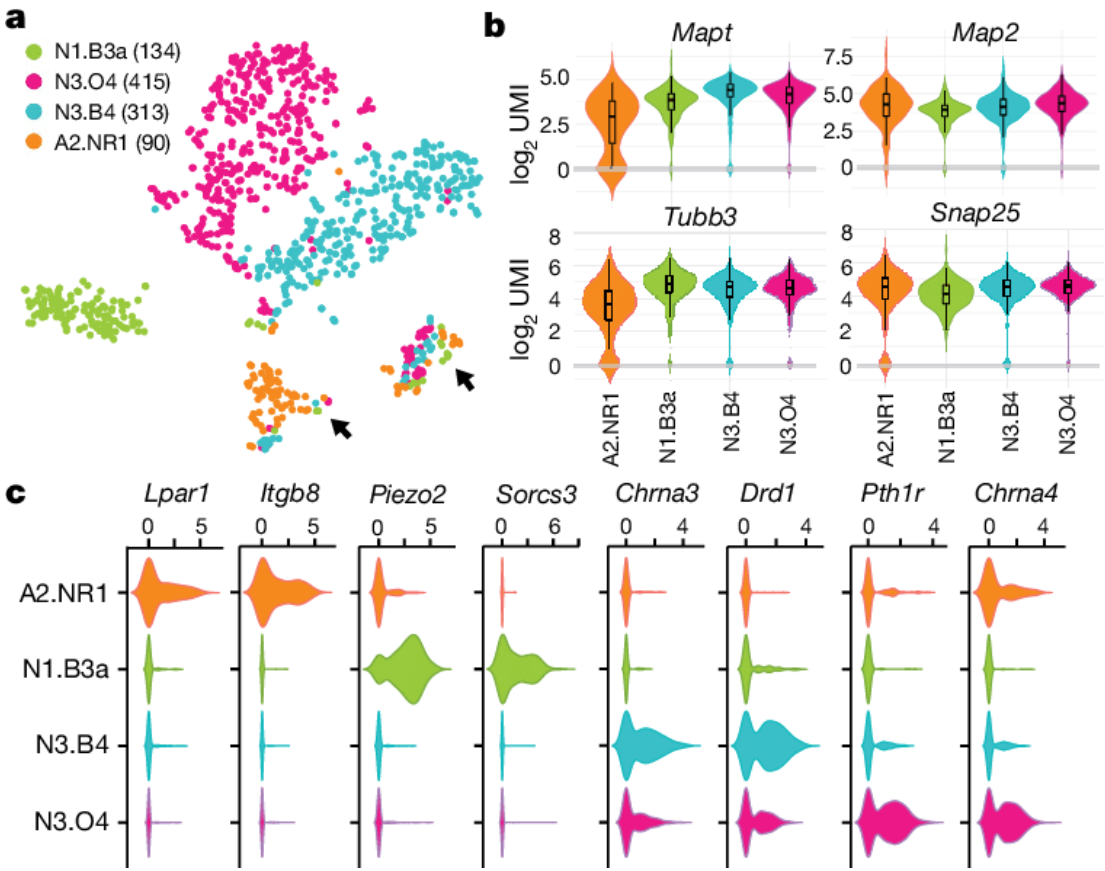
SCNet: Automatic Multi-Channel Genome Network Inference from Single-Cell RNA Sequences

Shiyin Wang (wangshiy16@mails.tsinghua.edu.cn)

December 23, 2019

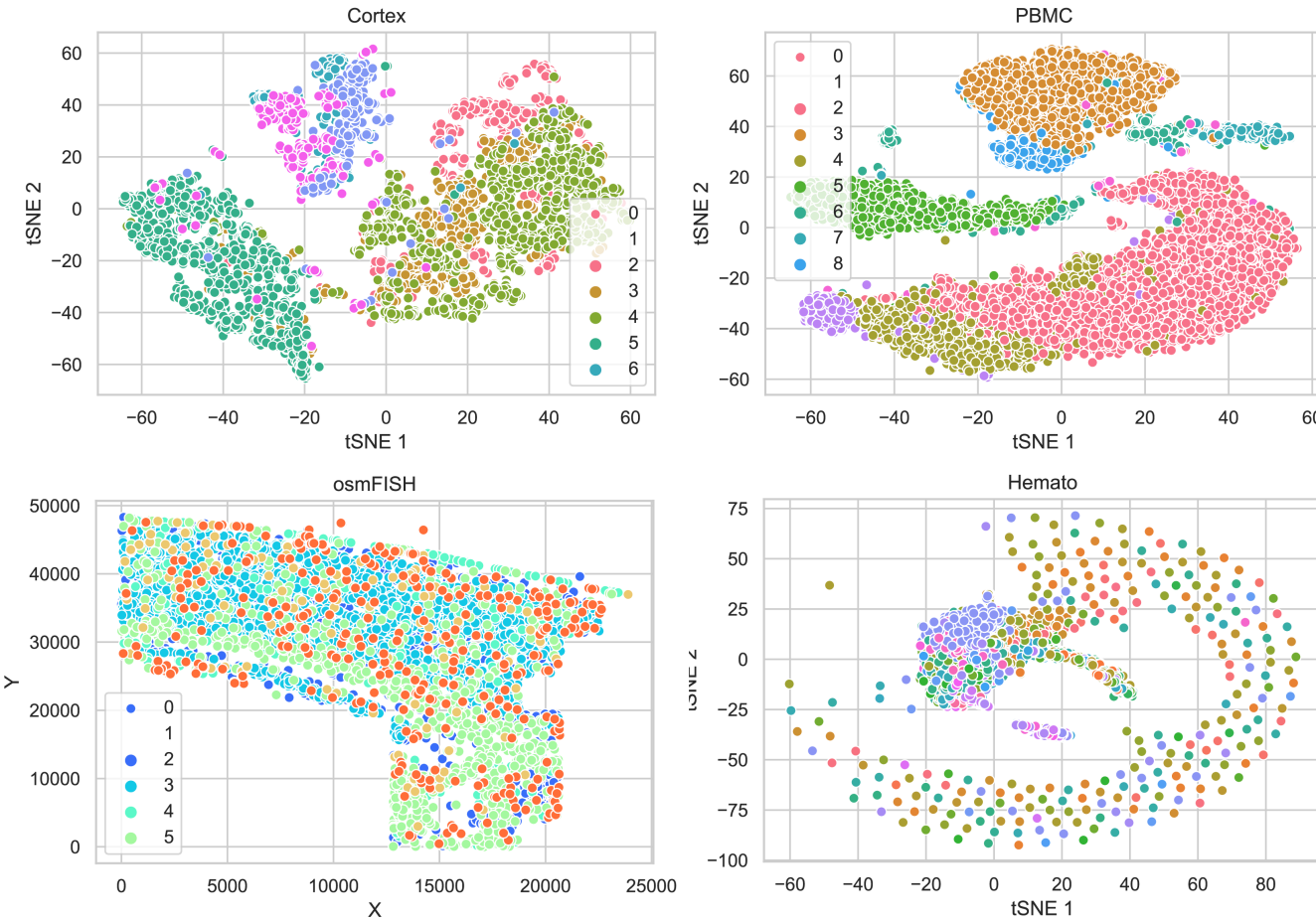
Outline

What is single-cell RNA sequencing?



Tsunemoto, Rachel, et al. "Diverse reprogramming codes for neuronal identity." Nature 557.7705 (2018): 375.

Datasets


Cortex
tSNE 1
tSNE 2
PBMC
tSNE 1
tSNE 2
osmFISH
X
Y
Hemato
tSNE 1

Bayesian Methods

Likelihood

How probable is the evidence given that our hypothesis is true?

Prior

How probable was our hypothesis *before* observing the evidence?

$$\mathrm{P}(H \mid e) = \frac{\mathrm{P}(e \mid H)\,\mathrm{P}(H)}{\mathrm{P}(e)}$$

Posterior

How probable is our hypothesis given the observed evidence? (Not directly computable)

Marginal

How probable is the new evidence under all possible hypotheses?

$\mathrm{P}(e) = \sum \mathrm{P}(e \mid H_i)\,\mathrm{P}(H_i)$

Markov Chain Monte Carlo

- Initialise $x^{(0)}$.
- For $i = 0$ to $N - 1$
  - Sample $u \sim U_{[0,1]}$.
  - Sample $x^\star \sim q(x^\star | x^{(i)})$.
  - If $u < A(x^{(i)}, x^\star) = \min\left\{1, \frac{p(x^\star)q(x^{(i)}|x^\star)}{p(x^{(i)})q(x^\star|x^{(i)})}\right\}$

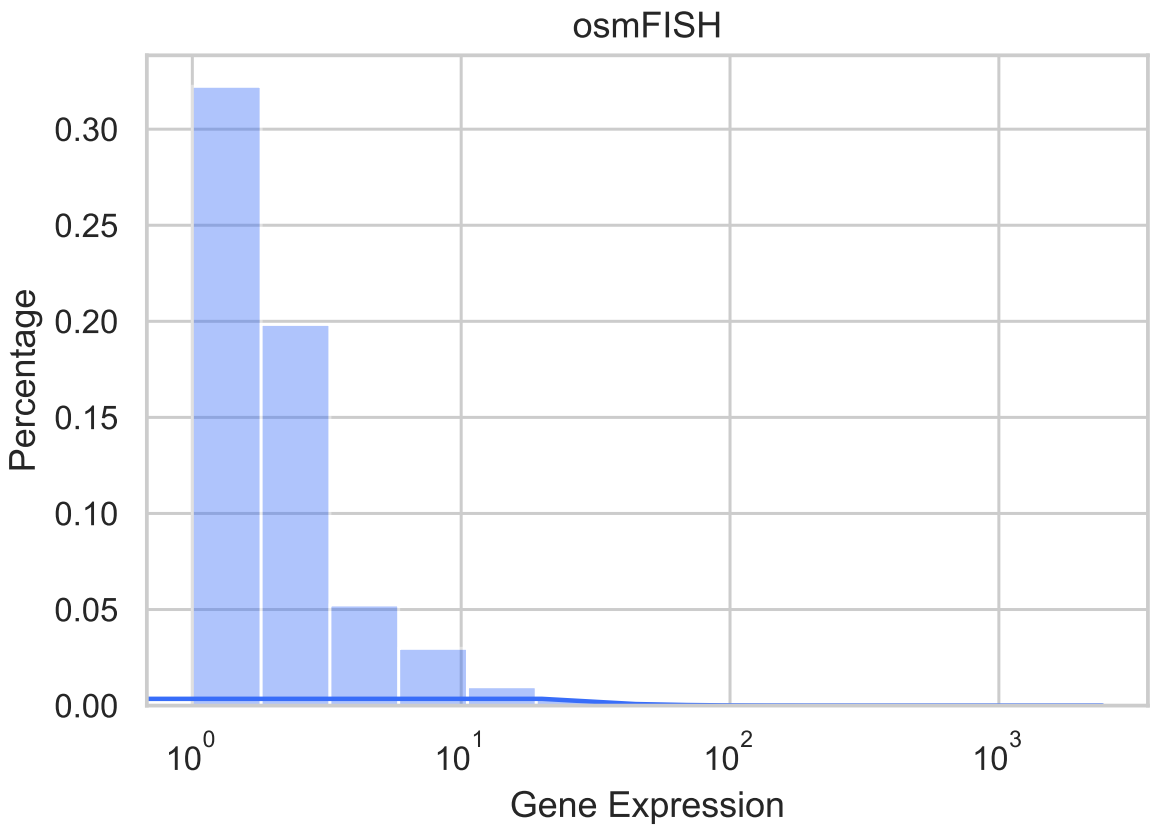
    $$x^{(i+1)} = x^\star$$

    else

    $$x^{(i+1)} = x^{(i)}$$

Slide in CPSC 540, taught in 2013 at UBC by Nando de Freitas

Choose Log-Scale to Model Gene Expression



Bayesian Hierarchical Linear Model

X and Y are the expression profiles of two genes. The edge weights in the desired network correspond to the regression coefficient k in the equation.

$$\begin{aligned} logY &= klogX + \epsilon \\ k &\sim N(\beta, \sigma) \\ \epsilon &\sim N(0, \gamma^2) \end{aligned} \tag{1}$$

This model estimates k from single-cell RNA sequencing records.

Define Transition Probability through Edge Weights

Once we have a weighted network, we can define the association probability between two nodes X and Y by a very trivial model.

$$Pr(X \rightarrow Y) = 1 - (1 - w_{X,Y})\Pi_{a \in V}(1 - w_{X,a}w_{a,Y})\Pi_{a \in V, b \in V}(1 - w_{X,a}w_{a,b}w_{b,Y}) \quad (2)$$

For simplicity, I expanded the search for three steps. Walk length is flexible to choose.

Maximal Likelihood Optimization

Bayesian hierarchical linear model and transition probability explain the network dynamics from two different angles. Now we can combine them together to infer the edge weights of networks (V, E, W).

$$maximize\ L(W;_k, \Sigma_k) = \sum_{v_1 \in V} \sum_{v_2 \in V} Pr_{N(_k, \Sigma_k)}(w = Pr(v_1 \to v_2)) \tag{3}$$

Add regulation term to restrain the number of edges in the network.

$$maximize\ L(W;_k, \Sigma_k) = \sum_{v_1 \in V} \sum_{v_2 \in V} Pr_{N(_k, \Sigma_k)}(w = Pr(v_1 \to v_2)) - \lambda |V| \tag{4}$$

Markov Chain Monte Carlo

Possible Directions

- Integrate existing knowledge from protein-protein interaction networks (STRING, OmniPATH, ConsensusPath, etc) as priors
- Design better probability models
- Make interactive transition videos

Resources

- 10x Genomics: Datasets providing single cell and spatial views of biological systems (`https://www.10xgenomics.com`)

Questions & Answers