

L'ESSENTIEL

> En 2001, deux équipes ont indépendamment publié la première carte du génome humain, fournissant ainsi la liste de tous les gènes codant des protéines chez l'humain.

> Depuis, les recherches sur le génome humain ont explosé, notamment sur les régions du génome qui ne codent pas de protéines.

> L'analyse des publications scientifiques associées révèle qu'une grande partie concerne un petit nombre de gènes et qu'il reste encore de nombreuses cibles potentielles de médicaments à exploiter.

LES AUTEURS

ALEXANDER J. GATES
bio-informaticien,
chercheur associé
à l'université Northeastern,
à Boston, aux États-Unis

DEISY MORSELLI GYSI
bio-informaticienne en
postdoctorat à l'université
Northeastern, à Boston,
aux États-Unis

MANOLIS KELLIS
professeur d'informatique
à l'institut de technologie du
Massachusetts, à Cambridge,
aux États-Unis

ALBERT-LÁSZLÓ BARABÁSI
professeur de science
des réseaux à l'université
Northeastern, à Boston,
aux États-Unis

Cet article est une traduction de « *A wealth of discovery built on the Human Genome Project – by the numbers* », publié sur *Nature.com* le 10 février 2021.

Le séquençage du génome humain a eu un impact considérable sur la génomique en termes de publications scientifiques, de nouveaux médicaments et de compréhension des maladies.

Le vingtième anniversaire de la publication de la première carte du génome humain (indépendamment par une équipe internationale autour du biologiste américain Craig Venter et par le Consortium international pour le séquençage du génome humain) offre l'occasion d'examiner en quoi ce projet a aidé la recherche des racines génétiques de maladies, favorisé la découverte de nouveaux médicaments et contribué à revoir notre compréhension du génome.

Pour illustrer cet impact, nous avons combiné plusieurs ensembles de données afin de quantifier les différents éléments génétiques qui ont été découverts et ont donné lieu à des publications, et visualiser comment le rythme et les sujets de publication ont évolué au fil des ans. Notre analyse rassemble des données telles que 38546 transcriptions d'ARN, environ 1 million de polymorphismes touchant un seul nucléotide [les nucléotides sont les constituants de l'ADN, ndt] ou SNP, 1660 maladies ayant des racines génétiques documentées, 7712 produits pharmaceutiques autorisés ou expérimentaux et 704515 articles scientifiques publiés entre 1900 et 2017.

Les résultats mettent en évidence comment le projet *Génome humain*, avec la liste exhaustive des gènes codant des protéines qu'il a permis d'établir, a marqué le début d'une nouvelle ère – celle de l'étude de la partie non codante du génome – et a ouvert la voie à des développements thérapeutiques. Surtout, ces résultats

permettent de suivre l'émergence d'une vision systémique de la biologie, parallèlement à la perspective classique focalisée sur chaque gène individuel, à mesure que les chercheurs cartographiaient les interactions des éléments constitutifs des cellules.

UN APERÇU DES PUBLICATIONS DU DOMAINE

Notre analyse a cependant ses limites. Par exemple, il n'existe pas de consensus sur l'endroit où un gène commence et se termine ni même, étonnamment, sur la séquence exacte codant certains gènes. De multiples conventions sont utilisées en parallèle pour nommer certains éléments du génome, de sorte que notre méthodologie a parfois échoué à les relier. Et d'autres liens entre publications et éléments génomiques n'ont peut-être pas été ajoutés aux bases de données par leurs auteurs. Enfin, nos graphiques s'arrêtent en 2017, car il peut y avoir un décalage entre la publication d'un article et son entrée dans les bases de données que nous avons utilisées.

Toutefois, nous ne pensons pas que ces limitations modifient de façon notable les tendances globales constatées dans l'évolution de la recherche sur le génome au fil du temps. Et ces tendances persistent lorsque nous prenons en compte la croissance de la production scientifique en biologie au cours de la même période. Nous n'avons pas effectué ces contrôles sur toute la période qui a suivi la découverte des gènes, mais nous estimons que cela n'aurait pas modifié nos conclusions.

>