

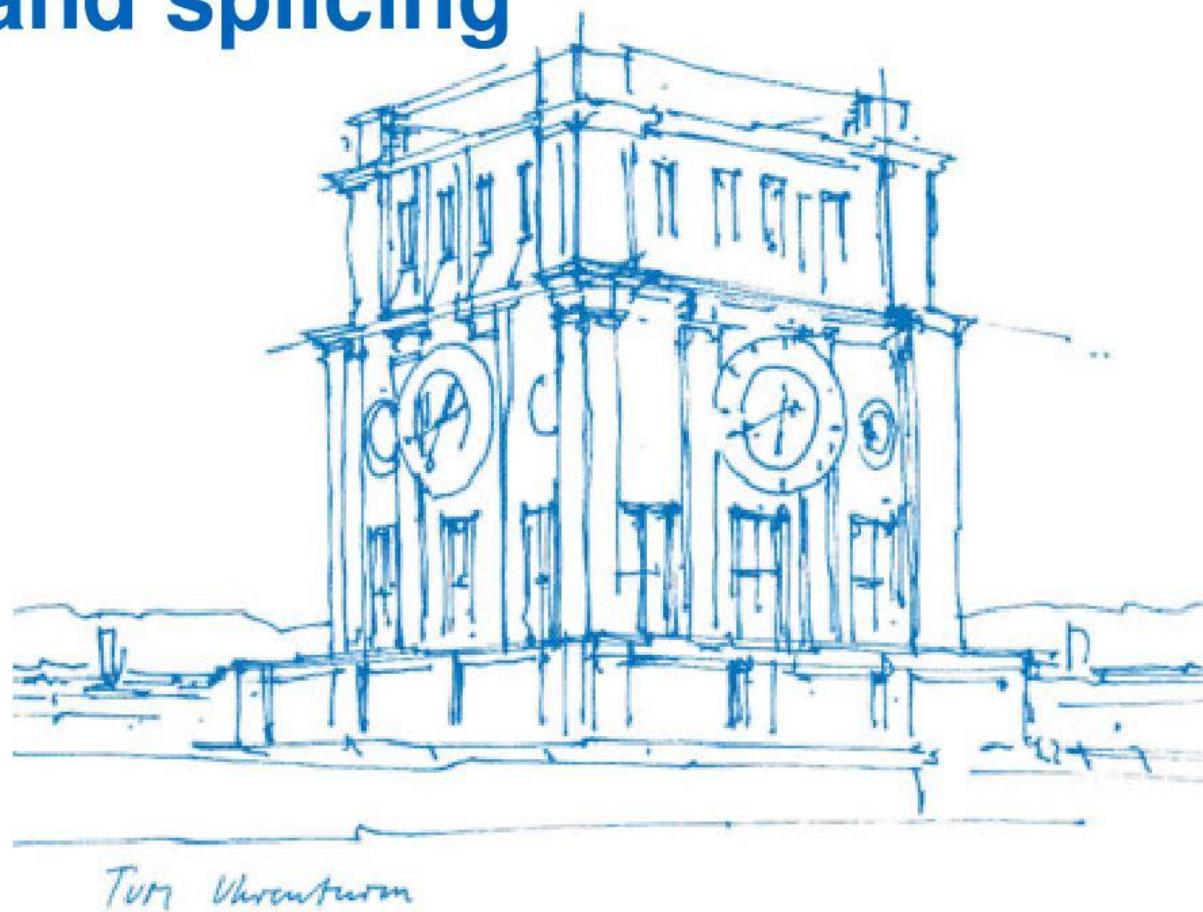
A multi-modal model of RNA velocity describing nuclear export and splicing kinetics

Robin Mittas

Chair for mathematical modelling of biological systems
(Prof. Dr. Theis)

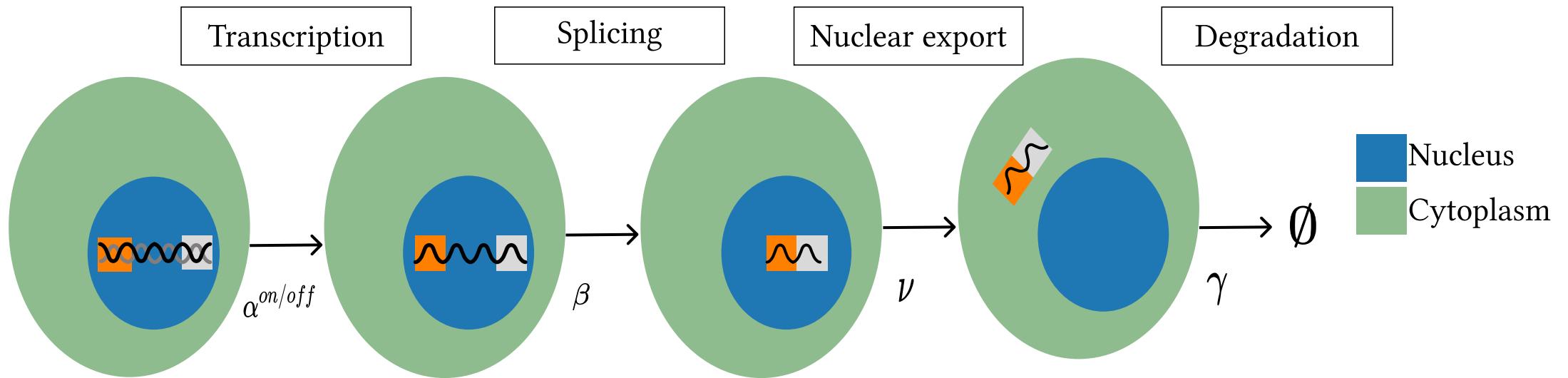
Supervisors: Philipp Weiler, Dr. Fabiola Curion
Computation, Information and Technology (MA)
Technical University of Munich

October 24th, 2023

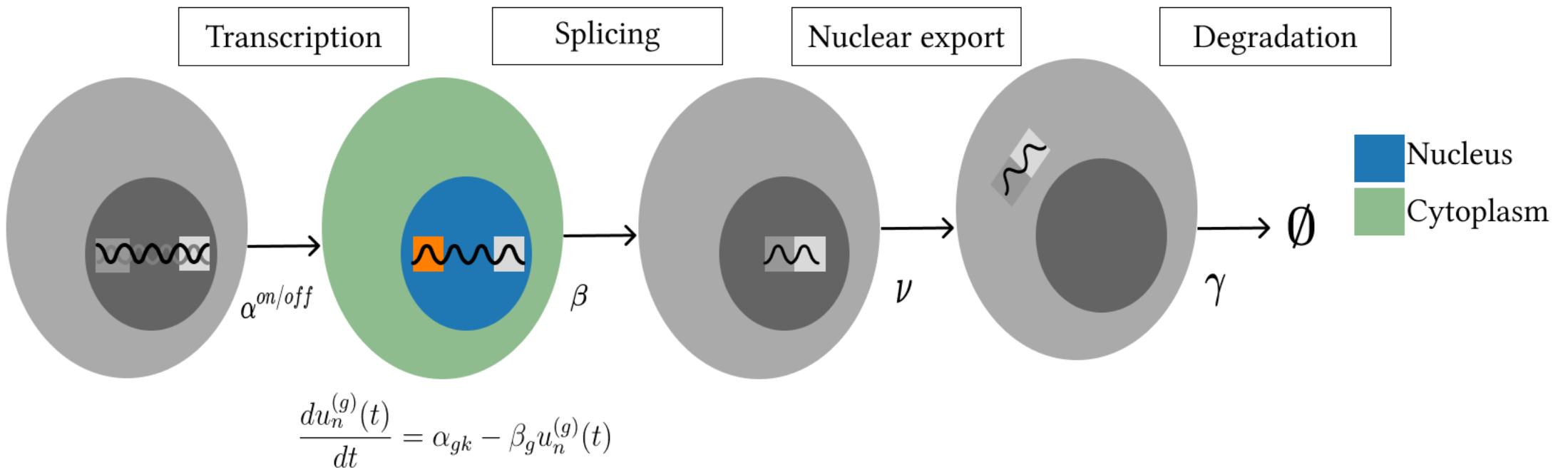


Motivation

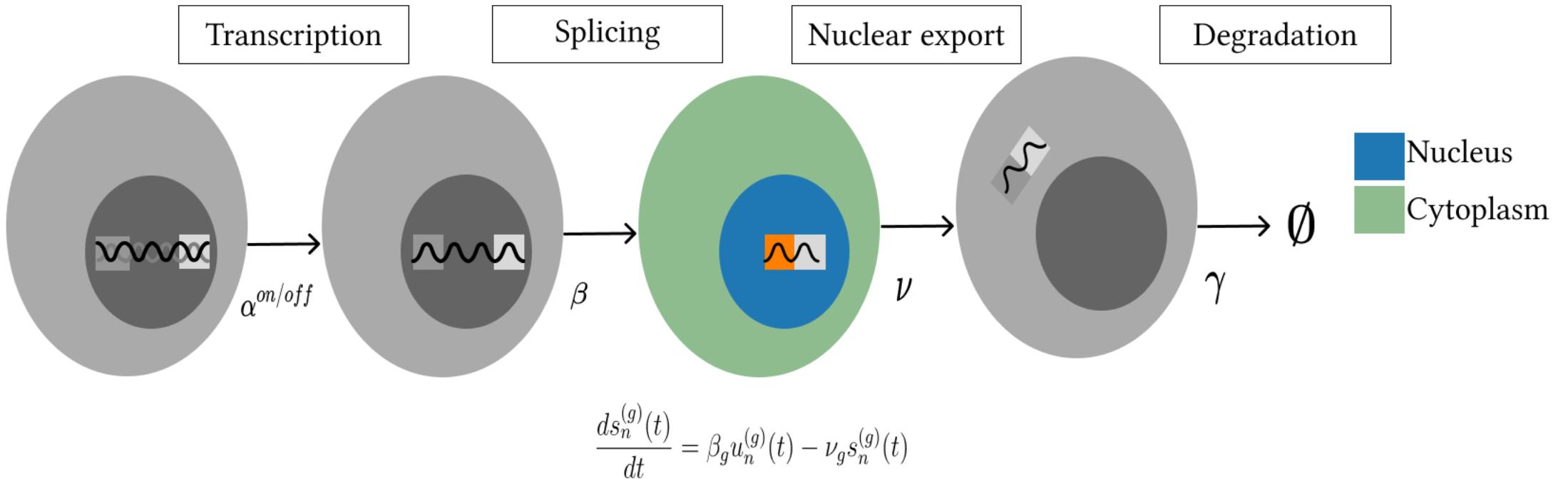
Benchmarking	Model extension
<ul style="list-style-type: none">• Single-cell vs. single-nucleus• Applicability to estimate single-cell velocities based on snRNA-seq measurements	<ul style="list-style-type: none">• Integration of snRNA and scRNA transcripts• Unified system• Biological more accurate description• Estimation of nuclear export rates• Estimation of cytosolic RNA abundance



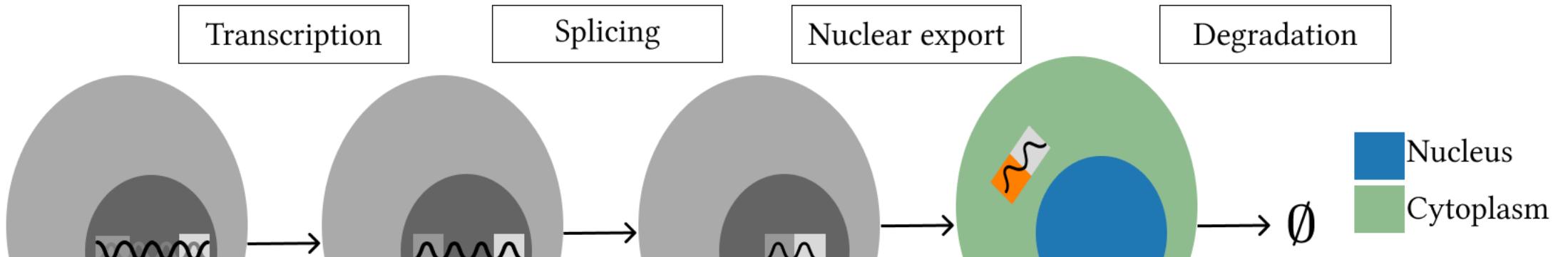
Modelled dynamics



Modelled dynamics

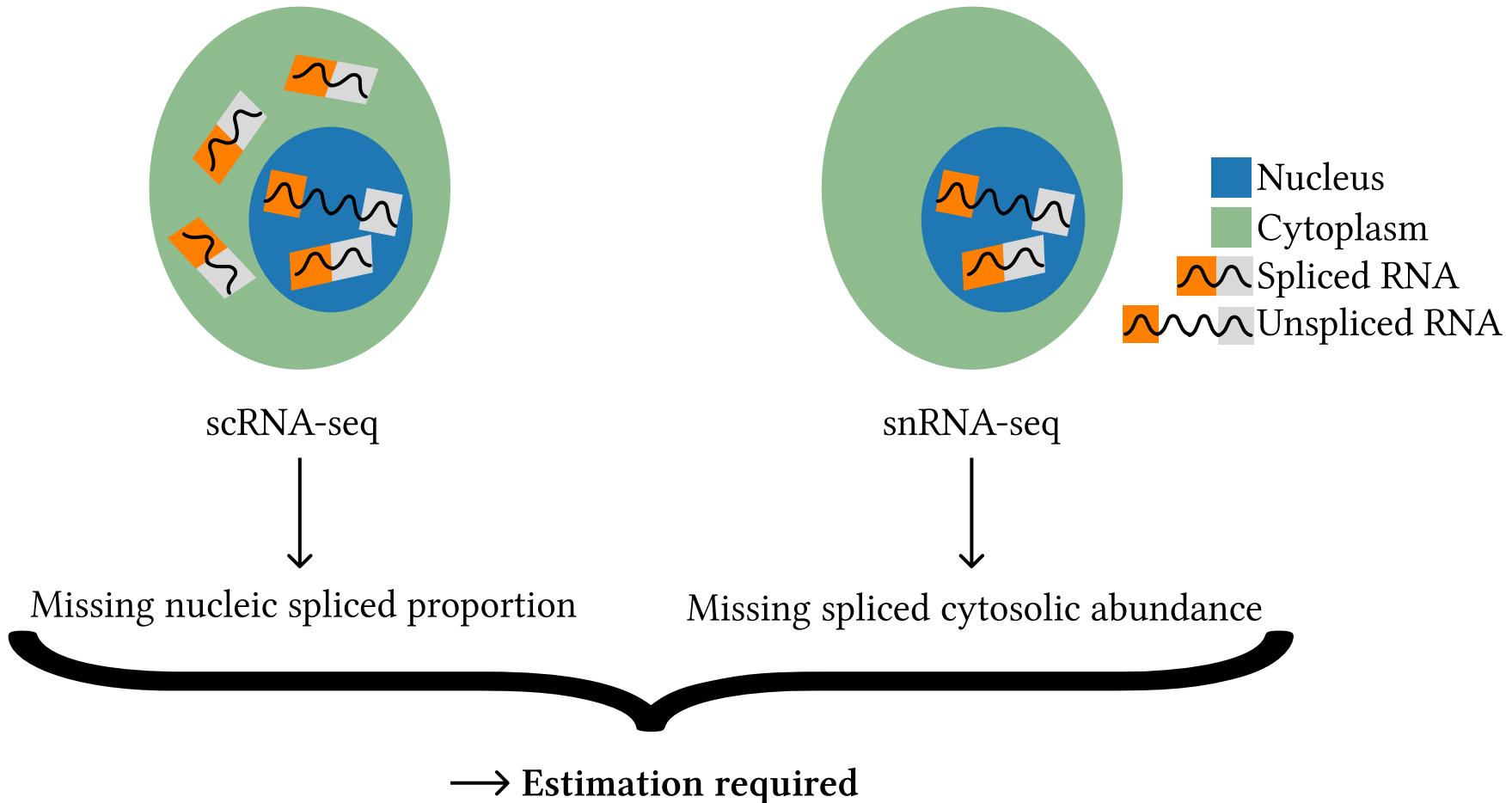


Modelled dynamics



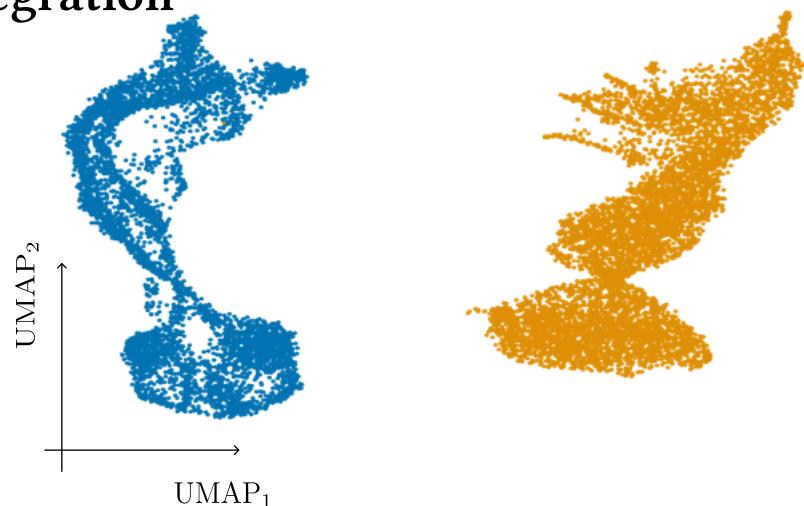
$$\frac{ds_c^{(g)}(t)}{dt} = \nu_g s_n^{(g)}(t) - \gamma_g s_c^{(g)}(t)$$

Methods motivation

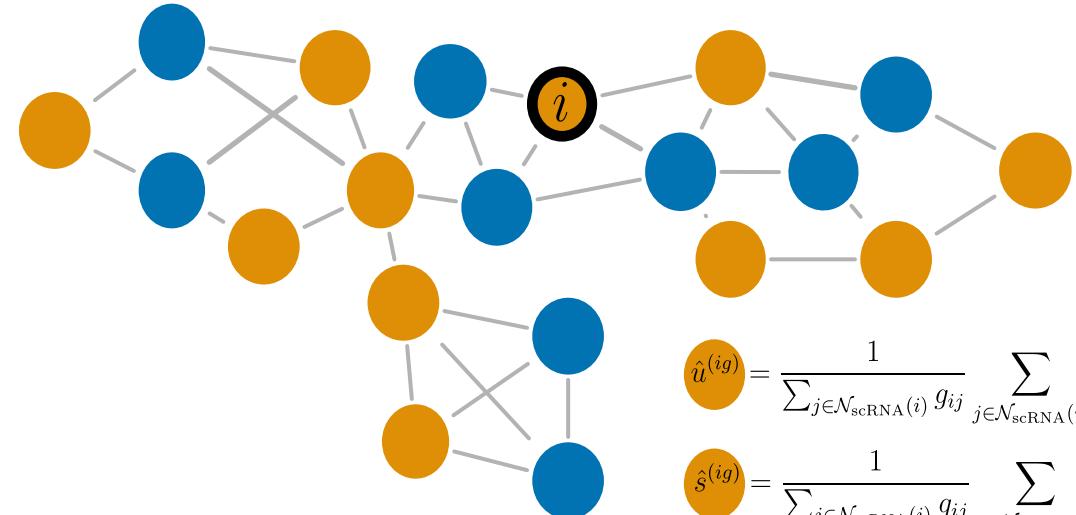


Methods outline

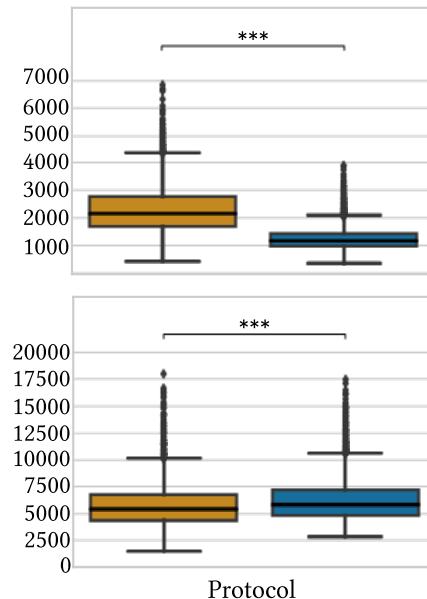
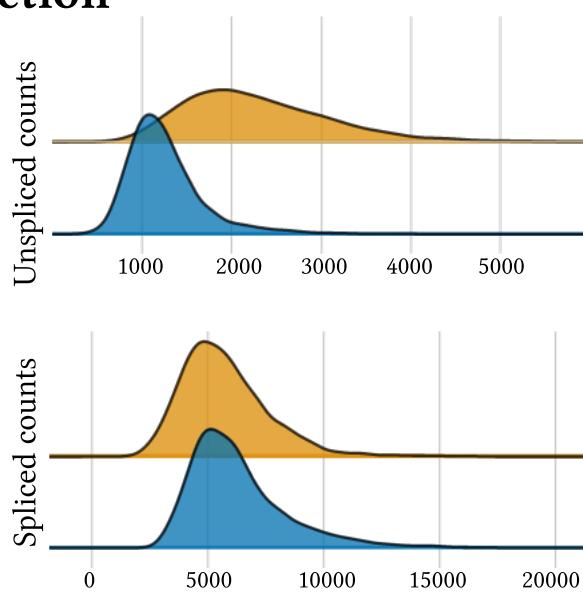
1. Integration



2. Estimation



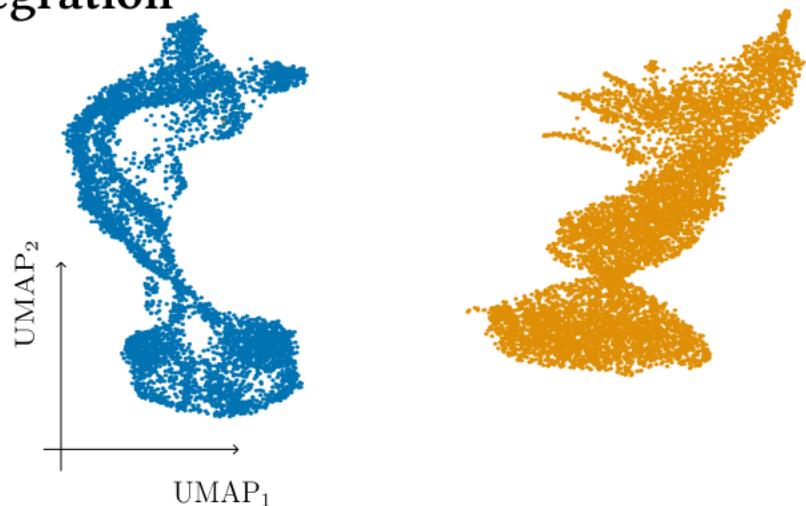
3. Lambda correction



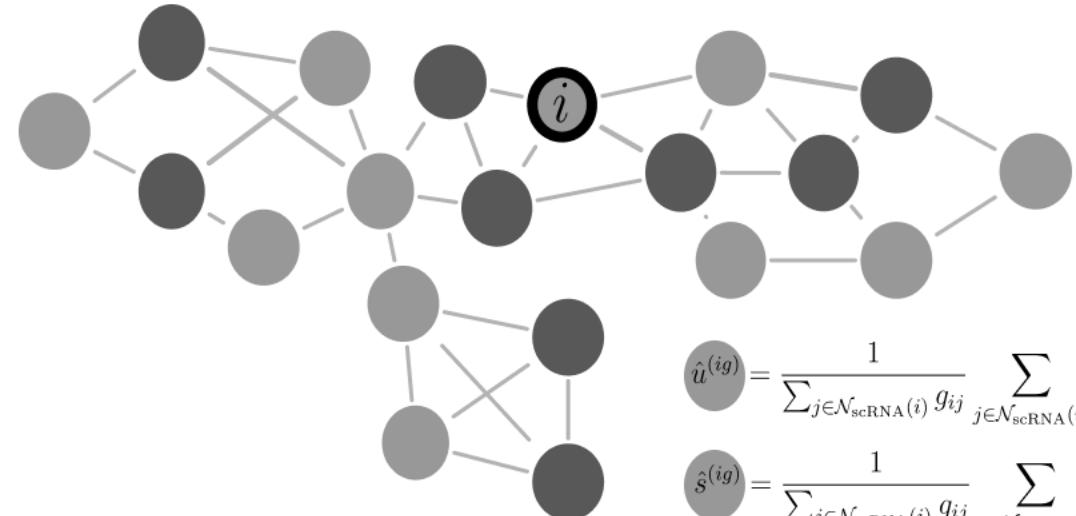
Welchs t -test significance *** $P < 10^{-3}$

Methods outline

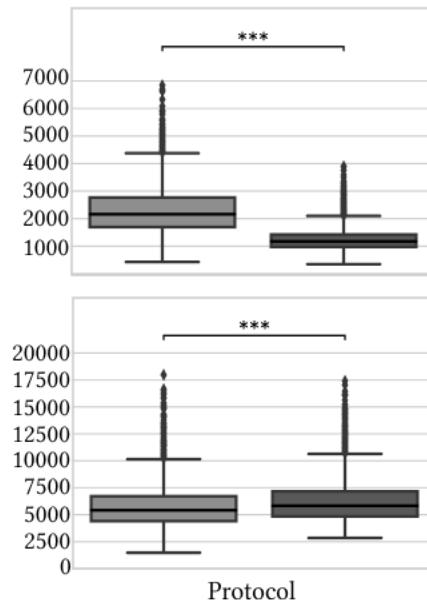
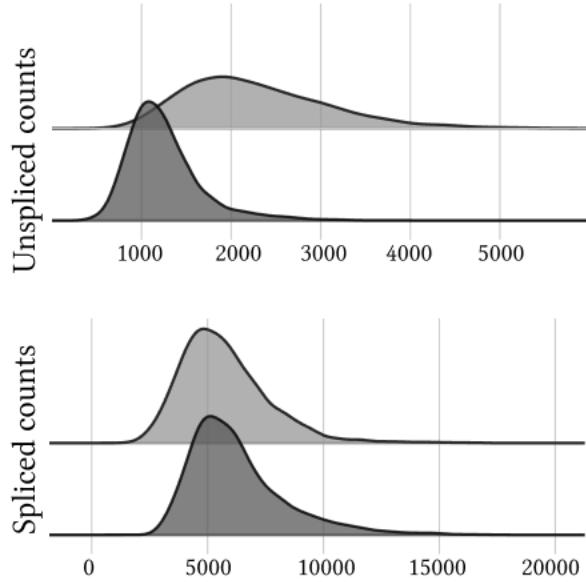
1. Integration



2. Estimation



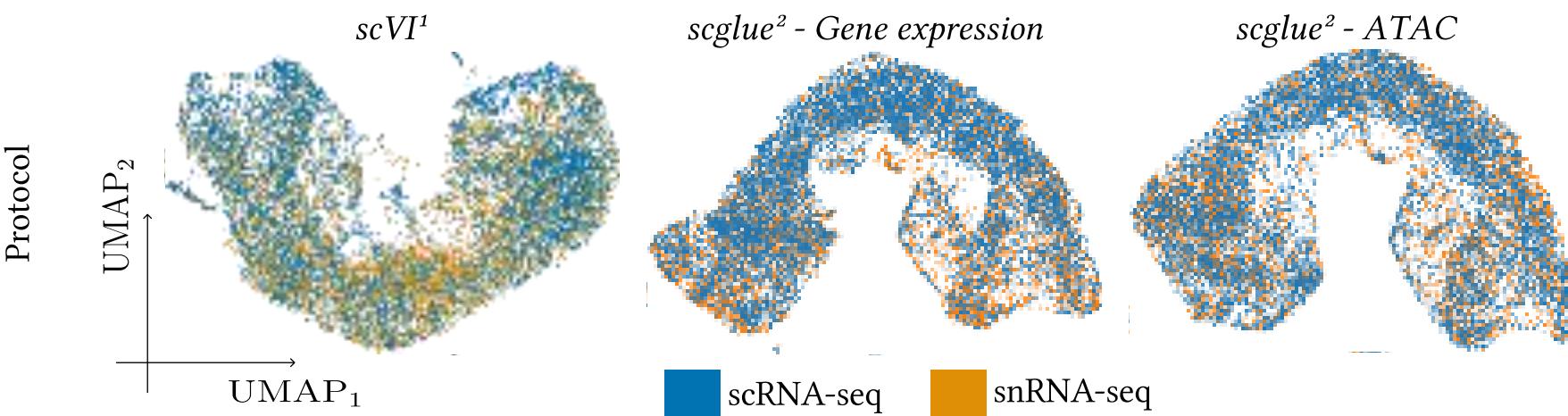
3. Lambda correction



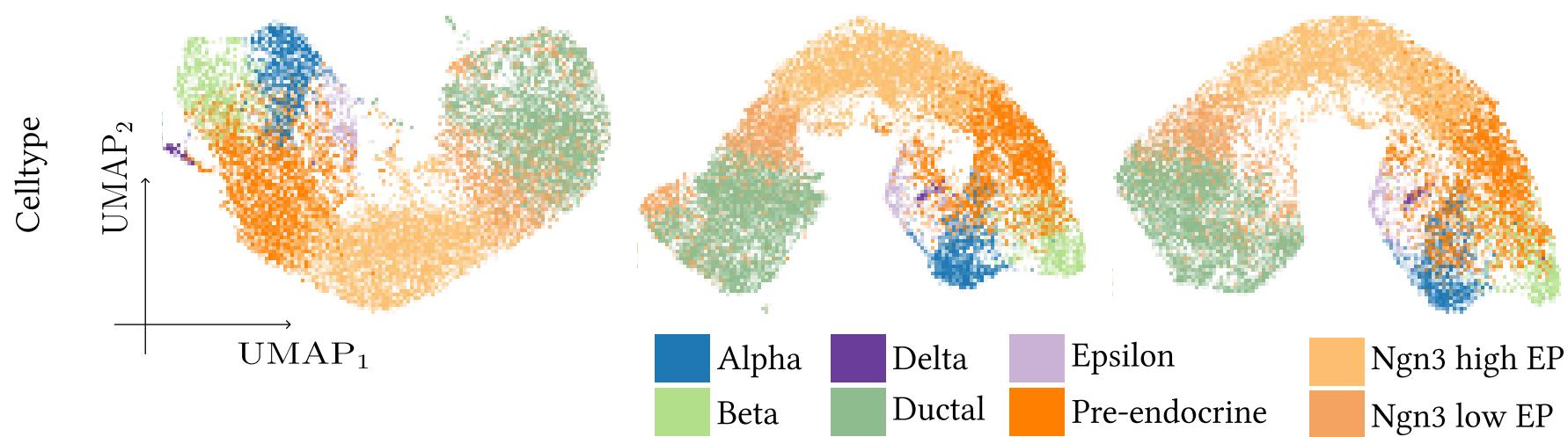
Welchs t -test significance *** $P < 10^{-3}$

Integration of modalities

a



b



[1] R. Lopez et al. "Deep generative modeling for single-cell transcriptomics", Nature Methods (2018)

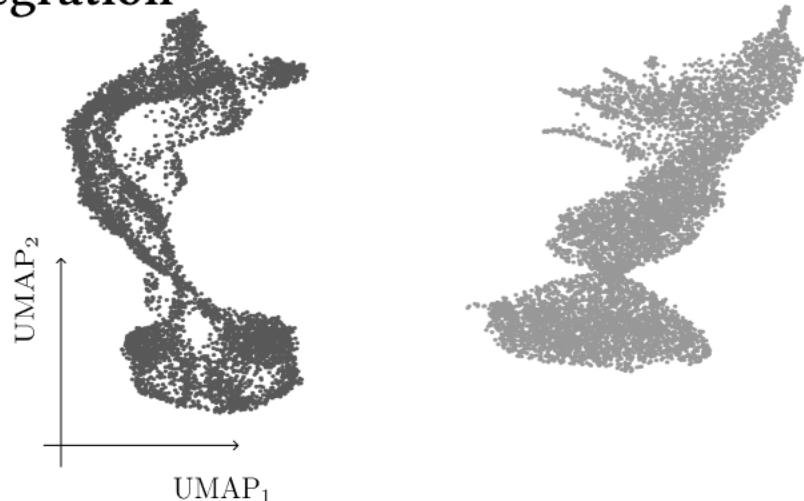
[2] Z.-J. Cao and G. Gao. "Multi-omics single-cell data integration and regulatory inference with graph-linked embedding", Nature Biotechnology (2022)

scIB metrics³

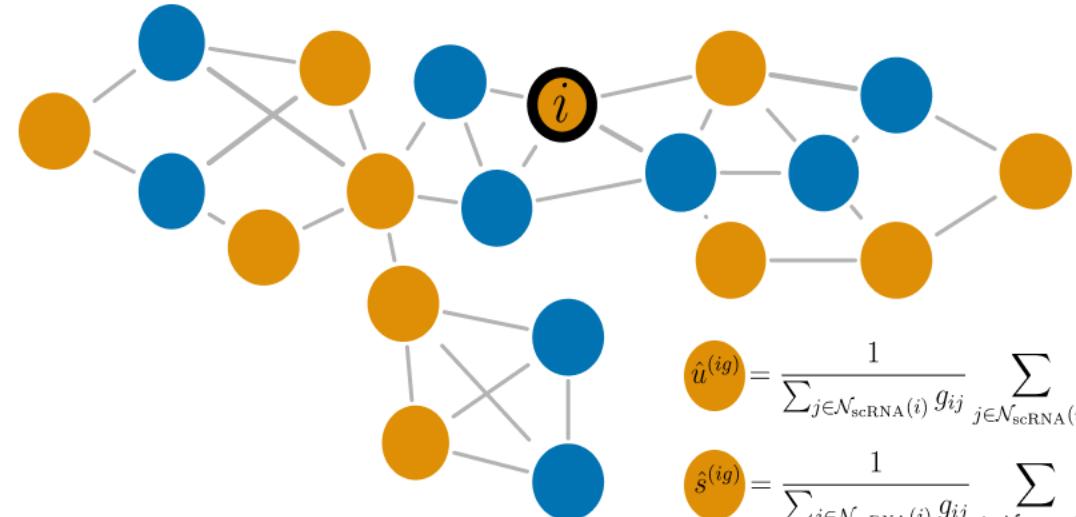


Methods outline

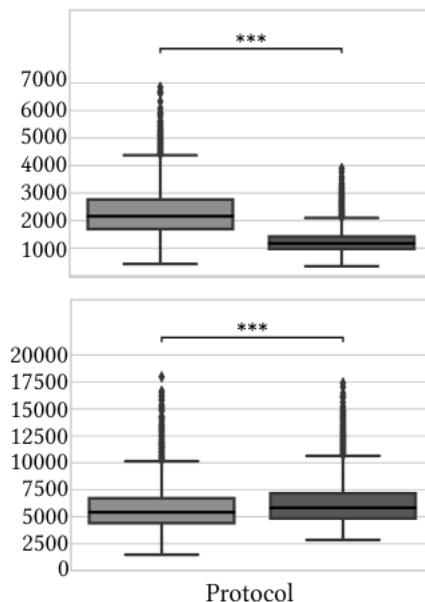
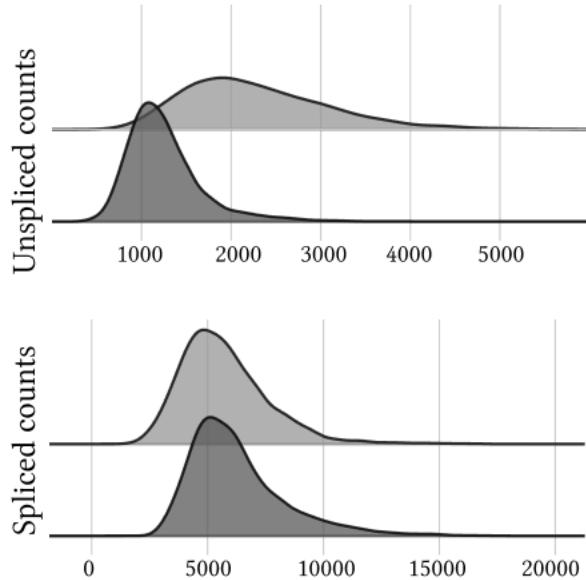
1. Integration



2. Estimation

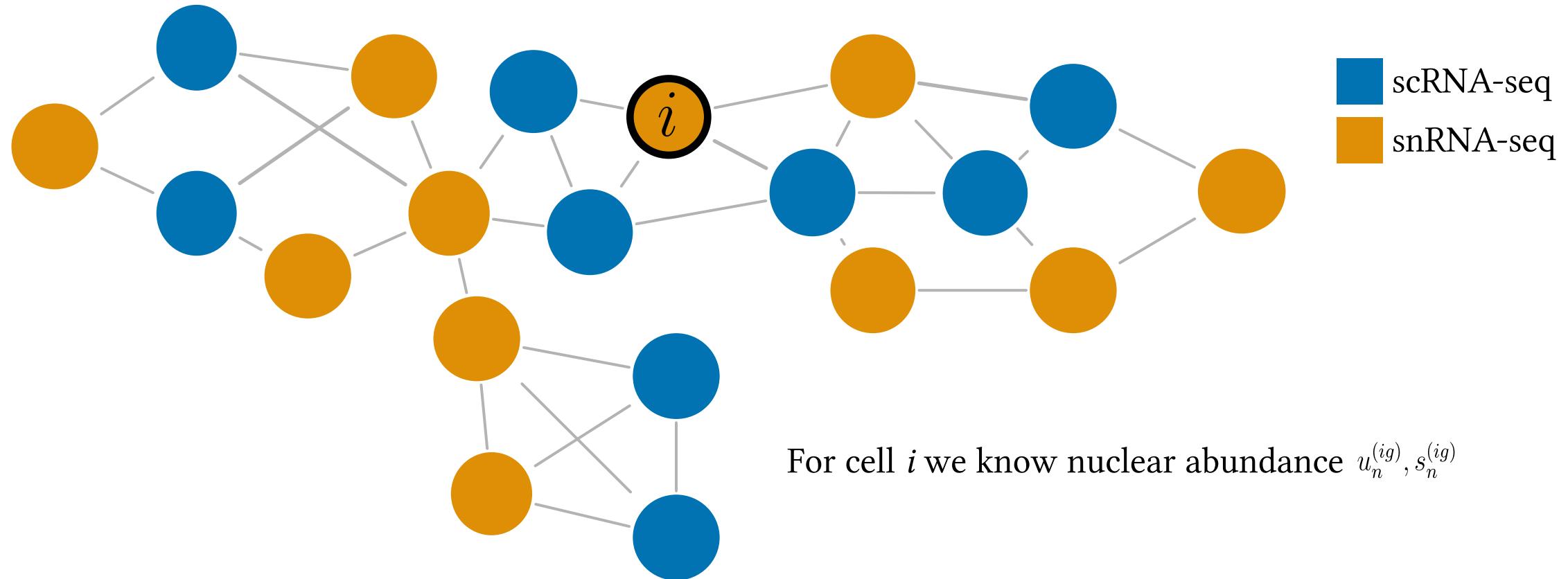


3. Lambda correction

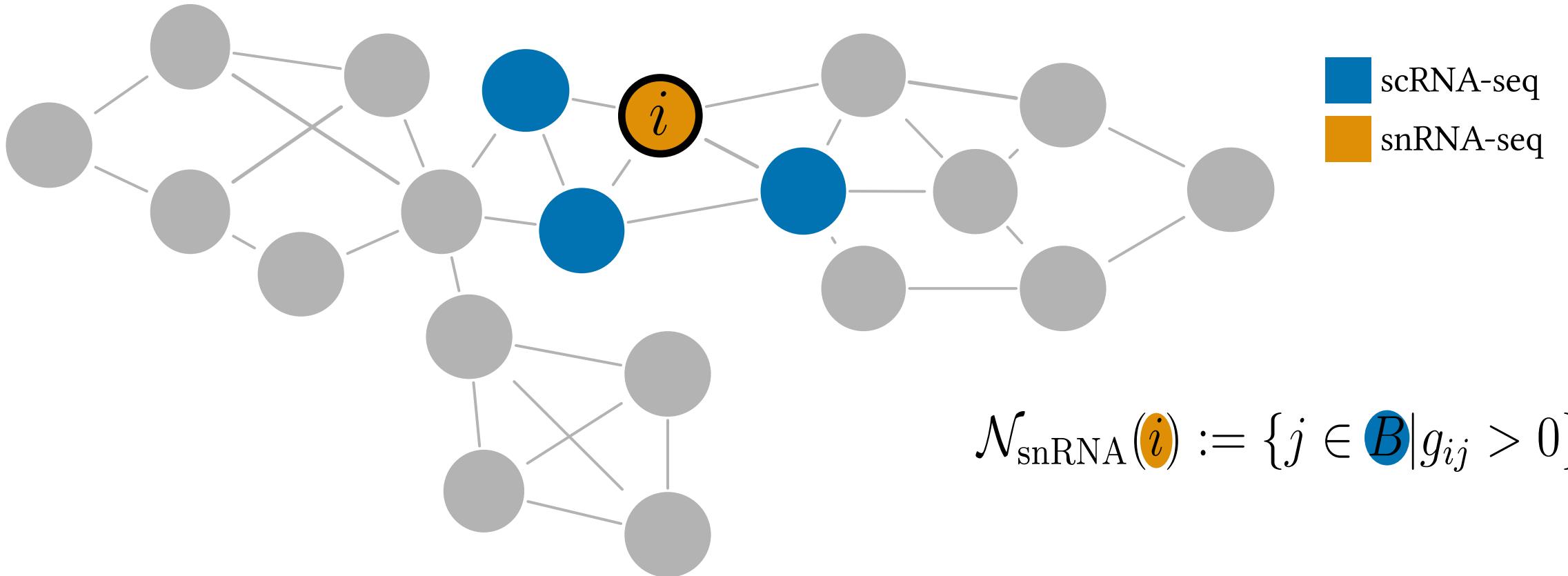


Welchs t -test significance *** $P < 10^{-3}$

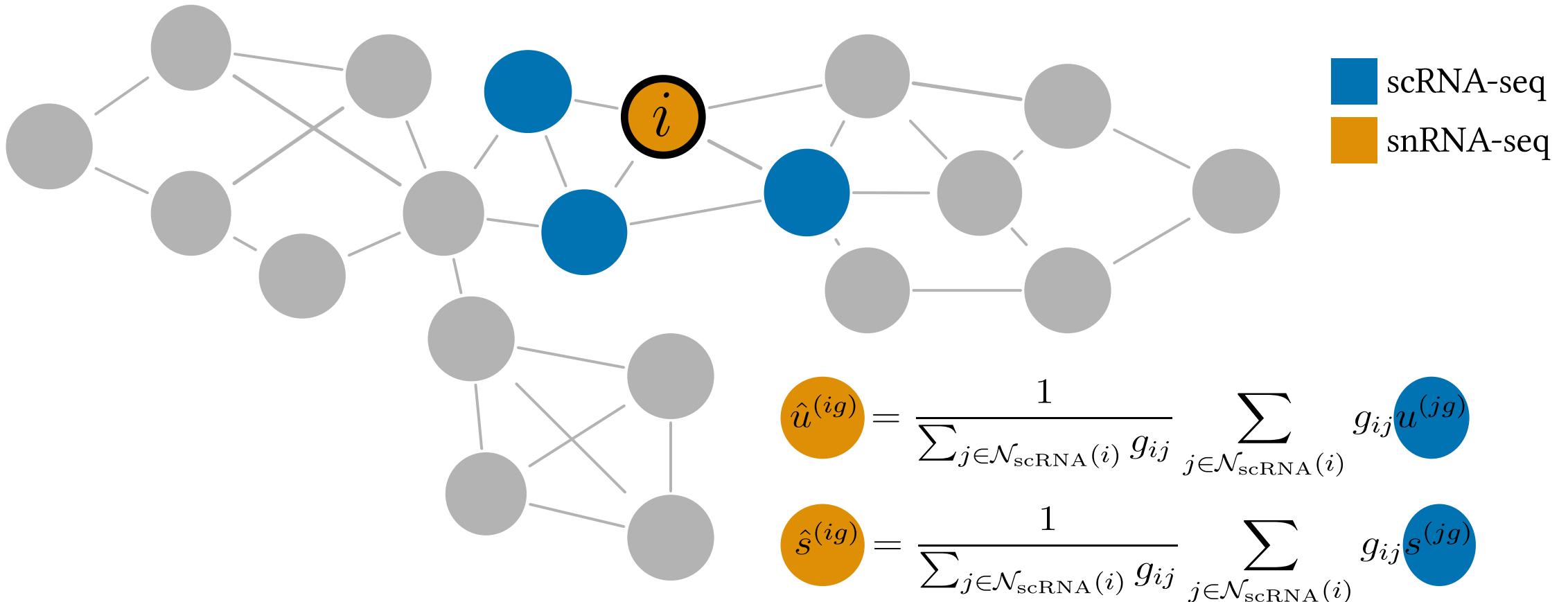
Estimation of nuclear and cytosolic abundances



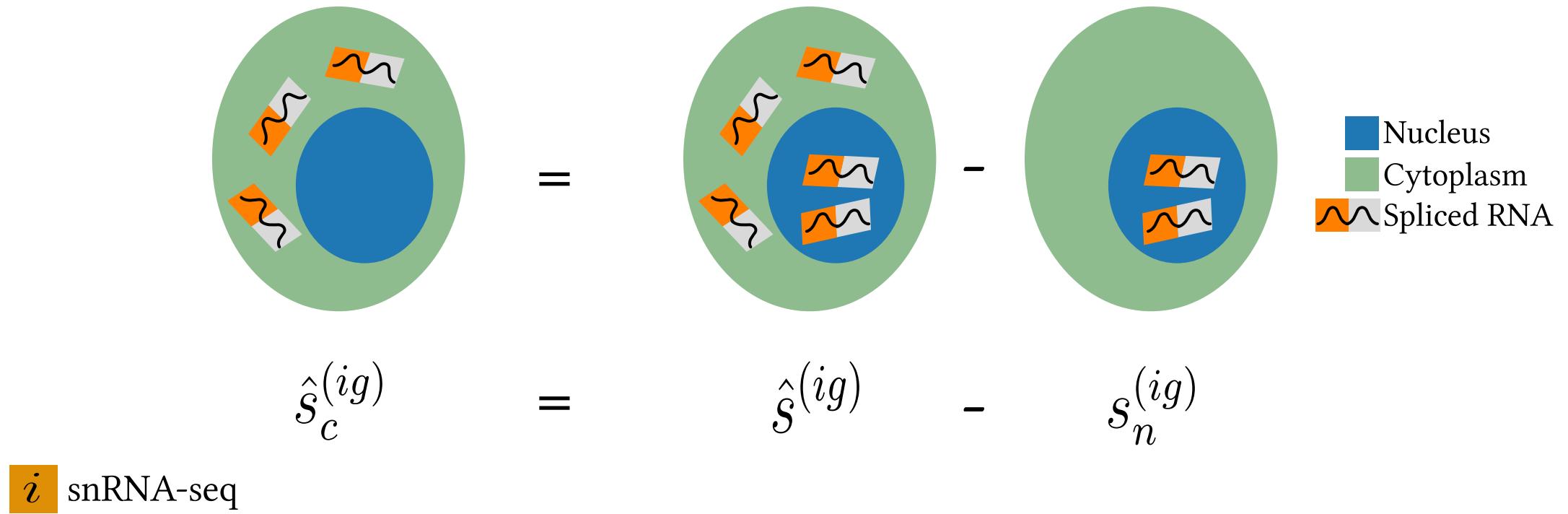
Estimation of cellular abundances



Estimation of cellular abundances

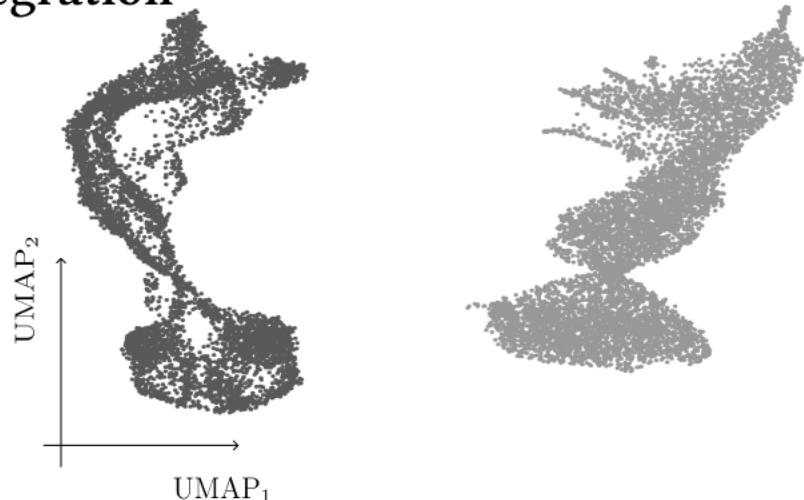


Estimation of spliced cytosolic abundance

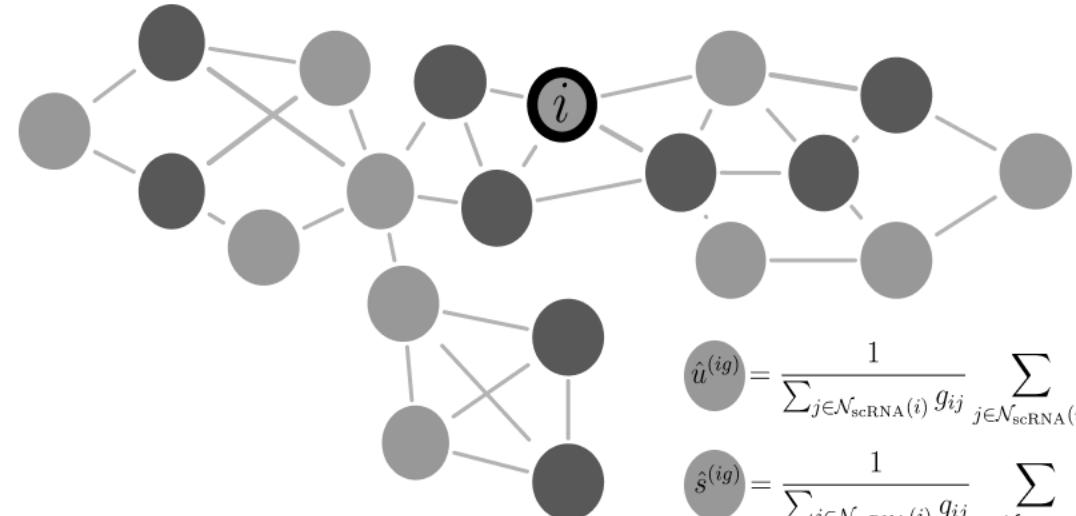


Methods outline

1. Integration

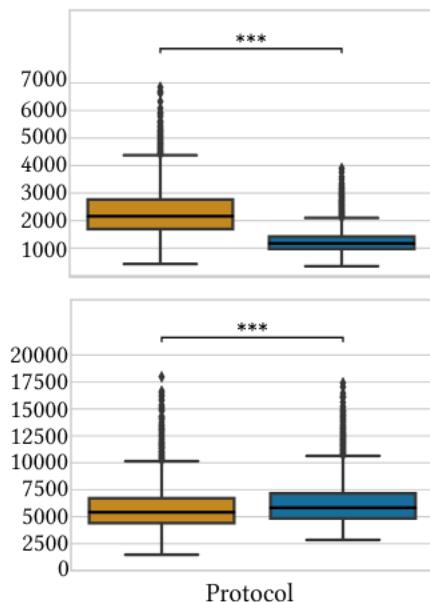
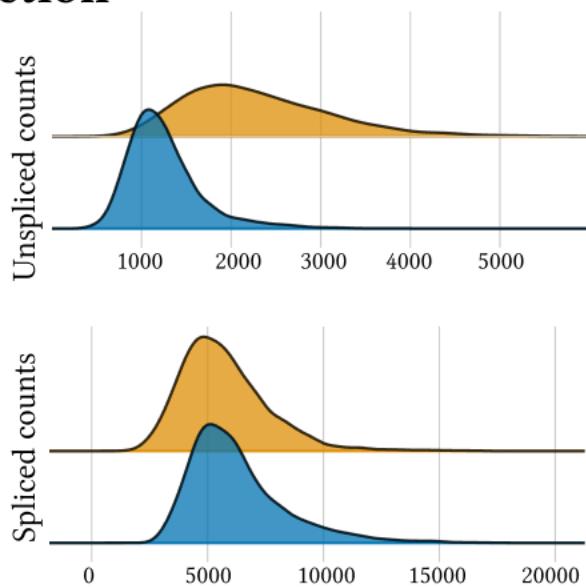


2. Estimation



$$\hat{u}^{(ig)} = \frac{1}{\sum_{j \in \mathcal{N}_{\text{scRNA}}(i)} g_{ij}} \sum_{j \in \mathcal{N}_{\text{scRNA}}(i)} g_{ij} u^{(jg)}$$
$$\hat{s}^{(ig)} = \frac{1}{\sum_{j \in \mathcal{N}_{\text{scRNA}}(i)} g_{ij}} \sum_{j \in \mathcal{N}_{\text{scRNA}}(i)} g_{ij} s^{(jg)}$$

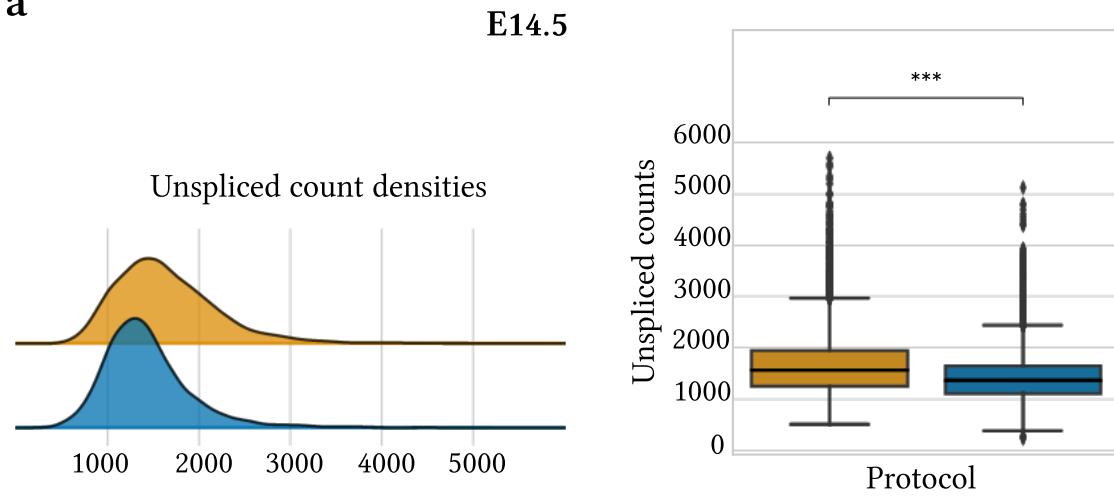
3. Lambda correction



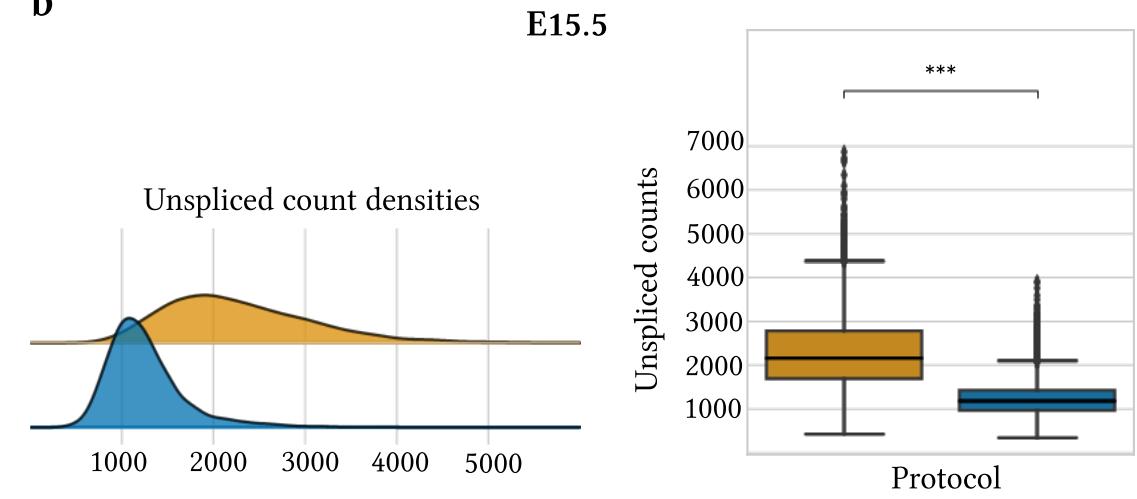
Welchs *t*-test significance *** $P < 10^{-3}$

Lambda Correction - pancreatic endocrinogenesis datasets^{4,5}

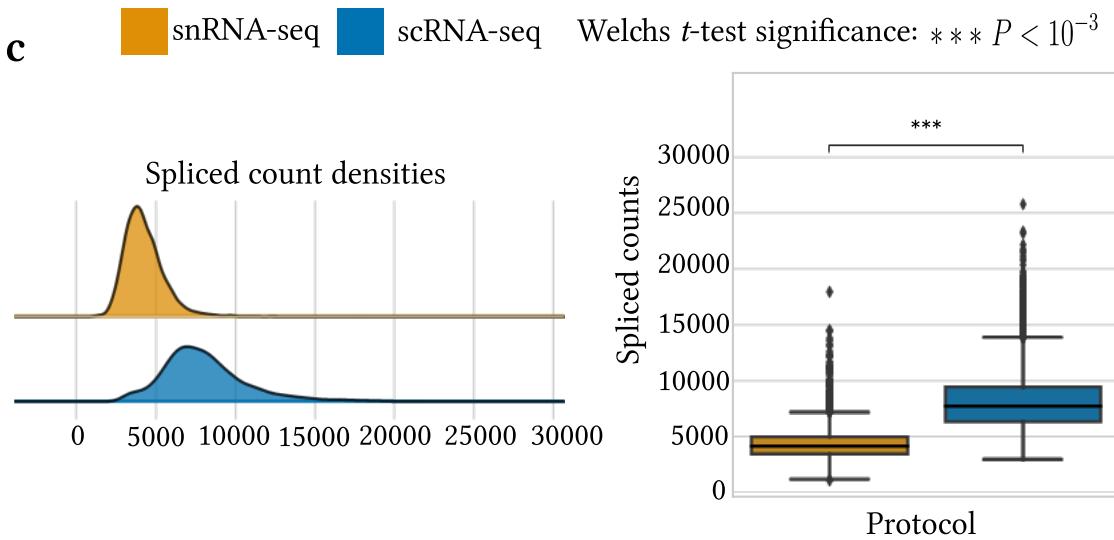
a



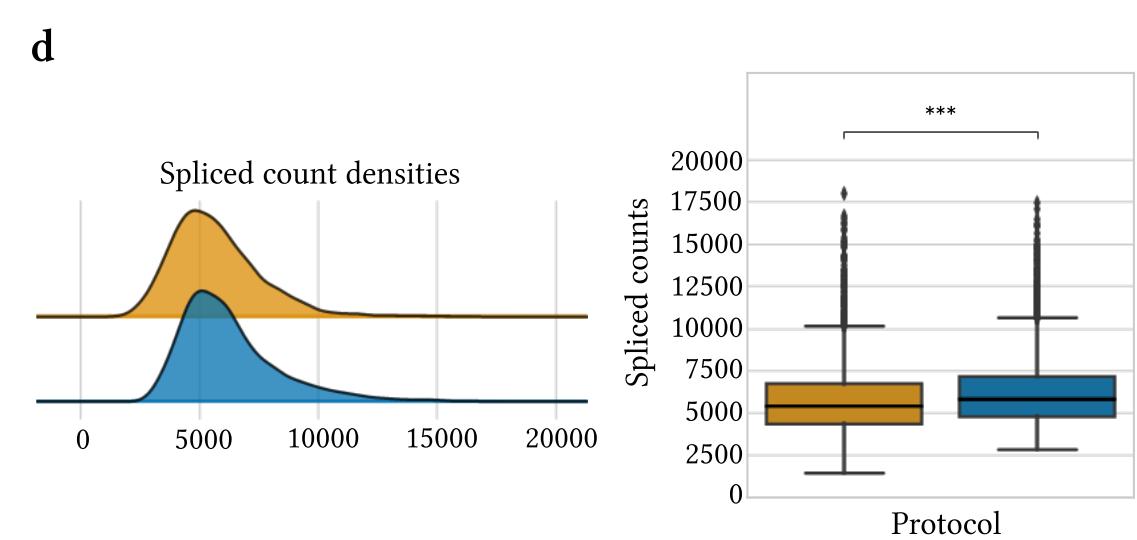
b



c



d



[4] A. Bastidas-Ponce et al. "Massive single-cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis", In: Development (2019).

[5] D. Klein et al. "Mapping cells through time and space with moscot", In: Development (2023)

Lambda Correction - Steps

1. Estimated/ Measured count matrices

$$U = \underbrace{(\hat{u}, u)^T}_{\text{Estimate based on scRNA}}, U_n = \underbrace{(u_n, \hat{u}_n)^T}_{\text{Estimate based on snRNA}}, S_n = (s_n, \hat{s}_n)^T \in \mathbb{R}^{N \times G}$$

Lambda Correction - Steps

1. Estimated/ Measured count matrices

$$U = \underbrace{(\hat{u}, u)^T}_{\text{Estimate based on scRNA}}, U_n = \underbrace{(u_n, \hat{u}_n)^T}_{\text{Estimate based on snRNA}}, S_n = (s_n, \hat{s}_n)^T \in \mathbb{R}^{N \times G} \longrightarrow 2. \text{ Define scaling matrix}$$
$$\Lambda_{ij} = \begin{cases} \frac{U_{ij}}{(U_n)_{ij}} & \text{if } U_{ij}, (U_n)_{ij} \neq 0 \\ 0 & \text{else} \end{cases}$$

Lambda Correction - Steps

1. Estimated/ Measured count matrices

$$U = \underbrace{(\hat{u}, u)^T}_{\text{Estimate based on scRNA}}, U_n = \underbrace{(u_n, \hat{u}_n)^T}_{\text{Estimate based on snRNA}}, S_n = (s_n, \hat{s}_n)^T \in \mathbb{R}^{N \times G}$$

—————>

$$\Lambda_{ij} = \begin{cases} \frac{U_{ij}}{(U_n)_{ij}} & \text{if } U_{ij}, (U_n)_{ij} \neq 0 \\ 0 & \text{else} \end{cases}$$

2. Define scaling matrix



3. Scale nucleic count matrices

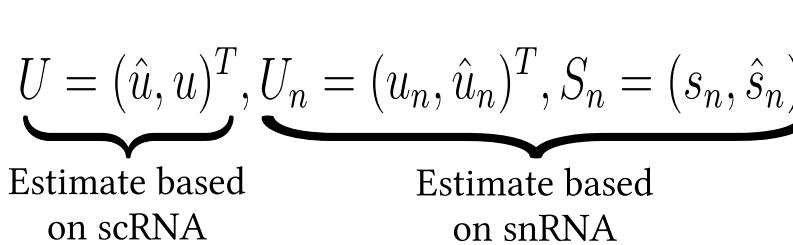
$$(S_n)_{ij} = \Lambda_{ij}(S_n)_{ij}$$

$$(U_n)_{ij} = \Lambda_{ij}(U_n)_{ij}$$

Lambda Correction - Steps

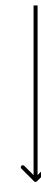
1. Estimated/ Measured count matrices

$$U = (\hat{u}, u)^T, U_n = (u_n, \hat{u}_n)^T, S_n = (s_n, \hat{s}_n)^T \in \mathbb{R}^{N \times G}$$



2. Define scaling matrix

$$\Lambda_{ij} = \begin{cases} \frac{U_{ij}}{(U_n)_{ij}} & \text{if } U_{ij}, (U_n)_{ij} \neq 0 \\ 0 & \text{else} \end{cases}$$



4. Recalculate cytosolic abundance

$$\hat{s}_c^{(ig)} = \hat{s}^{(ig)} - s_n^{(ig)}$$

3. Scale nucleic count matrices

$$(S_n)_{ij} = \Lambda_{ij}(S_n)_{ij}$$

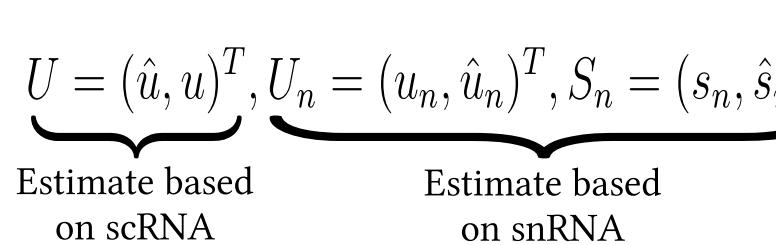
$$(U_n)_{ij} = \Lambda_{ij}(U_n)_{ij}$$



Lambda Correction - Steps

1. Estimated/ Measured count matrices

$$U = (\hat{u}, u)^T, U_n = (u_n, \hat{u}_n)^T, S_n = (s_n, \hat{s}_n)^T \in \mathbb{R}^{N \times G}$$



\longrightarrow

2. Define scaling matrix

$$\Lambda_{ij} = \begin{cases} \frac{U_{ij}}{(U_n)_{ij}} & \text{if } U_{ij}, (U_n)_{ij} \neq 0 \\ 0 & \text{else} \end{cases}$$



5. Drop genes where

$$\lambda_g = \sum_{i=1}^N \Lambda_{ig} == 0$$

\leftarrow

4. Recalculate cytosolic abundance

$$\hat{s}_c^{(ig)} = \hat{s}^{(ig)} - s_n^{(ig)}$$

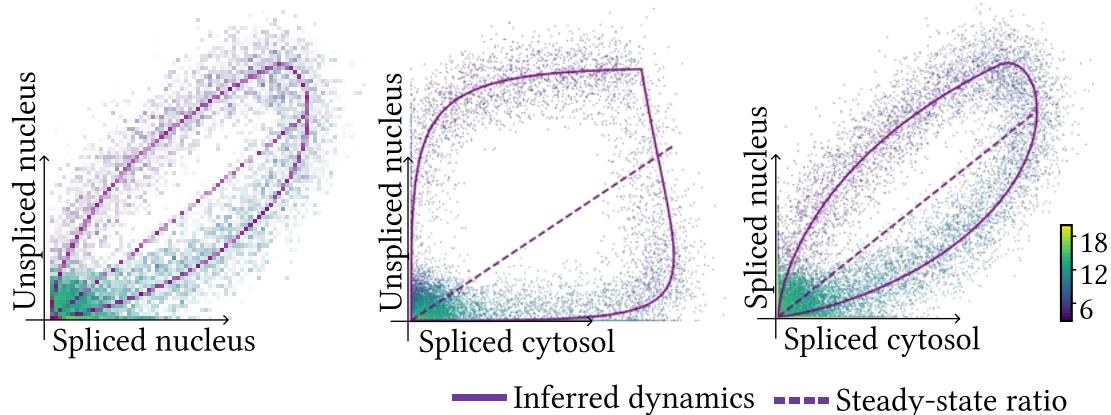
\leftarrow

3. Scale nucleic count matrices

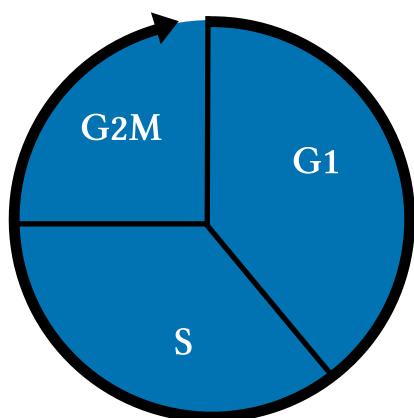
$$(S_n)_{ij} = \Lambda_{ij}(S_n)_{ij}$$
$$(U_n)_{ij} = \Lambda_{ij}(U_n)_{ij}$$

Results

1. Data simulation



3. Cell cycle analysis - Pancreas E14.5^{4,5}

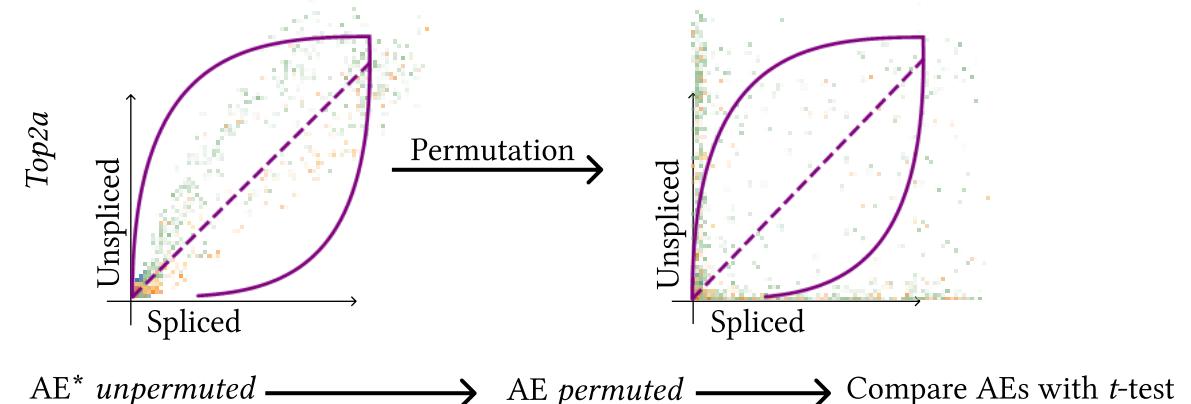


*AE = Absolute error

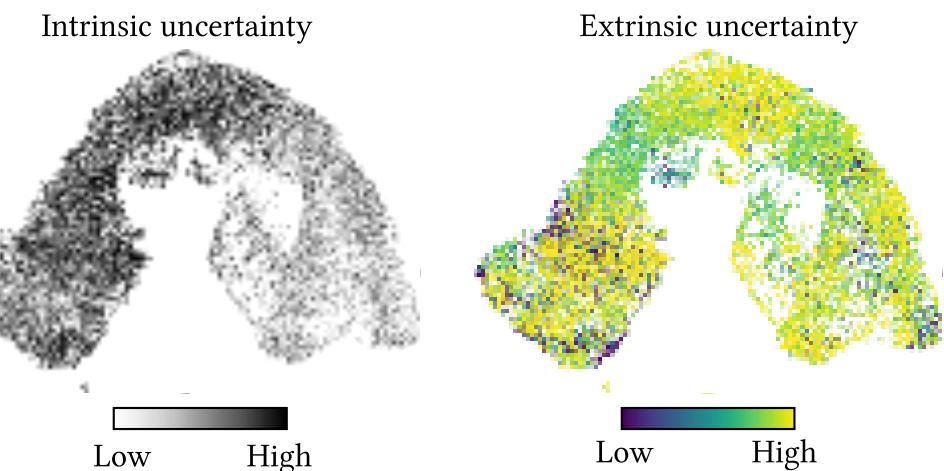
[4] A. Bastidas-Ponce et al. "Massive single-cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis", In: Development (2019).

[5] D. Klein et al. "Mapping cells through time and space with moscot", In: Development (2023)

2. Permutation score analysis - Pancreas E14.5^{4,5}

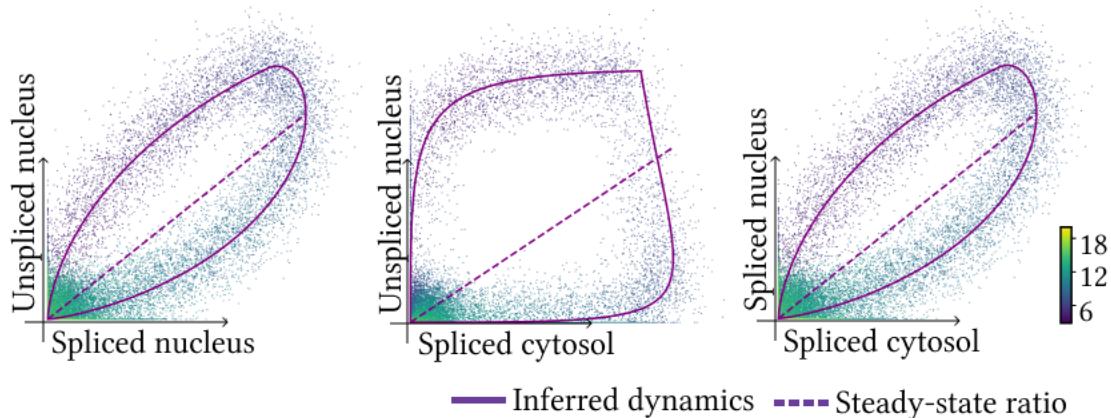


4. Uncertainty analysis Pancreas E14.5^{4,5}

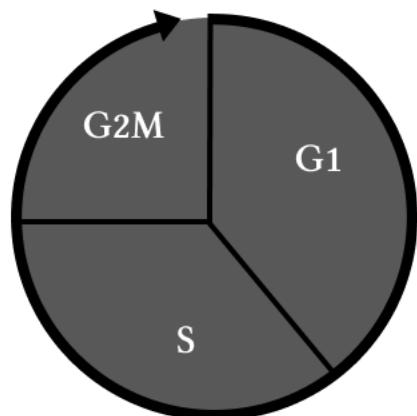


Results

1. Data simulation



3. Cell cycle analysis - Pancreas E14.5^{4,5}

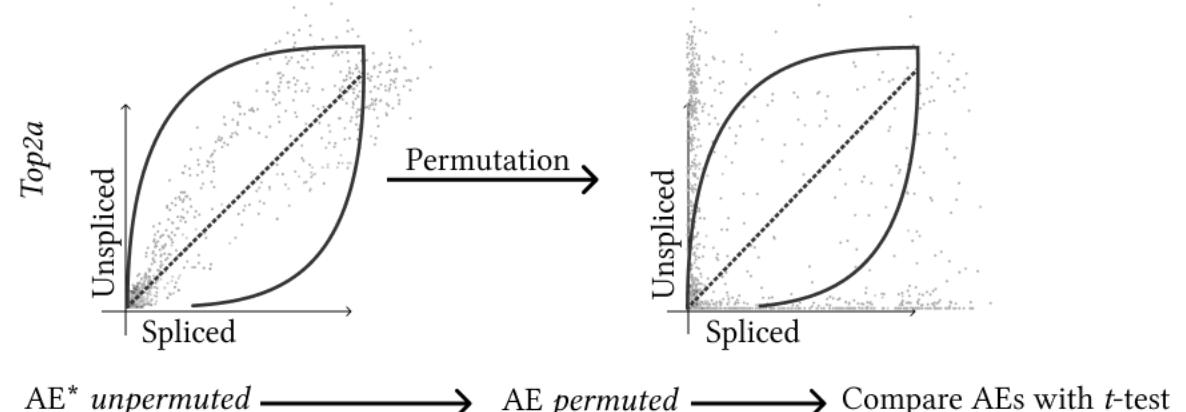


*AE = Absolute error

[4] A. Bastidas-Ponce et al. "Massive single-cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis", In: Development (2019).

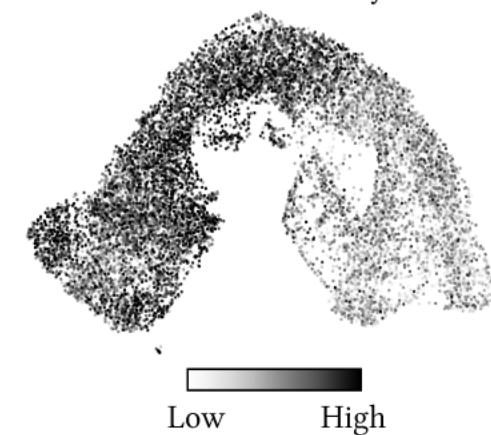
[5] D. Klein et al. "Mapping cells through time and space with moscot", In: Development (2023)

2. Permutation score analysis - Pancreas E14.5^{4,5}

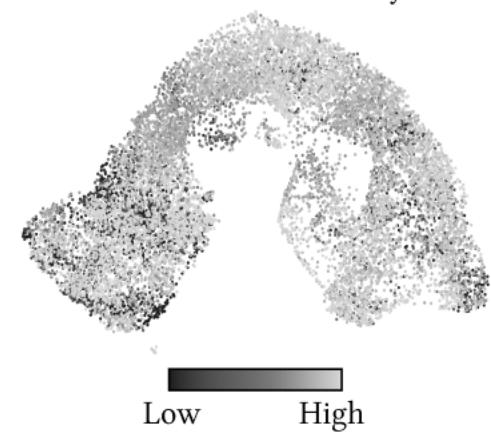


4. Uncertainty analysis Pancreas E14.5^{4,5}

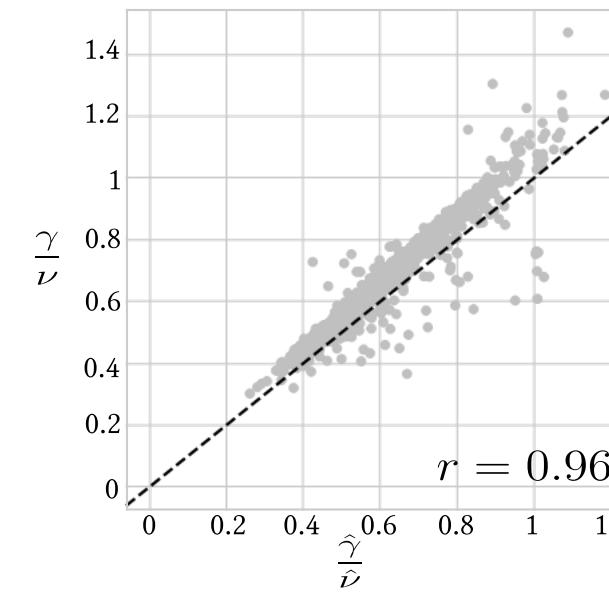
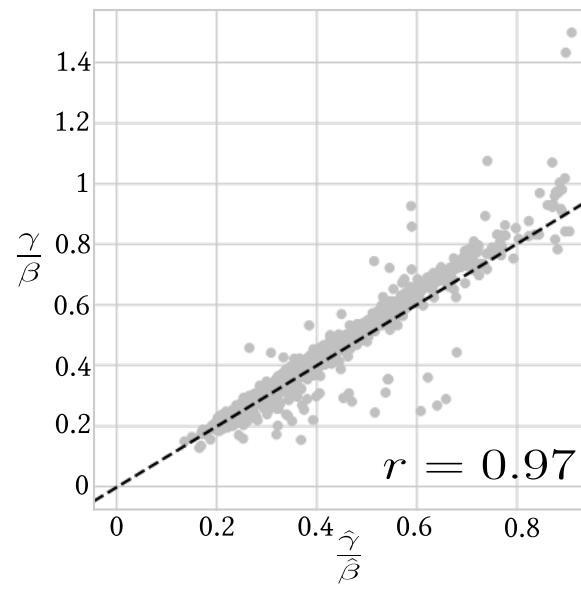
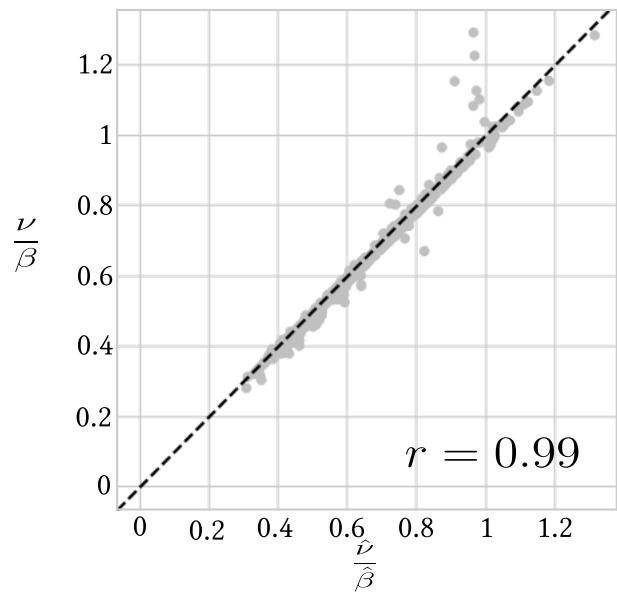
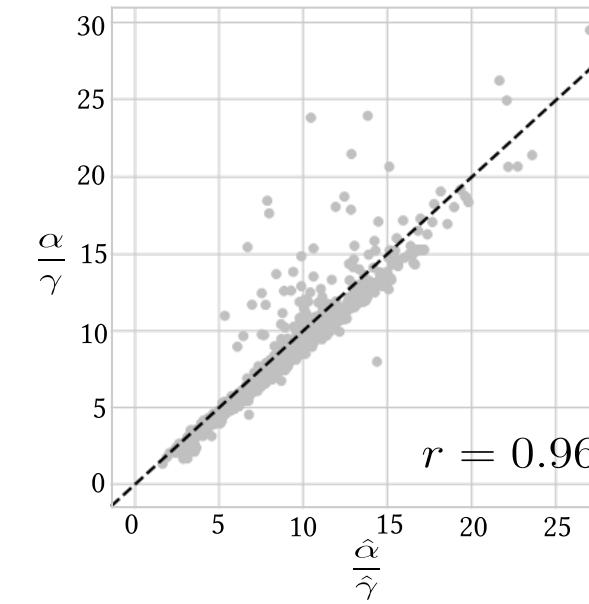
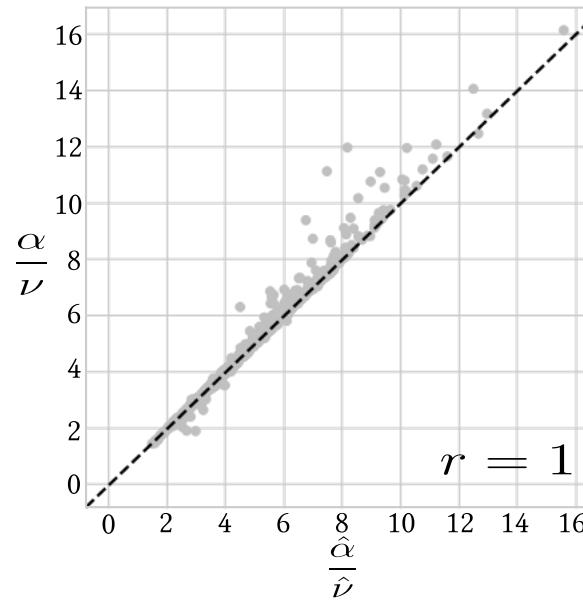
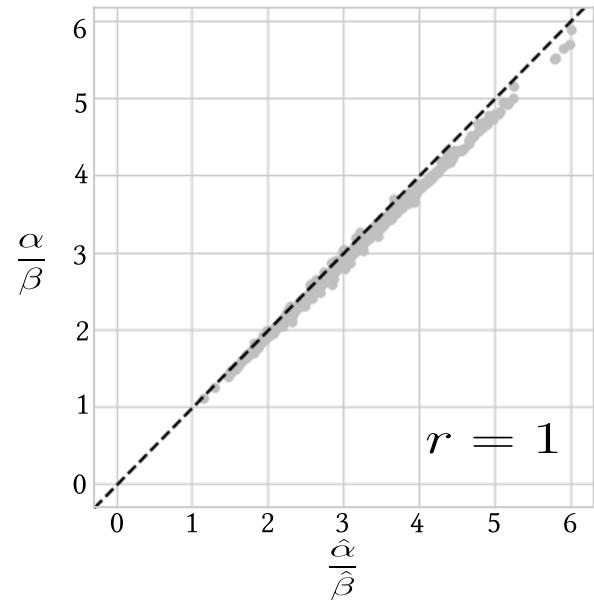
Intrinsic uncertainty



Extrinsic uncertainty

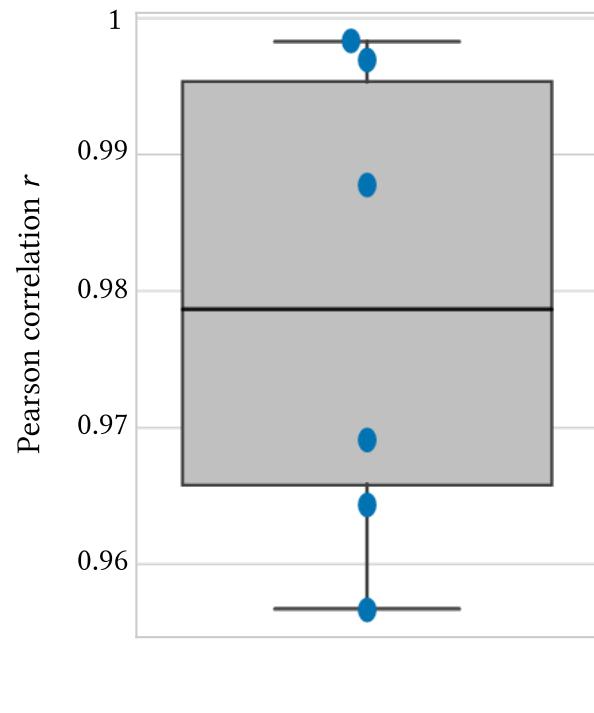


Data simulation - Rate parameter inference

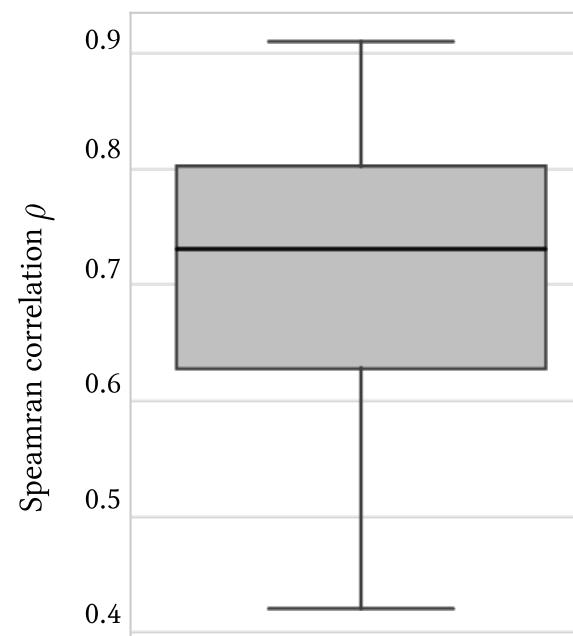


Data simulation - Correlations and time assignments

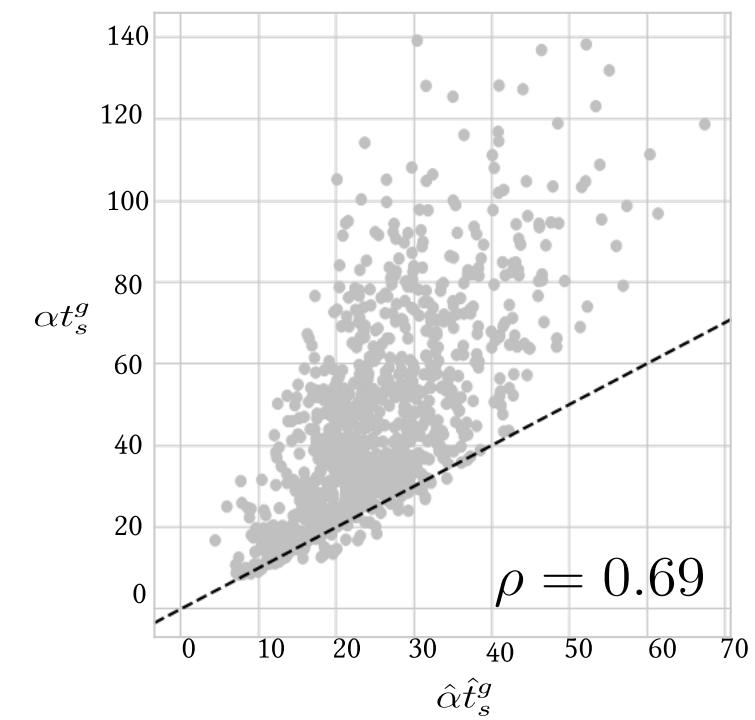
Rate parameter correlations



Latent/ true time correlations

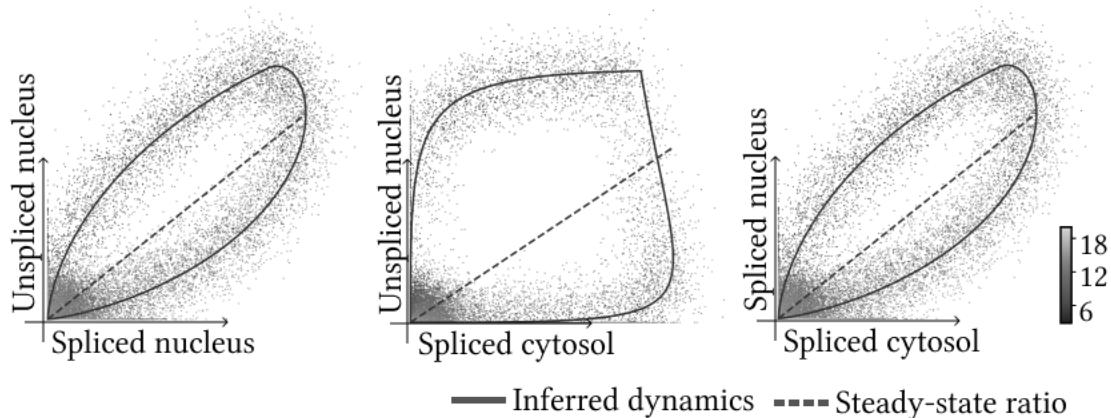


Switching times

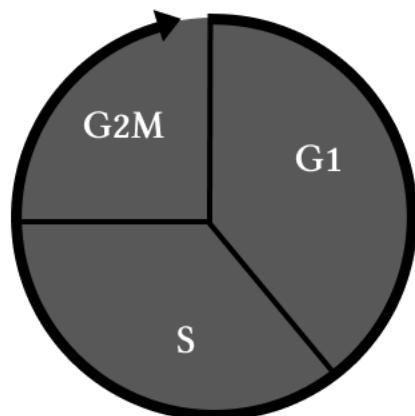


Results

1. Data simulation



3. Cell cycle analysis - Pancreas E14.5^{4,5}

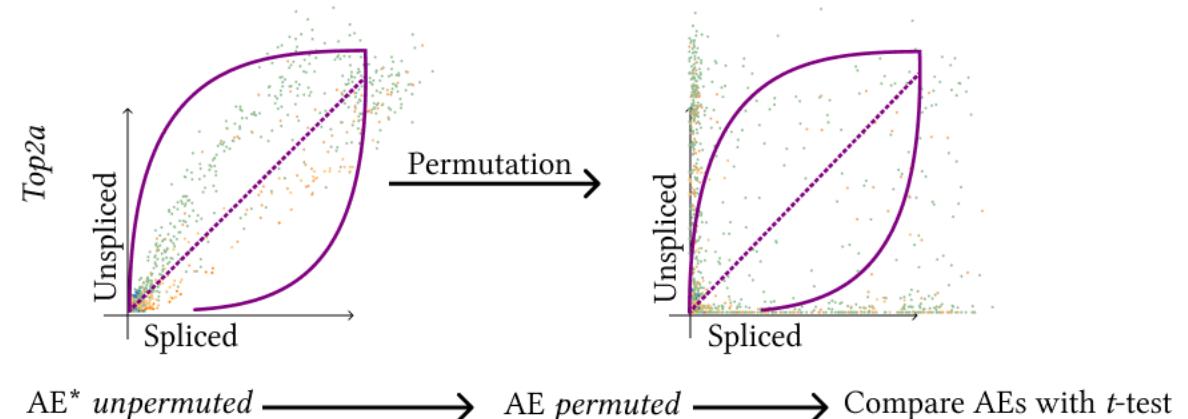


*AE = Absolute error

[4] A. Bastidas-Ponce et al. "Massive single-cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis", In: Development (2019).

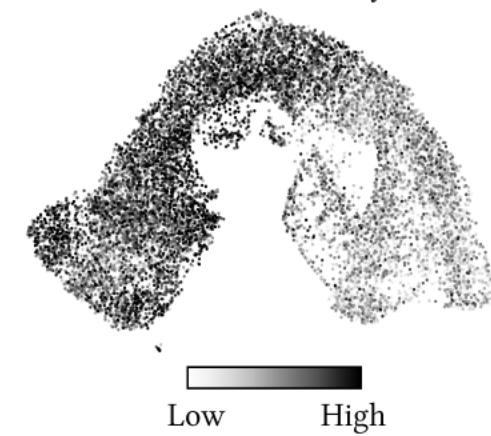
[5] D. Klein et al. "Mapping cells through time and space with moscot", In: Development (2023)

2. Permutation score analysis - Pancreas E14.5^{4,5}

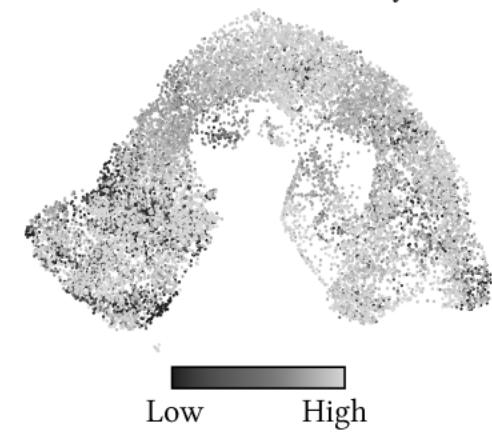


4. Uncertainty analysis Pancreas E14.5^{4,5}

Intrinsic uncertainty

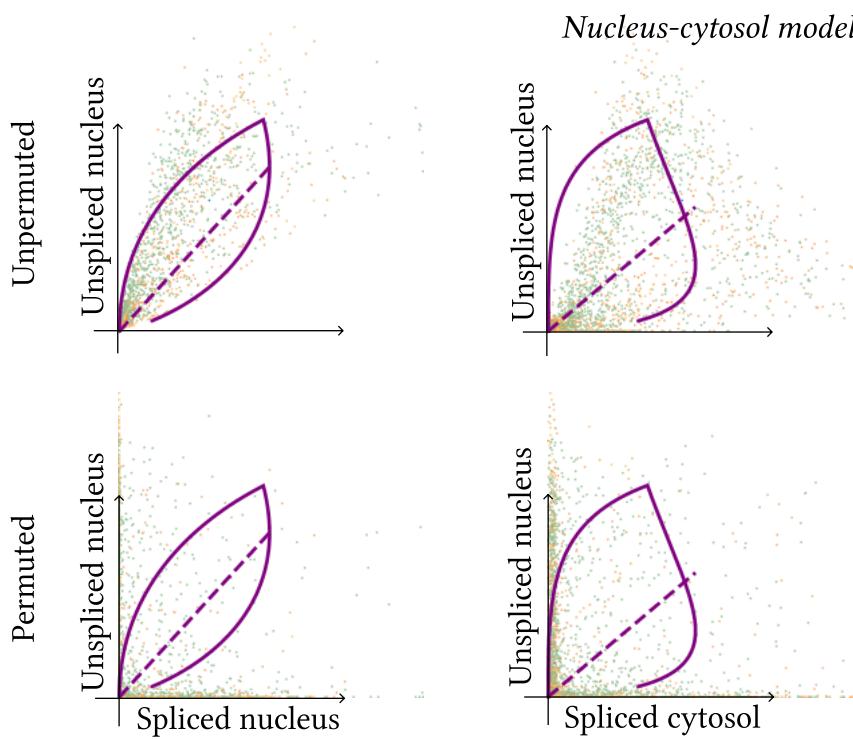


Extrinsic uncertainty

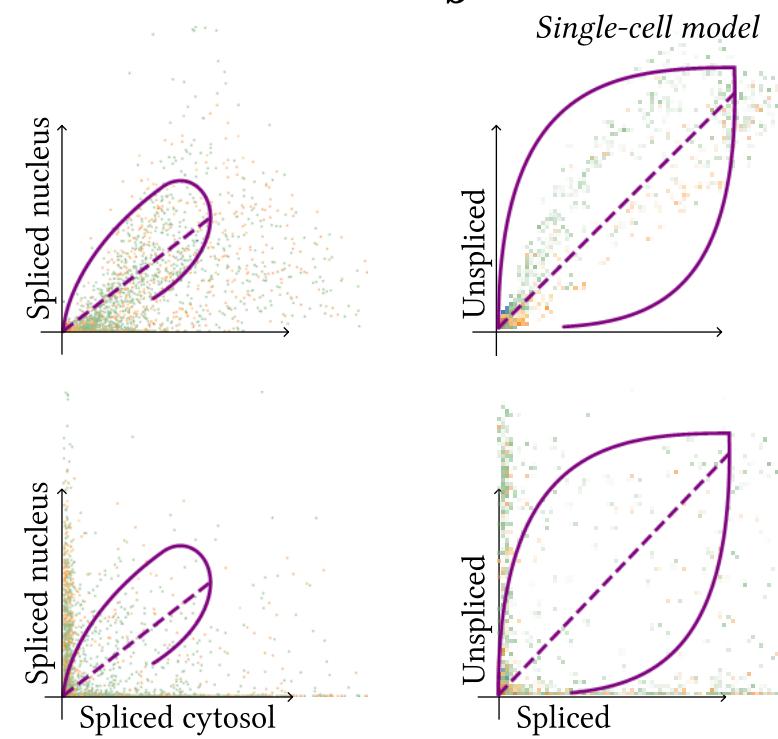


Permutation score analysis - *Top2a*

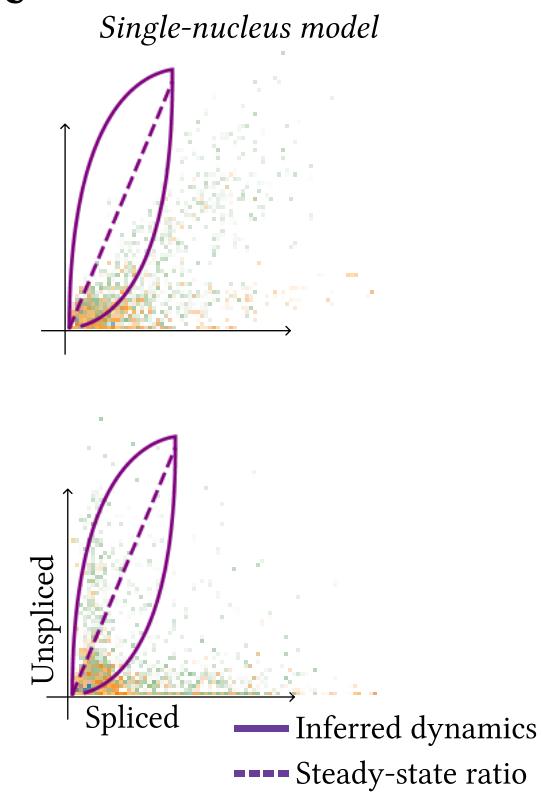
a



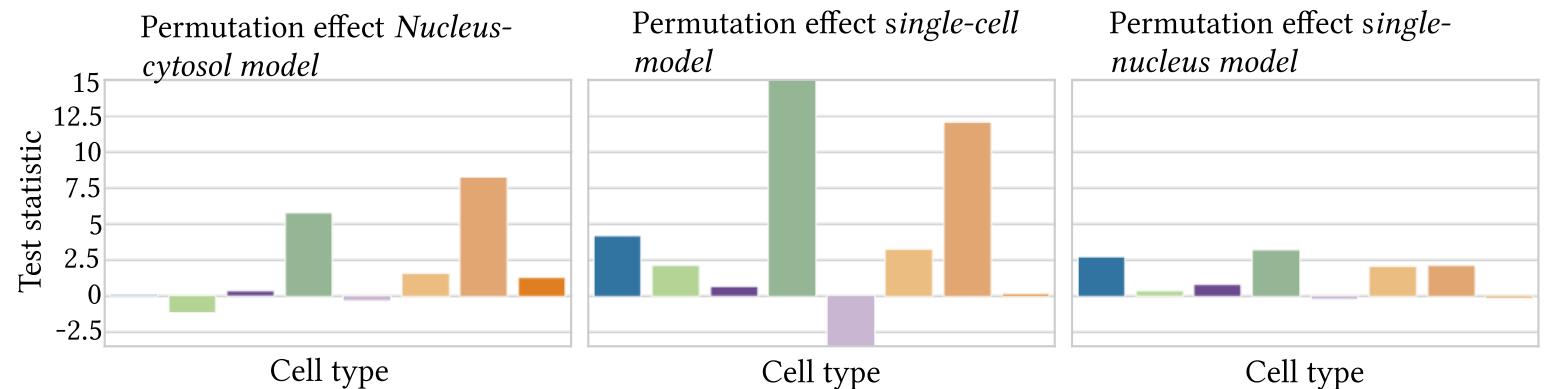
b



c



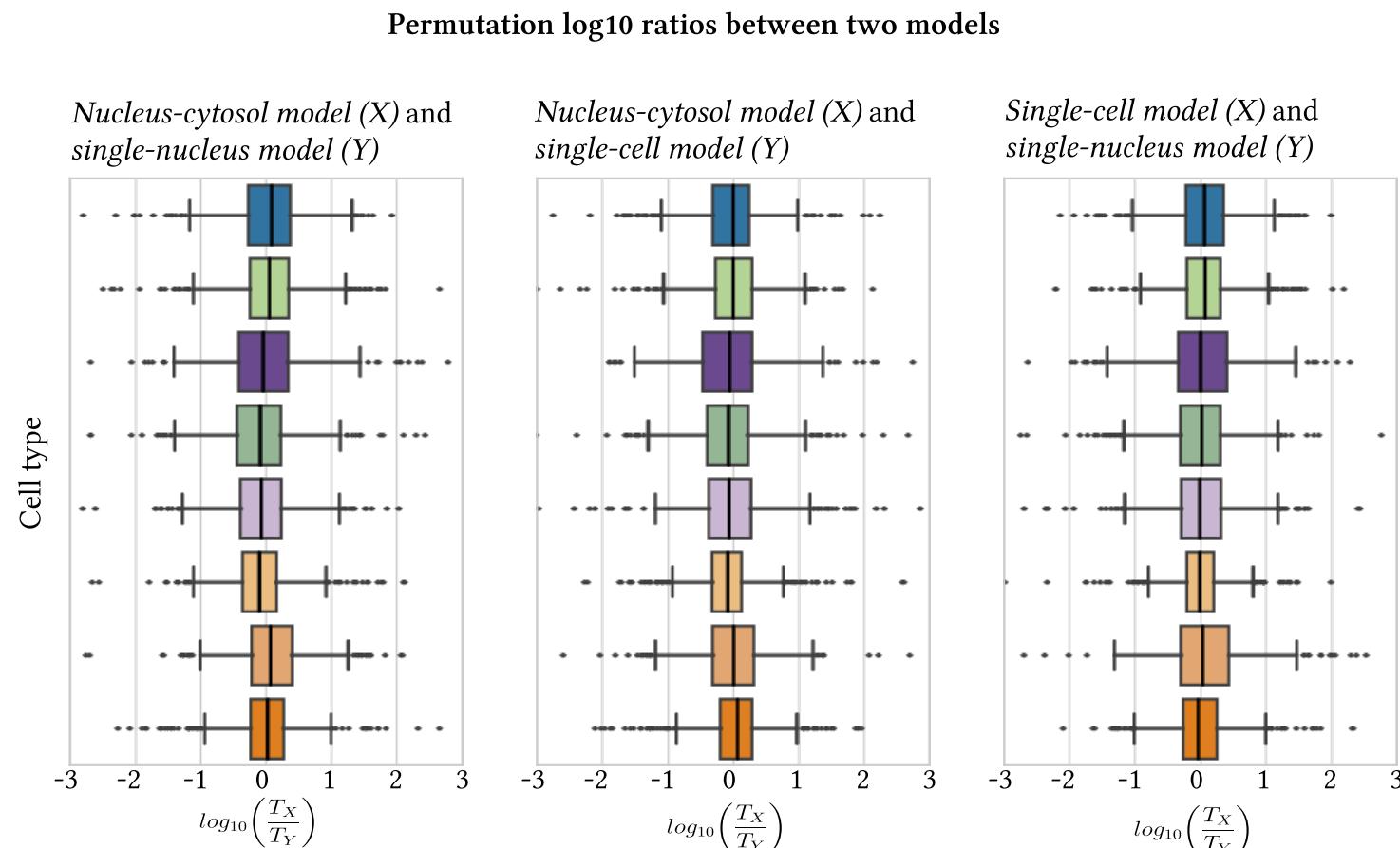
d



Alpha	Epsilon
Beta	Pre-endocrine
Delta	Ngn3 high EP
Ductal	Ngn3 low EP

Permutation score analysis

a

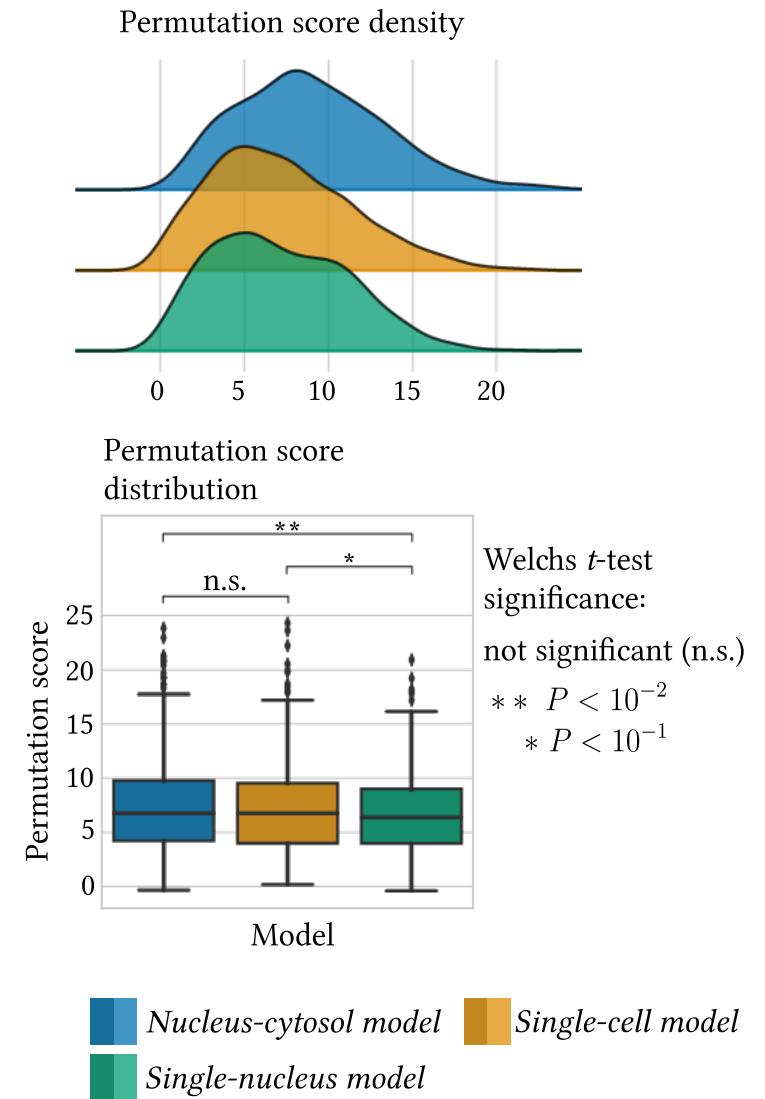


T_X : Permutation score of *Model X* per gene and cell type

T_Y : Permutation score of *Model Y* per gene and cell type

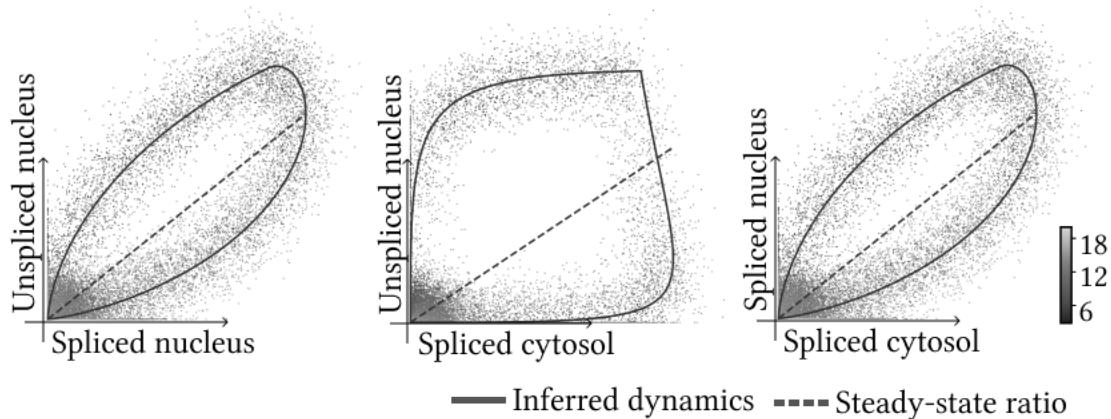
Alpha	Delta	Epsilon	Ngn3 high EP
Beta	Ductal	Pre-endocrine	Ngn3 low EP

b

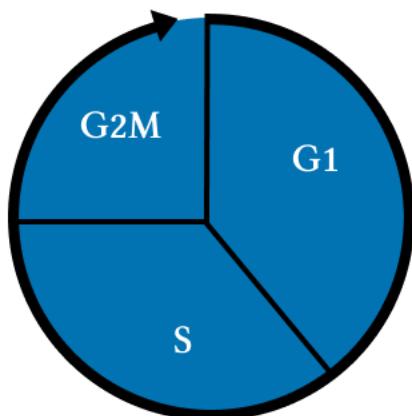


Results

1. Data simulation



3. Cell cycle analysis - Pancreas E14.5^{4,5}

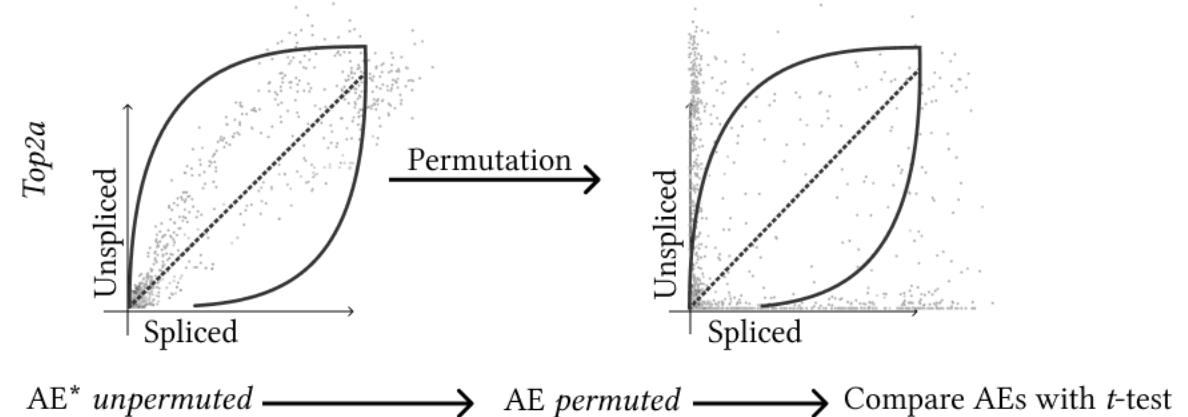


*AE = Absolute error

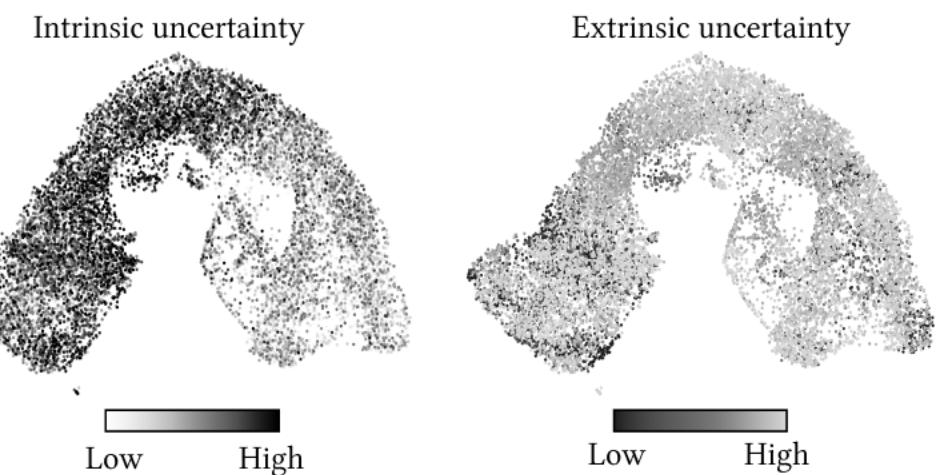
[4] A. Bastidas-Ponce et al. "Massive single-cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis", In: Development (2019).

[5] D. Klein et al. "Mapping cells through time and space with moscot", In: Development (2023)

2. Permutation score analysis - Pancreas E14.5^{4,5}

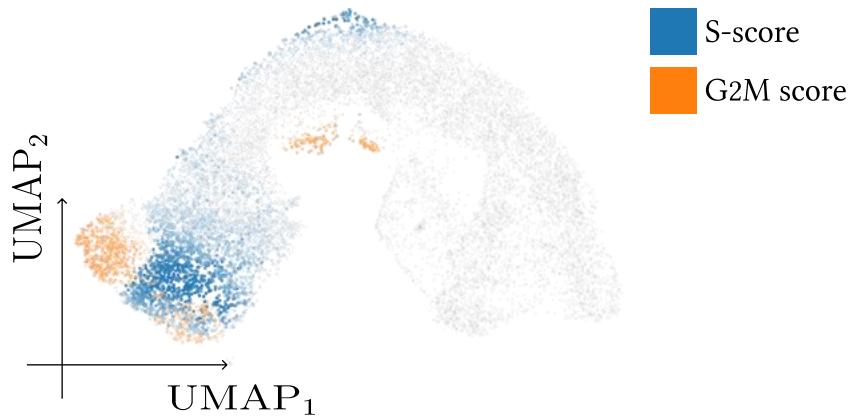


4. Uncertainty analysis Pancreas E14.5^{4,5}

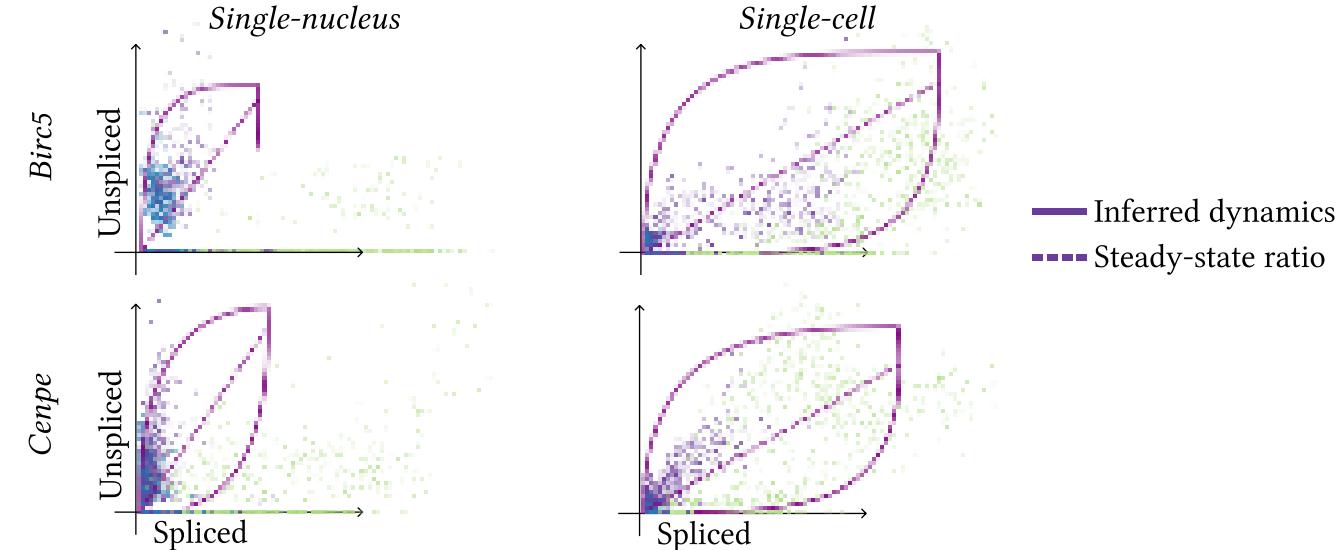


Cell cycle analysis

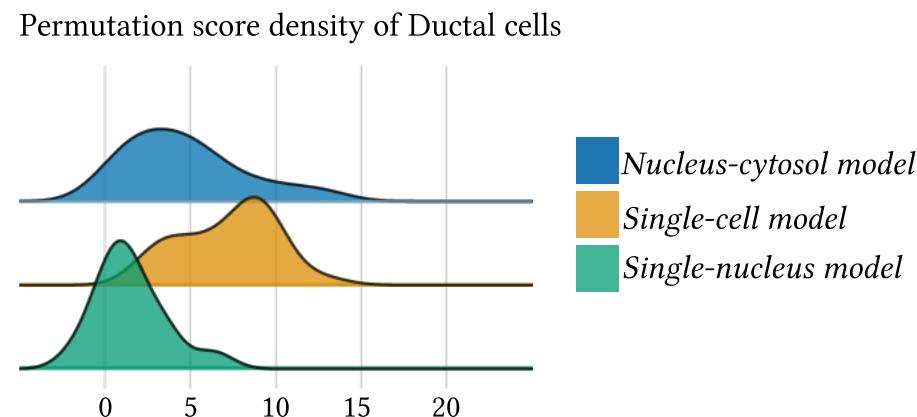
1. Cell cycle phase assignment



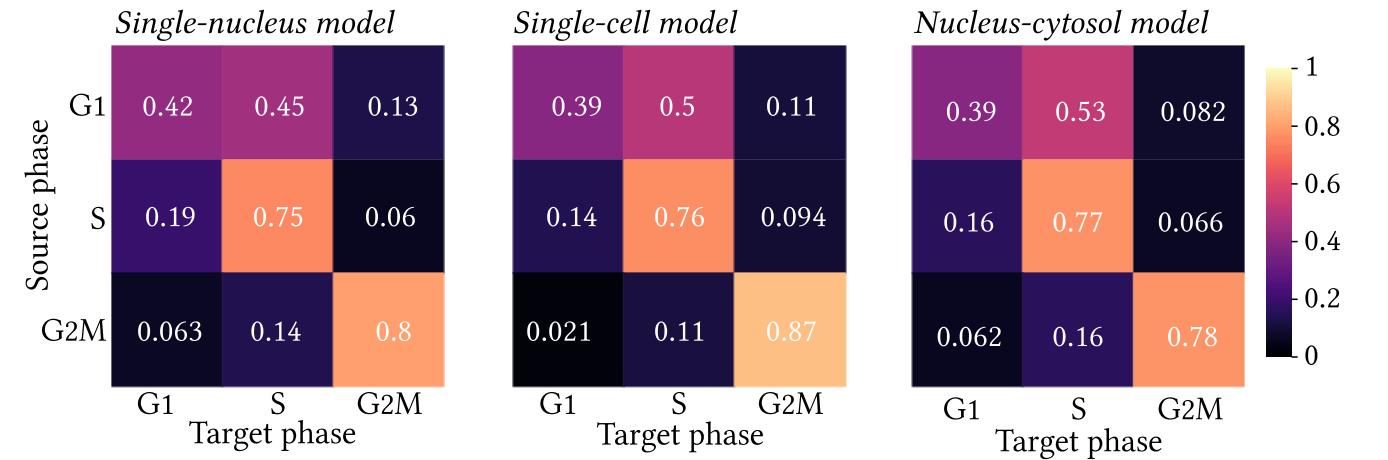
2. Investigate phase portraits



3. Compare permutation scores



4. Cell cycle transition probabilities

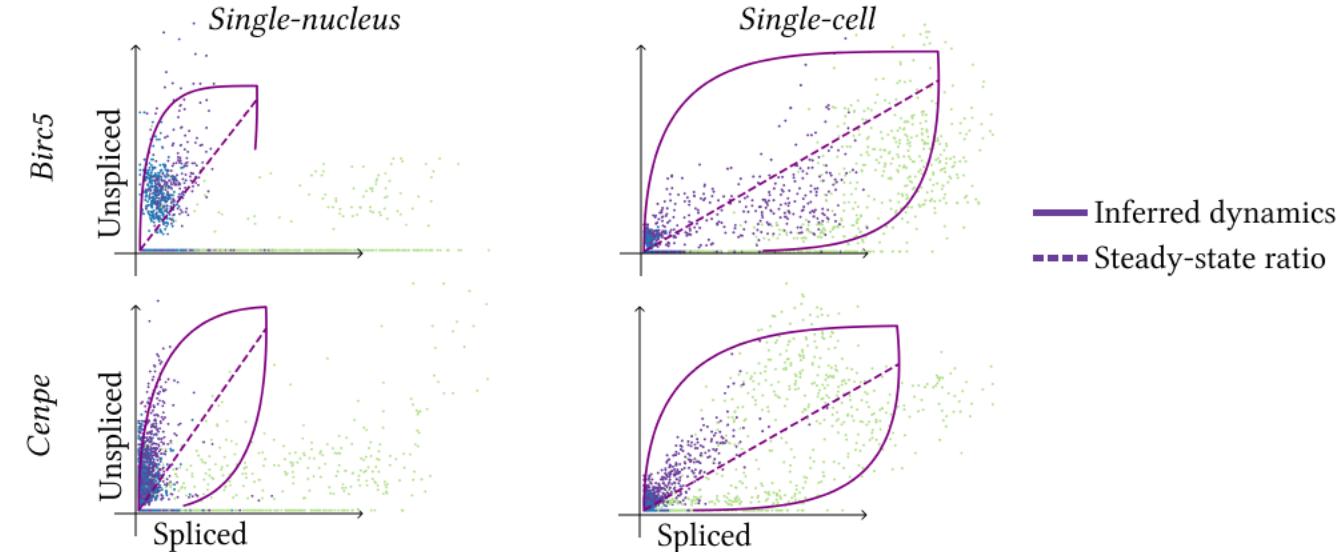


Cell cycle analysis

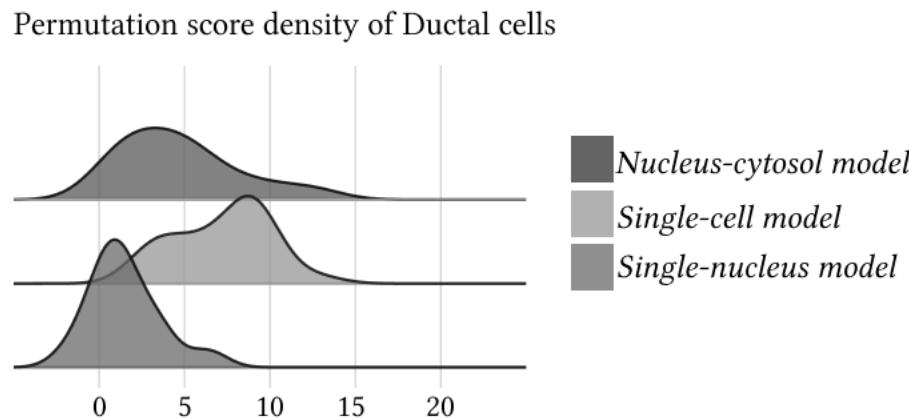
1. Cell cycle phase assignment



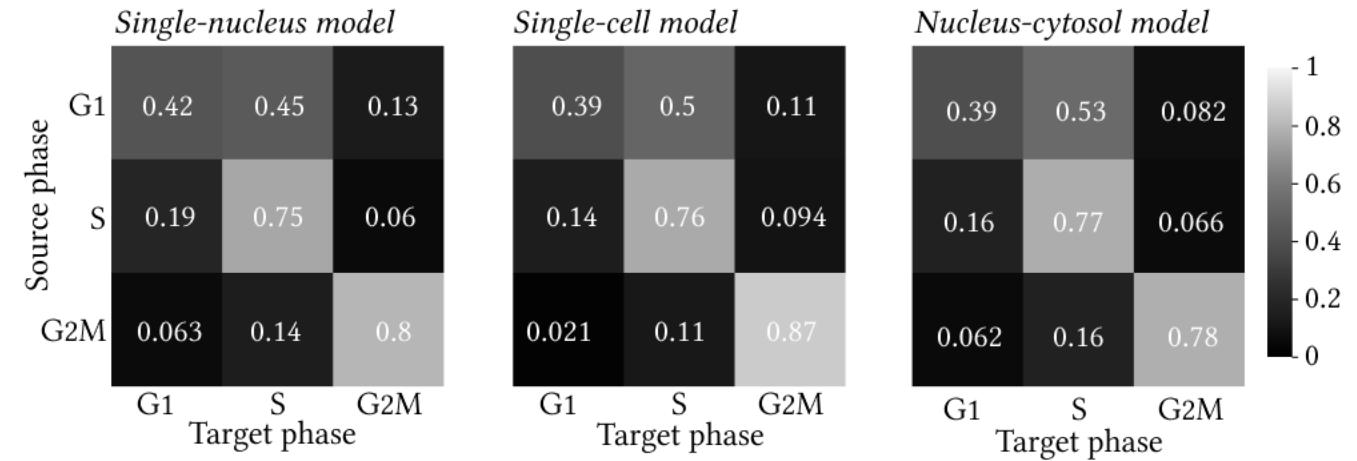
2. Investigate phase portraits



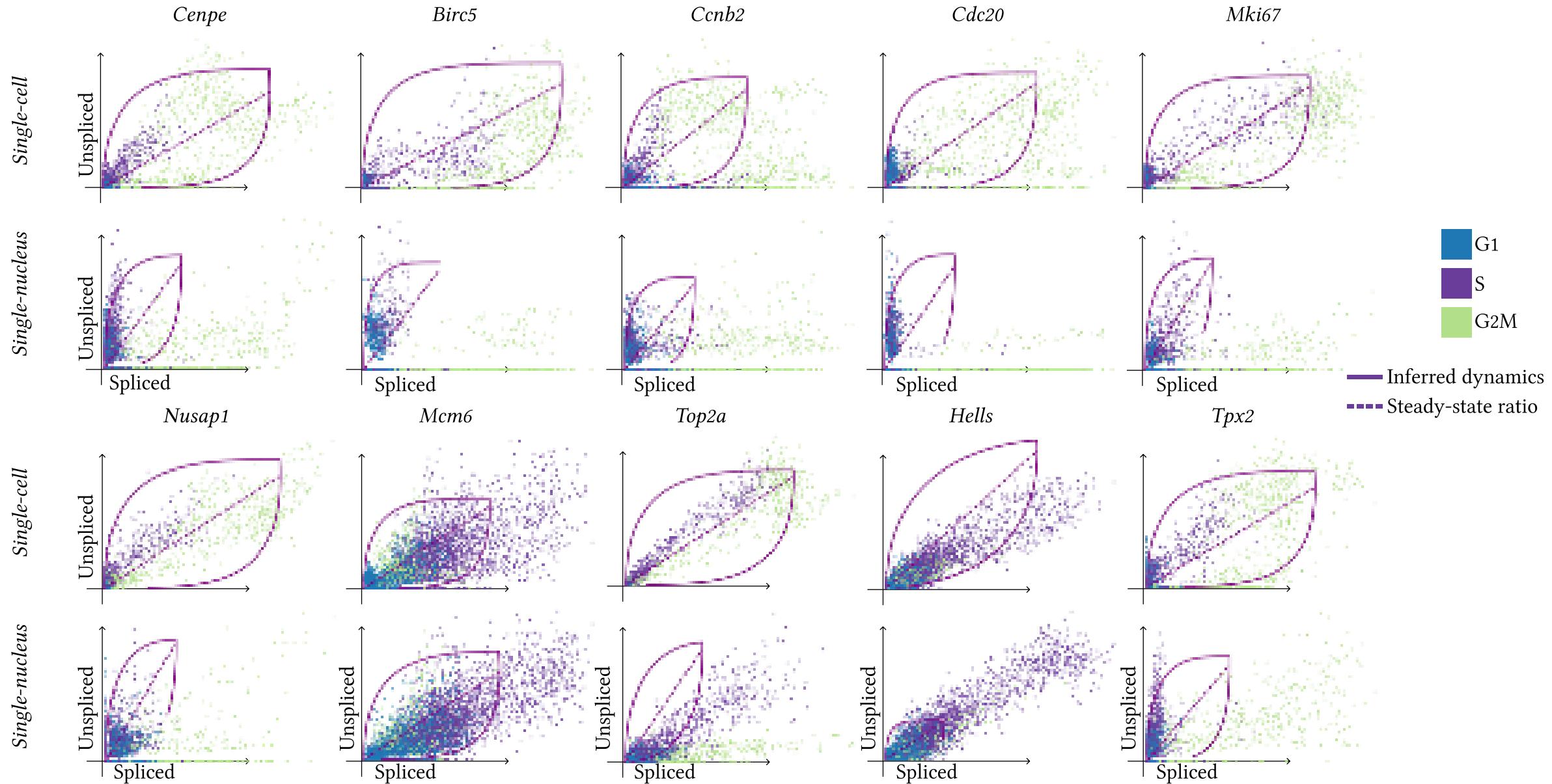
3. Compare permutation scores



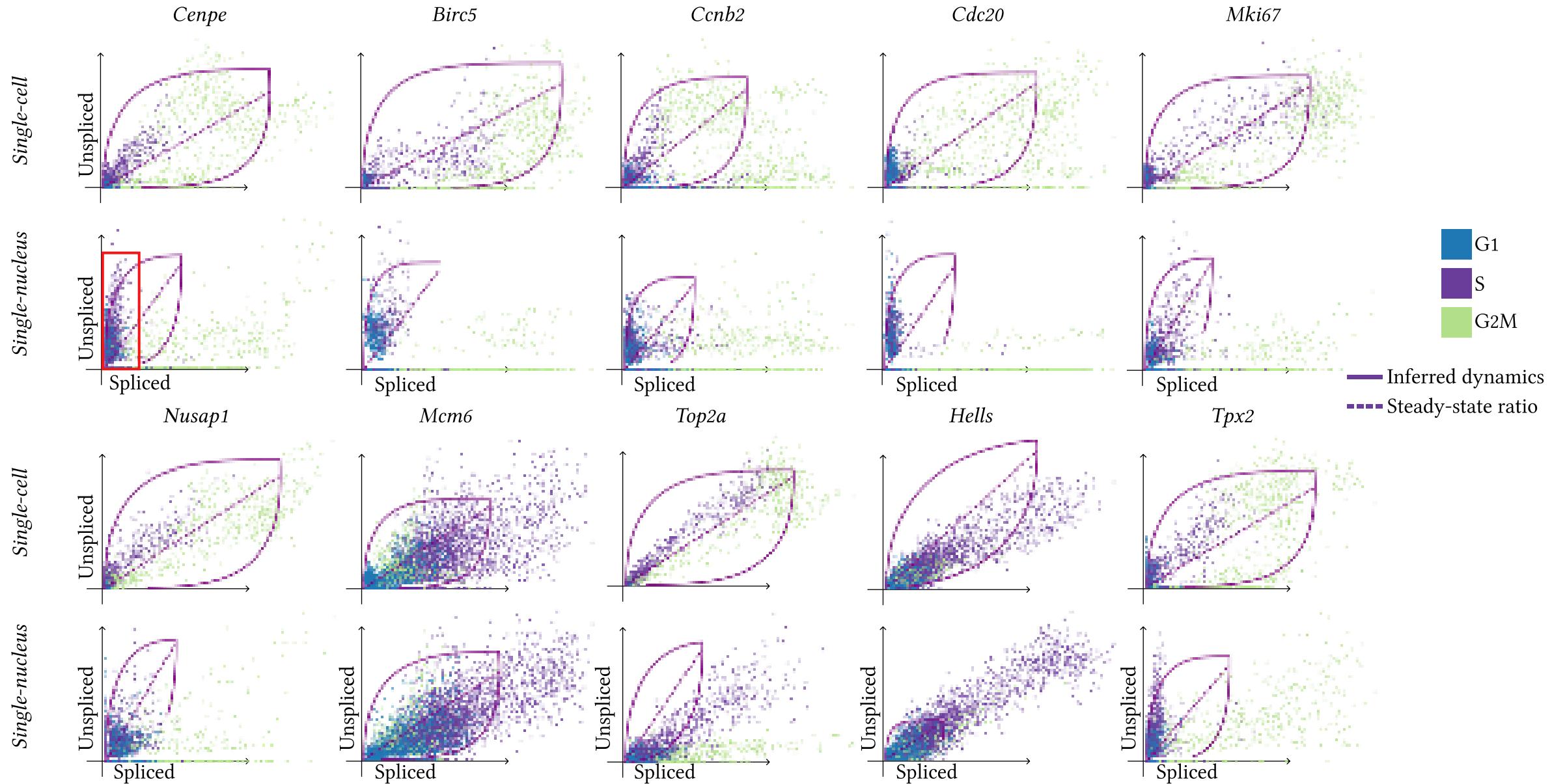
4. Cell cycle transition probabilities



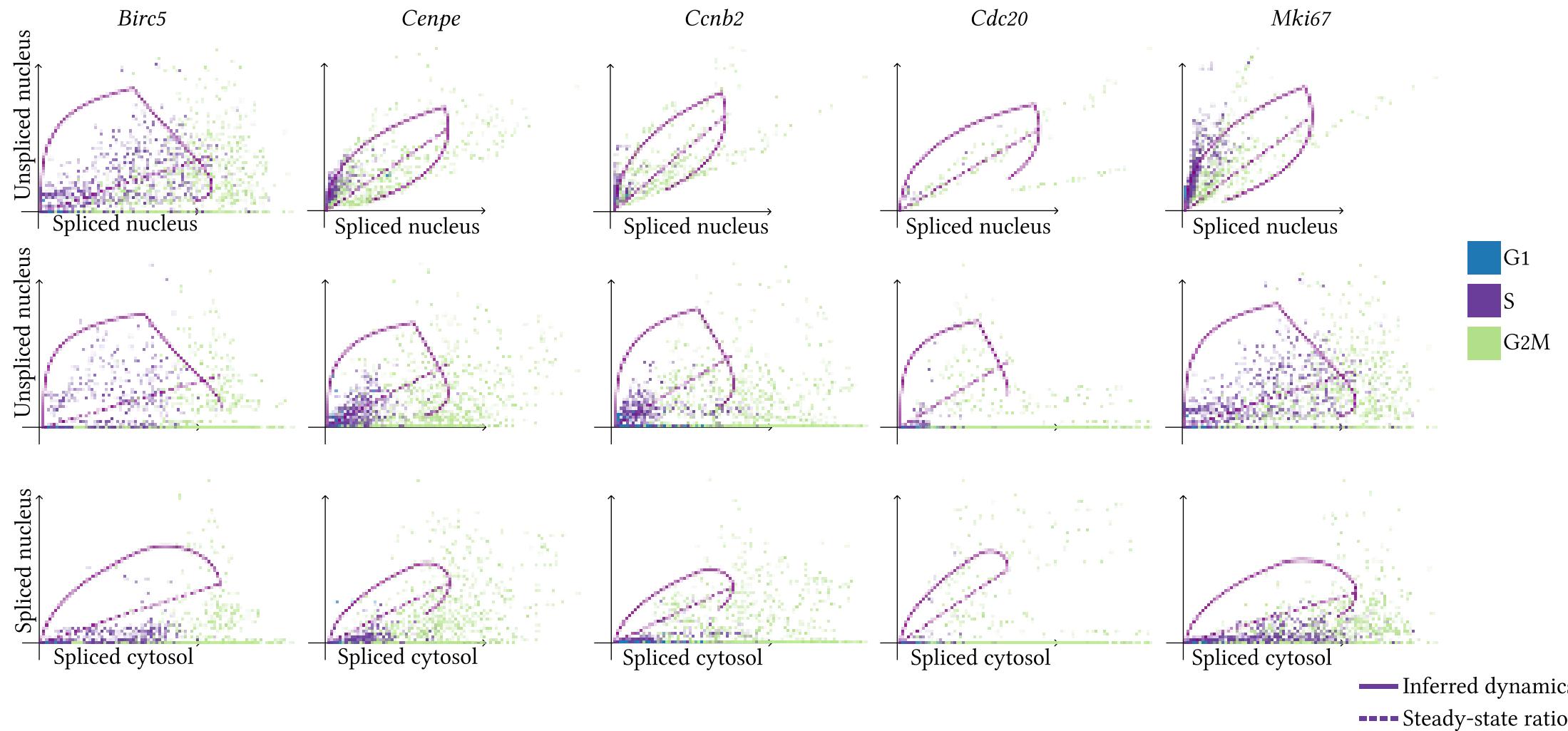
Phase portraits cycling genes



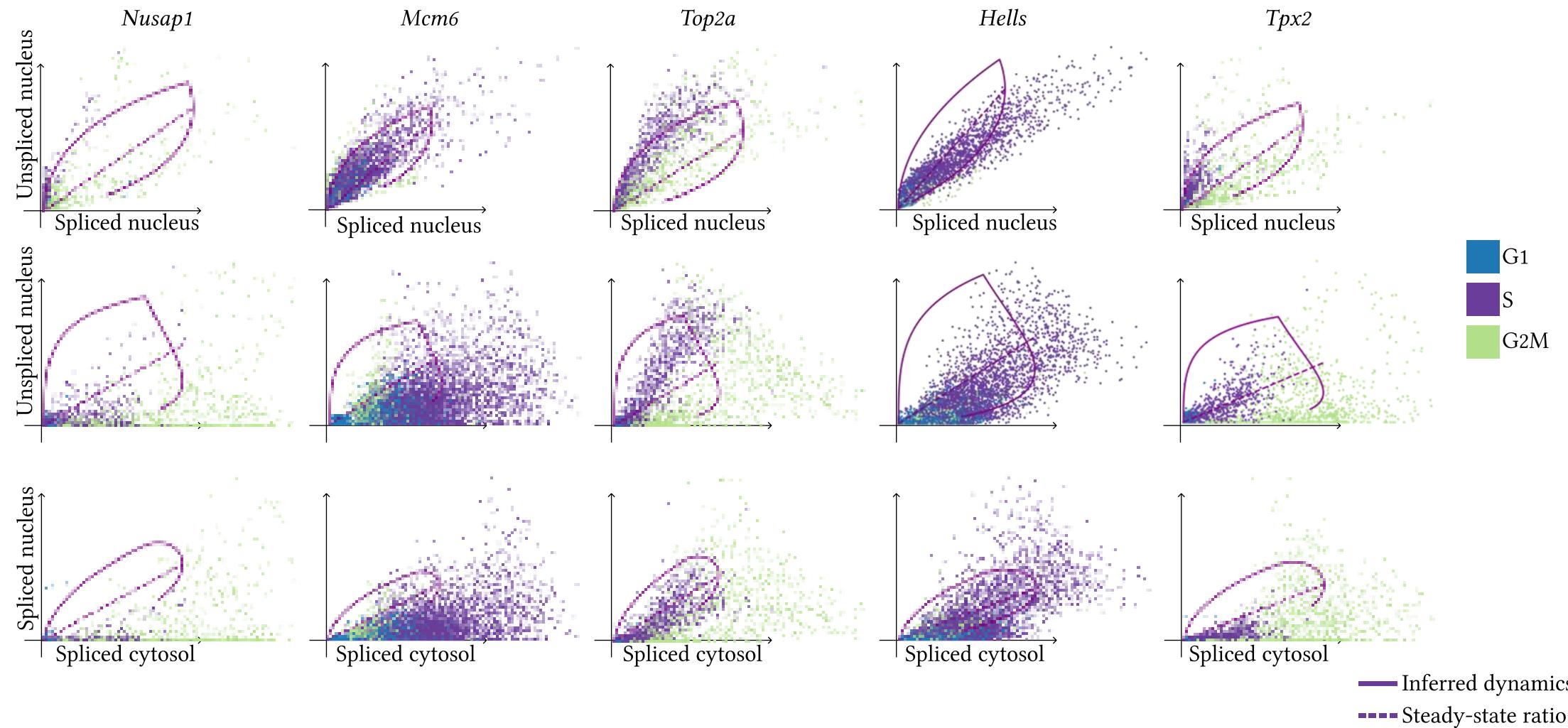
Phase portraits cycling genes



Phase portraits cycling genes - Nucleus cytosol model



Phase portraits cycling genes - Nucleus cytosol model

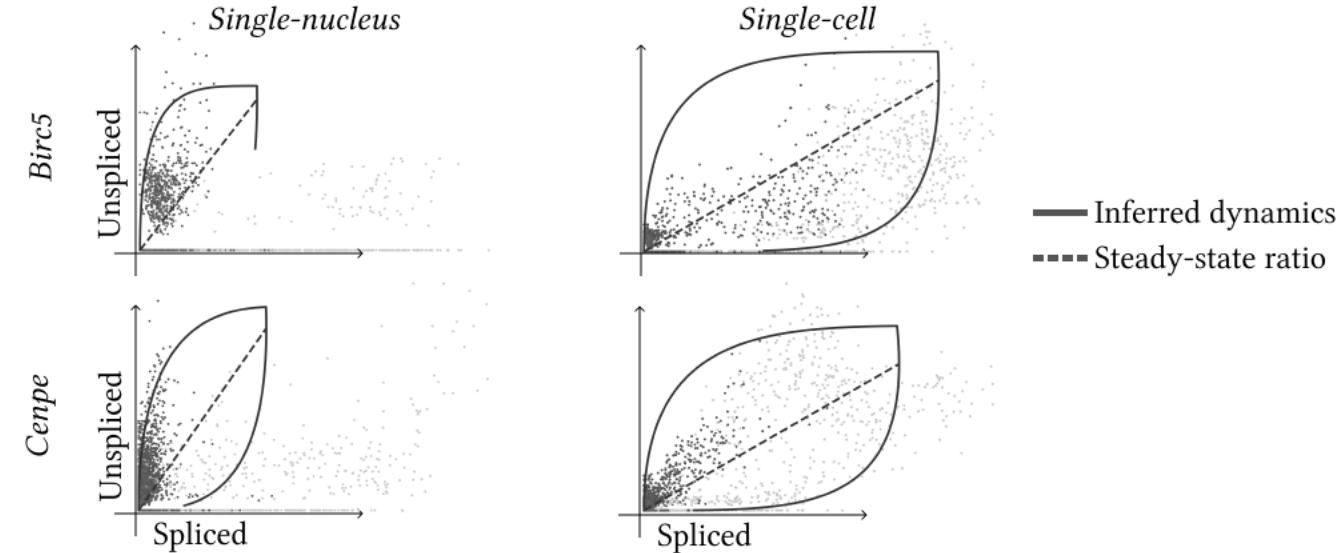


Cell cycle analysis

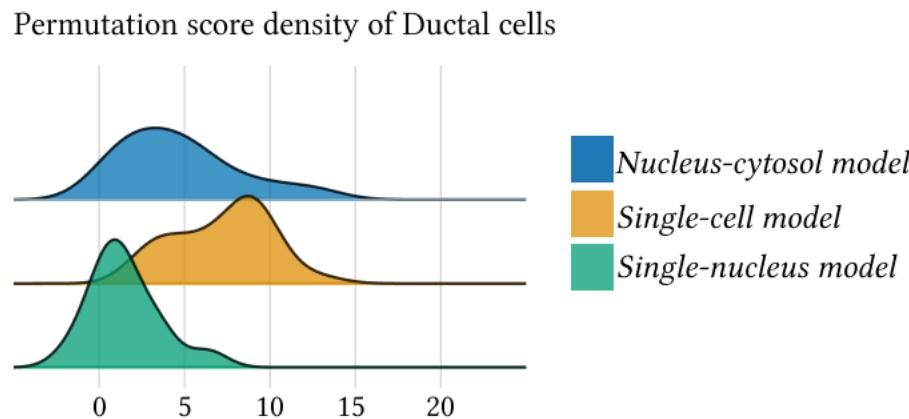
1. Cell cycle phase assignment



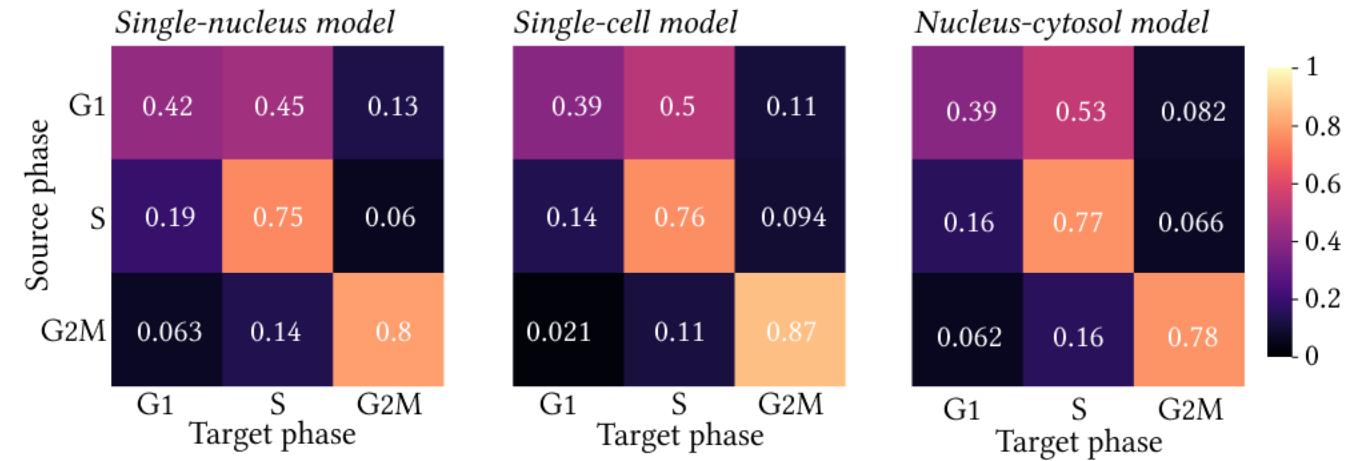
2. Investigate phase portraits



3. Compare permutation scores

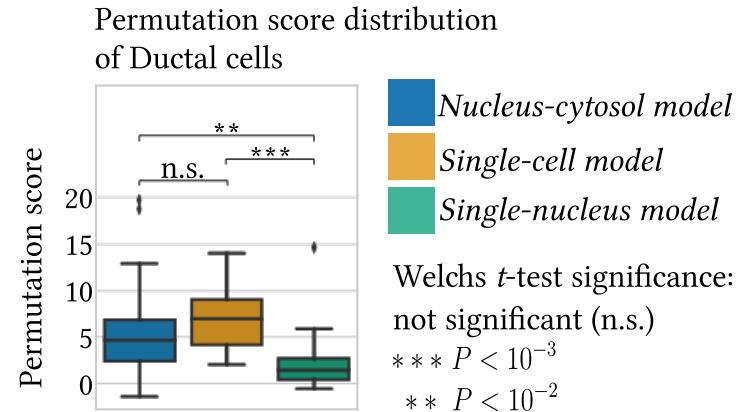
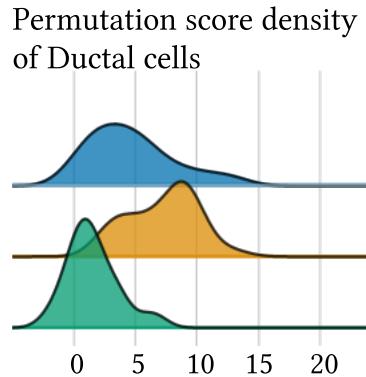


4. Cell cycle transition probabilities



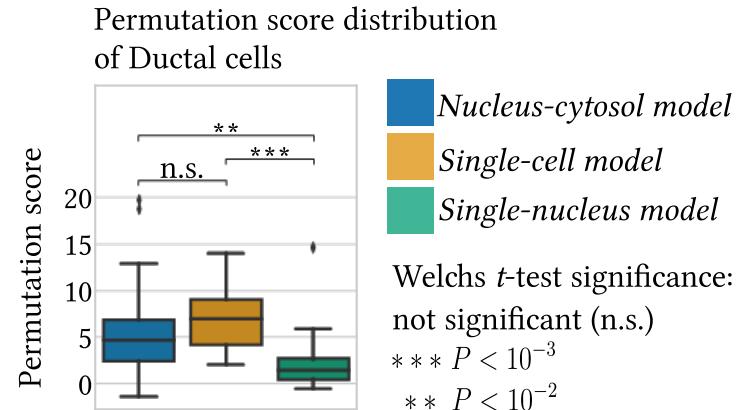
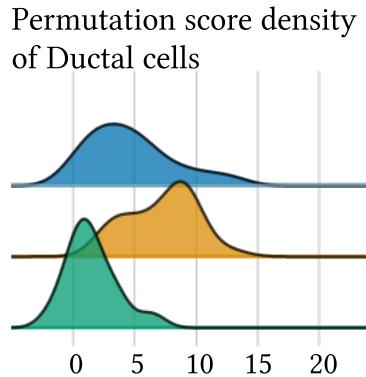
Cell cycle analysis - Permutation score and transition probabilities

a



Cell cycle analysis - Permutation score and transition probabilities

a



b

X = Source phase (S, G1, G2M)

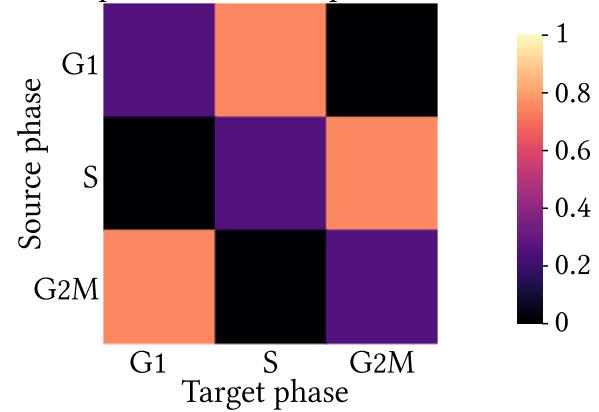
Y = Target phase (S, G1, G2M)

C_X, C_Y = Set of cells assigned to source and target phase resp.

$P(i \rightarrow j) = T_{ij}$ = Cell-cell transition matrix computed with *Cellrank*⁶

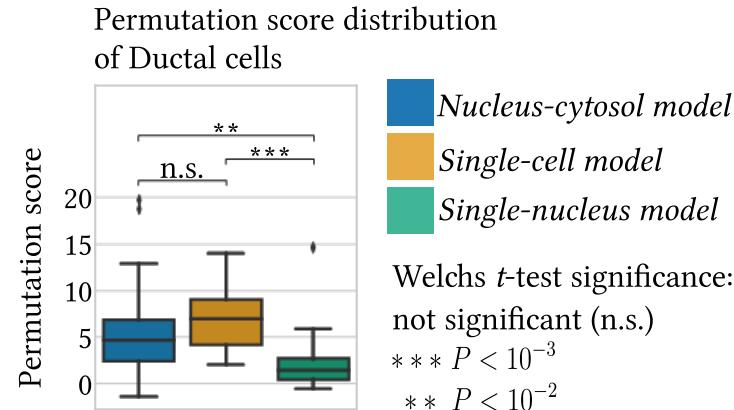
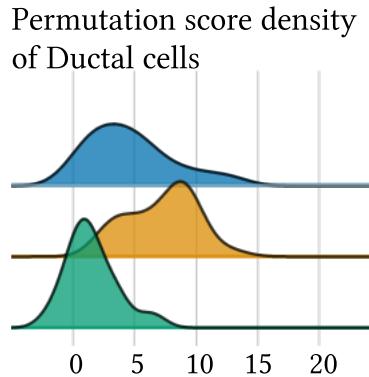
$$P(X \rightarrow Y) := \frac{1}{|C_X|} \sum_{i \in C_X} \sum_{j \in C_Y} P(i \rightarrow j)$$

Expected transition probabilities



Cell cycle analysis - Permutation score and transition probabilities

a



b

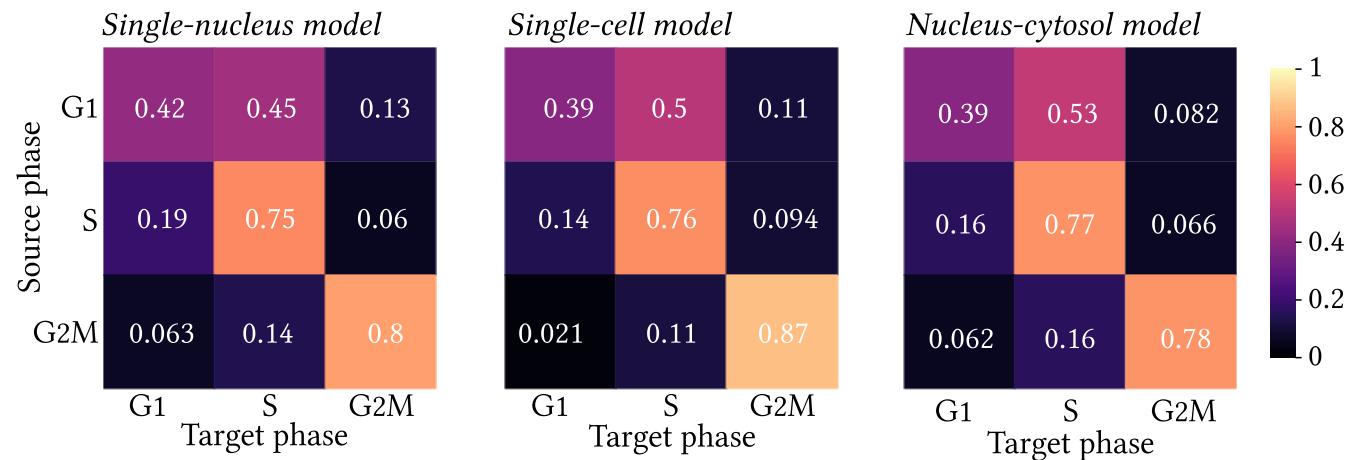
X = Source phase (S, G1, G2M)

Y = Target phase (S, G1, G2M)

C_X, C_Y = Set of cells assigned to source and target phase resp.

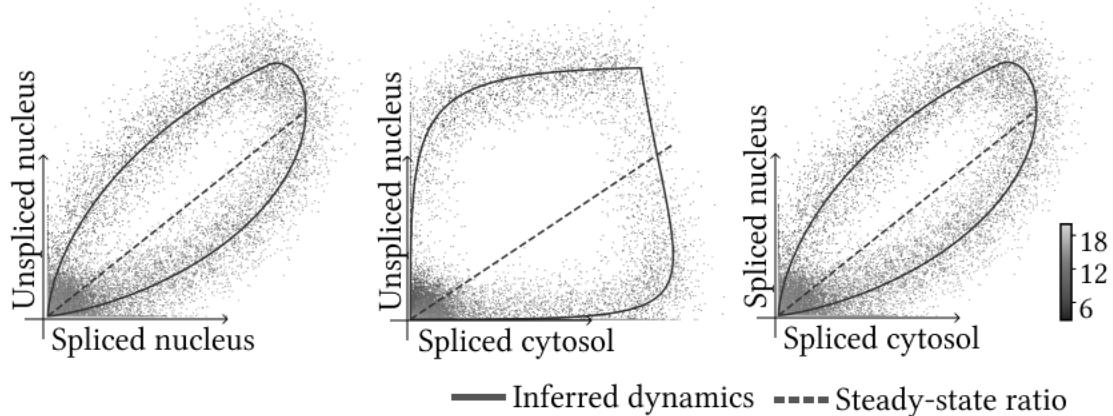
$P(i \rightarrow j) = T_{ij}$ = Cell-cell transition matrix computed with *Cellrank*⁶

$$P(X \rightarrow Y) := \frac{1}{|C_X|} \sum_{i \in C_X} \sum_{j \in C_Y} P(i \rightarrow j)$$

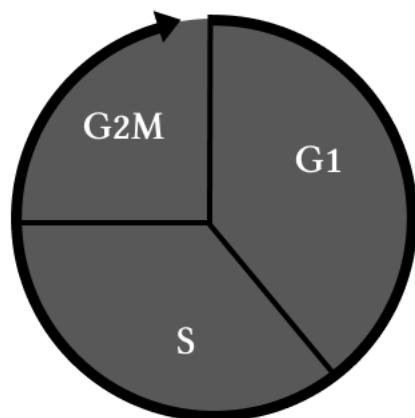


Results

1. Data simulation



3. Cell cycle analysis - Pancreas E14.5^{4,5}

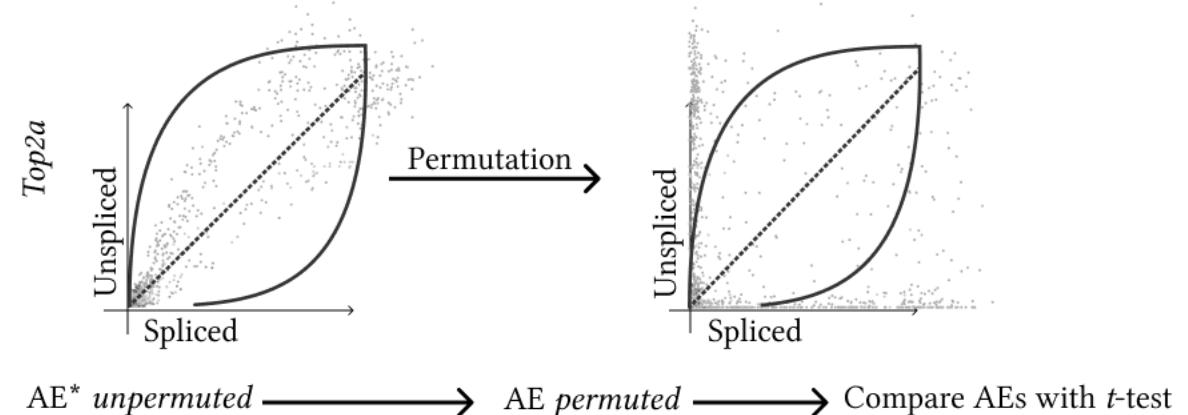


*AE = Absolute error

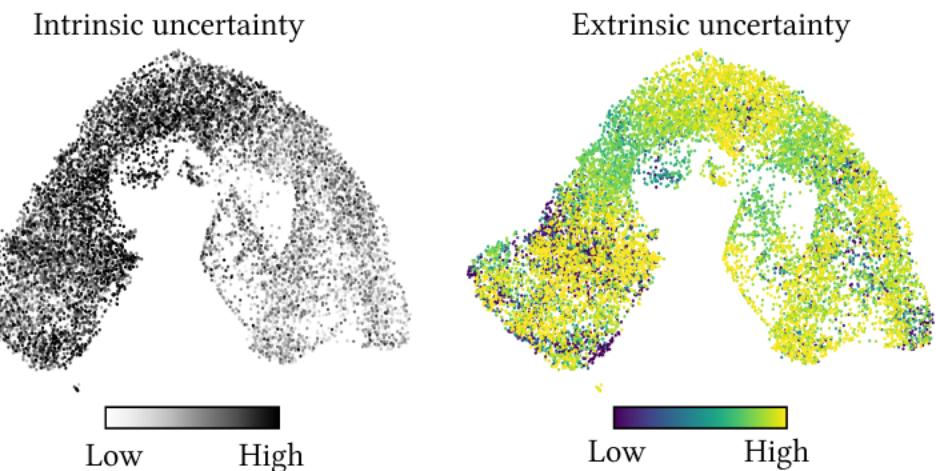
[4] A. Bastidas-Ponce et al. "Massive single-cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis", In: Development (2019).

[5] D. Klein et al. "Mapping cells through time and space with moscot", In: Development (2023)

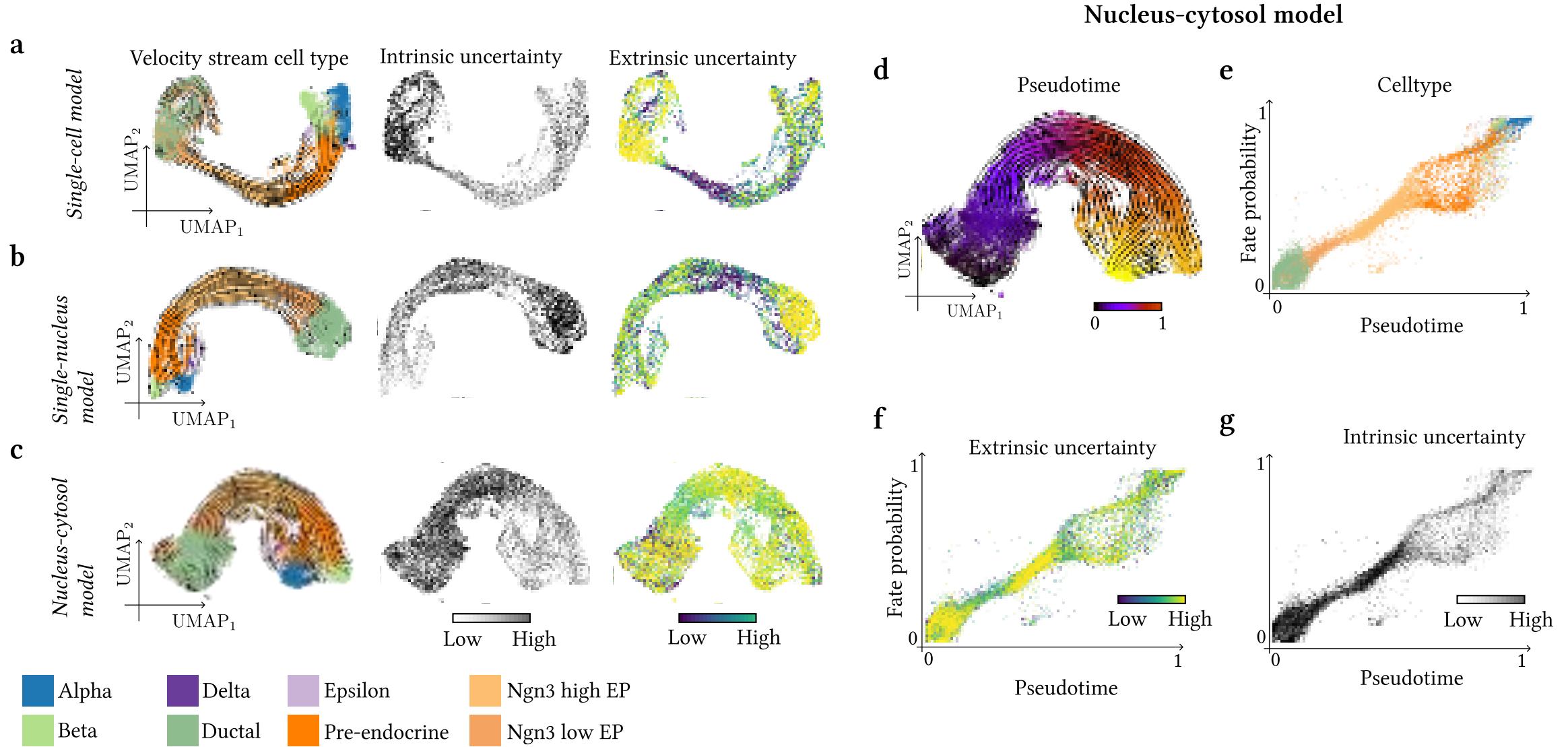
2. Permutation score analysis - Pancreas E14.5^{4,5}



4. Uncertainty analysis Pancreas E14.5^{4,5}



Uncertainty analysis



Conclusion

Benchmarking Single-cell vs. single-nucleus	Model extension
<ul style="list-style-type: none">• Smaller permutation scores• Cell cycle genes do not follow modelling assumptions for snRNA• Similar cell cycle transition probabilities	<ul style="list-style-type: none">• Accurately fits simulated data• Dependence on estimation process:<ul style="list-style-type: none">• Noisy phase portraits• Similar to single-cell model:<ul style="list-style-type: none">• Permutation scores• Cell cycle transitions• Uncertainty:<ul style="list-style-type: none">• Insightful information of cell's fate and future state

Conclusion

Benchmarking Single-cell vs. single-nucleus	Model extension
<ul style="list-style-type: none">• Smaller permutation scores• Cell cycle genes do not follow modelling assumptions for snRNA• Similar cell cycle transition probabilities	<ul style="list-style-type: none">• Accurately fits simulated data• Dependence on estimation process:<ul style="list-style-type: none">• Noisy phase portraits• Similar to single-cell model:<ul style="list-style-type: none">• Permutation scores• Cell cycle transitions• Uncertainty:<ul style="list-style-type: none">• Insightful information of cell's fate and future state



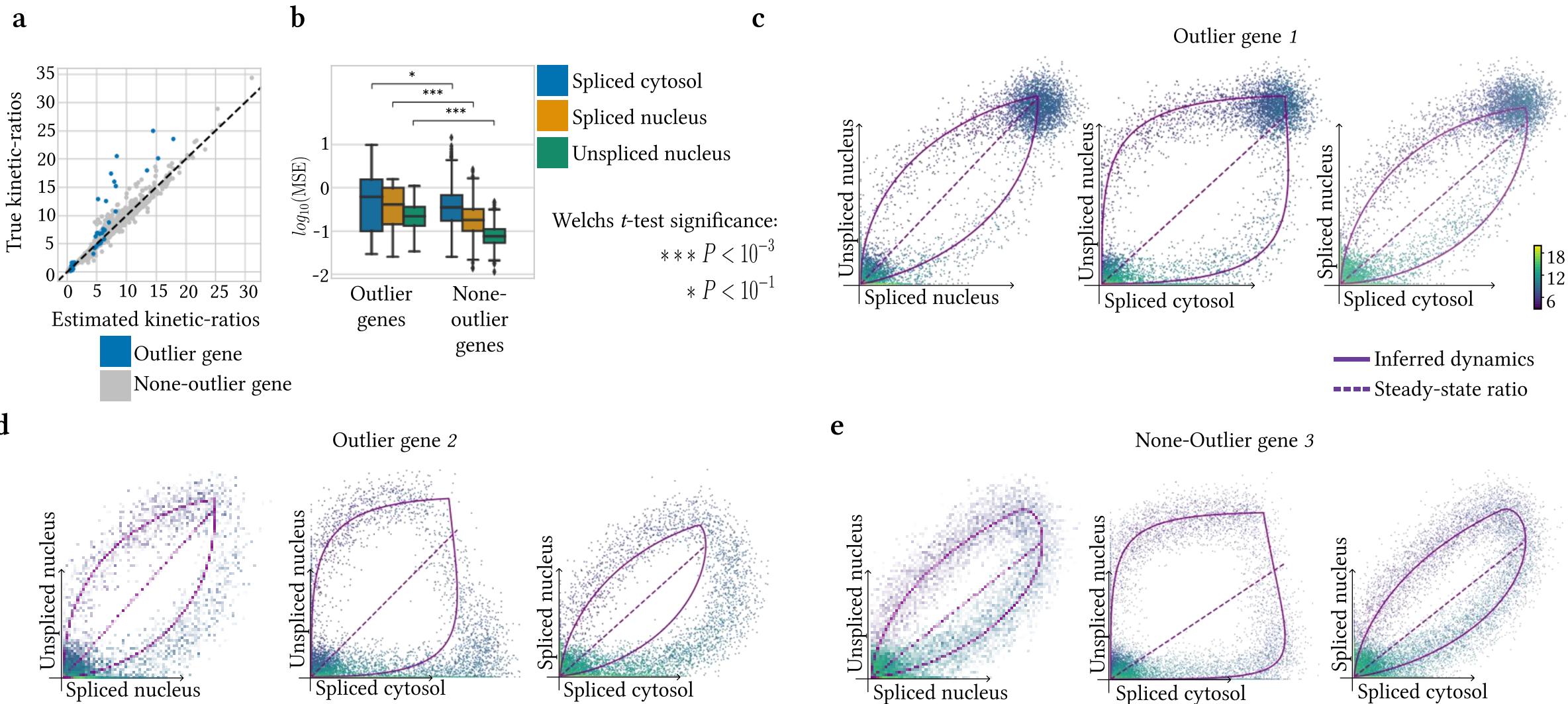
Future perspectives
<ul style="list-style-type: none">• Test model pipeline on additional datasets• Cell type specific transcription rates• Transcription rate based on open chromatin profiles⁷• Batch correction

[7] C. Li et al. "Single-cell multi-omic velocity infers dynamic and decoupled gene regulation". (2021).

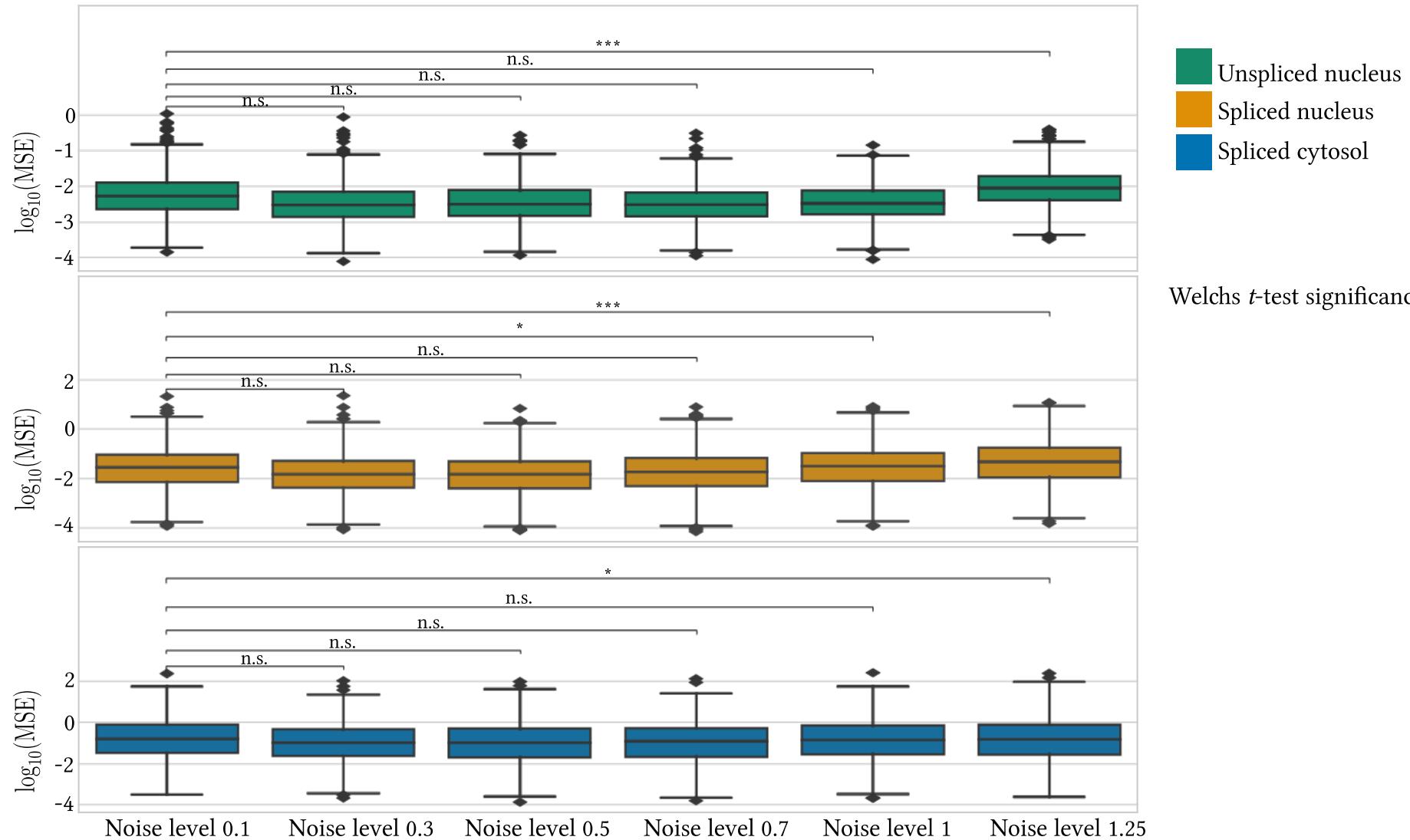
[8] D. Mahdessian et al. "Spatiotemporal dissection of the cell cycle with single-cell proteogenomics". Nature (2022)

Thanks for your attention! ☺

Appendix: Data simulation - Outlier genes



Appendix: Data simulation - Effect of introduced noise (MSE)

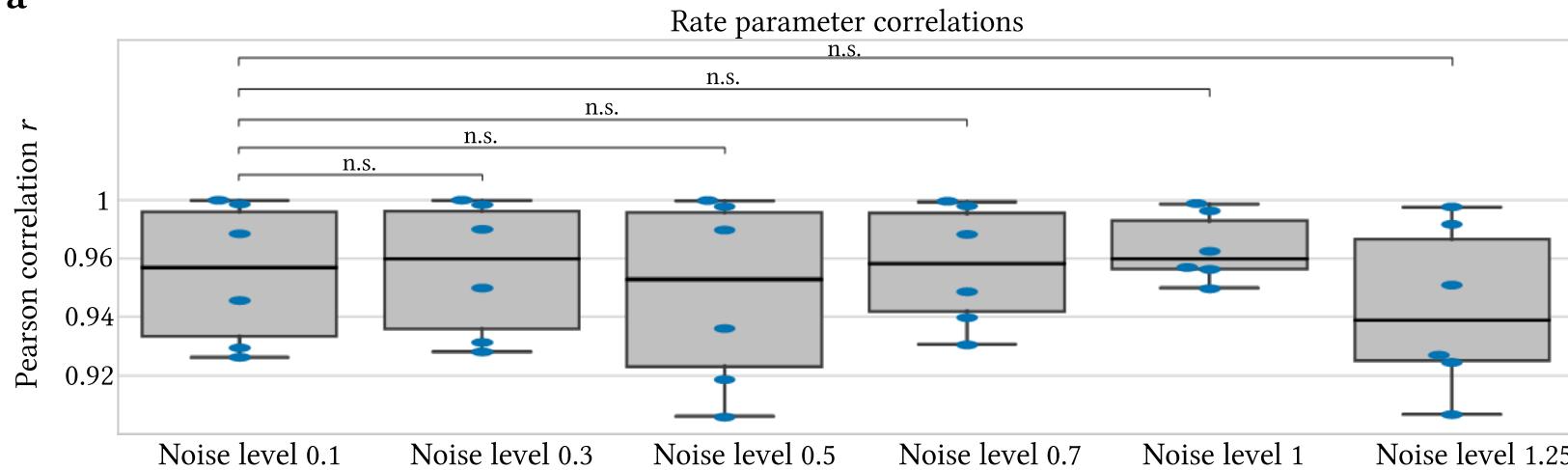


Welches *t*-test significance: not significant (n.s.)

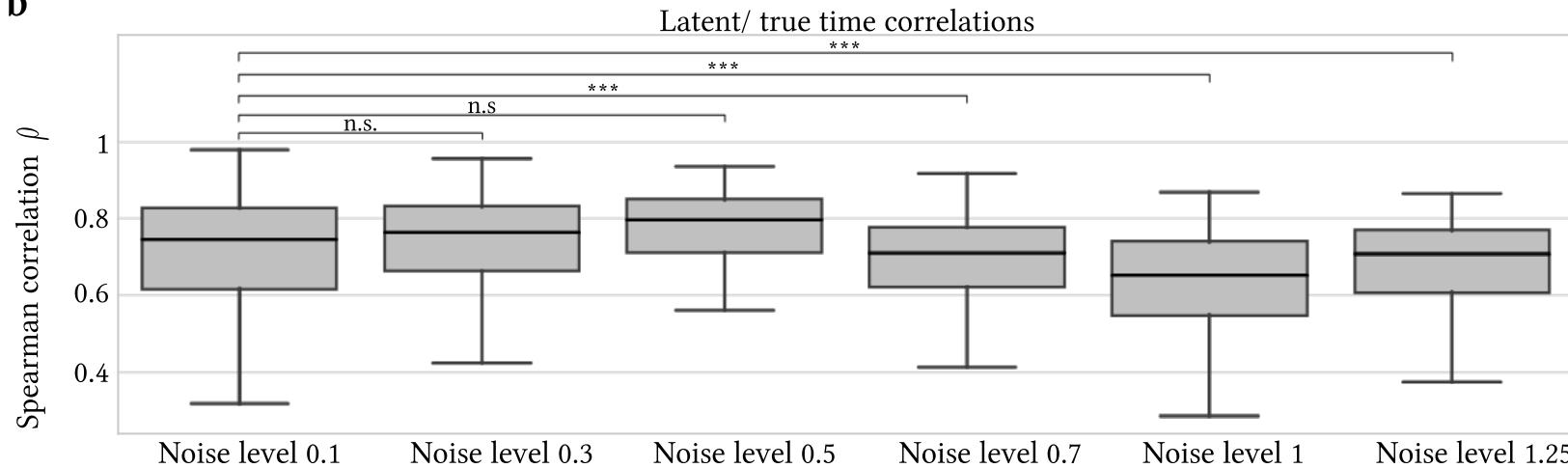
*** $P < 10^{-3}$
** $P < 10^{-2}$
* $P < 10^{-1}$

Appendix: Data simulation - Effect of introduced noise (latent variables)

a

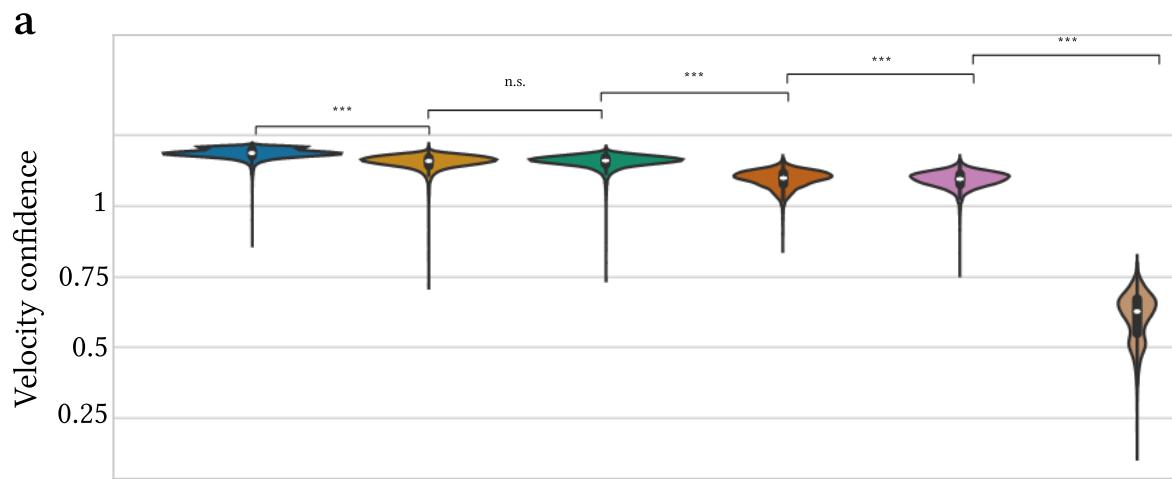


b



Appendix: Velocity confidence

Single-cell against single-nucleus neighbors velocity confidence: $c_i = \frac{1}{k} \sum_{j \in \mathcal{N}_{\text{scRNA}}(i)}^k \text{corr}(v_i, v_j)$ i snRNA-seq

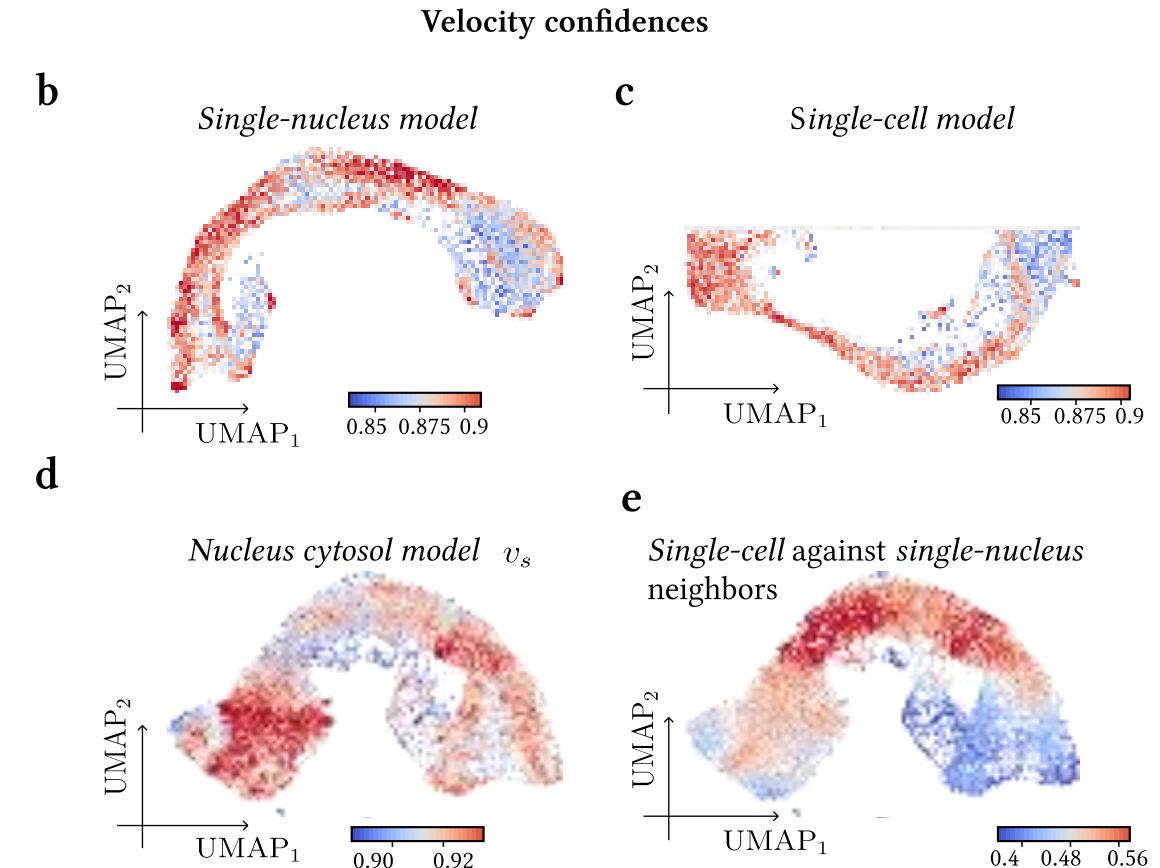


Welches t -test significance: not significant (n.s.)

*** $P < 10^{-3}$

** $P < 10^{-2}$

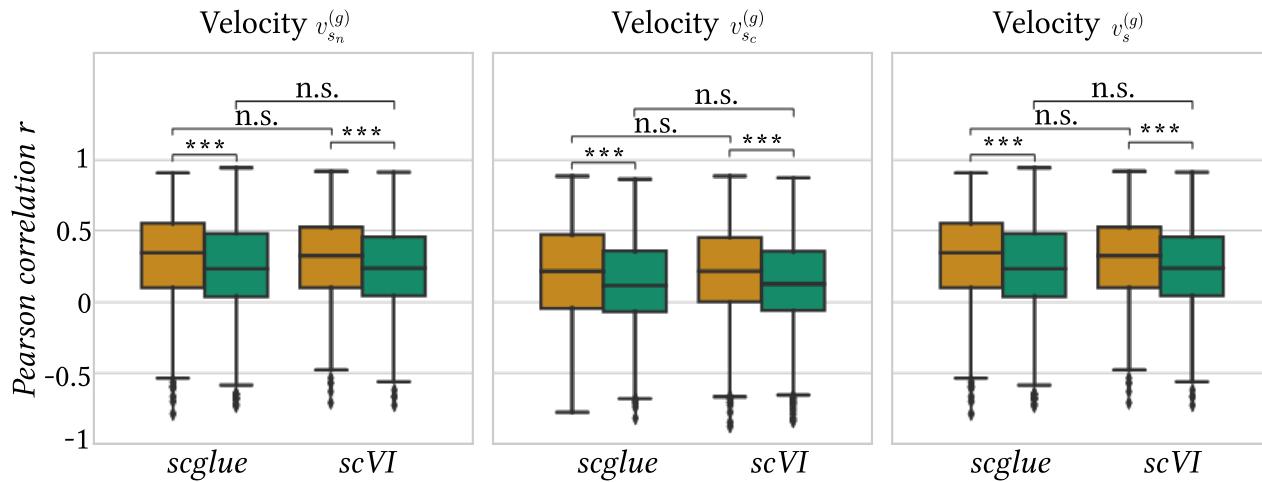
* $P < 10^{-1}$



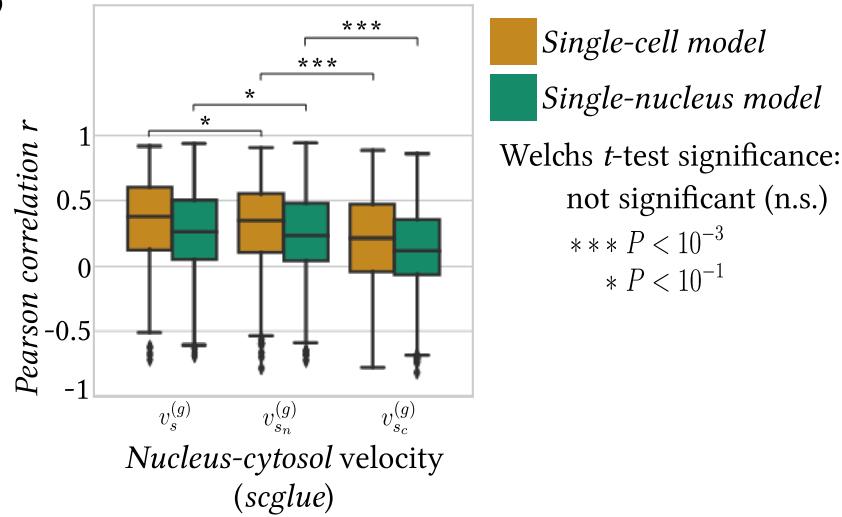
Appendix: Velocity comparison

a

Velocity correlations *Nucleus-cytosol model/ single-modal models*

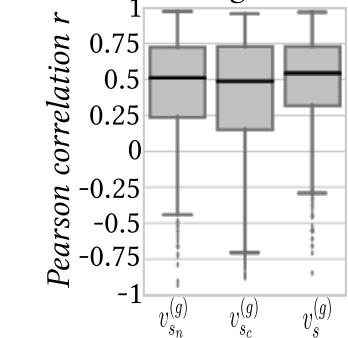


b

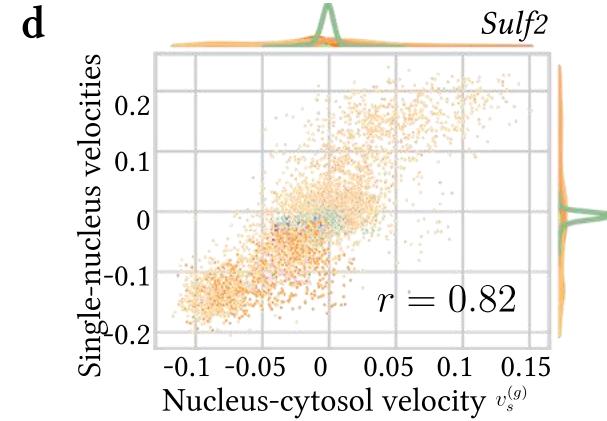


c

Velocity correlations between *scVI* and *scglue*



d



e

