

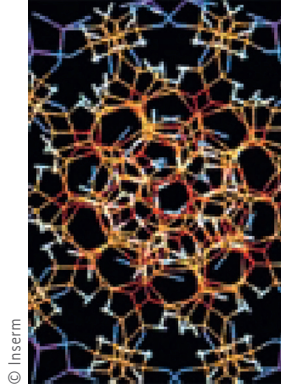
Nanopore et télomères

Résumé

Les télomères sont des structures nucléoprotéiques constituées de séquences répétées de type TTAGGG coiffant les chromosomes eucaryotes. Ces structures extrêmement conservées entre les espèces jouent un rôle majeur dans la stabilité du génome. La longueur des télomères, pouvant aller jusqu'à 20 kb, joue un rôle important dans le vieillissement, l'oncogenèse mais également chez l'homme, dans les troubles cardiométaboliques et neurologiques. Le séquençage de l'ADN télomérique est un défi technique du fait des répétitions TTAGGG et de leur longueur. Cependant, l'émergence des nouvelles technologies de séquençage de 4^e génération telles que le Nanopore développé par la société Oxford a permis de séquencer pour la première fois cette structure répétitive complexe. Tan et al. ont montré que des erreurs d'interprétation du signal brut en séquence nucléotidique (*base-calling*) sont observées dans les séquences télomériques sur l'ensemble des données issues de l'Oxford Nanopore [1]. Ces erreurs sont induites par des similitudes dans les profils du signal brut entre les différents types de répétitions. Les modèles de *base-calling* et les algorithmes d'interprétation jouent un rôle majeur dans la caractérisation finale de la structure nucléotidique séquencée. Les auteurs ont développé une nouvelle stratégie bio-informatique améliorant considérablement le *base-calling* en utilisant un algorithme entraîné à reconnaître les séquences télomériques. Grâce à cette approche, ils ont considérablement diminué les erreurs et détecté avec précision les hexamères TTAGGG des télomères. De manière plus générale, leur étude souligne l'importance de vérifier avec précision l'interprétation des signaux bruts générés dans les régions longues, répétitives et mal définies du génome.

Commentaire

Le développement des nouvelles technologies de séquençage à lecture longue par Pacific Bioscience et Oxford



© Inserm

Sorbonne Université, Inserm, Institut de Myologie, Centre de Recherche en Myologie, Paris, France.
stephanie.tome@inserm.fr

Nanopore a permis d'améliorer le génome de référence Hg38 dans lequel des milliers d'erreurs structurelles ont été corrigées (génome T2T-CHM13). Ces deux plateformes de séquençage à lecture longue sont extrêmement puissantes et permettent de séquencer des régions complexes telles que les régions répétées du génome. Cependant, un support bio-informatique solide ne peut être dissocié de ces nouvelles stratégies moléculaires de séquençage où l'analyse des variants structuraux des séquences répétées telles que les télomères reste difficile comme observé dans l'article cité en référence. Cette étude montre l'importance des algorithmes utilisés pour interpréter les différents signaux bruts en séquences nucléotidiques. La lecture de notre ADN dépend non seulement des prouesses technologiques actuelles mais également des modèles de *base-calling* et des algorithmes développés par les experts en bio-informatique. Le plus surprenant dans cet article est cette notion d'entraînement de ces algorithmes pour les inciter à mieux interpréter les signaux. Allons-nous vers une intelligence artificielle qui pourra demain interpréter avec précision les signaux des séquenceurs de PacBio et Oxford Nanopore ? Ce n'est pas exclu. Quelles sont les limites de cette approche ? Pourra-t-on détecter les événements *de novo* et rares de notre génome ? Tout ceci reste à prouver. ♦

Nanopore and telomeres

RÉFÉRENCE

1. Tan KT, Slevin MK, Meyerson M, et al. Identifying and correcting repeat-calling errors in nanopore sequencing of telomeres. *Genome Biol* 2022 ; 23 : 180.

LIENS D'INTÉRÊT

L'auteure déclare n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

TIRÉS À PART

S. Tomé