Logic programs to infer computational models of human pre-implantation embryonic development

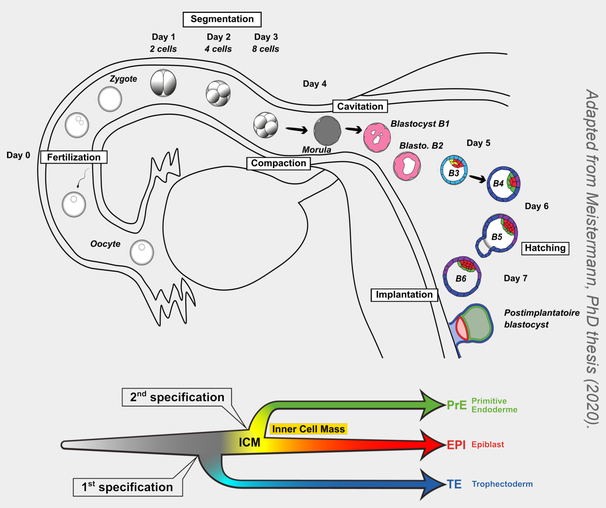
Mathieu Bolteau[1], Jérémie Bourdon[1], Laurent David[2] and Carito Guziolowski[1]

[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

Our project is to develop a **computational model** of the dynamics of a complex biological system using **single-cell RNA-Seq data**. We apply our work on the modeling of human pre-implantation embryonic development. *In vitro* fertilization (IVF) allows infertile couples to have babies, but **only 25% of IVF cycles are successful**. Because of the limited access to human embryos, the field of IVF needs *in silico* models in order to improve our understanding of embryo development. Ultimately, our model would allow to predict how embryos respond to specific perturbations of the system, such as changes in the culture media composition.

HUMAN PRE-IMPLANTATION EMBRYONIC DEVELOPMENT

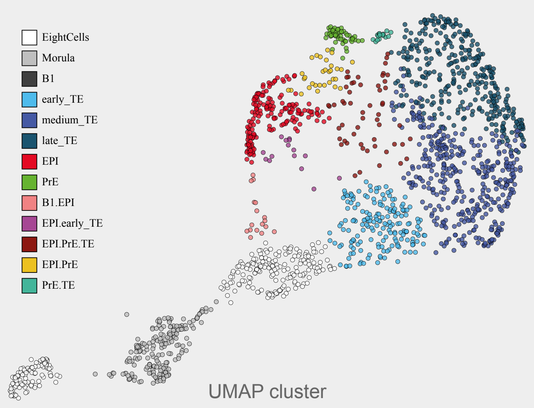
The embryo goes through different stages of development. The cells that compose it are of different types and play different roles. Many events occur during development, including two specifications that lead to three cell fates (EPI, PrE, TE).



Adapted from Meistermann, PhD thesis (2020).

DATA

We use single-cell RNA-seq data of cells extracted from different embryos at different developmental stages. The dataset is composed of the expression of ~20,000 genes for an atlas of ~1,700 cells from 128 embryos. Cells are clustered according to their cellular types [1].

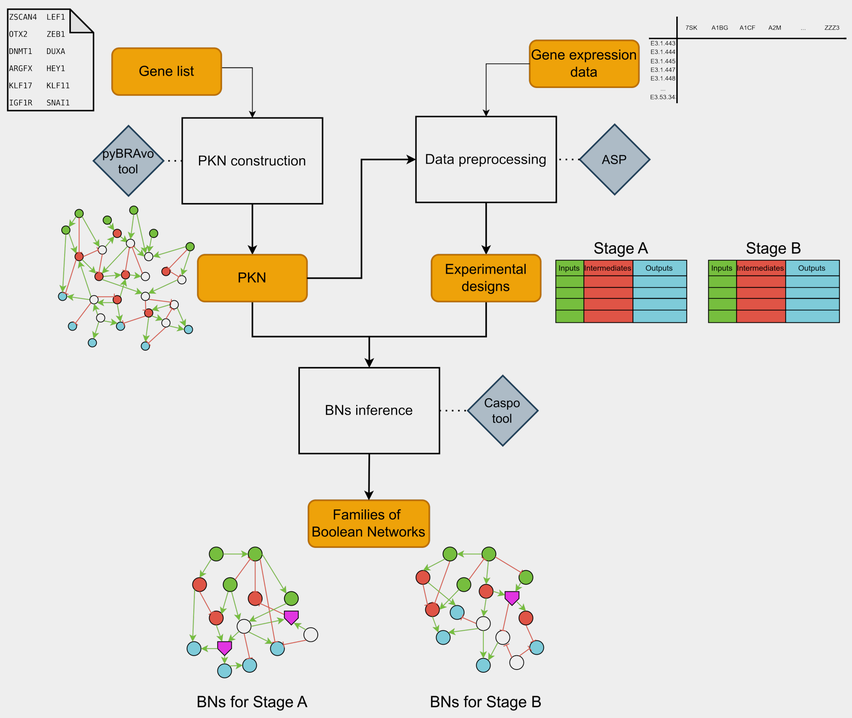


Meistermann et al. Cell Stem Cell (2021).

OBJECTIVES

1. Infer Boolean models for each developmental stage
2. Integrate dynamics between each Boolean networks (BNs) to model the pre-implantation embryonic development

OUR PROPOSED METHOD



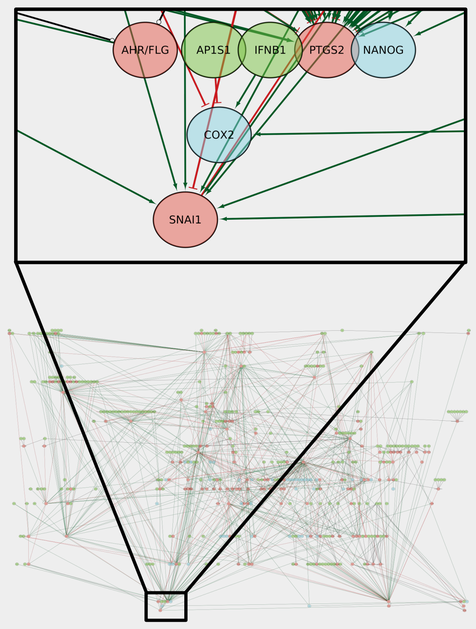
PKN CONSTRUCTION

From a list of genes the aim is to reconstruct automatically a **PKN** (Prior-Knowledge Network), composed of gene interactions, using the software pyBRAvo [6] which queries the Pathway Commons database.

Preliminary results:

	pyBRAvo depth		
	2	3	4
Nodes	463	752	890
Edges	912	1923	2563
Nodes without predecessors	271	304	292
Nodes without successors	49	56	63

Nodes without predecessors are noted as **inputs** whereas nodes without successors are noted as **outputs**. The remaining nodes are considered as **intermediates**.

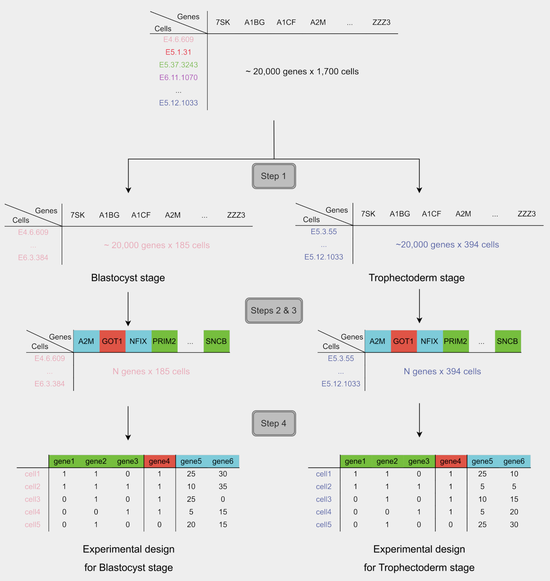


PKN generated by pyBRAvo (depth=2)

DATA PREPROCESSING

This step consists of inferring binarized pseudo-perturbations and discrete readouts from the scRNA-seq data. Using a method based on Answer-Set Programming [7], we construct experimental designs for two selected developmental stages. This preprocessing follows the steps below:

1. Isolate cells of two chosen developmental stages
2. Only genes included in the PKN will be kept. Previous matrices are reduced.
3. Binarize possible **inputs** and **intermediates** genes values. Discretize **output** genes values.
4. Reduce the dimension of discrete matrices using logic programming. Keep cells where, for both matrices, **inputs** and **intermediates** binarized measurements are identical. Maximize differences between the **output** measurements for both matrices. These two matrices constitute experimental designs.



BOOLEAN NETWORK INFERENCE

The last step of our method has the objective of learning families of BNs compatible with a PKN and experimental designs. Using Capso [8], implemented in Answer-Set Programming, BNs are inferred for both development stages, modeling respectively each stage. Some post-analyses could be performed, such as identify similarities or key genes for instance.

RELATED WORKS

Different groups are interested in the dynamics of the embryonic development using scRNA-Seq data with various perspectives. Statistical methods using manifolds [2] or graphs [3] allow to extract the temporal aspects hidden in the scRNA-Seq data. Additionally, other groups focus on the modeling side of the development process, using logic programming solver [4] or using satisfiability modulo theories solvers [5].

PERSPECTIVES

Our proposed method needs to be fully implemented. Currently, we have proceeded to the PKNs construction step which allows us to have at our disposal some PKNs that can be used to test our method. The following steps will be:

- Implementation and test of the data preprocessing and BNs inference step
- Infer a Boolean model for each developmental stages

To answer the second objective, the challenge is the integration of dynamics between each Boolean models to model the pre-implantation embryonic development.

REFERENCES

[1] Meistermann *et al*. *Cell Stem Cell* (2021).
[2] McInnes *et al. arXiv (2018).*
[3] Qiu *et al*. Nature Methods (2017).
[4] Chevalier *et al*. *IEEE – ICTAI* (2019).
[5] Shavit *et al.* Biosystems (2016).
[6] Lefebvre *et al*. *Database* (2021).
[7] Chebouba *et al*. *BMC Bioinformatics* (2018).
[8] 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.

Contact : mathieu.bolteau@univ-nantes.fr