

> focalisées sur le gène *TP53* augmente, il devient d'autant plus facile d'obtenir des financements, un encadrement, des outils et d'être cité pour les travaux futurs sur *TP53*. Dans la science des réseaux, ce phénomène est appelé « attachement préférentiel ». De fait, on constate que le nombre de nouvelles publications annuelles portant sur un gène donné est directement proportionnel à l'abondance de la littérature pré-existante sur ce gène.

Un défi pour la biologie consiste maintenant à déterminer ce qu'il faudra étudier ensuite. Les chercheurs consacrent-ils de l'argent, du temps et des efforts à ce qui est le plus important et le plus urgent, ou à ce qui est en terrain connu, parce que c'est une recette fiable pour obtenir des subventions et de la reconnaissance ?

DES TRÉSORS DANS LES POUBELLES

Un grand débat a précédé le lancement du projet *Génome humain* : était-il utile de cartographier les vastes régions non codantes du génome, que l'on appelait alors « ADN poubelle » ou « matière noire » du génome ? Grâce en grande partie au projet *Génome humain*, on sait maintenant que la majorité des séquences fonctionnelles du génome humain ne codent pas de protéines. Au contraire, des éléments tels que les ARN non codants, les promoteurs, les séquences stimulatrices [de petits segments d'ADN nécessaires pour activer les gènes, *ndt*] et d'innombrables motifs de régulation des gènes travaillent ensemble pour donner vie au génome. Les variations dans ces régions n'altèrent pas les protéines, mais peuvent perturber les réseaux qui régissent l'expression de ces dernières.

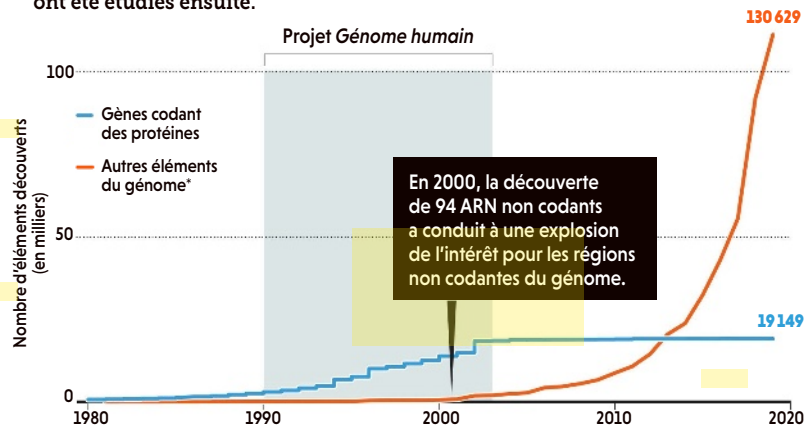
Avec la carte du génome humain en main, la découverte d'éléments non codants a explosé. Jusqu'à présent, le rythme a été cinq fois plus rapide que celui de la découverte des gènes codant des protéines et il ne montre aucun signe de ralentissement. De même, le nombre de publications portant sur ces constituants du génome a augmenté au cours de la période couverte par notre ensemble de données (1900 à 2017). Par exemple, il existe des milliers d'articles sur les ARN non codants, qui régulent l'expression des gènes.

Le projet *Génome humain* a également offert un moyen de cataloguer les variations génétiques humaines, y compris celles des polymorphismes touchant un seul nucléotide, les SNP. D'autres efforts importants ont permis de réduire le coût de l'identification de variations communes à des milliers d'individus, notamment celui du projet international *HapMap* (dont la troisième et dernière phase s'est achevée en 2010) et du projet *1000 génomes* (achevé en 2015). Ces ensembles de données, combinés aux progrès de l'analyse statistique, ont

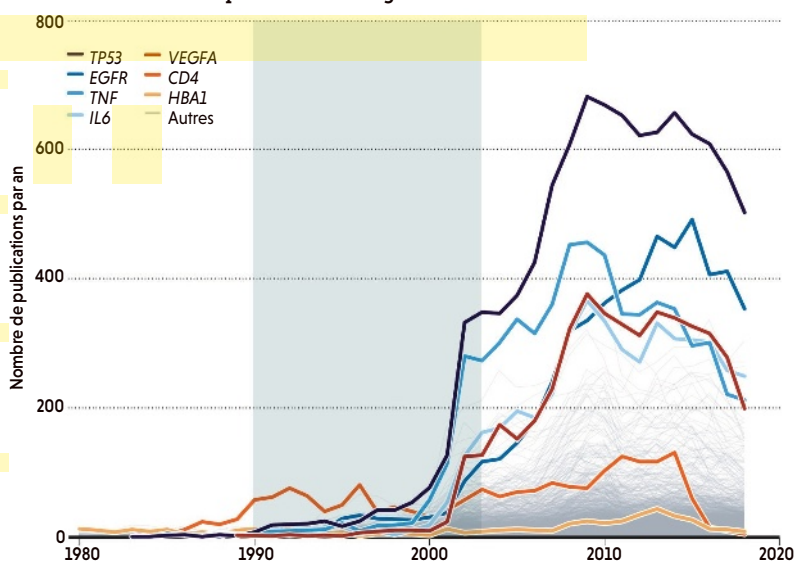
VINGT ANS DE POUBELLES, DE STARS ET DE MÉDICAMENTS

Ce que les chercheurs en génomique ont étudié, quand et pourquoi, tel que le retrace l'analyse bibliométrique.

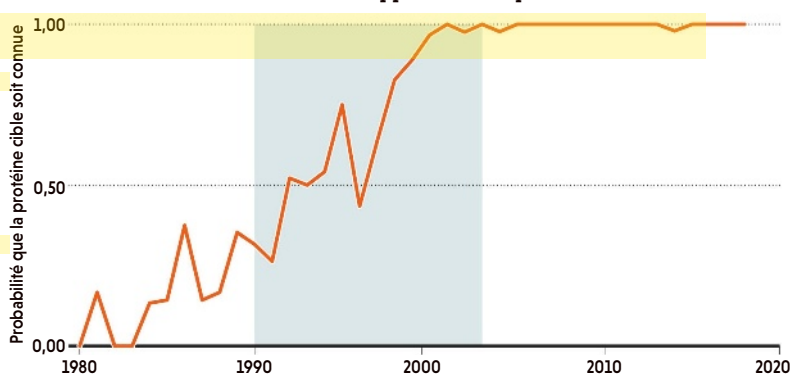
Éléments non codants. La plupart des gènes codant des protéines ont été découverts avant la première publication du projet *Génome humain* en 2001. De nombreux autres éléments du génome, auparavant nommés « ADN-poubelle », ont été étudiés ensuite.



Gènes stars. Le projet *Génome humain* a conduit à une explosion des recherches sur un petit nombre de gènes.



Cibles thérapeutiques. Depuis 2001, on a identifié les protéines cibles de près de 100 % des nouveaux médicaments approuvés chaque année aux États-Unis.

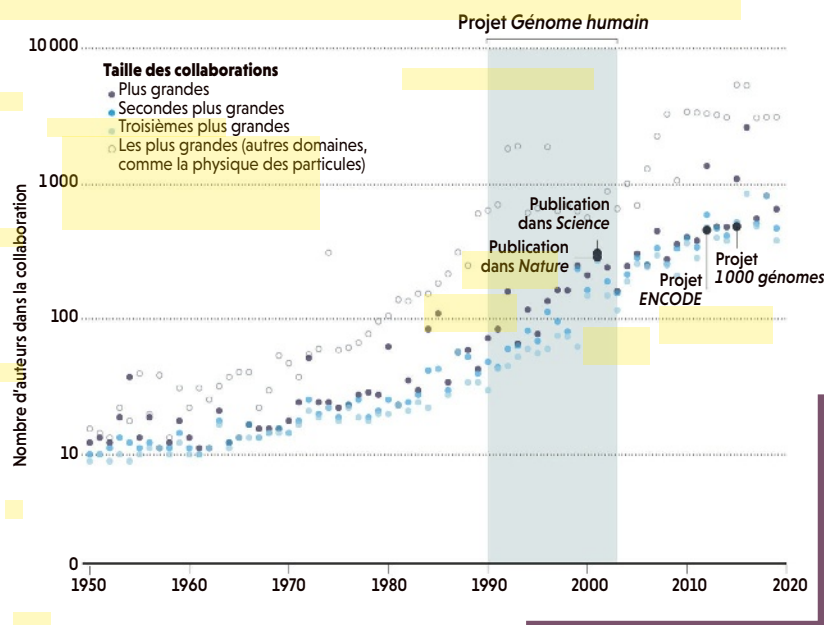


* Polymorphismes touchant un seul nucléotide (SNP), pseudogènes, ARN non codant, promoteurs, etc.

© Nature ; source : Barabási lab

PAS DE BOND DANS LA TAILLE DES ÉQUIPES

Selon une idée répandue, le nombre d'auteurs collaborant dans les publications liées au projet de séquençage du génome humain a brusquement augmenté après 2001. En réalité, la taille des équipes en biologie a progressé de façon continue depuis les années 1950.



inauguré les études d'association sur le génome entier (Gwas) pour d'innombrables traits phénotypiques tels que la taille, l'obésité et la sensibilité à des maladies complexes telles que la schizophrénie [ces études consistent à comparer le génome d'un grand nombre d'individus témoins à celui d'un grand nombre d'individus présentant le trait ou la maladie considérés, ndt]. Plus de 30 000 articles sont aujourd'hui publiés par an sur les liens entre SNP et traits phénotypiques. Une grande partie de ces associations se trouvent dans les régions non codantes du génome autrefois négligées.

Par ailleurs, les fonctions cellulaires reposent sur des liens plus ou moins directs entre le matériel génétique et les protéines. La cartographie de ce réseau complète désormais la perspective de l'hérédité mendélienne. Aujourd'hui, plus de 300 000 interactions ont été cartographiées – il s'agit de protéines se liant à des régions non codantes du génome ou à d'autres protéines.

DE NOUVEAUX MÉDICAMENTS

Avant les années 1980, la découverte de médicaments devait beaucoup au hasard. Leurs cibles moléculaires et protéiques étaient généralement inconnues. Jusqu'en 2001, la probabilité de connaître toutes les protéines ciblées par un médicament était inférieure à 50%. Le

projet *Génome humain* a changé la donne. Aujourd'hui, les cibles protéiques de presque tous les médicaments autorisés aux États-Unis chaque année sont connues.

Sur les quelque 20 000 protéines révélées par le projet *Génome humain* comme cibles potentielles de médicaments, notre analyse révèle que seulement 10% environ (2 149) jusqu'à présent ont été ciblées par des médicaments approuvés. Cela laisse près de 90% du protéome non exploré par la pharmacologie. Si on inclut les médicaments encore en phase d'expérimentation, le nombre de protéines ciblées passe à 3 119. Là encore, l'attention accordée à ces cibles est très inégale. Cinq pour cent de tous les médicaments actuellement approuvés (99 molécules distinctes) ciblent par exemple la protéine ADRA1A, impliquée dans la croissance et la prolifération cellulaires.

Comme pour l'étude des gènes, de bonnes raisons expliquent ce déséquilibre. Certaines protéines sont plus importantes que d'autres pour la santé ou plus susceptibles d'être utilisées comme cibles de médicaments. D'autres au contraire pourraient ne pas être des cibles exploitables. Mais peut-être que beaucoup plus de protéines mériteraient d'être étudiées comme cibles potentielles de médicaments si seulement les chercheurs, les bailleurs de fonds et les revues académiques étaient moins réticents à prendre des risques.

Cela dit, la majorité des médicaments efficaces ne visent pas directement les gènes individuels liés à une maladie. Ils ciblent plutôt des protéines situées une ou deux interactions plus loin dans le réseau de molécules impliquées, en modulant les conséquences de composants défectueux. Par exemple, un examen à grande échelle de médicaments existants pour estimer s'ils pourraient être redirigés contre le Covid-19 a révélé que seuls 1% des candidats prometteurs visaient une protéine du virus SARS-CoV-2 – la majorité modulaient des protéines humaines qui n'étaient pas directement impliquées dans son activité. Ces médicaments présentent un potentiel énorme.

UN APERÇU DU RÉSEAU

En résumé, nous pensons que le projet *Génome humain* est plus remarquable pour la nouvelle ère génomique qu'il a inaugurée que pour le catalogue de gènes codant des protéines lui-même. Comme le montre la théorie des systèmes complexes, une étude précise des composants est nécessaire mais pas suffisante pour comprendre un système. La complexité découle de la diversité des interactions des composants. Après vingt ans de recherches s'appuyant sur le projet *Génome humain*, les biologistes ont maintenant un aperçu de la structure et de la dynamique des réseaux qui définissent la vie. ■

BIBLIOGRAPHIE

D. M. Gysi et al., **Network medicine framework for identifying drug repurposing opportunities for COVID-19**, *arXiv*, 2020.

D. S. Wishart et al., **DrugBank 5.0 : A major update to the DrugBank database for 2018**, *Nucleic Acids Res.*, vol. 46, pp. D1074-D1082, 2018.

J. C. Venter et al., **The sequence of the human genome**, *Science*, vol. 291, pp. 1304-1351, 2001.

International Human Genome Sequencing Consortium, **Initial sequencing and analysis of the human genome**, *Nature*, vol. 409, pp. 860-921, 2001.

G. Bianconi et A.-L. Barabási, **Competition and multiscaling in evolving networks**, *Europhys. Lett.*, vol. 54, pp. 436-442, 2001.