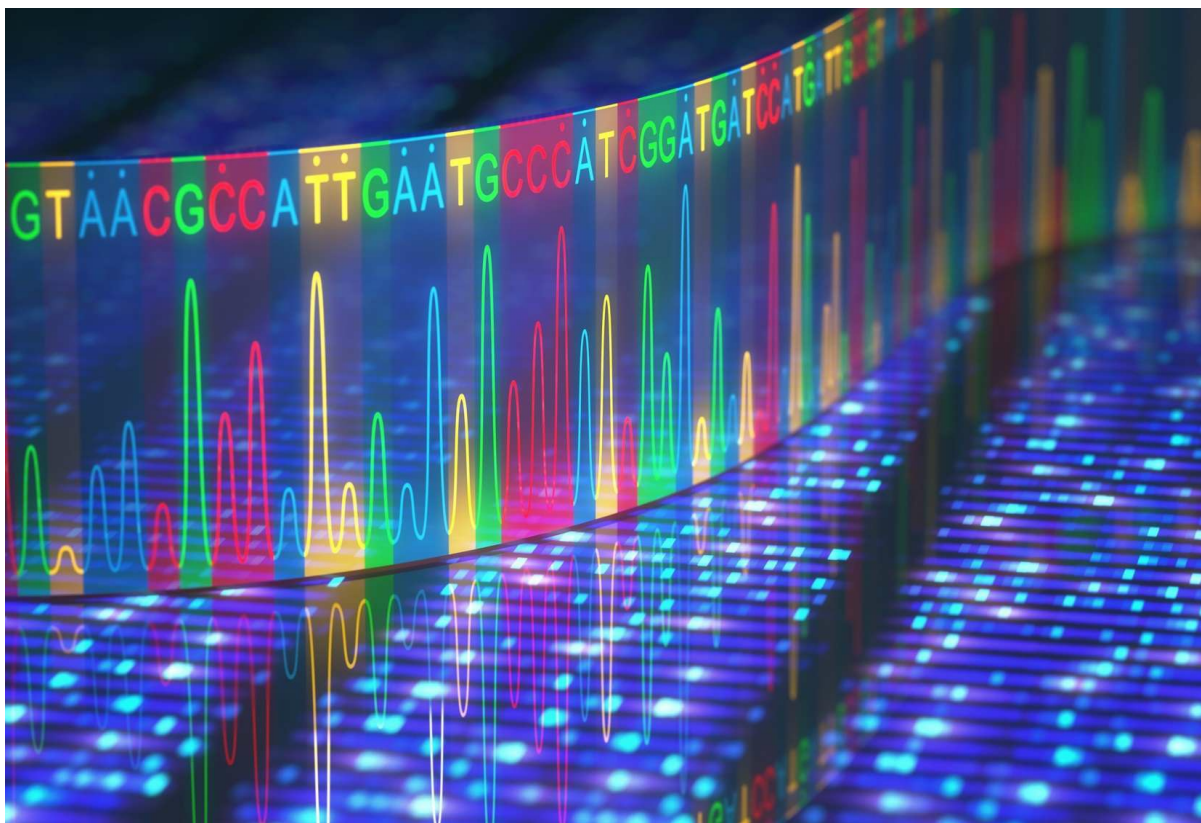


Des régions inexplorées du génome humain enfin dévoilées

Plus aucune base de l'ADN d'Homo sapiens n'échappe aux scientifiques ! Grâce au progrès des techniques de séquençage, le consortium Telomere-to-Telomere a décrypté les derniers 8 % du génome humain qui restaient encore inconnus. Une recherche longue de 20 ans résumée en six publications évènements dans Science.



Le génome humain est désormais séquencé à 100 %. © ktsdesign, Fotolia

Il a fallu 20 ans pour venir à bout des 8 % du génome humain qui restaient inexplorés, soit le temps nécessaire pour séquencer les premiers 92 %. Ce travail colossal a mobilisé une centaine de chercheurs à travers le monde, sous le consortium *Telomere-to-Telomere* (T2T), et vient d'être publié dans pas moins de six publications scientifiques différentes, toutes parues dans *Science*, le 1^{er} avril 2022 pour un numéro spécial sur le sujet.

Il y a 20 ans, l'Homme séquençait son génome pour la première fois

Décortiquer les longues séquences répétées

Pourquoi ces petits 8 % ont été si difficiles à séquencer ? Car ils sont situés dans des régions chromosomiques connues

pour leur structure particulière. Les télomères - la région tout au bout des bras chromosomiques - et le centromère - au centre - sont faits de longues séquences répétées. Comme leur nom l'indique, ce sont des régions de l'ADN où une base (A, T, C ou G) ou une courte suite de bases se répète un grand nombre de fois. Les techniques antérieures de séquençage ne parvenaient pas à gérer ces très longs morceaux d'ADN, appelés aussi *reads*. Pour séquençer l'ADN, il faut d'abord le couper en petits morceaux, puis en lire la séquence et replacer chaque pièce d'ADN au bon endroit dans le génome, comme un puzzle. *The Human Genome Project* a publié le tout premier séquençage du génome humain en 2001, mais certaines pièces du puzzle étaient impossibles à replacer.

Le T2T n'a pas passé à la moulinette l'ADN d'un être humain - un volontaire qui aurait accepté que son patrimoine génétique soit la référence - mais d'une cellule humaine en culture ; une môle hydatiforme précisément. Une môle est une anomalie de fécondation : un ovule sans matériel génétique est fécondé par un spermatozoïde. La cellule qui en résulte porte deux paires identiques de chaque chromosome (46 XX dans 76 % des cas et 46 XY dans 25 %). Malgré cette origine un peu étrange, rien ne permet de penser que le génome d'une môle est différent de celui des autres cellules, précise Megan Dennis, un scientifique de l'université de Californie, membre du T2T.

Un chromosome humain entièrement séquençé pour la première fois



Les télomères et le centromère sont des régions chromosomiques qui vont révéler leurs secrets grâce au nouveau séquençage humain. © Kateryna Kon, Fotolia

Les dernières pièces du puzzle

Les progrès de la technologie ont permis de lire en intégralité les 3.054.815.472 paires de bases de l'ADN chromosomique et les 16.569 supplémentaires de l'ADN mitochondrial. Les scientifiques du T2T ont utilisé l'*Oxford Nanopore Technologies*, un séquenceur qui fait passer les morceaux d'ADN à lire à travers une membrane trouée de minuscules pores qui ne font que 1,5 nanomètre de diamètre à leur point le plus étroit. Ils ont pu ainsi lire des *reads* d'ADN, longs de plus de 100.000 paires de bases avec une très grande précision, et donc de couvrir les longues séquences répétées qui posaient problème jusqu'ici. Les données du séquençage ont été créées par *Pacific Biosciences*, qui possède une plateforme de séquençage spécifique aux longues séquences. Les scientifiques ont pu ainsi placer les dernières pièces dans le grand puzzle génétique qu'ils étaient en train d'assembler.

Grâce à cela, le génome humain s'est enrichi de presque 200 millions de paires de bases - dont 90 % sont situées dans les centromères - desquelles les scientifiques ont repéré 99 nouveaux gènes, potentiellement codants et 2.000 autres candidats qui doivent encore être vérifiés. Des erreurs présentes dans l'ancien génome de référence ont aussi été corrigées. Cette version 2.0 du génome humain, T2T-CHM13, sera sûrement au cœur de nombreuses découvertes en génétique dans le futur.

Les scientifiques ont déjà identifié des variants génétiques dans la région des centromères - reste à savoir si ces subtilités participent à l'émergence des maladies. D'autres régions désormais accessibles pourraient révéler les secrets de l'évolution de l'espèce humaine. Structuellement, l'analyse des séquences centromériques pourrait permettre aux scientifiques de comprendre pourquoi le centromère se forme ici et pas ailleurs, alors que rien de particulier ne semble les guider. En résumé, le champ des possibles semble infini.

Enrichir la base de données

Le T2T ne va pas arrêter ses investigations ici. Le consortium a déjà intégré le *Human Pangenome Reference Consortium* qui a pour objectif de séquencer l'ADN de 350 individus. « *La pangénomique consiste à capturer la diversité de la population humaine, et il s'agit également de s'assurer que nous avons correctement capturé l'ensemble du génome* », explique Benedict Paten, co-auteur des articles parus dans *Science*.

En attendant, le génome de référence T2T-CHM13 complète celui de 2001 (nommé GRCh38) et est disponible intégralement annoté sur le *UCSC Genome Browser*, prêt à être décortiqué dans tous les sens par les scientifiques du monde entier.

