



rien de plus qu'un index, mais un index sur une collection gigantesque de pages qui évoluent en permanence. Une prouesse algorithmique quand on considère les échelles: aujourd'hui, le moteur de recherche Google indexe 30000 milliards de pages et, chaque jour, 20 milliards de sites sont indexés et 3,3 milliards de requêtes sont effectuées.

Comment trouver, parmi les éventuelles centaines de millions de pages qui correspondent à une requête, les quelques pages qui contiennent l'information pertinente pour le lecteur? C'est la seconde grande fonction du moteur de recherche, le classement, qui fait émerger les pages «intéressantes». Dans le cas du moteur de Google Search, le principe utilisé historiquement pour le classement est celui de la «popularité sur le graphe du Web», c'est-à-dire la probabilité pour un flâneur qui se baladerait au hasard sur la Toile de se trouver sur une page particulière. L'algorithme PageRank permet de calculer cette popularité. Aujourd'hui, de nombreuses autres informations sont utilisées pour le classement, notamment les clics des internautes, qui votent ainsi avec leurs doigts. Le moteur de recherche sait ainsi faire surgir, de la masse du Web et des choix des internautes, une intelligence collective pour évaluer, classer, filtrer les informations.

Dans les années 2000, le développement de nouvelles techniques, désignées sous le terme de Web 2.0, a permis à tout un chacun non seulement d'accéder à des informations publiées par des autorités – administrations, universités, entreprises... – mais aussi de publier lui-même des informations. Ce mouvement a conduit quelques années plus tard à des réseaux sociaux entre humains.

Au fil des ans, il est devenu évident qu'au-delà d'une collection universelle de documents, le Web offrait une gamme infinie d'applications à inventer. Nous avons vu arriver le Web de la 3D avec ses mondes virtuels, le Web des objets. Peut-être l'évolution la plus fondamentale sera-t-elle le passage d'un Web des humains à un Web où humains et machines coexistent, en s'appuyant sur les services web, les données semi-structurées, et les connaissances. À chaque fois, l'existence d'un langage standardisé est un composant essentiel du progrès.

Si le Web a été conçu initialement comme outil pour la recherche, il est devenu quasiment impossible aujourd'hui de vivre sans: de trouver du travail, de travailler, de se loger, de gérer ses comptes bancaires, presque d'avoir des amis, etc. On peut cependant regretter le Web des débuts, libre, gratuit, universel, ouvert et anarchique. Le Web évolue vers des espaces plus fermés, notamment sous la pression de la monétisation des contenus ou des biais des systèmes de recommandation, vers des espaces plus hostiles avec ses rumeurs et fausses nouvelles, ses sites racistes ou criminels... Aujourd'hui, plus que jamais, le Web est une utopie à défendre: c'est à nous de décider ce que sera le Web demain. ■

BIBLIOGRAPHIE

S. Abiteboul et G. Dowek,
Le Temps des algorithmes,
Le Pommier, 2017.

J. Gillies et R. Cailliau,
**How the Web was born :
The story of the World
Wide Web**, Oxford
University Press, 2000.

LE GÈNE *BRCA1*, IMPLIQUÉ DANS LE CANCER DU SEIN

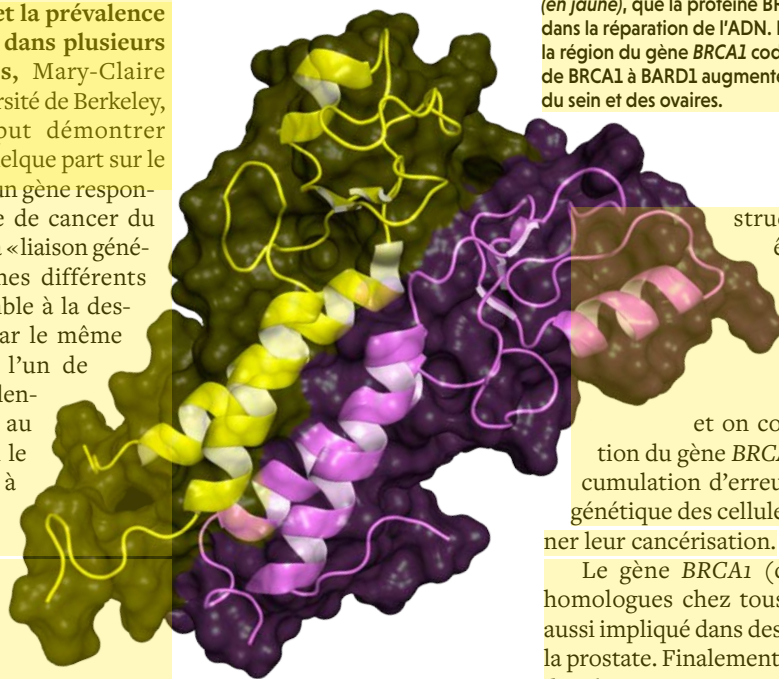
En cherchant des corrélations entre 183 gènes connus et la prévalence du cancer du sein dans plusieurs lignées familiales, Mary-Claire King, alors à l'université de Berkeley, aux États-Unis, put démontrer en 1990 qu'il devait y avoir, quelque part sur le bras long du chromosome 17, un gène responsable d'un risque héréditaire de cancer du sein. L'analyse se fondait sur la «liaison génétique» : le fait que deux gènes différents peuvent être transmis ensemble à la descendance s'ils sont portés par le même chromosome et assez près l'un de l'autre. Cette découverte déclencha une véritable course au séquençage, à une époque où le Projet génome humain venait à peine d'être lancé (voir page 20).

Les années suivantes, de nombreux travaux, notamment à Cambridge, en Angleterre, précisèrent la position probable du gène, désormais appelé *BRCA1* (pour *breast cancer*). Finalement, en 1994, Mark Skolnick, fondateur de la start-up Myriad Genetics en 1991, conjointement avec des chercheurs de l'université de l'Utah, des Instituts américains de santé et de l'université McGill, à Montréal, identifia et séquença le gène *BRCA1*.

Le dépôt d'un brevet sur *BRCA1* par Myriad Genetics a provoqué à l'époque un scandale qui a quelque peu occulté la portée scientifique du résultat. De nombreuses objections ont été soulevées, comme le fait que le brevet plaçait Myriad Genetics dans une situation de monopole qui empêchait les patients de demander un second avis en cas de dépistage positif, ou encore que la découverte dans la nature d'un gène ne pouvait pas donner plus de droits que, par exemple, la découverte d'une espèce animale. Dans les faits, les brevets sur *BRCA1* (ainsi que ceux déposés un an plus tard sur le gène apparenté *BRCA2*) se sont révélés difficiles à faire valoir en dehors des États-Unis.

Dans sa version normale, le gène *BRCA1* code une protéine (notée *BRCA1*) comportant une structure appelée «doigt de zinc» (où un ion zinc sert en quelque sorte «d'épingle» pour plier correctement la protéine). Ce type de

C'est notamment associée à une protéine, *BARD1* (en jaune), que la protéine *BRCA1* (en rose) intervient dans la réparation de l'ADN. Plusieurs mutations dans la région du gène *BRCA1* codant le domaine de liaison de *BRCA1* à *BARD1* augmentent le risque de cancer du sein et des ovaires.



structure est connu pour être capable de se lier à l'ADN. On sait désormais que la protéine *BRCA1* est un des outils du kit de réparation du génome et on comprend qu'une mutation du gène *BRCA1* puisse favoriser l'accumulation d'erreurs dans le programme génétique des cellules concernées et entraîner leur cancérisation.

Le gène *BRCA1* (dont on a trouvé des homologues chez tous les mammifères) est aussi impliqué dans des cancers ovariens ou de la prostate. Finalement, il apparaît que lorsque de tels cancers apparaissent chez des patients pourtant dotés d'une forme fonctionnelle de *BRCA1*, c'est parfois que d'autres facteurs empêchent son expression normale. Autrement dit, la découverte du gène *BRCA1* ne concerne pas seulement les quelque 5 à 10% de cancers du sein héréditaires.

Avant ces travaux, on pensait qu'en dehors de quelques maladies dont la transmission est évidemment mendélienne (comme la drépanocytose ou anémie falciforme, une maladie héréditaire touchant l'hémoglobine – la protéine qui transporte l'oxygène dans le sang –, due à une mutation sur les deux copies héritées du gène qui code la protéine), il était vain de chercher une origine génétique pour des pathologies vraisemblablement causées par un enchevêtrement de facteurs génétiques et environnementaux. Désormais, l'oncogénétique, l'étude des mécanismes du cancer à l'échelle moléculaire et génétique, permet non seulement des dépistages, mais aussi de comprendre les mécanismes moléculaires à l'œuvre dans certains cancers. ■

R. C.

Y. Miki *et al.*, A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*, *Science*, vol. 266, pp. 66-71, 1994