Logic programs to infer computational models of human embryonic development

Programmes logiques pour déduire des modèles informatiques du développement embryonnaire humain

Mathieu Bolteau

Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France

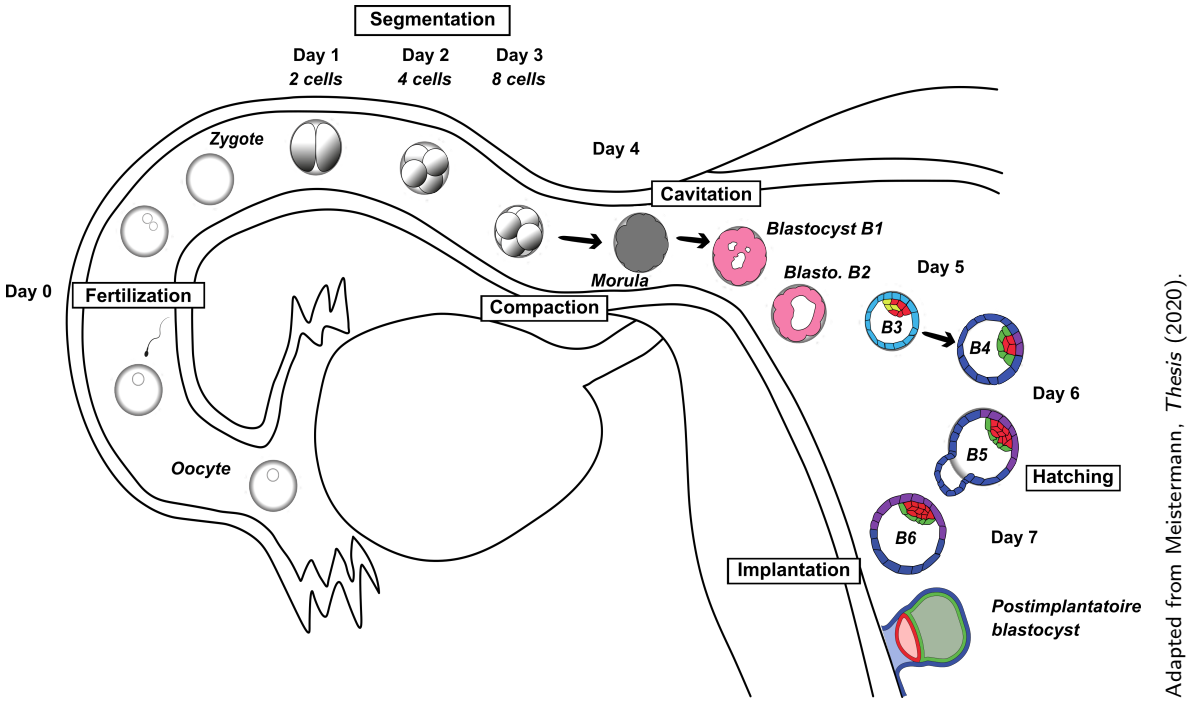
Ph.D. Defense – Friday, October 4th 2024

Referees & Examiners:	Loïc Paulevé	Directeur de recherche, CNRS, Bordeaux, France
	Delphine Ropers	Directrice de recherche, Inria, Grenoble, France
Examiners:	Clémence Frioux	Chargée de recherche, Inria, Bordeaux, France
	Cédric Lhoussaine	Professeur, Université de Lille, Lille, France
	Richard Redon	Directeur de recherche, Inserm, Nantes, France
Ph.D. Director:	Jérémie Bourdon	Professeur, Nantes Université, Nantes, France
Ph.D. Advisor:	Carito Guziolowski	Maîtresse de conférence, École Centrale de Nantes, Nantes, France
Invited Member:	Laurent David	Maître de conférence-Praticien hospitalier, Nantes Université, Nantes, France

LS2N
Laboratoire des Sciences du Numérique de Nantes

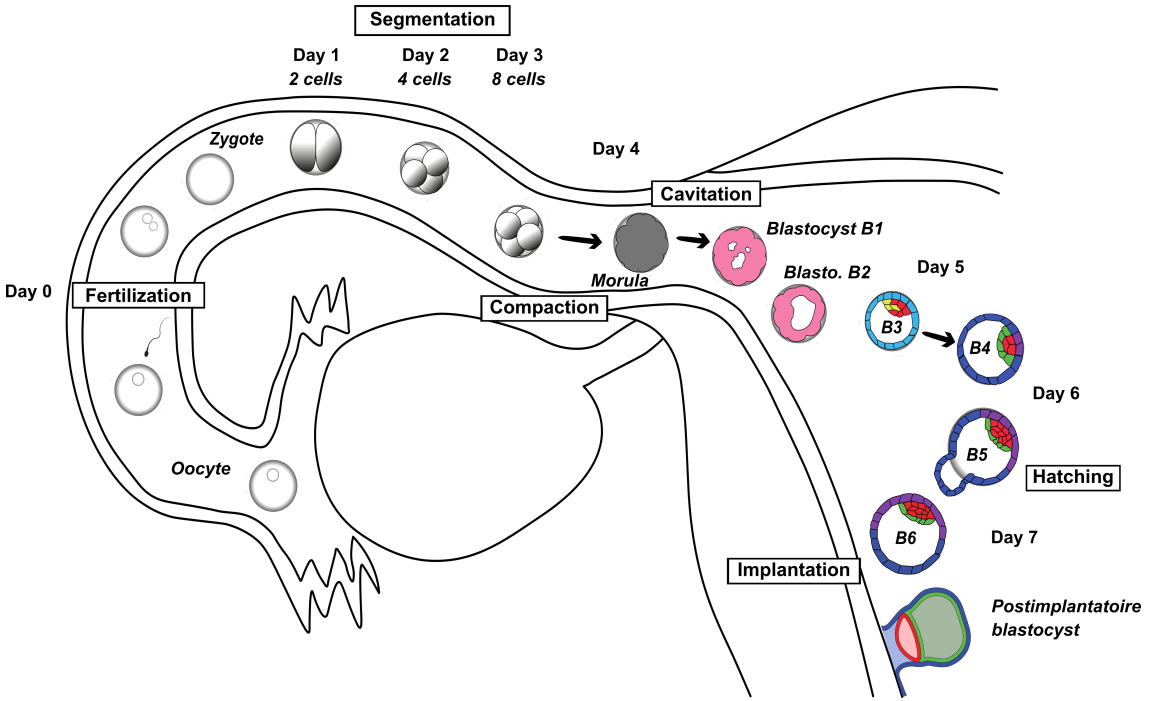
Nantes Université

Human preimplantation embryonic development



Adapted from Meistermann, *Thesis* (2020).

Human preimplantation embryonic development



[Meistermann et al., *Cell Stem Cell* (2021)]

Context

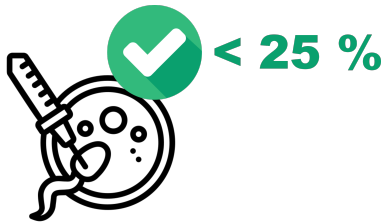
Study human embryo is complex

- Biological mechanisms
- Legal constraints
- Experimental concerns

Context

Study human embryo is complex

- Biological mechanisms
- Legal constraints
- Experimental concerns



Poor success of IVF

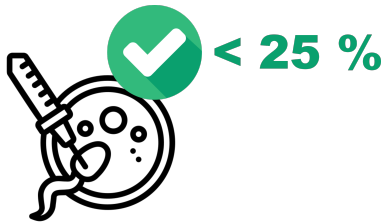
[De Geyter et al., *Human Reproduction* (2024)]

IVF = in vitro fertilization

Context

Study human embryo is complex

- Biological mechanisms
- Legal constraints
- Experimental concerns



Poor success of IVF

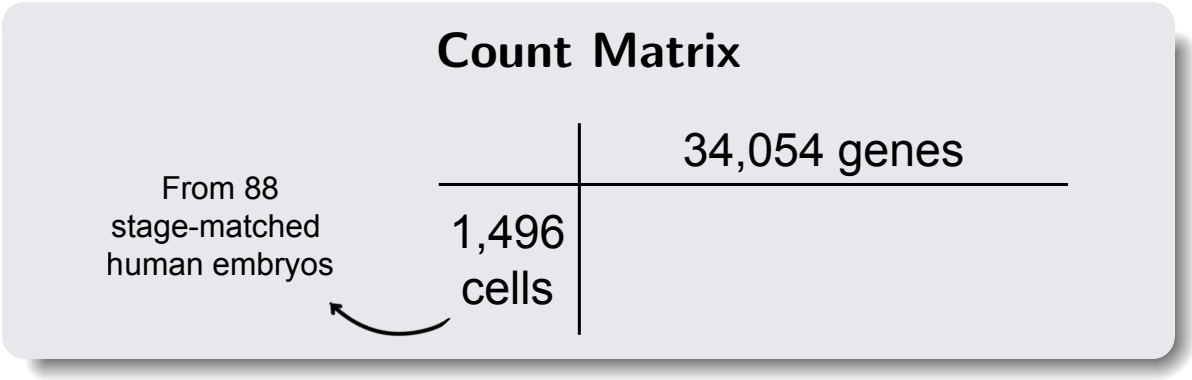
[De Geyter et al., *Human Reproduction* (2024)]

→ **In silico model of human embryonic development**

IVF = in vitro fertilization

Data and preliminary study

[Meistermann et al., *Cell Stem Cell* (2021)]



Data and preliminary study

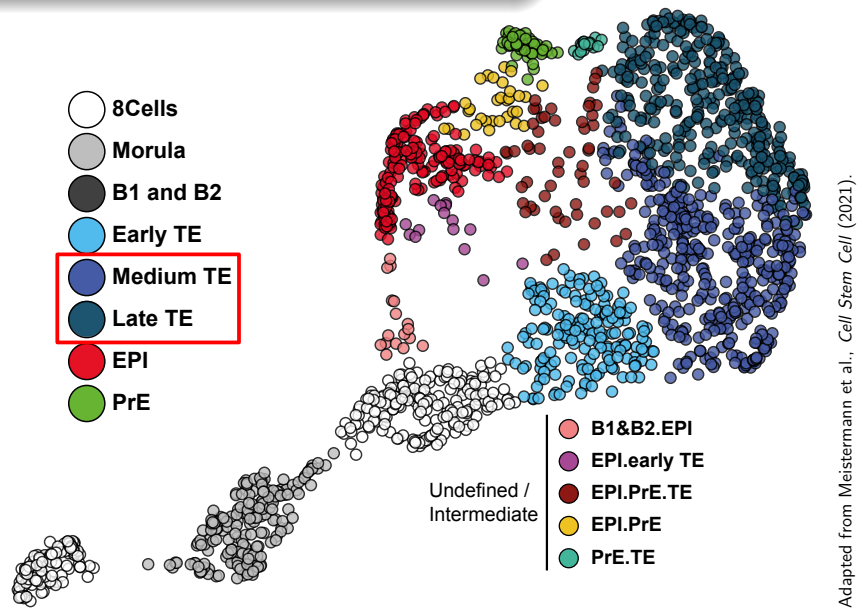
[Meistermann et al., *Cell Stem Cell* (2021)]

Count Matrix

	34,054 genes
1,496 cells	

From 88 stage-matched human embryos

- **Clustering** of cells (8 stages)



Data and preliminary study

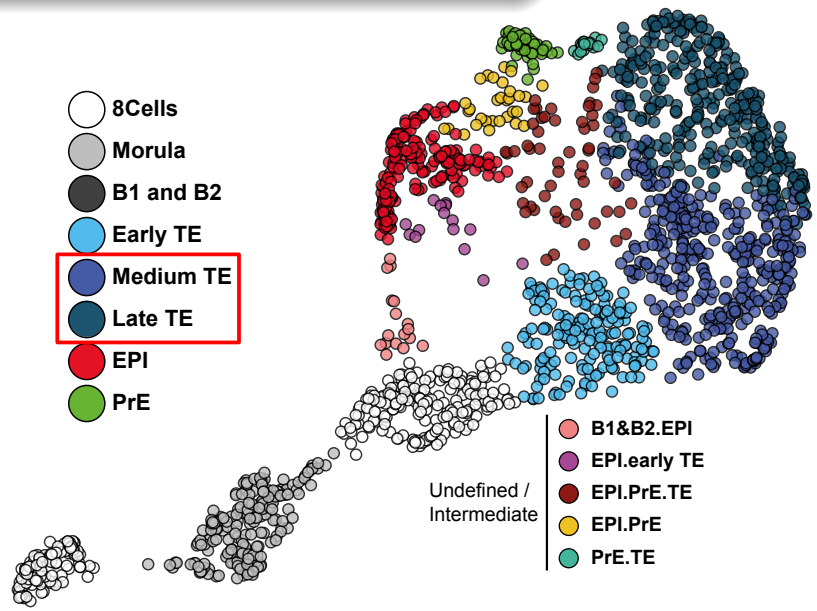
[Meistermann et al., *Cell Stem Cell* (2021)]

Count Matrix

	34,054 genes
1,496 cells	

From 88 stage-matched human embryos

- **Clustering** of cells (8 stages)
- Identification of **gene modules**
  $\rightarrow$ 438 transcription factors



Adapted from Meistermann et al., *Cell Stem Cell* (2021).

Data and preliminary study

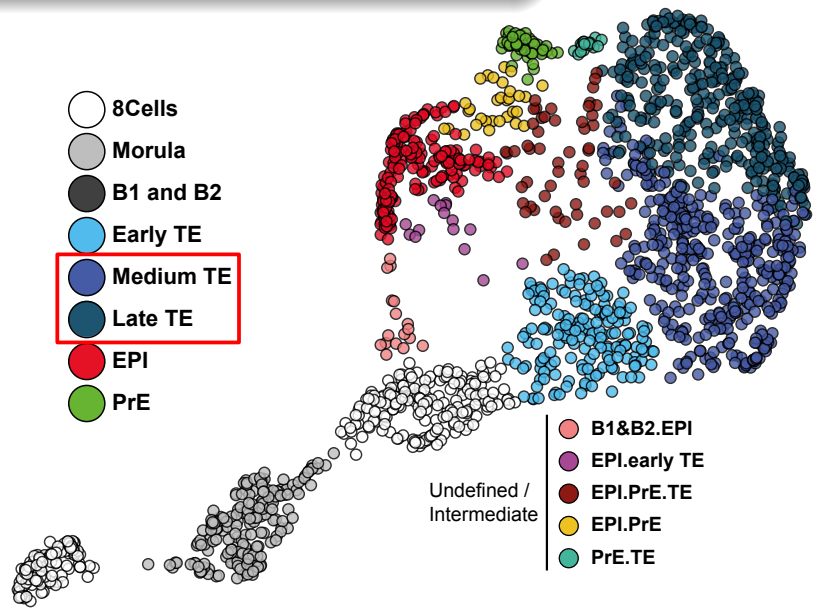
[Meistermann et al., *Cell Stem Cell* (2021)]

Count Matrix

	34,054 genes
1,496 cells	

From 88 stage-matched human embryos

- **Clustering** of cells (8 stages)
- Identification of **gene modules** $\rightarrow$ 438 transcription factors
- **Pseudotime** evolution of cells at different developmental stages



Adapted from Meistermann et al., *Cell Stem Cell* (2021).

Modeling single-cell data: existing tools

SCNS		[Moignard et al., *Nature biotech.* (2015)]
Gene expression state changes	Cell ordering	State transition graph: high combinatory (require small number of studied genes)

BoNesis		[Chevalier et al., *ICTAI* (2019)]
Prior gene interactions (Dorothea database)	Cell ordering and dynamical constraints	Mean of cells and gene expression

RE:IN		[Dunn et al., *EMBO journal* (2019)]
Prior gene interactions (gene expression correlation)	Perturbations experiments	Biological system allowing perturbations / Limited number of perturbations

State-of-the-art tool review

Method	System size	Cell heterogeneity	Cellular dynamic evolution	Exhaustive enumeration	Validation
SCNS	• $\approx$ 40 genes • $\approx 4,000$ cells	✓	✓	✓	✓ *
BoNesis	• $\approx 1,000$ genes • $\approx$ 600 cells	✗	✓	✗	✗
RE:IN	• $\approx$ 20 genes • $\approx$ 30 perturbations	✗	✗	✗	✓

* Thanks to experimental perturbations

State-of-the-art tool review

Method	System size	Cell heterogeneity	Cellular dynamic evolution	Exhaustive enumeration	Validation
SCNS	• $\approx$ 40 genes • $\approx 4,000$ cells	✓	✓	✓	✓ *
BoNesis	• $\approx 1,000$ genes • $\approx$ 600 cells	✗	✓	✗	✗
RE:IN	• $\approx$ 20 genes • $\approx$ 30 perturbations	✗	✗	✗	✓
Our goal	• $\approx$ **150 genes** • $\approx$ **700 cells**	✓	✓	✓	✓

* Thanks to experimental perturbations

Thesis Objective

Develop a new method to model regulatory mechanisms occurring in developmental stages

- Cell heterogeneity
- Cellular dynamic evolution
- Exhaustive enumeration
- Validation

Thesis Objective

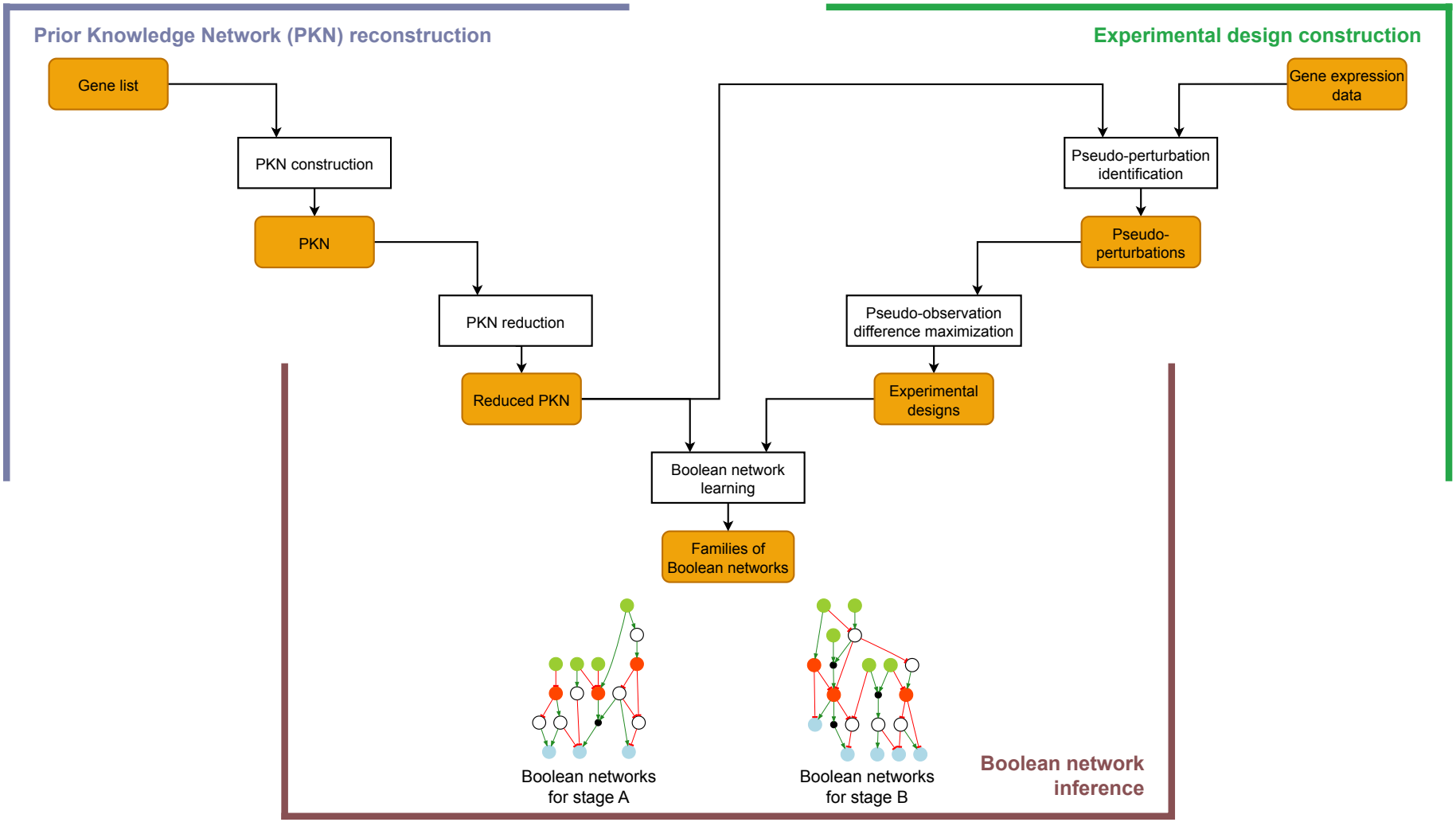
Develop a new method to model regulatory mechanisms occurring in developmental stages

- Cell heterogeneity
- Cellular dynamic evolution
- Exhaustive enumeration
- Validation

SCIBORG

Using single-cell data to infer Boolean networks modeling regulation of genes

Overview of SCIBORG

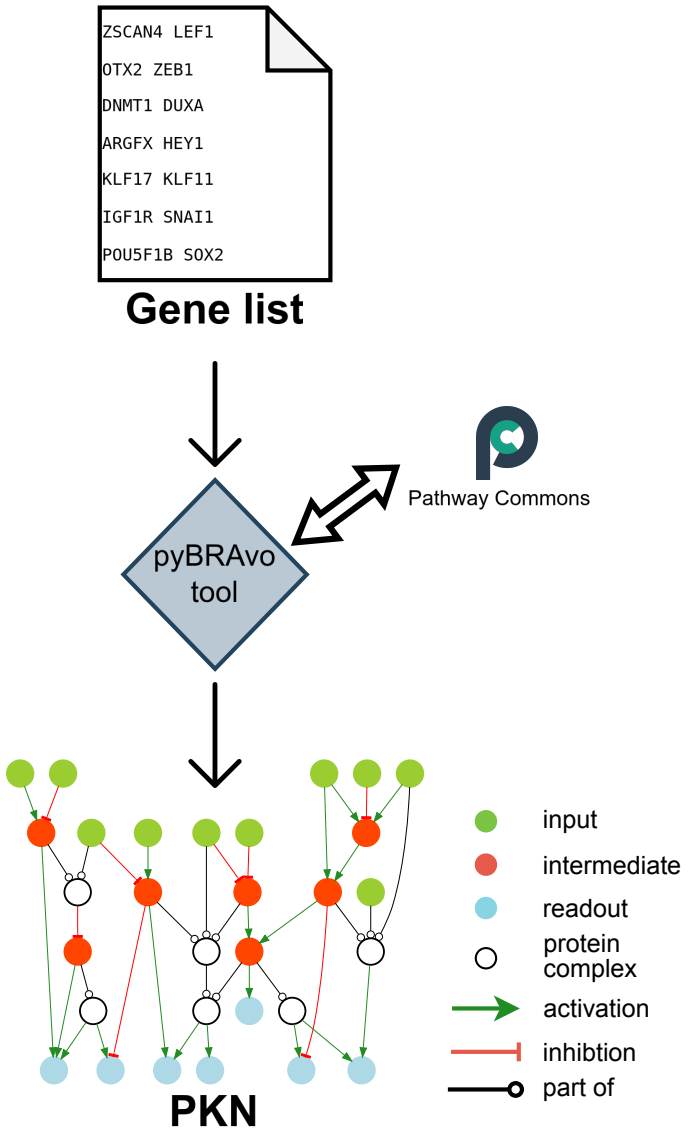


[Bolteau et al., *ISBRA* (2023)] [Bolteau et al., *J. of Computational Biology* (2024)]

Step 1. PKN reconstruction

Recursive queries on Pathway Commons database, via pyBRAvo [Lefebvre et al., *Database* (2021)]

Input	Output	Parameter
Gene list	PKN	Depth



PKN = prior knowledge network

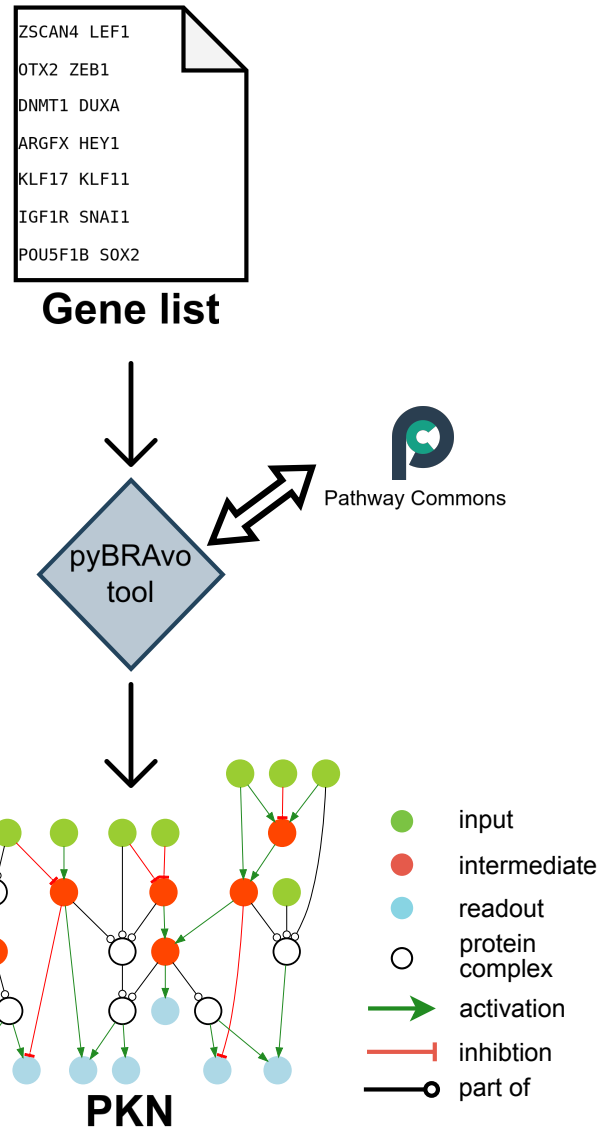
Step 1. PKN reconstruction

Recursive queries on Pathway Commons database, via pyBRAvo [Lefebvre et al., *Database* (2021)]

Input	Output	Parameter
Gene list	PKN	Depth

input & intermediate genes → **possible pseudo-perturbations**

readout genes → **pseudo-observations**



PKN = prior knowledge network

Step 2. Experimental design construction

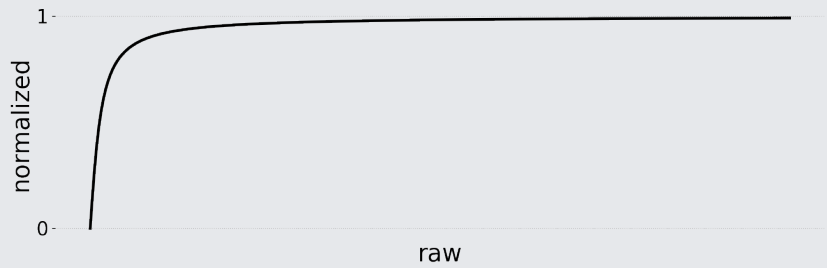
Data preprocessing

- Binarization of input + intermediate genes

$$binarized = \begin{cases} 0, & \text{if } raw < 2, \\ 1, & \text{otherwise.} \end{cases}$$

- Normalization of readout genes

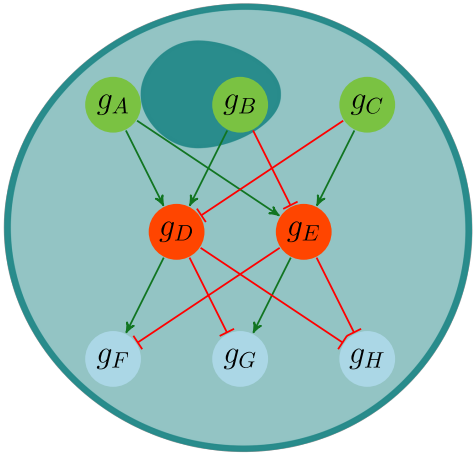
$$normalized = \frac{2}{\pi} \times \arctan(raw)$$



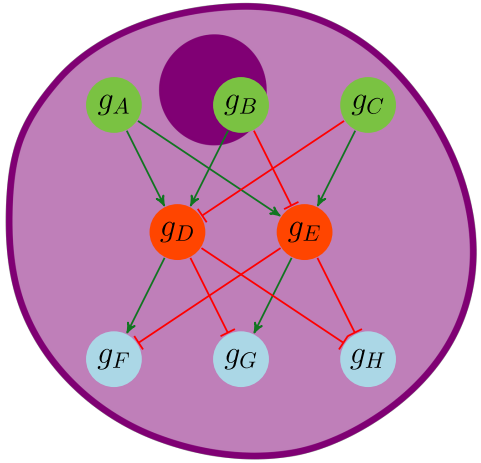
[Bolteau et al., *ISBRA* (2023)] [Bolteau et al., *J. of Computational Biology* (2024)] [Bolteau et al., *in prep.*]

Step 2. Experimental design construction

stage A

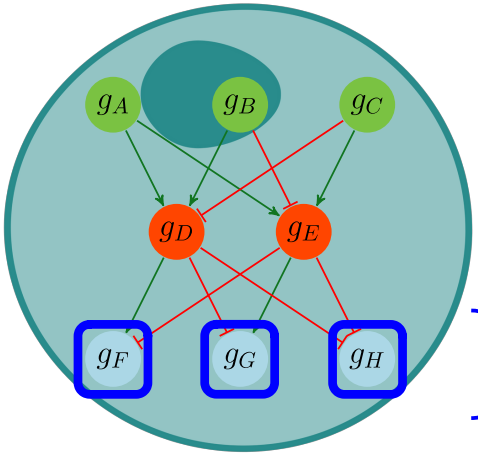


stage B



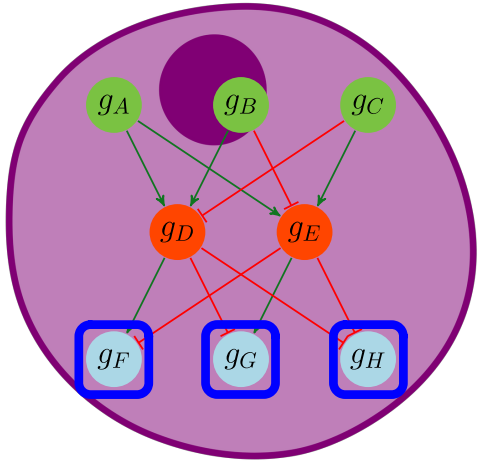
Step 2. Experimental design construction

stage A

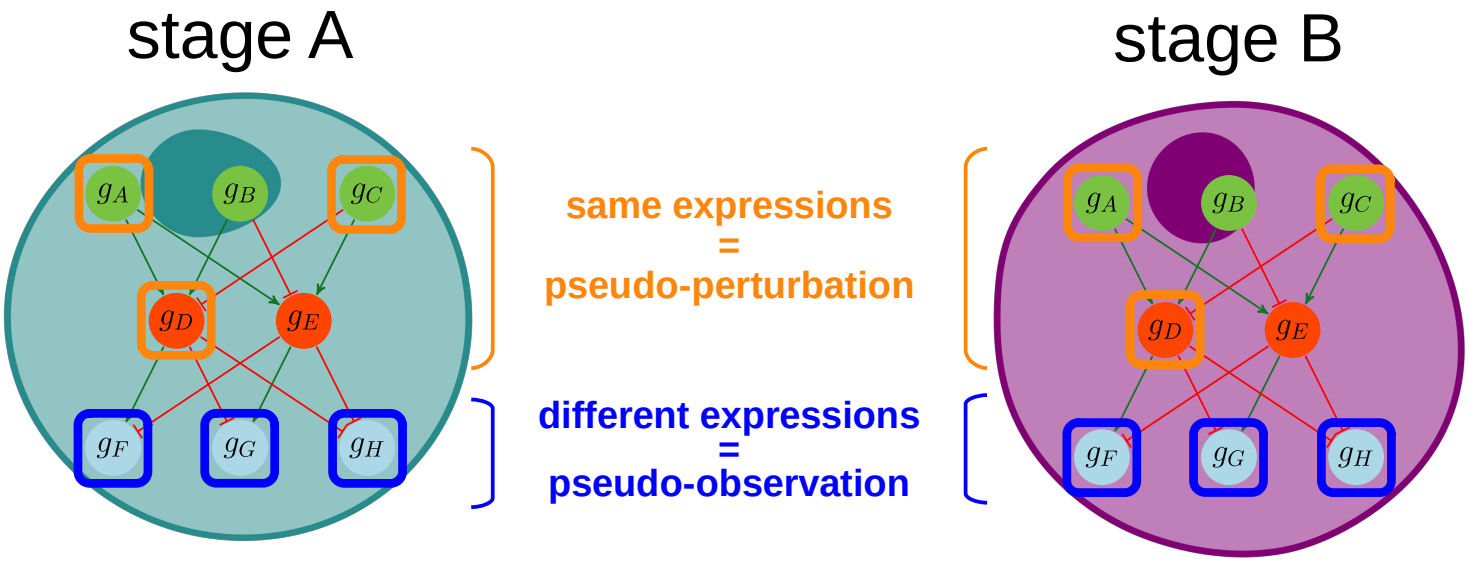


different expressions = pseudo-observation

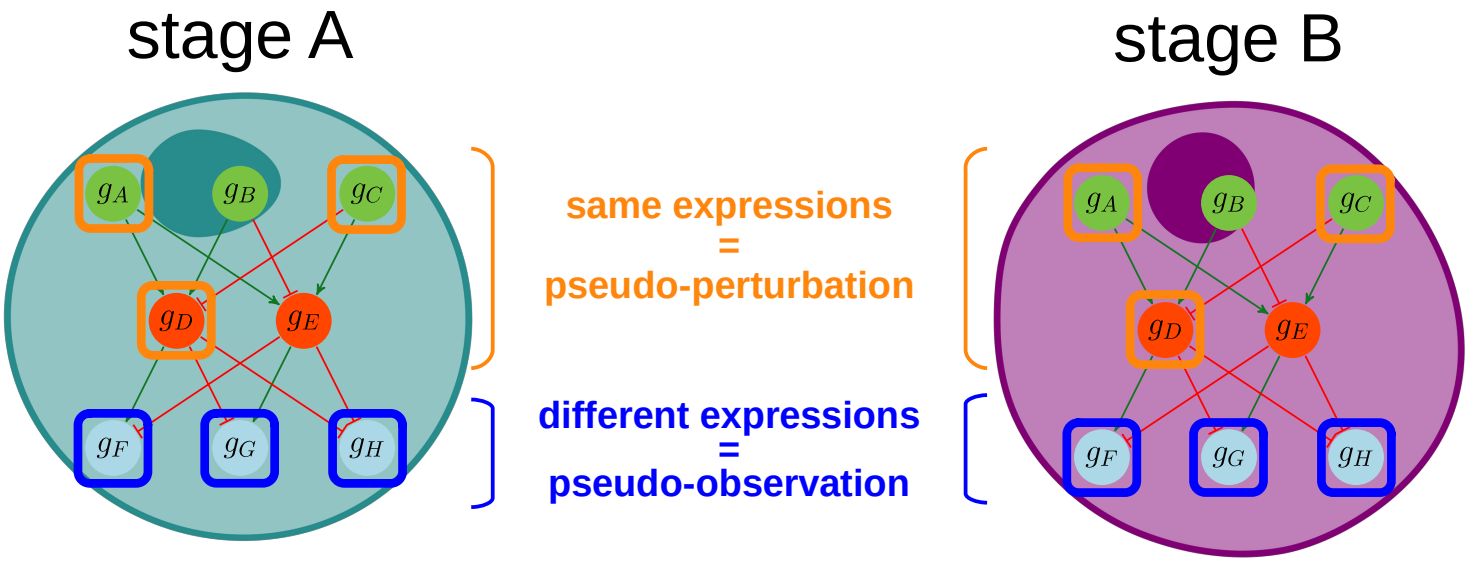
stage B



Step 2. Experimental design construction



Step 2. Experimental design construction



Pseudo-perturbation identification problem statement

Given k,
select k **genes** from gene population,
that maximize the number of **pairs of cells** from stages A and B,
having the **same expression for the k-genes**.

pseudo-perturbations

Step 2. Pseudo-perturbation identification

	possible pseudo-perturbations					pseudo-observations			
cell	A	B	C	D	E	F	G	H	stage
c_1	1	1	0	1	0	0.8	0.4	0.6	A
c_2	1	0	1	0	0	0.2	0.5	0.3	A
c_3	1	1	0	1	1	0.8	0.3	0.9	A
c_4	1	1	1	0	1	0.6	0.1	0.2	B
c_5	1	0	0	1	1	0.7	0.8	0.5	B
c_6	0	0	1	0	1	0.7	0.2	0.3	B

parameter *k*

Step 2. Pseudo-perturbation identification

pseudo-perturbations

pseudo-observations

cell	A	B	C	D	E	F	G	H	stage
c_1	1	1	0	1	0	0.8	0.4	0.6	A
c_2	1	0	1	0	0	0.2	0.5	0.3	A
c_3	1	1	0	1	1	0.8	0.3	0.9	A
c_4	1	1	1	0	1	0.6	0.1	0.2	B
c_5	1	0	0	1	1	0.7	0.8	0.5	B
c_6	0	0	1	0	1	0.7	0.2	0.3	B

parameter *k*

***k*=3**

selected genes: A, C & D

experimental design A

cell	A	C	D	F	G	H	stage
c_1	1	0	1	0.8	0.4	0.6	A
c_2	1	1	0	0.2	0.5	0.3	A

match

match

experimental design B

cell	A	C	D	F	G	H	stage
c_5	1	0	1	0.7	0.8	0.5	B
c_4	1	1	0	0.6	0.1	0.2	B

Step 2. Pseudo-perturbation identification

pseudo-perturbations | pseudo-observations

cell	A	B	C	D	E	F	G	H	stage
c_1	1	1	0	1	0	0.8	0.4	0.6	A
c_2	1	0	1	0	0	0.2	0.5	0.3	A
c_3	1	1	0	1	1	0.8	0.3	0.9	A
c_4	1	1	1	0	1	0.6	0.1	0.2	B
c_5	1	0	0	1	1	0.7	0.8	0.5	B
c_6	0	0	1	0	1	0.7	0.2	0.3	B

parameter *k*

***k*=3**

selected genes: A, C & D

experimental design A

cell	A	C	D	F	G	H	stage
c_1	1	0	1	0.8	0.4	0.6	A
c_2	1	1	0	0.2	0.5	0.3	A

match

match

experimental design B

cell	A	C	D	F	G	H	stage
c_5	1	0	1	0.7	0.8	0.5	B
c_4	1	1	0	0.6	0.1	0.2	B

- Optimal number of pseudo-perturbations: 2
- 2 pairs of matching cells: (c_1, c_5), (c_2, c_4)

Step 2. Pseudo-perturbation identification program

Program encoded in answer set programming (ASP) [Baral, *Cambridge University Press* (2003)]
Use of Clingo solver (Potassco suite) [Gebser et al., *AI Communications* (2011)]

Step 2. Pseudo-perturbation identification program

Program encoded in answer set programming (ASP) [Baral, *Cambridge University Press* (2003)]
Use of Clingo solver (Potassco suite) [Gebser et al., *AI Communications* (2011)]

Main rules (x4)

- k-genes: Select k genes among all possible combinations of input + intermediate genes
- Reachability: input $\rightarrow$ intermediate
- Matching cells: Select pairs of cells (c_1, c_2), $c_1 \in A$, $c_2 \in B$, for which the (binarized) expression matches for each of the *k-genes*
- Filter redundancy: The set of k (binarized) expressions should differ for all pseudo-perturbations of the same stage

Optimization

- Maximize the number of pseudo-perturbations of either A or B stage

[Bolteau et al., *ISBRA* (2023)] [Bolteau et al., *J. of Computational Biology* (2024)] [Bolteau et al., *in prep.*]

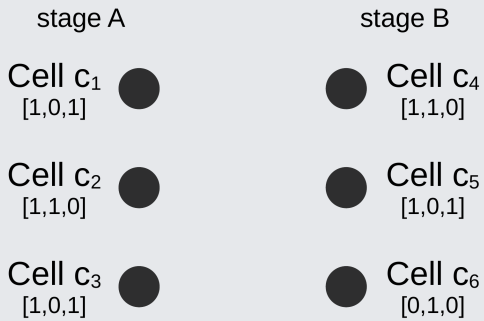
Step 2. Pseudo-perturbation identification program

Dataset	Chebouba [Chebouba et al., *BMC Bioinformatics* (2018)]		SCIBORG first version [Bolteau et al., *ISBRA* (2023)]		SCIBORG current version [Bolteau et al., *in prep*]	
	Execution time	Pseudo-perturbations	Execution time	Pseudo-perturbations	Execution time	Pseudo-perturbations
A	0.008s	3	0.008s	3	0.009s	3
B	0.048s	1	0.223s	4	0.060s	3
C	1.424s	1	10 min*	11	15min*	13
D	10 min*	10	10 min*	22	15min*	35
SC	5h 2 min	3	65h*	20	7h*	92
P	50h*	23	50h*	25		

Dataset A: artificial toy dataset
Datasets B to D: toy dasatets of single-cell RNA-seq data from human emb. dev.
Dataset SC: single-cell dataset of medium and late TE stages
Dataset P: phosphoproteomics dataset from [Chebouba et al., *BMC Bioinformatics* (2018)]
* Execution time corresponds to the fixed timeout.

Step 2. Maximizing the pseudo-observation difference

Redundancy

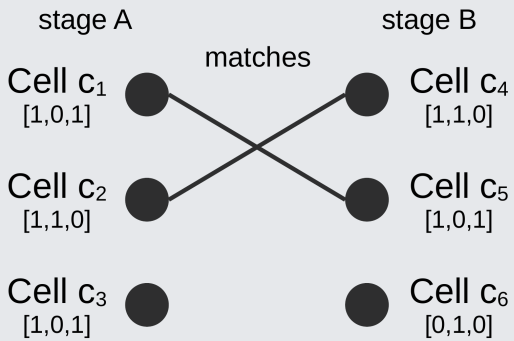


2 solutions:

[Bolteau et al., *ISBRA* (2023)] [Bolteau et al., *J. of Computational Biology* (2024)]

Step 2. Maximizing the pseudo-observation difference

Redundancy



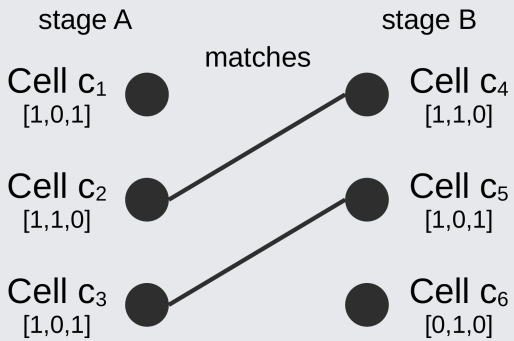
2 solutions:

1. (c_1, c_5), (c_2, c_4)

[Bolteau et al., *ISBRA* (2023)] [Bolteau et al., *J. of Computational Biology* (2024)]

Step 2. Maximizing the pseudo-observation difference

Redundancy



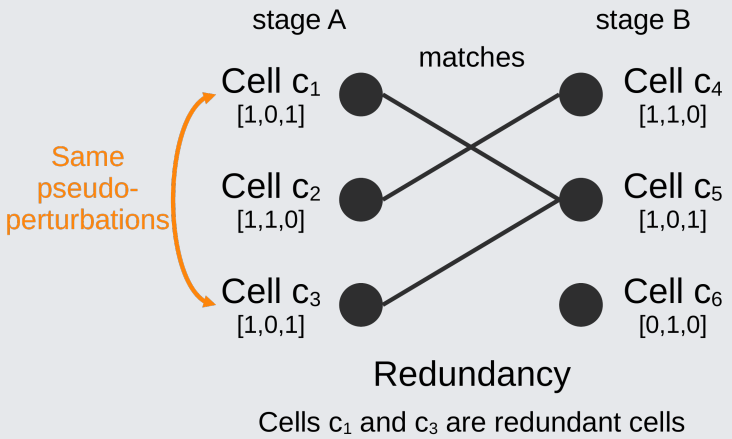
2 solutions:

1. (c_1, c_5), (c_2, c_4)
2. (c_3, c_5), (c_2, c_4)

[Bolteau et al., *ISBRA* (2023)] [Bolteau et al., *J. of Computational Biology* (2024)]

Step 2. Maximizing the pseudo-observation difference

Redundancy



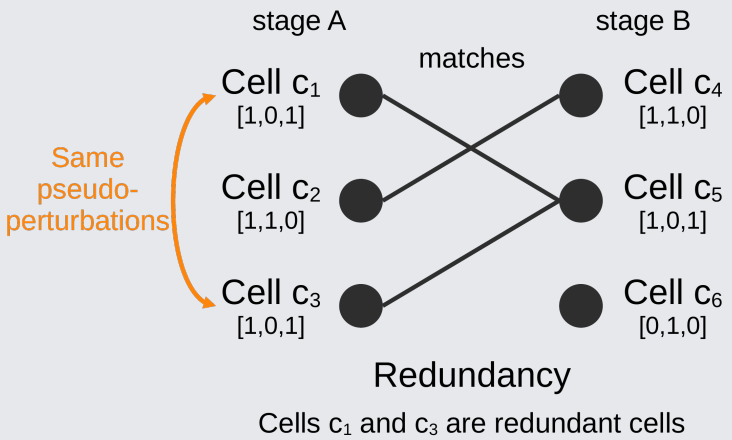
2 solutions:

1. (c_1, c_5), (c_2, c_4)
2. (c_3, c_5), (c_2, c_4)

[Bolteau et al., *ISBRA* (2023)] [Bolteau et al., *J. of Computational Biology* (2024)]

Step 2. Maximizing the pseudo-observation difference

Redundancy



2 solutions:

1. (c_1, c_5), (c_2, c_4)
2. (c_3, c_5), (c_2, c_4)

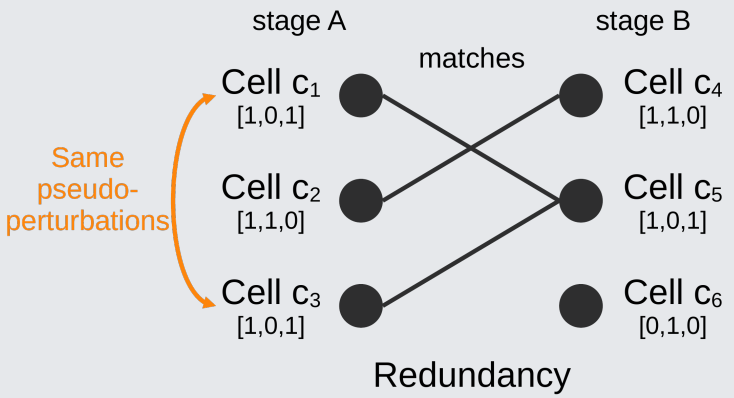
Representativity score of pseudo-perturbations:

- Stage A: 100% (3/3)
- Stage B: 66% (2/3)

[Bolteau et al., *ISBRA* (2023)] [Bolteau et al., *J. of Computational Biology* (2024)]

Step 2. Maximizing the pseudo-observation difference

Redundancy



Cells c_1 and c_3 are redundant cells

2 solutions:

1. (c_1, c_5), (c_2, c_4)
2. (c_3, c_5), (c_2, c_4)

Representativity score of pseudo-perturbations:

- Stage A: 100% (3/3)
- Stage B: 66% (2/3)

Pseudo-observation difference maximization

cell	A	C	D	F	G	H	stage
c_1	1	0	1	0.8	0.4	0.6	A
c_2	1	1	0	0.2	0.5	0.3	A

match

match

cell	A	C	D	F	G	H	stage
c_5	1	0	1	0.7	0.8	0.5	B
c_4	1	1	0	0.6	0.1	0.2	B

Solution 1

vs.

cell	A	C	D	F	G	H	stage
c_3	1	0	1	0.8	0.3	0.9	A
c_2	1	1	0	0.2	0.5	0.3	A

match

match

cell	A	C	D	F	G	H	stage
c_5	1	0	1	0.7	0.8	0.5	B
c_4	1	1	0	0.6	0.1	0.2	B

Solution 2

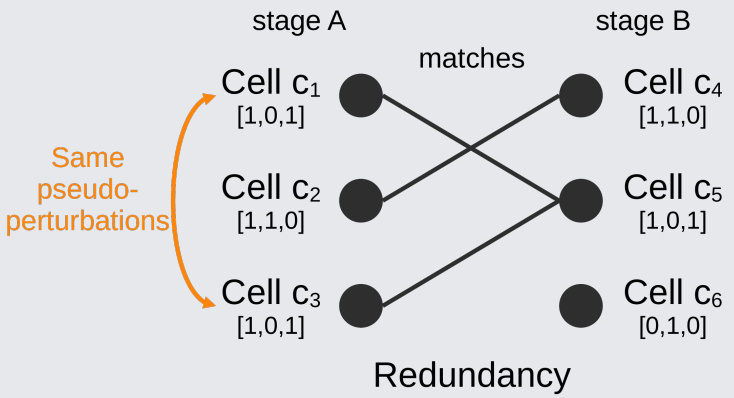
experimental design for stage A

experimental design for stage B

[Bolteau et al., *ISBRA* (2023)] [Bolteau et al., *J. of Computational Biology* (2024)]

Step 2. Maximizing the pseudo-observation difference

Redundancy



Cells c_1 and c_3 are redundant cells

2 solutions:

1. (c_1, c_5), (c_2, c_4)
2. (c_3, c_5), (c_2, c_4)

Representativity score of pseudo-perturbations:

- Stage A: 100% (3/3)
- Stage B: 66% (2/3)

Pseudo-observation difference maximization

cell	A	C	D	F	G	H	stage
c_1	1	0	1	0.8	0.4	0.6	A
c_2	1	1	0	0.2	0.5	0.3	A

match, match

cell	A	C	D	F	G	H	stage
c_5	1	0	1	0.7	0.8	0.5	B
c_4	1	1	0	0.6	0.1	0.2	B

Solution 1

$\text{diff}(c_1, c_5) = |0.8-0.7| + |0.4-0.8| + |0.6-0.5| = 0.6$

vs.

cell	A	C	D	F	G	H	stage
c_3	1	0	1	0.8	0.3	0.9	A
c_2	1	1	0	0.2	0.5	0.3	A

match, match

cell	A	C	D	F	G	H	stage
c_5	1	0	1	0.7	0.8	0.5	B
c_4	1	1	0	0.6	0.1	0.2	B

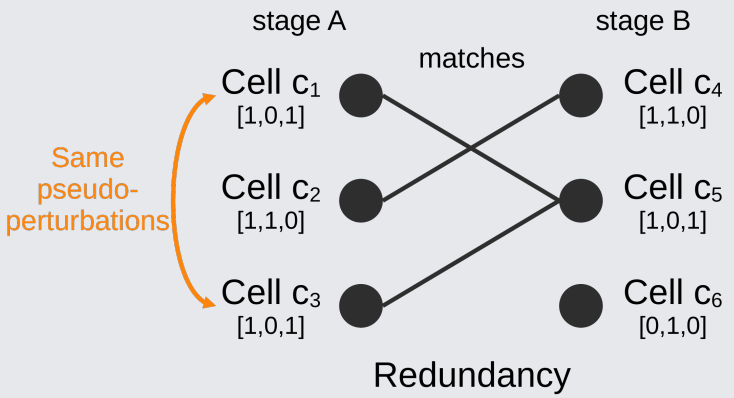
Solution 2

experimental design for stage A — experimental design for stage B

[Bolteau et al., *ISBRA* (2023)] [Bolteau et al., *J. of Computational Biology* (2024)]

Step 2. Maximizing the pseudo-observation difference

Redundancy



Cells c_1 and c_3 are redundant cells

2 solutions:

1. (c_1, c_5), (c_2, c_4)
2. (c_3, c_5), (c_2, c_4)

Representativity score of pseudo-perturbations:

- Stage A: 100% (3/3)
- Stage B: 66% (2/3)

Pseudo-observation difference maximization

cell	A	C	D	F	G	H	stage
c_1	1	0	1	0.8	0.4	0.6	A
c_2	1	1	0	0.2	0.5	0.3	A

match, match

cell	A	C	D	F	G	H	stage
c_5	1	0	1	0.7	0.8	0.5	B
c_4	1	1	0	0.6	0.1	0.2	B

Solution 1

$\text{diff}(c_1, c_5) = |0.8-0.7| + |0.4-0.8| + |0.6-0.5| = 0.6$

vs.

$\text{diff}(c_3, c_5) = |0.8-0.7| + |0.3-0.8| + |0.9-0.5| = 0.9$

cell	A	C	D	F	G	H	stage
c_3	1	0	1	0.8	0.3	0.9	A
c_2	1	1	0	0.2	0.5	0.3	A

match, match

cell	A	C	D	F	G	H	stage
c_5	1	0	1	0.7	0.8	0.5	B
c_4	1	1	0	0.6	0.1	0.2	B

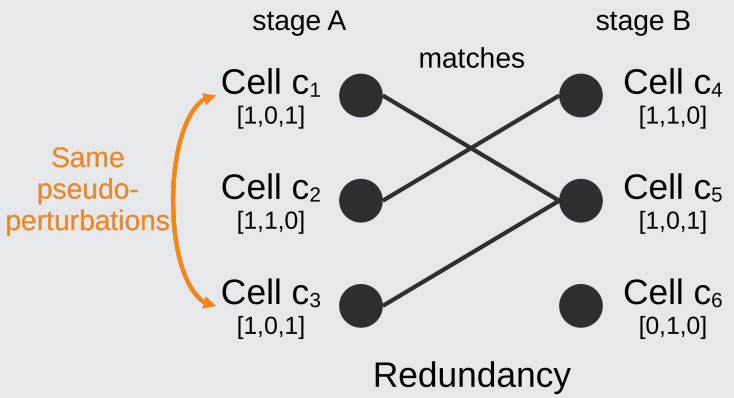
Solution 2

experimental design for stage A — experimental design for stage B

[Bolteau et al., *ISBRA* (2023)] [Bolteau et al., *J. of Computational Biology* (2024)]

Step 2. Maximizing the pseudo-observation difference

Redundancy



Cells c_1 and c_3 are redundant cells

2 solutions:

1. (c_1, c_5), (c_2, c_4)
2. (c_3, c_5), (c_2, c_4)

Representativity score of pseudo-perturbations:

- Stage A: 100% (3/3)
- Stage B: 66% (2/3)

Pseudo-observation difference maximization

cell	A	C	D	F	G	H	stage
c_1	1	0	1	0.8	0.4	0.6	A
c_2	1	1	0	0.2	0.5	0.3	A

match, match

cell	A	C	D	F	G	H	stage
c_5	1	0	1	0.7	0.8	0.5	B
c_4	1	1	0	0.6	0.1	0.2	B

Solution 1

$\text{diff}(c_1, c_5) = |0.8-0.7| + |0.4-0.8| + |0.6-0.5| = 0.6$

vs.

$\text{diff}(c_3, c_5) = |0.8-0.7| + |0.3-0.8| + |0.9-0.5| = 0.9$

cell	A	C	D	F	G	H	stage
c_3	1	0	1	0.8	0.3	0.9	A
c_2	1	1	0	0.2	0.5	0.3	A

match, match

cell	A	C	D	F	G	H	stage
c_5	1	0	1	0.7	0.8	0.5	B
c_4	1	1	0	0.6	0.1	0.2	B

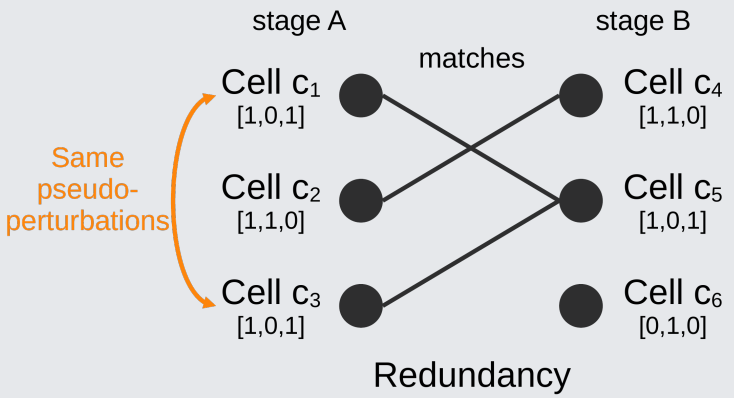
Solution 2

experimental design for stage A — experimental design for stage B

[Bolteau et al., *ISBRA* (2023)] [Bolteau et al., *J. of Computational Biology* (2024)]

Step 2. Maximizing the pseudo-observation difference

Redundancy



2 solutions:

1. (c_1, c_5), (c_2, c_4)
2. (c_3, c_5), (c_2, c_4)

Representativity score of pseudo-perturbations:

- Stage A: 100% (3/3)
- Stage B: 66% (2/3)

Pseudo-observation difference maximization

cell	A	C	D	F	G	H	stage
c_1	1	0	1	0.8	0.4	0.6	A
c_2	1	1	0	0.2	0.5	0.3	A

match / match

cell	A	C	D	F	G	H	stage
c_5	1	0	1	0.7	0.8	0.5	B
c_4	1	1	0	0.6	0.1	0.2	B

Solution 1

$\text{diff}(c_1, c_5) = |0.8-0.7| + |0.4-0.8| + |0.6-0.5| = 0.6$

vs.

$\text{diff}(c_3, c_5) = |0.8-0.7| + |0.3-0.8| + |0.9-0.5| = 0.9$

cell	A	C	D	F	G	H	stage
c_3	1	0	1	0.8	0.3	0.9	A
c_2	1	1	0	0.2	0.5	0.3	A

match / match

cell	A	C	D	F	G	H	stage
c_5	1	0	1	0.7	0.8	0.5	B
c_4	1	1	0	0.6	0.1	0.2	B

Solution 2

experimental design for stage A — experimental design for stage B

[Bolteau et al., *ISBRA* (2023)] [Bolteau et al., *J. of Computational Biology* (2024)]

Step 3. BN learning using Caspo

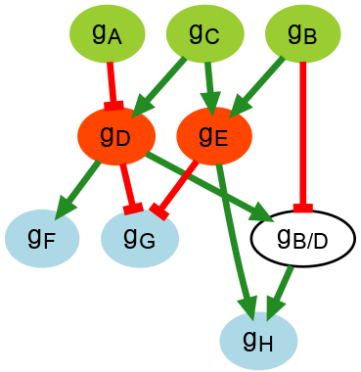
[Guziolowski et al., *Bioinformatics* (2013)]

cell	A	C	D	F	G	H	stage
c_3	1	0	1	0.8	0.3	0.9	A
c_2	1	1	0	0.2	0.5	0.3	A

experimental design of stage A

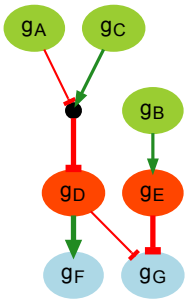
cell	A	C	D	F	G	H	stage
c_5	1	0	1	0.7	0.8	0.5	B
c_4	1	1	0	0.6	0.1	0.2	B

experimental design of stage B

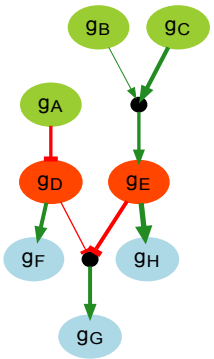


PKN

learning
Boolean networks
(Caspo)



family of Boolean networks of stage A

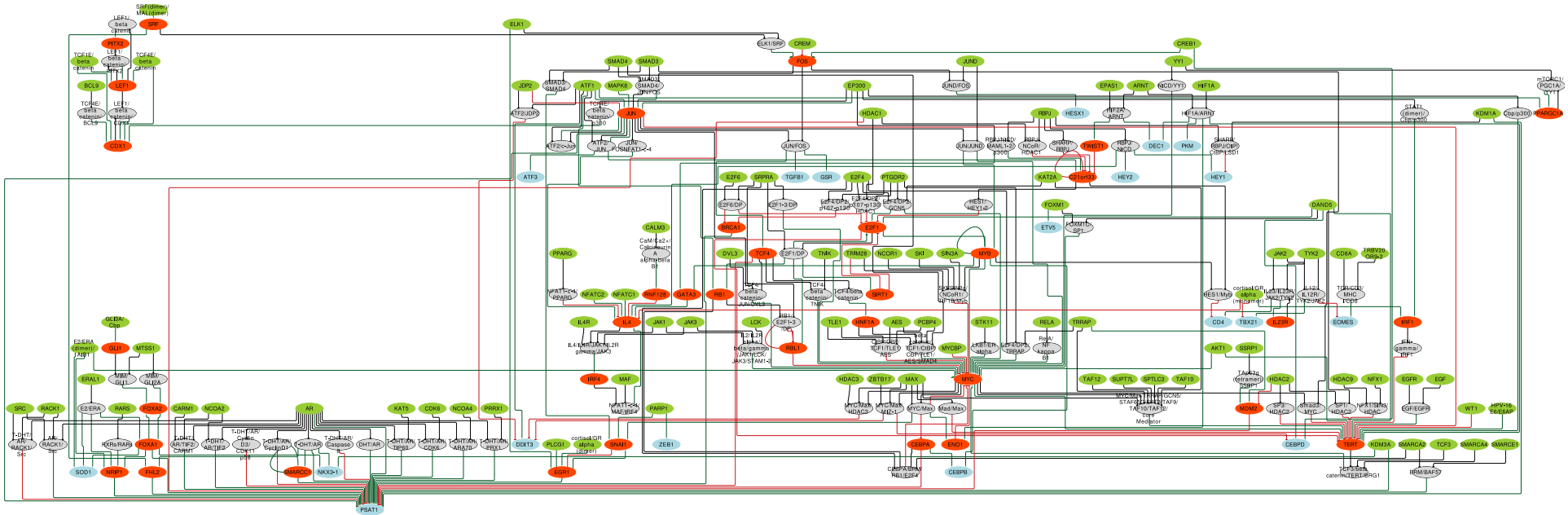


family of Boolean networks of stage B

Reconstructed PKN

Input	Output					Parameter
348 transcription factors from gene module analysis	233 nodes				369 edges	No depth (total reconstruction)
	85 inputs	36 intermediates	19 readouts	85 protein complexes		

[Meistermann et al., *Cell Stem Cell* (2021)]

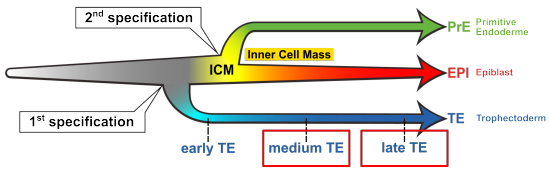


[Bolteau et al., *in prep.*]

Pseudo-perturbation identification

Focus on medium and late TE stages

(TE maturation)

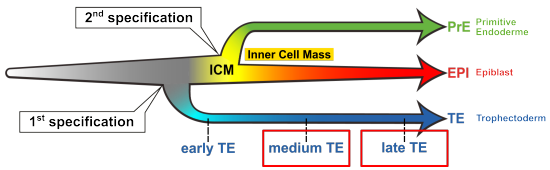


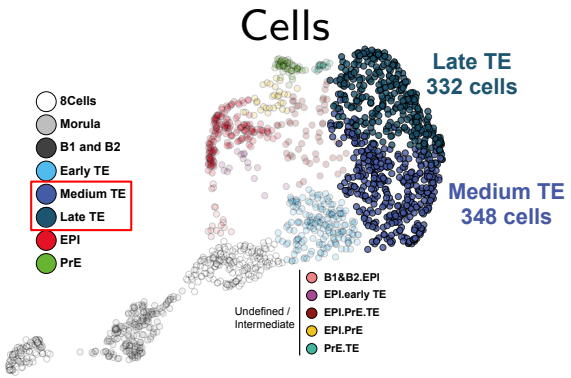
TE = trophectoderm

Pseudo-perturbation identification

Focus on medium and late TE stages

(TE maturation)



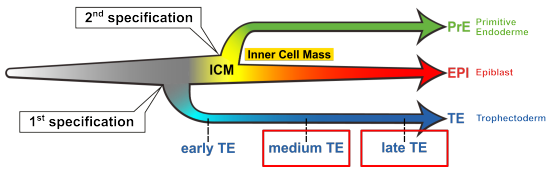


TE = trophectoderm

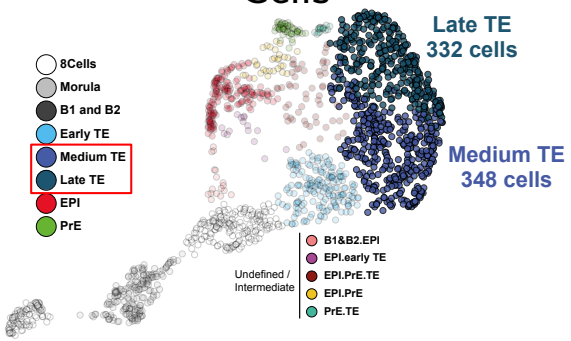
Pseudo-perturbation identification

Focus on medium and late TE stages

(TE maturation)



Cells



Genes

Choose k genes
from 85 inputs and 36 intermediates
to identify pseudo-perturbations

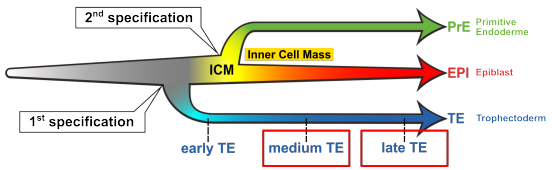
$$\binom{85+36}{k} \iff \binom{121}{k}$$

TE = trophectoderm

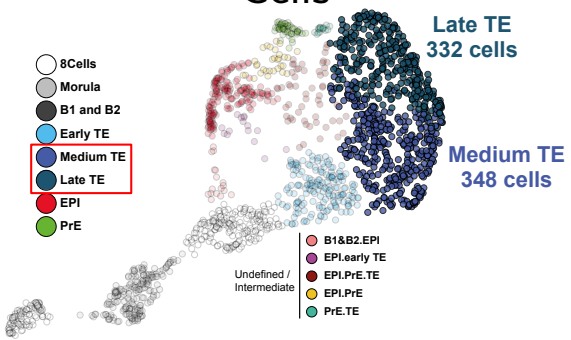
Pseudo-perturbation identification

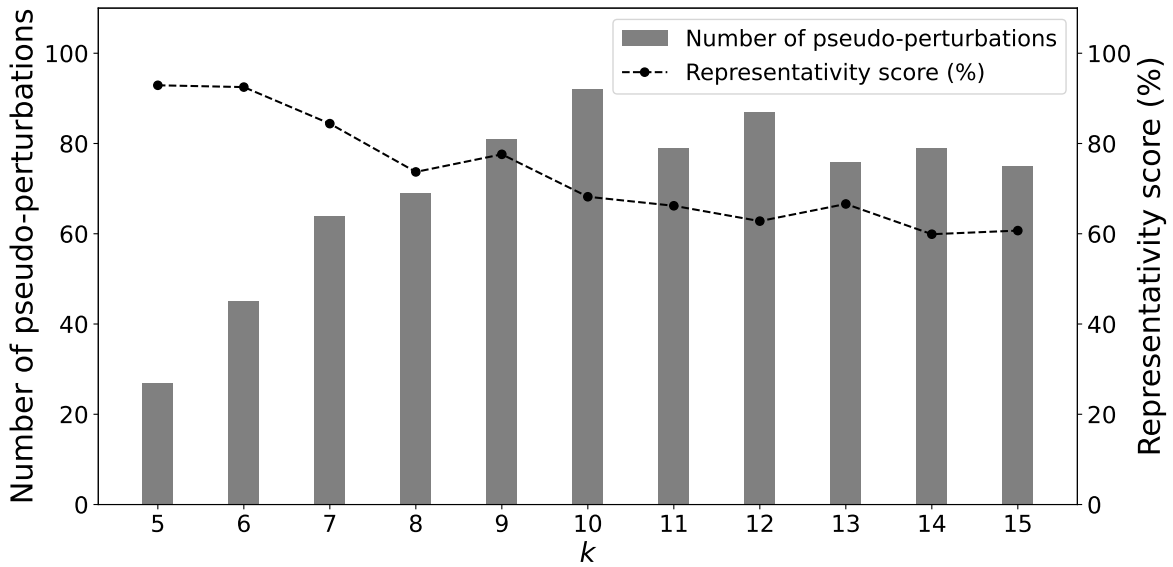
Focus on medium and late TE stages

(TE maturation)



Cells





Genes

Choose k genes
from 85 inputs and 36 intermediates
to identify pseudo-perturbations

$$\binom{85+36}{k} \iff \binom{121}{k}$$

TE = trophectoderm

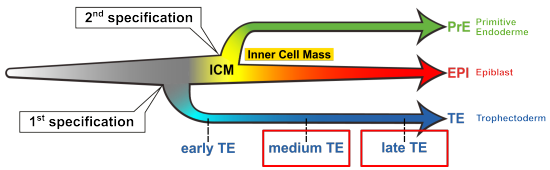
Execution time: 30 hours on a computer cluster (1.5 To RAM)

[Bolteau et al., *in prep.*]

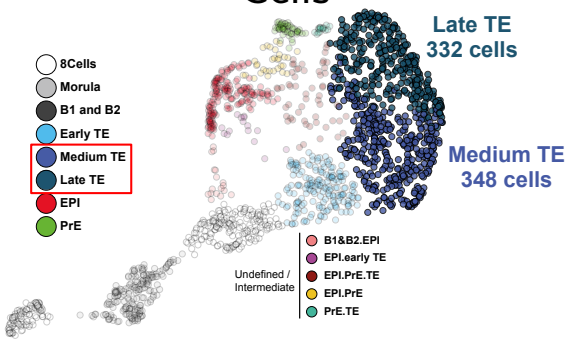
Pseudo-perturbation identification

Focus on medium and late TE stages

(TE maturation)



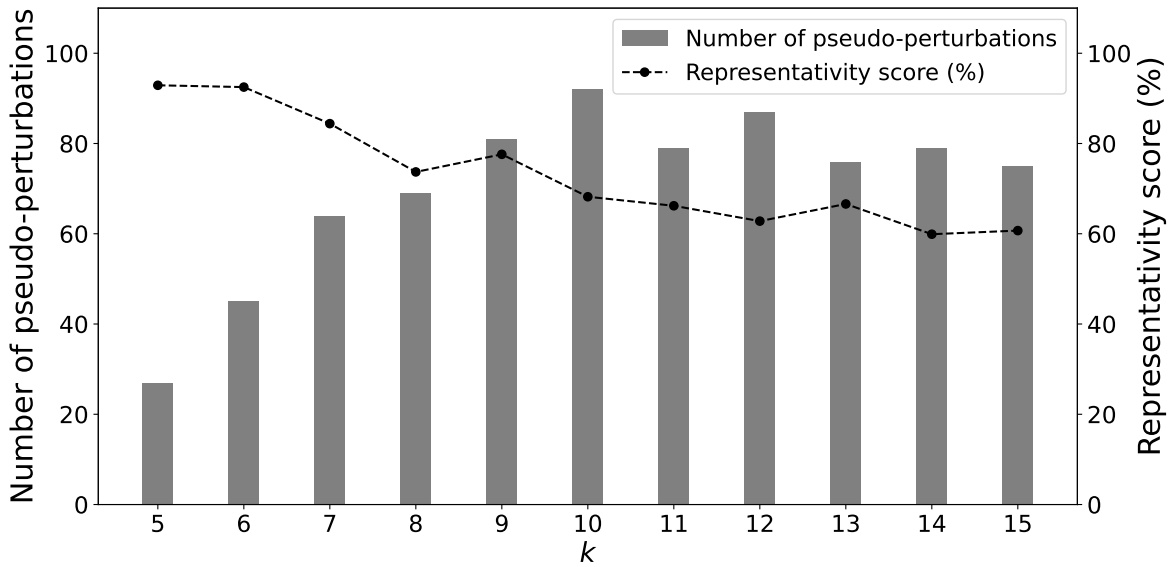
Cells



Genes

Choose k genes
from 85 inputs and 36 intermediates
to identify pseudo-perturbations

$$\binom{85+36}{k} \iff \binom{121}{k}$$



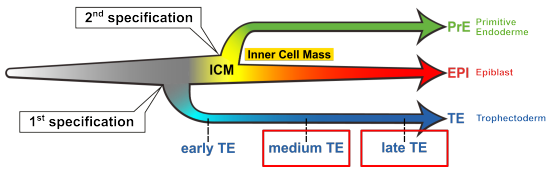
$k = 10$ the best parameter value for pseudo-perturbation identification

TE = trophectoderm Execution time: 30 hours on a computer cluster (1.5 To RAM) [Bolteau et al., *in prep.*]

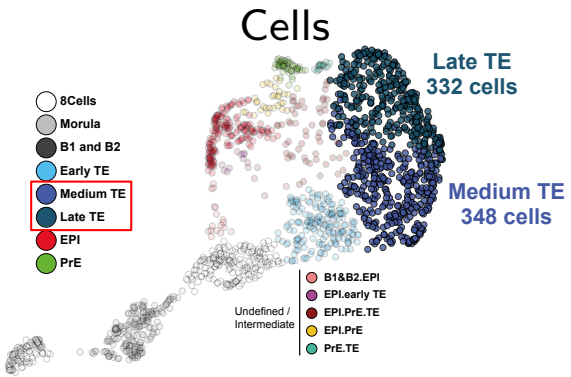
Pseudo-perturbation identification

Focus on medium and late TE stages

(TE maturation)



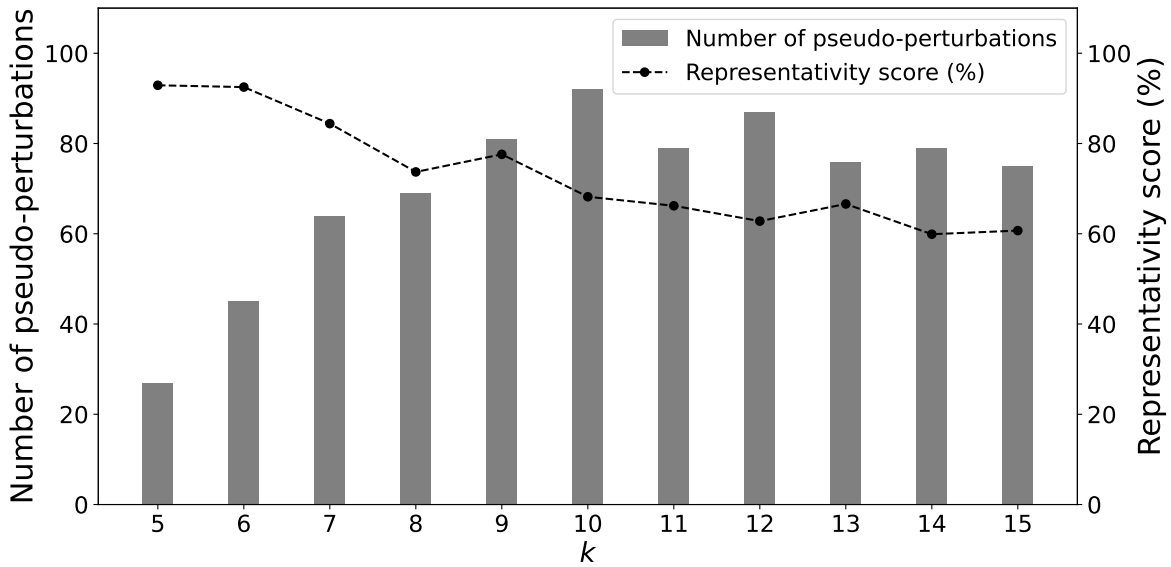
Cells



Genes

Choose k genes
from 85 inputs and 36 intermediates
to identify pseudo-perturbations

$$\binom{85+36}{k} \iff \binom{121}{k}$$



Research space

$$\binom{|genes|}{k} \times 2^{|cells_A| \times |cells_B|}$$

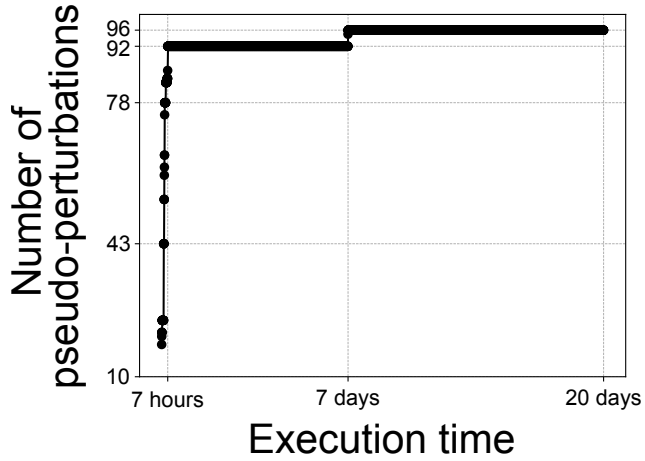
$$= \binom{121}{10} \times 2^{348 \times 332}$$

$$= 3.26 \times 10^{34793}$$

$k = 10$ the best parameter value for pseudo-perturbation identification

TE = trophectoderm Execution time: 30 hours on a computer cluster (1.5 To RAM) [Bolteau et al., *in prep.*]

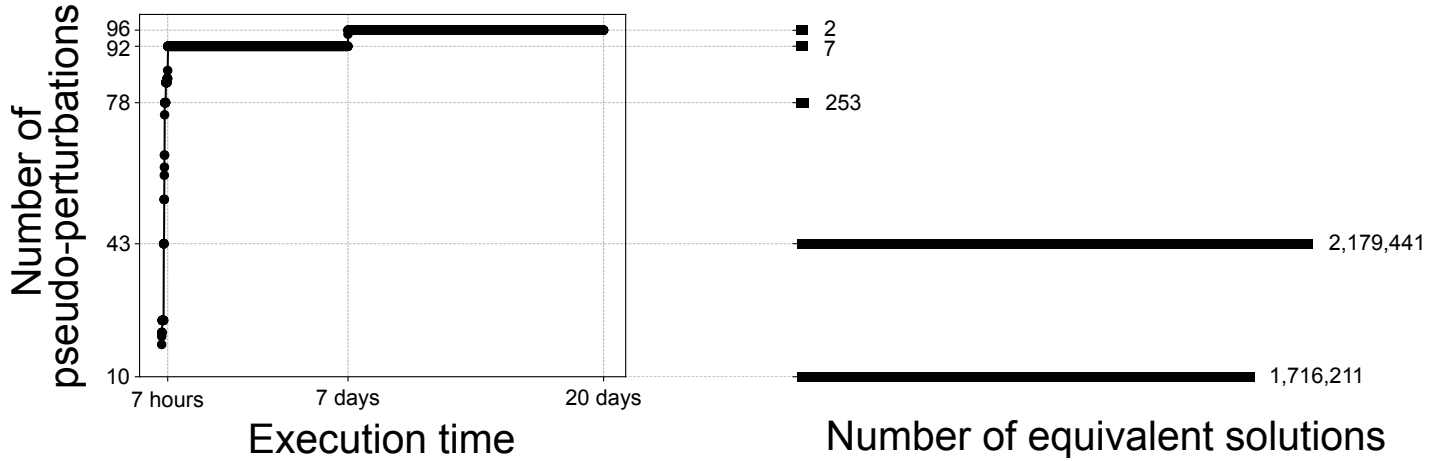
Convergence of the identified pseudo-perturbations



Convergence of the number of pseudo-perturbations

[Bolteau et al., *in prep.*]

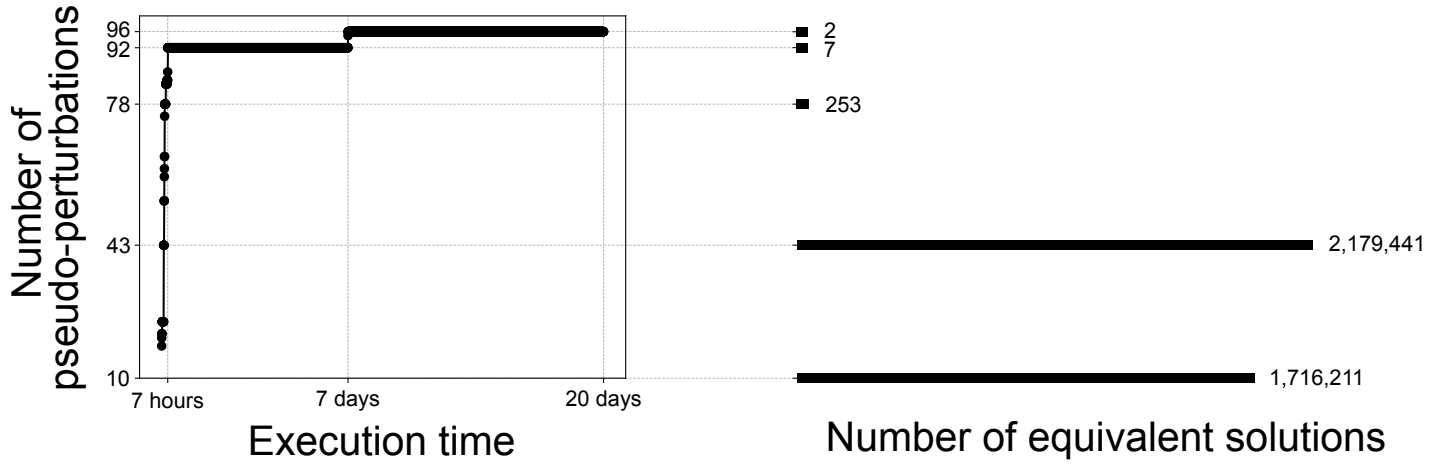
Convergence of the identified pseudo-perturbations



Convergence of the number of pseudo-perturbations and equivalent solutions

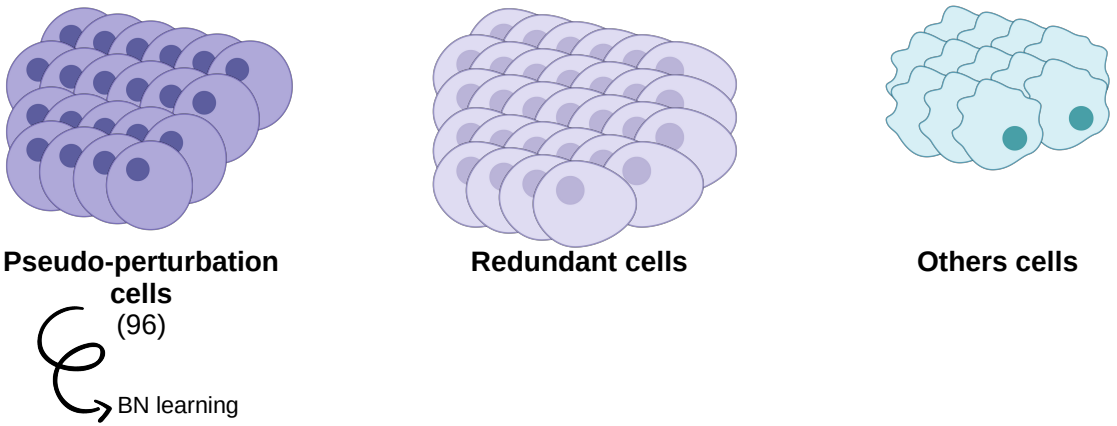
[Bolteau et al., *in prep.*]

Convergence of the identified pseudo-perturbations



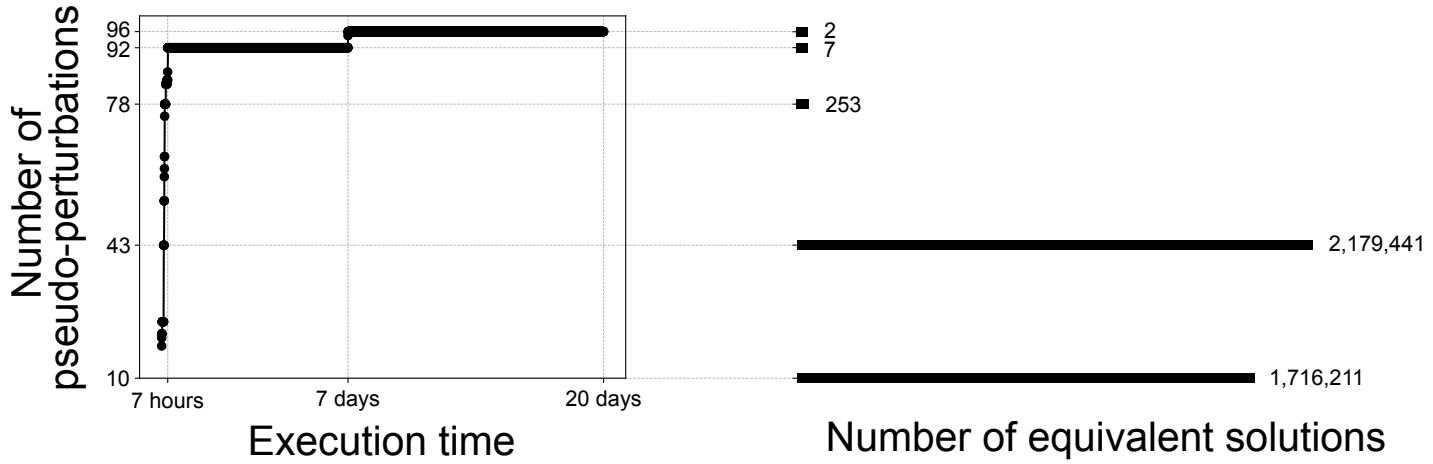
Execution time: 7 days on a computer cluster (1.5 To RAM)

Convergence of the number of pseudo-perturbations and equivalent solutions



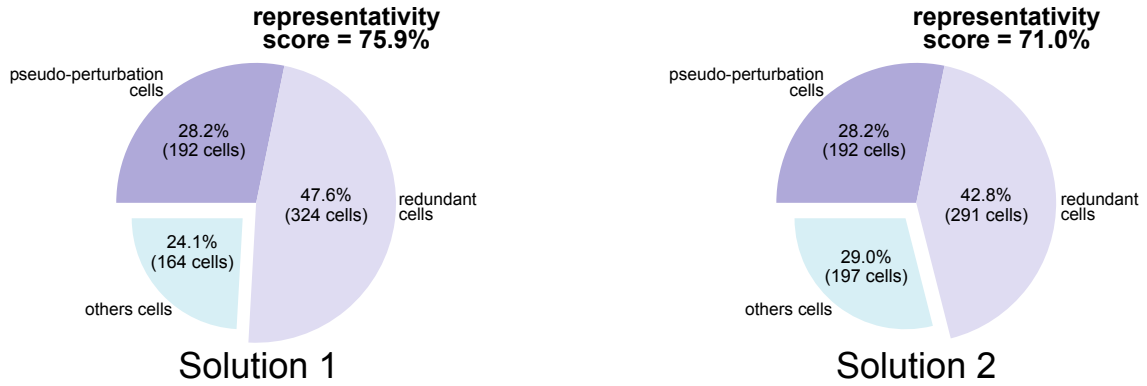
[Bolteau et al., *in prep.*]

Convergence of the identified pseudo-perturbations



Execution time: 7 days on a computer cluster (1.5 To RAM)

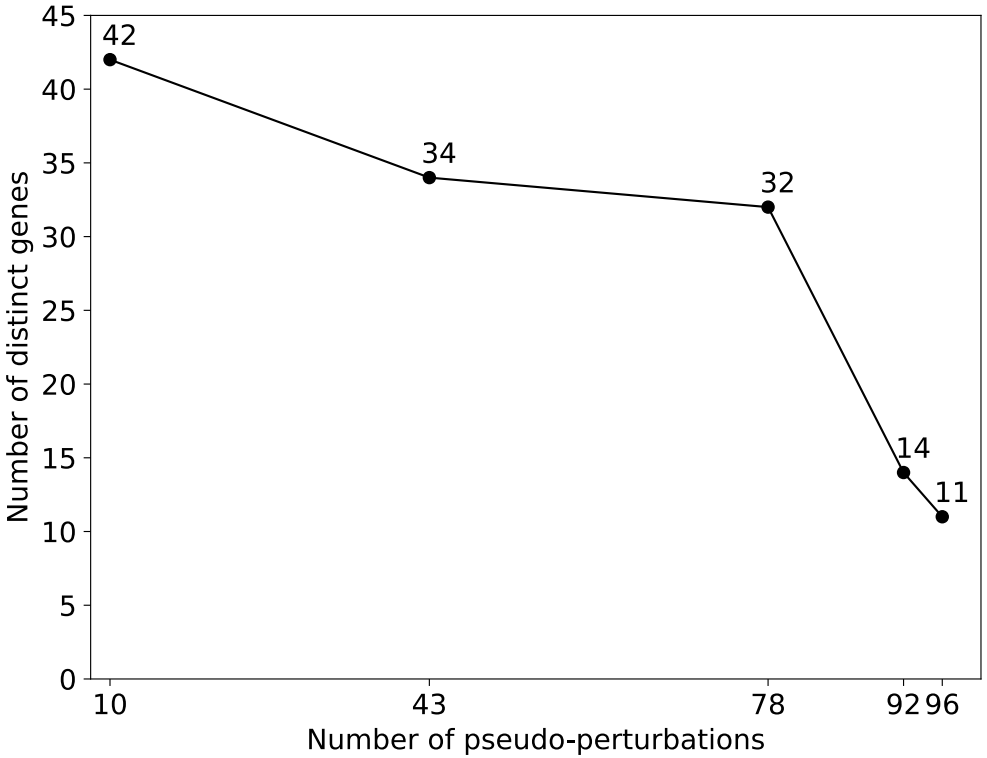
Convergence of the number of pseudo-perturbations and equivalent solutions



The 96 pseudo-perturbations are representative of more than 70% of the cell population

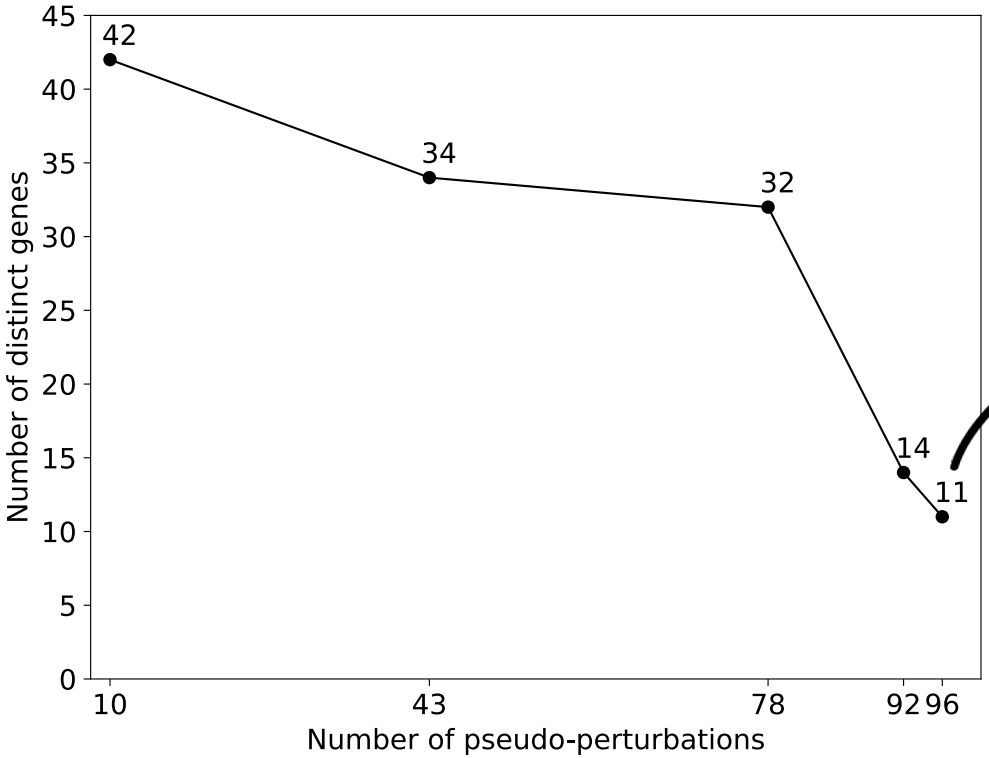
[Bolteau et al., *in prep.*]

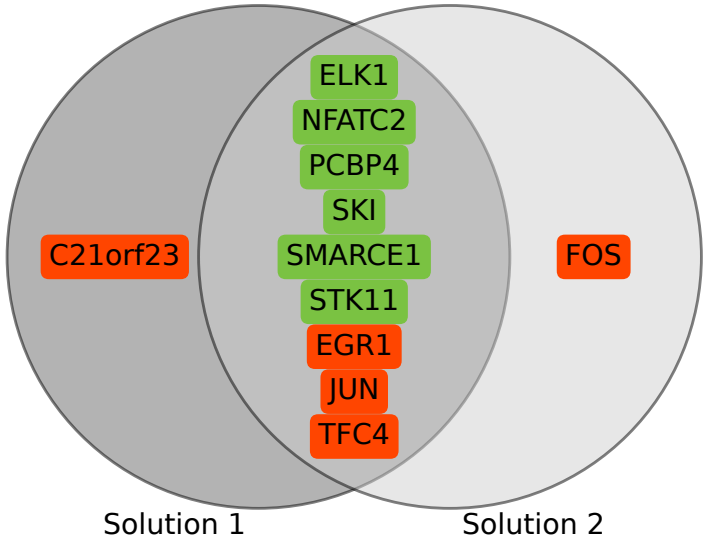
Robustness of the equivalent solutions' composition



[Bolteau et al., *in prep.*]

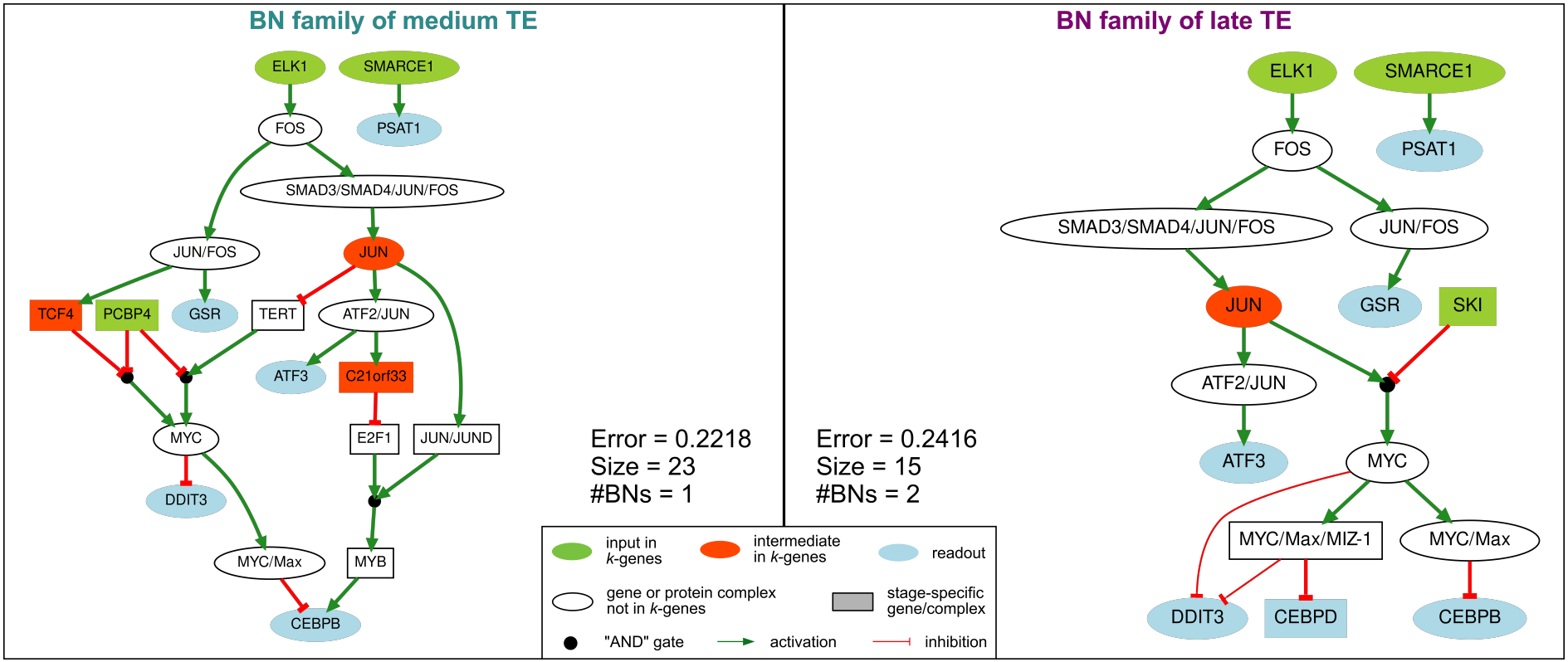
Robustness of the equivalent solutions' composition





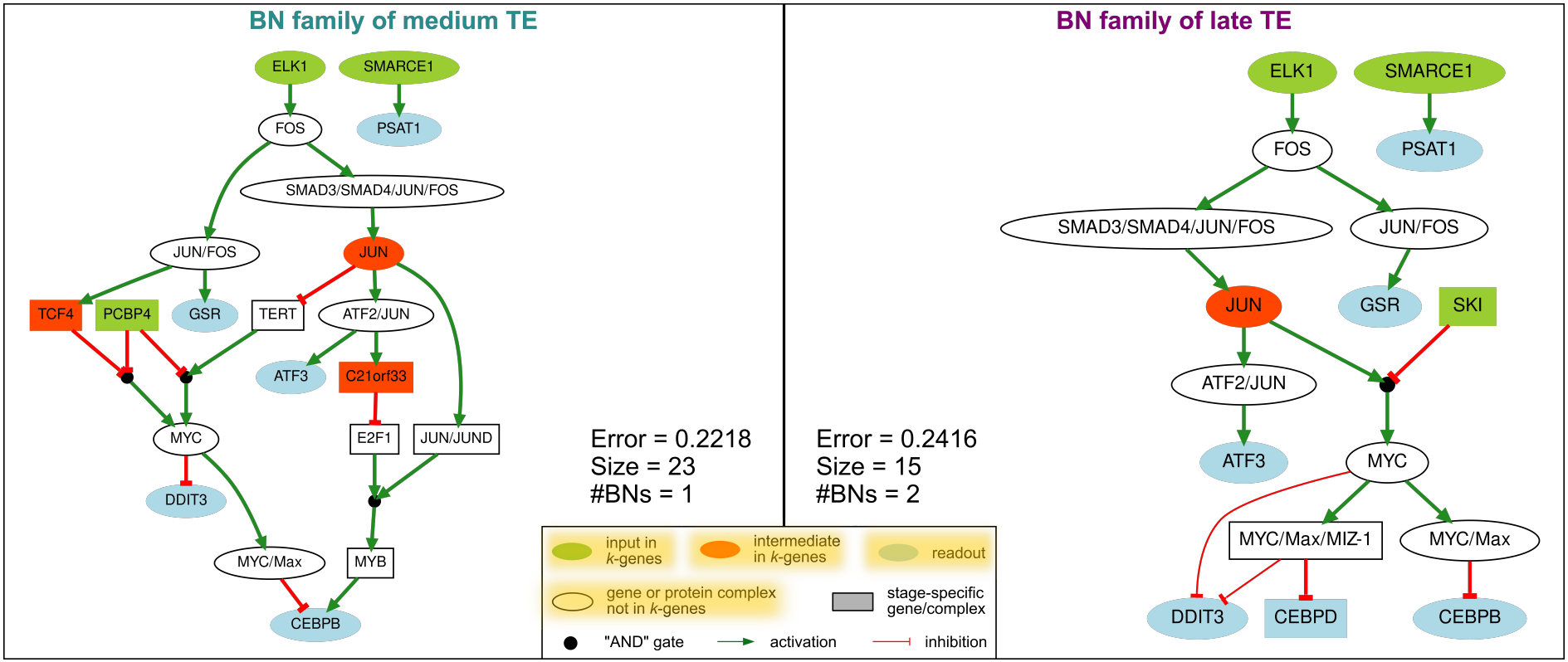
[Bolteau et al., *in prep.*]

Regulatory mechanisms in the learned models



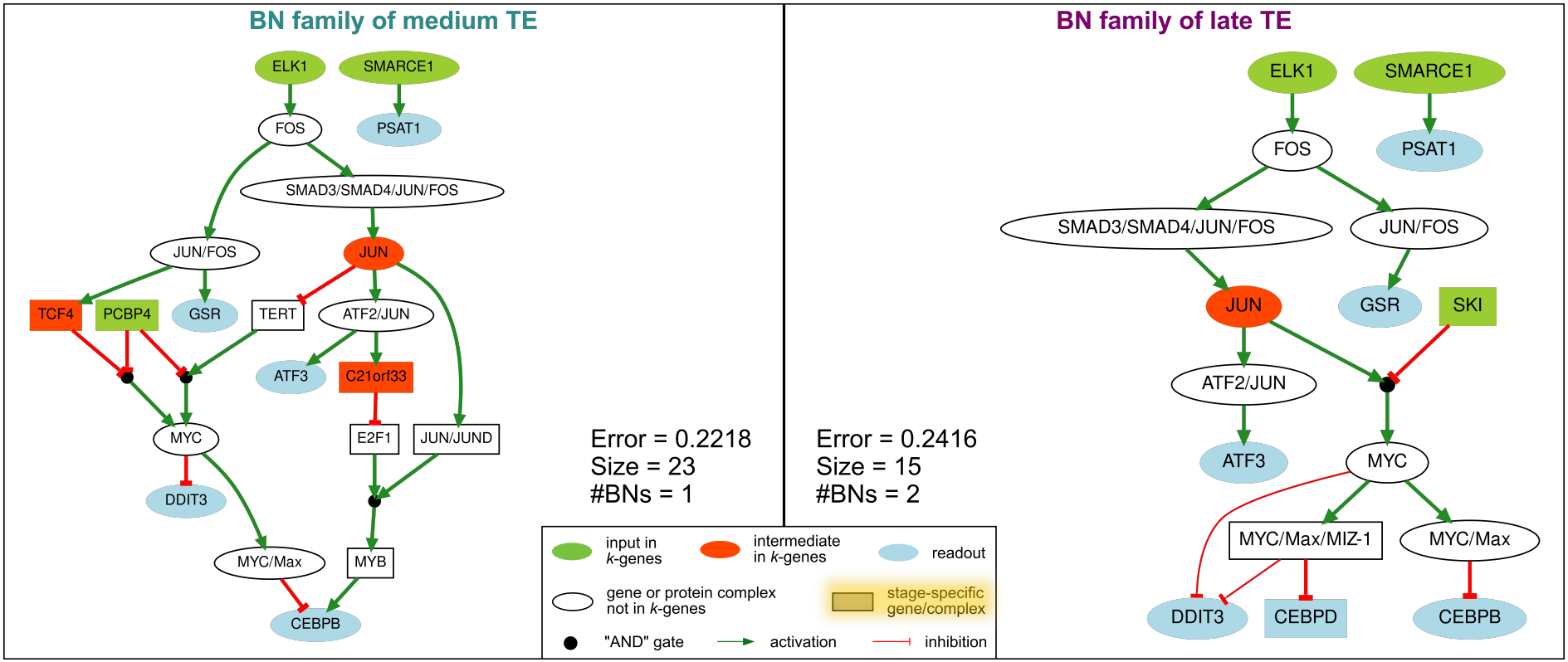
[Bolteau et al., *in prep.*]

Regulatory mechanisms in the learned models



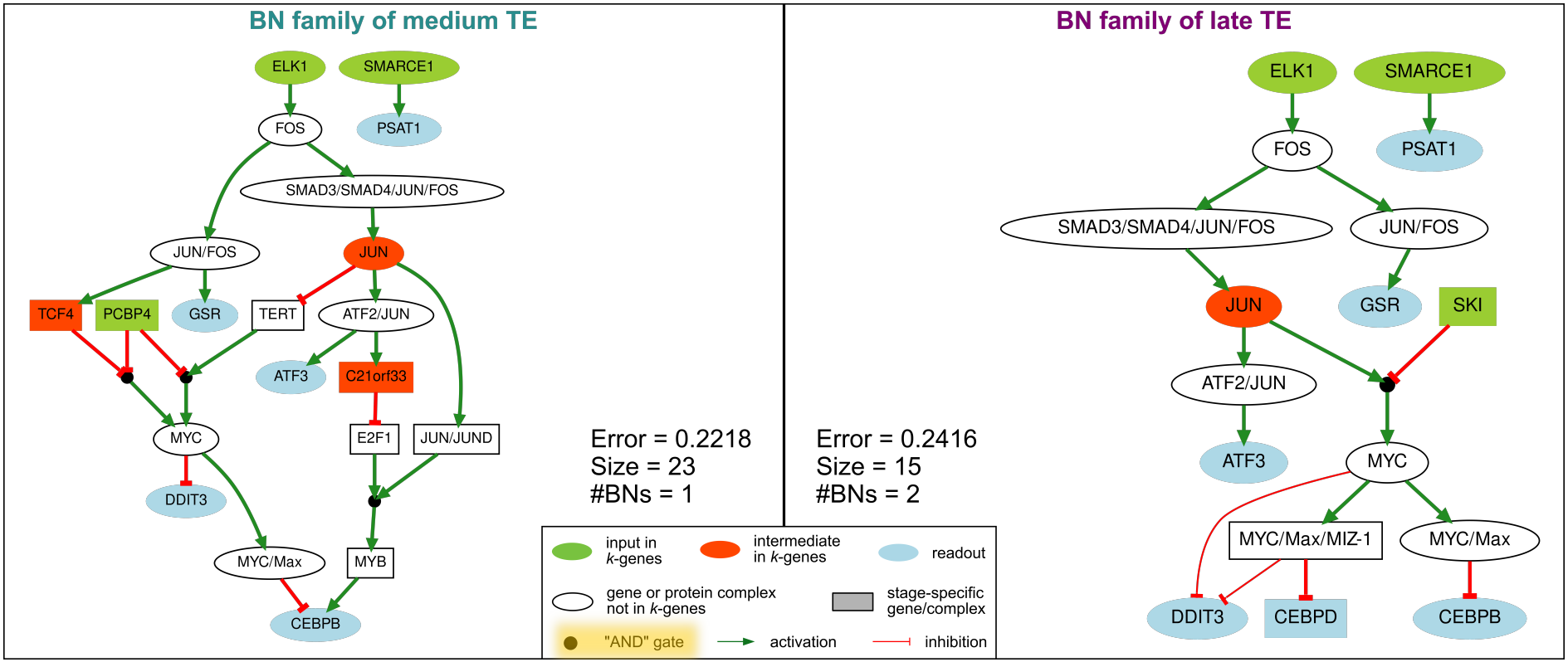
[Bolteau et al., *in prep.*]

Regulatory mechanisms in the learned models



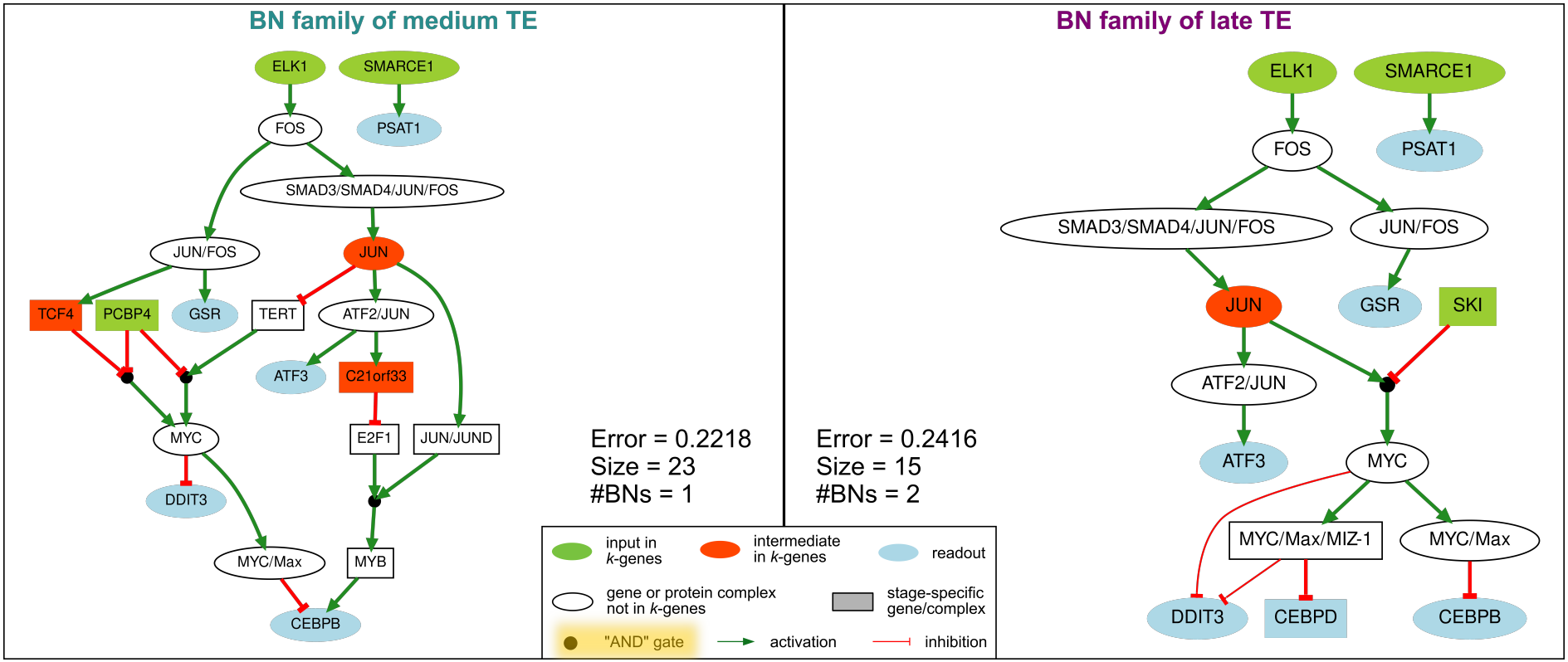
[Bolteau et al., *in prep.*]

Regulatory mechanisms in the learned models



[Bolteau et al., *in prep.*]

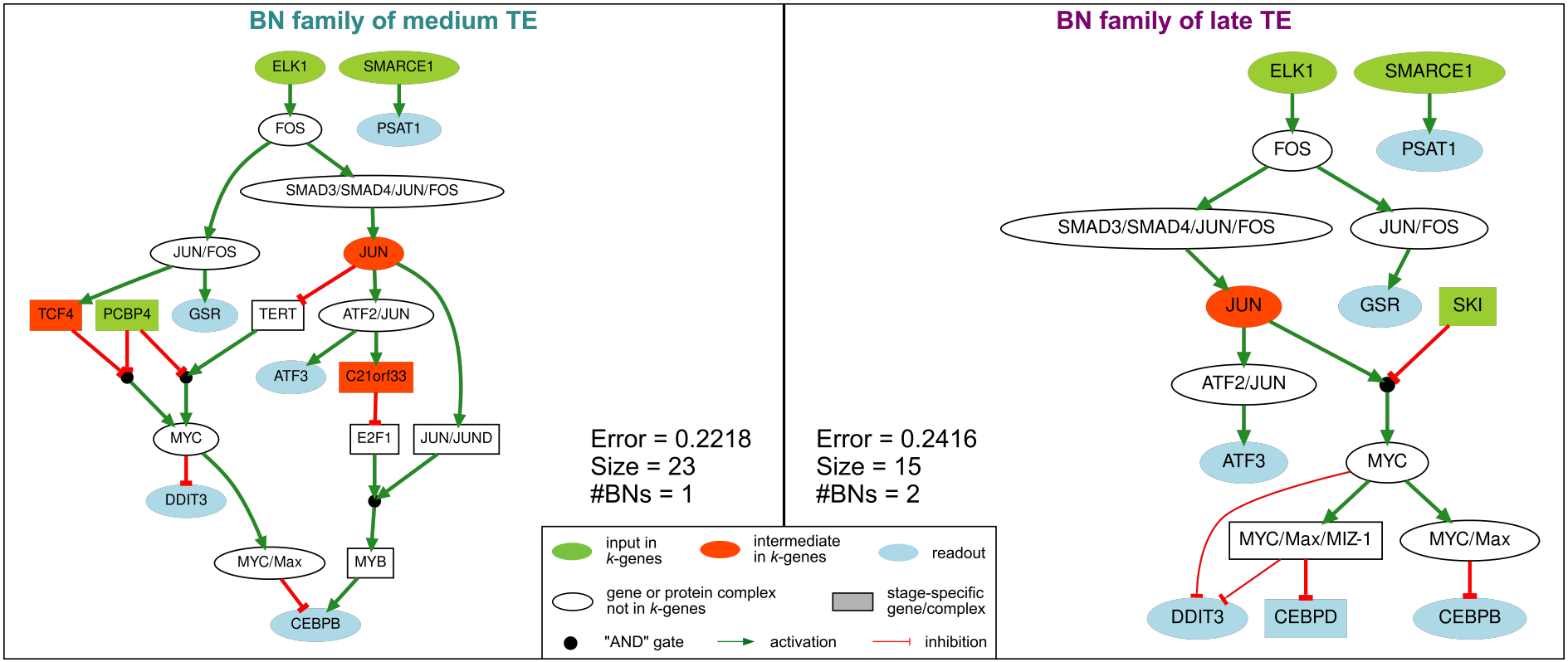
Regulatory mechanisms in the learned models



$$\mathrm{MYC} = (\neg \mathrm{TCF4} \wedge \neg \mathrm{PCBP4}) \vee (\neg \mathrm{PCBP4} \wedge \mathrm{TERT})$$

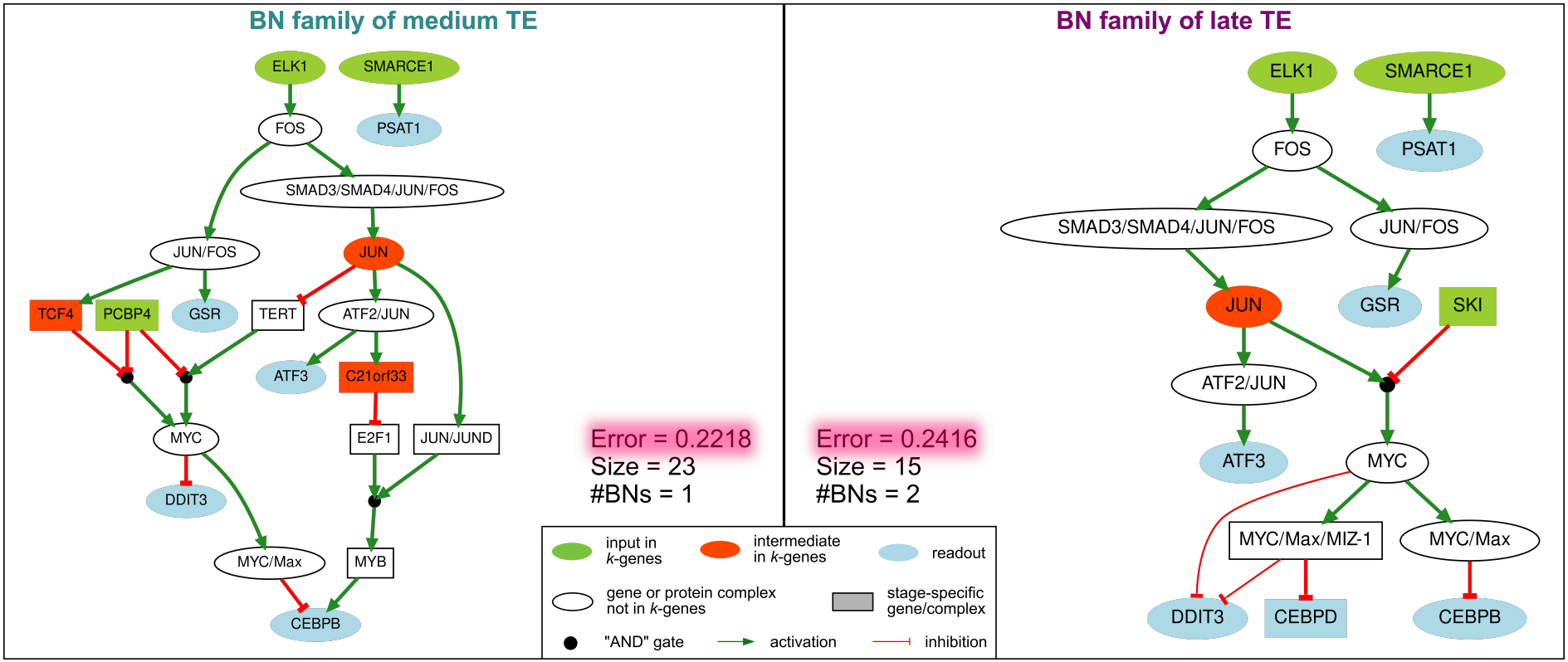
[Bolteau et al., *in prep.*]

Regulatory mechanisms in the learned models



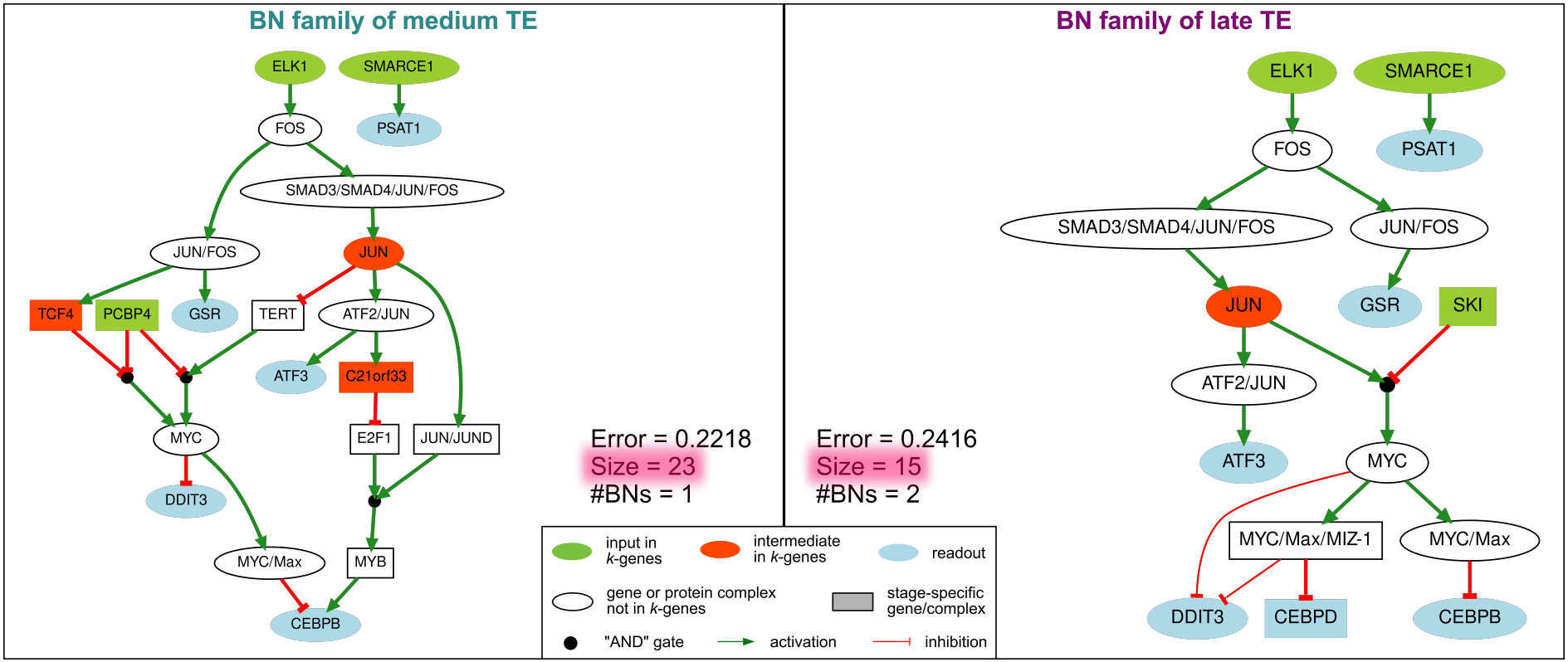
[Bolteau et al., *in prep.*]

Regulatory mechanisms in the learned models



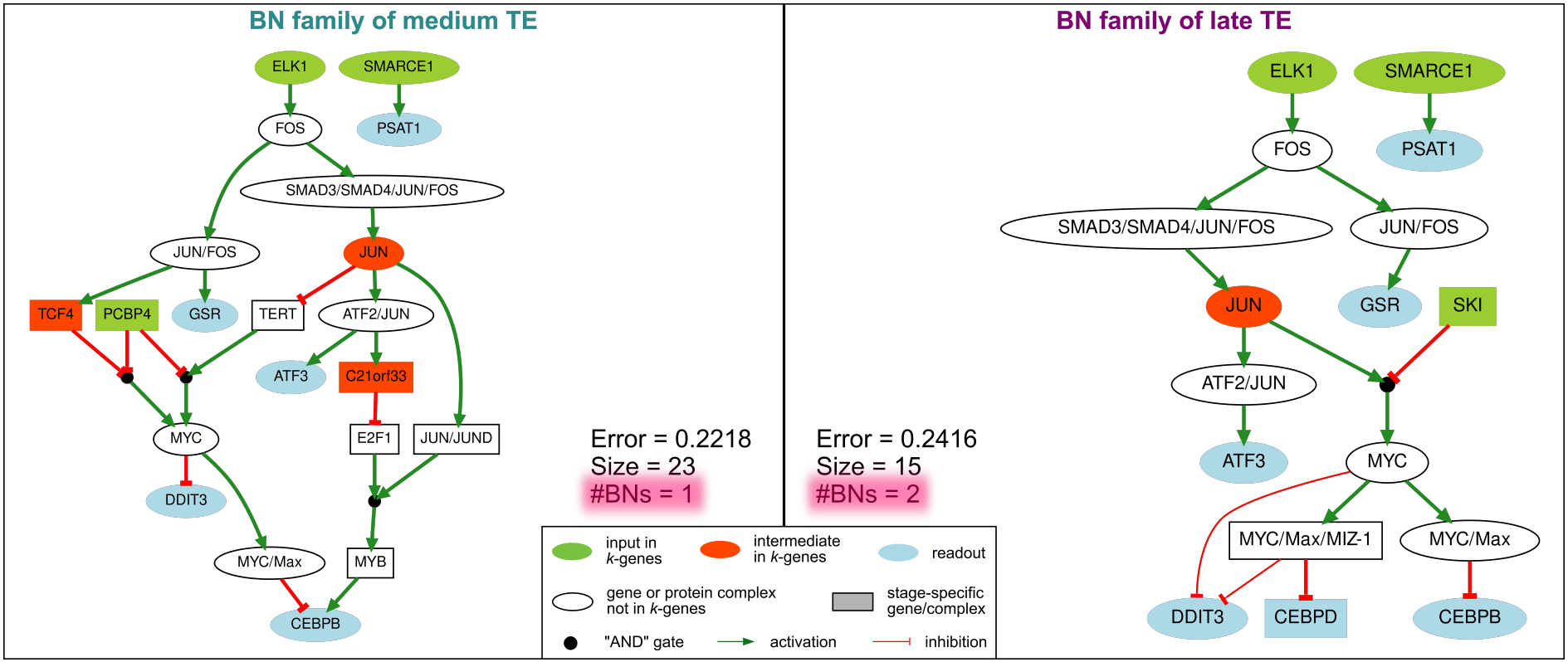
[Bolteau et al., *in prep.*]

Regulatory mechanisms in the learned models



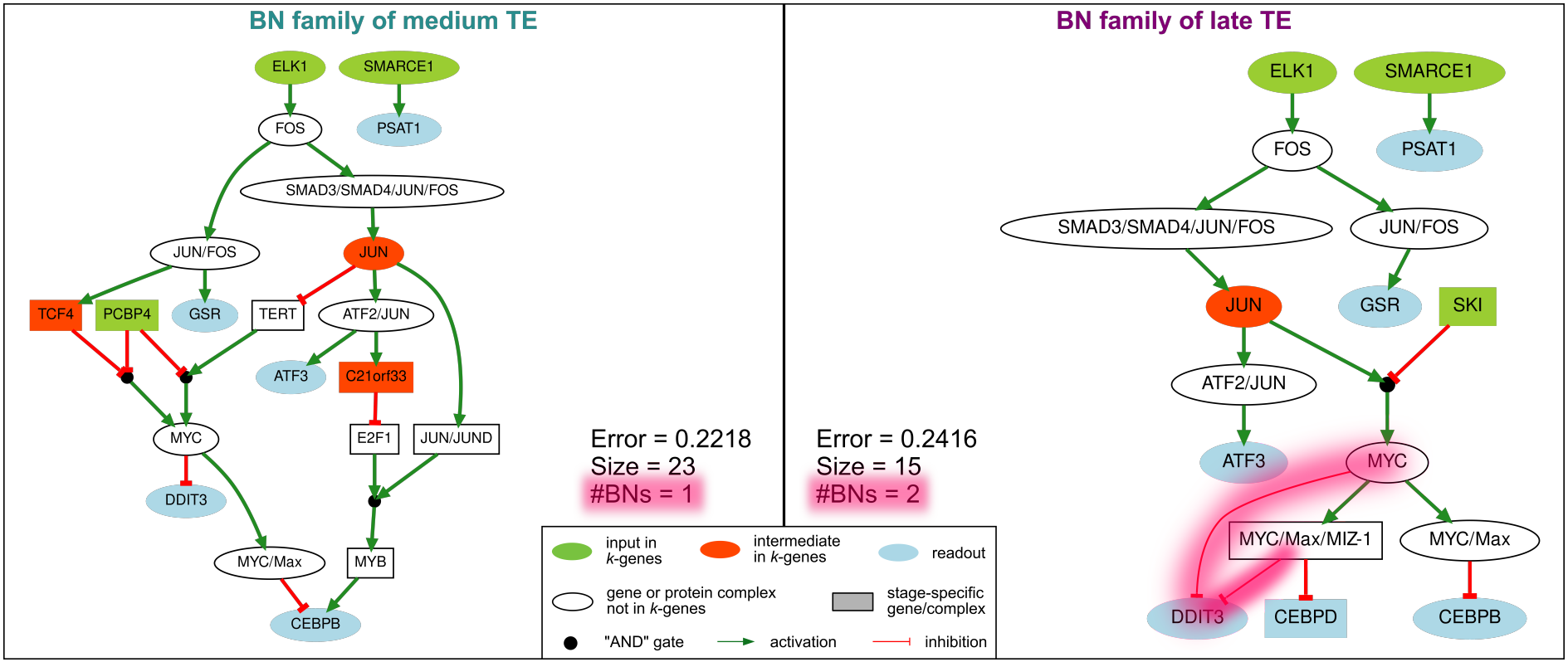
[Bolteau et al., *in prep.*]

Regulatory mechanisms in the learned models



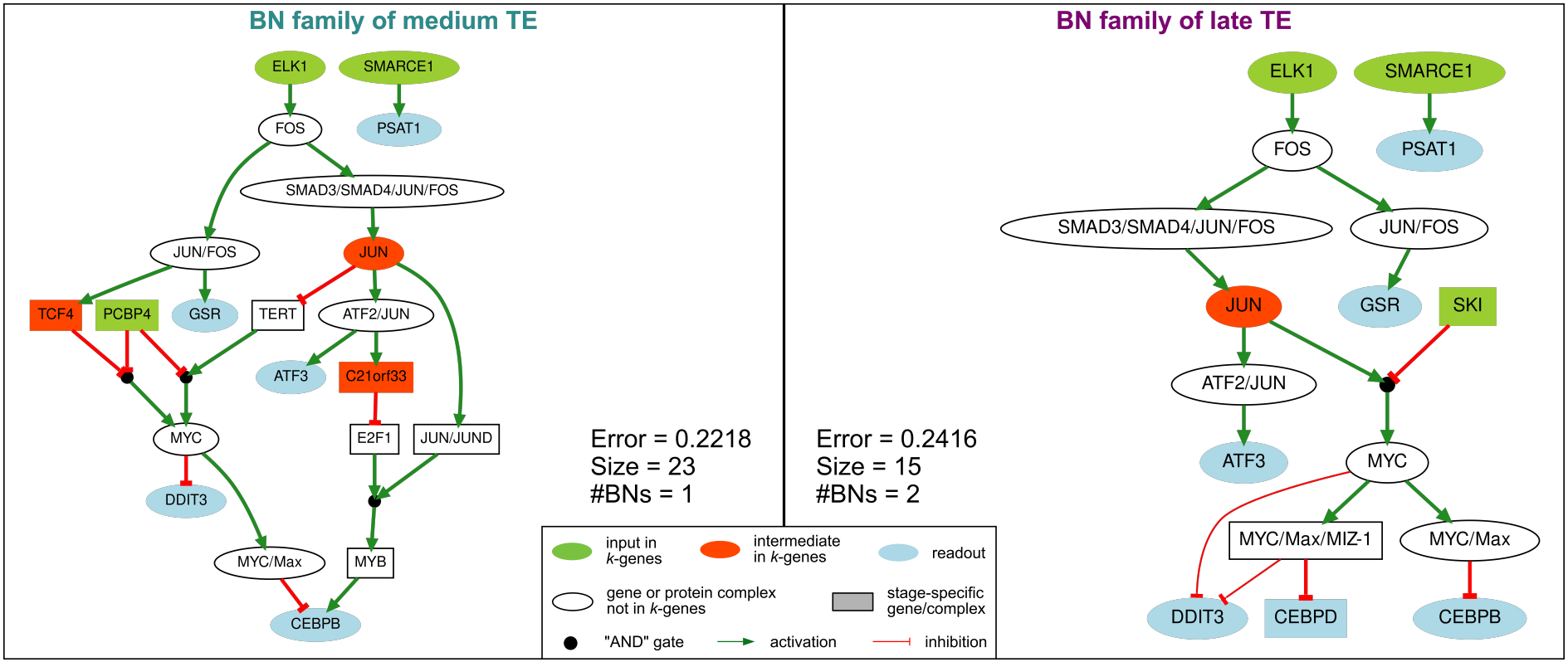
[Bolteau et al., *in prep.*]

Regulatory mechanisms in the learned models



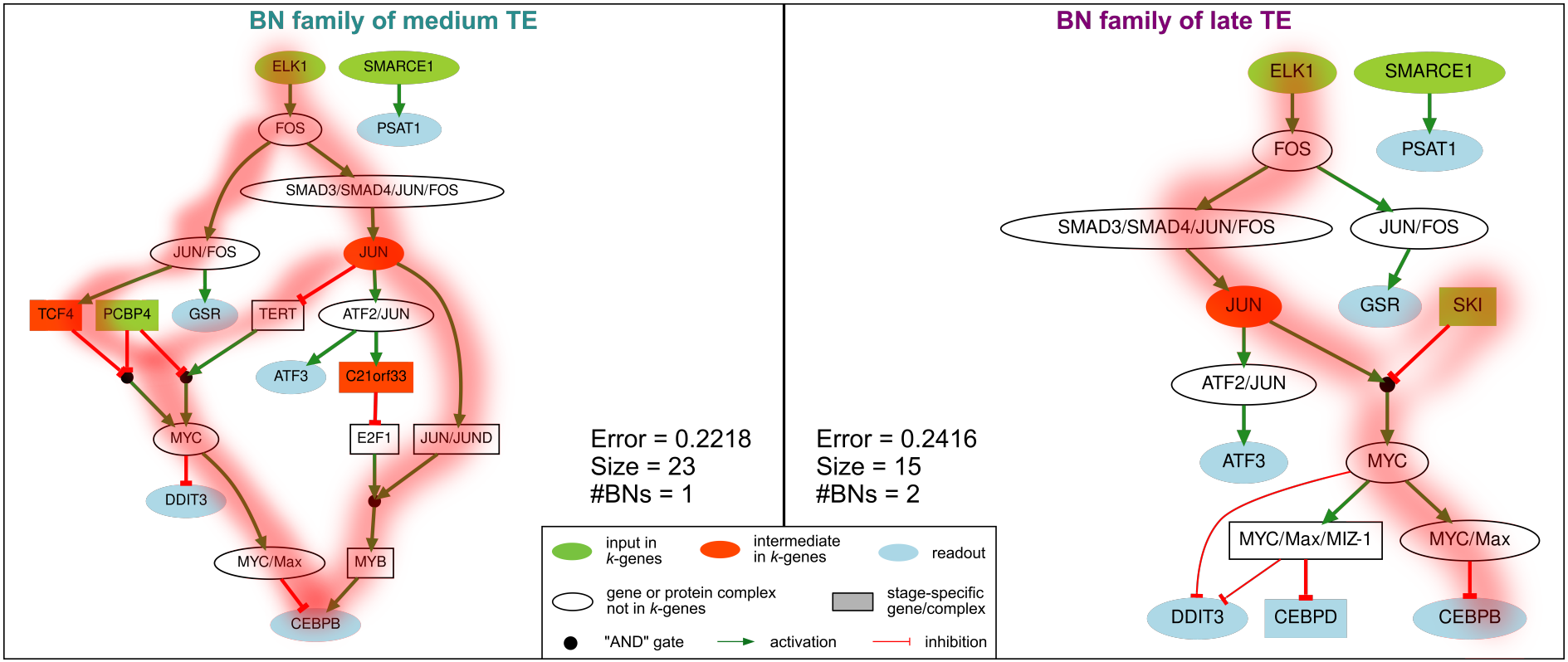
[Bolteau et al., *in prep.*]

Regulatory mechanisms in the learned models



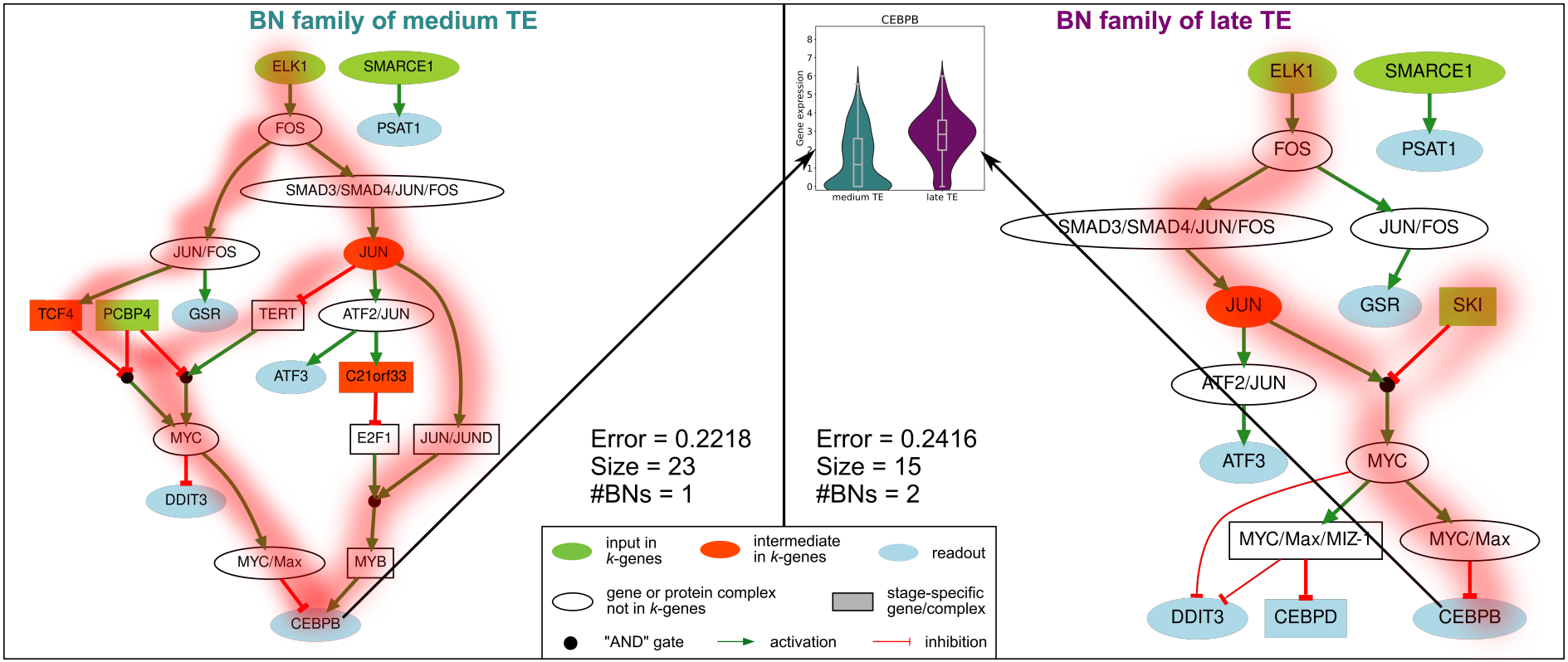
[Bolteau et al., *in prep.*]

Regulatory mechanisms in the learned models



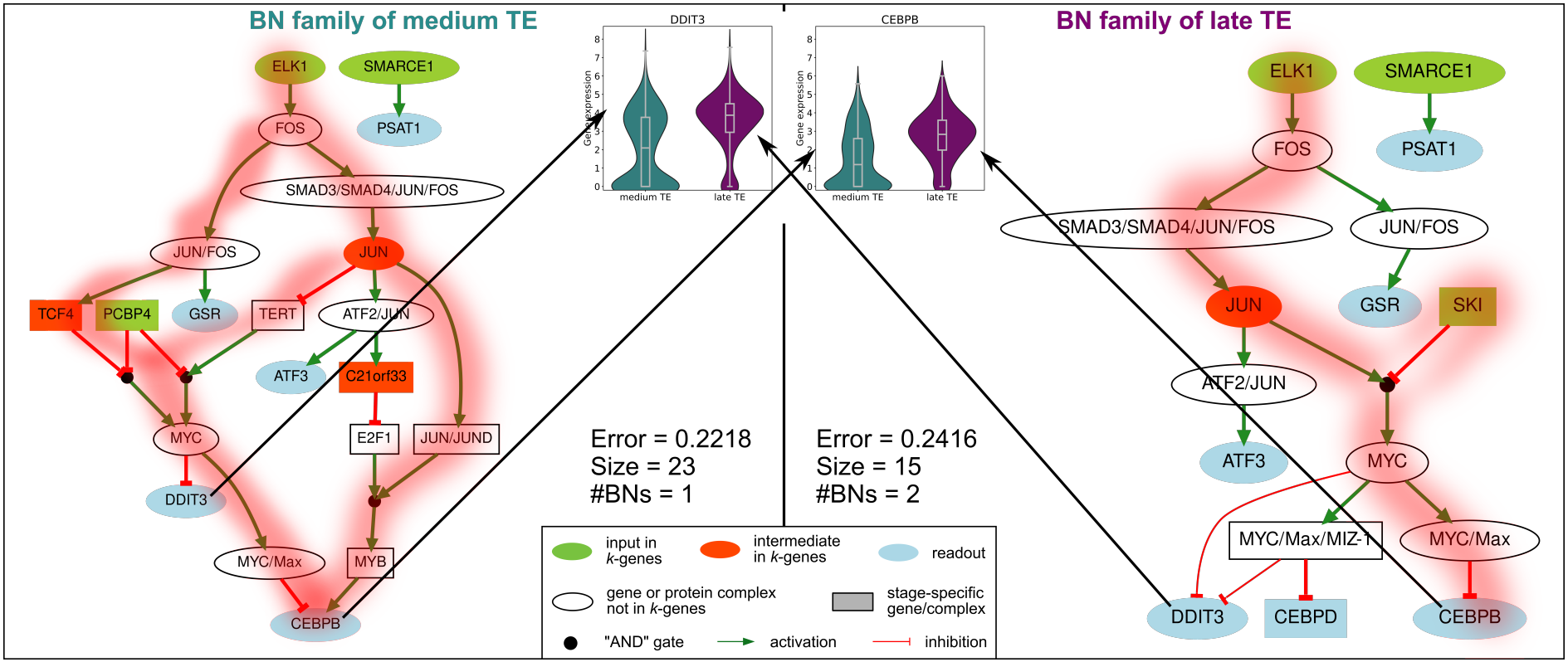
[Bolteau et al., *in prep.*]

Regulatory mechanisms in the learned models



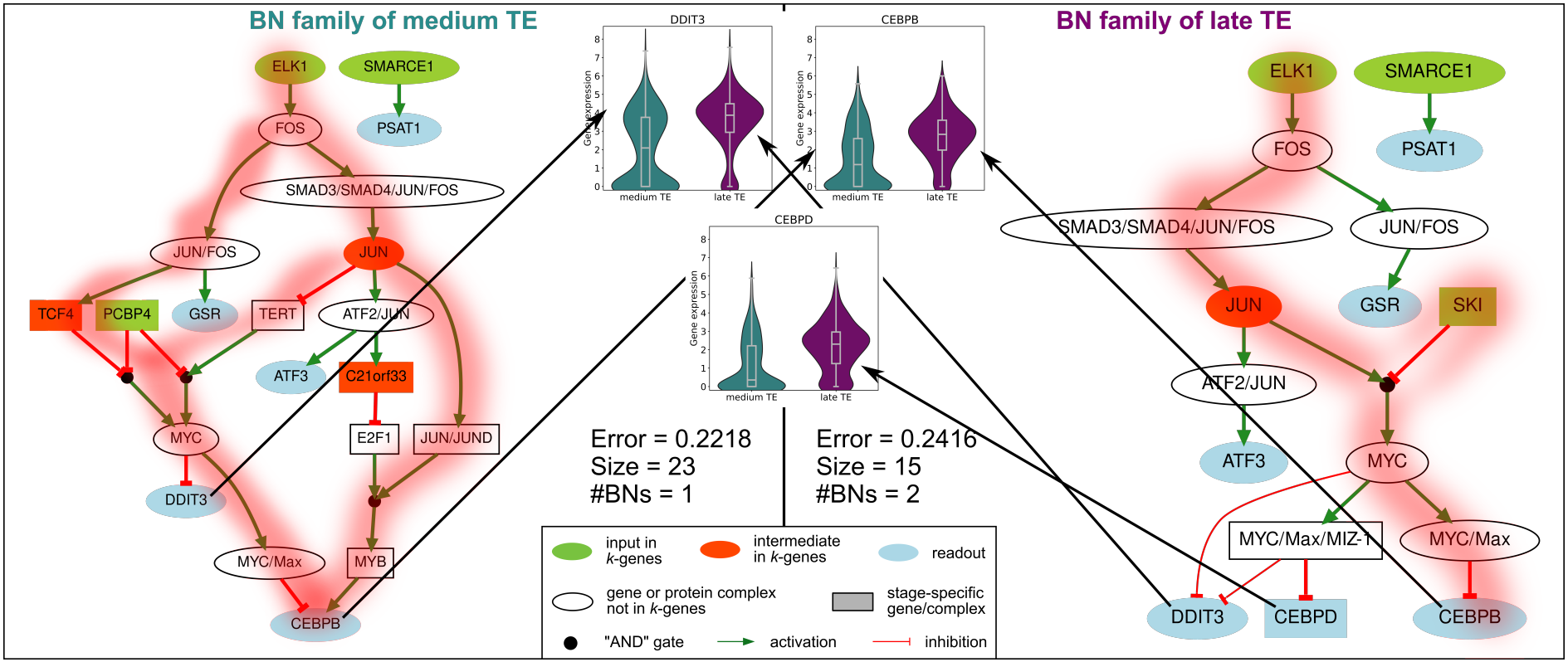
[Bolteau et al., *in prep.*]

Regulatory mechanisms in the learned models



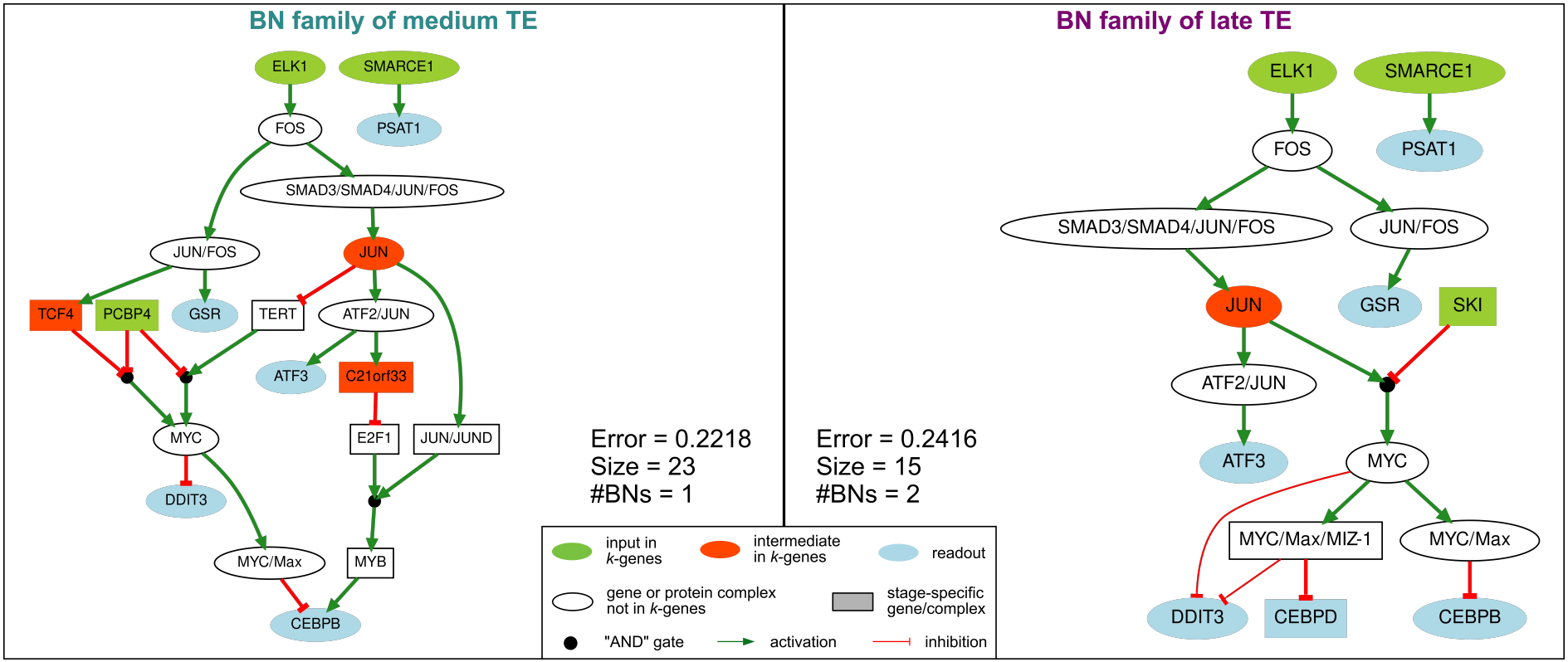
[Bolteau et al., *in prep.*]

Regulatory mechanisms in the learned models



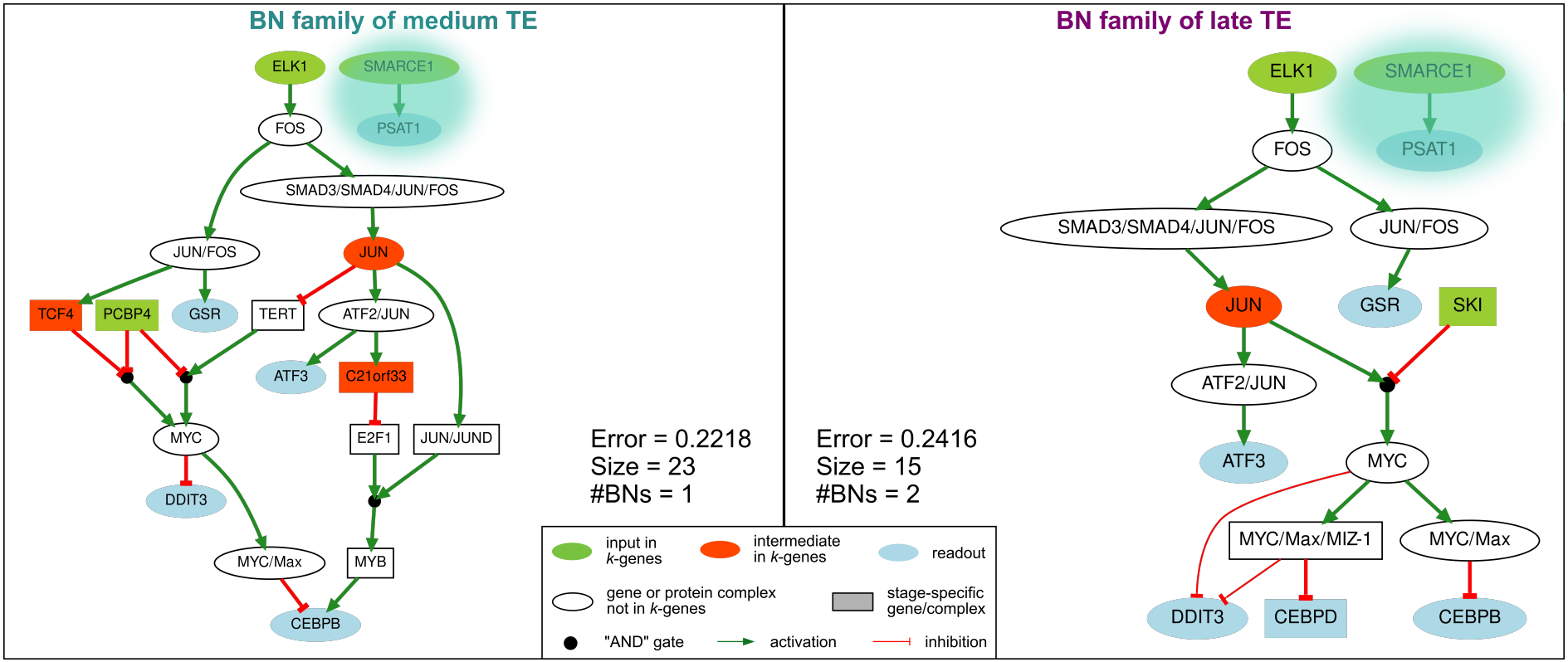
[Bolteau et al., *in prep.*]

Regulatory mechanisms in the learned models



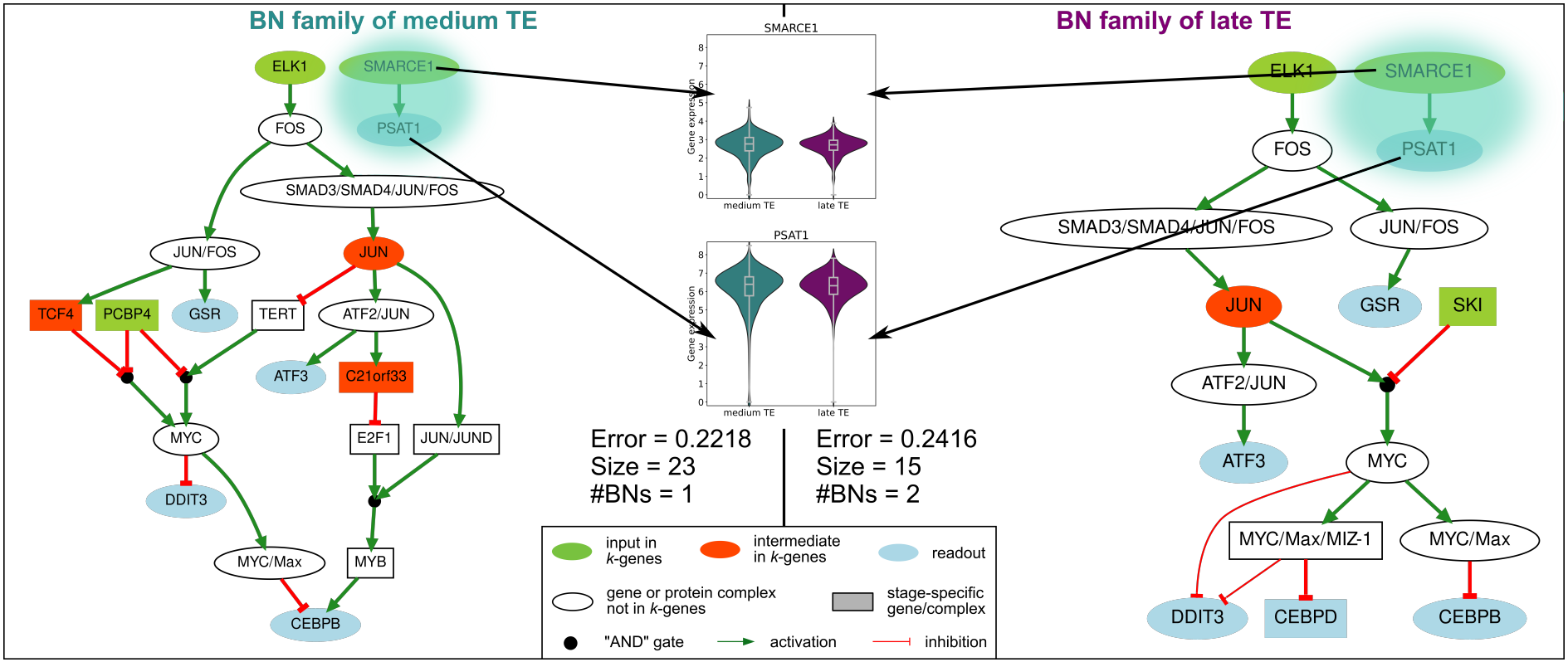
[Bolteau et al., *in prep.*]

Regulatory mechanisms in the learned models



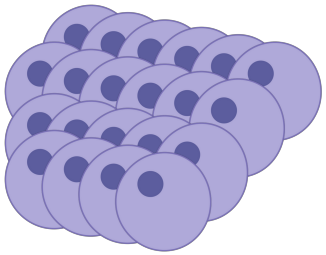
[Bolteau et al., *in prep.*]

Regulatory mechanisms in the learned models

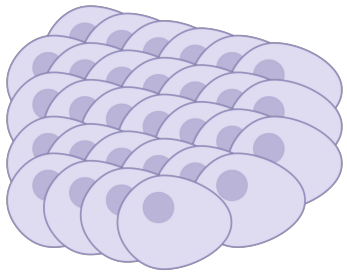


[Bolteau et al., *in prep.*]

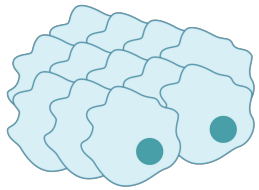
Validation of the learned models: Cell classifier



Pseudo-perturbation cells
(training set)



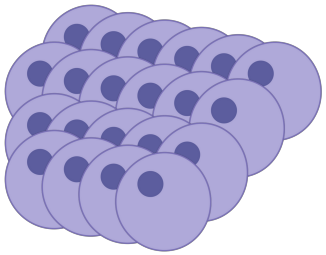
Redundant cells
(testing set 1)



Other cells
(testing set 2)

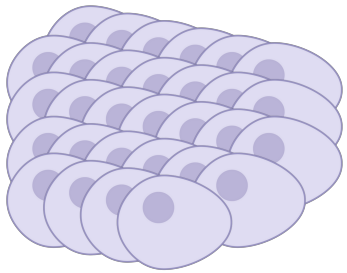
[Bolteau et al., *in prep.*]

Validation of the learned models: Cell classifier



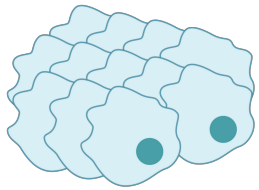
Pseudo-perturbation cells
(training set)

Accuracy	BAC	Medium TE accuracy	Late TE accuracy
72 %	72 %	64 %	81 %



Redundant cells
(testing set 1)

Accuracy	BAC	Medium TE accuracy	Late TE accuracy
66 %	67 %	52 %	81 %



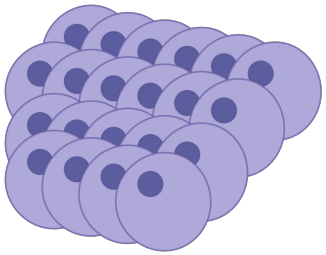
Other cells
(testing set 2)

Accuracy	BAC	Medium TE accuracy	Late TE accuracy
68 %	68 %	62 %	73 %

BAC = Balanced accuracy

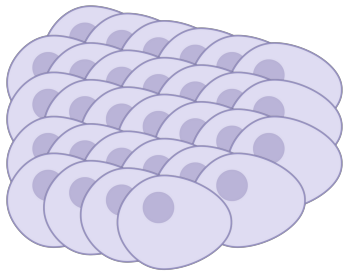
[Bolteau et al., *in prep.*]

Validation of the learned models: Cell classifier



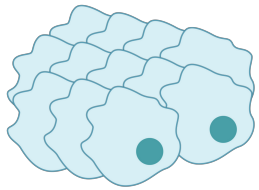
Pseudo-perturbation cells
(training set)

Accuracy	BAC	Medium TE accuracy	Late TE accuracy
72 %	72 %	64 %	81 %



Redundant cells
(testing set 1)

Accuracy	BAC	Medium TE accuracy	Late TE accuracy
66 %	67 %	52 %	81 %



Other cells
(testing set 2)

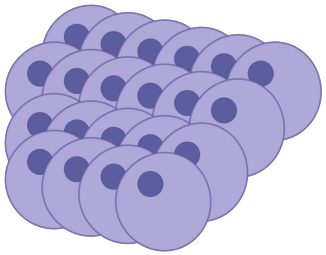
Accuracy	BAC	Medium TE accuracy	Late TE accuracy
68 %	68 %	62 %	73 %

SCIBORG learns accurate models

BAC = Balanced accuracy

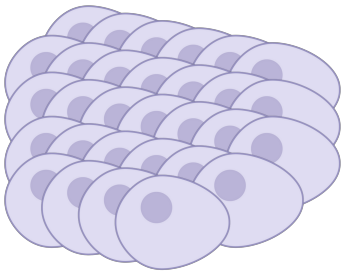
[Bolteau et al., *in prep.*]

Validation of the learned models: Cell classifier



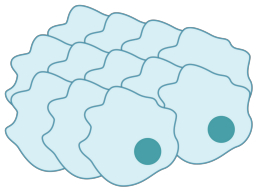
Pseudo-perturbation cells
(training set)

Accuracy	BAC	Medium TE accuracy	Late TE accuracy
72 %	72 %	64 %	81 %



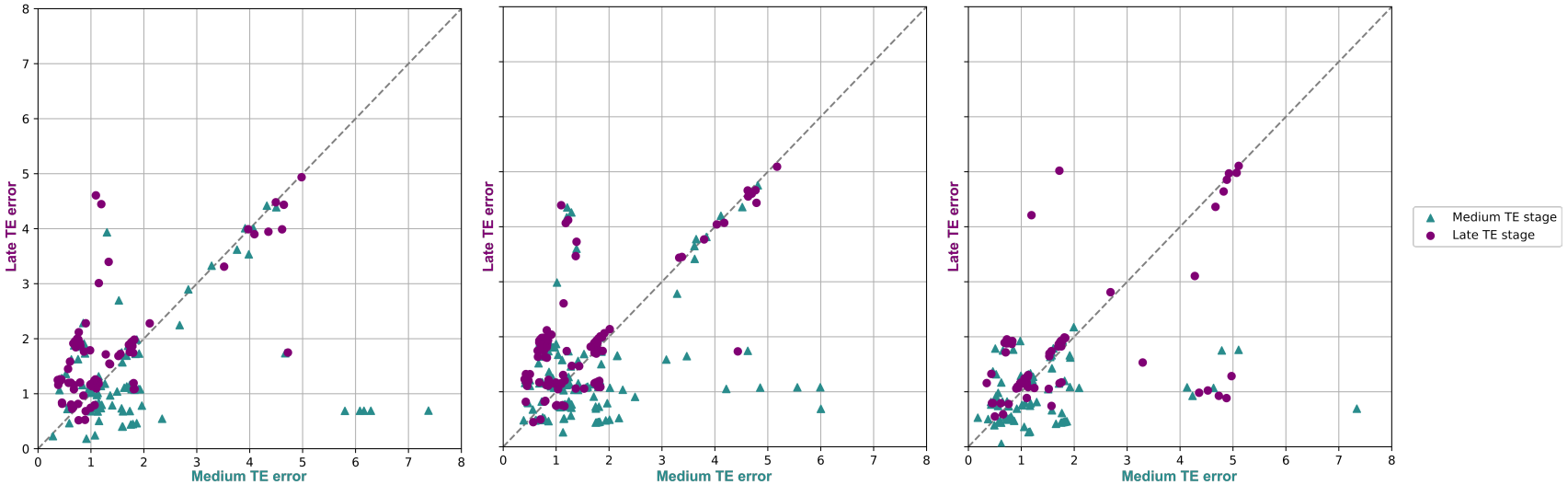
Redundant cells
(testing set 1)

Accuracy	BAC	Medium TE accuracy	Late TE accuracy
66 %	67 %	52 %	81 %



Other cells
(testing set 2)

Accuracy	BAC	Medium TE accuracy	Late TE accuracy
68 %	68 %	62 %	73 %

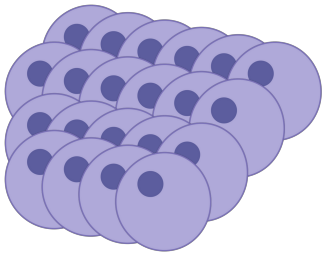


SCIBORG learns accurate models

BAC = Balanced accuracy

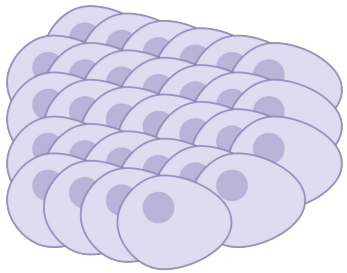
[Bolteau et al., *in prep.*]

Validation of the learned models: Cell classifier



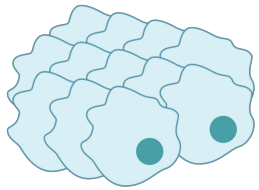
Pseudo-perturbation cells
(training set)

Accuracy	BAC	Medium TE accuracy	Late TE accuracy
72 %	72 %	64 %	81 %



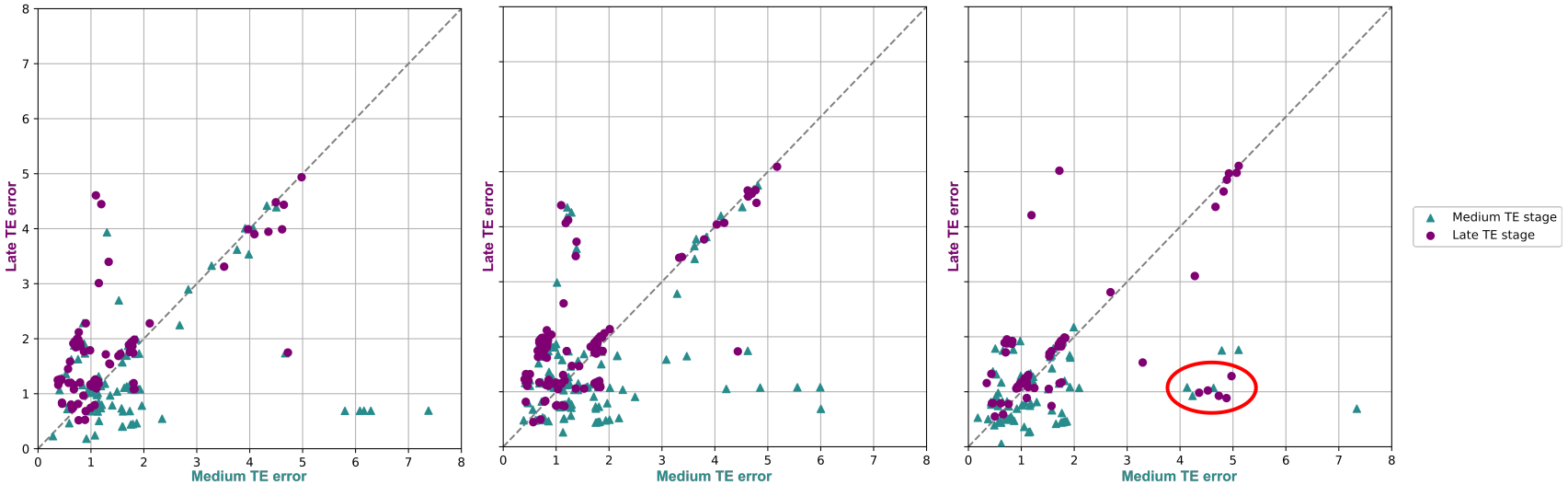
Redundant cells
(testing set 1)

Accuracy	BAC	Medium TE accuracy	Late TE accuracy
66 %	67 %	52 %	81 %



Other cells
(testing set 2)

Accuracy	BAC	Medium TE accuracy	Late TE accuracy
68 %	68 %	62 %	73 %



SCIBORG learns accurate models

BAC = Balanced accuracy

[Bolteau et al., *in prep.*]

Conclusion

Convergence of pseudo-perturbation identification

- Efficient program
- Limited number of equivalent solutions
- Pseudo-perturbations representativity

Conclusion

Convergence of pseudo-perturbation identification

- Efficient program
- Limited number of equivalent solutions
- Pseudo-perturbations representativity

Robustness in both equivalent solutions

- Distinct genes

Conclusion

Convergence of pseudo-perturbation identification

- Efficient program
- Limited number of equivalent solutions
- Pseudo-perturbations representativity

Robustness in both equivalent solutions

- Distinct genes

Gene regulatory mechanisms

- Distinguishing 2 stages
- Potential key genes

Conclusion

Complementarity with the state-of-the-art methods

Method	Cell heterogeneity	Cellular dynamic evolution	Exhaustive enumeration	Validation
SCIBORG		 *		

* Secondary objective

Conclusion

Complementarity with the state-of-the-art methods

Method	Cell heterogeneity	Cellular dynamic evolution	Exhaustive enumeration	Validation
SCIBORG	✓	✗*	✓	✓

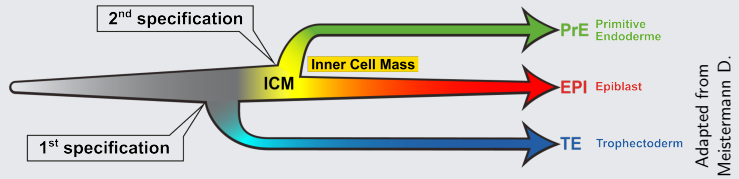
* Secondary objective

Publications

- Bolteau, M., Bourdon, J., David, L. Guziolowski, C., Inferring Boolean Networks from Single-Cell Human Embryo Datasets, *International Symposium of Bioinformatics and Research Applications (ISBRA)*, 2023.
- Bolteau, M., Chebouba, L., David, L., Bourdon, J. Guziolowski, C., Boolean Network Models of Human Preimplantation Development, *Journal of Computational Biology*, 2024.
- Bolteau, M., Bourdon, J., David, L. Guziolowski, C., *in prep.*
- Le Bars, S., Bolteau, M., Bourdon, J. Guziolowski, C., Predicting weighted unobserved nodes in a regulatory network using answer set programming, *BMC Bioinformatics*, 2023.

Perspectives

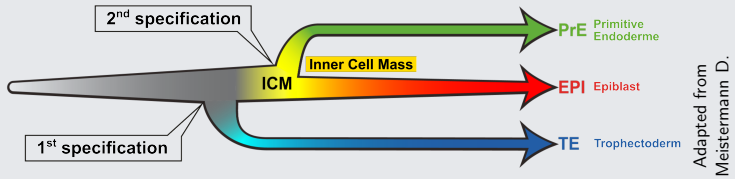
Study other developmental stages via exploration other cell fates (PrE, EPI)



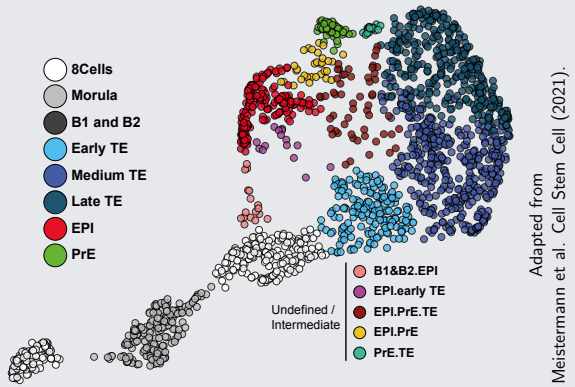
PrE = primitive endorderm ; EPI = epiblast ; TE = trophectoderm

Perspectives

Study other developmental stages via exploration other cell fates (PrE, EPI)



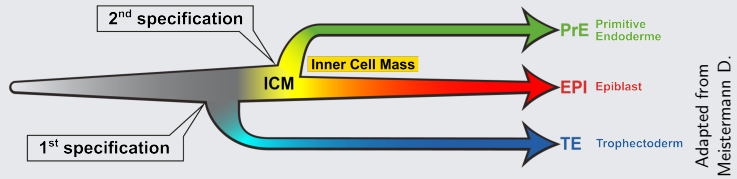
A tool for classifying undefined cells



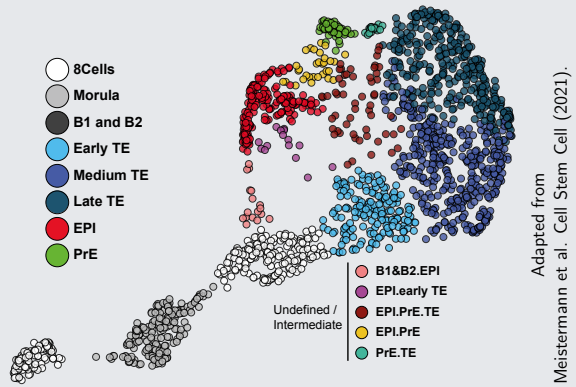
PrE = primitive endorderm ; EPI = epiblast ; TE = trophectoderm

Perspectives

Study other developmental stages via exploration other cell fates (PrE, EPI)



A tool for classifying undefined cells



Biological validation using blastoids

PrE = primitive endorderm ; EPI = epiblast ; TE = trophectoderm

Perspectives

Apply SCIBORG on other biological studies	Inner lymphoid cell development (CRCI2NA lab)	Cell differentiation in Duchenne muscular dystrophy (TaRGeT lab)

Perspectives

Apply SCIBORG
on other biological studies

Inner lymphoid cell development (CRCI2NA lab)

Cell differentiation in Duchenne muscular dystrophy (TaRGeT lab)

Modeling dynamic processes
to deal with more than two stages
(caspo-ts)

[Razzaq et al., *PLOS Comp. Bio.* (2018)]

- Preliminary work has been conducted
  (Centrale Nantes students supervision)
- Work to be continued with another Ph.D. thesis

Logic programs to infer computational models of human embryonic development

Programmes logiques pour déduire des modèles informatiques du développement embryonnaire humain

Mathieu Bolteau

Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France

Ph.D. Defense – Friday, October 4th 2024

ANR BOOSTIVF