

# EVALUATING THE REPRODUCIBILITY OF SINGLE-CELL NETWORK INFERENCE ALGORITHMS

---

**LAURA CANTINI**

CNRS Researcher  
Computational Systems biology Lab  
Institut de Biologie de l'ENS  
Paris, France



# Laura Cantini

CNRS researcher in IBENS-ENS and chair PRAIRIE

---



# Laura Cantini

CNRS researcher in IBENS-ENS and chair PRAIRIE

---



## **METHODS FOR SINGLE-CELL MULTI-MODAL DATA INTEGRATION**

# Laura Cantini

CNRS researcher in IBENS-ENS and chair PRAIRIE

---



## METHODS FOR SINGLE-CELL MULTI-MODAL DATA INTEGRATION



# Laura Cantini

CNRS researcher in IBENS-ENS and chair PRAIRIE

---



## METHODS FOR SINGLE-CELL MULTI-MODAL DATA INTEGRATION

Jules Samaran  
(PhD)



Geert-Jan  
Huizing (PhD)



Ina Maria  
Deutschmann  
(Postdoc)

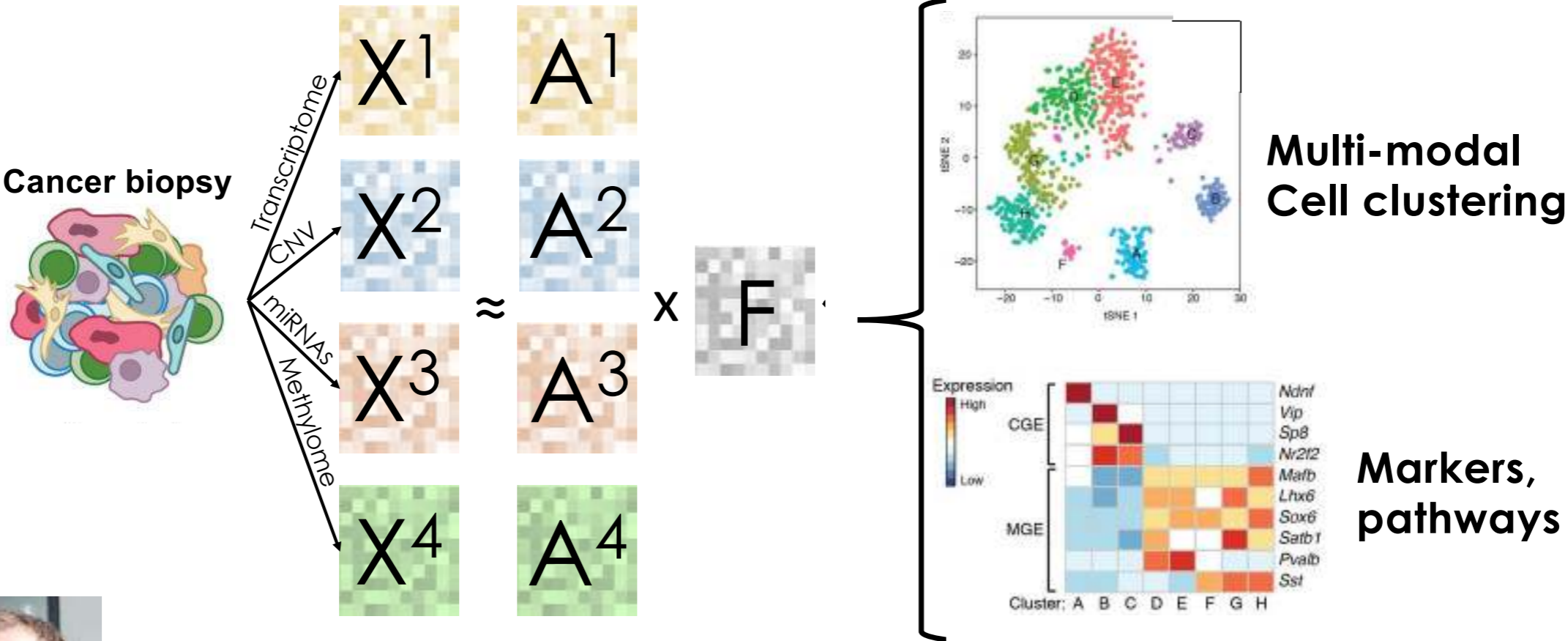


Remi Trimbour  
(M2)



# METHODS FOR SINGLE-CELL MULTI-MODAL DATA INTEGRATION

## Single-cell multi-modal joint Dimensionality Reduction (jDR)



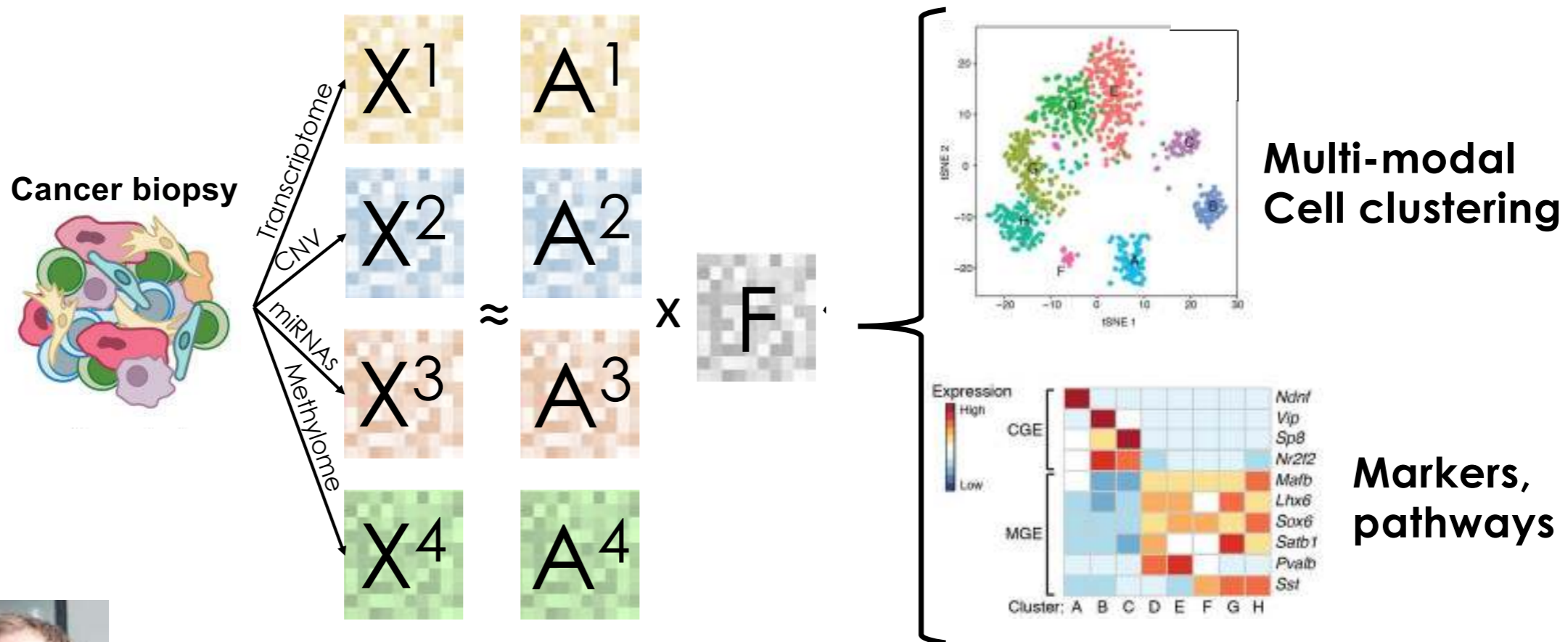
Jules Samaran (PhD)



Geert-Jan Huizing (PhD)

# METHODS FOR SINGLE-CELL MULTI-MODAL DATA INTEGRATION

## Single-cell multi-modal joint Dimensionality Reduction (jDR)



Jules Samaran (PhD)



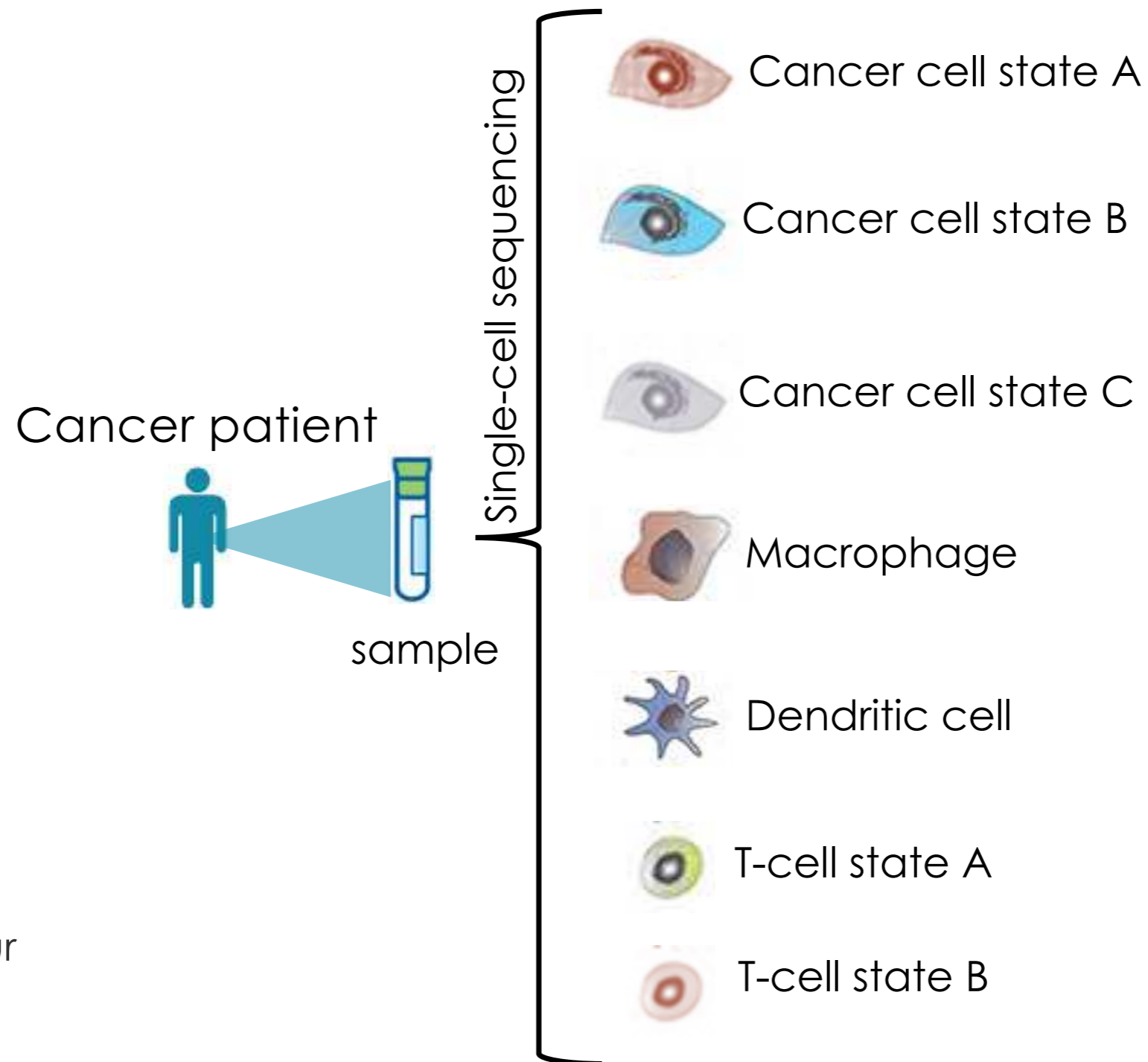
Geert-Jan Huizing (PhD)

Paired data: Multiome, CITE-seq

Unpaired data: scATAC, scRNA

# APPLICATION TO BIOLOGY

## Cell states and types identification

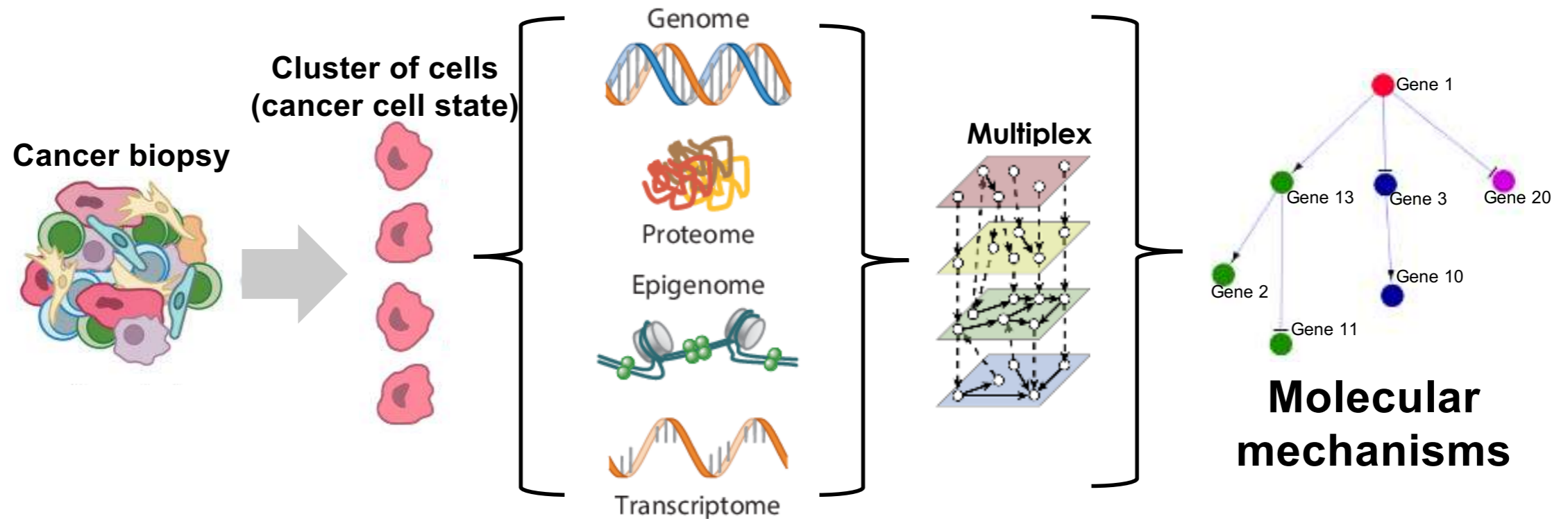


Claire  
Lassonneur  
(PhD)



# METHODS FOR SINGLE-CELL MULTI-MODAL DATA INTEGRATION

## Single-cell multi-omics multi-layer networks



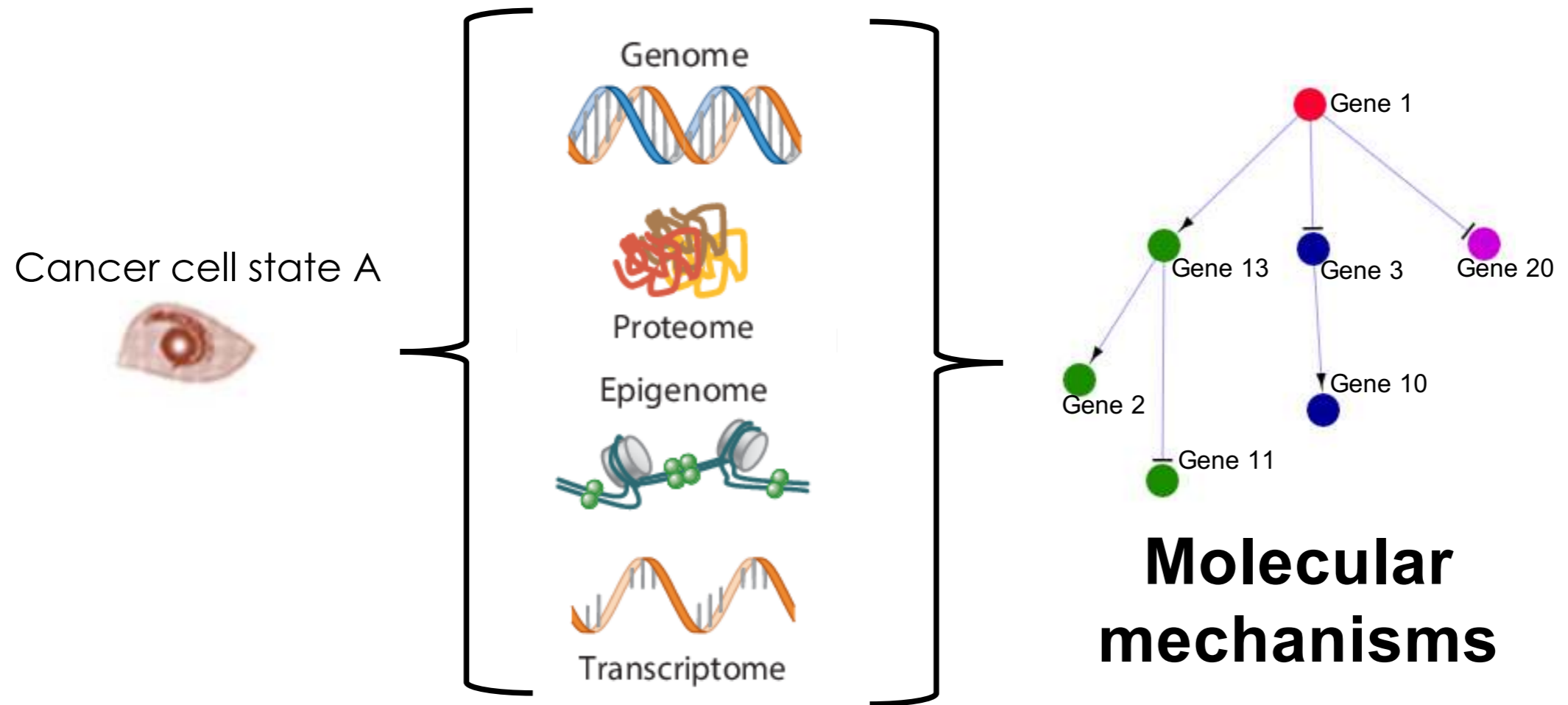
Ina Maria Deutschmann (Postdoc)



Remi Trimbour (M2)

# APPLICATION TO BIOLOGY

## Underlying molecular mechanisms



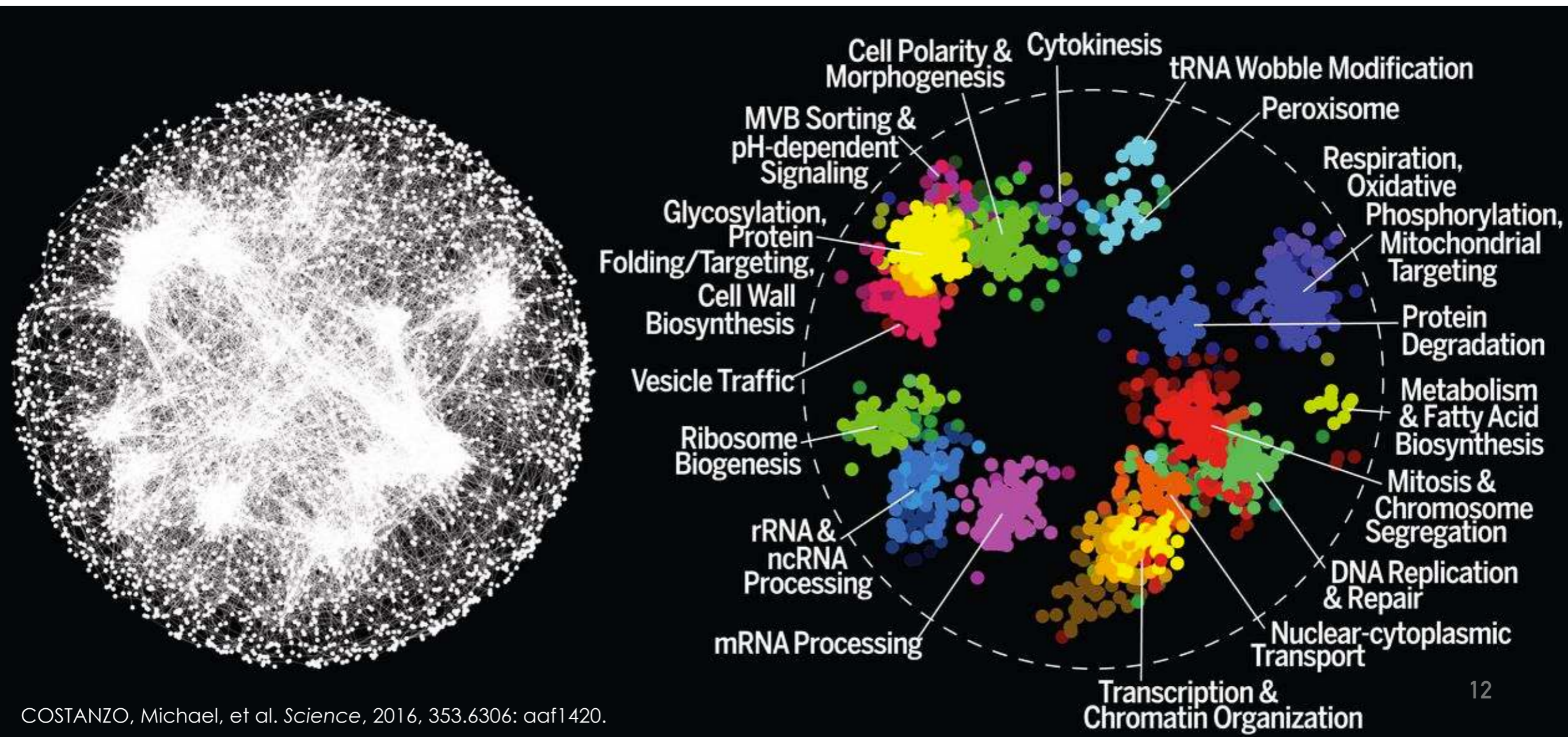
Claire  
Lassonneur  
(PhD)



# BENCHMARKING SINGLE-CELL NETWORK INFERENCE ALGORITHMS

# NETWORKS DISENTANGLE BIOLOGICAL COMPLEXITY

The cell's phenotypic behavior depends on a variety of biological macromolecules interacting at different layers of regulation



# NETWORKS IN BIOLOGY

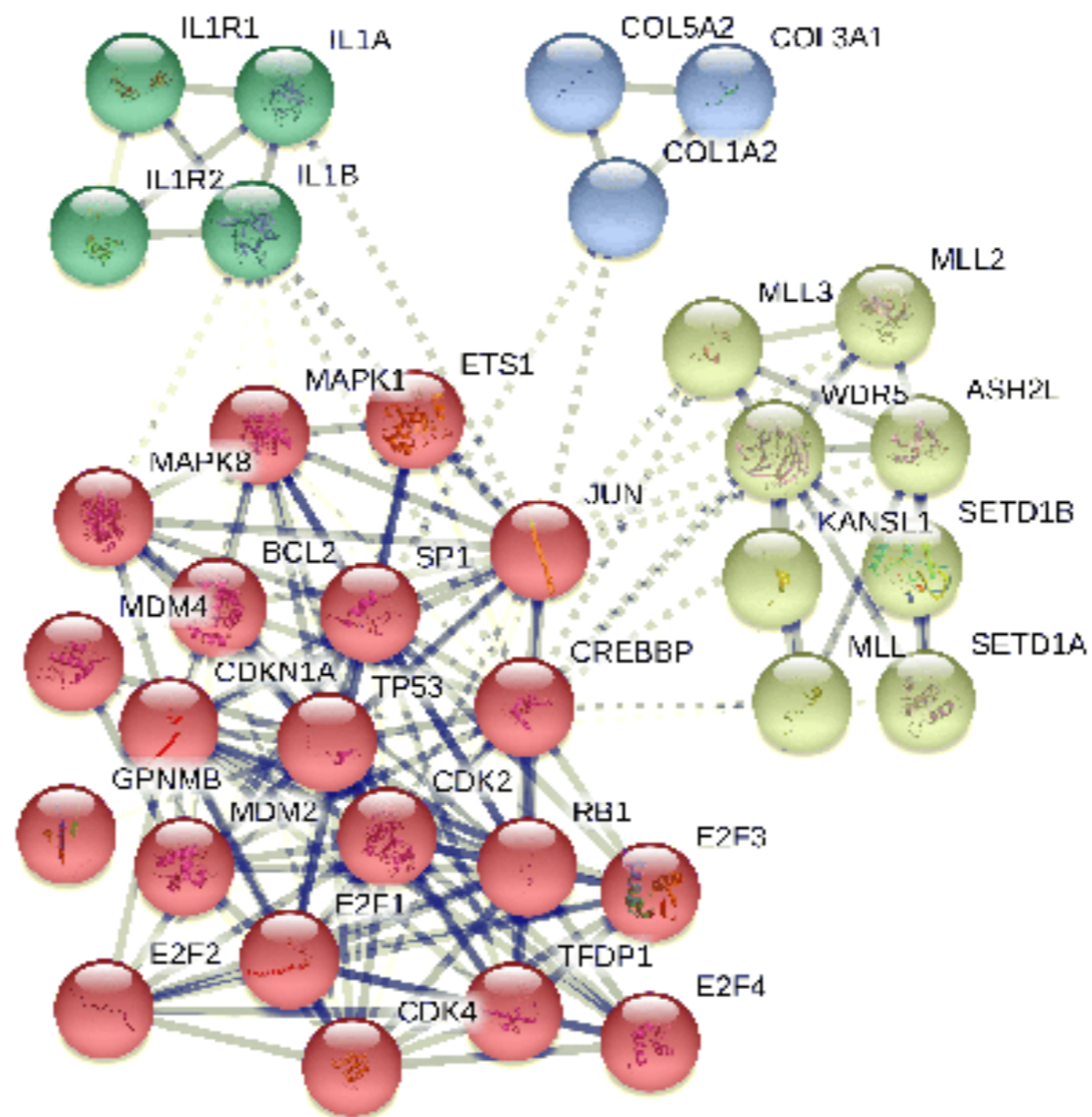
---

Networks used to model interactions between biological macromolecules (genes, proteins).

# NETWORKS IN BIOLOGY

Networks used to model interactions between biological macromolecules (genes, proteins).

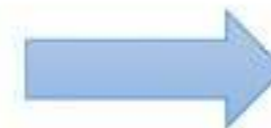
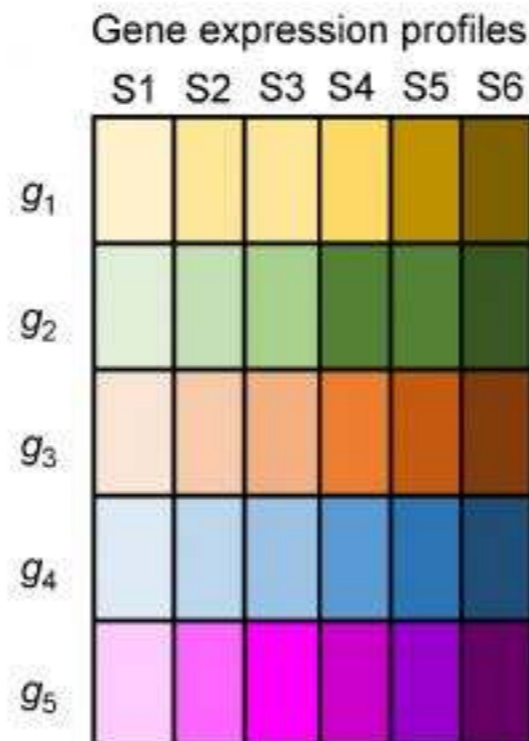
**Networks derived from experiments**



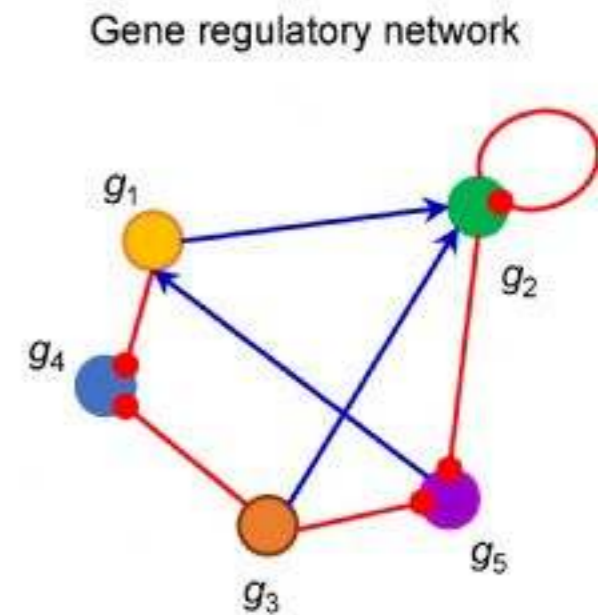
# NETWORKS IN BIOLOGY

Networks used to model interactions between biological macromolecules (genes, proteins).

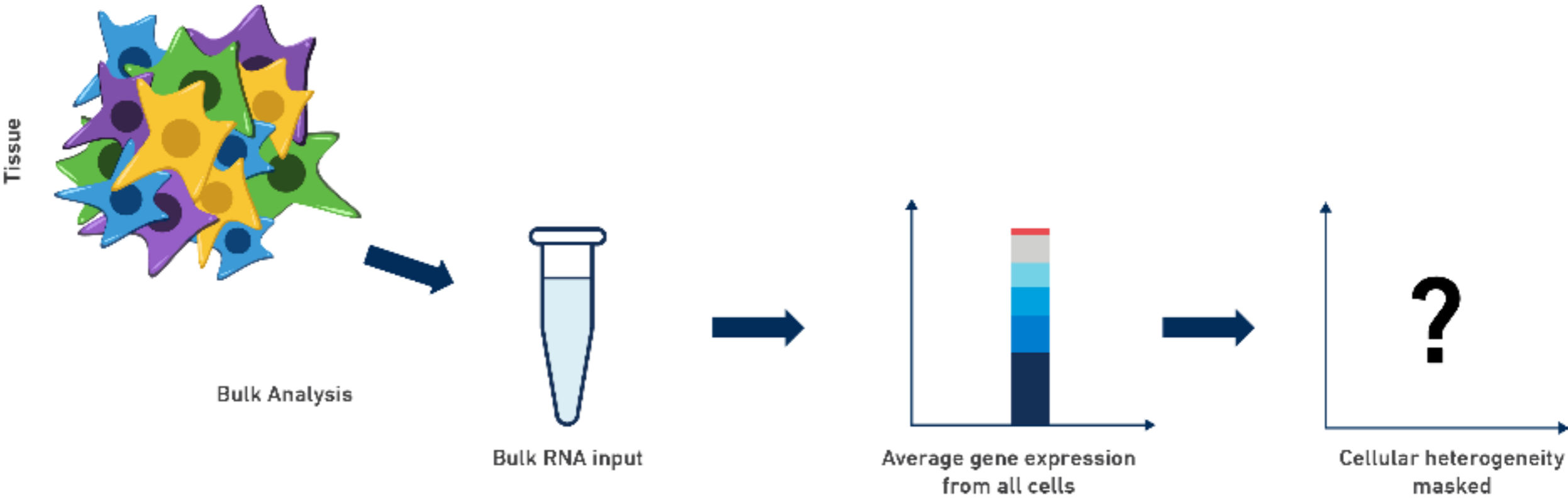
Networks inferred from data



Reverse engineering

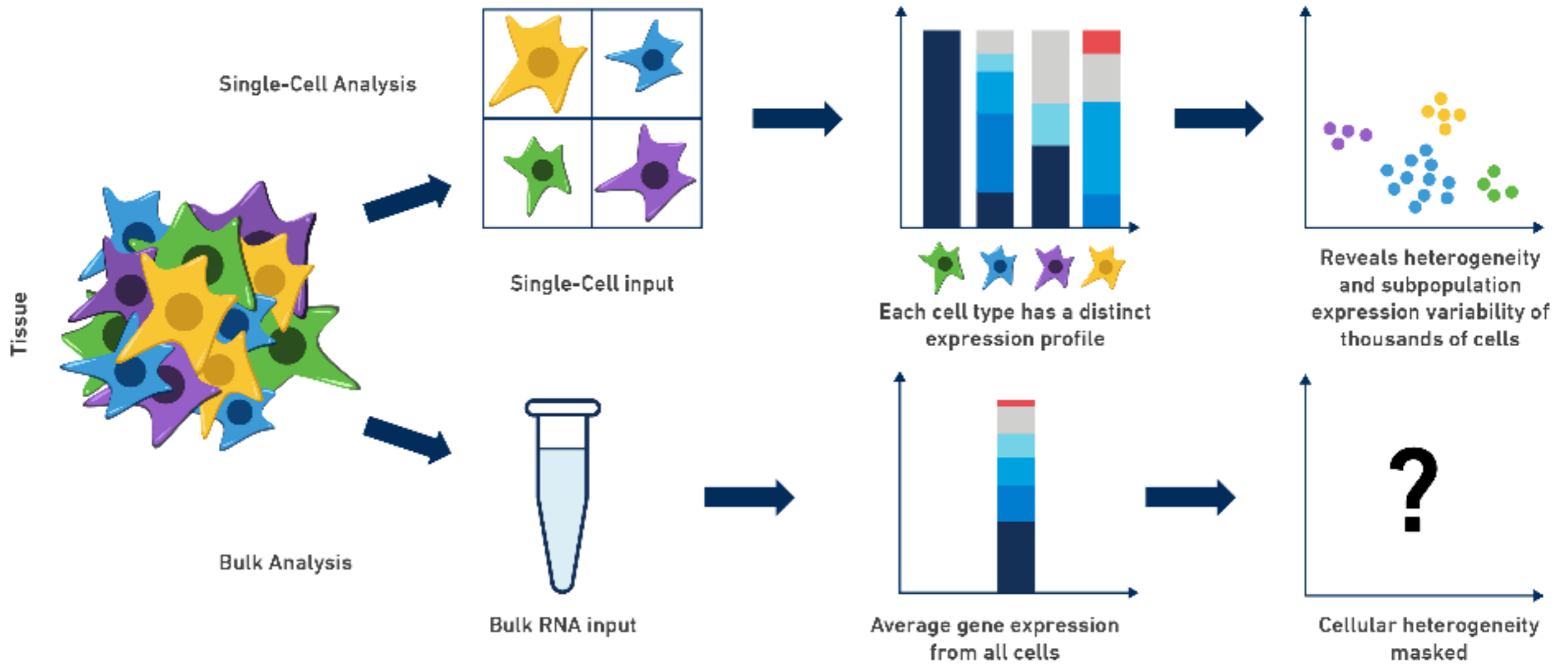


# THE ADVENT OF SCRNA-SEQ





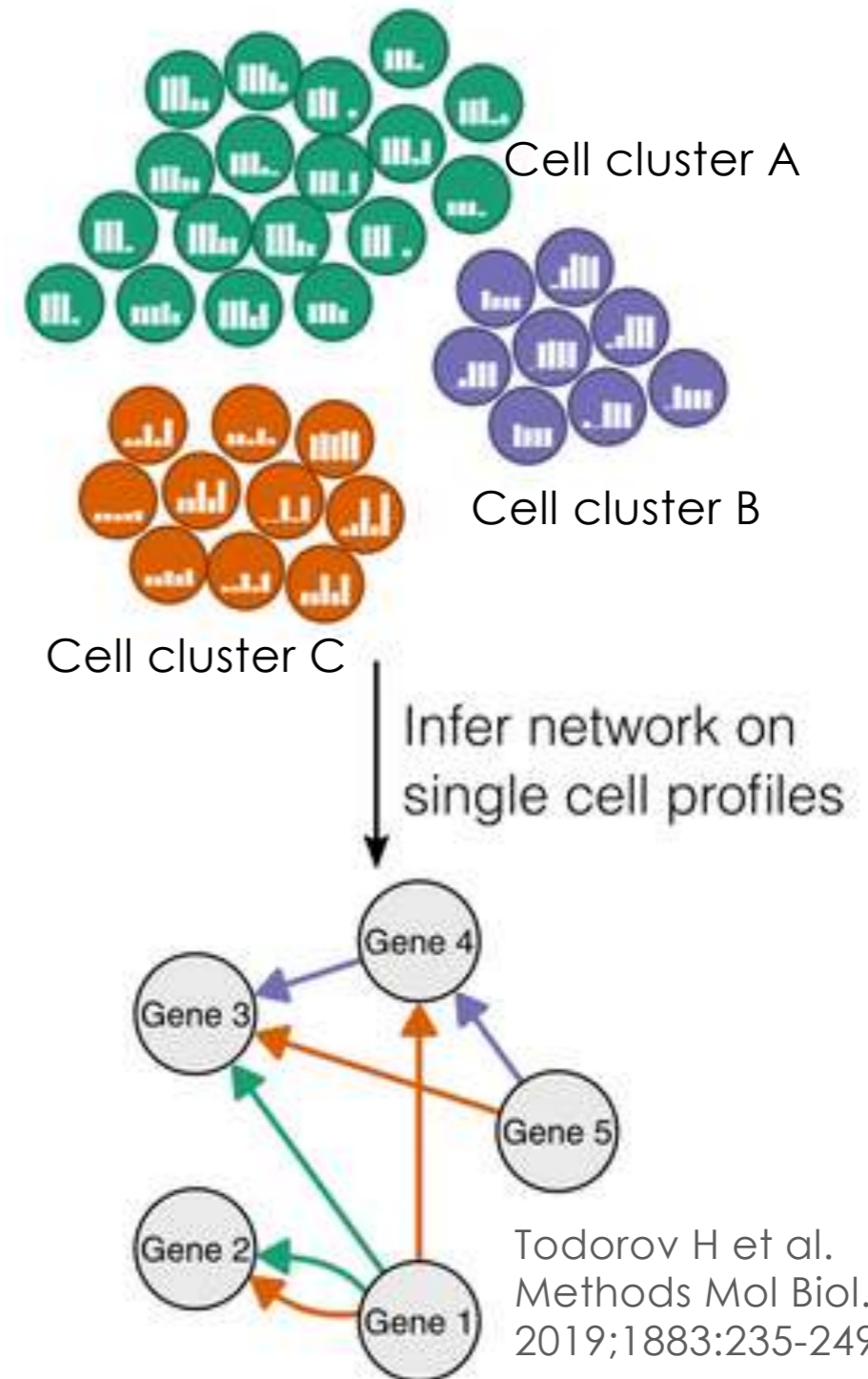
# THE ADVENT OF SCRNA-SEQ



# PROMISES AND CHALLENGES OF SCRNA-SEQ IN NETWORK INFERENCE

---

- Cell type/state regulatory programs
- Easy access to many observations
- Dealing with specificities of scRNA-seq

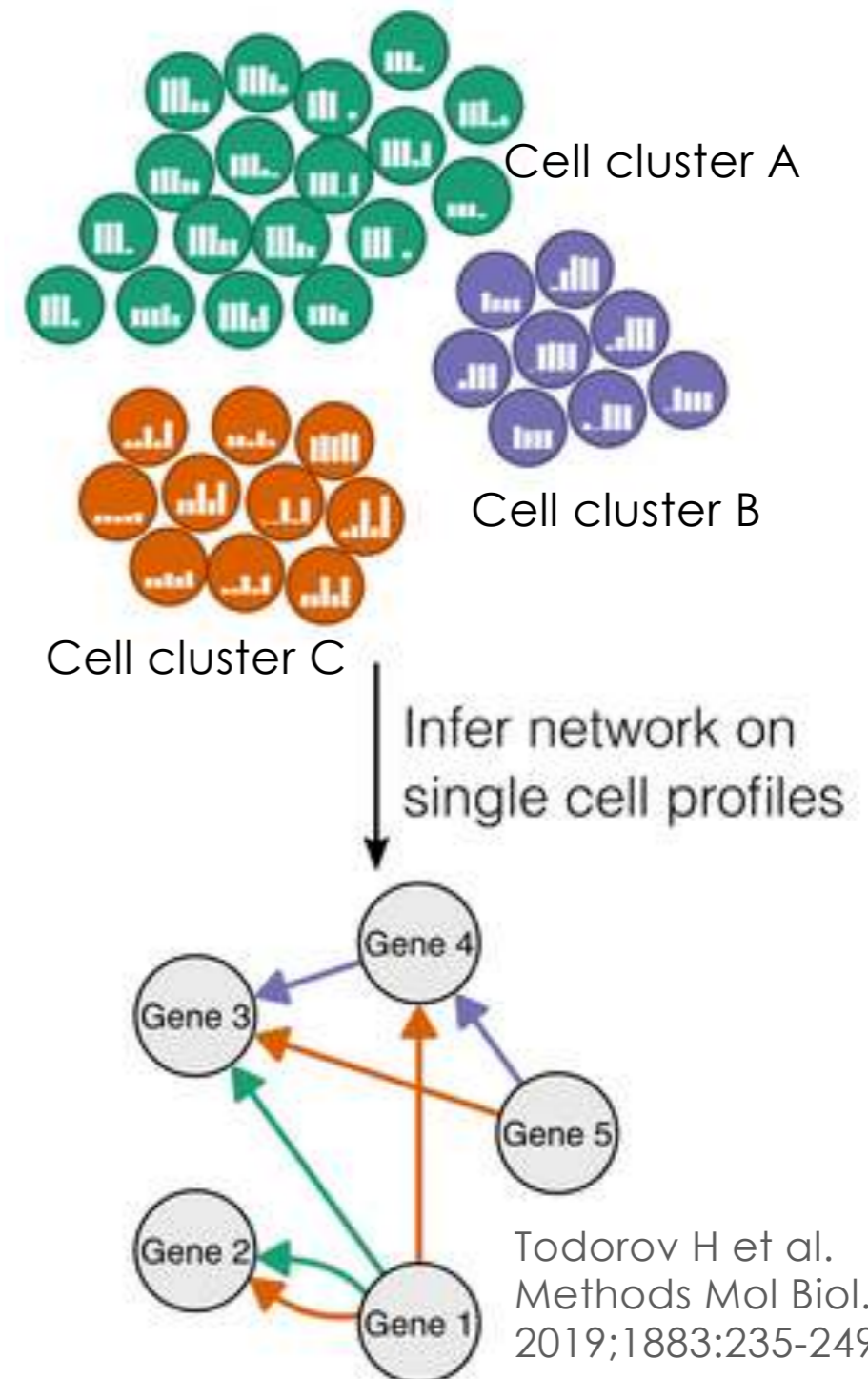


# EXISTING METHODS FOR NETWORK INFERENCE IN SCRNA-SEQ

---

Existing methods are based on:

- Use/not use pseudotime
- Adapted/ not adapted from bulk data
- Random forests (GENIE3, GRNBoost)
- Information theory (PIDC)
- Correlation (PPCOR, GeneNet)
- ODEs (GRISLIE, SCODE)



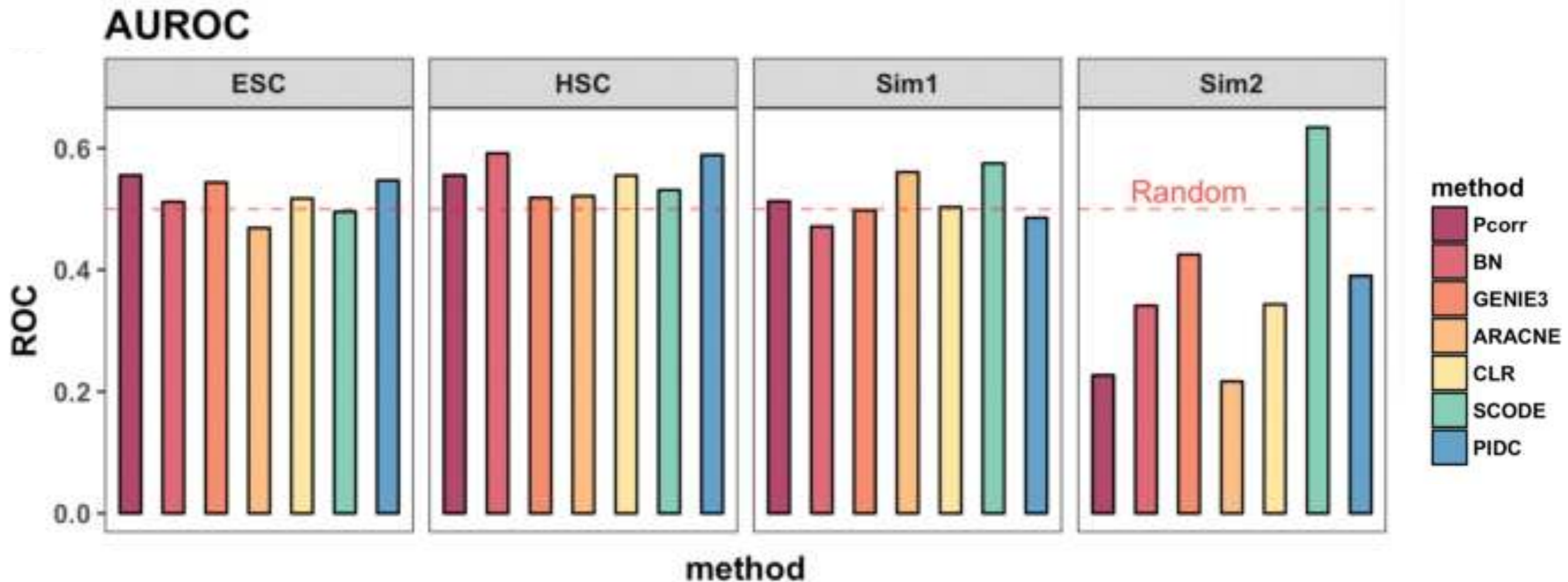
# BENCHMARKING NETWORK INFERENCE

---

- One data simulation strategy plus two real data
- In real data ground-truth: STRING
- Few genes considered 10-100

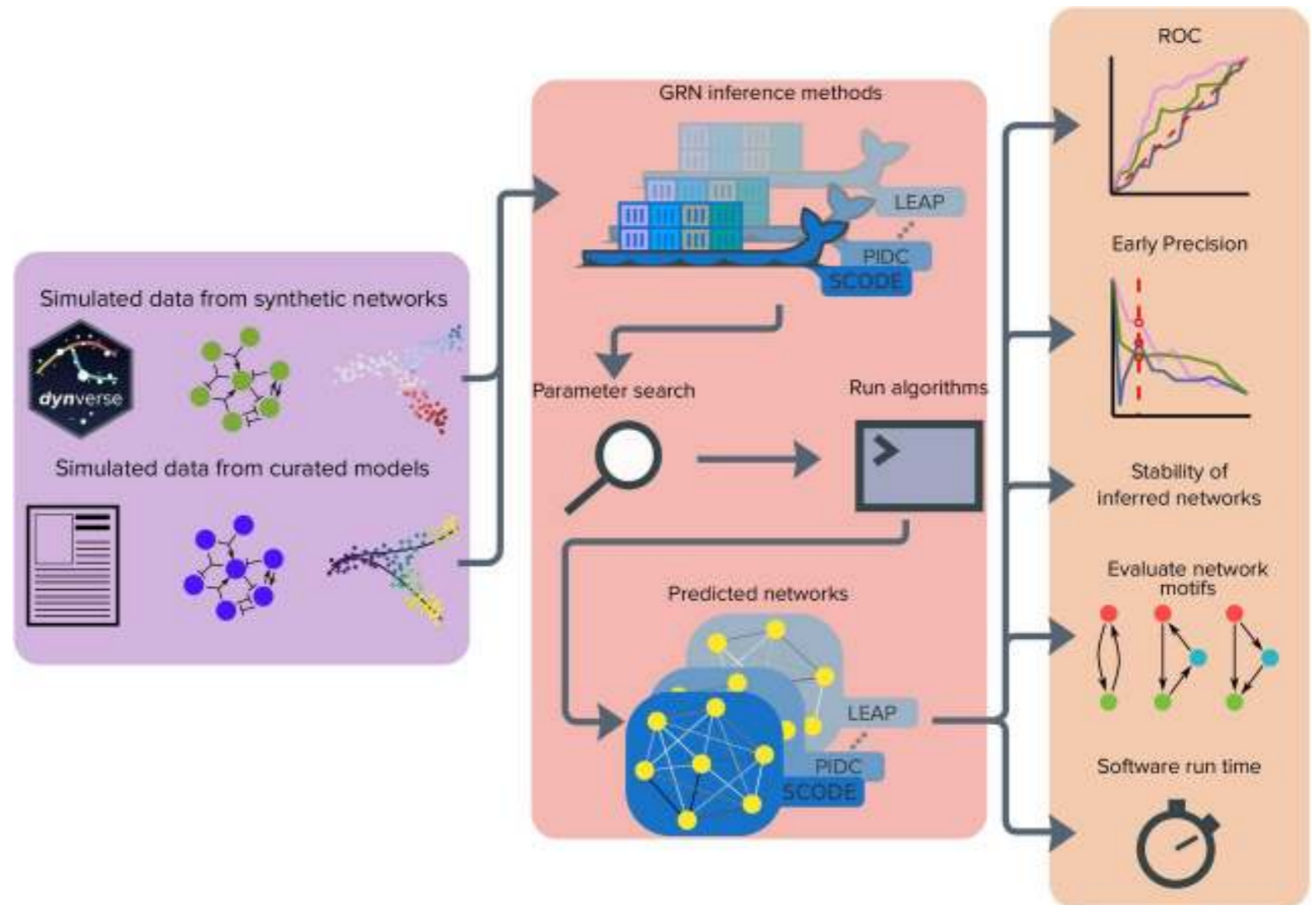
# BENCHMARKING NETWORK INFERENCE

- One data simulation strategy plus two real data
- In real data ground-truth: STRING
- Few genes considered 10-100



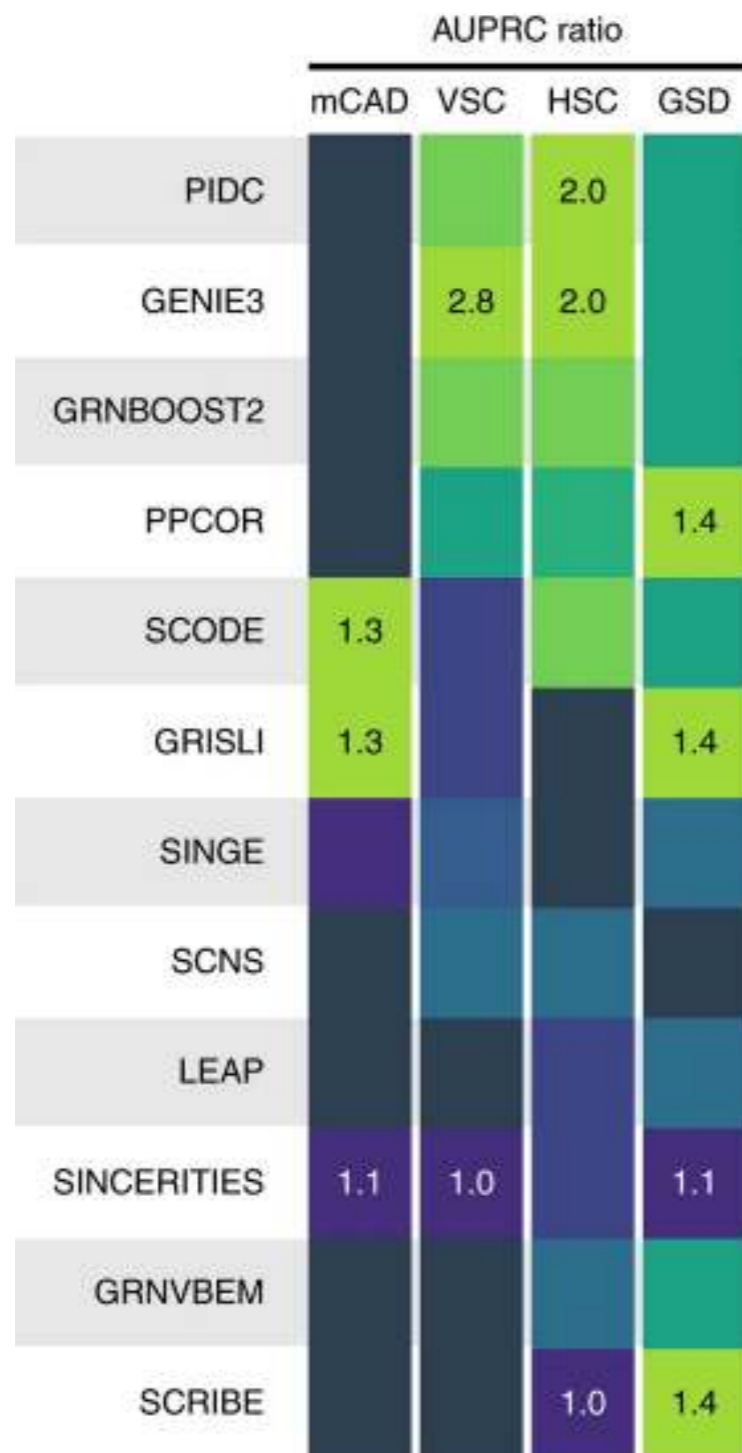
# BENCHMARKING NETWORK INFERENCE

- Two data simulation strategies plus real data
- In real data ground-truth: STRING or ChIP-seq
- Few genes considered 100-2000

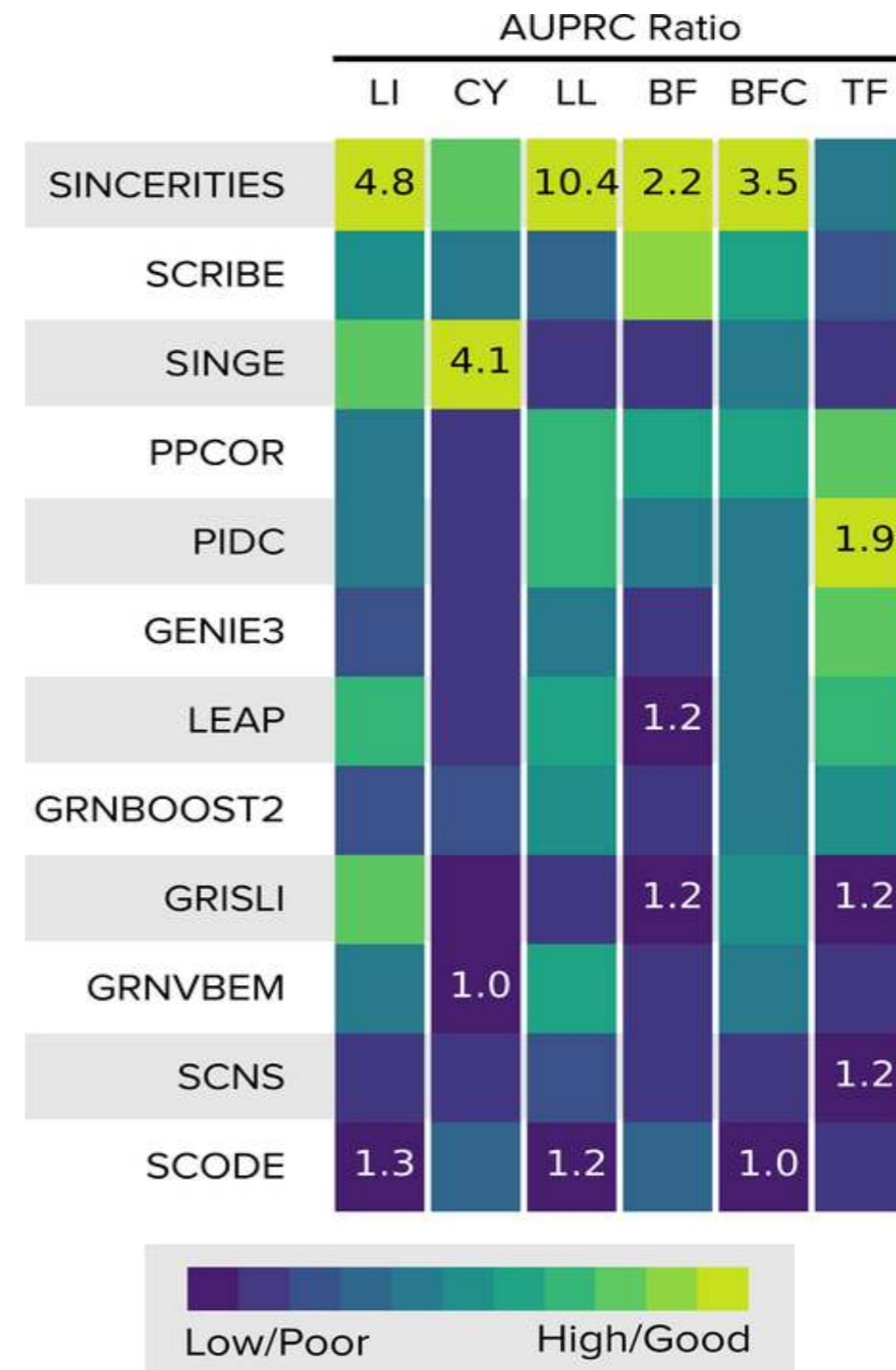


# POOR PERFORMANCES IN SIMULATED DATA

BoolODE simulated data



Curated models data



# PERFORMANCES ON REAL DATA ARE DIFFICULT TO EVALUATE

	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	
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	
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



**AIM: BENCHMARKING  
ON REAL SCRNA DATA  
WITH MANY GENES**

# BENCHMARKING NETWORK INFERENCE REPRODUCIBILITY

---

Using the reproducibility between biological replicates to evaluate the quality of network inference methods in real scRNA-seq data

Independent scRNA-seq datasets



*Retina dataset 1*



*Retina dataset 2*

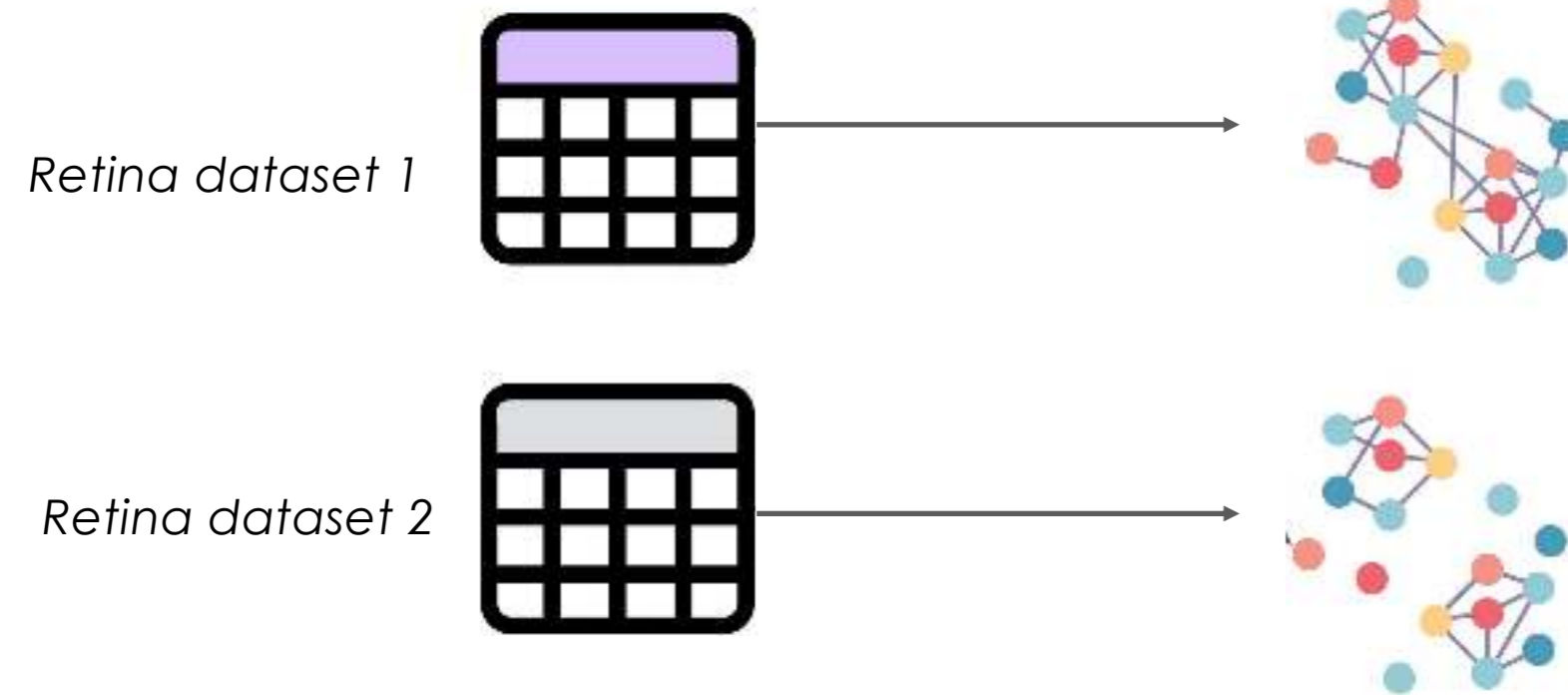
# BENCHMARKING NETWORK INFERENCE REPRODUCIBILITY

---

Using the reproducibility between biological replicates to evaluate the quality of network inference methods in real scRNA-seq data

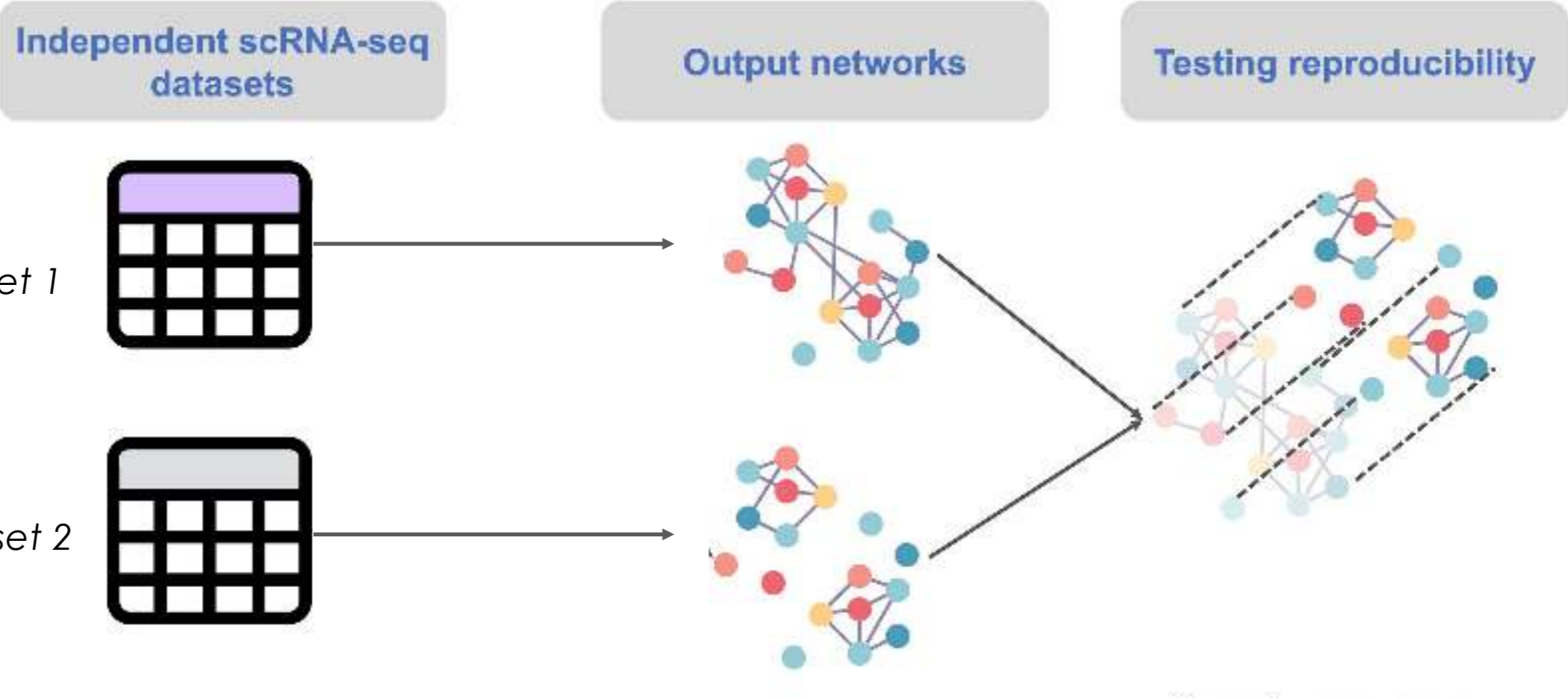
Independent scRNA-seq datasets

Output networks



# BENCHMARKING NETWORK INFERENCE REPRODUCIBILITY

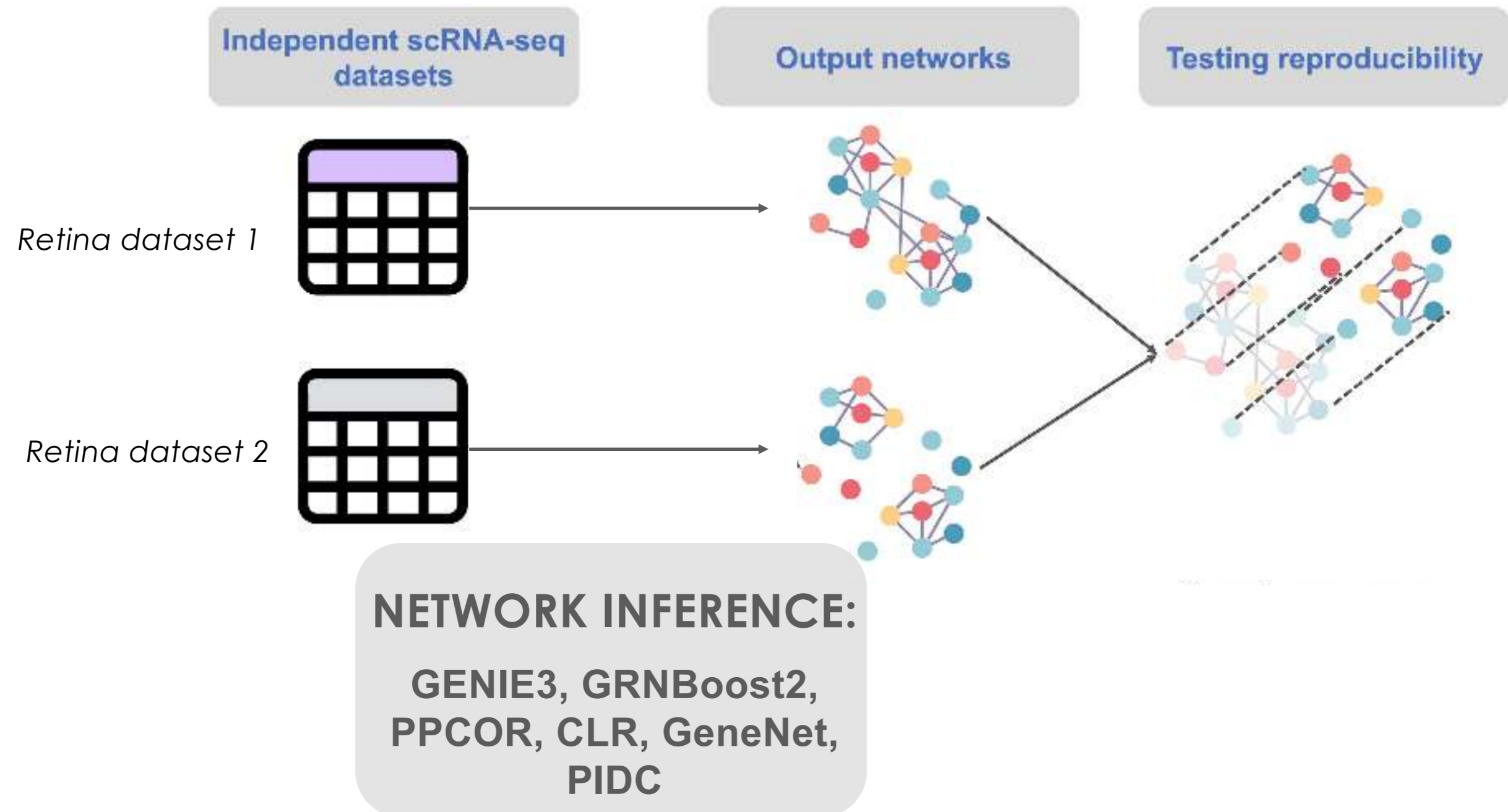
Using the reproducibility between biological replicates to evaluate the quality of network inference methods in real scRNA-seq data



# BENCHMARKING NETWORK INFERENCE REPRODUCIBILITY

---

Using the reproducibility between biological replicates to evaluate the quality of network inference methods in real scRNA-seq data



# BENCHMARKING NETWORK INFERENCE REPRODUCIBILITY

Using the reproducibility between biological replicates to evaluate the quality of network inference methods in real scRNA-seq data



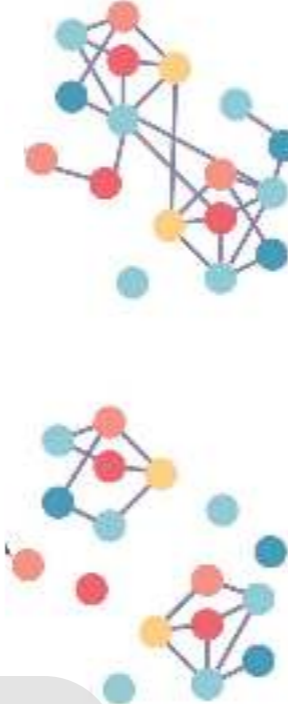
Retina dataset 1



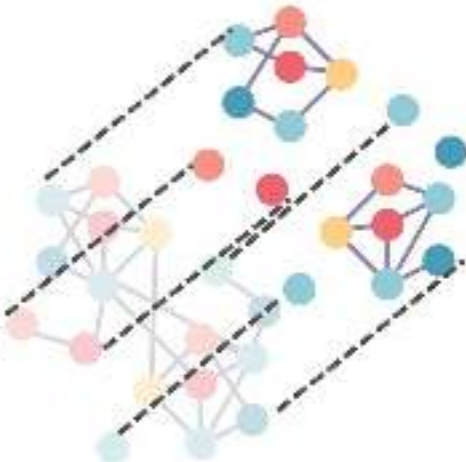
Retina dataset 2



Output networks



Testing reproducibility



**NETWORK INFERENCE:**

**GENIE3, GRNBoost2,  
PPCOR, CLR, GeneNet,  
PIDC**

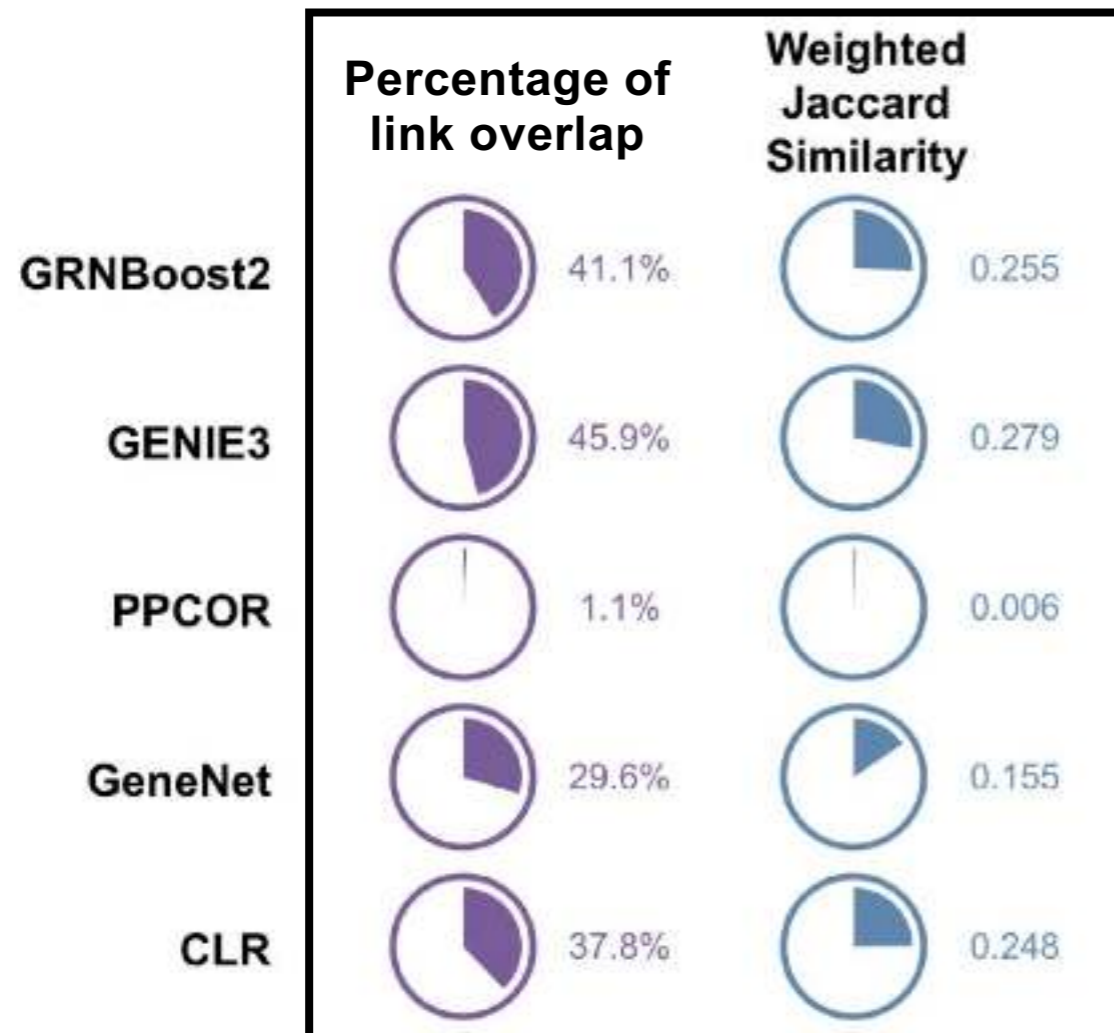
**Data considered:**

retina, colorectal cancer,  
hematopoiesis (HSC, CLP,  
Erythrocytes, Monocytes,  
Dendritic cells)

# GENIE3 AND GRNBOOST2 PROVED BEST PERFORMANCES IN RETINA

---

## Reproducibility scores



# GENIE3 AND GRNBOOST2 PROVED BEST PERFORMANCES IN RETINA

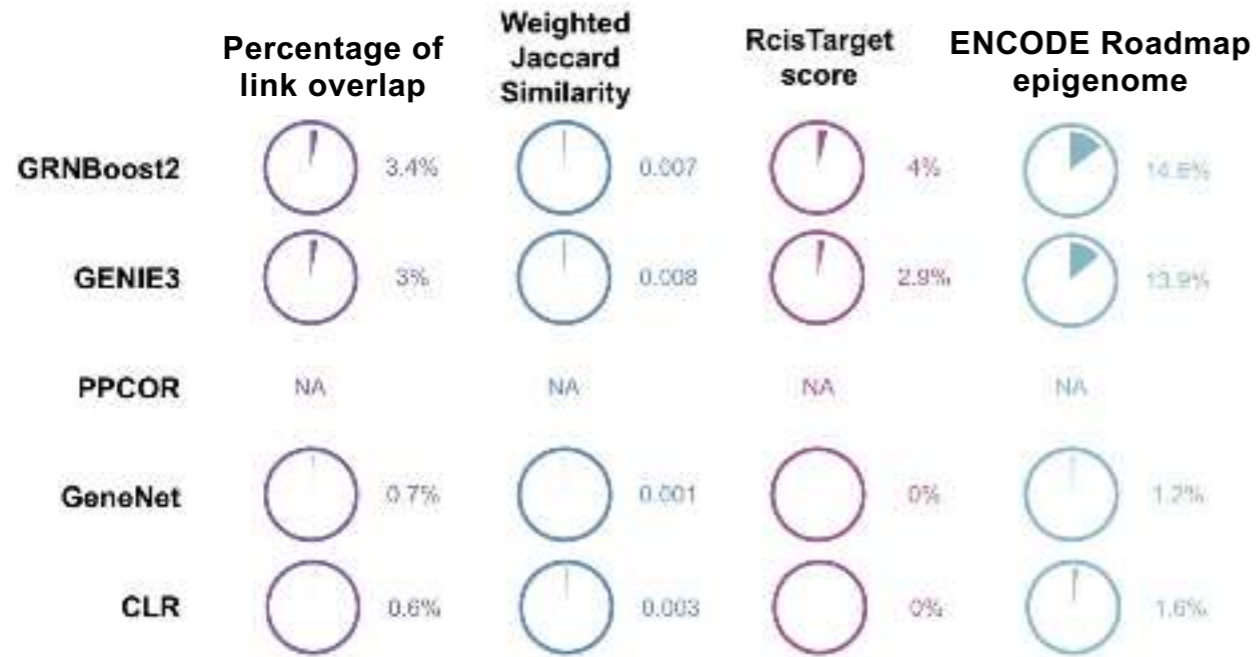
## Intersection with known interactions





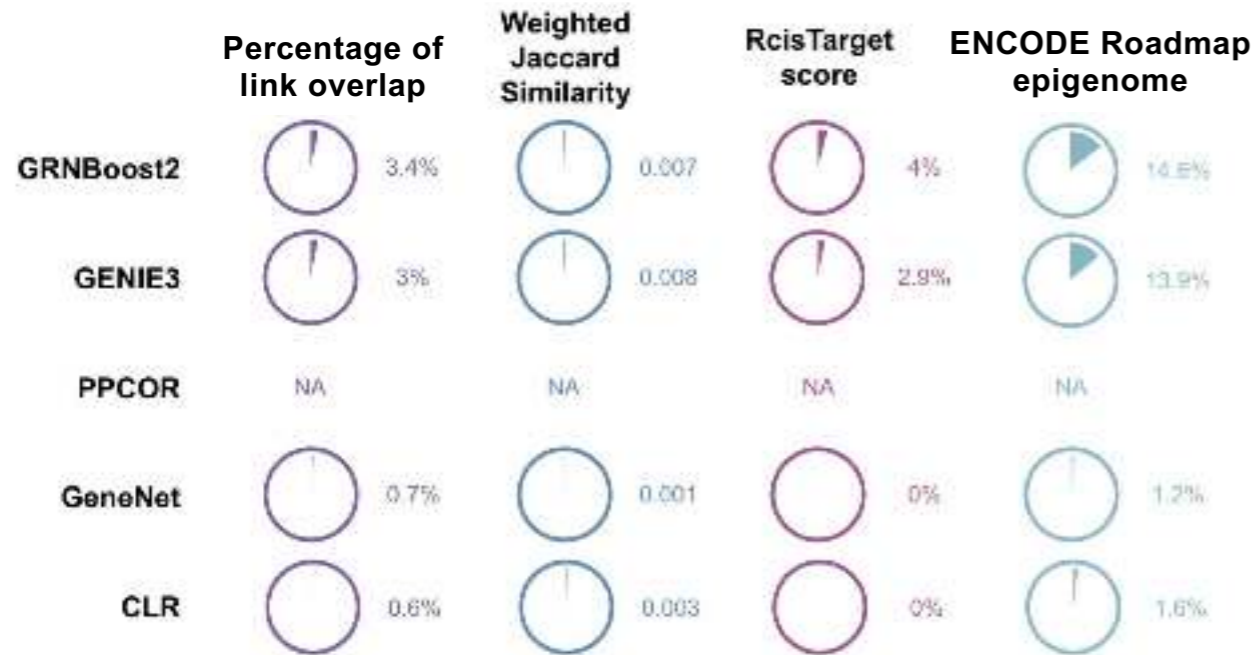
# GENIE3 AND GRNBOOST2 PROVED BEST PERFORMANCES IN CRC

## Colorectal cancer (CRC)

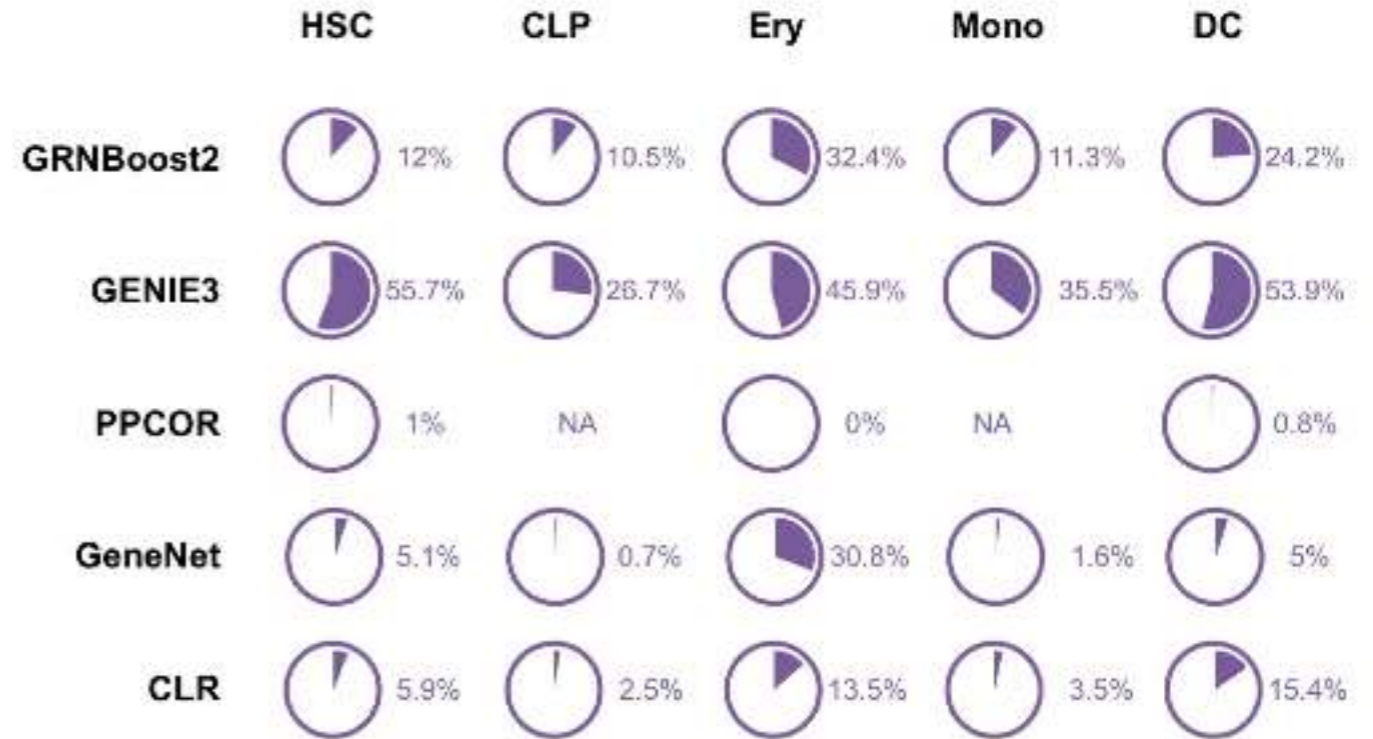


# GENIE3 PROVED BEST PERFORMANCES IN HEMATOPOIESIS

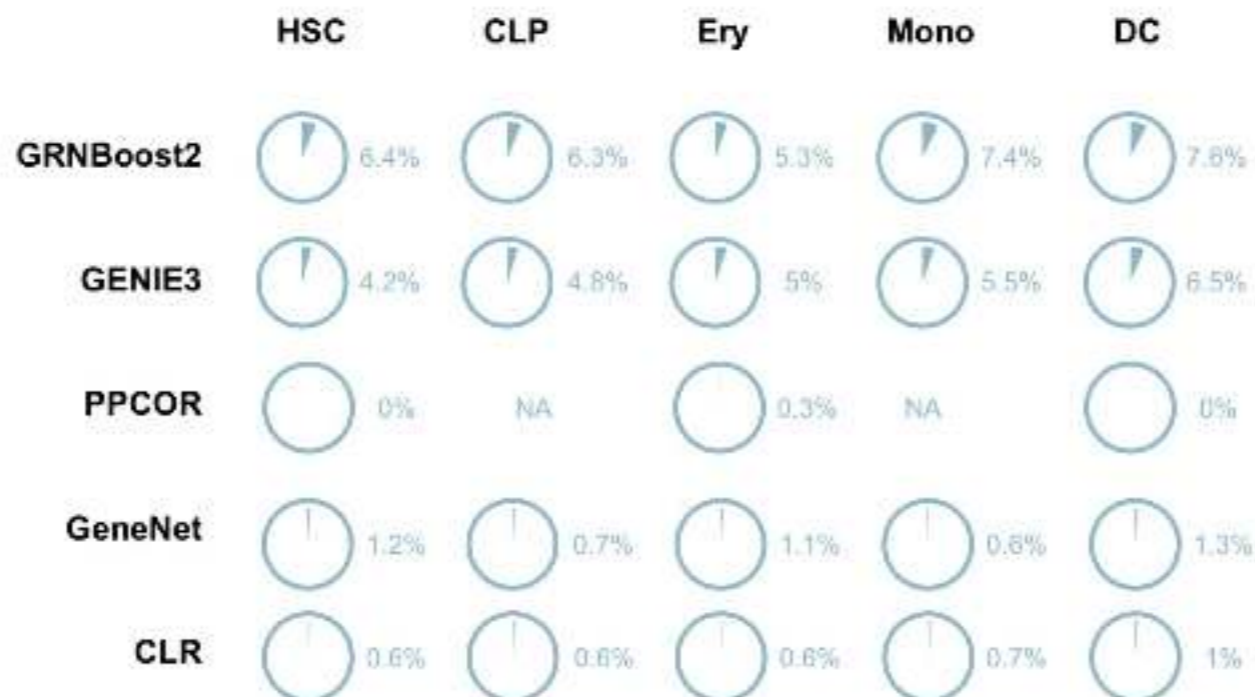
## Colorectal cancer (CRC)



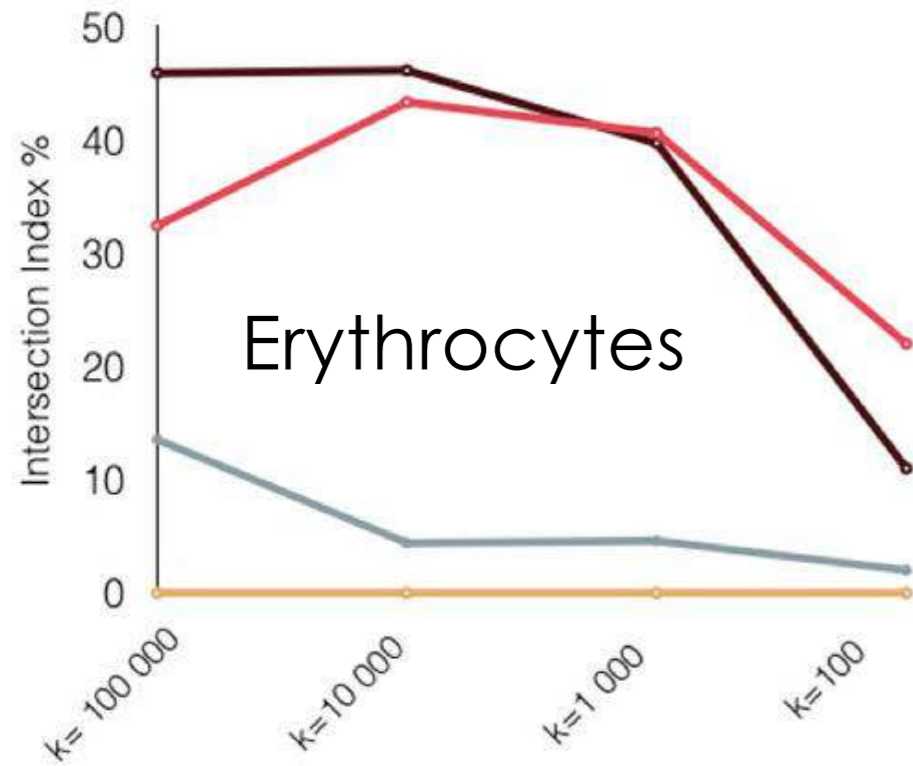
## Hematopoiesis reproducibility score



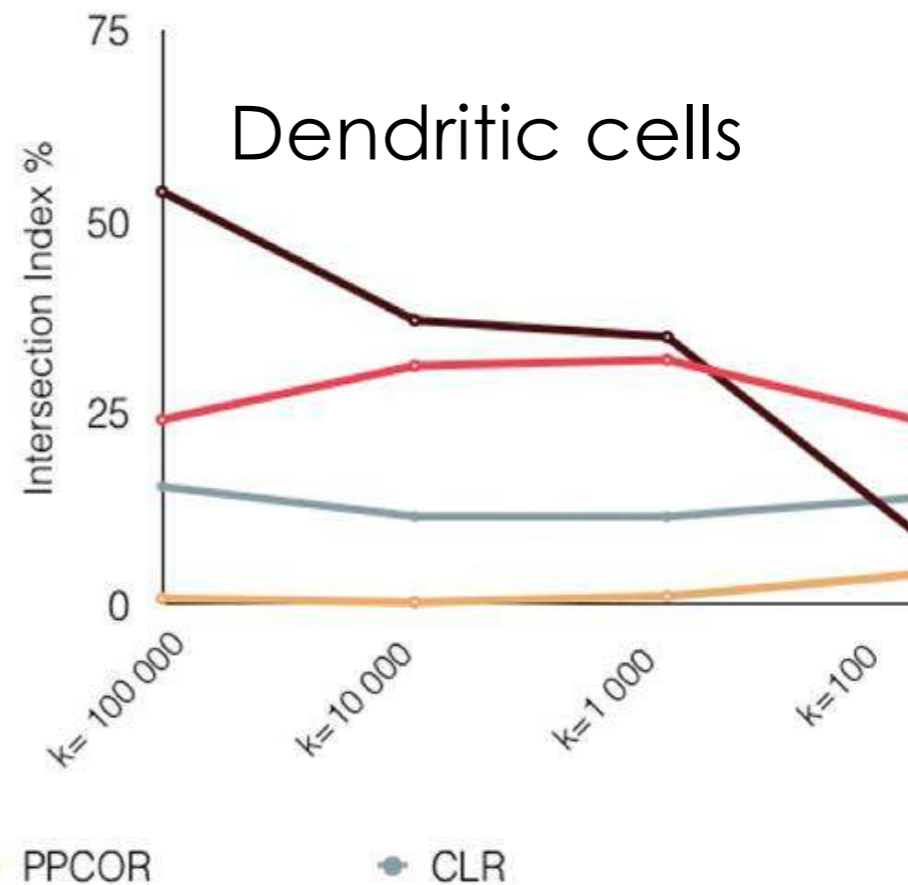
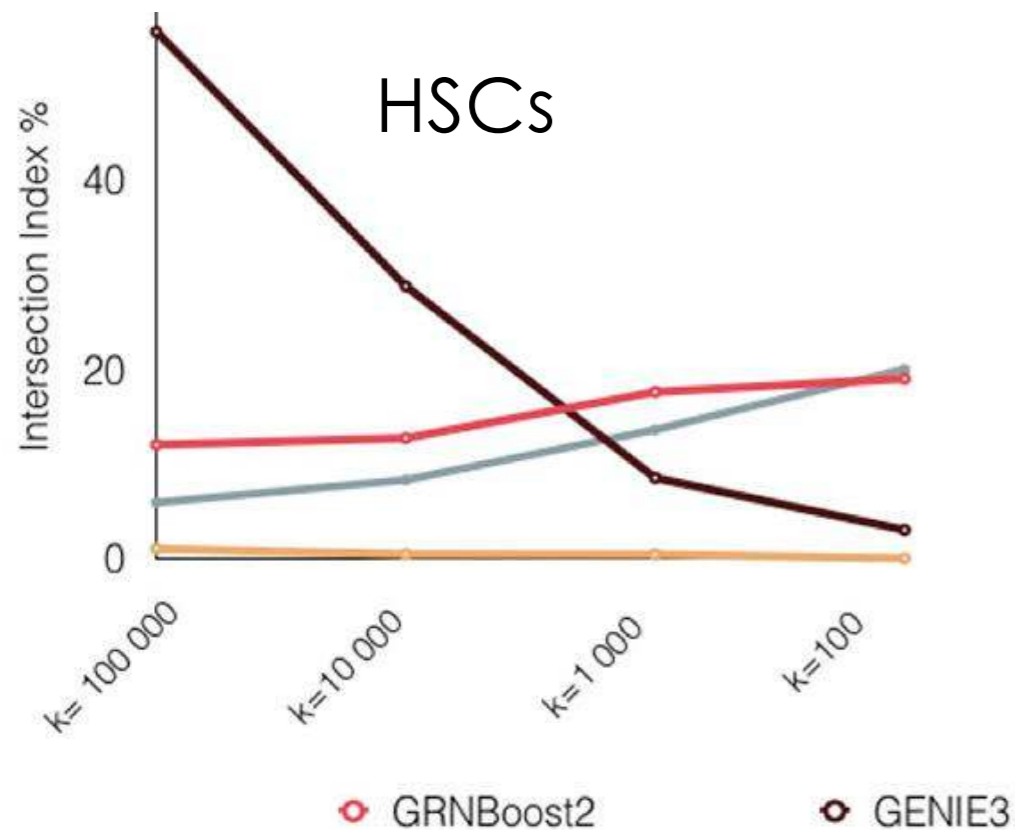
## Hematopoiesis intersection score



# THE LINK THRESHOLD DOES NOT AFFECT REPRODUCIBILITY



- Previous results obtained for k=100,000
- Changing threshold does not alter much results
- GRNBoost performs better at 1000/100 links, but low density network



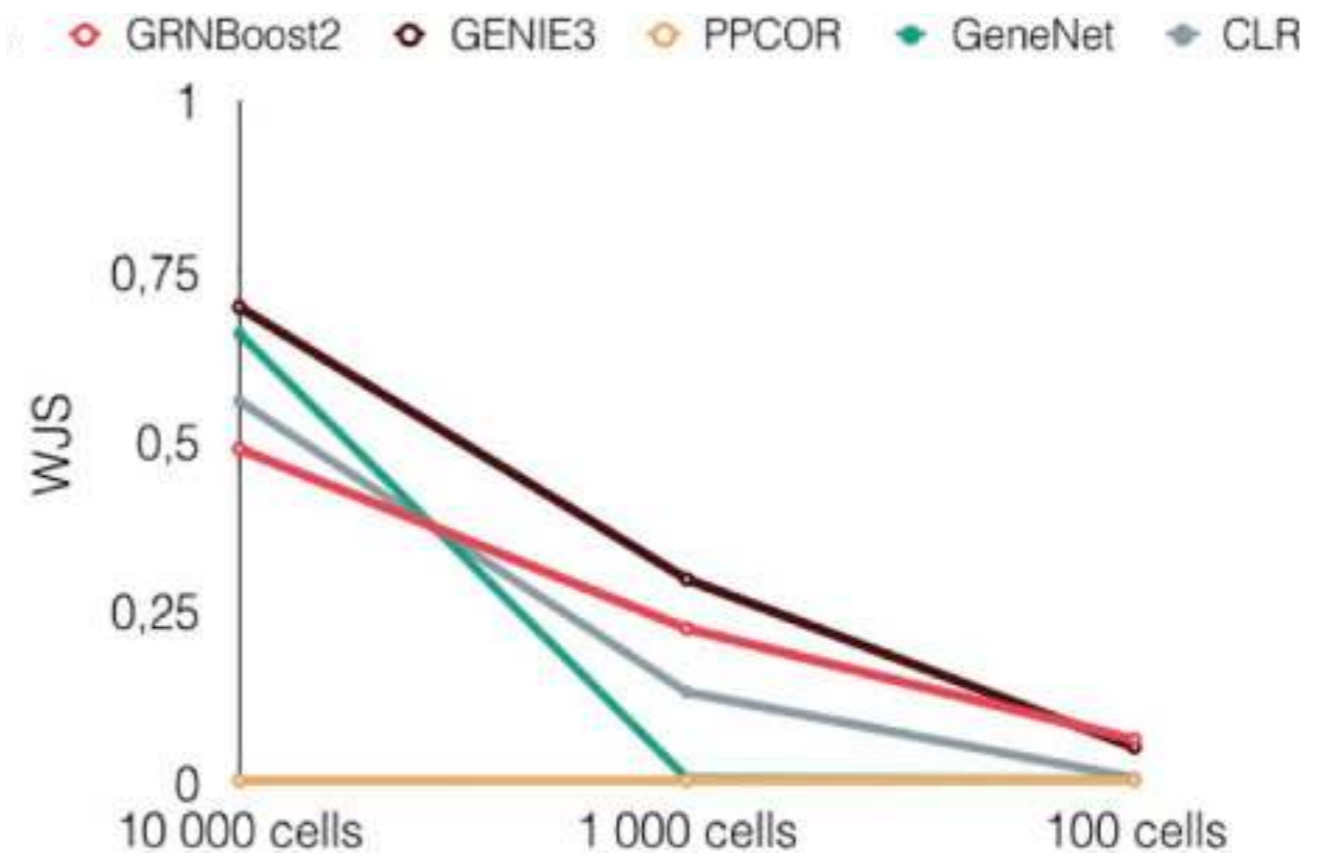
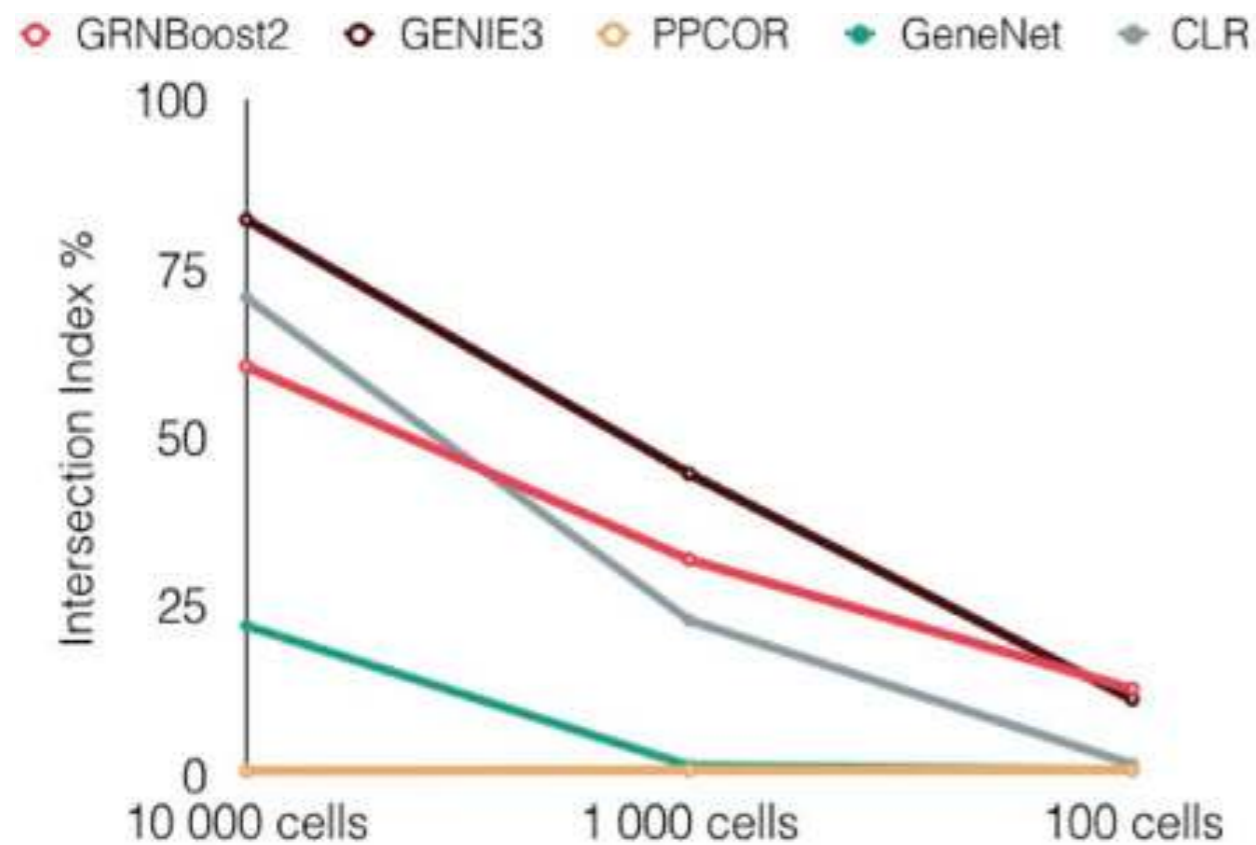
# CELLS AND SEQUENCING PLATFORM DO NOT AFFECT RESULTS

---

To test impact of sequencing platform and number of cells, we subset the retina dataset

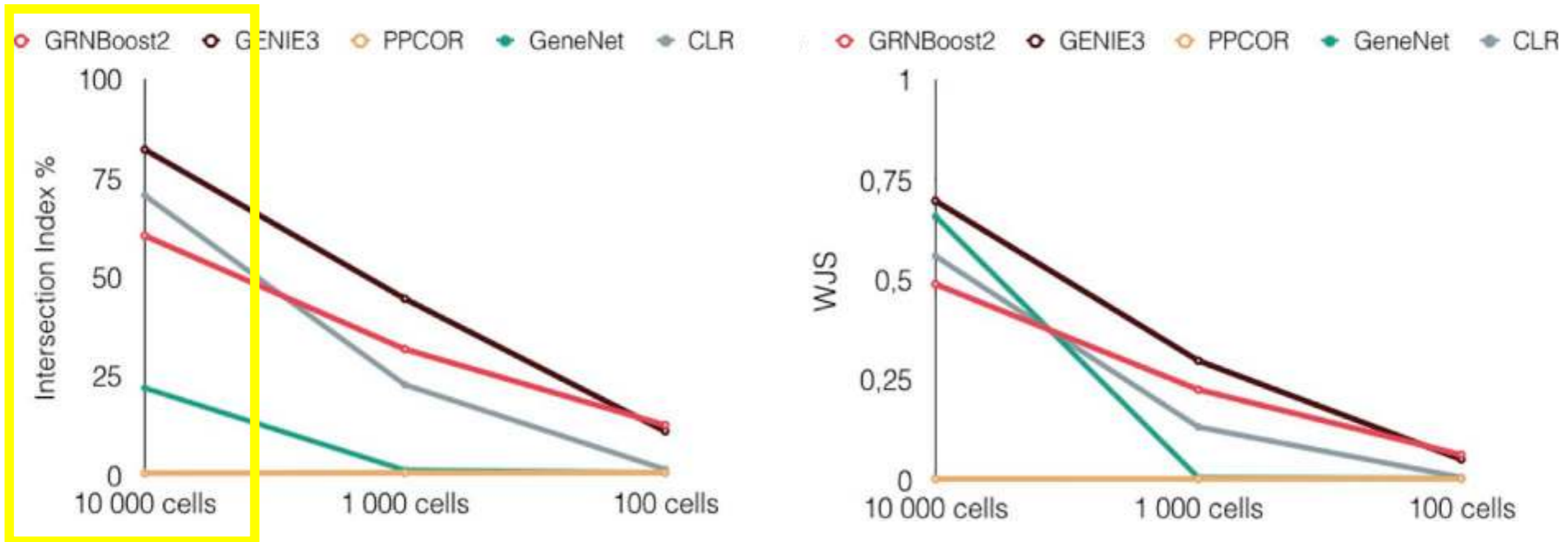
# CELLS AND SEQUENCING PLATFORM DO NOT AFFECT RESULTS

To test impact of sequencing platform and number of cells, we subset the retina dataset



# CELLS AND SEQUENCING PLATFORM DO NOT AFFECT RESULTS

To test impact of sequencing platform and number of cells, we subset the retina dataset



# SCNET JUPYTER NOTEBOOK TO REPRODUCE RESULTS

ComputationalSystemsBiology / scNET Public

<> Code Issues 2 Pull requests 1 Actions Projects Wiki Security Insights Settings

master 1 branch 0 tags

Go to file

Add file

Code

ykang-m updated scNET environment

3b23007 on Sep 15 18 commits

Algorithm_Comparison.ipynb	notebook edited for additional algorithms	3 months ago
Functions.R	Updated for CLR and GeneNet comparison	3 months ago
README.md	Update alg notebook	11 months ago
scNET.yml	updated scNET environment	3 months ago

README.md

## Evaluating the reproducibility of single-cell gene regulatory network inference algorithms

We here benchmark three single-cell network inference algorithms based on their reproducibility, i.e. their ability to infer similar networks once applied to two independent datasets from the same biological condition.

The benchmarked methods are:

# CONCLUSIONS

---

- Benchmarking of scRNA-seq networks inference based on reproducibility
- GENIE3 shows better performances
- Our results agree with previous benchmarks



# CONCLUSIONS

---

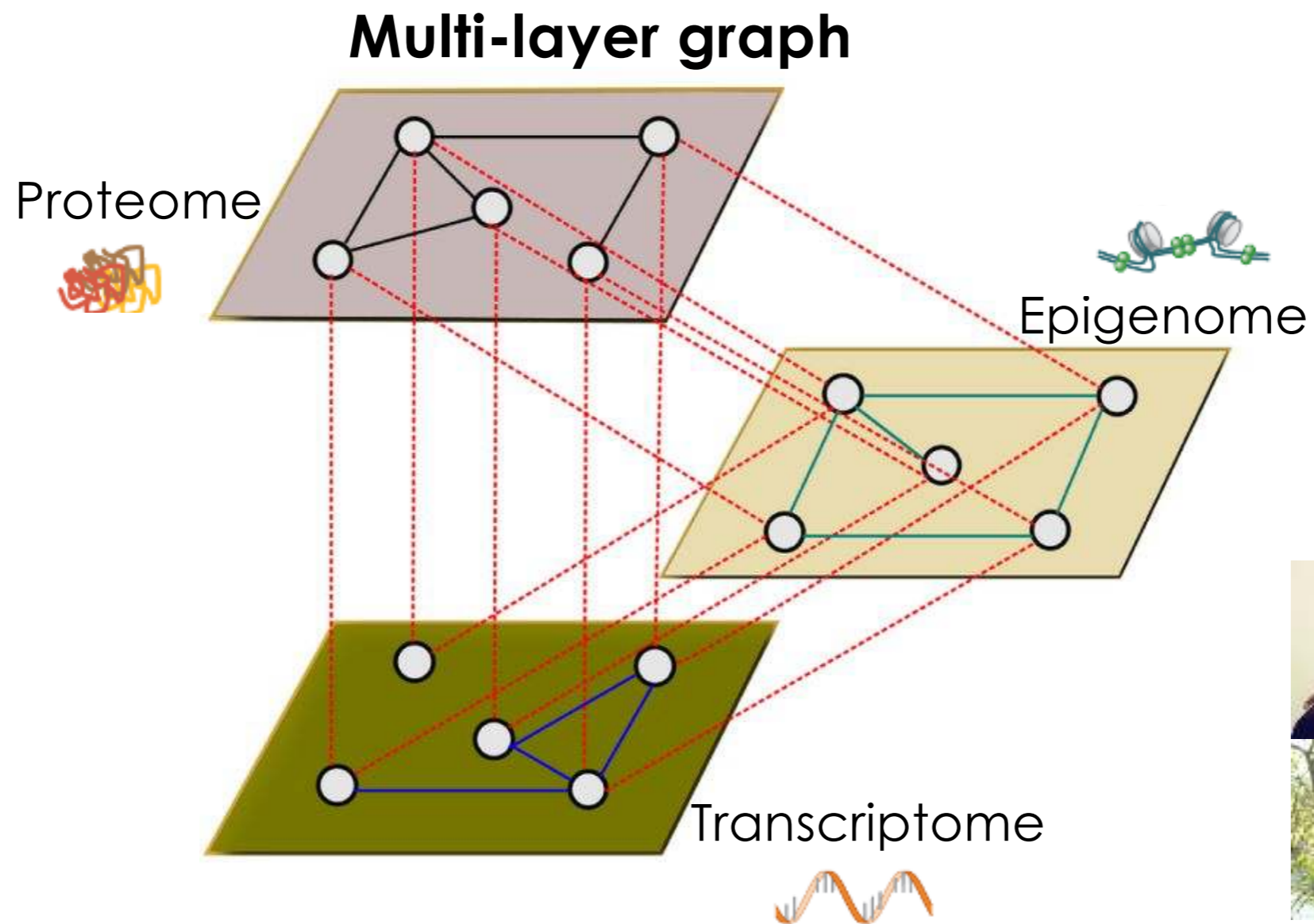
- Benchmarking of scRNA-seq networks inference based on reproducibility
- GENIE3 shows better performances
- Our results agree with previous benchmarks

## **Pay attention!**

- Combining benchmarks is fundamental
- Methods based on pseudotime are missing

# PERSPECTIVES: SINGLE-CELL MULTI-LAYER GRAPHS

Combining multi-modal data to improve quality of single-cell graphs



Ina Maria  
Deutschmann

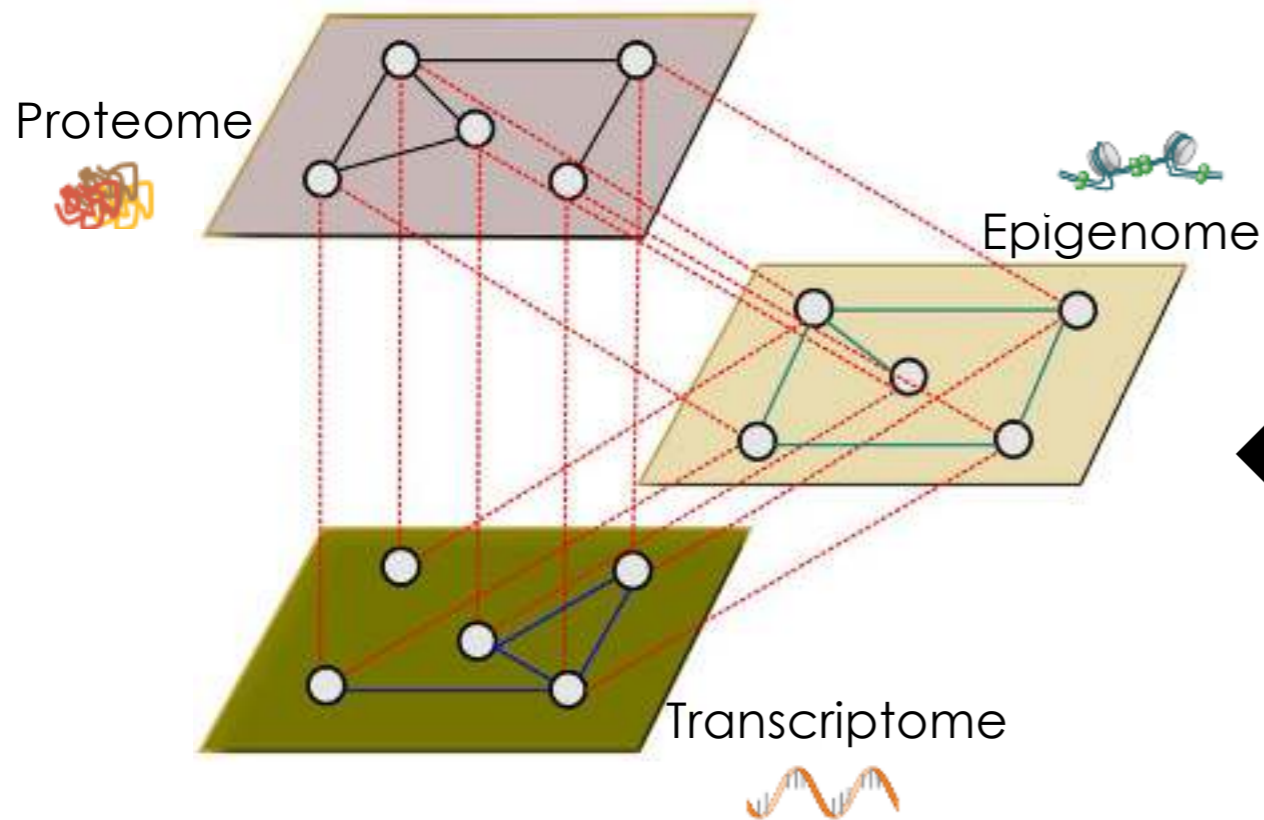


Remi  
Trimbour

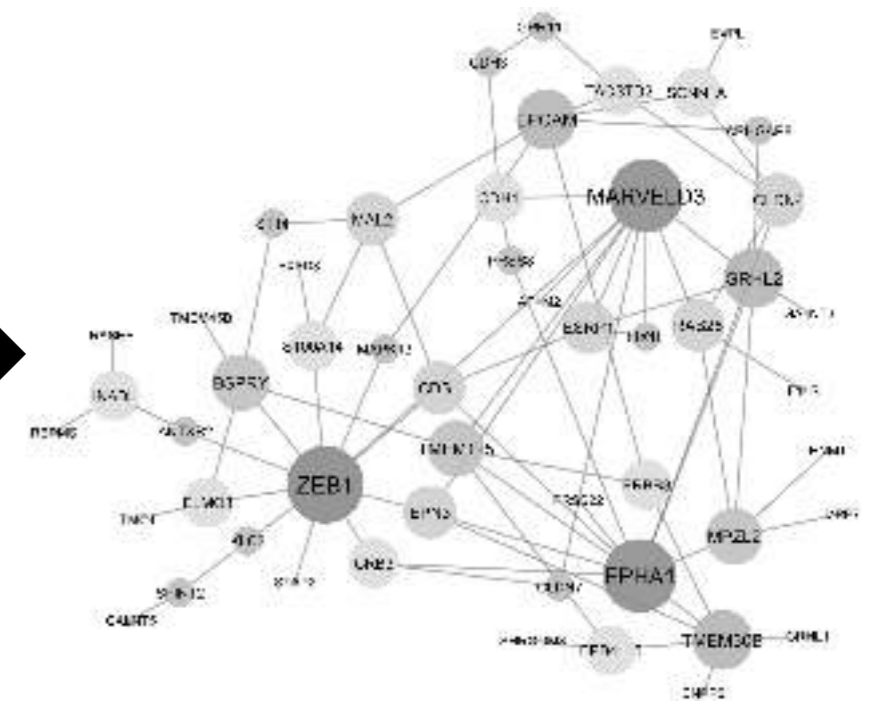
# PERSPECTIVES: SINGLE-CELL MULTI-LAYER GRAPHS

Combining multi-modal data to improve quality of single-cell graphs

## Multi-layer graph



**Currently used approach:**  
transcriptome alone



# ACKNOWLEDGEMENTS

---

## CSB Team



Denis Thieffry, Morgane Thomas-Chollier, Marika Kapsimali, Nathalie Lehmann, **Yoonjee Kang**, Geert-Jan Huizing, Claire Lansonneur, Ina Maria Deutschmann, Jules Samaran, Remi Trimbour

## Funding



## Twitter



**Associated publication:** Kang Y, Thieffry D, **Cantini L**. Frontiers in genetics. 2021  
<https://github.com/ComputationalSystemsBiology/scNET>

# EXISTING METHODS SHOW POOR PERFORMANCES

