

Éditorial

Le génome humain... encore ?

Bertrand Jordan

► Un numéro récent de la prestigieuse revue *Science* est largement consacré à « *La séquence complète d'un génome humain* » [1]. Surprise... La séquence du génome humain n'a-t-elle pas été annoncée en grande pompe en 2000, publiée en 2001, puis, sous une forme « achevée », en 2004 ? En fait, cette séquence était très imparfaite – quoiqu'elle se soit révélée incroyablement utile. Depuis 2004, elle a été maintes fois retouchée (on en était à la 38^e version [2]), mais elle restait incomplète et par endroits lacunaire. La version publiée le 1^{er} avril de cette année, appelée CHM3 (*complete hyaditiform mole 3*), est, elle, complète, quasiment sans erreur, et elle constitue un réel saut qualitatif.

Mais revenons un peu en arrière... Le *Programme Génome*, discuté dès les années 1980 et officiellement lancé en 1990, fut, à ses débuts, l'objet de vives critiques. Pour beaucoup de biologistes (sans doute même la majorité), ce programme apparaissait comme une énorme et dispendieuse entreprise technologique dont l'utilité réelle était très contestable, et qui allait, par l'énormité de l'investissement requis, mettre en péril les budgets disponibles pour la véritable recherche en biologie moléculaire et en génétique. De plus, il n'était, en fait, pas du tout certain que celui-ci soit réalisable. À l'époque, la plus longue séquence jamais lue était celle du chromosome III de la levure : 315 000 bases, dont la détermination avait mobilisé une vingtaine de laboratoires durant plusieurs années. Les trois milliards de bases (gigabases) du génome humain (dix mille fois plus, donc !) semblaient, à beaucoup, hors de portée, d'autant plus qu'un projet de séquençage de l'ADN de la bactérie *Escherichia coli* (cinq millions de bases), lancé au Japon quelques années plus tôt, avait piteusement échoué. Le programme démarra pourtant, surtout aux États-Unis et au Royaume Uni.

Comme ses premières étapes impliquaient l'établissement de cartes détaillées et globales du génome avant le séquençage proprement dit, les données de cartographie obtenues s'avèrent extrêmement utiles pour les équipes de génétique médicale, ce qui facilita l'acceptation du programme par la communauté scientifique... La prise de conscience de la nécessité d'une organisation de type industriel, quasiment militaire, permit des progrès rapides, soulignés par le succès du Généthon français qui construisit la première carte génétique détaillée de l'ensemble du génome humain [3]. À la fin de la décennie 1990, la concurrence quelque peu déloyale d'une entreprise américaine, *Celera Genomics*¹, visant à obtenir sa propre séquence du

génome humain et à breveter de très nombreux gènes, stimula encore davantage les efforts et aboutit à la séquence du génome, annoncée en 2000 et publiée (en libre accès !) début 2001 [4].

Cette séquence allait révolutionner la biologie. Pour les affections génétiques « mendéliennes » (liées à un seul gène), elle allait permettre d'accéder directement à la région du génome désignée par l'analyse génétique des familles, et donnerait immédiatement le catalogue des gènes présents dans cette région. La technique des *microarrays* (les puces à ADN) rendit possible l'étude de maladies dont le déterminisme génétique est complexe, grâce aux « balayages » larges du génome, ou GWAS (*genome-wide association studies*), mettant en évidence les gènes impliqués. Les mêmes *microarrays*, utilisés pour les analyses d'expression des gènes en cancérologie, permirent d'établir des « signatures » prospectives (indiquant le pronostic) et prédictives (prédisant l'efficacité d'un traitement donné). Pour le cancer du sein, par exemple, les signatures d'expression, développées dans le milieu académique puis commercialisées par deux entreprises (*Agendia* et *Oncocyte*), permirent ainsi de prévoir l'évolution de la maladie, mais aussi d'éviter des chimiothérapies inutiles pour une partie des patientes dont le risque de récurrence prédict était très faible. L'analyse de l'ADN des tumeurs mit en évidence les gènes *drivers* impliqués dans la tumorigenèse (comme le gène *EGFR* [*epidermal growth factor receptor*] et ses diverses mutations). À partir de 2005, avec le bond en avant des techniques de séquençage (NGS, *new generation sequencing*), il devint possible de séquencer, en clinique, l'ADN de tumeurs à partir de biopsies et, parfois, d'en déduire la thérapie ciblée à utiliser : c'est ce que l'on a appelé l'oncologie de précision. Pour le cancer du poumon, par exemple, la mise en évidence d'une mutation du gène *EGFR* indique la possibilité d'emploi d'inhibiteurs de l'EGFR, comme l'afatinib (*Giotrif*®), l'erlotinib (*Tarceva*®) ou le gefitinib (*Iressa*®). L'oncologie de précision n'a pas répondu à tous les espoirs mis en elle, mais elle a néanmoins significativement amélioré l'efficacité des traitements. Enfin, le séquençage de l'ADN de différentes populations et, plus récemment, celui de l'ADN « ancien », parfois extrait d'ossements vieux de plus de cinquante mille ans, a réellement révolutionné l'anthropologie et révélé toute la complexité de l'histoire humaine [5].

¹ *Celera Genomics* avait accès en temps réel à toutes les données du projet public et les utilisait pour ordonner ses séquences, l'inverse n'étant pas vrai.



Toutes ces avancées ont très largement reposé sur l'existence d'une séquence de référence du génome humain. Pourtant, même dans sa dernière version appelée GRCh38 (*genome research consortium h38*) [2], cette dernière présentait de nombreuses imperfections. Les centromères, les télomères et les bras courts des chromosomes acrocentriques n'avaient pas pu être séquencés, en raison de la présence de multiples séquences répétées. Dans ces régions, la continuité de la séquence n'avait pu être obtenue : elle comportait encore près de mille « trous » (*gaps*) et plus de dix millions de bases d'ADN dont la position n'avait pas été déterminée. Du coup, l'utilisation de cette séquence de référence dans divers projets de génétique moléculaire aboutissait parfois à une impasse. L'utilisation de l'ADN d'une lignée cellulaire totalement homozygote, et la mise en œuvre de nouvelles techniques de séquençage permettant la lecture de dizaines, voire même de centaines de milliers de bases d'un seul tenant, ont permis à un consortium, appelé T2T (*telomere to telomere*) d'obtenir une séquence qui, cette fois, est réellement complète [6] (→). Chaque chromosome est représenté par une séquence continue, allant d'une extrémité à l'autre du chromosome ; télomères et centromères sont désormais intégralement séquencés, sans ambiguïté ; et il n'y a plus aucun trou ni aucune base non assignée. Enfin, le taux d'erreur estimé est de un pour dix millions de bases, mille fois plus bas que précédemment. C'est donc bien plus qu'une nouvelle version : c'est, cette fois, la véritable séquence complète d'un génome.

La séquence d'un génome : il s'agit bien d'un génome en particulier, celui des cellules choisies pour ce travail. Pas plus que celle qui l'a précédée (CRCh38), cette séquence ne constitue « la » séquence du génome : la diversité génétique humaine, plus importante qu'on ne l'avait imaginée en 2001 [7], est telle que deux personnes prises au hasard se distinguent non seulement par quelques millions de différences ponctuelles, mais aussi par un grand nombre d'insertions, de délétions et de duplications, dont une partie seulement a été précisément répertoriée, faute de séquences réellement complètes. En d'autres termes, il faudra lire, avec la même précision, l'ADN d'un certain nombre de personnes pour avoir une véritable évaluation de la diversité génétique de notre espèce. C'est possible, à un coût acceptable, grâce aux techniques moléculaires et informatiques rapportées dans l'article de *Science* [1], même si l'on est loin, dans ce cas, du mythique « génome à mille dollars », et cet objectif va, à coup sûr, faire l'objet de nombreux travaux. Dès aujourd'hui, l'ouverture des « boîtes noires » que constituaient centromères, télomères et bras acrocentriques va permettre de

comprendre des situations compliquées². D'une manière générale, cette représentation précise et complète d'un génome humain va améliorer l'efficacité de la « médecine de précision ». Et, dans la perspective de modifications programmées par édition du génome, une connaissance plus précise et complète de la séquence du génome ciblé ne sera pas inutile... ♦

The human genome – again?

LIENS D'INTÉRÊT

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

RÉFÉRENCES

1. Nurk S, Koren S, Rhie A *et al.* The complete sequence of a human genome. *Science* 2022 ; 376 : 44-53.
2. GRCh38.p14 2022/02/03 (Genome Reference Consortium). https://www.ncbi.nlm.nih.gov/assembly/GCA_000001405.29
3. Weissenbach J, Gyapay G, Dib C, *et al.* A second-generation linkage map of the human genome. *Nature* 1992 ; 359 : 794-801.
4. Olson MV. The Human Genome Project: a player's perspective. *J Mol Biol* 2002 ; 319 : 931-42.
5. Slatkin M, Racimo F. Ancient DNA and human history. *Proc Natl Acad Sci USA* 2016 ; 113 : 6380-7.
6. Jordan B. Séquence du génome : la fin du commencement. *Med Sci (Paris)* 2022 ; 38 : 609-11.
7. Collins FS, Mansoura MK. The Human Genome Project. Revealing the shared inheritance of all humankind. *Cancer* 2001 ; 91 : 221-5.



Bertrand Jordan

Biologiste, généticien et immunologiste,

Président d'Aprogène

(Association pour la promotion de la génomique),

13007 Marseille, France.

brjordan@orange.fr

TIRÉS À PART

B. Jordan

² Comme celle de la région FHSD (*Facioscapulohumeral Muscular Dystrophy*) qui contient des gènes impliqués dans une dystrophie musculaire, située dans une zone subtélomérique du chromosome 4 et dupliquée sur de nombreux autres chromosomes, ce qui en rend l'étude très ardue si l'on ne dispose pas d'une séquence complète et fiable.



Tarifs d'abonnement m/s - 2022

Abonnez-vous

à médecine/sciences

> Grâce à m/s, vivez en direct les progrès des sciences biologiques et médicales

Bulletin d'abonnement page 612 dans ce numéro de m/s

