CS 5854: Benchmarking Algorithms for Gene Regulatory Network Inference from Single-Cell Transcriptomic Data

T. M. Murali

February 20, 22, 27, March 1, 2023

Plethora of Computational Analysis Techniques

Mark Robinson
@markrobinsonca

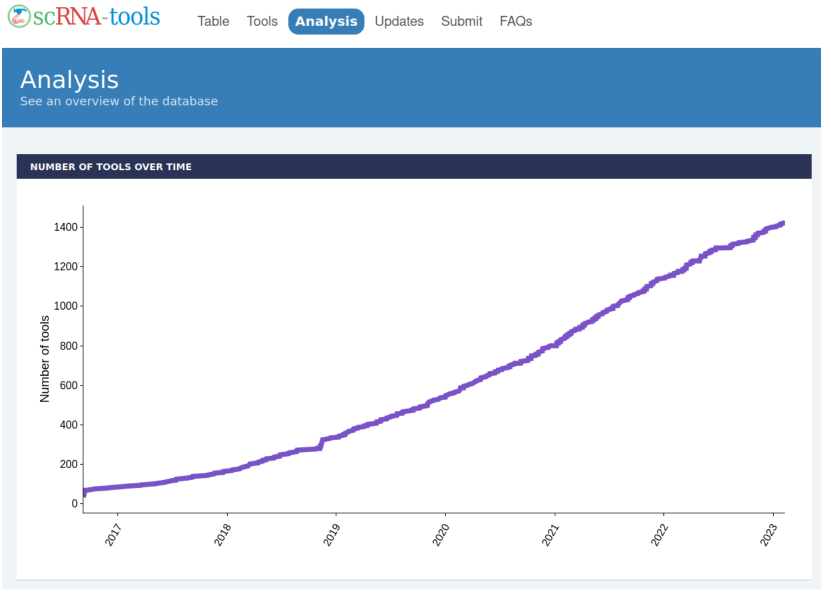
Dear #academictwitter, I'm looking for a list of benchmarks for single cell RNA-seq data analysis. Generally, so normalization, DE, clustering, cell assignment, etc. I'm OFC aware of the "Methods comparisons" section in @seandavis12's awesome list (bit.ly/2Rhq1fY) ..

seandavi/awesome-single-cell
Community-curated list of software packages and data resources for single-cell, including RNA-seq, ATAC-seq, etc...
github.com

1:18 PM · Jan 22, 2020 · Twitter Web App

https://docs.google.com/spreadsheets/d/
1Gqn0eZ8oiNh8-9ovyh4D_ZsoWcKYoKTJ0C1AfrJrdRA/edit#gid=0

Plethora of Computational Analysis Techniques



Analysis

See an overview of the database

NUMBER OF TOOLS OVER TIME

Plethora of Computational Analysis Techniques

Technology Feature | Published: 20 February 2020

Bench pressing with genomics benchmarkers

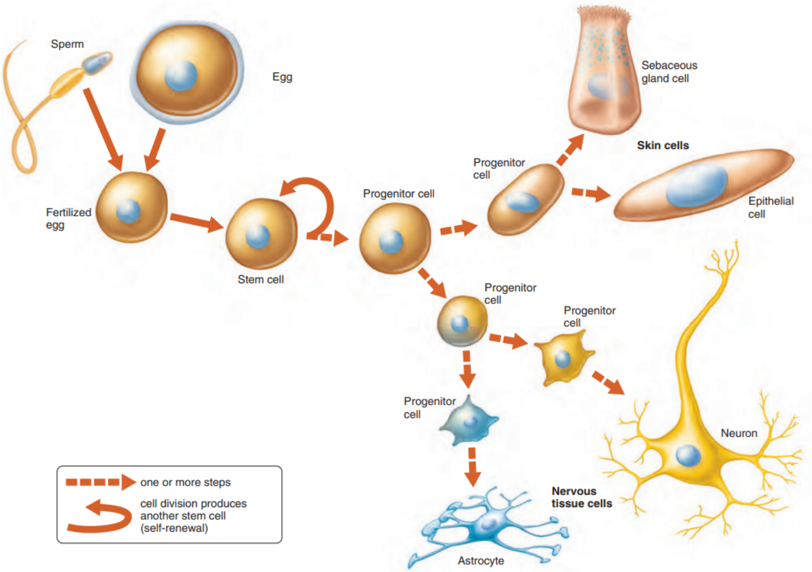
Vivien Marx

Nature Methods **17**, 255–258 (2020) | Cite this article

5868 Accesses | **8** Citations | **59** Altmetric | Metrics

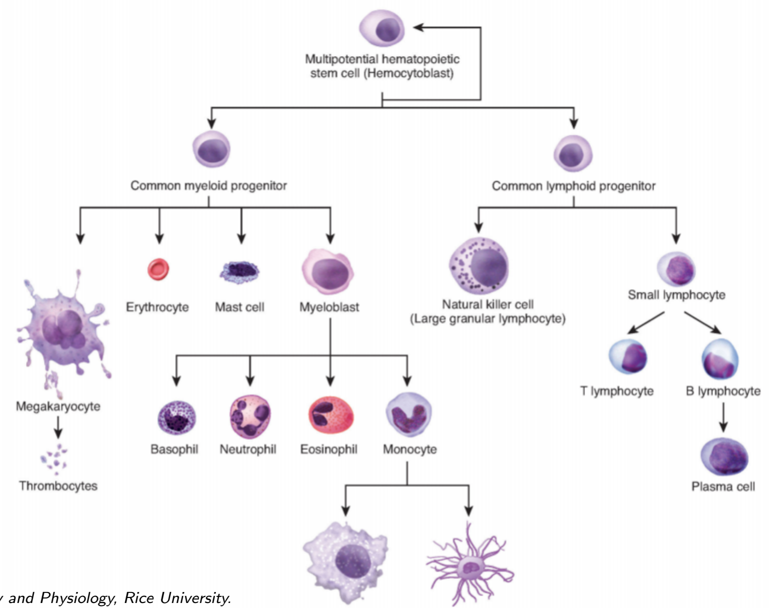
Some -omics tools can be more accurate, sensitive or efficient than others. Yet benchmarking is no tell-all.

Cellular Differentiation



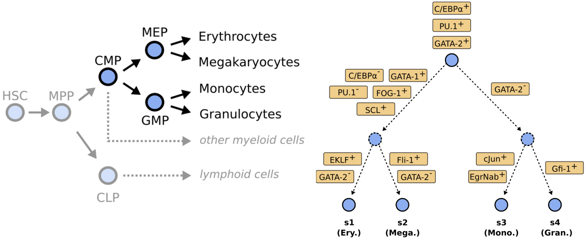
Shier, *et al.*, (2015) "*Hole's Essentials of Human Anatomy and Physiology*", McGraw-Hill

Cellular Differentiation



Anatomy and Physiology, Rice University.

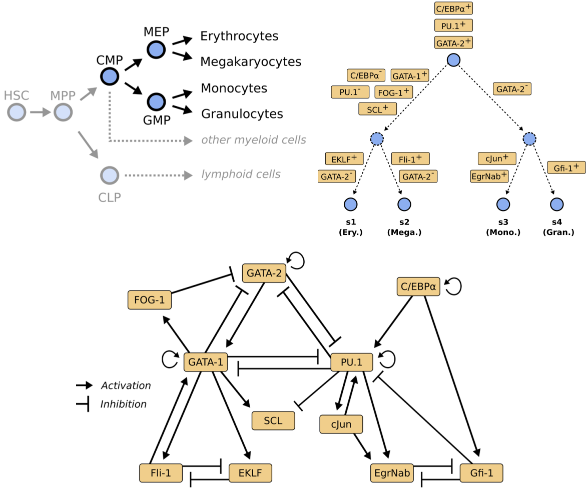
Cellular Differentiation



- Cells in different states express different sets of genes.
- Cells move from one “state” to another.

Krumsiek et al. (2010). “Hierarchical Differentiation of Myeloid Progenitors...” PLoS ONE

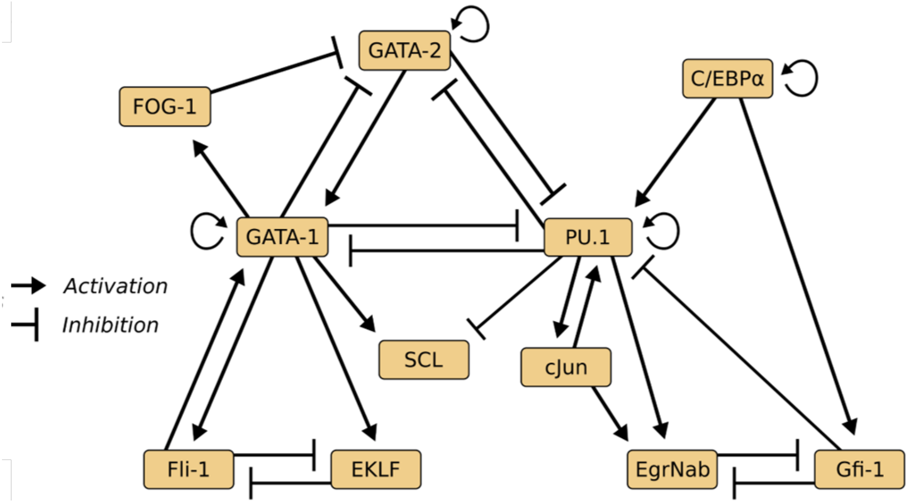
Cellular Differentiation



- Transcription factors activate/inhibit genes to effect cell transition from one state to another.

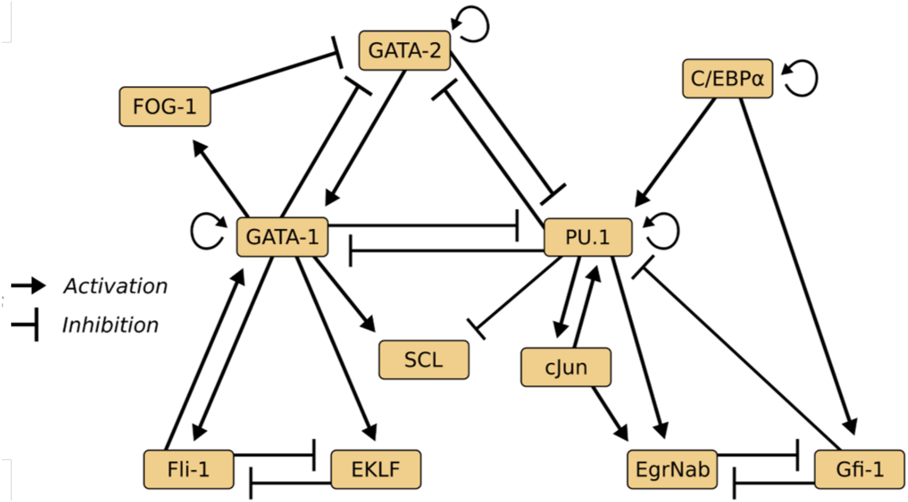
Krumsiek et al. (2010). "Hierarchical Differentiation of Myeloid Progenitors..." PLoS ONE

Gene Regulatory Network (GRN)



Krumsiek et al. (2010). "Hierarchical Differentiation of Myeloid Progenitors..." PLoS ONE

Gene Regulatory Network (GRN)

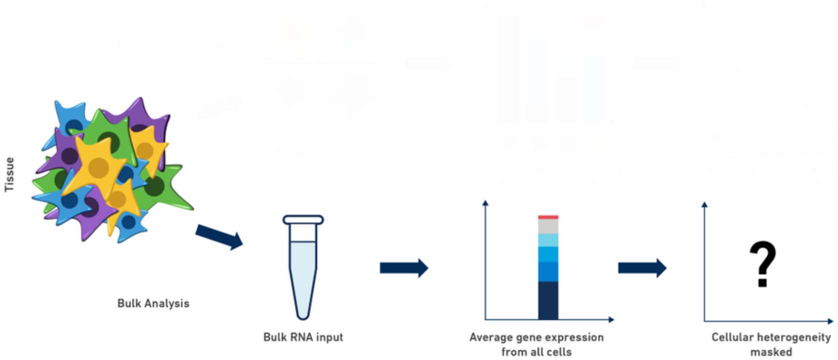


How do we build GRNs using computational techniques?

Krumsiek et al. (2010). "Hierarchical Differentiation of Myeloid Progenitors..." PLoS ONE

Bulk RNA Sequencing

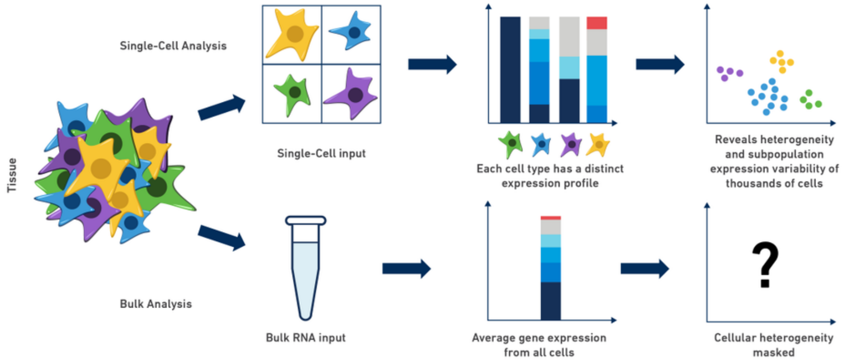
- A population of cells isolated at the same time may correspond to multiple, distinct intermediate differentiation states.
- Averages gene expression and masks cellular heterogeneity.
- Difficult to experimentally purify cells in intermediate states.



10x Genomics

Single-cell RNA Sequencing (scRNA-seq)

- Produce thousands of independent measurements.
- Computational ordering of cells along "lineages" provide a high-resolution "pseudotemporal" view of gene expression kinetics.
- Richness of these datasets may facilitate inference.



10x Genomics

Trapnell et al., "The dynamics and regulators of cell fate decisions ...", Nat. Biotech., 2014.

Over a Dozen Methods Have Already Been Developed

	Properties	
	Category	Addl. Inputs
PIDC	MI	–
GENIE3	RF	–
GRNBOOST2	RF	–
SCODE	ODE+Reg	ODE parameters
PPCOR	Corr	–
SINCERITIES	Reg	–
SCRIBE	MI	Type of RDI
SINGE	GC	Regression parameters
LEAP	Corr	Lag
GRISLI	ODE+Reg	Regression parameters
GRNVBEM	Reg	–
SCNS	Bool	Boolean model

Over a Dozen Methods Have Already Been Developed

	Properties	
	Category	Addl. Inputs
PIDC	MI	–
GENIE3	RF	–
GRNBOOST2	RF	–
SCODE	ODE+Reg	ODE parameters
PPCOR	Corr	–
SINCERITIES	Reg	–
SCRIBE	MI	Type of RDI
SINGE	GC	Regression parameters
LEAP	Corr	Lag
GRISLI	ODE+Reg	Regression parameters
GRNVBEM	Reg	–
SCNS	Bool	Boolean model

Over a Dozen Methods Have Already Been Developed

	Properties	
	Category	Addl. Inputs
PIDC	MI	-
GENIE3	RF	-
GRNBOOST2	RF	-
SCODE	ODE+Reg	ODE parameters
PPCOR	Corr	-
SINCERITIES	Reg	-
SCRIBE	MI	Type of RDI
SINGE	GC	Regression parameters
LEAP	Corr	Lag
GRISLI	ODE+Reg	Regression parameters
GRNVBEM	Reg	-
SCNS	Bool	Boolean model

Over a Dozen Methods Have Already Been Developed

	Properties	
	Category	Addl. Inputs
PIDC	MI	–
GENIE3	RF	–
GRNBOOST2	RF	–
SCODE	ODE+Reg	ODE parameters
PPCOR	Corr	–
SINCERITIES	Reg	–
SCRIBE	MI	Type of RDI
SINGE	GC	Regression parameters
LEAP	Corr	Lag
GRISLI	ODE+Reg	Regression parameters
GRNVBEM	Reg	–
SCNS	Bool	Boolean model

Over a Dozen Methods Have Already Been Developed

	Properties		
	Category	Addl. Inputs	Time ordered?
PIDC	MI	–	✗
GENIE3	RF	–	✗
GRNBOOST2	RF	–	✗
SCODE	ODE+Reg	ODE parameters	✓
PPCOR	Corr	–	✗
SINCERITIES	Reg	–	✓
SCRIBE	MI	Type of RDI	✓
SINGE	GC	Regression parameters	✓
LEAP	Corr	Lag	✓
GRISLI	ODE+Reg	Regression parameters	✓
GRNVBEM	Reg	–	✓
SCNS	Bool	Boolean model	✓

Over a Dozen Methods Have Already Been Developed

	Properties			
	Category	Addl. Inputs	Time ordered?	Directed?
PIDC	MI	–	✗	✗
GENIE3	RF	–	✗	✓
GRNBOOST2	RF	–	✗	✓
SCODE	ODE+Reg	ODE parameters	✓	✓
PPCOR	Corr	–	✗	✗
SINCERITIES	Reg	–	✓	✓
SCRIBE	MI	Type of RDI	✓	✓
SINGE	GC	Regression parameters	✓	✓
LEAP	Corr	Lag	✓	✓
GRISLI	ODE+Reg	Regression parameters	✓	✓
GRNVBEM	Reg	–	✓	✓
SCNS	Bool	Boolean model	✓	✓

Over a Dozen Methods Have Already Been Developed

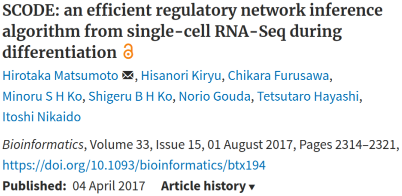
	Properties				
	Category	Addl. Inputs	Time ordered?	Directed?	Signed?
PIDC	MI	–	✗	✗	✗
GENIE3	RF	–	✗	✓	✗
GRNBOOST2	RF	–	✗	✓	✗
SCODE	ODE+Reg	ODE parameters	✓	✓	✓
PPCOR	Corr	–	✗	✗	✓
SINCERITIES	Reg	–	✓	✓	✓
SCRIBE	MI	Type of RDI	✓	✓	✗
SINGE	GC	Regression parameters	✓	✓	✗
LEAP	Corr	Lag	✓	✓	✗
GRISLI	ODE+Reg	Regression parameters	✓	✓	✗
GRNVBEM	Reg	–	✓	✓	✓
SCNS	Bool	Boolean model	✓	✓	✓

Over a Dozen Methods Have Already Been Developed

	Properties				
	Category	Addl. Inputs	Time ordered?	Directed?	Signed?
PIDC	MI	–	✗	✗	✗
GENIE3	RF	–	✗	✓	✗
GRNBOOST2	RF	–	✗	✓	✗
SCODE	ODE+Reg	ODE parameters	✓	✓	✓
PPCOR	Corr	–	✗	✗	✓
SINCERITIES	Reg	–	✓	✓	✓
SCRIBE	MI	Type of RDI	✓	✓	✗
SINGE	GC	Regression parameters	✓	✓	✗
LEAP	Corr	Lag	✓	✓	✗
GRISLI	ODE+Reg	Regression parameters	✓	✓	✗
GRNVBEM	Reg	–	✓	✓	✓
SCNS	Bool	Boolean model	✓	✓	✓

How accurately do these methods infer GRNs?

Performance of Current Algorithms



SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation

Hirotaka Matsumoto ✉, Hisanori Kiryu, Chikara Furusawa, Minoru S H Ko, Shigeru B H Ko, Norio Gouda, Tetsutaro Hayashi, Itoshi Nikaido

Bioinformatics, Volume 33, Issue 15, 01 August 2017, Pages 2314–2321, https://doi.org/10.1093/bioinformatics/btx194

Published: 04 April 2017 **Article history ▾**

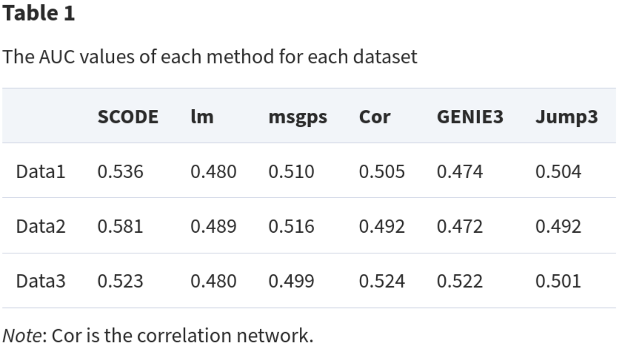


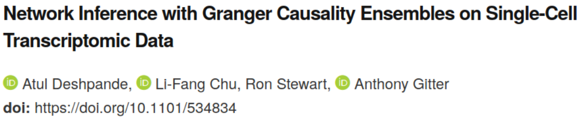
Table 1

The AUC values of each method for each dataset

	SCODE	lm	msgps	Cor	GENIE3	Jump3
Data1	0.536	0.480	0.510	0.505	0.474	0.504
Data2	0.581	0.489	0.516	0.492	0.472	0.492
Data3	0.523	0.480	0.499	0.524	0.522	0.501

Note: Cor is the correlation network.

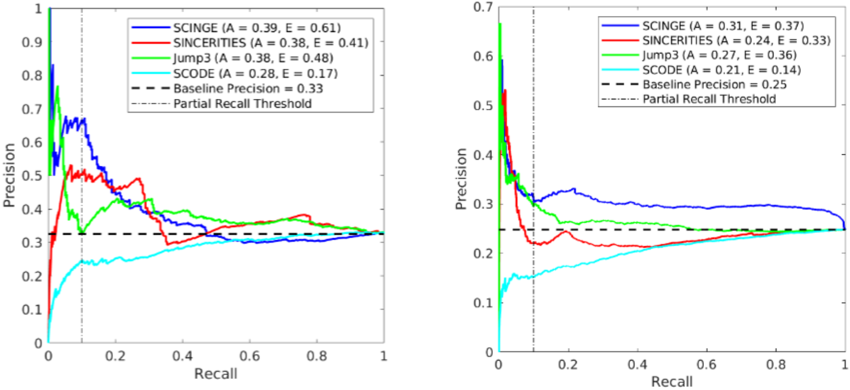
Performance of Current Algorithms



Network Inference with Granger Causality Ensembles on Single-Cell Transcriptomic Data

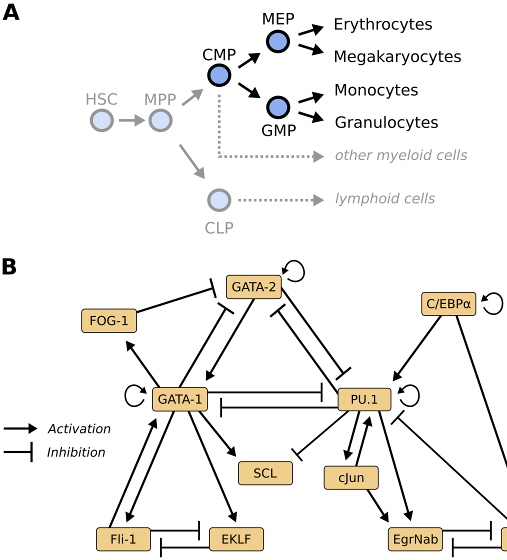
Atul Deshpande, Li-Fang Chu, Ron Stewart, Anthony Gitter

doi: https://doi.org/10.1101/534834



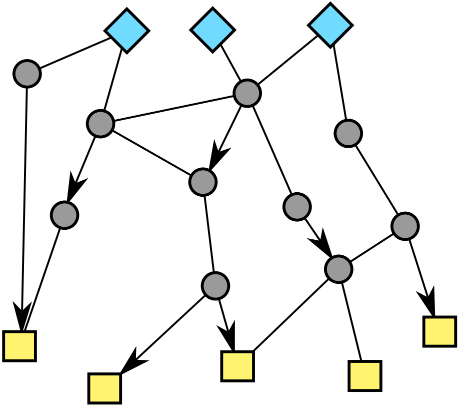
Performance is close to that of a random predictor!

What is a GRN?



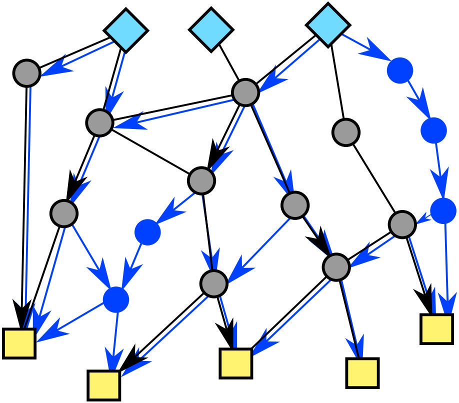
Krumsiek et al., Hierarchical differentiation of myeloid progenitors is encoded in the transcription factor network. PLoS One 6, 2011.

Evaluation of Inferred GRNs



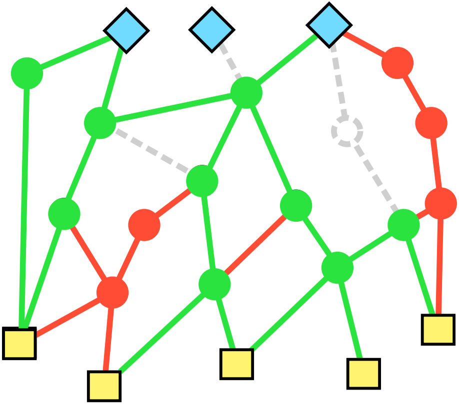
Ground-truth GRN

Evaluation of Inferred GRNs



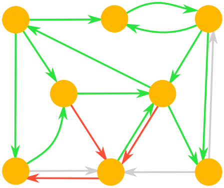
Ground-truth GRN and Inferred Reconstruction

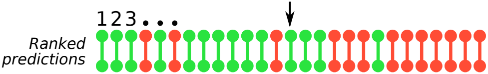
Evaluation of Inferred GRNs



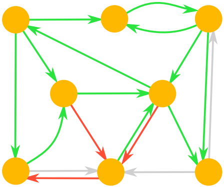
Ground-truth GRN and Inferred Reconstruction

Evaluation of Inferred GRNs





Evaluation of Inferred GRNs

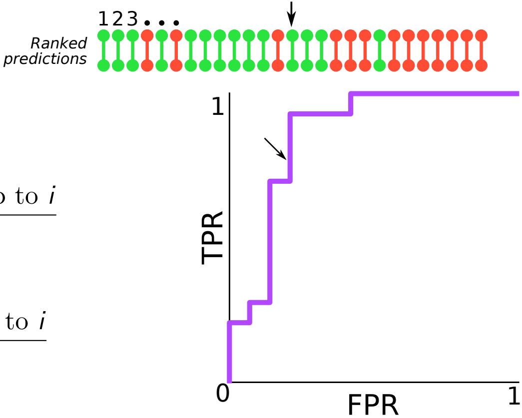


Specificity/False positive rate:

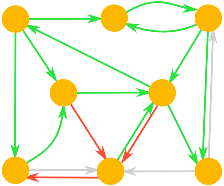
$$sp_i = \frac{\#\text{false positives up to } i}{\#\text{negatives}}$$

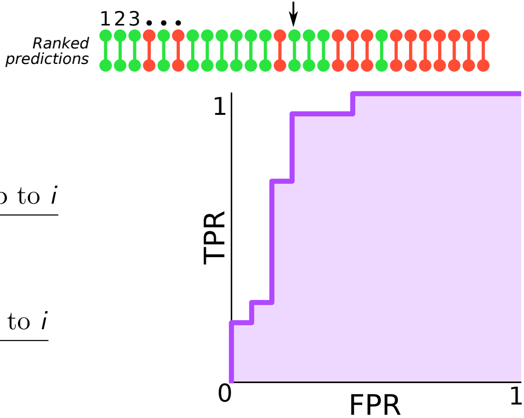
Recall/True positive rate:

$$r_i = \frac{\#\text{true positives up to } i}{\#\text{positives}}$$



Evaluation of Inferred GRNs





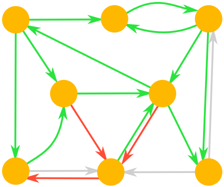
Specificity/False positive rate:

$$sp_i = \frac{\#\text{false positives up to } i}{\#\text{negatives}}$$

Recall/True positive rate:

$$r_i = \frac{\#\text{true positives up to } i}{\#\text{positives}}$$

Evaluation of Inferred GRNs

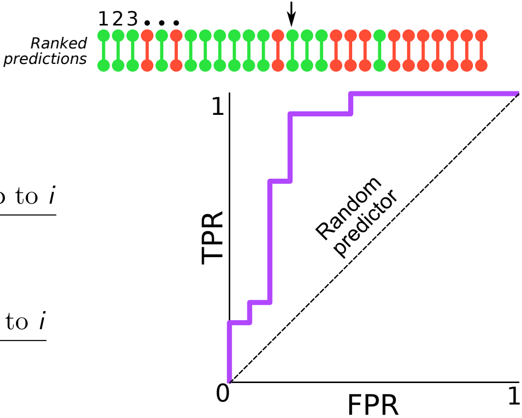


Specificity/False positive rate:

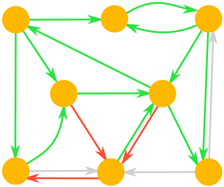
$$sp_i = \frac{\#\text{false positives up to } i}{\#\text{negatives}}$$

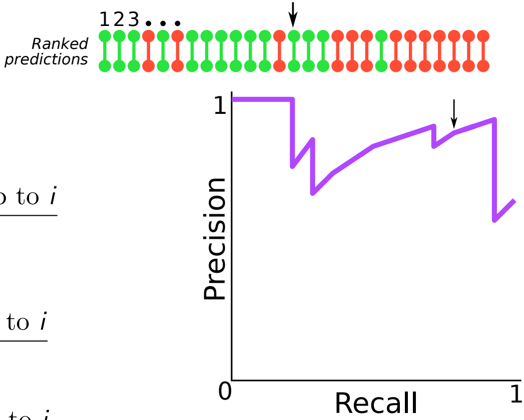
Recall/True positive rate:

$$r_i = \frac{\#\text{true positives up to } i}{\#\text{positives}}$$



Evaluation of Inferred GRNs





Specificity/False positive rate:

$$sp_i = \frac{\#\text{false positives up to } i}{\#\text{negatives}}$$

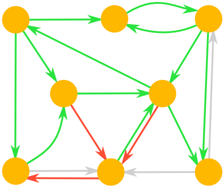
Recall/True positive rate:

$$r_i = \frac{\#\text{true positives up to } i}{\#\text{positives}}$$

Precision:

$$p_i = \frac{\#\text{true positives up to } i}{i}$$

Evaluation of Inferred GRNs



Specificity/False positive rate:

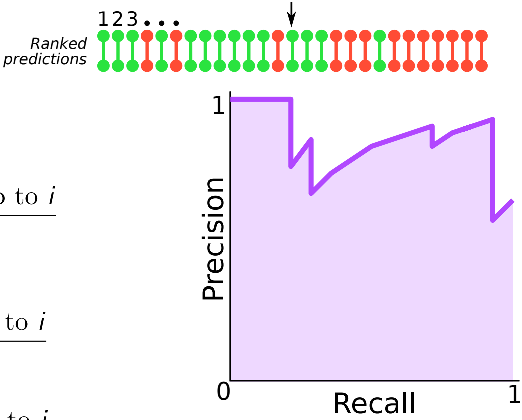
$$sp_i = \frac{\#\text{false positives up to } i}{\#\text{negatives}}$$

Recall/True positive rate:

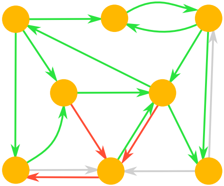
$$r_i = \frac{\#\text{true positives up to } i}{\#\text{positives}}$$

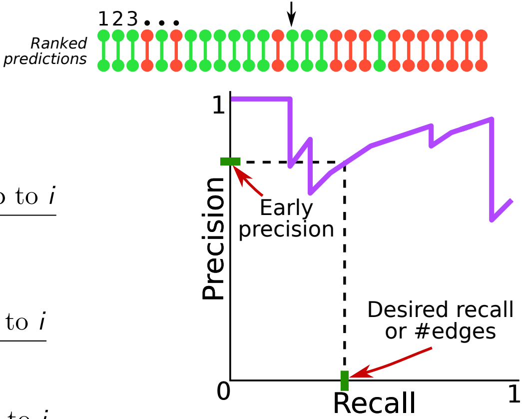
Precision:

$$p_i = \frac{\#\text{true positives up to } i}{i}$$



Evaluation of Inferred GRNs





Specificity/False positive rate:

$$sp_i = \frac{\#\text{false positives up to } i}{\#\text{negatives}}$$

Recall/True positive rate:

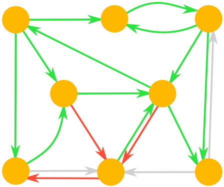
$$r_i = \frac{\#\text{true positives up to } i}{\#\text{positives}}$$

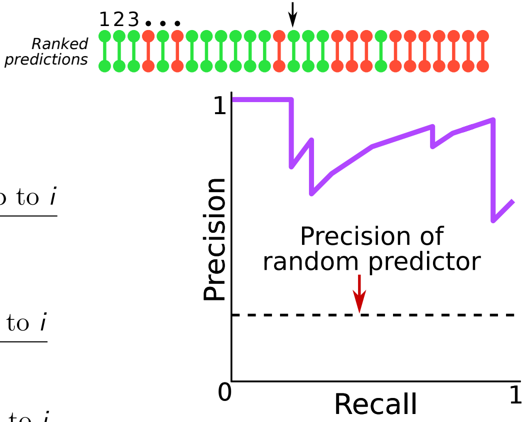
Precision:

$$p_i = \frac{\#\text{true positives up to } i}{i}$$

- What will a random predictor output? Random permutation of all possible edges.

Evaluation of Inferred GRNs





Specificity/False positive rate:

$$sp_i = \frac{\#\text{false positives up to } i}{\#\text{negatives}}$$

Recall/True positive rate:

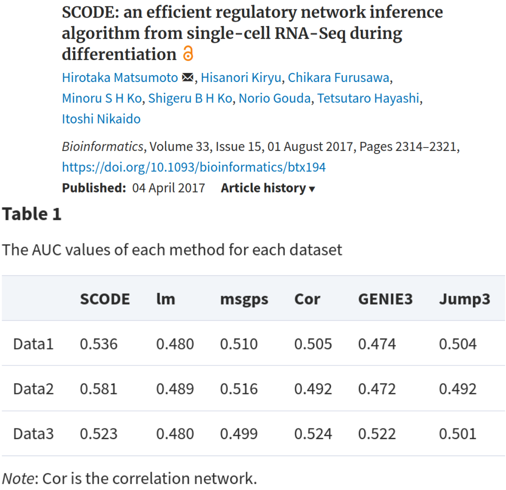
$$r_i = \frac{\#\text{true positives up to } i}{\#\text{positives}}$$

Precision:

$$p_i = \frac{\#\text{true positives up to } i}{i}$$

- What will a random predictor output? Random permutation of all possible edges.
- What is the precision of a random predictor? At any recall, ratio of #edges in ground-truth network and total number of possible edges.

Performance of Current Algorithms



SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation

Hirotaka Matsumoto ✉, Hisanori Kiryu, Chikara Furusawa, Minoru S H Ko, Shigeru B H Ko, Norio Gouda, Tetsutaro Hayashi, Itoshi Nikaido

Bioinformatics, Volume 33, Issue 15, 01 August 2017, Pages 2314–2321, https://doi.org/10.1093/bioinformatics/btx194

Published: 04 April 2017 **Article history** ▾

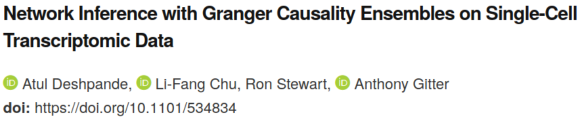
Table 1

The AUC values of each method for each dataset

	SCODE	lm	msgps	Cor	GENIE3	Jump3
Data1	0.536	0.480	0.510	0.505	0.474	0.504
Data2	0.581	0.489	0.516	0.492	0.472	0.492
Data3	0.523	0.480	0.499	0.524	0.522	0.501

Note: Cor is the correlation network.

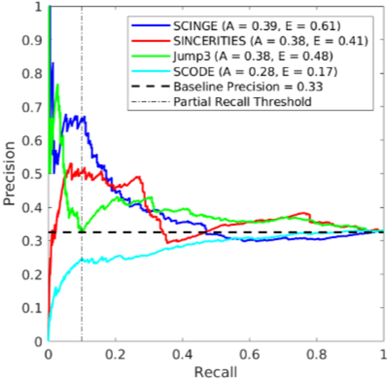
Performance of Current Algorithms

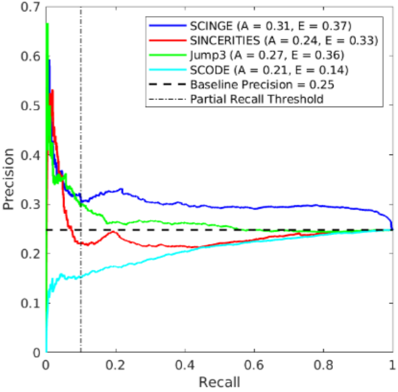


Network Inference with Granger Causality Ensembles on Single-Cell Transcriptomic Data

Atul Deshpande, Li-Fang Chu, Ron Stewart, Anthony Gitter

doi: https://doi.org/10.1101/534834





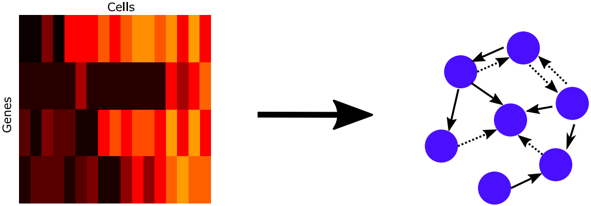
A = average precision, E = early precision

Performance of Current Algorithms

Performance of many algorithms is close to that of a random predictor!

Motivation for Benchmarking GRN Inference Methods

- Criteria for evaluation and comparison of methods vary from one paper to another.
- There are no standard ground-truth datasets.
- No existing framework for systematic comparison of methods.



scRNA-seq expression

Predicted GRN

BEELINE's Strategy

- How did BEELINE avoid these pitfalls (especially the second)?

BEELINE's Strategy

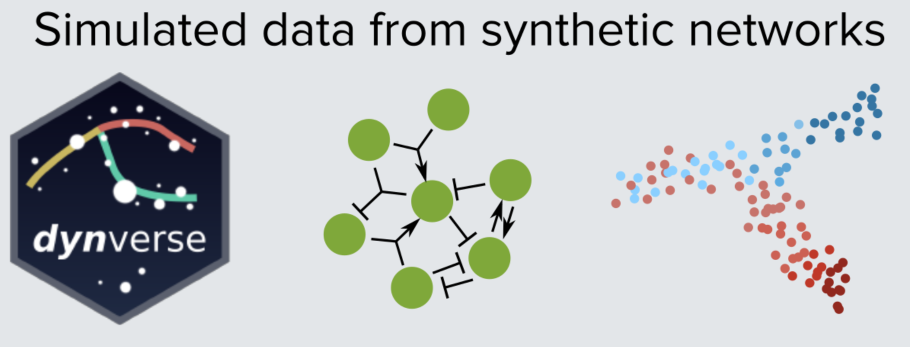
- How did BEELINE avoid these pitfalls (especially the second)?
- Used ground-truth networks with predictable trajectories.

BEELINE's Strategy

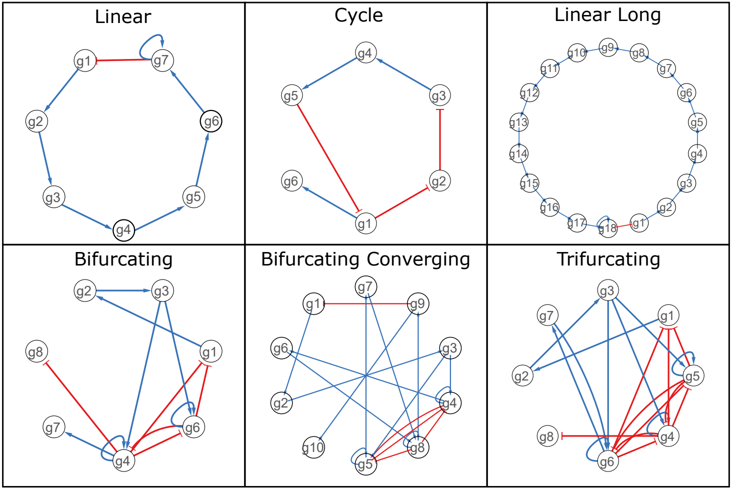
- How did BEELINE avoid these pitfalls (especially the second)?
- Used ground-truth networks with predictable trajectories.
  - Synthetic "toy" networks from Dynverse.
  - Boolean models curated from the literature.

Saelens et al., A comparison of single-cell trajectory inference methods Nat. Biotech., 2019

Input Type 1: Synthetic Networks

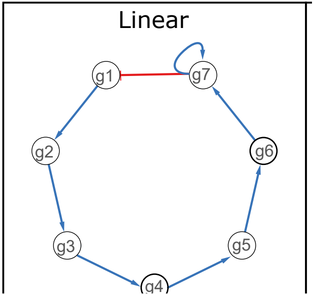


Dynamics of Synthetic Networks



- A gene is ON if and only if at least one activator is ON and every inhibitor is OFF. If there are no activators, a gene is ON iff every inhibitor is OFF.

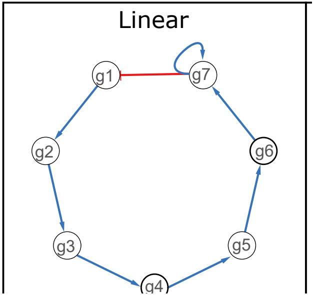
Dynamics of Synthetic Networks



Work out dynamics on your own and on the board.

- A gene is ON if and only if at least one activator is ON and every inhibitor is OFF. If there are no activators, a gene is ON iff every inhibitor is OFF.

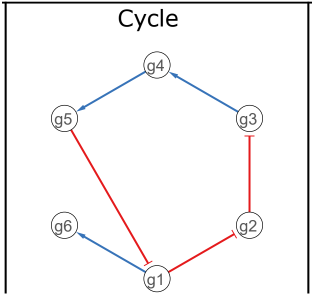
Dynamics of Synthetic Networks



Work out dynamics on your own and on the board.

- A gene is ON if and only if at least one activator is ON and every inhibitor is OFF. If there are no activators, a gene is ON iff every inhibitor is OFF.

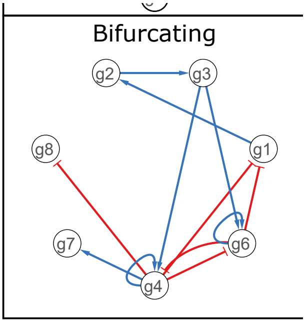
Dynamics of Synthetic Networks



Work out dynamics on your own and on the board.

- A gene is ON if and only if at least one activator is ON and every inhibitor is OFF. If there are no activators, a gene is ON iff every inhibitor is OFF.

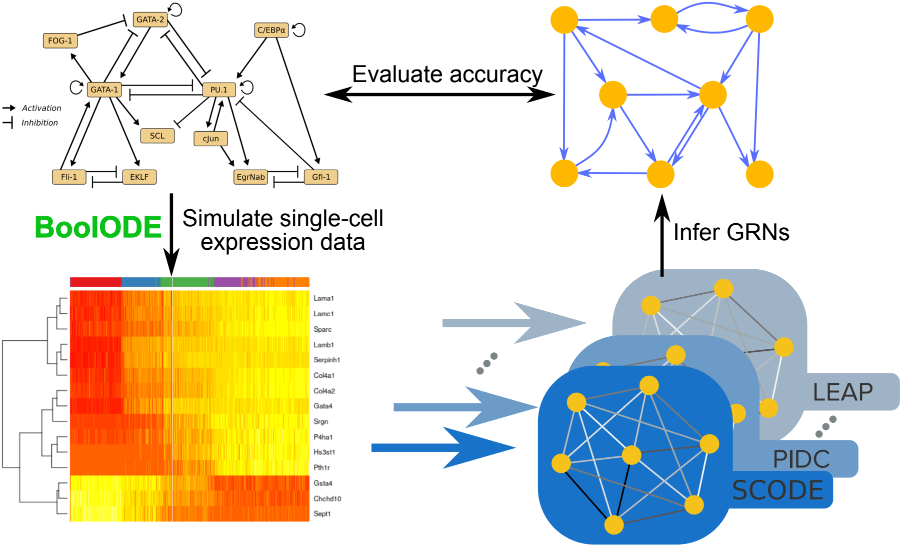
Dynamics of Synthetic Networks



Work out dynamics on your own and on the board.

- A gene is ON if and only if at least one activator is ON and every inhibitor is OFF. If there are no activators, a gene is ON iff every inhibitor is OFF.

Strategy for Ground-Truth Networks



Evaluate accuracy

BoolODE

Simulate single-cell expression data

Infer GRNs

BoolODE for Simulation

Read relevant part of "Methods" to answer these questions.

- What does BoolODE do?
- How long do we perform each simulation?
- How do we define (expression profile) for a cell?
- How did we create simulated datasets?
- Do we need to compute pseudotime?

BoolODE for Simulation

Read relevant part of "Methods" to answer these questions.

- What does BoolODE do? *Converts Boolean network into a system of stochastic differential equations and simulates it.*
- How long do we perform each simulation?
- How do we define (expression profile) for a cell?
- How did we create simulated datasets?
- Do we need to compute pseudotime?

BoolODE for Simulation

Read relevant part of “Methods” to answer these questions.

- What does BoolODE do? *Converts Boolean network into a system of stochastic differential equations and simulates it.*
- How long do we perform each simulation? *Depends on the network. See Supplementary Table 7.*
- How do we define (expression profile) for a cell?
- How did we create simulated datasets?
- Do we need to compute pseudotime?

BoolODE for Simulation

Read relevant part of “Methods” to answer these questions.

- What does BoolODE do? *Converts Boolean network into a system of stochastic differential equations and simulates it.*
- How long do we perform each simulation? *Depends on the network. See Supplementary Table 7.*
- How do we define (expression profile) for a cell? *Select a random timepoint from each simulation.*
- How did we create simulated datasets?

- Do we need to compute pseudotime?

BoolODE for Simulation

Read relevant part of “Methods” to answer these questions.

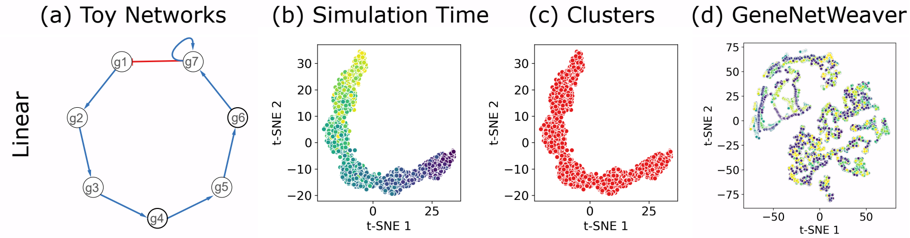
- What does BoolODE do? *Converts Boolean network into a system of stochastic differential equations and simulates it.*
- How long do we perform each simulation? *Depends on the network. See Supplementary Table 7.*
- How do we define (expression profile) for a cell? *Select a random timepoint from each simulation.*
- How did we create simulated datasets? *See details in “Pre-processing datasets from synthetic networks for GRN algorithms.”*
- Do we need to compute pseudotime?

BoolODE for Simulation

Read relevant part of “Methods” to answer these questions.

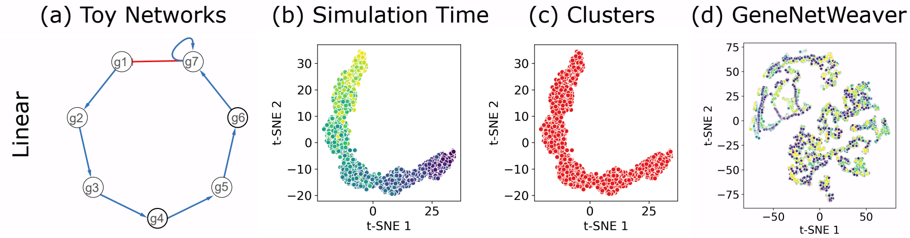
- What does BoolODE do? *Converts Boolean network into a system of stochastic differential equations and simulates it.*
- How long do we perform each simulation? *Depends on the network. See Supplementary Table 7.*
- How do we define (expression profile) for a cell? *Select a random timepoint from each simulation.*
- How did we create simulated datasets? *See details in “Pre-processing datasets from synthetic networks for GRN algorithms.”*
- Do we need to compute pseudotime? *Decided not to do so for synthetic networks. Why?*

Results of Simulations



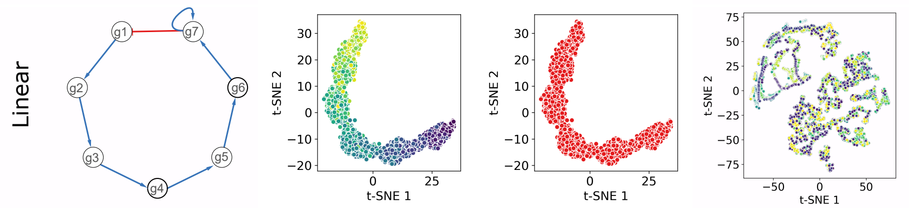
- What did we cluster in column (c)? Which clustering algorithm did we use?
- What value did we use if a GRN method required time-ordered cells? Why?
- What do we learn by comparing plots in column (b) to those in column (c)?
- Does GeneNetWeaver (column (d)) replicate trajectories expected from synthetic networks?

Results of Simulations



- What did we cluster in column (c)? Which clustering algorithm did we use? *We clustered entire simulations using the k-means algorithm with k set to the expected number of steady states.*
- What value did we use if a GRN method required time-ordered cells? Why?
- What do we learn by comparing plots in column (b) to those in column (c)?
- Does GeneNetWeaver (column (d)) replicate trajectories expected from synthetic networks?

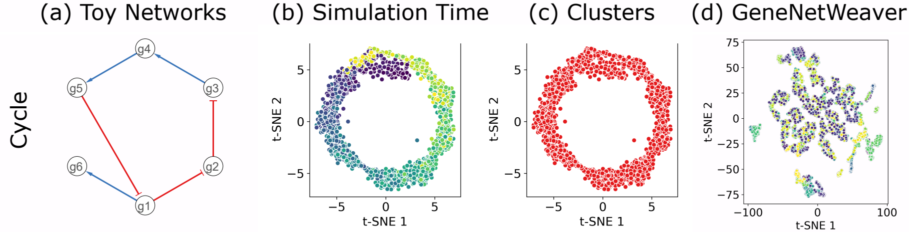
Results of Simulations



(a) Toy Networks (b) Simulation Time (c) Clusters (d) GeneNetWeaver

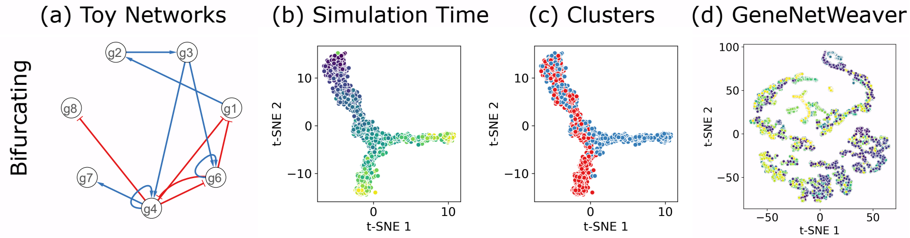
- What did we cluster in column (c)? Which clustering algorithm did we use? *We clustered entire simulations using the k-means algorithm with k set to the expected number of steady states.*
- What value did we use if a GRN method required time-ordered cells? Why? *We used the simulation time.*
- What do we learn by comparing plots in column (b) to those in column (c)?
- Does GeneNetWeaver (column (d)) replicate trajectories expected from synthetic networks?

Results of Simulations



- What did we cluster in column (c)? Which clustering algorithm did we use? *We clustered entire simulations using the k-means algorithm with k set to the expected number of steady states.*
- What value did we use if a GRN method required time-ordered cells? Why? *We used the simulation time.*
- What do we learn by comparing plots in column (b) to those in column (c)?
- Does GeneNetWeaver (column (d)) replicate trajectories expected from synthetic networks?

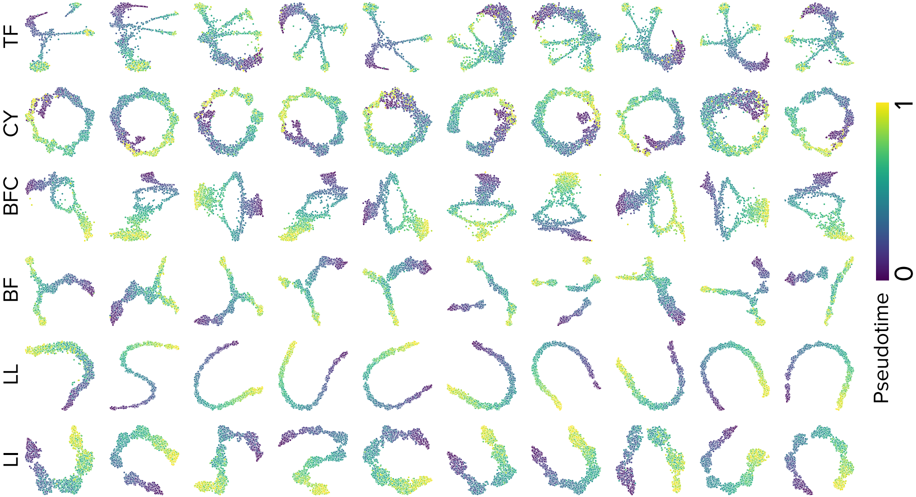
Results of Simulations



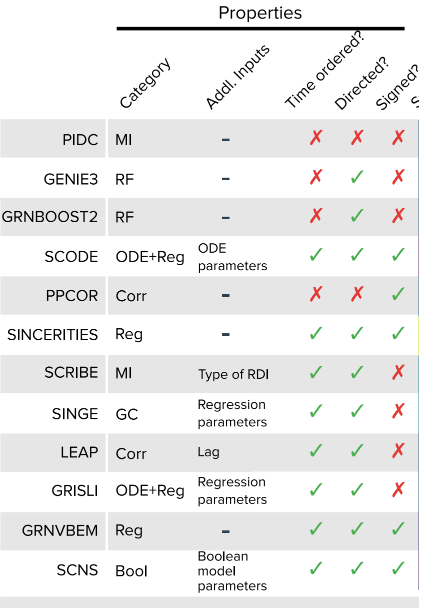
- What did we cluster in column (c)? Which clustering algorithm did we use? *We clustered entire simulations using the k-means algorithm with k set to the expected number of steady states.*
- What value did we use if a GRN method required time-ordered cells? Why? *We used the simulation time.*
- What do we learn by comparing plots in column (b) to those in column (c)?
- Does GeneNetWeaver (column (d)) replicate trajectories expected from synthetic networks?

Results for Multiple Simulations

Used 10 sets of simulation parameters.



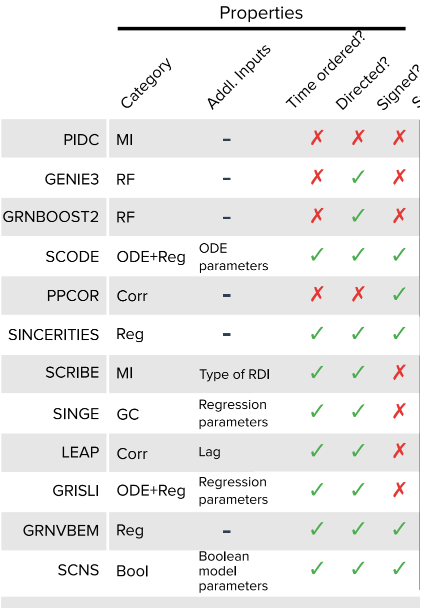
Algorithms included in BEELINE



	Properties				
	Category	Addl. Inputs	Time ordered?	Directed?	Signed?
PIDC	MI	–	✗	✗	✗
GENIE3	RF	–	✗	✓	✗
GRNBOOST2	RF	–	✗	✓	✗
SCODE	ODE+Reg	ODE parameters	✓	✓	✓
PPCOR	Corr	–	✗	✗	✓
SINCERITIES	Reg	–	✓	✓	✓
SCRIBE	MI	Type of RDI	✓	✓	✗
SINGE	GC	Regression parameters	✓	✓	✗
LEAP	Corr	Lag	✓	✓	✗
GRISLI	ODE+Reg	Regression parameters	✓	✓	✗
GRNVBEM	Reg	–	✓	✓	✓
SCNS	Bool	Boolean model parameters	✓	✓	✓

- Which algorithms were developed for bulk RNA-seq data?
- How did we evaluate methods that output undirected GRNs?
- How did we handle methods with parameters?

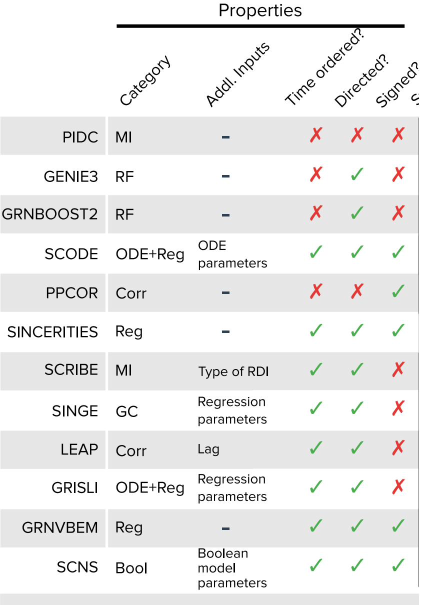
Algorithms included in BEELINE



	Properties				
	Category	Addl. Inputs	Time ordered?	Directed?	Signed?
PIDC	MI	–	✗	✗	✗
GENIE3	RF	–	✗	✓	✗
GRNBOOST2	RF	–	✗	✓	✗
SCODE	ODE+Reg	ODE parameters	✓	✓	✓
PPCOR	Corr	–	✗	✗	✓
SINCERITIES	Reg	–	✓	✓	✓
SCRIBE	MI	Type of RDI	✓	✓	✗
SINGE	GC	Regression parameters	✓	✓	✗
LEAP	Corr	Lag	✓	✓	✗
GRISLI	ODE+Reg	Regression parameters	✓	✓	✗
GRNVBEM	Reg	–	✓	✓	✓
SCNS	Bool	Boolean model parameters	✓	✓	✓

- Which algorithms were developed for bulk RNA-seq data? *GENEI3, GRNBOOST2, PPCOR.*
- How did we evaluate methods that output undirected GRNs?
- How did we handle methods with parameters?

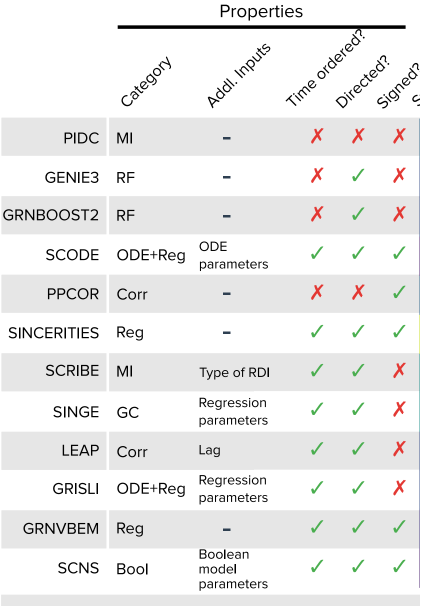
Algorithms included in BEELINE



	Properties				
	Category	Addl. Inputs	Time ordered?	Directed?	Signed?
PIDC	MI	–	✗	✗	✗
GENIE3	RF	–	✗	✓	✗
GRNBOOST2	RF	–	✗	✓	✗
SCODE	ODE+Reg	ODE parameters	✓	✓	✓
PPCOR	Corr	–	✗	✗	✓
SINCERITIES	Reg	–	✓	✓	✓
SCRIBE	MI	Type of RDI	✓	✓	✗
SINGE	GC	Regression parameters	✓	✓	✗
LEAP	Corr	Lag	✓	✓	✗
GRISLI	ODE+Reg	Regression parameters	✓	✓	✗
GRNVBEM	Reg	–	✓	✓	✓
SCNS	Bool	Boolean model parameters	✓	✓	✓

- Which algorithms were developed for bulk RNA-seq data? *GENEI3, GRNBOOST2, PPCOR.*
- How did we evaluate methods that output undirected GRNs? *Gave both edge directions the same rank.*
- How did we handle methods with parameters?

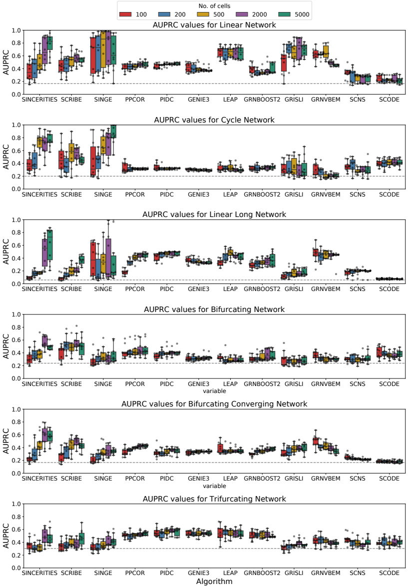
Algorithms included in BEELINE



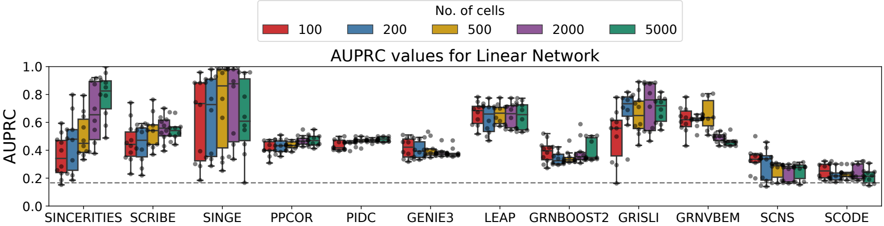
	Properties				
	Category	Addl. Inputs	Time ordered?	Directed?	Signed?
PIDC	MI	–	✗	✗	✗
GENIE3	RF	–	✗	✓	✗
GRNBOOST2	RF	–	✗	✓	✗
SCODE	ODE+Reg	ODE parameters	✓	✓	✓
PPCOR	Corr	–	✗	✗	✓
SINCERITIES	Reg	–	✓	✓	✓
SCRIBE	MI	Type of RDI	✓	✓	✗
SINGE	GC	Regression parameters	✓	✓	✗
LEAP	Corr	Lag	✓	✓	✗
GRISLI	ODE+Reg	Regression parameters	✓	✓	✗
GRNVBEM	Reg	–	✓	✓	✓
SCNS	Bool	Boolean model parameters	✓	✓	✓

- Which algorithms were developed for bulk RNA-seq data? *GENEI3, GRNBOOST2, PPCOR.*
- How did we evaluate methods that output undirected GRNs? *Gave both edge directions the same rank.*
- How did we handle methods with parameters? *Performed parameter sweeps.*

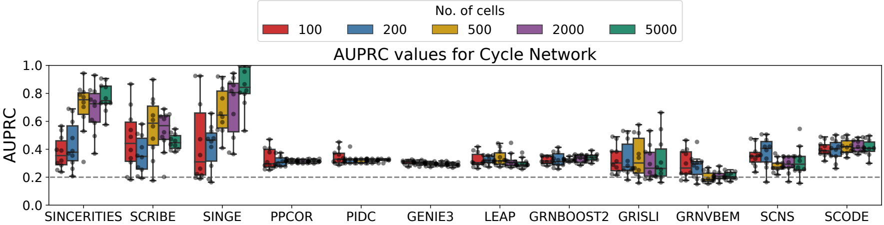
AUPRC Results



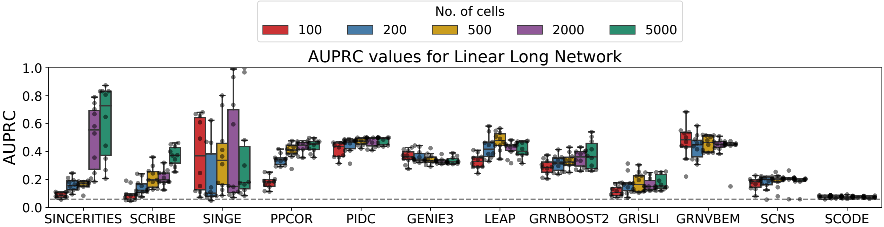
AUPRC Results



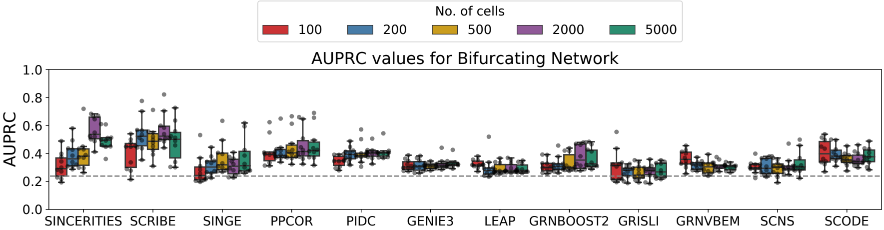
AUPRC Results



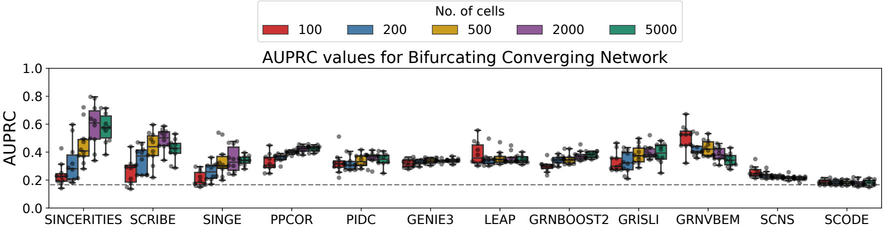
AUPRC Results



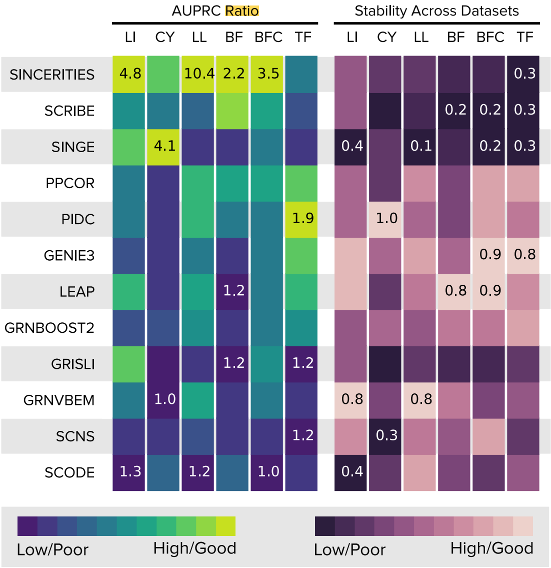
AUPRC Results



AUPRC Results



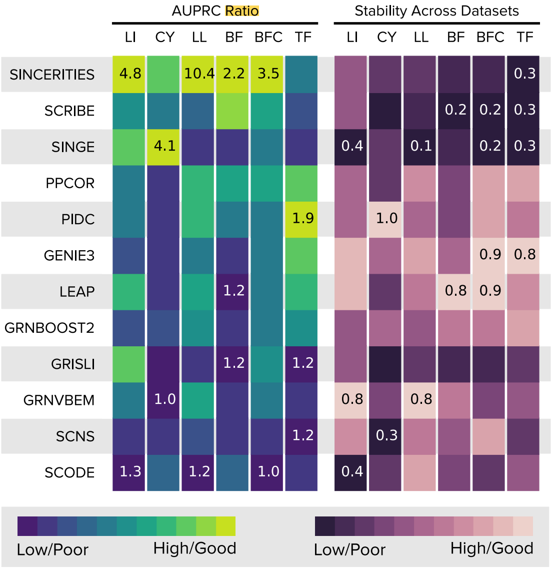
Summary of Results



	AUPRC Ratio						Stability Across Datasets					
	LI	CY	LL	BF	BFC	TF	LI	CY	LL	BF	BFC	TF
SINCERITIES	4.8		10.4	2.2	3.5							0.3
SCRIBE										0.2	0.2	0.3
SINGE		4.1					0.4		0.1		0.2	0.3
PPCOR												
PIDC						1.9		1.0				
GENIE3											0.9	0.8
LEAP				1.2						0.8	0.9	
GRNBOOST2												
GRISLI				1.2		1.2						
GRNVBEM		1.0					0.8		0.8			
SCNS						1.2		0.3				
SCODE	1.3		1.2		1.0		0.4					

- What is AUPRC ratio?
- What is a top-*k* GRN?
- How did we compute stability?

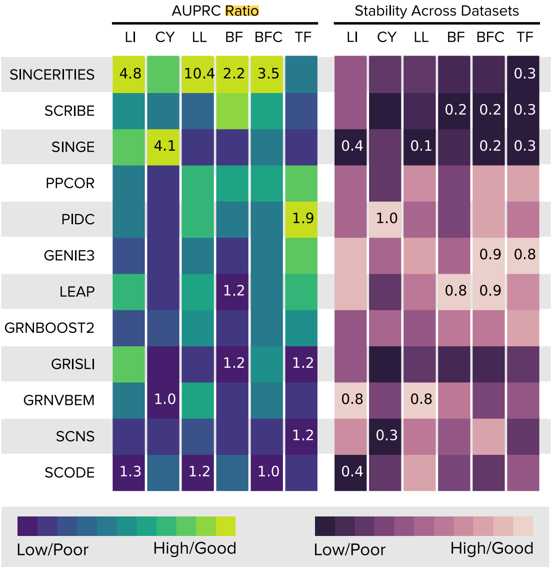
Summary of Results



	AUPRC Ratio						Stability Across Datasets					
	LI	CY	LL	BF	BFC	TF	LI	CY	LL	BF	BFC	TF
SINCERITIES	4.8		10.4	2.2	3.5							0.3
SCRIBE										0.2	0.2	0.3
SINGE		4.1					0.4		0.1		0.2	0.3
PPCOR												
PIDC						1.9		1.0				
GENIE3											0.9	0.8
LEAP				1.2						0.8	0.9	
GRNBOOST2												
GRISLI				1.2		1.2						
GRNVBEM		1.0					0.8		0.8			
SCNS						1.2		0.3				
SCODE	1.3		1.2		1.0		0.4					

- What is AUPRC ratio? *AUPRC of method divided by the AUPRC of a random predictor.*
- What is a top-*k* GRN?
- How did we compute stability?

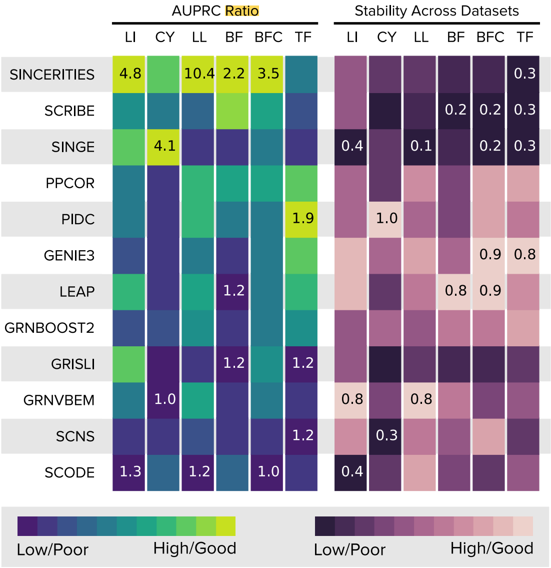
Summary of Results



	AUPRC Ratio						Stability Across Datasets					
	LI	CY	LL	BF	BFC	TF	LI	CY	LL	BF	BFC	TF
SINCERITIES	4.8		10.4	2.2	3.5							0.3
SCRIBE										0.2	0.2	0.3
SINGE		4.1					0.4		0.1		0.2	0.3
PPCOR												
PIDC						1.9		1.0				
GENIE3											0.9	0.8
LEAP				1.2						0.8	0.9	
GRNBOOST2												
GRISLI				1.2		1.2						
GRNVBEM		1.0					0.8		0.8			
SCNS						1.2		0.3				
SCODE	1.3		1.2		1.0		0.4					

- What is AUPRC ratio? *AUPRC of method divided by the AUPRC of a random predictor.*
- What is a top-k GRN? *Predicted network containing as many edges as the ground-truth GRN.*
- How did we compute stability?

Summary of Results



	AUPRC Ratio						Stability Across Datasets					
	LI	CY	LL	BF	BFC	TF	LI	CY	LL	BF	BFC	TF
SINCERITIES	4.8		10.4	2.2	3.5							0.3
SCRIBE										0.2	0.2	0.3
SINGE		4.1					0.4		0.1		0.2	0.3
PPCOR												
PIDC						1.9		1.0				
GENIE3											0.9	0.8
LEAP				1.2						0.8	0.9	
GRNBOOST2												
GRISLI				1.2		1.2						
GRNVBEM		1.0					0.8		0.8			
SCNS						1.2		0.3				
SCODE	1.3		1.2		1.0		0.4					

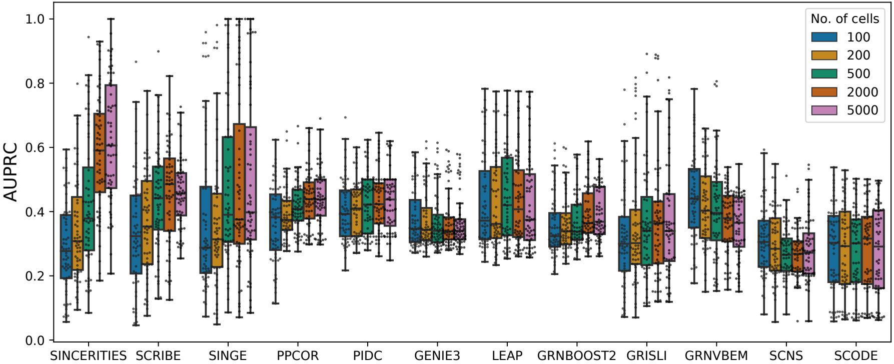
- What is AUPRC ratio? *AUPRC of method divided by the AUPRC of a random predictor.*
- What is a top-*k* GRN? *Predicted network containing as many edges as the ground-truth GRN.*
- How did we compute stability? *"We repeated this procedure on ten different sampled parameter sets to obtain 50 datasets." For each (algorithm, synthetic network) pair, computed Jaccard coefficient of all pairs of GRNs from top-k edges.*

Other Analyses:

- Earlier results were for 2,000 cells.
- We also sampled 100, 500, 1,000, and 5,000 cells as well.

Other Analyses: Effect of Number of Cells

- Earlier results were for 2,000 cells.
- We also sampled 100, 500, 1,000, and 5,000 cells as well.



Other Analyses: Effect of Number of Cells

- Earlier results were for 2,000 cells.
- We also sampled 100, 500, 1,000, and 5,000 cells as well.

#cells compared	SINCERITIES	SCRIBE	SINGE	PPCOR	PIDC	GENIE3
100–5K	$\mathbf{9.2 \times 10^{-10}}$	$\mathbf{7.9 \times 10^{-6}}$	$\mathbf{2.9 \times 10^{-4}}$	$\mathbf{5.3 \times 10^{-6}}$	$\mathbf{8.2 \times 10^{-4}}$	1.0
200–5K	$\mathbf{3 \times 10^{-9}}$	$\mathbf{2.5 \times 10^{-5}}$	$\mathbf{2.7 \times 10^{-5}}$	$\mathbf{2.5 \times 10^{-6}}$	$\mathbf{2.6 \times 10^{-5}}$	1.0
500–5K	$\mathbf{6.9 \times 10^{-8}}$	0.30	0.35	$\mathbf{2.4 \times 10^{-6}}$	$\mathbf{2.9 \times 10^{-2}}$	0.67
2K–5K	0.12	0.71	0.31	0.35	0.40	1.0

# cells compared	LEAP	GRNBoost2	GRISLI	GRNVBEM	SCNS	SCODE
100–5K	1.0	$\mathbf{5.6 \times 10^{-4}}$	$\mathbf{2.3 \times 10^{-3}}$	1.0	1.0	1.0
200–5K	1.0	$\mathbf{2. \times 10^{-4}}$	0.41	1.0	1.0	1.0
500–5K	1.0	$\mathbf{5.7 \times 10^{-4}}$	0.59	1.0	0.91	0.88
2K–5K	1.0	0.87	1.0	1.0	0.66	1.0

Other Analyses: Effect of Number of Cells

- Earlier results were for 2,000 cells.
- We also sampled 100, 500, 1,000, and 5,000 cells as well.

#cells compared	SINCERITIES	SCRIBE	SINGE	PPCOR	PIDC	GENIE3
100–5K	$\mathbf{9.2 \times 10^{-10}}$	$\mathbf{7.9 \times 10^{-6}}$	$\mathbf{2.9 \times 10^{-4}}$	$\mathbf{5.3 \times 10^{-6}}$	$\mathbf{8.2 \times 10^{-4}}$	1.0
200–5K	$\mathbf{3 \times 10^{-9}}$	$\mathbf{2.5 \times 10^{-5}}$	$\mathbf{2.7 \times 10^{-5}}$	$\mathbf{2.5 \times 10^{-6}}$	$\mathbf{2.6 \times 10^{-5}}$	1.0
500–5K	$\mathbf{6.9 \times 10^{-8}}$	0.30	0.35	$\mathbf{2.4 \times 10^{-6}}$	$\mathbf{2.9 \times 10^{-2}}$	0.67
2K–5K	0.12	0.71	0.31	0.35	0.40	1.0

# cells compared	LEAP	GRNBoost2	GRISLI	GRNVBEM	SCNS	SCODE
100–5K	1.0	$\mathbf{5.6 \times 10^{-4}}$	$\mathbf{2.3 \times 10^{-3}}$	1.0	1.0	1.0
200–5K	1.0	$\mathbf{2. \times 10^{-4}}$	0.41	1.0	1.0	1.0
500–5K	1.0	$\mathbf{5.7 \times 10^{-4}}$	0.59	1.0	0.91	0.88
2K–5K	1.0	0.87	1.0	1.0	0.66	1.0

- Fidelity of simulations from inferred GRNs (Supplementary Note 1.4)

Other Analyses: Effect of Number of Cells

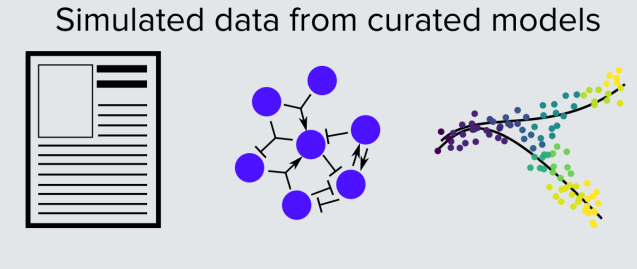
- Earlier results were for 2,000 cells.
- We also sampled 100, 500, 1,000, and 5,000 cells as well.

#cells compared	SINCERITIES	SCRIBE	SINGE	PPCOR	PIDC	GENIE3
100–5K	**9.2×10^{-10}**	**7.9×10^{-6}**	**2.9×10^{-4}**	**5.3×10^{-6}**	**8.2×10^{-4}**	1.0
200–5K	**3×10^{-9}**	**2.5×10^{-5}**	**2.7×10^{-5}**	**2.5×10^{-6}**	**2.6×10^{-5}**	1.0
500–5K	**6.9×10^{-8}**	0.30	0.35	**2.4×10^{-6}**	**2.9×10^{-2}**	0.67
2K–5K	0.12	0.71	0.31	0.35	0.40	1.0

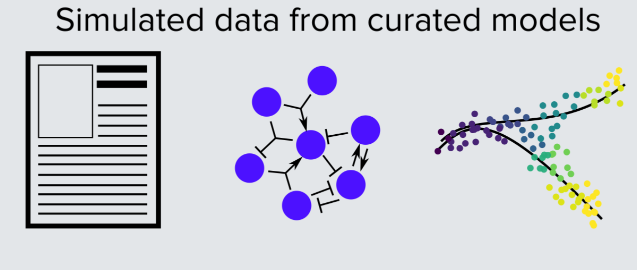
# cells compared	LEAP	GRNBoost2	GRISLI	GRNVBEM	SCNS	SCODE
100–5K	1.0	**5.6×10^{-4}**	**2.3×10^{-3}**	1.0	1.0	1.0
200–5K	1.0	**$2. \times 10^{-4}$**	0.41	1.0	1.0	1.0
500–5K	1.0	**5.7×10^{-4}**	0.59	1.0	0.91	0.88
2K–5K	1.0	0.87	1.0	1.0	0.66	1.0

- Fidelity of simulations from inferred GRNs (Supplementary Note 1.4)
- Effect of using trajectory information (Supplementary Note 1.5)

Input Type 2: Curated Models

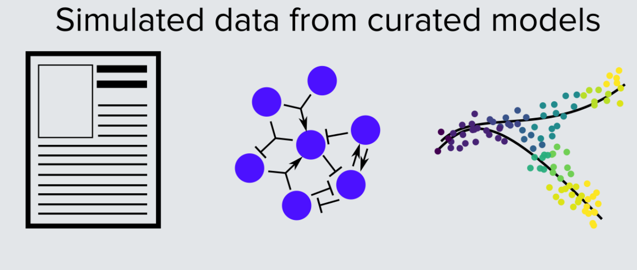


Input Type 2: Curated Models



- What was our motivation for curating Boolean models?
  - Simple structures of synthetic networks may be unrealistic.
  - Existing simulators (e.g., GeneNetWeaver) simulate dense subnetworks of large-scale GRNs grown around random seeds.

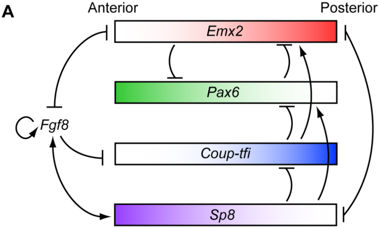
Input Type 2: Curated Models



- We selected four published Boolean models:
  1. Mammalian cortical area development (mCAD).
  2. Ventral spinal cord (VSC) development.
  3. Hematopoietic stem cell (HSC) differentiation.
  4. Gonadal sex determination (GSD).

Name of model	**#genes**	**#edges**		**#Steady States**
		Activation	Inhibition	
mCAD	5	5	9	2
VSC	8	0	15	5
HSC	11	15	15	4
GSD	19	27	59	2

Example of Curated Model



C

$$
\begin{aligned}
F(t+1) &= \mathrm{F}(t)\ \&\ \mathrm{NOT\ E}(t)\ \&\ \mathrm{S}(t)\\
E(t+1) &= \mathrm{NOT\ F}(t)\ \&\ \mathrm{NOT\ P}(t)\ \&\ \mathrm{C}(t)\\
&\quad\ \&\ \mathrm{NOT\ S}(t)\\
P(t+1) &= \mathrm{NOT\ E}(t)\ \&\ \mathrm{NOT\ C}(t)\ \&\ \mathrm{S}(t)\\
C(t+1) &= \mathrm{NOT\ F}(t)\ \&\ \mathrm{NOT\ S}(t)\\
S(t+1) &= \mathrm{F}(t)\ \&\ \mathrm{NOT\ E}(t)
\end{aligned}
$$

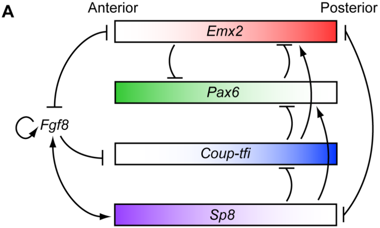
$$
\begin{aligned}
\mathrm{F}(t+1) &= F(t)\\
\mathrm{E}(t+1) &= E(t)\\
\mathrm{P}(t+1) &= P(t)\\
\mathrm{C}(t+1) &= C(t)\\
\mathrm{S}(t+1) &= S(t)
\end{aligned}
$$

B

Interaction	References
Fgf8→*Fgf8*	[52, 57, 58]
Emx2⊣*Fgf8*	[20, 27], but see also [22]
Sp8→*Fgf8*	[24, 26]
Fgf8⊣*Emx2*	[20, 21, 27, 52, 53, 54]
Pax6⊣*Emx2*	[55]
Coup-tfi→*Emx2*	[23], but see also [28]
Sp8⊣*Emx2*	[26]
Emx2⊣*Pax6*	[22, 55]
Coup-tfi⊣*Pax6*	[23, 28]
Sp8→*Pax6*	[26]
Fgf8⊣*Coup-tfi*	[20, 21, 27], but see also [54]
Sp8⊣*Coup-tfi*	[26]
Fgf8→*Sp8*	[24], but see also [27, 54]
Emx2⊣*Sp8*	[24, 27]

Giacomantonio and Goodhill, A Boolean model of the gene regulatory network underlying mammalian cortical area development, PLoS Comput. Biol., 2010.

Example of Curated Model



C

$$\begin{aligned} F(t+1) &= \text{F}(t)\ \&\ \text{NOT E}(t)\ \&\ \text{S}(t) \\ E(t+1) &= \text{NOT F}(t)\ \&\ \text{NOT P}(t)\ \&\ \text{C}(t) \\ &\quad \&\ \text{NOT S}(t) \\ P(t+1) &= \text{NOT E}(t)\ \&\ \text{NOT C}(t)\ \&\ \text{S}(t) \\ C(t+1) &= \text{NOT F}(t)\ \&\ \text{NOT S}(t) \\ S(t+1) &= \text{F}(t)\ \&\ \text{NOT E}(t) \end{aligned}$$

$$\begin{aligned} \text{F}(t+1) &= F(t) \\ \text{E}(t+1) &= E(t) \\ \text{P}(t+1) &= P(t) \\ \text{C}(t+1) &= C(t) \\ \text{S}(t+1) &= S(t) \end{aligned}$$

B

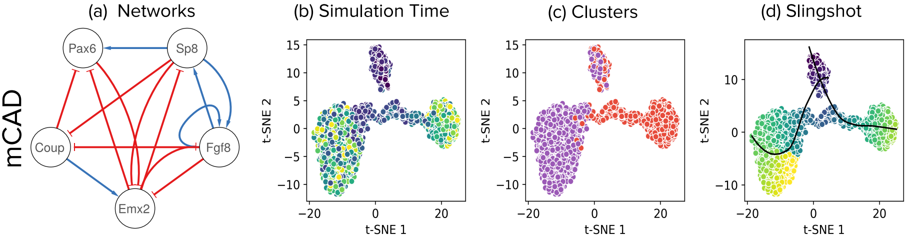
Only direct interaction → Sp8→*Fgf8*

Possible indirect interaction → Coup-tfi→*Emx2*

Interaction	References
Fgf8→*Fgf8*	[52, 57, 58]
Emx2⊣*Fgf8*	[20, 27], but see also [22]
Sp8→*Fgf8*	[24, 26]
Fgf8⊣*Emx2*	[20, 21, 27, 52, 53, 54]
Pax6⊣*Emx2*	[55]
Coup-tfi→*Emx2*	[23], but see also [28]
Sp8⊣*Emx2*	[26]
Emx2⊣*Pax6*	[22, 55]
Coup-tfi⊣*Pax6*	[23, 28]
Sp8→*Pax6*	[26]
Fgf8⊣*Coup-tfi*	[20, 21, 27], but see also [54]
Sp8⊣*Coup-tfi*	[26]
Fgf8→*Sp8*	[24], but see also [27, 54]
Emx2⊣*Sp8*	[24, 27]

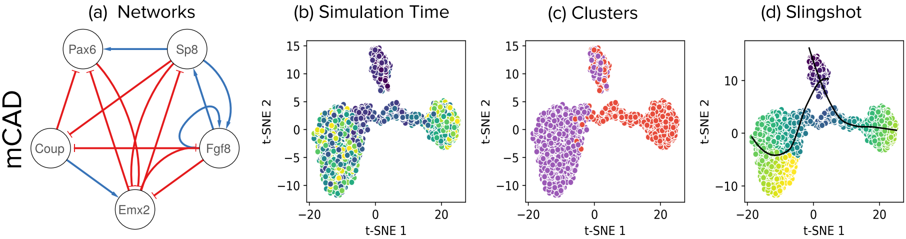
Giacomantonio and Goodhill, A Boolean model of the gene regulatory network underlying mammalian cortical area development, PLoS Comput. Biol., 2010.

Results of Simulations



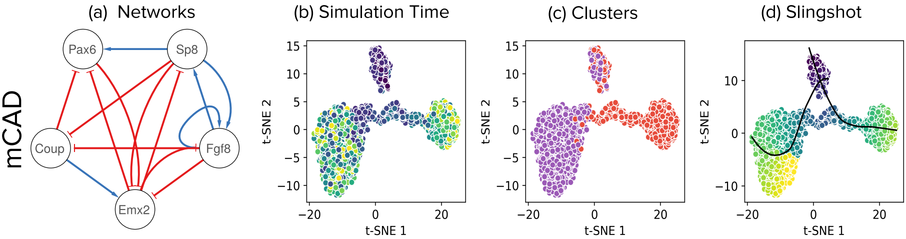
- Why did we do two types of clustering here (column (c) and "behind the scenes" for column (d))?
- What did we compute in column (d)?
- What do learn by comparing plots in column (c) to those in column (b)?
- How well do pseudotime values computed by Slingshot (column (d)) correspond to simulation times of the cells (column (b))?

Results of Simulations



- Why did we do two types of clustering here (column (c) and “behind the scenes” for column (d))? *Read “Pseudotime inference using Slingshot.”*
- What did we compute in column (d)?
- What do learn by comparing plots in column (c) to those in column (b)?
- How well do pseudotime values computed by Slingshot (column (d)) correspond to simulation times of the cells (column (b))?

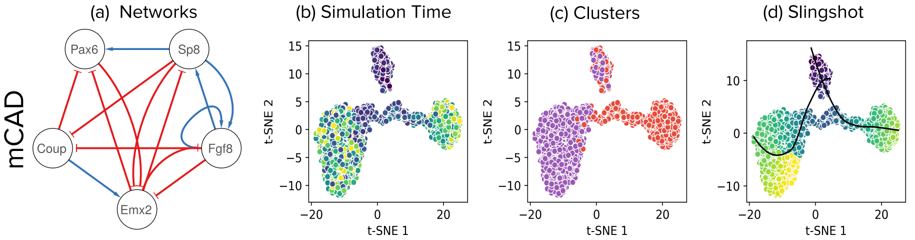
Results of Simulations



- Why did we do two types of clustering here (column (c) and “behind the scenes” for column (d))? *Read “Pseudotime inference using Slingshot.”*
- What did we compute in column (d)? *Slingshot trajectories using clusters of cells with #clusters = 1+ #steady states.*
- What do learn by comparing plots in column (c) to those in column (b)?

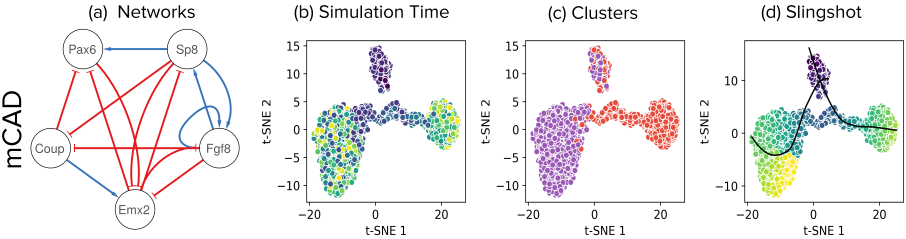
- How well do pseudotime values computed by Slingshot (column (d)) correspond to simulation times of the cells (column (b))?

Results of Simulations



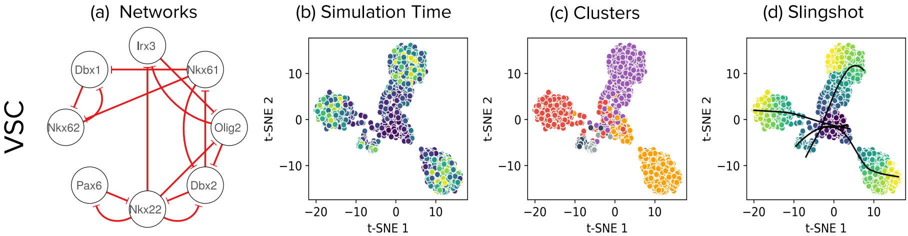
- Why did we do two types of clustering here (column (c) and "behind the scenes" for column (d))? *Read "Pseudotime inference using Slingshot."*
- What did we compute in column (d)? *Slingshot trajectories using clusters of cells with #clusters = 1+ #steady states.*
- What do learn by comparing plots in column (c) to those in column (b)? *Clusters of simulations (as for synthetic models) correspond well to Slingshot trajectories.*
- How well do pseudotime values computed by Slingshot (column (d)) correspond to simulation times of the cells (column (b))?

Results of Simulations



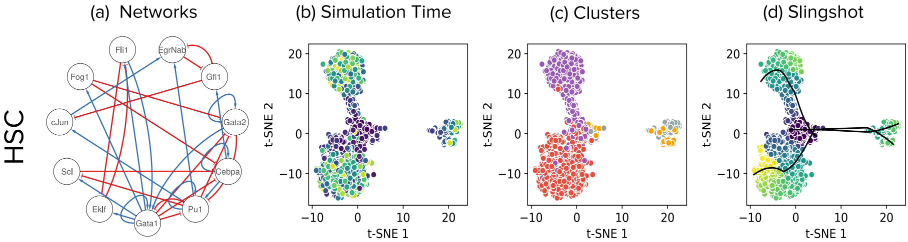
- Why did we do two types of clustering here (column (c) and “behind the scenes” for column (d))? *Read “Pseudotime inference using Slingshot.”*
- What did we compute in column (d)? *Slingshot trajectories using clusters of cells with #clusters = 1+ #steady states.*
- What do learn by comparing plots in column (c) to those in column (b)? *Clusters of simulations (as for synthetic models) correspond well to Slingshot trajectories.*
- How well do pseudotime values computed by Slingshot (column (d)) correspond to simulation times of the cells (column (b))? *Correlation is high (Supplementary Table 8).*

Results of Simulations



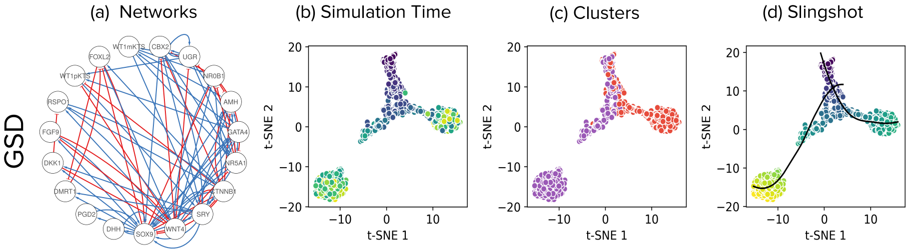
- Why did we do two types of clustering here (column (c) and "behind the scenes" for column (d))? *Read "Pseudotime inference using Slingshot."*
- What did we compute in column (d)? *Slingshot trajectories using clusters of cells with #clusters = 1+ #steady states.*
- What do learn by comparing plots in column (c) to those in column (b)? *Clusters of simulations (as for synthetic models) correspond well to Slingshot trajectories.*
- How well do pseudotime values computed by Slingshot (column (d)) correspond to simulation times of the cells (column (b))? *Correlation is high (Supplementary Table 8).*

Results of Simulations


(a) Networks (b) Simulation Time (c) Clusters (d) Slingshot

- Why did we do two types of clustering here (column (c) and “behind the scenes” for column (d))? *Read “Pseudotime inference using Slingshot.”*
- What did we compute in column (d)? *Slingshot trajectories using clusters of cells with #clusters = 1+ #steady states.*
- What do learn by comparing plots in column (c) to those in column (b)? *Clusters of simulations (as for synthetic models) correspond well to Slingshot trajectories.*
- How well do pseudotime values computed by Slingshot (column (d)) correspond to simulation times of the cells (column (b))? *Correlation is high (Supplementary Table 8).*

Results of Simulations



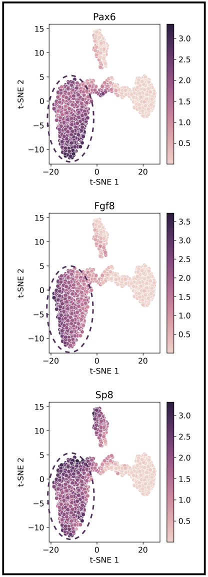
(a) Networks (b) Simulation Time (c) Clusters (d) Slingshot

- Why did we do two types of clustering here (column (c) and “behind the scenes” for column (d))? *Read “Pseudotime inference using Slingshot.”*
- What did we compute in column (d)? *Slingshot trajectories using clusters of cells with #clusters = 1+ #steady states.*
- What do learn by comparing plots in column (c) to those in column (b)? *Clusters of simulations (as for synthetic models) correspond well to Slingshot trajectories.*
- How well do pseudotime values computed by Slingshot (column (d)) correspond to simulation times of the cells (column (b))? *Correlation is high (Supplementary Table 8).*

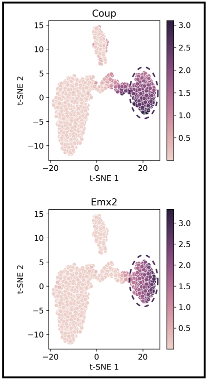
Further Validation of BoolODE Simulations

- Each publication specifies a unique gene expression pattern that characterizes each steady state of that model.
- Does simulation match these patterns?

Anterior

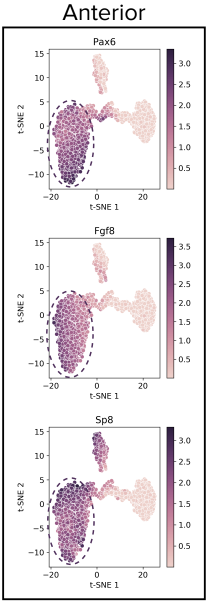


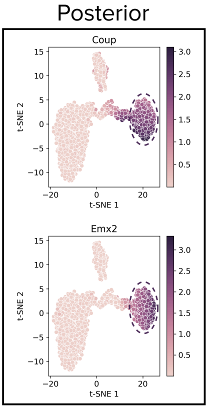
Posterior



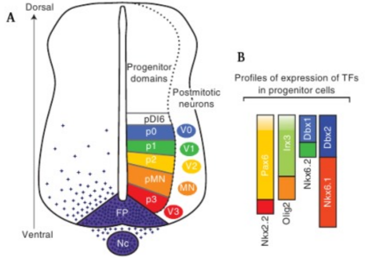
Further Validation of BoolODE Simulations

- Each publication specifies a unique gene expression pattern that characterizes each steady state of that model.
- Does simulation match these patterns?
- Plot expression of each gene and visually correlate "clusters" in t-SNE plots to steady state.

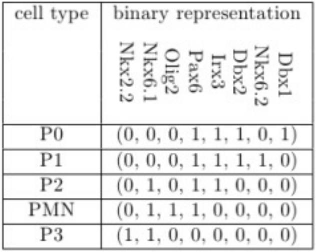




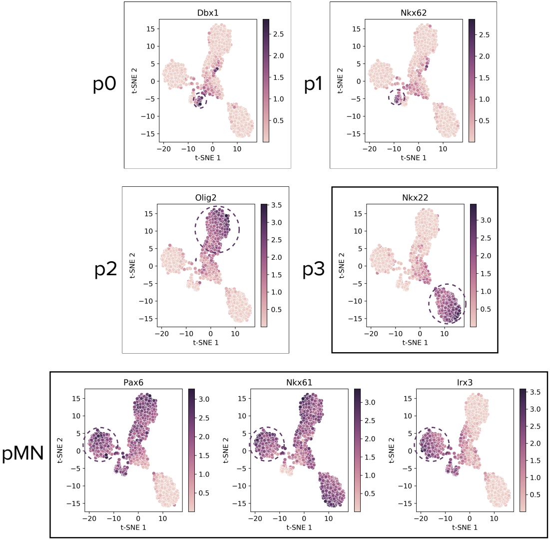
BoolODE Simulations for VSC model



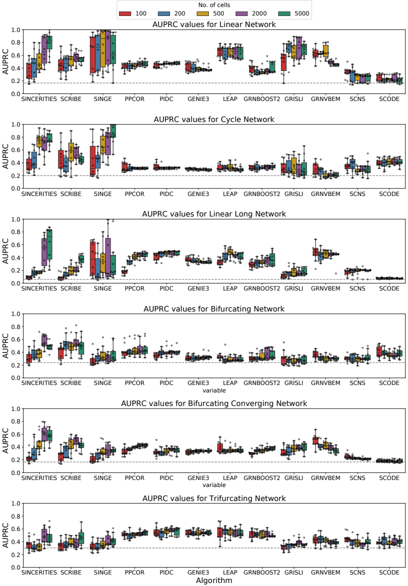
C



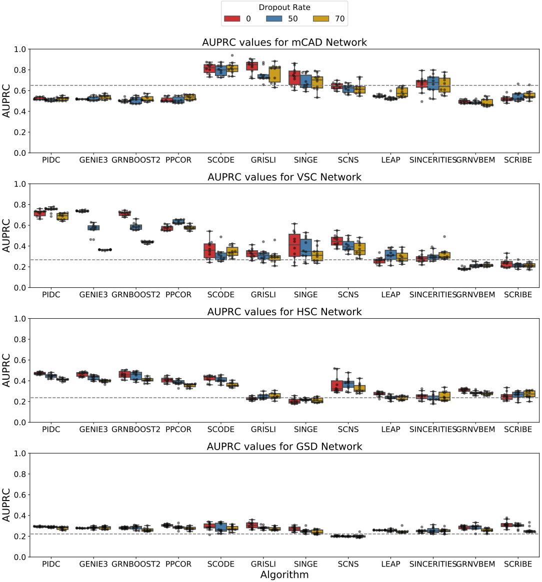
cell type	binary representation
	Nkx2.2 Nkx6.1 Olig2 Pax6 Irx3 Dbx2 Nkx6.2 Dbx1
P0	(0, 0, 0, 1, 1, 1, 0, 1)
P1	(0, 0, 0, 1, 1, 1, 1, 0)
P2	(0, 1, 0, 1, 1, 0, 0, 0)
PMN	(0, 1, 1, 1, 0, 0, 0, 0)
P3	(1, 1, 0, 0, 0, 0, 0, 0)



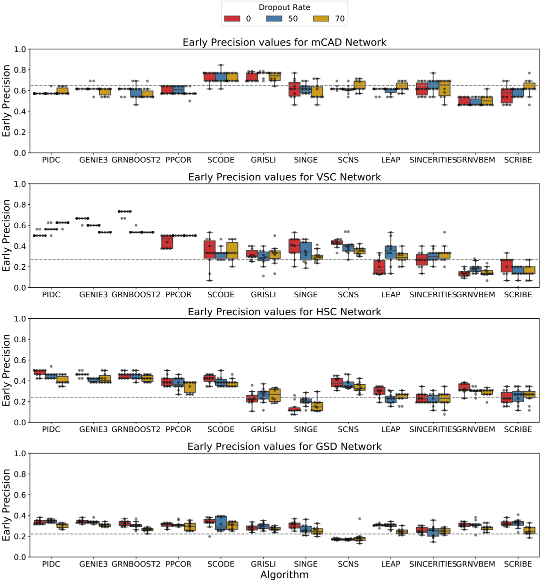
AUPRC Results for Synthetic Networks



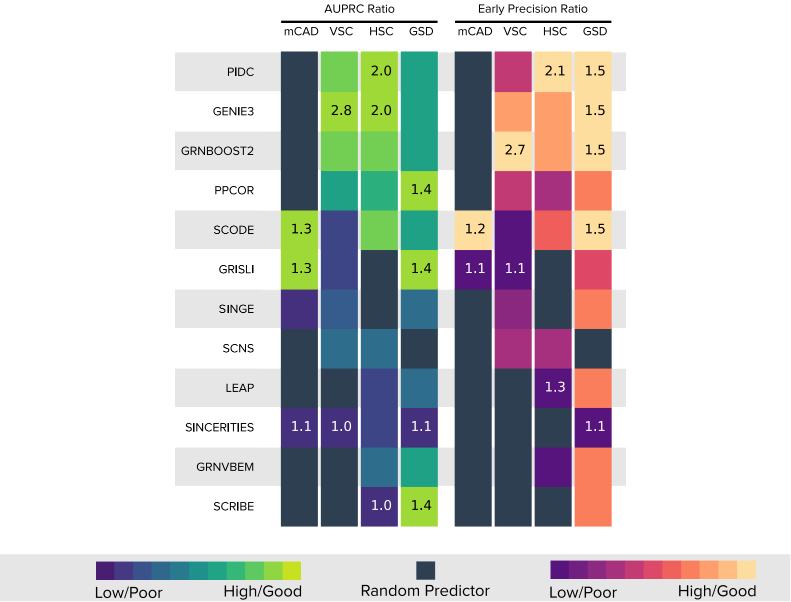
AUPRC Results for Curated Models



EPR Results for Curated Models

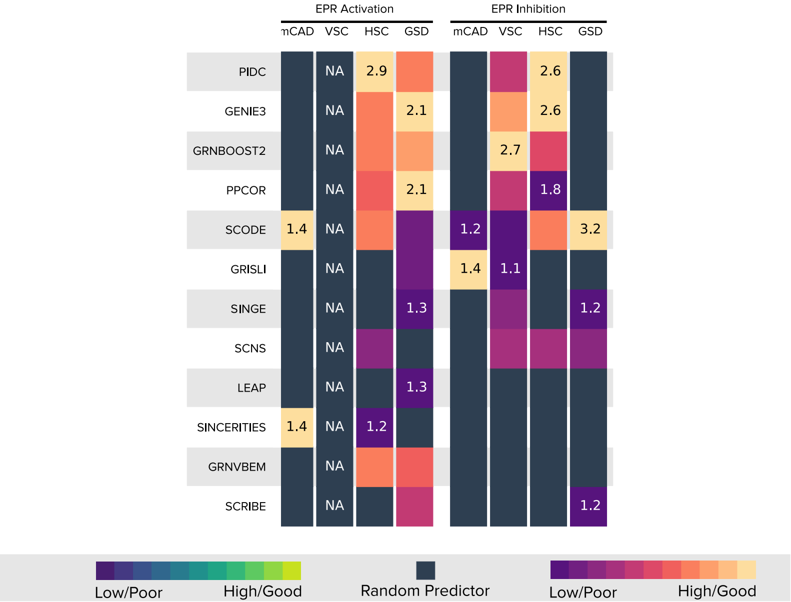


AUPRC and EPR Ratios



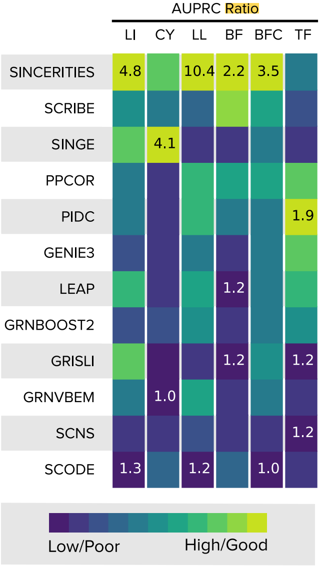
	AUPRC Ratio				Early Precision Ratio			
	mCAD	VSC	HSC	GSD	mCAD	VSC	HSC	GSD
PIDC			2.0				2.1	1.5
GENIE3		2.8	2.0					1.5
GRNBOOST2						2.7		1.5
PPCOR				1.4				
SCODE	1.3				1.2			1.5
GRISLI	1.3			1.4	1.1	1.1		
SINGE								
SCNS								
LEAP							1.3	
SINCERITIES	1.1	1.0		1.1				1.1
GRNVBEM								
SCRIBE			1.0	1.4				

AUPRC and EPR Ratios



	EPR Activation				EPR Inhibition			
	mCAD	VSC	HSC	GSD	mCAD	VSC	HSC	GSD
PIDC		NA	2.9				2.6	
GENIE3		NA		2.1			2.6	
GRNBOOST2		NA				2.7		
PPCOR		NA		2.1			1.8	
SCODE	1.4	NA			1.2			3.2
GRISLI		NA			1.4	1.1		
SINGE		NA		1.3				1.2
SCNS		NA						
LEAP		NA		1.3				
SINCERITIES	1.4	NA	1.2					
GRNVBEM		NA						
SCRIBE		NA						1.2

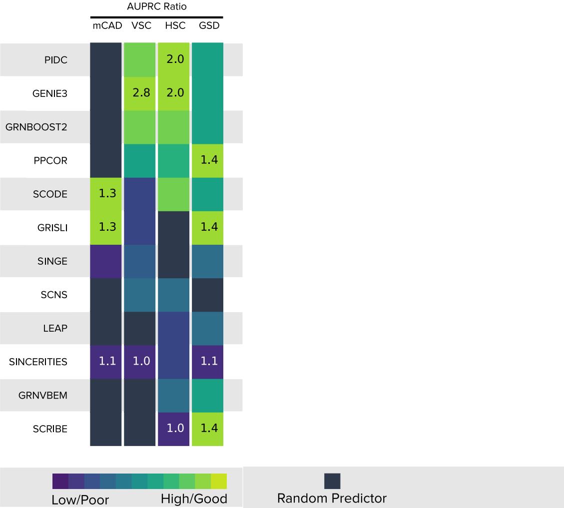
Results for Synthetic Models



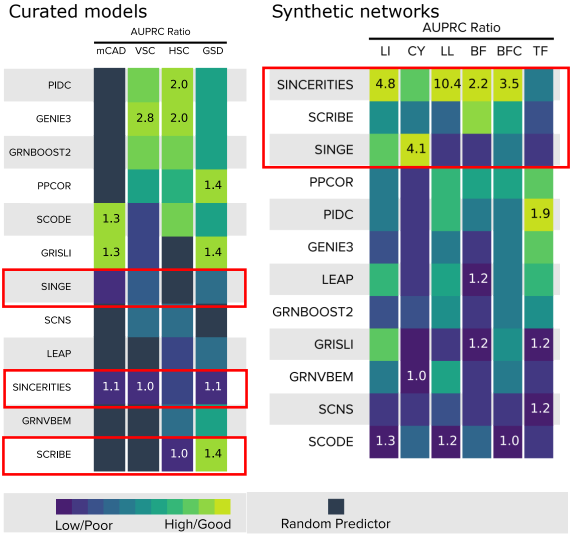
	AUPRC Ratio					
	LI	CY	LL	BF	BFC	TF
SINCERITIES	4.8		10.4	2.2	3.5	
SCRIBE						
SINGE		4.1				
PPCOR						
PIDC						1.9
GENIE3						
LEAP				1.2		
GRNBOOST2						
GRISLI				1.2		1.2
GRNVBEM		1.0				
SCNS						1.2
SCODE	1.3		1.2		1.0	

Results for Curated Models

Curated models

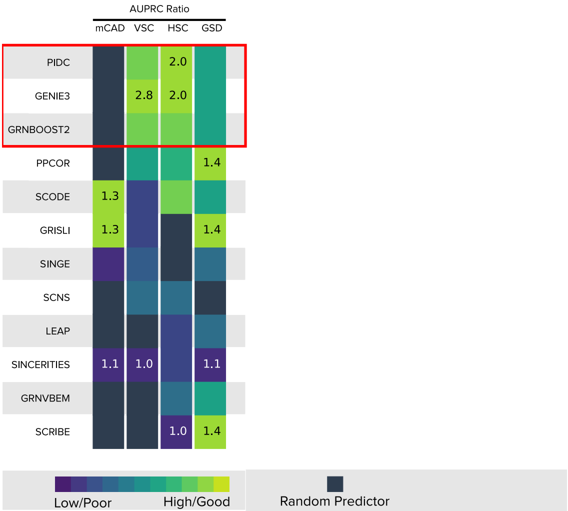


Results for Curated Models



Results for Curated Models

Curated models



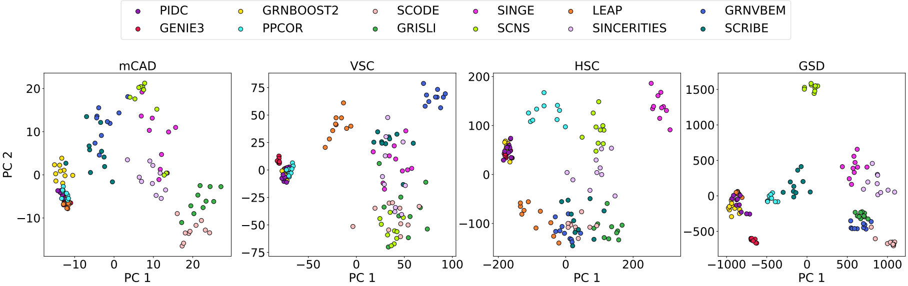
Methods that do not require pseudotime have the highest AUPRC ratios.

Other Analyses

Other Analyses

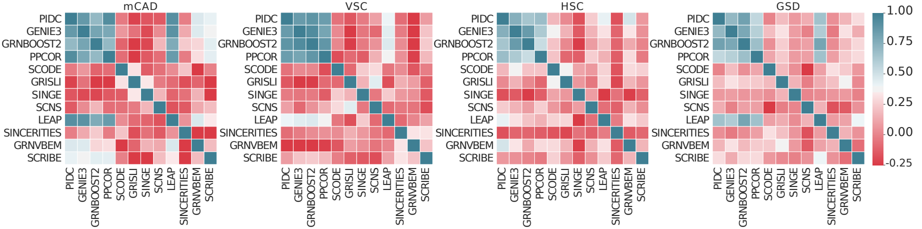
- Effect of dropouts (Supplementary Note 2.3).
- Similarity across algorithms (Supplementary Note 2.4)
- Motif analysis (Supplementary Note 2.5) and indirect edges (Supplementary Note 2.6)

Similarity of Reconstructions

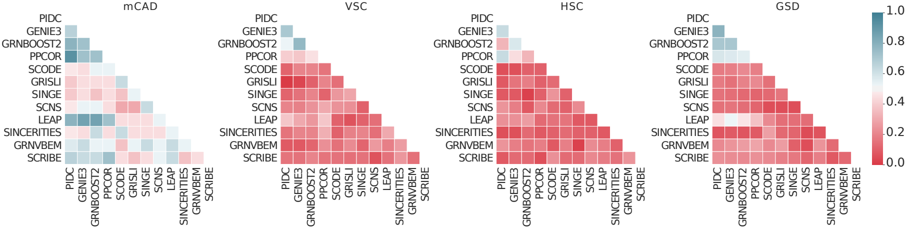


Similarity of Reconstructions

(a) Spearman correlation

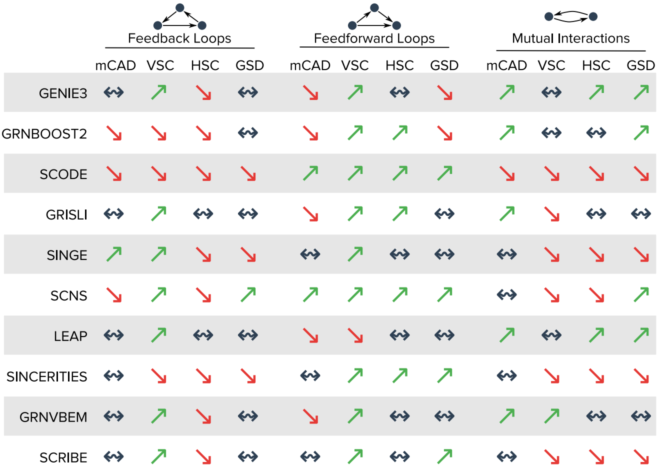


(b) Jaccard index



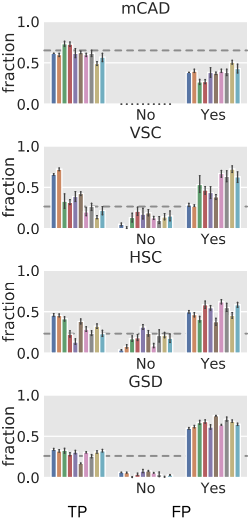
Motifs in Inferred GRNs

(a) Network Motifs



	Feedback Loops				Feedforward Loops				Mutual Interactions			
	mCAD	VSC	HSC	GSD	mCAD	VSC	HSC	GSD	mCAD	VSC	HSC	GSD
GENIE3	↔	↗	↘	↔	↘	↗	↔	↘	↗	↔	↗	↗
GRNBOOST2	↘	↘	↘	↔	↘	↗	↗	↘	↗	↔	↔	↗
SCODE	↘	↘	↘	↘	↗	↗	↗	↗	↘	↘	↘	↘
GRISLI	↔	↗	↔	↔	↘	↗	↗	↔	↗	↘	↔	↔
SINGE	↗	↗	↘	↘	↔	↗	↔	↔	↔	↘	↘	↘
SCNS	↘	↗	↘	↗	↗	↗	↗	↗	↔	↘	↘	↗
LEAP	↔	↗	↔	↔	↘	↘	↔	↔	↗	↔	↗	↗
SINCERITIES	↔	↘	↘	↘	↔	↗	↗	↗	↔	↘	↘	↘
GRNVBEM	↔	↗	↘	↔	↘	↗	↔	↔	↗	↗	↔	↔
SCRIBE	↔	↗	↘	↔	↔	↗	↔	↗	↔	↘	↘	↘

(b) Statistics of top-k edges



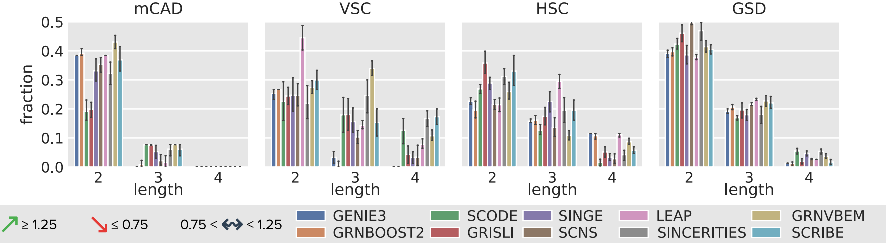
(c) Shortest path length distributions

↗ ≥ 1.25 ↘ ≤ 0.75 0.75 < ↔ < 1.25

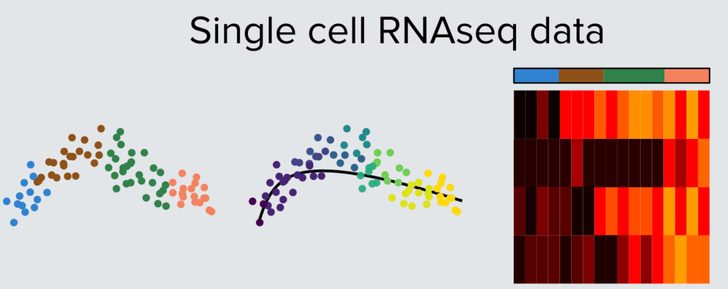
GENIE3, GRNBOOST2, SCODE, GRISLI, SINGE, SCNS, LEAP, SINCERITIES, GRNVBEM, SCRIBE

Indirect Edges in Inferred GRNs

(c) Shortest path length distributions

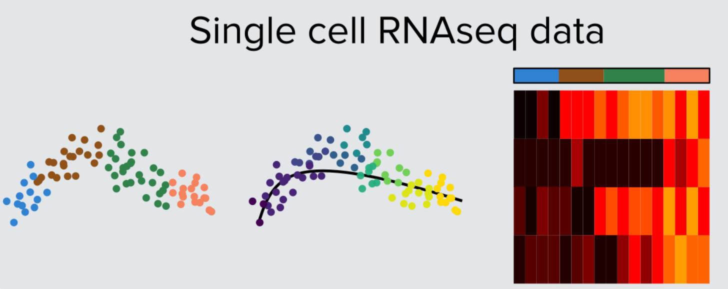


Input Type 3: Experimental sc-RNA Seq Data



- What is the motivation? Haven't we already demonstrated GRN methods, by and large, do not make good predictions?

Input Type 3: Experimental sc-RNA Seq Data

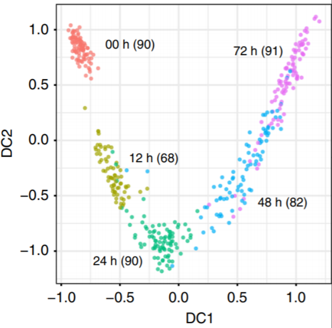


- What is the motivation? Haven't we already demonstrated GRN methods, by and large, do not make good predictions?
- *"The evaluation of GRN inference methods on experimental scRNA-seq data sets is essential."*
- *"Scale the simulations to more genes and sparser networks."*
- *"Test at least the top-performing tools on real data."*

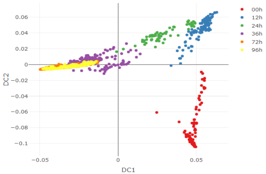
Experimental scRNA-seq Datasets

Mouse

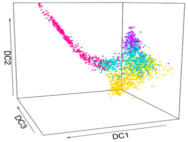
Embryonic stem cells



Dendritic cells

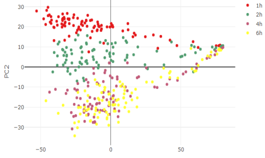


Hematopoietic stem cells

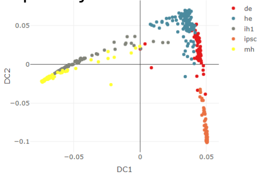


Human

Embryonic stem cells



Hepatocytes



[1] *Nestorowa, et al. (2016) "A single-cell ...". Blood, 128, 20–31.*

[2] *Hayashi et al. (2018) "Single-cell ..." Nat. Commun,. 9, 619.*

[3] *Shalek et al. (2014) "Single-cell RNA-seq ..." Nature, 510, 363–369.*

[4] *Camp et al. (2017) "Multilineage communication ..." Nature, 546, 533–538 .*

[5] *Chu et al. (2016) "Single-cell ..." Genome Biol, 17, 173.*

Characteristics of scRNA-Seq Datasets

Dataset	**Species**	**Starting cell type**	**Ending cell type(s)**	**#Cells**	**#Genes**	**# TFs**
mHSC-E[1]	Mouse	HSCs	Erythroid	1,071	2,634	204
mHSC-L[1]			Lymphoid	847	692	60
mHSC-GM[1]			Granulocyte-Macrophage	889	1,595	132
mESC[2]	Mouse	mESCs	Primitive Endoderm	421	8,150	620
mDC[3]	Mouse	DCs	Dendritic cells	383	3,755	321
hHep[4]	Human	iPSCs	Mature Hepatocytes	425	4,336	311
hESC[5]	Human	hESCs	Definitive Endoderm	758	4,406	330

[1] *Nestorowa, S. et al. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood (2016).*

[2] *Hayashi, T. et al. Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. Nat. Commun. (2018).*

[3] *Shalek, A. K. et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature (2014).*

[4] *Camp, J. G. et al. Multilineage communication regulates human liver bud development from pluripotency. Nature (2017).*

[5] *Chu, L. F. et al. Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. Genome Biol. (2016).*

Characteristics of scRNA-Seq Datasets

Dataset	**Species**	**Starting cell type**	**Ending cell type(s)**	**#Cells**	**#Genes**	**# TFs**
mHSC-E[1]	Mouse	HSCs	Erythroid	1,071	2,634	204
mHSC-L[1]			Lymphoid	847	692	60
mHSC-GM[1]			Granulocyte-Macrophage	889	1,595	132
mESC[2]	Mouse	mESCs	Primitive Endoderm	421	8,150	620
mDC[3]	Mouse	DCs	Dendritic cells	383	3,755	321
hHep[4]	Human	iPSCs	Mature Hepatocytes	425	4,336	311
hESC[5]	Human	hESCs	Definitive Endoderm	758	4,406	330

- Why do we need to specify starting cell type?

[1] *Nestorowa, S. et al. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood (2016).*

[2] *Hayashi, T. et al. Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. Nat. Commun. (2018).*

[3] *Shalek, A. K. et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature (2014).*

[4] *Camp, J. G. et al. Multilineage communication regulates human liver bud development from pluripotency. Nature (2017).*

[5] *Chu, L. F. et al. Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. Genome Biol. (2016).*

Characteristics of scRNA-Seq Datasets

Dataset	**Species**	**Starting cell type**	**Ending cell type(s)**	**#Cells**	**#Genes**	**# TFs**
mHSC-E[1]	Mouse	HSCs	Erythroid	1,071	2,634	204
mHSC-L[1]			Lymphoid	847	692	60
mHSC-GM[1]			Granulocyte-Macrophage	889	1,595	132
mESC[2]	Mouse	mESCs	Primitive Endoderm	421	8,150	620
mDC[3]	Mouse	DCs	Dendritic cells	383	3,755	321
hHep[4]	Human	iPSCs	Mature Hepatocytes	425	4,336	311
hESC[5]	Human	hESCs	Definitive Endoderm	758	4,406	330

- Why do we need to specify starting cell type?
- Were there any multi-lineage datasets?

---

[1] *Nestorowa, S. et al. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood (2016).*

[2] *Hayashi, T. et al. Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. Nat. Commun. (2018).*

[3] *Shalek, A. K. et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature (2014).*

[4] *Camp, J. G. et al. Multilineage communication regulates human liver bud development from pluripotency. Nature (2017).*

[5] *Chu, L. F. et al. Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. Genome Biol. (2016).*

Characteristics of scRNA-Seq Datasets

Dataset	**Species**	**Starting cell type**	**Ending cell type(s)**	**#Cells**	**#Genes**	**# TFs**
mHSC-E[1]	Mouse	HSCs	Erythroid	1,071	2,634	204
mHSC-L[1]			Lymphoid	847	692	60
mHSC-GM[1]			Granulocyte-Macrophage	889	1,595	132
mESC[2]	Mouse	mESCs	Primitive Endoderm	421	8,150	620
mDC[3]	Mouse	DCs	Dendritic cells	383	3,755	321
hHep[4]	Human	iPSCs	Mature Hepatocytes	425	4,336	311
hESC[5]	Human	hESCs	Definitive Endoderm	758	4,406	330

- Why do we need to specify starting cell type?
- Were there any multi-lineage datasets?
- Were there any time-series datasets?

[1] *Nestorowa, S. et al. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood (2016).*

[2] *Hayashi, T. et al. Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. Nat. Commun. (2018).*

[3] *Shalek, A. K. et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature (2014).*

[4] *Camp, J. G. et al. Multilineage communication regulates human liver bud development from pluripotency. Nature (2017).*

[5] *Chu, L. F. et al. Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. Genome Biol. (2016).*

Characteristics of scRNA-Seq Datasets

Dataset	**Species**	**Starting cell type**	**Ending cell type(s)**	**#Cells**	**#Genes**	**# TFs**
mHSC-E[1]	Mouse	HSCs	Erythroid	1,071	2,634	204
mHSC-L[1]			Lymphoid	847	692	60
mHSC-GM[1]			Granulocyte-Macrophage	889	1,595	132
mESC[2]	Mouse	mESCs	Primitive Endoderm	421	8,150	620
mDC[3]	Mouse	DCs	Dendritic cells	383	3,755	321
hHep[4]	Human	iPSCs	Mature Hepatocytes	425	4,336	311
hESC[5]	Human	hESCs	Definitive Endoderm	758	4,406	330

- Why do we need to specify starting cell type?
- Were there any multi-lineage datasets?
- Were there any time-series datasets?
- What do the numbers in "#genes" and "#TFs" denote?

---

[1] *Nestorowa, S. et al. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood (2016).*

[2] *Hayashi, T. et al. Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. Nat. Commun. (2018).*

[3] *Shalek, A. K. et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature (2014).*

[4] *Camp, J. G. et al. Multilineage communication regulates human liver bud development from pluripotency. Nature (2017).*

[5] *Chu, L. F. et al. Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. Genome Biol. (2016).*

Characteristics of scRNA-Seq Datasets

Dataset	Species	Starting cell type	Ending cell type(s)	#Cells	#Genes	# TFs
mHSC-E[1]	Mouse	HSCs	Erythroid	1,071	2,634	204
mHSC-L[1]			Lymphoid	847	692	60
mHSC-GM[1]			Granulocyte-Macrophage	889	1,595	132
mESC[2]	Mouse	mESCs	Primitive Endoderm	421	8,150	620
mDC[3]	Mouse	DCs	Dendritic cells	383	3,755	321
hHep[4]	Human	iPSCs	Mature Hepatocytes	425	4,336	311
hESC[5]	Human	hESCs	Definitive Endoderm	758	4,406	330

- Why do we need to specify starting cell type?
- Were there any multi-lineage datasets?
- Were there any time-series datasets?
- What do the numbers in “#genes” and “#TFs” denote?
- How did we select genes for analysis?

[1] *Nestorowa, S. et al. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood (2016).*

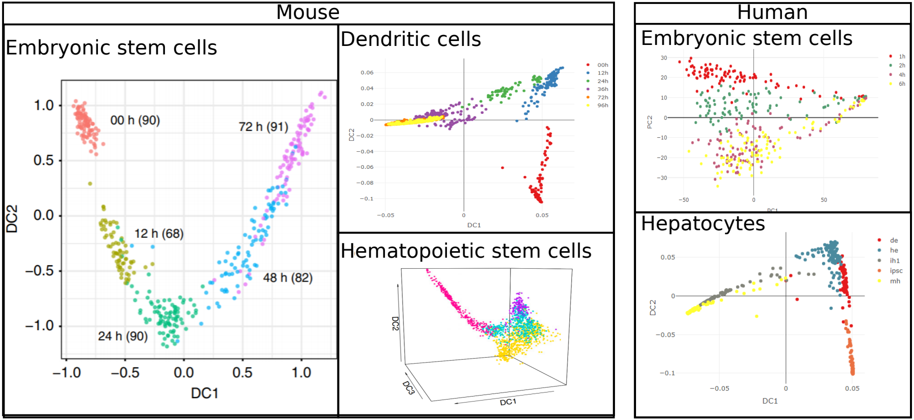
[2] *Hayashi, T. et al. Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. Nat. Commun. (2018).*

[3] *Shalek, A. K. et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature (2014).*

[4] *Camp, J. G. et al. Multilineage communication regulates human liver bud development from pluripotency. Nature (2017).*

[5] *Chu, L. F. et al. Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. Genome Biol. (2016).*

Experimental scRNA-seq Datasets


Mouse

Embryonic stem cells

Dendritic cells

Hematopoietic stem cells

Human

Embryonic stem cells

Hepatocytes

No standard ground-truth networks

Ground-Truth Networks

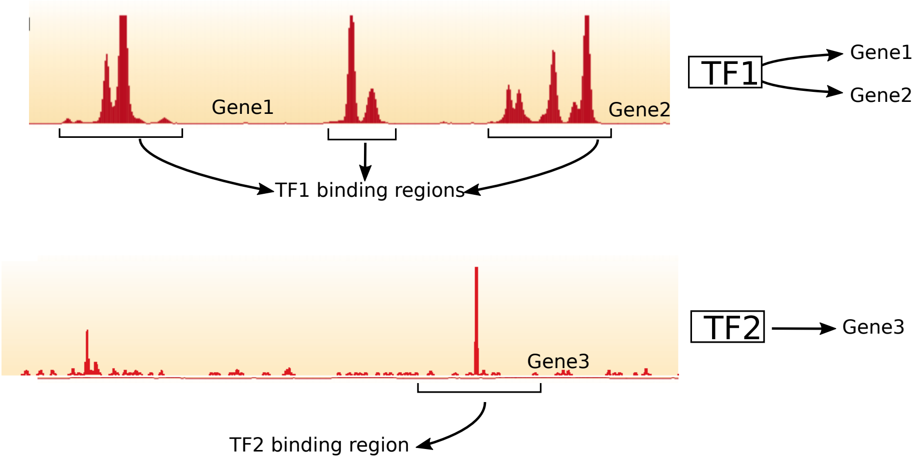
- How many types of networks?
- What was the rationale for selecting them?
- What are the differences between these networks?

Ground-Truth Networks

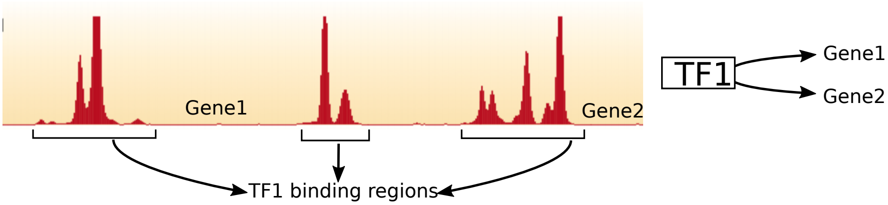
- How many types of networks?
- What was the rationale for selecting them?
- What are the differences between these networks?

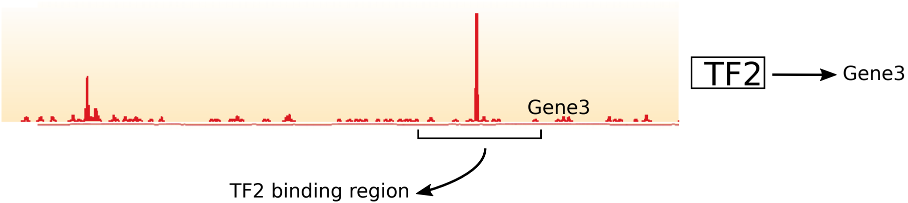
1. Cell-type-specific GRNs
2. Non-specific GRNs
3. STRING

Cell-Type Specific ChIP-seq Network



Cell-Type Specific ChIP-seq Network





Very few TFs tested for each cell type.

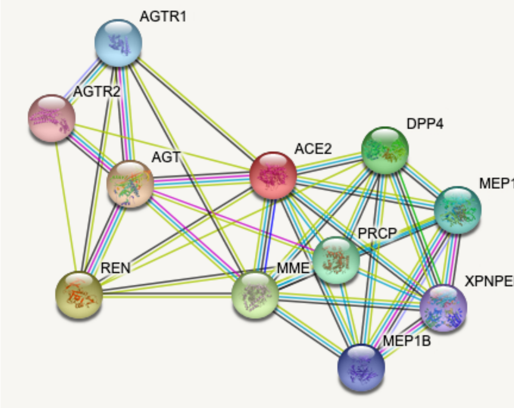
Non-Specific ChIP-Seq Network

- Curated TF-gene interactions from
  1. RegNetwork
  2. TRRUST
  3. DoRothEA
- Network is the union of these datasets.
- 292 TFs, 1,142 genes, 4,597 interactions.

Liu et al., "RegNetwork: an integrated ..." Database, 2015
Han et al., "TRRUSTv2 ..." NAR, 46(D1):D380–D386, 2018
Garcia-Alonso et al., "Benchmark ...", Gen. Res., 29:1363–1375, 2019

STRING

- Organism-specific STRING protein interaction network
- Both direct and indirect interactions
- Experimental and computed interactions

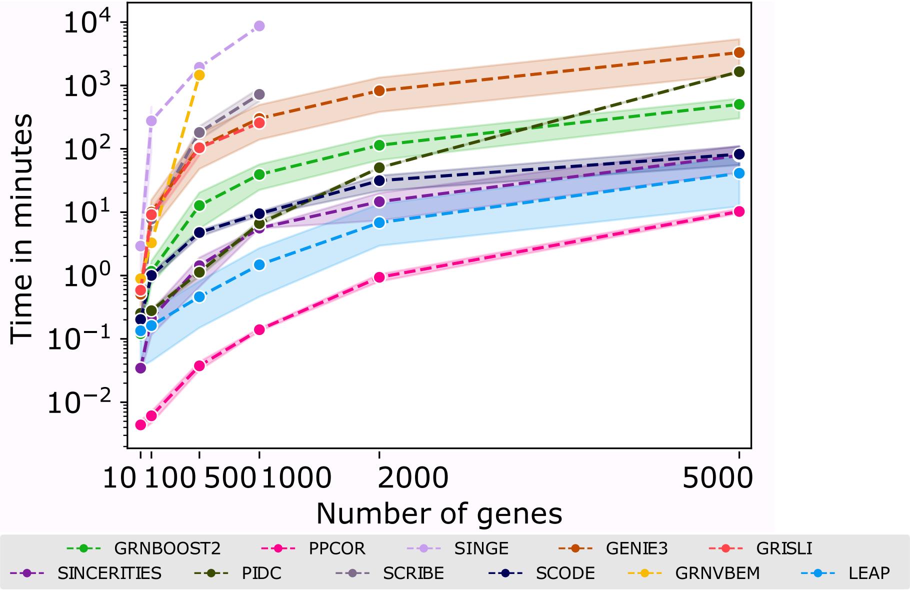


Szklarczyk et al. (2019) "STRING v11..." NAR, 47:D607–613

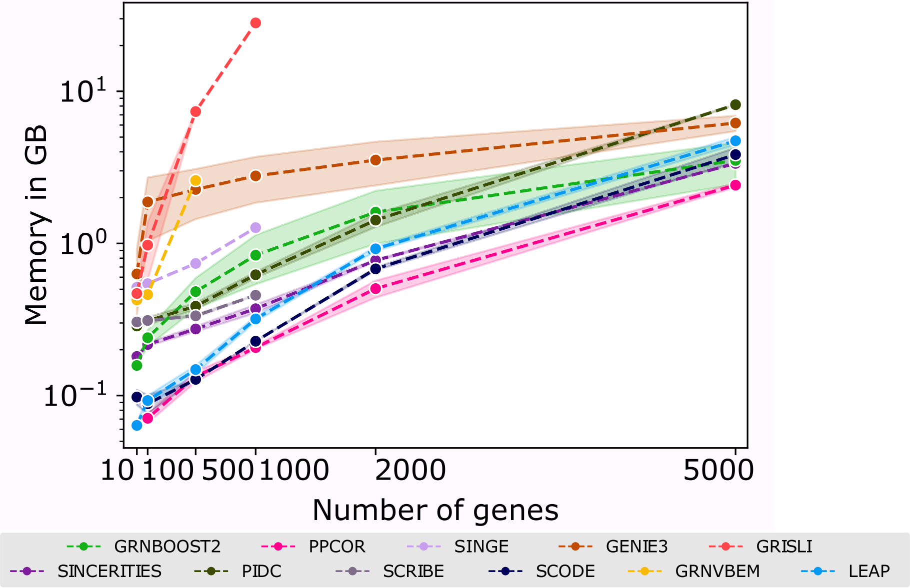
Statistics on Ground-Truth Networks

Species	**Source**	**#TFs**	**#Genes (incl. TFs)**	**#Edges**	**Density**
Mouse	HSC, E, L, G-M ChIP-Atlas	137	19,324	1,078,888	0.407
Mouse	DC, ChIP-Atlas	36	11,092	30,658	0.077
Mouse	mESC, ESCAPE+ ChIP-Atlas	247	25,703	6,348,394	0.154
Mouse	mESC, LOGOF, ESCAPE	57	18,427	104,797	0.100
Mouse	TRRUST + RegNetwork	1,455	17,852	100,139	0.004
Mouse	STRING	1,350	7,771	157,134	0.015
Human	hESC, ChEA + ChIP-Atlas	130	18,104	436,563	0.186
Human	HEPG2, ChEA + ChIP-Atlas	84	16,822	342,862	0.243
Human	TRRUST + RegNetwork +DoRothEA	2,165	23,566	386,293	0.008
Human	STRING	1,489	8,806	198,285	0.015

Running Time and Memory Usage



Running Time and Memory Usage



Algorithms Evaluated

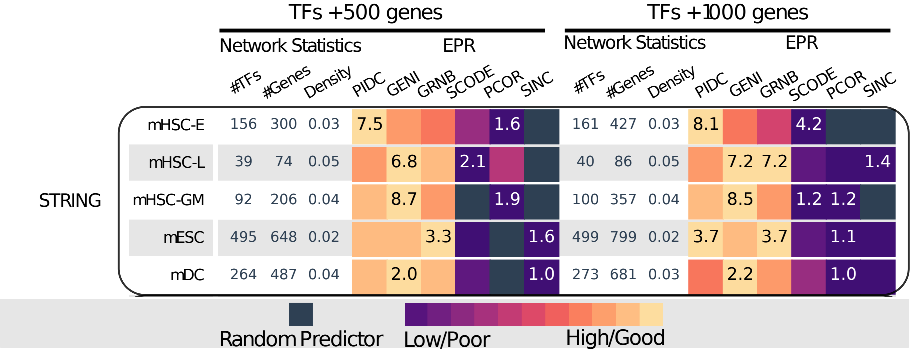
- Top five methods for synthetic networks.
- Top five methods for curated networks.

Algorithms Evaluated

- Top five methods for synthetic networks.
- Top five methods for curated networks.
- Ignored SCRIBE and SINGE because of parameter search time.
- Six methods left: GENIE3, GRNBoost2, PPCOR, PIDC, SCODE, SINCERITIES.

Results on Mouse Datasets

- Why did we use the EPR?
- How did we measure network density?
- What trends did we observe in the EPR?
- Why did PPCOR's performance drop for experimental datasets?

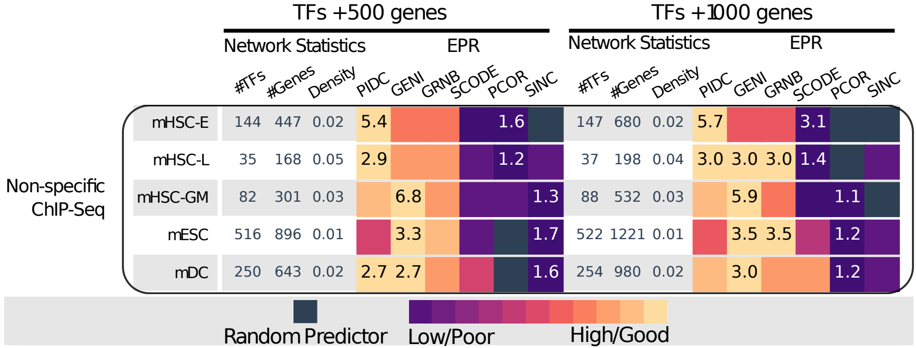


		TFs +500 genes										TFs +1000 genes									
		Network Statistics			EPR							Network Statistics			EPR						
		#TFs	#Genes	Density	PIDC	GENI	GRNB	SCODE	PCOR	SINC		#TFs	#Genes	Density	PIDC	GENI	GRNB	SCODE	PCOR	SINC	
STRING	mHSC-E	156	300	0.03	7.5				1.6			161	427	0.03	8.1			4.2			
	mHSC-L	39	74	0.05		6.8		2.1				40	86	0.05		7.2	7.2			1.4	
	mHSC-GM	92	206	0.04		8.7			1.9			100	357	0.04		8.5		1.2	1.2		
	mESC	495	648	0.02			3.3			1.6		499	799	0.02	3.7		3.7		1.1		
	mDC	264	487	0.04		2.0				1.0		273	681	0.03		2.2			1.0		

Random Predictor Low/Poor High/Good

Results on Mouse Datasets

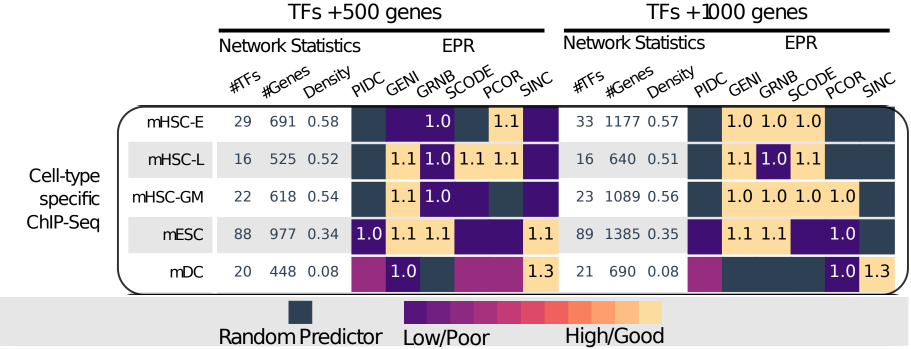
- Why did we use the EPR?
- How did we measure network density?
- What trends did we observe in the EPR?
- Why did PPCOR's performance drop for experimental datasets?



		TFs +500 genes									TFs +1000 genes								
		Network Statistics			EPR						Network Statistics			EPR					
		#TFs	#Genes	Density	PIDC	GENI	GRNB	SCODE	PCOR	SINC	#TFs	#Genes	Density	PIDC	GENI	GRNB	SCODE	PCOR	SINC
Non-specific ChIP-Seq	mHSC-E	144	447	0.02	5.4				1.6		147	680	0.02	5.7			3.1		
	mHSC-L	35	168	0.05	2.9				1.2		37	198	0.04	3.0	3.0	3.0	1.4		
	mHSC-GM	82	301	0.03		6.8				1.3	88	532	0.03		5.9			1.1	
	mESC	516	896	0.01		3.3				1.7	522	1221	0.01		3.5	3.5		1.2	
	mDC	250	643	0.02	2.7	2.7				1.6	254	980	0.02		3.0			1.2	

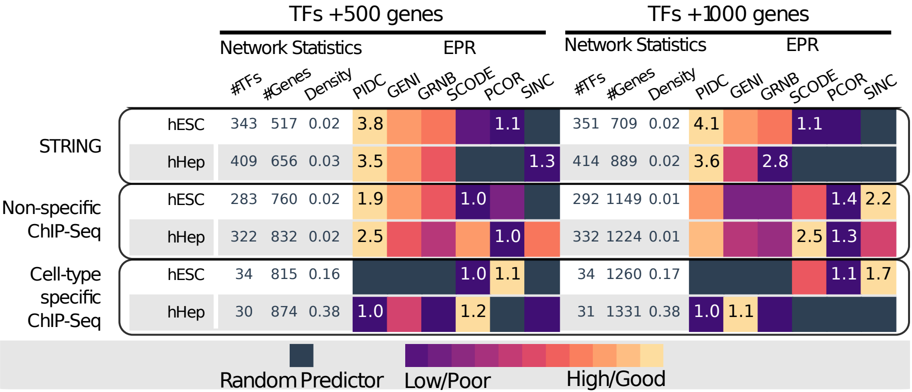
Results on Mouse Datasets

- Why did we use the EPR?
- How did we measure network density?
- What trends did we observe in the EPR?
- Why did PPCOR's performance drop for experimental datasets?



		TFs +500 genes									TFs +1000 genes								
		Network Statistics			EPR						Network Statistics			EPR					
		#TFs	#Genes	Density	PIDC	GENI	GRNB	SCODE	PCOR	SINC	#TFs	#Genes	Density	PIDC	GENI	GRNB	SCODE	PCOR	SINC
Cell-type specific ChIP-Seq	mHSC-E	29	691	0.58			1.0		1.1		33	1177	0.57		1.0	1.0	1.0		
	mHSC-L	16	525	0.52		1.1	1.0	1.1	1.1		16	640	0.51		1.1	1.0	1.1		
	mHSC-GM	22	618	0.54		1.1	1.0				23	1089	0.56		1.0	1.0	1.0	1.0	
	mESC	88	977	0.34	1.0	1.1	1.1			1.1	89	1385	0.35		1.1	1.1		1.0	
	mDC	20	448	0.08		1.0				1.3	21	690	0.08					1.0	1.3

Results on Human Datasets



		TFs +500 genes									TFs +1000 genes								
		Network Statistics			EPR						Network Statistics			EPR					
		#TFs	#Genes	Density	PIDC	GENI	GRNB	SCODE	PCOR	SINC	#TFs	#Genes	Density	PIDC	GENI	GRNB	SCODE	PCOR	SINC
STRING	hESC	343	517	0.02	3.8				1.1		351	709	0.02	4.1			1.1		
	hHep	409	656	0.03	3.5					1.3	414	889	0.02	3.6		2.8			
Non-specific ChIP-Seq	hESC	283	760	0.02	1.9			1.0			292	1149	0.01					1.4	2.2
	hHep	322	832	0.02	2.5				1.0		332	1224	0.01				2.5	1.3	
Cell-type specific ChIP-Seq	hESC	34	815	0.16				1.0	1.1		34	1260	0.17					1.1	1.7
	hHep	30	874	0.38	1.0			1.2			31	1331	0.38	1.0	1.1				

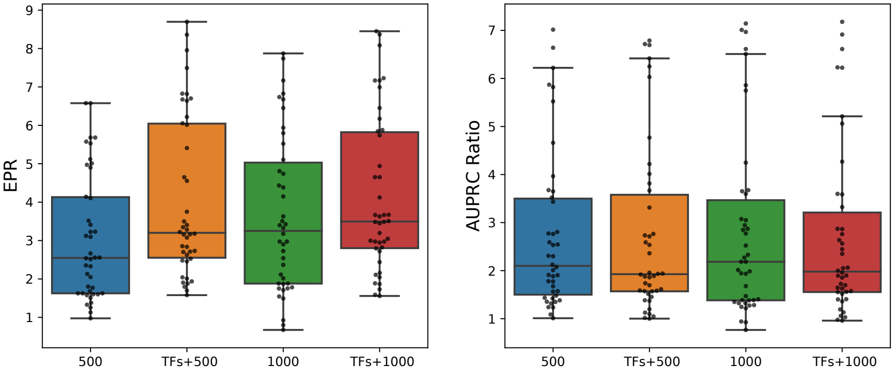
Other Analyses

Other Analyses

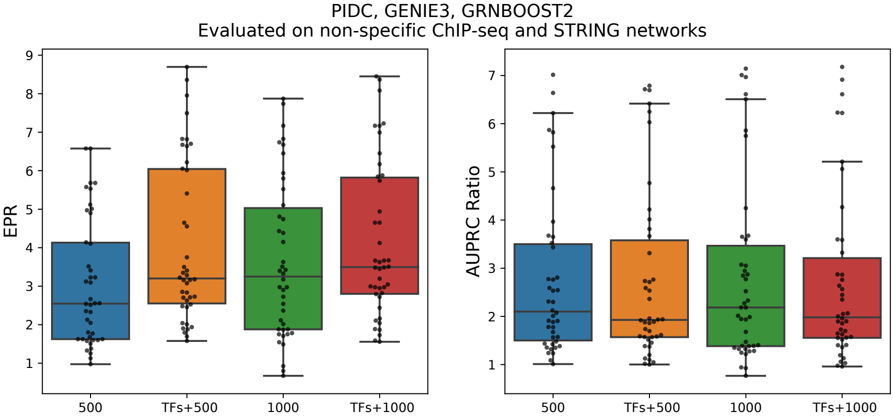
- Consensus with direct clustering of gene expression data (Supplementary Note 3.2).
- Gene selection strategy (Supplementary Note 3.3).
- Stability across runs (Supplementary Note 3.4).
- Similarity across algorithms (Supplementary Note 3.5).

Gene Selection Strategy

PIDC, GENIE3, GRNBOOST2
Evaluated on non-specific ChIP-seq and STRING networks



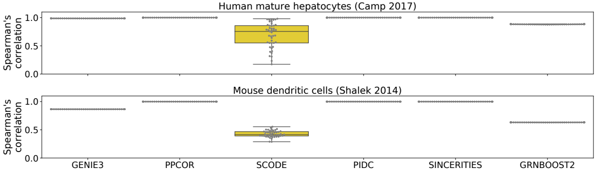
Gene Selection Strategy



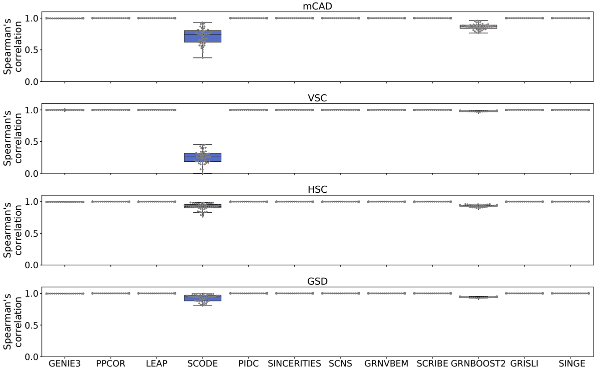
- Only evaluated algorithms with three highest median EPR values.
- Including all significantly-varying TFs had a statistically significant improvement on the EPR.
- Changing the number of genes from 500 to 1,000 did not make a significant difference.

Stability Across Runs

(a) Experimental single-cell RNA-seq datasets

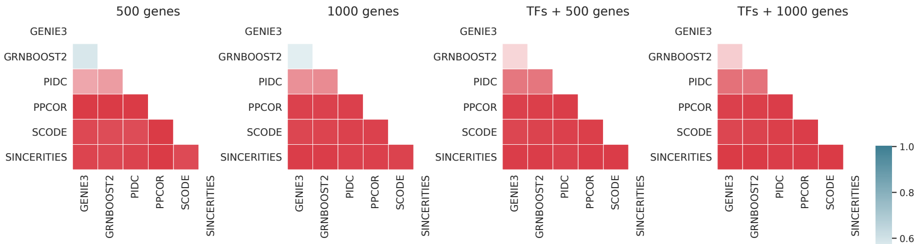


(b) Boolean networks

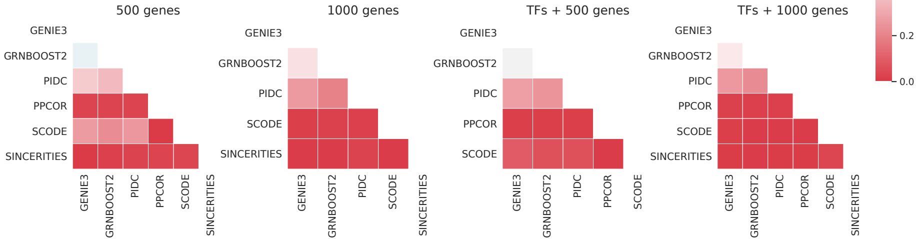


Similarity Across Algorithms

(a) Mouse embryonic stem cells (mESC)

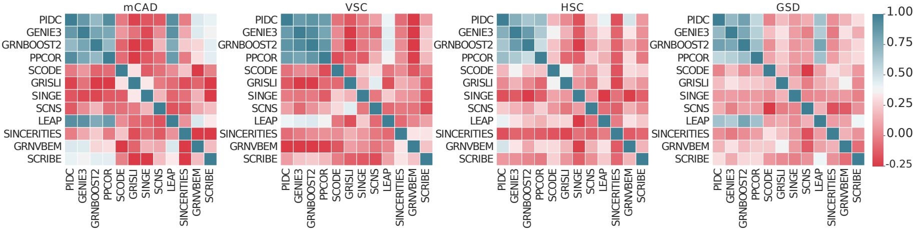


(b) Human definitive endoderm single-cell RNA-seq dataset (hESC)

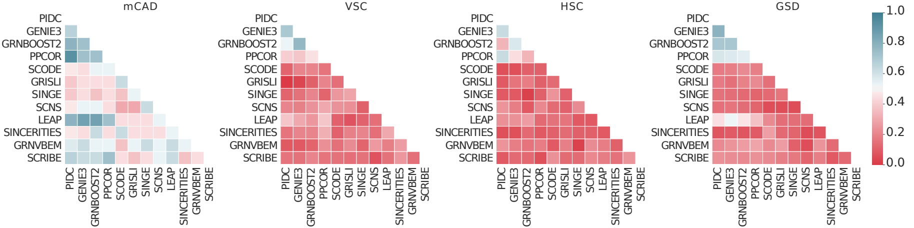


Similarity Across Algorithms

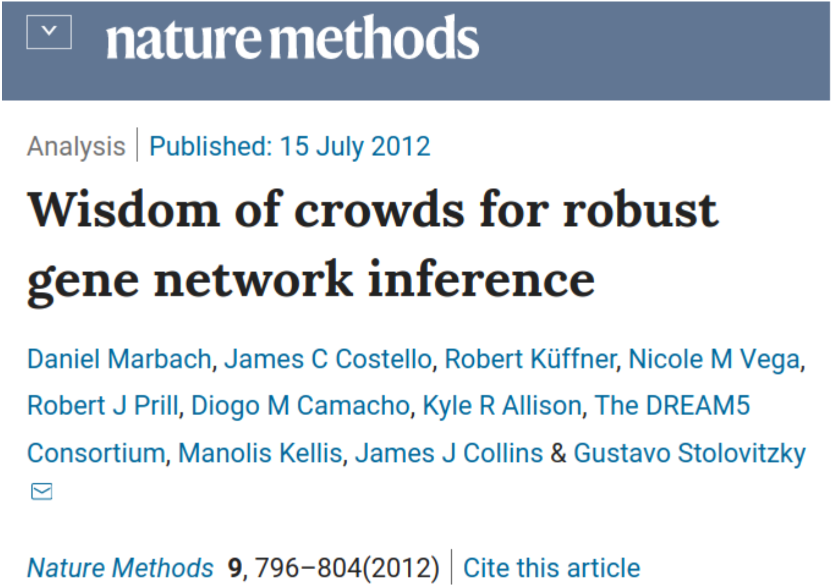
(a) Spearman correlation



(b) Jaccard index



Wisdom of Crowds for GRN Inference (Bulk RNA-Seq)



nature methods

Analysis | Published: 15 July 2012

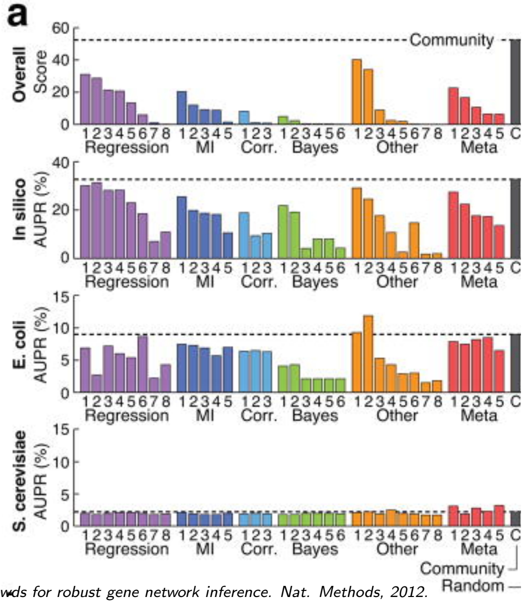
Wisdom of crowds for robust gene network inference

Daniel Marbach, James C Costello, Robert Küffner, Nicole M Vega, Robert J Prill, Diogo M Camacho, Kyle R Allison, The DREAM5 Consortium, Manolis Kellis, James J Collins & Gustavo Stolovitzky

Nature Methods **9**, 796–804(2012) | Cite this article

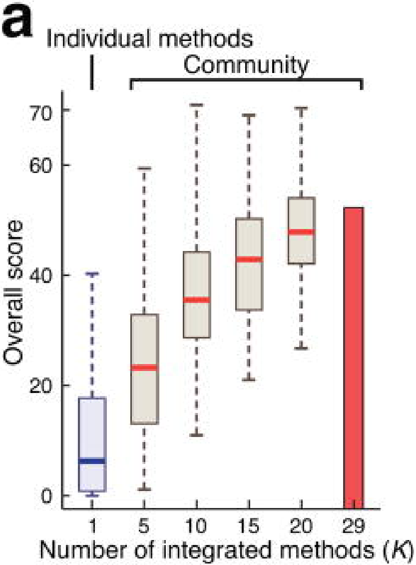
Marbach et al. Wisdom of crowds for robust gene network inference. Nat. Methods, 2012.

Wisdom of Crowds for GRN Inference (Bulk RNA-Seq)



Marbach et al. Wisdom of crowds for robust gene network inference. Nat. Methods, 2012.

Wisdom of Crowds for GRN Inference (Bulk RNA-Seq)



Marbach et al. Wisdom of crowds for robust gene network inference. Nat. Methods, 2012.

Implementing Wisdom of Crowds

Used four methods to compute ensembles.

Borda: rank of an edge = averages of its ranks computed by each GRN inference algorithm.

mBorda: modification of the Borda method: assign a weight of $1/n^2$ to rank n.

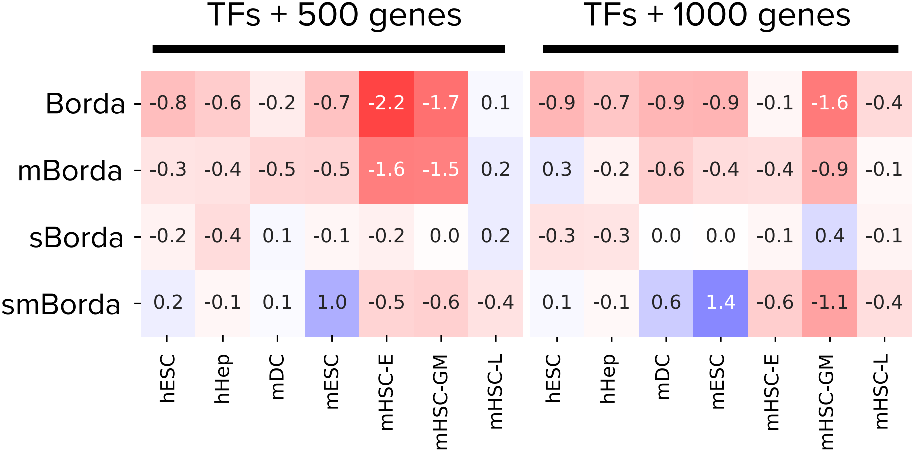
sBorda: use only the three methods with the highest AUPRC or EPR.

smBorda: combine sBorda and mBorda.

Deshpande et al. Network inference with Granger causality ensembles on single-cell transcriptomic data, bioRxiv, 2019.
Marbach et al. Wisdom of crowds for robust gene network inference. Nat. Methods, 2012.

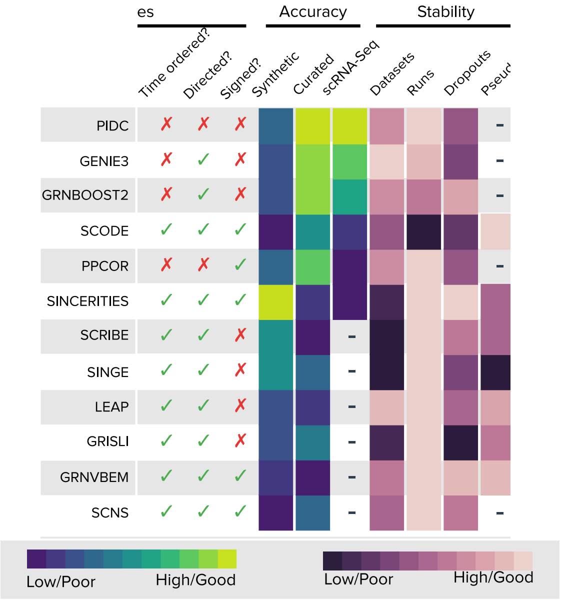
Wisdom of Crowds for GRN Inference (scRNA-Seq)

Difference in EPR between ensemble method and best individual GRN algorithm.



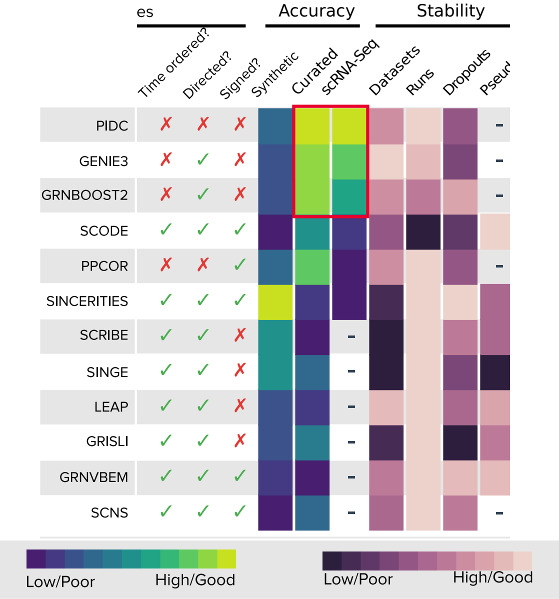
	TFs + 500 genes							TFs + 1000 genes						
	hESC	hHep	mDC	mESC	mHSC-E	mHSC-GM	mHSC-L	hESC	hHep	mDC	mESC	mHSC-E	mHSC-GM	mHSC-L
Borda	-0.8	-0.6	-0.2	-0.7	-2.2	-1.7	0.1	-0.9	-0.7	-0.9	-0.9	-0.1	-1.6	-0.4
mBorda	-0.3	-0.4	-0.5	-0.5	-1.6	-1.5	0.2	0.3	-0.2	-0.6	-0.4	-0.4	-0.9	-0.1
sBorda	-0.2	-0.4	0.1	-0.1	-0.2	0.0	0.2	-0.3	-0.3	0.0	0.0	-0.1	0.4	-0.1
smBorda	0.2	-0.1	0.1	1.0	-0.5	-0.6	-0.4	0.1	-0.1	0.6	1.4	-0.6	-1.1	-0.4

Summary of Results and Recommendations



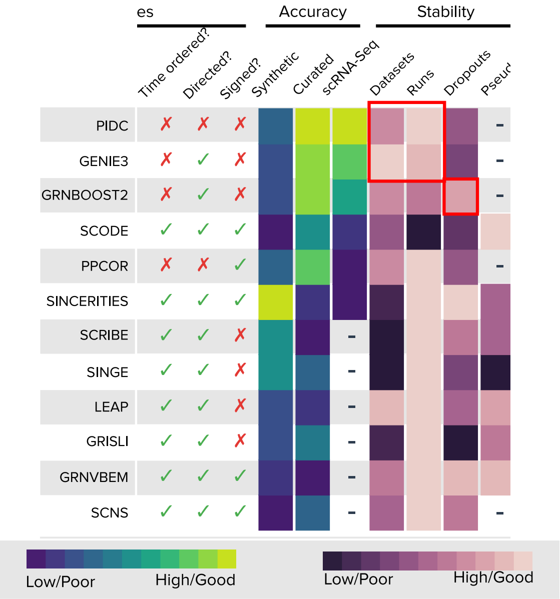
	es			Accuracy			Stability			
	Time ordered?	Directed?	Signed?	Synthetic	Curated	scRNA-Seq	Datasets	Runs	Dropouts	Pseud
PIDC	✗	✗	✗							–
GENIE3	✗	✓	✗							–
GRNBOOST2	✗	✓	✗							–
SCODE	✓	✓	✓							
PPCOR	✗	✗	✓							–
SINCERITIES	✓	✓	✓							
SCRIBE	✓	✓	✗			–				
SINGE	✓	✓	✗			–				
LEAP	✓	✓	✗			–				
GRISLI	✓	✓	✗			–				
GRNVBEM	✓	✓	✓			–				
SCNS	✓	✓	✓			–				–

Summary of Results and Recommendations



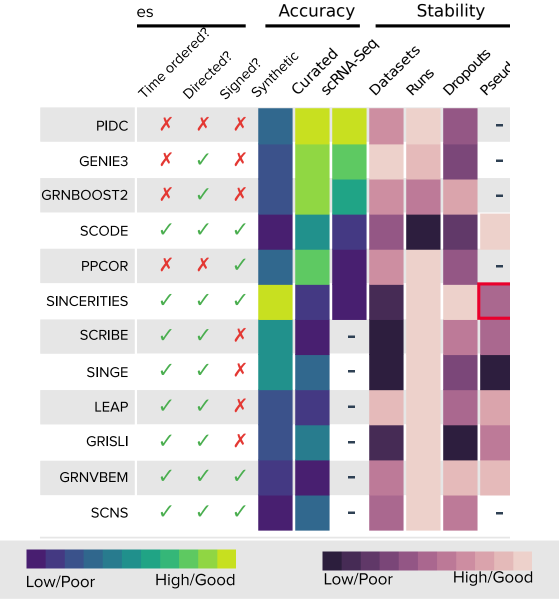
	es			Accuracy			Stability			
	Time ordered?	Directed?	Signed?	Synthetic	Curated	scRNA-Seq	Datasets	Runs	Dropouts	Pseud
PIDC	✗	✗	✗							-
GENIE3	✗	✓	✗							-
GRNBOOST2	✗	✓	✗							-
SCODE	✓	✓	✓							
PPCOR	✗	✗	✓							-
SINCERITIES	✓	✓	✓							
SCRIBE	✓	✓	✗			-				
SINGE	✓	✓	✗			-				
LEAP	✓	✓	✗			-				
GRISLI	✓	✓	✗			-				
GRNVBEM	✓	✓	✓			-				
SCNS	✓	✓	✓			-				-

Summary of Results and Recommendations



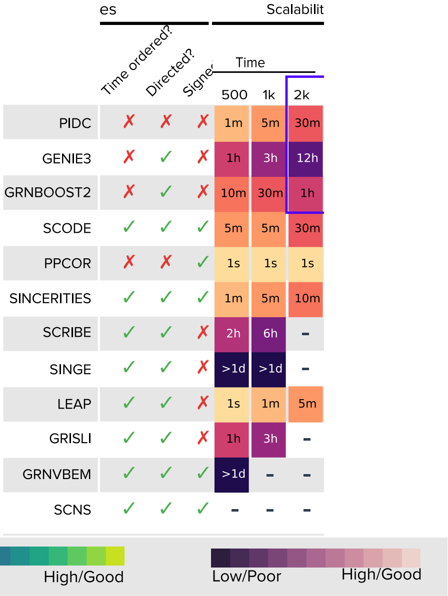
	es			Accuracy			Stability			
	Time ordered?	Directed?	Signed?	Synthetic	Curated	scRNA-Seq	Datasets	Runs	Dropouts	Pseud
PIDC	✗	✗	✗							-
GENIE3	✗	✓	✗							-
GRNBOOST2	✗	✓	✗							-
SCODE	✓	✓	✓							
PPCOR	✗	✗	✓							-
SINCERITIES	✓	✓	✓							
SCRIBE	✓	✓	✗			-				
SINGE	✓	✓	✗			-				
LEAP	✓	✓	✗			-				
GRISLI	✓	✓	✗			-				
GRNVBEM	✓	✓	✓			-				
SCNS	✓	✓	✓			-				-

Summary of Results and Recommendations



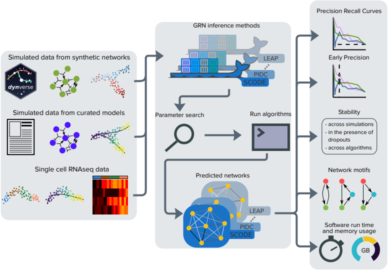
	es			Accuracy			Stability			
	Time ordered?	Directed?	Signed?	Synthetic	Curated	scRNA-Seq	Datasets	Runs	Dropouts	Pseud
PIDC	✗	✗	✗							-
GENIE3	✗	✓	✗							-
GRNBOOST2	✗	✓	✗							-
SCODE	✓	✓	✓							
PPCOR	✗	✗	✓							-
SINCERITIES	✓	✓	✓							
SCRIBE	✓	✓	✗			-				
SINGE	✓	✓	✗			-				
LEAP	✓	✓	✗			-				
GRISLI	✓	✓	✗			-				
GRNVBEM	✓	✓	✓			-				
SCNS	✓	✓	✓			-				-

Summary of Results and Recommendations



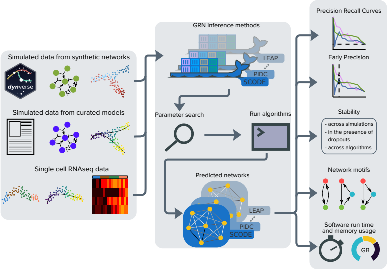
	es			Scalabilit		
				Time		
	Time ordered?	Directed?	Signe	500	1k	2k
PIDC	✗	✗	✗	1m	5m	30m
GENIE3	✗	✓	✗	1h	3h	12h
GRNBOOST2	✗	✓	✗	10m	30m	1h
SCODE	✓	✓	✓	5m	5m	30m
PPCOR	✗	✗	✓	1s	1s	1s
SINCERITIES	✓	✓	✓	1m	5m	10m
SCRIBE	✓	✓	✗	2h	6h	-
SINGE	✓	✓	✗	>1d	>1d	-
LEAP	✓	✓	✗	1s	1m	5m
GRISLI	✓	✓	✗	1h	3h	-
GRNVBEM	✓	✓	✓	>1d	-	-
SCNS	✓	✓	✓	-	-	-

Contributions of BEELINE



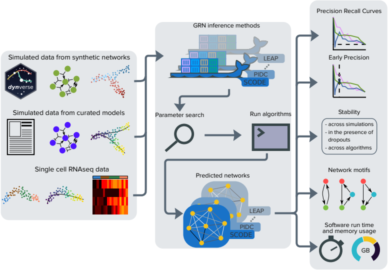
- Framework for benchmarking algorithms that infer GRNs from single-cell gene expression data.
- Facilitates reproducible, rigorous and extensible evaluations of GRN inference methods.

Contributions of BEELINE



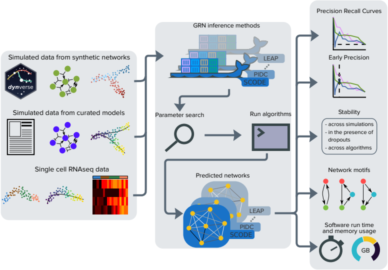
- Framework for benchmarking algorithms that infer GRNs from single-cell gene expression data.
- Facilitates reproducible, rigorous and extensible evaluations of GRN inference methods.
- BoolODE accurately simulates Boolean models with predictable trajectories.

Contributions of BEELINE



- Framework for benchmarking algorithms that infer GRNs from single-cell gene expression data.
- Facilitates reproducible, rigorous and extensible evaluations of GRN inference methods.
- BoolODE accurately simulates Boolean models with predictable trajectories.
- Recommendations to users of GRN inference algorithms.

Contributions of BEELINE



- Framework for benchmarking algorithms that infer GRNs from single-cell gene expression data.
- Facilitates reproducible, rigorous and extensible evaluations of GRN inference methods.
- BoolODE accurately simulates Boolean models with predictable trajectories.
- Recommendations to users of GRN inference algorithms.
- Software available at `https://github.com/Murali-group/BEELINE`.

BEELINE Usage in the Community

- 15 papers have used BEELINE (the full pipeline, Docker images, recommendations, BoolODE).
- 10 papers have used datasets released or collected by BEELINE.
- Most papers have developed new GRN inference algorithms.
- Small number of other applications.

BEELINE Usage in the Community

- 15 papers have used BEELINE (the full pipeline, Docker images, recommendations, BoolODE).
- 10 papers have used datasets released or collected by BEELINE.
- Most papers have developed new GRN inference algorithms.
- Small number of other applications.
- Time may be ripe for another benchmarking effort!