

> l'entreprise de postproduction hollywoodienne Technicolor est parvenue à archiver 1 million de copies du vieux film *Le Voyage dans la Lune*, de Georges Méliès, sur des brins d'ADN de synthèse, au lieu des fragiles microfilms habituels. Plus récemment, des chercheurs ont utilisé l'ADN comme outil pour concevoir des thérapies géniques, accélérer le développement de médicaments anticancer et même produire le premier «enregistrement en direct» de l'expression des gènes d'une cellule. Dans ce domaine en constante évolution, l'ADN est au cœur de recherches non seulement pour ses capacités de stockage, mais aussi pour ses possibilités de production de données à des vitesses sans précédent.

DES CODES-BARRES ADN

Au cours des dernières années, de plus en plus de scientifiques ont utilisé l'ADN comme «enregistreur moléculaire» pour suivre les résultats de leurs expériences. Le procédé implique souvent l'emploi de «codes-barres ADN»: afin de suivre les résultats d'une expérience, les scientifiques ont recours à une séquence d'ADN connue qui endossera le rôle d'étiquette moléculaire. Ils associent par exemple un résultat expérimental à la séquence ACTATC de nucléotides (les constituants de l'ADN), un autre à la séquence TCTGAT, et ainsi de suite.

Le principe du code-barres ADN est né au début des années 1990, lorsque Richard Lerner et Sydney Brenner, alors tous deux en poste à l'institut de recherche Scripps, en Californie, ont eu l'idée de l'utiliser comme un moyen de suivre des réactions chimiques. Leur concept était extrêmement innovant, mais en avance sur leur temps: les techniques permettant de séquencer facilement l'ADN à des coûts réduits n'existaient pas encore. Le potentiel de cette méthode a donc été exploité plus tard, quand le séquençage de nouvelle génération est apparu.

En 2005, une étape majeure a été franchie: des chercheurs ont réussi à analyser 25 millions de nucléotides en seulement 4 heures. Les progrès se sont poursuivis rapidement: on sait désormais lire des millions de séquences d'ADN en même temps, ce qui signifie qu'il est techniquement possible de réaliser des milliers d'expériences simultanément. L'analyse des expériences à codes-barres ADN relève donc à présent de la gestion de données. Au lieu de réaliser leurs tests un à un, les chercheurs peuvent émettre 20000 hypothèses, puis les tester toutes en même temps afin de déterminer lesquelles sont correctes.

Les biologistes ont été les premiers à recourir aux méthodes de codes-barres ADN. Puis, à mesure que le procédé est devenu plus accessible, des chercheurs issus d'autres domaines, dont l'ingénierie chimique et la science des matériaux, ont commencé à user de ces

Avec les codes-barres ADN, on peut tester 20000 hypothèses en même temps

techniques pour suivre leurs expériences à des échelles inédites. Par exemple, dans mon laboratoire de l'institut de technologie de Géorgie, aux États-Unis, des ingénieurs utilisent des codes-barres ADN pour améliorer la conception et le fonctionnement de nanoparticules. Ces nanoparticules sont destinées, *in fine*, à agir comme vecteurs, c'est-à-dire comme moyens de transport d'un médicament jusqu'à des cellules malades. Les nanotechnologies, qui reposent principalement sur l'ingénierie physique et chimique, semblent *a priori* sans rapport avec l'ADN. Mais si on considère l'ADN comme un outil de suivi et d'archivage de n'importe quelle donnée, son intérêt devient évident.

TESTER 500 NANOPARTICULES EN MÊME TEMPS

Pour les chercheurs en nanomédecine, il est plus aisé de concevoir des expériences destinées à trouver de nouvelles thérapies que de les réaliser et d'analyser les résultats. En effet, de nombreux paramètres comme la taille, la forme, la charge et la composition chimique des nanoparticules influent sur l'efficacité avec laquelle ces dernières transporteront un médicament jusqu'à des cellules malades. En outre, ces facteurs interagissent les uns avec les autres, ce qui complique la tâche des chercheurs. Une solution évidente serait de tester chaque nanoparticule individuellement. Mais, selon des données issues d'entreprises pharmaceutiques qui ont développé de telles nanoparticules, ce type de tests coûte plusieurs centaines de millions de dollars.

C'est là que les capacités de stockage de l'ADN présentent un réel intérêt. Pour augmenter le nombre de nanoparticules testées, on peut concevoir des milliers de nanoparticules aux structures chimiques variées (de grandes sphères chargées positivement ou de petits triangles à charge neutre, par exemple) et assigner un code-barres ADN à chacune d'elles. La nanoparticule 1, de structure chimique 1, porte le code-barres 1. La nanoparticule 2, de

structure chimique 2, le code-barres 2, etc. On peut ensuite administrer des centaines de ces nanoparticules à des cellules malades. Afin d'identifier les nanoparticules les mieux adaptées pour livrer un médicament, il suffit de séquencer l'ADN de ces cellules pour y quantifier les codes-barres présents. L'échelle de ces expériences est complètement nouvelle en nanomédecine. Dans mon domaine, à l'issue d'une expérience traditionnelle – le test d'une seule condition expérimentale –, on recueille en général une à cinq données. Mais d'ici à la fin de l'année, à l'aide de tels codes-barres ADN, mon laboratoire espère tester l'efficacité de 500 nanoparticules comme vecteurs dans 40 types de cellules différentes. Cela revient à pratiquer 20 000 expériences en même temps.

Pour ce faire, nous avons aussi besoin d'un outil capable de gérer ces données et de nous aider à les analyser. Nous avons donc d'abord testé la fiabilité de notre jeu de données: à quel

point les résultats issus d'une expérience reproduite plusieurs fois nous permettaient-ils de prédire l'issue d'une autre expérience? Puis nous avons fait appel aux statistiques pour savoir si certaines caractéristiques des nanoparticules, comme leur taille, influent sur l'introduction du médicament dans les tissus cibles. Nous avons ainsi découvert que la chimie des nanoparticules, et non leur taille, affectait leur efficacité. Avec cette technique, nous espérons découvrir de nouvelles thérapies géniques plus rapidement, en utilisant moins de ressources. L'un de nos objectifs est d'identifier une nanoparticule capable de cibler spécifiquement les tumeurs. Cela réduirait les effets secondaires des traitements habituels, telles les nausées et la perte de cheveux.

Nous avons déjà eu quelques succès. En 2018, en utilisant cette approche, nous avons rapidement identifié des nanoparticules capables de livrer des médicaments aux cellules qui tapissent les vaisseaux sanguins (les cellules endothéliales), ainsi qu'à de nombreux types de cellules immunitaires. Or, avec les traitements existants, nombre de protéines des cellules immunitaires sont souvent hors de portée des agents thérapeutiques – de petites molécules ou des anticorps. Grâce à nos découvertes, les atteindre et modifier leur activité devient possible. Ces travaux ont intéressé de nombreux chercheurs en thérapie génique. Nous avons donc décidé de fonder GuideRx, une entreprise de codes-barres ADN destinée au développement de nouvelles thérapies géniques.

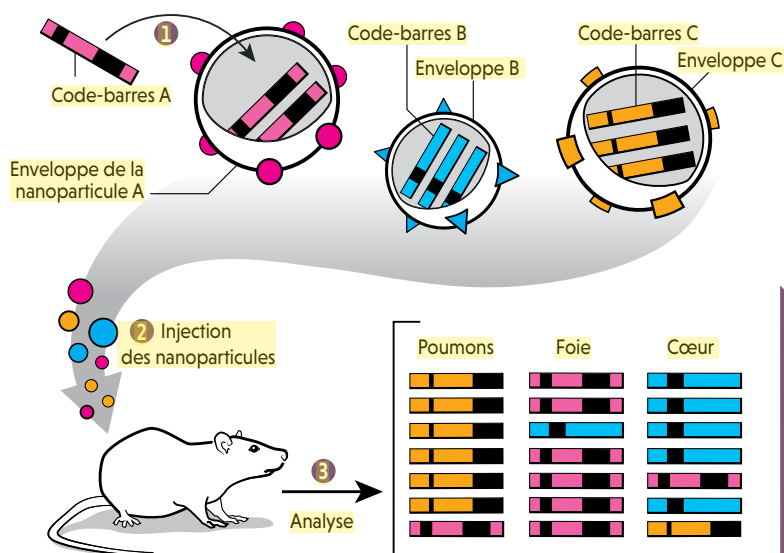
ÉTUDIER LES RÉSISTANCES DES CELLULES CANCÉREUSES

La technique des codes-barres ADN est devenue tellement commune que même au sein d'un seul domaine, elle est employée de diverses façons. C'est le cas, par exemple, en biologie du cancer, où l'on examine comment des mutations génétiques conduisent à des cancers et de quelle façon contrer les mécanismes en jeu. La résistance aux médicaments que les cellules cancéreuses développent chez les patients au fil du traitement est un défi majeur dans ce domaine.

En 2016, dans le laboratoire de Todd Golub, à l'université Harvard, des scientifiques ont eu recours aux codes-barres ADN pour étudier ces résistances. À l'aide d'un virus, ils ont inséré de façon permanente un code-barres ADN directement dans le génome de cellules cancéreuses. Une cellule cancéreuse de type A a reçu une séquence code-barres A, une cellule cancéreuse de type B, un code-barres B, etc. Les scientifiques ont ensuite mélangé les différentes cellules et les ont traitées avec un médicament contre le cancer. Une cellule sensible au traitement mourait ou ne se divisait plus. En revanche, si une cellule développait une

TRACER LES NANOPARTICULES AVEC DES CODES-BARRES ADN

Les thérapies géniques consistent à injecter du matériel génétique dans les cellules d'un organisme grâce à un vecteur – un moyen de transport capable de véhiculer le matériel génétique jusqu'au noyau des cellules. Si les vecteurs les plus courants sont dérivés de virus, on utilise aussi parfois des nanoparticules. Auparavant, tester l'efficacité de nanoparticules comme vecteurs était long et laborieux. Mais grâce aux codes-barres ADN, ces recherches sont beaucoup plus rapides: il est possible de réaliser des expériences avec des centaines de particules différentes en même temps. Pendant la phase de test (ci-dessous), on place un code-barres ADN différent à l'intérieur de l'enveloppe de chaque type de nanoparticule (1). Puis on injecte simultanément de nombreuses nanoparticules à un animal de laboratoire (2). Ensuite, on séquence l'ADN des cellules de l'animal pour déterminer quel code-barres elles contiennent: on sait ainsi quelle nanoparticule a atteint quel tissu (3). Grâce à cette méthode, les chercheurs identifient le modèle de nanoparticule qui convient le mieux selon l'objectif thérapeutique. Et le ciblage, plus spécifique, minimise les effets secondaires.



UN CHEVAL AU GALOP DANS UNE BACTÉRIE

Dans la plupart des recherches visant à coder des informations (images, textes, sons ou films) sur de l'ADN, les données sont archivées sur des brins d'ADN de synthèse. L'ADN est ensuite encapsulé dans des nanobilles de verre ou séché à l'air, ce qui permet de le conserver à très long terme. En 2017, Seth Shipman, alors à l'université Harvard, aux États-Unis, et ses

collègues ont testé une autre approche : stocker des images animées dans l'ADN de bactéries vivantes.

Leur image de départ était une photographie de main de 56 x 56 pixels de quatre couleurs de pixels différentes (quatre niveaux de gris, voir ci-dessous). En langage binaire, chaque couleur de pixel peut être codée par un duo de 0 et/ou de 1. Pour encoder l'image dans de l'ADN, les chercheurs ont associé chaque couleur de pixel à un nucléotide. Puis ils ont scindé les 3 136 pixels de l'image en 112 ensembles de 28 pixels, disséminés dans toute l'image. Chaque ensemble de pixels était ainsi codé par une succession de 28 nucléotides, à laquelle était ajouté un code-barres de 4 nucléotides permettant de reconnaître la zone et de la situer dans l'image. Les biologistes ont ensuite synthétisé les 112 fragments d'ADN obtenus en de nombreux exemplaires.

Pour intégrer ces fragments au sein du génome de bactéries *Escherichia coli*, ils ont détourné l'arsenal CRISPR de ces organismes – une sorte de système immunitaire qui enregistre, dans l'ADN bactérien, une trace des infections virales subies. Par électroporation, une technique qui consiste à appliquer un champ électrique sur les membranes cellulaires afin de les perméabiliser, les chercheurs ont introduit les fragments d'ADN dans les bactéries. Grâce à la machinerie CRISPR, dont les biologistes avaient forcé l'activation, les fragments d'ADN ont été copiés au sein du génome des bactéries. Le jour suivant, les chercheurs ont séquencé l'ADN de la population bactérienne,

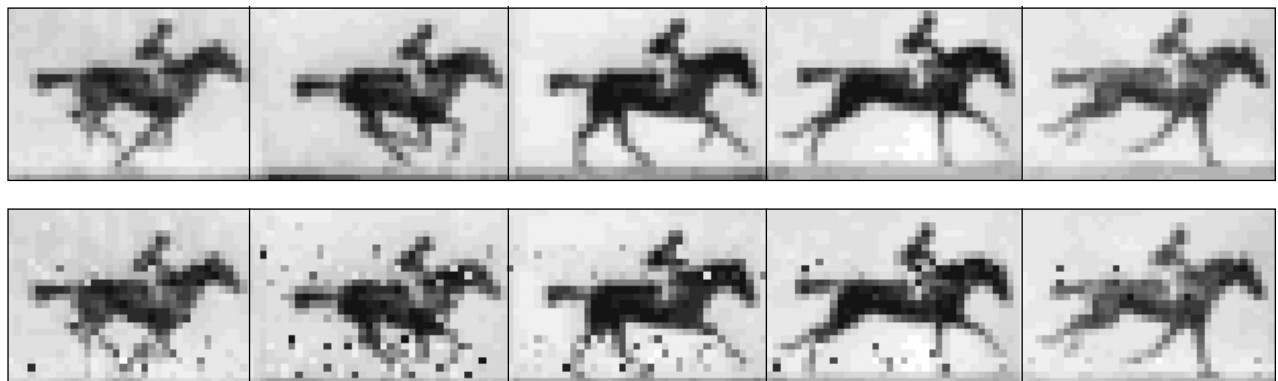
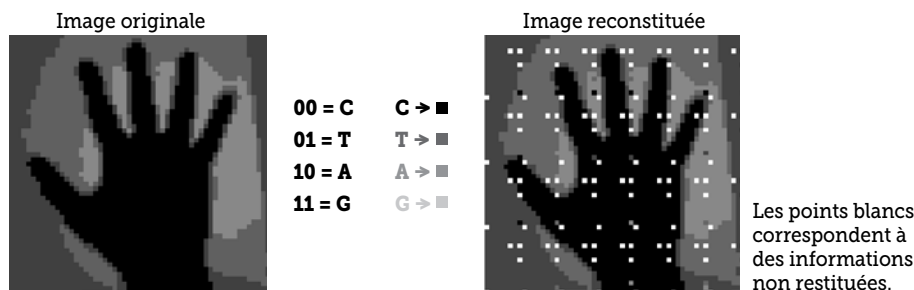
puis, grâce à un programme, ils ont converti les séquences nucléotidiques en langage binaire et ont ouvert l'image : 88 % des ensembles de pixels étaient correctement restitués. Les chercheurs ont ensuite réalisé la même expérience avec une image de main de 33 x 33 pixels comportant cette fois 21 couleurs de pixels. Les couleurs de pixels étaient codées non plus sur un nucléotide, mais sur un triplet. Avec cette méthode, 96 % des ensembles de pixels étaient fidèlement restitués 2 heures 40 après l'électroporation. Ensuite, le pourcentage de reconstitution de l'image diminuait légèrement dans les 24 heures, puis restait stable pendant les 6 jours suivants, au cours desquels 48 générations de bactéries se succèdent.

Les scientifiques ne se sont pas arrêtés à ce premier succès : ils ont souhaité enregistrer une séquence de cinq images issues du premier film d'Eadweard Muybridge, mettant en scène un cheval au galop et sa cavalière. Pour cela, ils ont exploité une caractéristique de l'arsenal CRISPR : lorsque les bactéries copient de l'ADN étranger au sein de leur propre ADN, l'ADN est copié dans un ordre précis selon la chronologie d'apparition. Ainsi, sur le même principe que pour l'image de main, ils ont d'abord introduit les brins d'ADN constituant la première image. Puis, le lendemain, ils ont refait de même avec ceux de la deuxième image, et ainsi de suite pendant cinq jours, réussissant de cette façon à coder une image animée dans de l'ADN bactérien.

En réalisant ces travaux, les chercheurs imaginaient-ils un futur où nos films favoris seraient stockés dans le génome de populations bactériennes ? Non, car les brins d'ADN de synthèse sont plus appropriés à cette fin. « Mais ce qui est vraiment intéressant dans cette étude, commente Christophe Antoniewski, directeur de la plateforme de bio-informatique ARTbio de l'institut de biologie Paris-Seine, c'est la perspective d'utiliser l'ADN comme enregistreur moléculaire *in vivo* de l'activité des cellules, afin d'étudier, par exemple, leur développement, ou leur dérèglement dans les cancers. »

HÉLÈNE GÉLOT
Journaliste scientifique

AVANT / APRÈS



La séquence originale (*en haut*) et celle reconstituée à partir des bactéries (*en bas*) ne diffèrent que de quelques pixels.