

Modélisation informatique du développement embryonnaire humain

1. Résumé du projet

Une meilleure compréhension du développement embryonnaire humain et des décisions concernant le devenir des cellules est nécessaire pour améliorer les technologies de procréation assistée, telles que la fécondation *in vitro*. Grâce aux nouvelles technologies en transcriptomique, une quantité considérable de données est générée, permettant la caractérisation d'embryons humains individuels au niveau unicellulaire. Ce projet vise à obtenir un modèle informatique permettant la discrimination des différents stades de développement embryonnaire, en tenant compte de ces données.

Concrètement, nous voudrions développer une méthode pour sélectionner d'abord les données d'expression de gènes concernant un sous-ensemble des cellules et des gènes à partir des données scRNAseq spécifiques à chaque stade de développement embryonnaire. Cette signature d'expression pourra être vue comme une dynamique de l'expression de certains gènes à travers les différents stades du développement. A partir de cette signature d'expression, nous souhaitons reconstruire des modèles Booléens qui vérifient cette dynamique d'expression. Finalement, nous souhaitons étudier la dynamique de ces modèles Booléens pour identifier des points d'attraction qui expliquent les différents destins cellulaires lors de l'évolution embryonnaire. Perturber *in-silico* un tel modèle Booléen, nous permettra de tester plusieurs hypothèses d'évolution cellulaire chez l'embryon humain.

2. Objectifs du projet

À ce jour, nous avons conçu une méthode utilisant la programmation logique pour interroger (1) les données scRNASeq liées au développement embryonnaire et (2) les bases de données publiques de connaissances sur la régulation génique, telles que Pathway Commons, et fournir en sortie des modèles booléens statiques qui expliquent la régulation génique en deux stades de développement embryonnaire : trophoctoderme moyen et tardif (Bolteau et al., 2023). Nous voulons maintenant aller de l'avant et modéliser des étapes consécutives et différents destins du développement embryonnaire. Les objectifs de ce projet de thèse sont :

- *Court terme.* Explorer d'autres étapes du développement embryonnaire. Pour cela, nous devons concevoir un programme logique pour extraire un ensemble de (pseudo) perturbations génétiques associées à l'expression discrète de gènes à travers les 9 stades de développement. Ces informations seront extraites à l'aide des ensembles de données scRNASeq pour toutes les étapes de développement. Une première ébauche du programme logique a été mise en œuvre. Cependant, il n'est pas encore appliqué à un environnement de données réel. Si ce programme réussit, la perspective devient très intéressante car nous pourrions appliquer une deuxième méthode, précédemment développée par nos soins dans (Razzaq et al. PloS Comp Biol 2018), pour obtenir des réseaux booléens dynamiques qui pourraient expliquer différentes trajectoires évolutives du développement embryonnaire.
- *Long terme.* Perturber le modèle informatique et modifier la dynamique d'un destin d'évolution donné. Une idée de méthode pour provoquer des perturbations dans ce système (également appelées perturbations), sera d'identifier les attracteurs dynamiques du système, puis de rechercher des ensembles de perturbations qui peuvent garantir l'arrivée à un destin plutôt qu'à un autre. Ceci est réalisable à mettre en œuvre avec des programmes logiques, dans une approximation. Cependant, d'autres types de solveurs de vérification peuvent

également être plus adaptés pour exclure les dynamiques faussement positives. Cette direction de recherche peut utiliser les résultats et les méthodes des études précédentes que nous avons proposées (Videla et al. Front. Bioeng. Biotechnol. 2015) (Fitime et al. Algorithms Mol Bio 2017), et celles proposées récemment par (Chevalier et al. LNBI , 2020), le tout utilisant la programmation logique.

3. Contexte

Cette thèse est financée par le programme I-SITE (Initiatives pour la science, l'innovation, les territoires et l'économie, label d'excellence des universités françaises) NEXt <https://next-isite.fr/>. Plus précisément sur le contexte pour favoriser les interactions entre sciences de l'ingénieur et santé. Le doctorant intégrera l'équipe ComBi du Laboratoire de Sciences du Numérique de Nantes (LS2N, <https://www.ls2n.fr/>), qui fédère les expertises de recherche nantaises en informatique et cybernétique pour développer les sciences du numérique, inclusives, des autres disciplines et en tenant compte des enjeux sociaux. L'équipe LS2N ComBi (Combinatoire et Bioinformatique) développe des méthodes algorithmiques et mathématiques pour l'étude de problèmes issus de la biologie. Les principaux thèmes de recherche de l'équipe portent sur la génomique comparée et la biologie des systèmes. Le contrat de doctorat se déroulera au sein de la grande école d'ingénieur, Centrale Nantes (<https://www.ec-nantes.fr/>). Cela offrira au doctorant la possibilité d'enseigner des séances de travaux pratiques au sein de cette institution. La date de début de cette thèse de doctorat est programmée à septembre 2024.

4. Prérequis

Le doctorant aura un profil Informatique ou Bioinformatique (Master ou équivalent) avec des connaissances en programmation logique ou en intelligence artificielle. Une expérience antérieure dans l'analyse (ou la modélisation informatique) d'ensembles de données massifs de nature biologique sera utile.

5. Candidature

N'hésitez pas à nous contacter si vous avez des questions concernant le projet, votre adéquation avec le profil ou la procédure de candidature à :

- carito[dot]guziolowski[at]ec-nantes[dot]fr
- laurent[dot]david[at]univ-nantes[dot]fr

Votre candidature doit contenir : (1) votre CV, (2) votre lettre de motivation précisant votre projet professionnel, (3) votre relevé de notes de Bac +3 à Bac +5 ou équivalent (pour les résultats de Master ou équivalent, joindre les documents en votre possession), et (4) les coordonnées de 2 références. Merci de nous adresser votre candidature par email avant le 30/06/2024.

6. Références

- (Meistermann *et al.*, 2021) Meistermann, D., Bruneau, A., Loubersac, S., Reignier, A., Firmin, J., et al. : Integrated pseudotime analysis of human pre-implantation embryo single-cell transcriptomes reveals the dynamics of lineage specification. *Cell Stem Cell* 28(9), 1625–1640.e6 (Sep 2021)
- (Bolteau *et al.* 2023) Bolteau, M., Bourdon, J., David, L., Guziolowski, C. (2023). Inferring Boolean Networks from Single-Cell Human Embryo Datasets. In: Guo, X., Mangul, S., Patterson, M., Zelikovsky, A. (eds) *Bioinformatics Research and Applications. ISBRA 2023. Lecture Notes in Computer Science*, vol 14248. Springer, Singapore.
- (Razzaq *et al.* *PLoS Comp Biol* 2018) Razzaq M, Paulevé L, Siegel A, Saez-Rodriguez J, Bourdon J, Guziolowski C. Computational discovery of dynamic cell line specific Boolean networks from multiplex time-course data. *PLoS Comput Biol.* 2018 Oct 29;14(10):e1006538.
- (Videla *et al.* *Front. Bioeng. Biotechnol.* 2015) S. Videla, I. Konokotina, L. Alexopoulos, J. Saez-Rodriguez, T. Schaub, A. Siegel, C. Guziolowski "Designing experiments to discriminate families of logic models". *Frontiers in Bioengineering and Biotechnology*, 2015, DOI=10.3389/fbioe.2015.00131
- (Fitime *et al.* *Algorithms Mol Bio* 2017) L. Fippo Fitime, O. Roux, C. Guziolowski, L. Paulevé. Identification of Bifurcation Transitions in Biological Regulatory Networks using Answer-Set Programming, *Algorithms Mol Biol.* 2017 Jul 20;12:19. doi: 10.1186/s13015-017-0110-3.
- (Chevalier *et al.* LNBI, 2020) Chevalier, S., Noël, V., Calzone, L., Zinovyev, A., Paulevé, L., Paulevé: Synthesis and simulation of ensembles of boolean networks for cell fate decision pp. 193–209 (2020)