

Chroniques génomiques

Séquençage d'ADN, la fin d'un quasi-monopole ?

Bertrand Jordan

De Sanger au NGS (*new generation sequencing*)

Le premier séquençage d'un génome humain, réalisé de 1990 à 2001, fut effectué en utilisant une version automatisée de la technique de Sanger¹, apparue dès 1977. Vers la fin de ce projet, qui impliqua de nombreux laboratoires à travers le monde et mobilisa des crédits totalisant près de trois milliards de dollars, on estimait que, compte tenu des progrès techniques et organisationnels, l'établissement d'une nouvelle séquence génomique humaine coûtait environ cent millions de dollars (environ 91 millions d'euros). À partir de 2002, de nouvelles techniques reposant sur une combinaison de micromécanique, d'enzymologie sophistiquée et d'optique de précision permirent la lecture simultanée de millions de molécules d'ADN et autorisèrent une accélération considérable des travaux et un abaissement spectaculaire des coûts de séquençage [1, 2] (→).

(→) Voir la Chronique
génomique de B. Jordan, *m/s*
n° 10, octobre 2008, page 869

Vers 2010, le coût de l'établissement d'une séquence génomique humaine était tombé à environ dix mille dollars (environ 9000 euros) pour une séquence de bonne qualité [3] (Figure 1), soit un génome haploïde (trois mille mégabases²), lu avec une redondance moyenne de trente fois (30X) et un taux d'erreurs inférieur à 1/1000.

Les acteurs de ce domaine appelé NGS (*new generation sequencing*) étaient assez nombreux au début, mais au fil des années, beaucoup disparurent (454, *Helicos*, *Solid*, *Complete Genomics*, etc.) pour ne laisser, vers la fin des années 2010, que trois intervenants principaux : l'entreprise nord-américaine *Illumina* en position dominante (environ 90 % du marché), et deux concurrents, le Californien *Pacific Biosciences* (*PacBio*) et le Britannique *Oxford Nanopore Technologies* (*ONT*). Reste aussi un acteur marginal, *Ion Torrent* (*Thermo*



Biologiste, généticien et immunologiste, Président d'Aprogène (Association pour la promotion de la Génomique), 13007 Marseille, France.
brjordan@orange.fr

Fisher), spécialisé dans les tests cliniques à bas débit, et un concurrent chinois, *MGI tech*, dont nous parlerons plus bas. Il est important de noter que la technologie *Illumina* est orientée vers la lecture à très haut débit de séquences courtes (cent à deux cents bases) alors que *PacBio* et *ONT* permettent, avec des approches différentes, des lectures très longues portant sur des dizaines de kilobases. En termes comparatifs, *Illumina* annonce les coûts de séquençage les plus bas, *PacBio* offre, outre les lectures longues, une très bonne qualité de lecture mais à un coût plus élevé. *ONT* présente des avantages au niveau de sa mise en œuvre et permet des lectures très longues, mais le coût du séquençage est relativement élevé et le taux d'erreur supérieur à celui de ses concurrents [4].

2023 : vers le génome à cent dollars ?

Avec l'augmentation des débits et l'abaissement des coûts, de nouveaux projets impliquant le séquençage de milliers de génomes humains deviennent possibles : citons, par exemple, la lecture intégrale du demi-million d'ADN collecté dans le cadre de la *UK Biobank*, déjà bien avancée [5] ou le projet de séquençage de toutes les espèces vivantes [6, 7]. En regard, les machines sont de plus en plus puissantes. Le dernier modèle d'*Illumina*, la *NovaSeq X plus*, peut lire en deux jours 16 téraoctets de séquence brute, soit près de deux cents génomes de séquence finie répondant aux critères déjà cités. Le coût annoncé est d'environ deux dollars par gigabase de séquence brute : un génome humain de trois gigabases lu avec une redondance de

¹ Voir l'article original : Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977 ; 74 : 5463-7. doi : 10.1073/pnas.74.12.5463.

² Pour la longueur des séquences, on parle de kilobases (mille bases), mégabases (un million), gigabases (un milliard) et téraoctets (mille milliards).



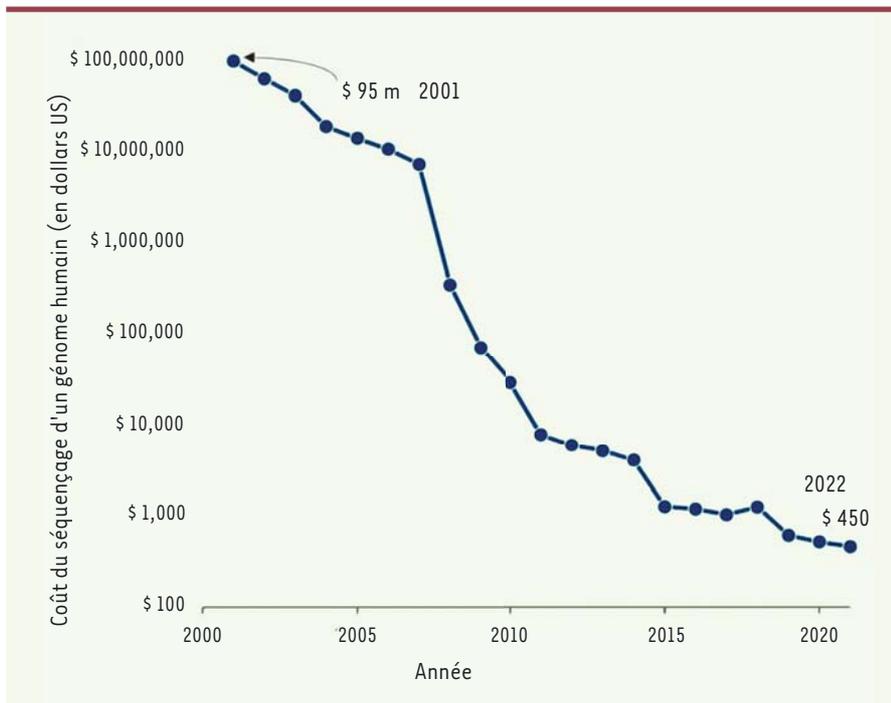


Figure 1. Évolution du coût du séquençage d'un génome humain au cours des vingt dernières années (noter l'échelle logarithmique). Le point d'inflexion vers 2005 correspond à l'arrivée des approches NGS (*new generation sequencing*). Début 2023, le coût revendiqué par *Illumina* est de deux cents dollars par génome. Source : *NHGRI (National Human Genome Research Institute)*.

MGI tech compte plus de deux mille employés et a vendu un bon millier de machines, mais est pour le moment très peu présente en Europe et en Amérique du Nord : *Illumina* a longtemps réussi à empêcher son implantation grâce à des procès portant sur la technique de séquençage, mais les tribunaux ont récemment rendu une décision favorable

à *MGI tech* qui peut désormais commercialiser ses systèmes sans restriction. Cette entreprise peut donc concurrencer *Illumina* sur son marché principal, avec de sérieuses références, des tarifs agressifs et une gamme complète d'appareils, dont des systèmes à très haut débit encore plus performants (sur le papier) que ceux d'*Illumina* (Figure 2). Reste que les usagers risquent d'hésiter à abandonner un fournisseur connu pour un nouveau venu dont le service après-vente en Occident est une inconnue⁴, et qui, de plus, émane d'un pays avec lequel les relations sont plus que délicates : il ne faudrait pas que *MGI tech* subisse le destin de *Huawei* ou de *TicTok*, produits chinois désormais interdits (ou presque) aux États-Unis... Mais *MGI tech* n'est pas le seul à s'aventurer sur le pré carré d'*Illumina* : plusieurs nouveaux arrivants annoncent ou commercialisent déjà des systèmes assez proches de la technologie *Illumina* mais promettant un coût inférieur et/ou une plus grande flexibilité. On peut citer à cet égard *Element Biosciences*, qui commercialise déjà une machine de moyenne gamme (lecture de 1 térabase par session, contre 16 pour le *NovaSeq X Illumina*, avec un coût d'acquisition raisonnable et une grande flexibilité), *Ultima Genomics* qui va mettre sur le marché une machine à très haut débit censée faire descendre le coût (toujours en réactifs) à un dollar par gigabase de séquence brute (soit un génome humain

Compétition sur les séquences courtes

30X reviendrait ainsi à moins de deux cents dollars. Il s'agit là du coût total en réactifs et informatique, depuis la préparation des bibliothèques jusqu'à l'assemblage de la séquence : l'amortissement du matériel et les salaires du personnel ne sont pas comptabilisés, mais compte tenu du volume de la production (plusieurs milliers de génomes par mois) les réactifs restent la principale dépense. *Illumina*, qui emploie plus de onze mille personnes et dispose d'une gamme complète de machines, en a vendu à ce jour plus de vingt mille, à un prix unitaire qui va d'un demi-million à plus d'un million de dollars. Avec une part de marché en 2022 qui reste supérieure à 70 %, cette firme est encore en position dominante. Mais elle fait désormais face à une concurrence sérieuse, tant de la part d'acteurs déjà existants que de plusieurs nouveaux venus [8].

à *MGI tech* qui peut désormais commercialiser ses systèmes sans restriction. Cette entreprise peut donc concurrencer *Illumina* sur son marché principal, avec de sérieuses références, des tarifs agressifs et une gamme complète d'appareils, dont des systèmes à très haut débit encore plus performants (sur le papier) que ceux d'*Illumina* (Figure 2).

Reste que les usagers risquent d'hésiter à abandonner un fournisseur connu pour un nouveau venu dont le service après-vente en Occident est une inconnue⁴, et qui, de plus, émane d'un pays avec lequel les relations sont plus que délicates : il ne faudrait pas que *MGI tech* subisse le destin de *Huawei* ou de *TicTok*, produits chinois désormais interdits (ou presque) aux États-Unis... Mais *MGI tech* n'est pas le seul à s'aventurer sur le pré carré d'*Illumina* : plusieurs nouveaux arrivants annoncent ou commercialisent déjà des systèmes assez proches de la technologie *Illumina* mais promettant un coût inférieur et/ou une plus grande flexibilité. On peut citer à cet égard *Element Biosciences*, qui commercialise déjà une machine de moyenne gamme (lecture de 1 térabase par session, contre 16 pour le *NovaSeq X Illumina*, avec un coût d'acquisition raisonnable et une grande flexibilité), *Ultima Genomics* qui va mettre sur le marché une machine à très haut débit censée faire descendre le coût (toujours en réactifs) à un dollar par gigabase de séquence brute (soit un génome humain

³ Aux États-Unis, *MGI Tech Co., Ltd.* se présente sous le nom de *Complete Genomics*, qui est sa propriété depuis 2013.

⁴ Et qui avait lancé il y a quelques années une machine très haut de gamme baptisée *Revelicity* avant d'abandonner ce projet.

30X pour cent dollars), et aussi *Singular Genomics* qui a commencé à vendre une machine de milieu de gamme, un peu comme *Element Biosciences*. Quelles sont les chances de ces nouveaux venus ? Ce sont des entreprises de petite taille, non encore implantées sur le marché, et qui s'attaquent au *market leader* en proposant des systèmes assez similaires mais un peu plus commodes ou un peu moins chers. On peut être assez sceptique sur leurs chances de succès. *MGI tech*, par contre, pourrait bien tailler des croupières à *Illumina* si les évolutions géopolitiques ne se mettent pas en travers de son chemin.

La montée en puissance des séquences longues

L'obtention de séquences longues (plusieurs dizaines ou même centaines de kilobases lues d'un seul tenant) présente de nombreux avantages. C'est particulièrement évident pour le séquençage *de novo* d'un génome encore inconnu, pour lequel on ne dispose pas de séquence de référence permettant l'assemblage des millions de courtes séquences lues au hasard ; mais, même pour un génome humain, cela permet de s'affranchir des problèmes liés aux séquences répétées, de repérer facilement délétions, duplications et inversions, et de définir les haplotypes⁵ en déterminant quels allèles sont associés sur la même molécule d'ADN. Ces avantages sont si évidents qu'*Illumina* propose des processus permettant (plus ou moins) d'associer plusieurs lectures courtes pour simuler une lecture longue (*synthetic reads*). Les systèmes *PacBio* et *ONT*, eux, lisent directement de très longs fragments d'ADN. Pour *PacBio*, cela repose sur un système optique très sophistiqué dont la sensibilité permet la lecture d'une seule molécule, et non d'un mélange de courts produits d'amplification par PCR (*polymerase chain reaction*). Fondée en 2004, la firme reste de dimensions modestes (sept cent cinquante employés, un demi-millier d'appareils installés) et est longtemps restée cantonnée à un marché de niche complémentaire à celui d'*Illumina* en raison de son coût élevé et d'un important taux d'erreurs. Mais des améliorations récentes lui ont permis d'atteindre le taux d'erreurs de référence, un débit élevé (une térabase de séquence par session) et un coût nettement abaissé correspondant à mille dollars (environ 900 euros) pour un génome humain [8]. C'est encore cinq fois plus cher que le

⁵ Groupe d'allèles de différents locus situés sur un même chromosome et habituellement transmis ensemble.



Figure 2. Du très haut de gamme : le système « DNBSEQ-T10x4RS » proposé par *MGI/Complete Genomics*, qui rassemble plusieurs appareils servis par un bras robotique et pour lequel l'entreprise revendique une production de 76 térabases (de séquence brute) en une session de quatre jours environ. Extrait du site de *MGI-Tech* (<https://en.mgi-tech.com/products/>).

dernier chiffre d'*Illumina*, mais les avantages des lectures longues peuvent justifier la différence. Le consortium *Telomere to Telomere*, qui vise le séquençage intégral et sans aucune interruption de tous les chromosomes humains [9, 10] (→) a, par exemple, choisi de faire reposer l'essentiel de ses travaux sur cette approche. À l'avenir, *PacBio* pourrait passer du statut de technologie complémentaire à celui de concurrent d'*Illumina*. Cette dernière firme l'a bien compris et a tenté d'acheter *PacBio* (pour 1,2 milliards de dollars, environ 1,08 milliards d'euros) en 2019, mais la transaction a été bloquée par la *Federal Trade Commission* et par l'Autorité de la Concurrence européenne en vertu des réglementations antitrust. L'autre intervenant dans le secteur des séquences longues est *Oxford Nanopore Technology (ONT)* qui, comme son nom l'indique, met en œuvre la technique des nanopores [11, 12] (→).

Il s'agit de faire passer les molécules d'ADN à travers des pores microscopiques, et d'interpréter les petites variations électriques produites par ce passage pour identifier les bases. *ONT* a réussi à mettre en œuvre cette approche sous forme d'une cartouche à peine plus grosse qu'une clef USB, contenant une membrane qui comporte un demi-millier de pores (*Figure 3*). Cette cartouche, qui coûte moins d'un millier d'euros, est branchée sur un ordinateur portable et l'échantillon est déposé sur la membrane. En quelques heures, ce système produit jusqu'à cinquante gigabases de séquence avec le cas échéant des lectures s'étendant sur des centaines de kilobases si l'ADN déposé est en bon état.

L'ensemble est ultra-portable (il existe des variants plus productifs associant jusqu'à 48 cartouches), rapide et sans investissement initial ; le coût par gigabase est de l'ordre de la centaine d'euros, bien plus élevé que les systèmes à très haut débit, mais compensé par l'absence

(→) Voir la Chronique génomique de B. Jordan, m/s n° 1, janvier 2019, page 81

(→) Voir la Chronique génomique de B. Jordan, m/s n° 8-9, août-septembre 2017, page 801 et la Synthèse de F. Montel, m/s n° 2, février 2018, page 161



Figure 3. Le « séquenceur » d'Oxford Nanopore Technology, une cartouche d'une centaine de grammes qui peut produire de dix à cinquante gigabases de séquence (brute). Il ne lui manque qu'un ordinateur pour être fonctionnel. On appréciera la différence avec le « monstre » de la Figure 2... Extrait du site d'ONT (<https://nanoporetech.com/products/specifications>).

d'investissement initial, la commodité (rapidité, portabilité) et la possibilité de lectures très longues (le record dépasse la mégabase). La qualité des séquences, assez mauvaise à l'origine [12], a été bien améliorée et se rapproche maintenant des standards du domaine. ONT reste une entreprise de taille moyenne (environ 400 employés) et a vendu presque dix mille systèmes allant de la cartouche individuelle aux systèmes multi-cartouches. Elle ne joue pas vraiment dans la même cour qu'Illumina ou même PacBio et va probablement principalement intervenir pour des séquençages ultra-rapides ou dans des conditions très rustiques (cf. son emploi dans le cadre de la récente épidémie d'Ebola [13]), pour la caractérisation de souches de SARS-CoV, ou encore pour des séquençages *de novo* d'organismes encore peu caractérisés.

Un champ réouvert

Après une dizaine d'années durant lesquelles la technologie Illumina, et la firme du même nom, ont dominé le marché du séquençage d'ADN, on voit que la compétition s'active à nouveau, et que le coût baisse encore, s'approchant aujourd'hui du « génome (humain) à cent dollars ». Cela ouvre un nouveau champ de possibilités puisque la lecture intégrale d'un ADN devient accessible pour le prix que coûtait un profil génétique par *microarray* il y a quelques années. Du coup, il devient envisageable, par exemple, de réaliser des balayages du génome (GWAS, *genome-wide association studies*) fondés sur la séquence plutôt que sur l'examen d'un million de SNP (*single nucleotide polymorphism*), de reprendre des questions d'anthropologie en séquençant des populations entières, ou encore d'appuyer une réinterprétation de la phylogénie sur la connaissance intégrale des génomes de très nombreuses espèces. Et bien sûr, l'utilisation de la séquence en clinique, déjà bien développée notamment pour la cancérologie, va pouvoir se généraliser.

Dans ce nouveau paysage, Illumina va sans doute rester en position dominante à moins d'erreurs stratégiques majeures⁶ : l'entreprise est très bien implantée et les chances de succès de nouveaux venus cassant les prix semblent assez faibles – à part le cas de MGI tech qui dispose de sérieux atouts mais risque de souffrir du contexte

géopolitique. Mais l'approche des séquences longues va certainement se développer en raison de ses avantages scientifiques et de l'existence depuis peu de systèmes réellement performants. À cet égard, PacBio semble le mieux placé et peut réellement concurrencer Illumina, ONT se positionnant plutôt sur les applications cliniques, sur le séquençage *de novo* et sur les travaux sur le terrain.

On a là un bel exemple de progrès technique fulgurant, passant du génome à cent millions de dollars au génome à cent dollars en vingt ans seulement, comparable mais encore plus rapide que celui de l'informatique. Un progrès qui repose sur la remarquable capacité d'innovation scientifique et technique du système étatsunien⁷ : il est à cet égard intéressant de constater l'absence du Japon qui pouvait *a priori* sembler bien armé et a d'ailleurs lancé des projets dès les années 1990 [14] – mais cette nation ne dispose pas de l'efficace écosystème⁸ présent en Californie ou dans la région de Boston. En tous cas, soixante-dix ans après la découverte de la structure en double hélice de l'ADN et vingt ans après le premier séquençage d'un génome humain, l'accès aisé aux séquences de ces molécules n'a pas fini de transformer notre compréhension de la biologie. ♦

SUMMARY

End of a near-monopoly on DNA sequencing

DNA sequencing costs have steadily decreased during the last decade, but the dominant technology (short-read sequencing, Illumina) has seen comparatively little competition after an initial flurry. This phase is now over, with serious competition involving both established and new companies as well as the growing importance of long-read sequencing. The hundred-dollar genome is in sight, and this will have a major impact on many fields of biology. ♦

LIENS D'INTÉRÊT

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

⁶ Comme l'absorption ratée de Pacific Biosciences et celle également ratée de GRAIL (diagnostic pré-symptomatique de cancers), toutes deux bloquées par les autorités anti-trust et qui ont coûté très cher à l'entreprise.

⁷ Et Britannique, appuyé sur une tradition d'excellence en recherche.

⁸ Laboratoires de recherche, entreprises innovantes et structures de capital-risque en interaction forte.

RÉFÉRENCES

1. Jordan B. Une révolution longuement attendue. *Med Sci (Paris)* 2008 ; 24 : 869-73.
2. Heather JM, Chain B. The sequence of sequencers : The history of sequencing DNA. *Genomics* 2016 ; 107 : 1-8.
3. <https://ourworldindata.org/grapher/cost-of-sequencing-a-full-human-genome>
4. <https://seekingalpha.com/article/4575505-gene-sequencing-market-illumina-pacific-bio-oxford-nanopore-comparison>
5. Halldorsson BV, Eggertsson HP, Moore CHS, et al. The sequences of 150,119 genomes in the UK Biobank. *Nature* 2022 ; 607 : 732-40.
6. A bid to sequence all species, in News at a glance, *Science* 2018 ; 362 : 504-6.
7. Rhie A, McCarthy SA, Fedrigo O, et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature* 2021 ; 592 : 737-46.
8. <https://www.genomeweb.com/sequencing/sequencing-instrument-launches-2022-portend-heated-competition#.ZA880B-Z0Fg>
9. Jarvis ED, Formenti G, Rhie A, et al. Semi-automated assembly of high-quality diploid human reference genomes. *Nature*. 2022 ; 611 : 519-31.
10. Jordan B. Actualités du séquençage d'ADN. *Med Sci (Paris)* 2019 ; 35 : 81-3.
11. Jordan B. Séquençage d'ADN : l'offensive des nanopores. *Med Sci (Paris)* 2017 ; 33 : 801-4.
12. Montel F. Séquençage de l'ADN par nanopores : résultats et perspectives. *Med Sci (Paris)* 2018 ; 34 : 161-5.
13. Quick J, Loman NJ, Duraffour S, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016 ; 530 : 228-32.
14. Endo I, Soeda E, Murakami Y, Nishi K. Human genome analysis system. *Nature* 1991 ; 352 : 89-90.

TIRÉS À PART

B. Jordan

www.myobase.org

Catalogue en ligne disponible gratuitement sur Internet publié par l'AFM-Téléthon.
Retrouvez facilement toutes les références bibliographiques sur les maladies neuromusculaires, les situations de handicap qu'elles génèrent et leurs aspects psychologiques.

Myobase donne un accès libre à 75 % du fonds documentaire collecté depuis 1990, représentant plus de 40 000 références spécifiques du domaine des maladies neuromusculaires.

> **articles** de la littérature biomédicale et psycho-sociale

> **livres, thèses**

> **guides** d'associations et **rapports** institutionnels d'agences internationales

> **brèves en français**, synthèses des articles médico-scientifiques internationaux les plus pertinents

> **publications AFM-Téléthon** destinées aux professionnels de santé ou aux personnes atteintes de maladie neuromusculaire et à leur entourage

UN OUTIL ERGONOMIQUE, UNE INTERFACE BILINGUE

- Laissez-vous guider par les **tutoriels**
- Lancez une **recherche** et affinez votre sélection grâce aux filtres

TOUT MYOBASE

Rechercher...



Recherche avancée

Historique

FILTRES

Type de document

- Article [3443]
- Publication AFM [176]
- Thèse/Mémoire [107]
- Brève [102]

► PUBLICATIONS AFM-Téléthon

► BRÈVES

► DOCUMENTS DE SYNTHÈSE

► INSTITUT DES BIOTHÉRAPIES PUBLICATIONS

- **Partagez** les résultats de votre recherche

UN ACCÈS facile et simple

Rechercher avec des opérateurs :

- guillemets pour une expression "**maladie de pompe**"
- **+** pour signifier **ET**, et retrouver tous les documents contenant les deux mots "**fauteuil +électrique**"
- **-** pour signifier **NON** et enlever le mot de la recherche : "**autonomie -établissement**"



Fils RSS
Les Fils RSS vous permettent de suivre quotidiennement les nouveautés de Myobase, mais aussi ...



Alertes Myobase
Les Alertes rassemblent une sélection des dernières acquisitions de Myobase et paraissent deux fois...



Veille Neuromusculaire
Publiée tous les 15 jours par le Service de documentation de l'AFM-Téléthon, La "V..."

- Cliquez sur l'**onglet thématique** qui vous convient (haut de la page d'accueil)

- Créez vos alertes personnalisées en ouvrant un **compte personnel**

- Téléchargez la **Veille Neuromusculaire**

- Abonnez-vous aux **flux RSS**

