Boolean Networks as a Framework to Model Human Preimplantation Development

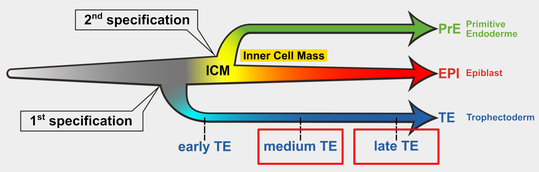
Mathieu Bolteau[1], Jérémie Bourdon[1], Laurent David[2] and Carito Guziolowski[1]

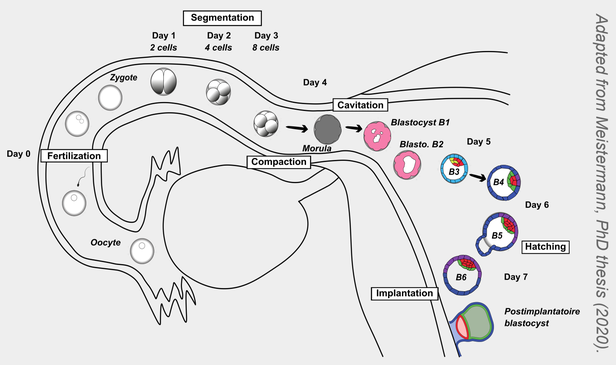
[1] Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France
[2] Nantes Université, CHU Nantes, INSERM, Center for Research in Transplantation and Translational Immunology, UMR 1064, F-44000 Nantes, France

We have developed an innovative method that combines **single-cell RNA sequencing (scRNAseq)** data and **prior biological knowledge** to accurately **infer Boolean networks (BNs)** in the context of human embryonic development. By integrating gene expression data and addressing computational challenges associated with heterogeneous scRNAseq data, our method sheds light on the **regulatory interactions that drive cellular decisions** during embryonic development. In contrast to existing statistical tools like pseudo-time analysis [1] or modeling methods [2,3], our approach allows for the distinction of different developmental stages by **identifying stage-specific regulatory mechanisms,** in the form Boolean network families, which consider heterogeneous and multiple cellular gene expression at each stage without the need for perturbations in the system.

HUMAN PRE-IMPLANTATION EMBRYONIC DEVELOPMENT

- Embryo goes through **different stages** during the development
- **Different cell types** contribute to the development
- **Two specifications** lead to **three distinct cell fates** (EPI, PrE, TE)

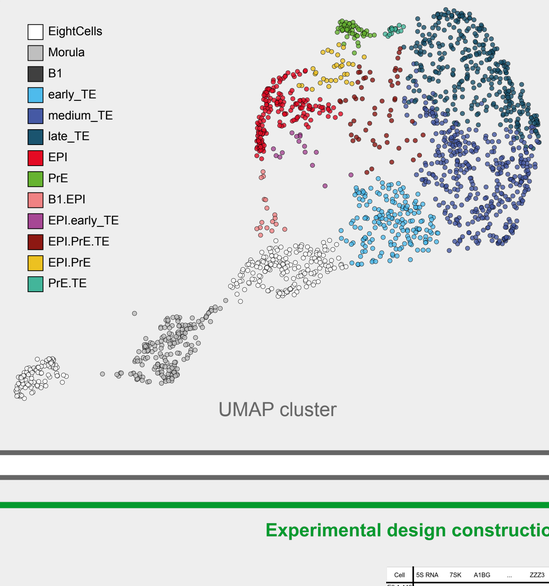




Adapted from Meistermann, PhD thesis (2020).

DATA

- **scRNAseq expression** for ~20,000 genes for ~1,700 cells from 128 multi-stage embryos.
- **Clustering cells** according to their cellular types [4]
- scRNAseq data specificities: **sparsity** and **redundancy**



Meistermann et al. Cell Stem Cell (2021).

OUR METHOD

1. PKN reconstruction

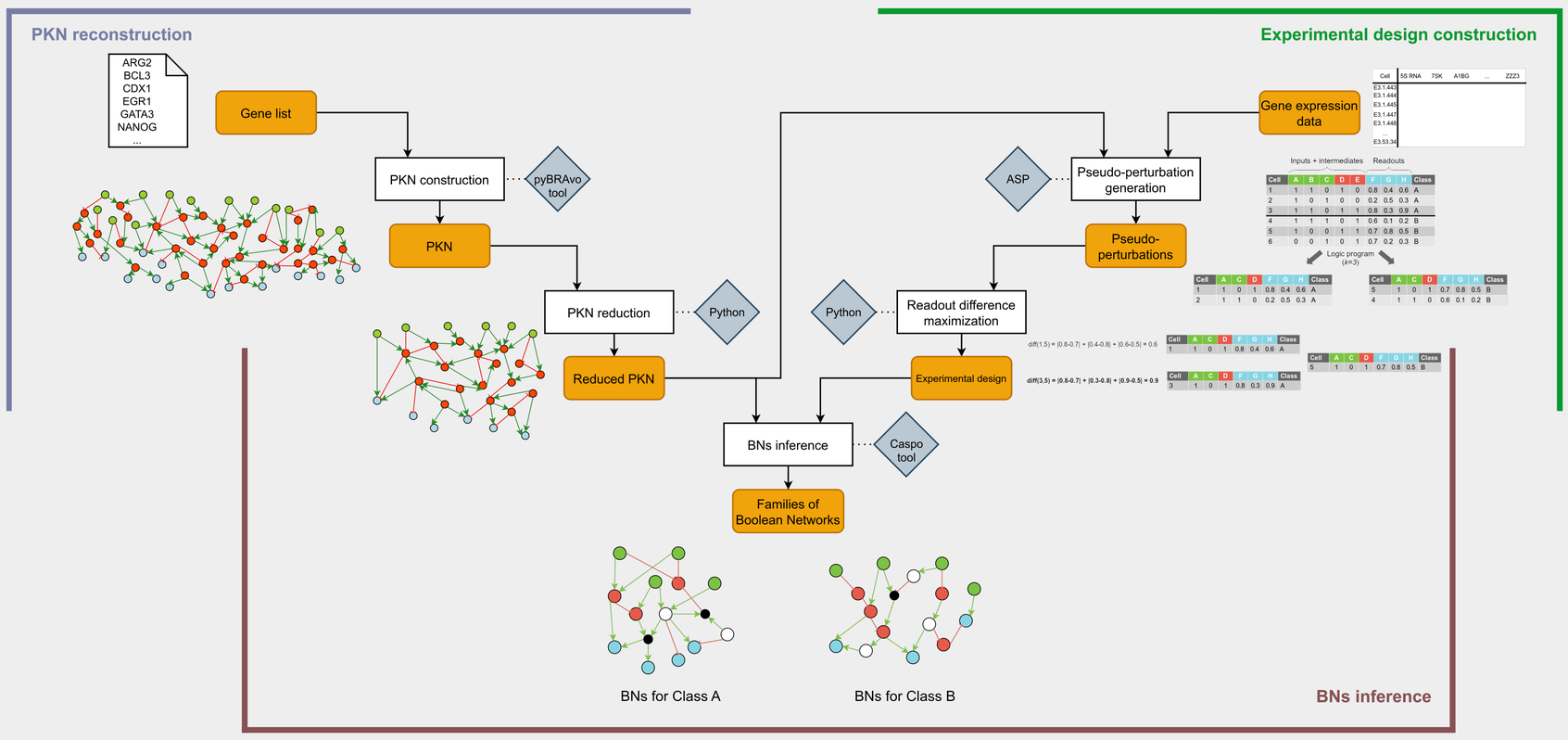
A Prior-Knowledge Network (**PKN**) is reconstructed using pyBRAvo [5] from a list of genes. This PKN is then reduced according to the scRNAseq data.

2. Experimental design construction

Given gene expression data of cells belonging to two classes, an **ASP program calculates pseudo-perturbations** for selected genes and cells. Pseudo-perturbations are used to maximize the readout differences; the output of this process is the optimal **experimental design**.

3. BNs inference

Caspo [6] is used to infer, given a PKN and an experimental design, **specific BNs to each class**.



APPLICATION ON THE TROPHECTODERM MATURATION

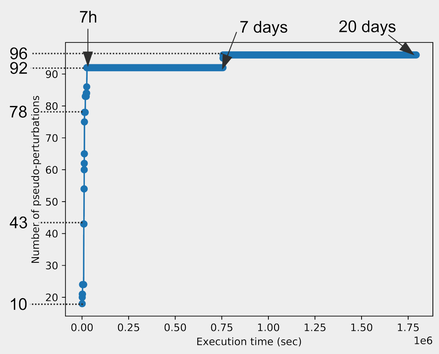
Objective: distinguish regulatory mechanisms between medium and late trophectoderm (TE).

Reconstructed PKN

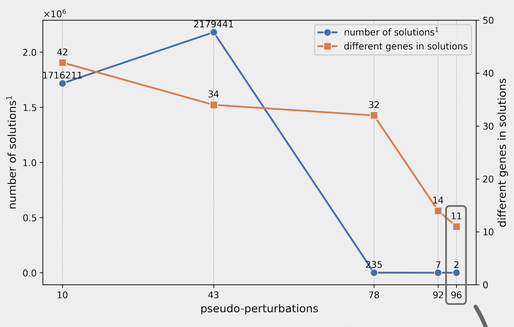
- Input: 438 transcription factors
- Output: PKN comprising
  - 233 nodes: inputs (85), intermediates (36), readouts (19)
  - 369 edges

Pseudo-perturbations search

$k=10 \rightarrow \binom{121}{10} = 1.24 \times 10^{14}$ possible choices



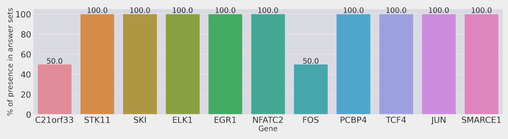
Robustness of solutions



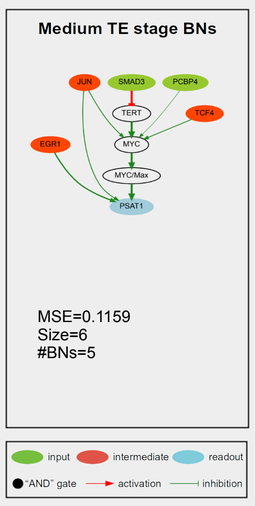
[1] 7 days of run on a computer cluster comprising 160 CPUs and 1.5 To of RAM

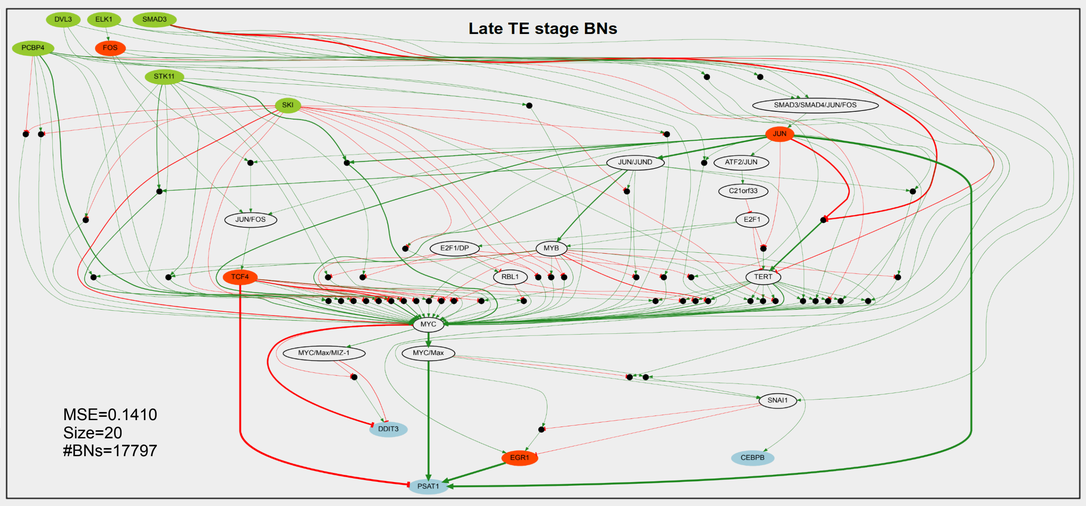
96 different pseudo-perturbations sub-optimal solution

Number of solution = 2
Different genes in solutions = 11



Inferred Boolean Networks (BNs)





- More readouts implicated in late TE stage
- Greater BNs variability for late TE → Gain of function
- Late TE seems more unstable → Transition from late TE to another stage

CONCLUSION

- **Efficient algorithm** to select cells and genes to generate pseudo-perturbations → 92 different pseudo-perturbations in 7h
- **Robustness** of the generated solutions → from +2 millions of solutions to only 2
- Discovery of **11 genes whose on/off values remain identical** for 96 cells across 2 classes
- A new method that deals **with single-cell data and its specificities** (redundancy and sparsity)
- A method that **learns Boolean networks** of 2 stages using scRNAseq data and Prior Knowledge and identify **mechanisms of TF-gene regulations** distinguishing 2 developmental stages

REFERENCES

[1] Qiu *et al. Nat Methods* (2017).
[2] Chevalier *et al. ICTAI* (2019).
[3] Dunn *et al. EMBO journal* (2019).
[4] Meistermann *et al*. *Cell Stem Cell* (2021).
[5] Lefebvre *et al*. *Database* (2021).
[6] Videla *et al*. *Bioinformatics* (2017).

This work was co-supported by ANR AIBY4 (ANR-20-THIA-0011) and ANR BOOSTIVF (ANR-20-CE17-0007) projects.

mathieu.bolteau@ls2n.fr
@MathieuBolteau