Introduction
OO
OO

Methods
OOOO
O

Results

Future Work
OO

Q & A
O

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

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

December 23, 2019

Outline

1 Introduction
- Single-Cell RNA Sequencing
- Mathematics Preliminaries

2 Methods
- Probabilistic models
- Optimization

3 Results

4 Future Work

5 Q & A

Introduction
Methods
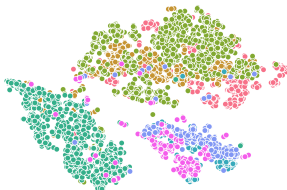Results
Future Work
Q & A

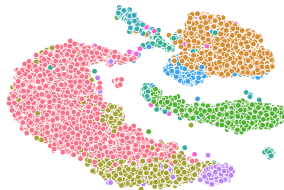Single-Cell RNA Sequencing

What is single-cell RNA sequencing?
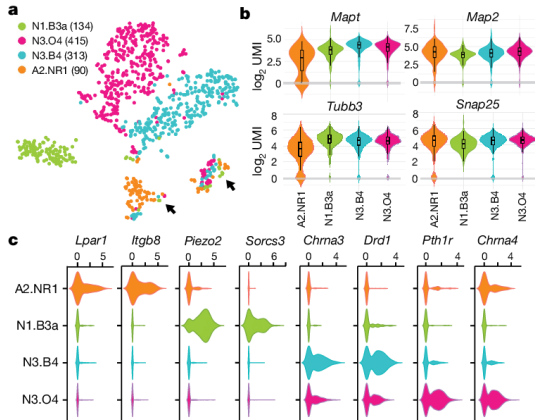
Tsunemoto, Rachel, et al. "Diverse reprogramming codes for neuronal identity." Nature 557.7705 (2018): 375.
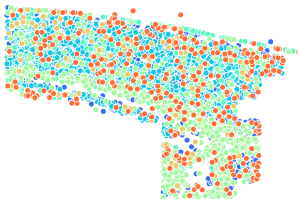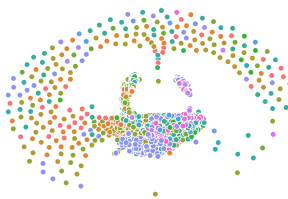
Datasets

Cortex (3005 cells, 55g genes, 7 types)

PBMC (11990 cells, 3345 genes, 9 types)

osmFISH (4530 cells, 33 genes, 6 types)

Hemato (4016 cells, 7397 genes, 101 types)

Introduction	Methods	Results	Future Work	Q & A
OO	OOOO		OO	O
●O	O			

Mathematics Preliminaries

Bayesian Methods

Introduction	Methods	Results	Future Work	Q & A
○○	○○○○		○○	○
○●	○			

Mathematics Preliminaries

Markov Chain Monte Carlo
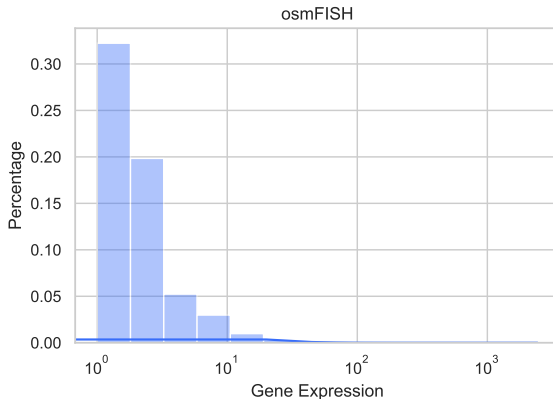
▶ Initialise $x^{(0)}$.

▶ 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

Introduction
○○
○○

Methods
●○○○
○

Results

Future Work
○○

Q & A
○

Probabilistic models

Choose Log-Scale to Model Gene Expression

The support of X and Y is a non-negative, we use $log(X + 1)$ instead of $log(X)$ to map it onto the log-scale non-negative space.

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.

$$log(Y + 1) = klog(X + 1) + \epsilon$$
$$k \sim N(\beta, \sigma) \tag{1}$$
$$\epsilon \sim N(0, \gamma^2)$$

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.

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

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

Introduction	Methods	Results	Future Work	Q & A
○○	○○○●		○○	○
○○	○			

Probabilistic models

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).

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

Add regulation term $\lambda|V|$ to restrain the number of edges in the
network.

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

Introduction	Methods	Results	Future Work	Q & A
○○	○○○○		○○	○
○○	●			

Optimization

Data Preprocessing - Normalization

1 Retained the top genes ordered by variance as in [Lopez *et al.*, 2018][1]

2 Normalize genes to standard Gaussian distribution $N(0, 1)$

Reason for normalization to standard deviation: So that the regression coefficient in $Y = kX + \epsilon$ is $\hat{k} = \frac{cov(X,Y)}{var(X)} = cov(X, Y)$, which is exchangeable and bidirectional.

---

[1]Lopez, R., Regier, J., Cole, M.B. et al. Deep generative modeling for single-cell transcriptomics. Nat Methods 15, 1053–1058 (2018) doi:10.1038/s41592-018-0229-2

Possible Directions

- Theoretical analysis and case study to compare with other methods that derive networks from covariance matrix directly
- Integrate existing knowledge from protein-protein interaction networks (STRING, OmniPATH, ConsensusPath, etc) as priors
- Design better probability models
- Make interactive transition videos

Introduction
oo
oo

Methods
oooo
o

Results

Future Work
o●

Q & A
o

Resources

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

Introduction
OO
OO

Methods
OOOO
O

Results

Future Work
OO

Q & A
●

Questions & Answers