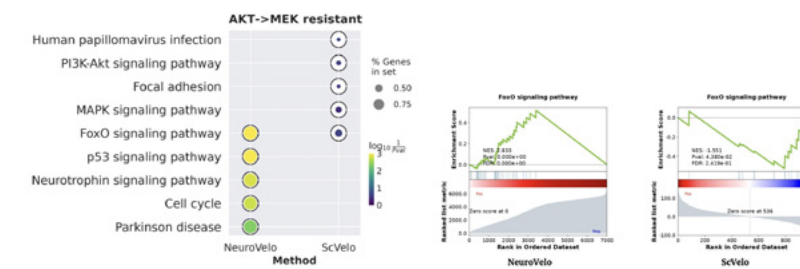
- NeuroVelo learns a low-dimensional nonlinear dynamical system
- Principal dynamics are given (locally) by the *eigenvectors* of the Jacobian matrix
- These eigenvectors can be decoded linearly to give a ranked list of genes
- Robustness is ensured by computing a stability index w.r.t. multiple initializations

## Interpreting Neurovelo cont'd

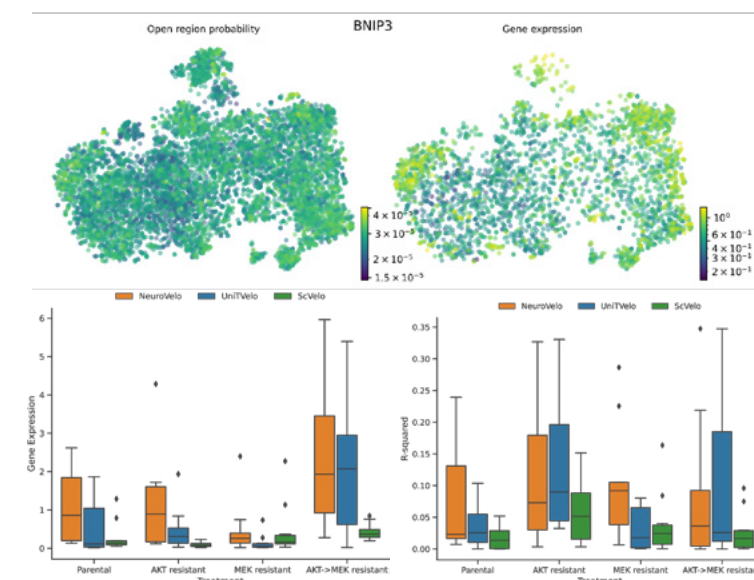
- Noise genes should have ranks uniformly distributed  
← Gaussian average
- Relevant genes should have consistently high ranks
- Expect bimodal distribution



## Validating NeuroVelo: enrichment



Validating NeuroVelo: multiome



### Thanks!

Collaborators and lab members/ alumni

#### SISSA

Riccardo Margiotta  
Rongrong Xie  
Viplove Arora

#### Nour el Kazwini

Alex Zhang  
**Idris Kouadri**  
**Boudjelthia**  
Federico Caretti

Katsiaryna  
Davydzenka

#### University of Edinburgh

Kashyap Chhatbar  
Kaan Ocal  
**Christos Maniatis**  
**Andreas Kapourani**

**Yuanhua Huang**  
**Catalina Vallejos**

#### Human Technopole

Andrea Sottoriva  
Salvatore Milite

Funding: ERC, AIRC, SISSA/ MUR

Thanks!

Thank You



### **Delitti in materia di violazione del diritto d'autore (Art. 25-novies, D.Lgs. n. 231/2001) [articolo aggiunto dalla L. n. 99/2009]**

- Messa a disposizione del pubblico, in un sistema di reti telematiche, mediante connessioni di qualsiasi genere, di un'opera dell'ingegno protetta, o di parte di essa (art. 171, legge n.633/1941 comma 1 lett. a) bis)
- Reati di cui al punto precedente commessi su opere altrui non destinate alla pubblicazione qualora ne risulti offeso l'onore o la reputazione (art. 171, legge n.633/1941 comma 3)
- Abusiva duplicazione, per trarne profitto, di programmi per elaboratore; importazione, distribuzione, vendita o detenzione a scopo commerciale o imprenditoriale o concessione in locazione di programmi contenuti in supporti non contrassegnati dalla SIAE; predisposizione di mezzi per rimuovere o eludere i dispositivi di protezione di programmi per elaboratori (art. 171-bis legge n.633/1941 comma 1)
- Riproduzione, trasferimento su altro supporto, distribuzione, comunicazione, presentazione o dimostrazione in pubblico, del contenuto di una banca dati; estrazione o reimpiego della banca dati; distribuzione, vendita o concessione in locazione di banche di dati (art. 171-bis legge n.633/1941 comma 2)
- Abusiva duplicazione, riproduzione, trasmissione o diffusione in pubblico con qualsiasi procedimento, in tutto o in parte, di opere dell'ingegno destinate al circuito televisivo, cinematografico, della vendita o del noleggio di dischi, nastri o supporti analoghi o ogni altro supporto contenente fonogrammi o videogrammi di opere musicali, cinematografiche o audiovisive assimilate o sequenze di immagini in movimento; opere letterarie, drammatiche, scientifiche o didattiche, musicali o drammatico musicali, multimediali, anche se inserite in opere collettive o composite o banche dati; riproduzione, duplicazione, trasmissione o diffusione abusiva, vendita o commercio, cessione a qualsiasi titolo o importazione abusiva di oltre cinquanta copie o esemplari di opere tutelate dal diritto d'autore e da diritti connessi; immissione in un sistema di reti telematiche, mediante connessioni di qualsiasi genere, di un'opera dell'ingegno protetta dal diritto d'autore, o parte di essa (art. 171-ter legge n.633/1941)
- Mancata comunicazione alla SIAE dei dati di identificazione dei supporti non soggetti al contrassegno o falsa dichiarazione (art. 171-septies legge n.633/1941)
- Fraudolenta produzione, vendita, importazione, promozione, installazione, modifica, utilizzo per uso pubblico e privato di apparati o parti di apparati atti alla decodificazione di trasmissioni audiovisive ad accesso condizionato effettuate via etere, via satellite, via cavo, in forma sia analogica sia digitale (art. 171-octies legge n.633/1941).

**[Torna all'inizio](#)**