

> 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.

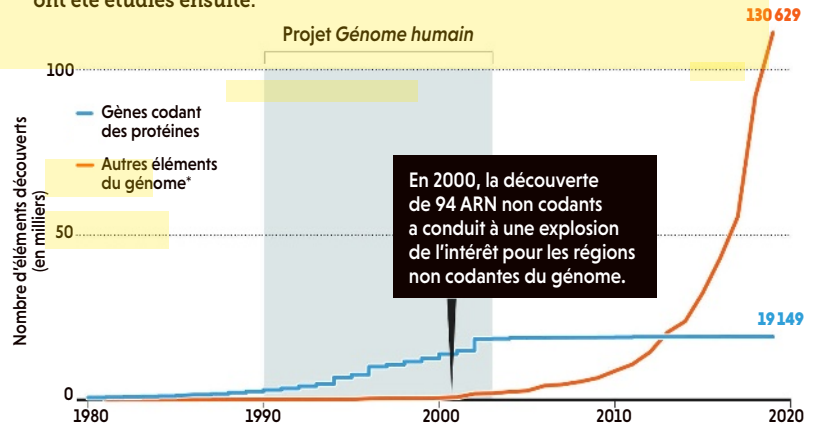
Avec la carte du génome humain en main, la découverte d'éléments non codants a explosé. Jusqu'à présent, le rythme a été cinq fois plus rapide que celui de la découverte des gènes codant des protéines et il ne montre aucun signe de ralentissement. De même, le nombre de publications portant sur ces constituants 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

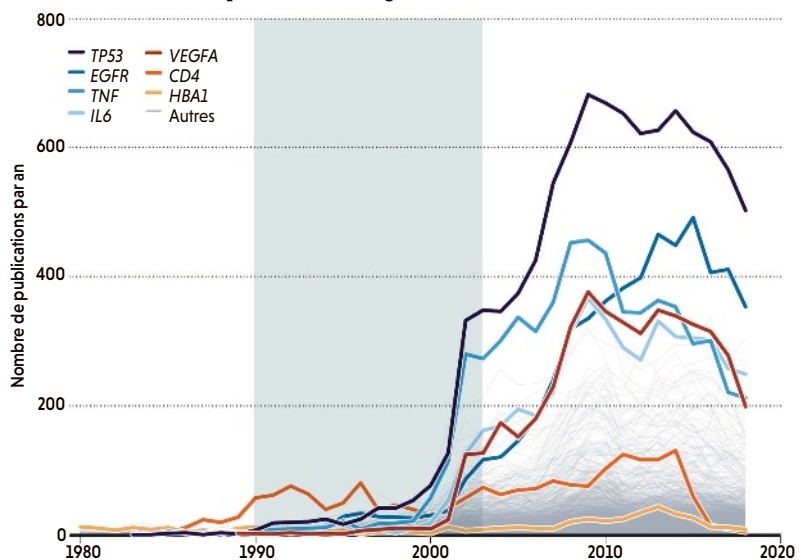
VINGT ANS DE POUBELLES, DE STARS ET DE MÉDICAMENTS

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

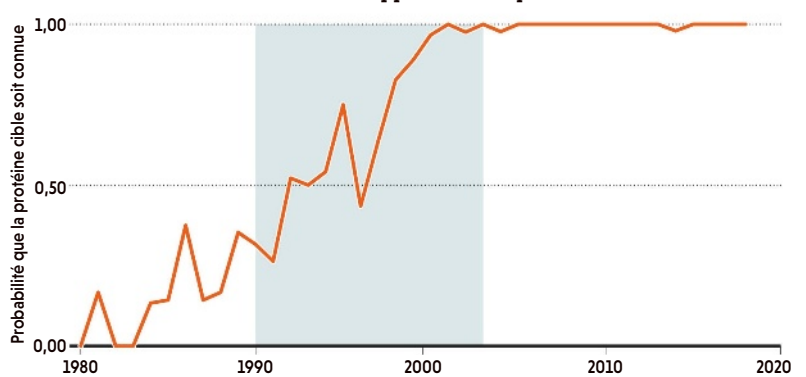
Éléments non codants. La plupart des gènes codant des protéines ont été découverts avant la première publication du projet *Génome humain* en 2001. De nombreux autres éléments du génome, auparavant nommés « ADN-poubelle », ont été étudiés ensuite.



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