

Bayesian model averaging et prédiction de pathologies par la distribution des longueurs des télomères

A. Gégout-Petit, L. Lenôtre et D. Villemonais

Introduction : Un télomère est une petite séquence non codante de macromolécules ATGC, composants de l'ADN, se trouvant aux extrémités des chromosomes. Le rôle de cette petite séquence non codante est double : d'une part, elle permet la réplication complète de l'ADN contenu dans les brins des chromosomes lors de la duplication du matériel génétique ayant lieu lors de la division cellulaire ; d'autre part, elle protège les extrémités des brins composant les chromosomes afin qu'ils ne soient pas considérés à tort comme abimés par les mécanismes de réparation de l'ADN. Chaque télomère présent aux extrémités des chromosomes est sujet à un mécanisme d'attrition agissant lors de la réplication cellulaire. Or, quand un télomère est trop court, il ne peut plus remplir convenablement sa fonction, ce qui conduit à des problèmes de conservation du matériel génétique. En conséquence, les cellules concernées arrêtent de se reproduire [3], amenant l'apparition de défaillances et donc de maladies lorsque les cellules atteignent un âge avancé [1].

Problématique générale : Des télomères courts vont impliquer l'apparition précoce d'un grand nombre de cellules vieillissantes dans un organisme ; et donc l'apparition de maladies dégénératives liées à l'âge, parfois de manière précoce. La distribution des longueurs des télomères d'un individu pourrait donc représenter potentiellement un bon indicateur afin de prévenir certaines pathologies liées au vieillissement. Pour confirmer ou infirmer cette hypothèse, des analyses statistiques fines sont nécessaires.

Problématique mathématique : La quantité d'informations ou de paramètres que l'on peut extraire de la distribution des longueurs des télomères d'un individu est *a priori* relativement grande. D'autre part, vu la modernité des travaux sur le sujet, pratiquement aucun paramètre n'a été identifié. En conséquence, il est difficile de donner des modèles de régression simples permettant de répondre au problème de la prédiction de pathologies par la distribution des longueurs des télomères.

Une manière de remédier à cette difficulté est d'utiliser un grand nombre de modèles de régression avec un nombre raisonnable de paramètres et d'agréger ces modèles. Une technique permettant de réaliser cela est le Bayesian Model Averaging développé dans [2].

L'idée du stage qui est proposé est d'étudier et d'implémenter un modèle simplifié, ou Toy Model, mettant en relation la distribution de la longueur des télomères et les pathologies

pouvant se développer afin de tester la pertinence et l'efficacité de la technique du Bayesian Model Averaging dans ce cas.

Contacts : lionel.lenotre@univ-lorraine.fr, anne.gegout-petit@univ-lorraine.fr et denis.villemonais@univ-lorraine.fr

Soutiens : Ce stage est soutenu par le projet Lorraine Université d'Excellence Geenage, l'Institut Elie Cartan de l'Université de Lorraine, l'Inria et le CHRU de Nancy - Hopitaux de Brabois.

Références

- [1] Athanase BENETOS et al. "Short Leukocyte Telomere Length Precedes Clinical Expression of Atherosclerosis". In : *Circulation Research* 122.4 (2018), p. 616-623.
- [2] Jennifer A. HOETING et al. "Bayesian Model Averaging : A Tutorial". In : *Statistical Science* 14.4 (1999), p. 382-401.
- [3] Zhou XU et al. "The Length of the Shortest Telomere as the Major Determinant of the Onset of Replicative Senescence". In : *Genetics* 194.4 (2013), p. 847-857.