

CES VARIATIONS QUI NOUS DISTINGUENT

Les tests génétiques reposent sur l'étude de la variation génétique entre individus, ou polymorphisme. Ce terme traduit le fait que tous les individus d'une même espèce ne sont pas identiques génétiquement : à certains endroits du génome, ils présentent différentes versions d'une même séquence, nommées allèles. Les endroits variables sont nombreux, mais très ciblés : un génome humain compte plusieurs millions d'endroits variables, mais deux génomes humains tirés au hasard ne diffèrent qu'au niveau de 0,1% de leur séquence d'ADN. La forme la plus commune de variation de séquence entre deux individus est le polymorphisme touchant un seul nucléotide (les nucléotides sont les constituants de l'ADN ou SNP (pour *single-nucleotide polymorphism*). Si quatre états alléliques sont théoriquement possibles à une position donnée (l'ADN n'est constitué que de quatre nucléotides distincts, notés A, T, G, C et assemblés en une longue chaîne), seuls

deux allèles sont observés pour la grande majorité des SNP. Les SNP sont stables, abondants et distribués uniformément dans tout le génome. L'une des méthodes les plus performantes actuellement pour mettre en évidence un polymorphisme SNP est l'usage des puces à ADN : sur une lame, on fixe des centaines de milliers de fragments d'ADN complémentaires des régions à SNP et l'on regarde à quelles versions de ces fragments l'ADN testé se lie. Un autre marqueur moléculaire très utilisé est le microsatellite. Il est constitué d'une répétition continue d'un motif court (de 1 à 4 nucléotides le plus souvent). Le nombre variable de répétitions du motif détermine les différents allèles du marqueur. Moins abondant dans le génome que le SNP, le microsatellite présente cependant l'avantage d'être plus informatif du fait de son multi-allélisme : un même marqueur peut avoir plusieurs dizaines d'allèles.

D'un individu à l'autre, le génome varie en des endroits très ciblés, comme les SNP (variations d'un seul nucléotide) ou les microsatellites (répétitions d'un motif court).

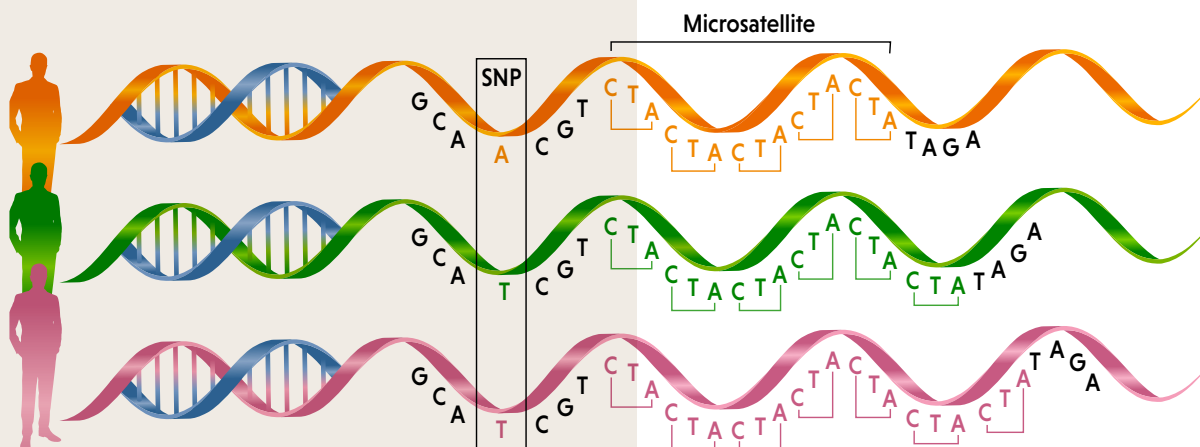
peu médiatisées. Mais en 2007, l'entrée sur le marché d'entreprises analysant l'ADN de façon beaucoup plus systématique a changé la donne.

POUR UNE POIGNÉE DE DOLLARS

«Genetics just got personal» : avec ce slogan, la société 23andMe s'est alors mise à vendre, pour 1000 dollars, un service fondé sur la caractérisation de plusieurs milliers de SNP répartis sur tout l'ADN. Les deux autres sociétés qui s'étaient lancées à la même époque n'étaient pas en reste. Quand l'islandaise DecodeMe affirmait : «C'est mon code», l'américaine Navigenics plaissait pour : «Mes gènes, ma santé, ma vie, mon guide». Ces masses de données génétiques étaient ensuite transformées en informations personnalisées portant sur la santé ou sur les origines.

Les sociétés 23andMe et DecodeMe ont aussi proposé aux utilisateurs de comparer leurs marqueurs génétiques avec ceux d'autres utilisateurs du service, de façon à détecter d'éventuels parents biologiques plus ou moins éloignés. En effet, la nouvelle quantité d'information génétique disponible permettait désormais de détecter des cousins au deuxième, voire parfois au troisième degré. Pour faire parler d'elle, la société 23andMe a même proposé à des célébrités de se soumettre au test, et a offert la possibilité d'obtenir le pourcentage d'ADN que l'on partage avec elles.

Cette attention médiatique nouvelle tenait aussi aux nombreuses fées qui se sont alors penchées sur le berceau de ces entreprises pionnières. On y trouve de nouveaux investisseurs comme Google, mais aussi des fonds de capital-risque. On y trouve également des personnalités scientifiques de premier plan qui acceptent d'entrer dans les conseils scientifiques des sociétés ou qui affichent publiquement leur soutien. Ainsi, au sommet économique mondial de Davos en 2007, Francis Collins, alors directeur des Instituts américains de la santé et ancien leader



DU TEST GÉNÉTIQUE AUX ORIGINES ETHNIQUES

Si les entreprises comme 23andMe vantent la simplicité de leurs tests génétiques portant sur les origines ethniques, l'analyse de ces tests est autrement plus complexe. La méthode consiste à examiner des centaines de milliers d'endroits variables du génome (des marqueurs SNP). L'opération revient ensuite, pour chaque segment du génome de l'individu, à rechercher la population humaine dans laquelle le profil de variation que présente l'individu est le plus fréquent. En considérant l'un après l'autre tous les segments du génome, des modèles statistiques permettent de sélectionner les populations les plus informatives pour reconstruire l'ADN entier de l'individu, comme si l'on reconstruisait un puzzle en prenant dans chaque population existante ce qui ressemble le plus à chaque pièce originale du puzzle. Ces modèles permettent ainsi de déterminer les proportions de chaque population qui optimisent la reconstruction. Ces proportions sont ensuite interprétées comme des probabilités d'origine de l'ADN de l'individu. Ainsi, une personne verra l'origine de ses gènes localisée par exemple à 45 % en Angleterre, 31 % en Irlande-Écosse, 18 % dans la péninsule ibérique et 6 % en Inde, avec parfois un degré de précision étonnant : « Vous avez 0,2 % d'origine sicilienne »...

> du projet *Génome humain*, est-il venu en personne encourager les deux fondatrices de 23andMe, Anne Wojcicki et Linda Avey.

LA COURSE AUX DONNÉES

Sous les feux des projecteurs, les tests génétiques destinés à la recherche des origines ont vite montré des limites. Elles n'ont cependant conduit à aucune critique ou interdiction par la FDA (l'agence américaine qui régleme l'accès au marché des produits alimentaires et des médicaments), dont le périmètre n'inclut pas les services génétiques portant sur les origines. Comme pour les tests génétiques de santé, ce n'est pas l'ADN entier qui est analysé, mais des centaines de milliers de marqueurs SNP. Les allèles identifiés chez l'individu testé sont ensuite comparés à la fréquence de ces mêmes allèles dans plusieurs populations vivant dans diverses régions du monde (voir l'encadré ci-dessus).

Pour constituer une telle base de référence, les laboratoires de généalogie génétique utilisent les données publiques des grands projets

de recherche internationaux explorant la diversité génétique de notre espèce tels que le projet HGDP (*Human genome diversity project*), comptant 1000 individus issus de 51 populations, le projet *Hapmap* (1200 personnes de 11 populations) ou le projet *1000 Genomes* (2500 individus de 26 populations).

En outre, les laboratoires enrichissent leur base de référence avec les données génétiques collectées directement auprès de leurs clients (80 % des utilisateurs de 23andMe autorisent le partage de leurs données, selon l'entreprise) ou auprès des clients des laboratoires concurrents. En témoigne le grand projet de collecte *One family one world*, lancé en 2017 par la société britannique LivingDNA, qui encourage toute personne ayant déjà réalisé un test génétique à télécharger ses données brutes sur sa plateforme afin de constituer le plus large panel possible d'origines mondiales.

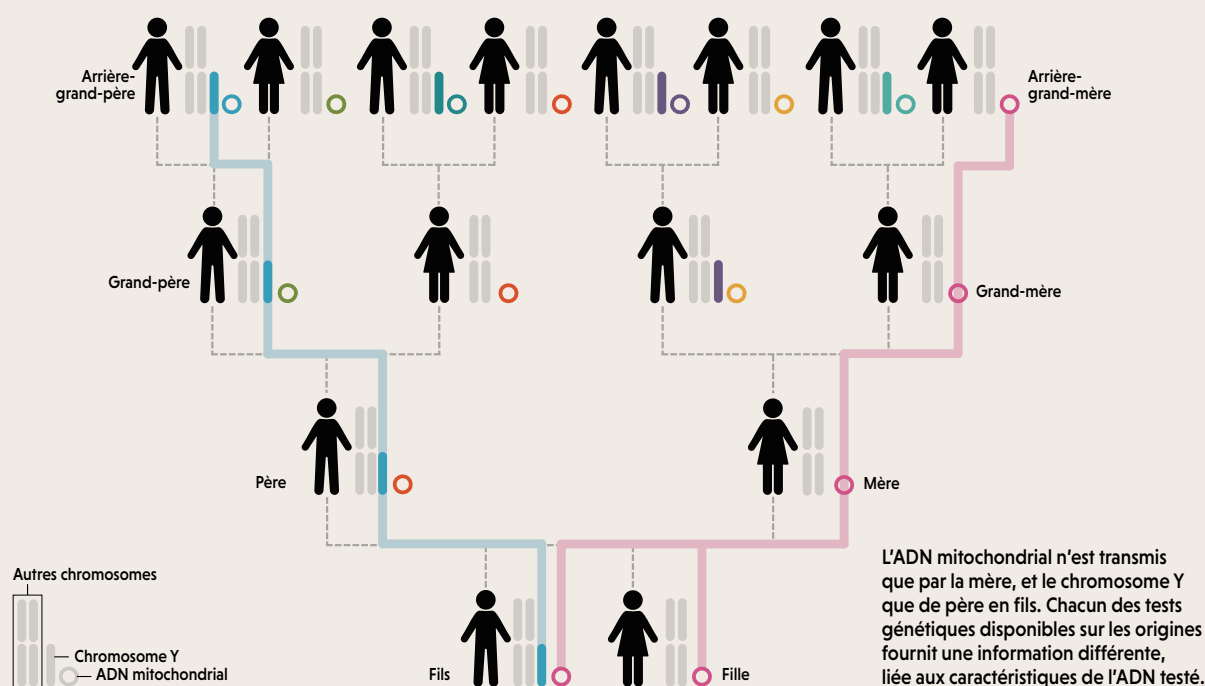
Enfin, certains laboratoires vont chercher directement sur place les échantillons d'ADN d'intérêt pour répondre à leurs problématiques spécifiques. C'est le cas de la société African Ancestry, spécialisée dans les origines africaines des Afro-Américains, qui dispose, dans sa base privée, de données sur plus de 200 populations ou ethnies vivant actuellement en Afrique. Certains laboratoires offrent d'ailleurs des kits gratuits aux individus dont les quatre grands-parents sont nés dans des régions sous-représentées dans leur base.

Ici se situe la première limite des analyses proposées : la précision et la fiabilité des résultats obtenus dépendent largement de la qualité et de la taille de la base de référence. Or, loin d'être exhaustives, ces bases ne contiennent qu'un nombre limité d'échantillons de personnes contemporaines, regroupées de façon arbitraire en ensembles géographiques ou ethniques sous le nom de « populations ». De fait, le sens exact de ces « populations » varie d'un laboratoire à l'autre selon les critères utilisés pour les définir (nationalité, langue, territoire, religion, histoire). En 2008 déjà, la Société américaine de génétique humaine considérait que même les meilleures bases de données actuelles ne représentaient qu'un échantillonnage médiocre de la diversité génétique humaine mondiale.

Aujourd'hui, les bases ont certes grossi, mais restent très insuffisantes. Ainsi, une jeune femme née au Vietnam et adoptée à l'âge de 1 mois qui souhaiterait connaître les ethnies dont elle pourrait être issue ne trouverait pas de réponse satisfaisante à ses interrogations, le panel le plus complet à l'heure actuelle ne distinguant qu'une seule origine pour le Vietnam sur la cinquantaine d'ethnies recensées dans ce pays. S'il paraît possible d'attribuer le génome d'une personne à un continent particulier avec une bonne probabilité statistique, tout résultat plus précis doit être considéré avec prudence.

TESTS GÉNÉTIQUES, MODE D'EMPLOI

L'histoire commence toujours de la même façon : un échantillon de salive prélevé soit en tamponnant l'intérieur de sa joue, soit en crachant dans un tube, à renvoyer au laboratoire qui avait fourni le kit de prélèvement. Puis l'entreprise explore environ 700 000 variations ponctuelles (SNP) de notre génome par une technique de génotypage à haut débit utilisant des puces à ADN, moins précise que le séquençage complet, mais moins chère et plus facile à utiliser. Certaines variations offrent des informations spécifiques sur la prédisposition à diverses maladies ou sur une multitude de caractéristiques physiologiques (couleur des cheveux, tolérance à la caféine, sensibilité au goût amer...) – informations révélées aux consommateurs qui ont souscrit aux options « santé » ou « traits » proposées par certaines sociétés. La majorité des autres variations sont utilisées à des fins de généalogie génétique. Trois principaux types de tests d'origine sont disponibles, chacun apportant une information différente.



LE TEST AUTOSOMAL

Très majoritaire, il consiste à analyser les variations génétiques situées sur les 22 paires de chromosomes non sexuels présents dans le noyau de nos cellules et transmis par nos parents. L'analyse d'un sous-ensemble de variations judicieusement sélectionnées, incluant notamment des marqueurs nommés AIM (*ancestry informative markers*), permet d'inférer l'origine biogéographique d'un individu, c'est-à-dire de fournir une estimation de ses régions d'origine. Ces AIM sont des SNP qui se caractérisent par de fortes variations de fréquence entre populations appartenant à différents groupes géographiques ou ethniques et ont par conséquent un fort pouvoir discriminant pour différencier ces groupes. Avec ce test, on est ainsi capable de remonter jusqu'à la huitième génération, soit environ 200 ans. On l'utilise aussi pour rechercher des cousins génétiques dans les bases de données ou estimer le pourcentage d'ADN provenant de l'homme de Néandertal.

LE TEST DE L'ADN MITOCHONDRIAL

Il se focalise sur les variations du génome mitochondrial, dont la transmission se fait quasi exclusivement par la mère. Distinct du génome contenu dans le noyau, ce petit segment circulaire d'ADN se situe dans les mitochondries, petits organites présents dans le cytoplasme de nos cellules. Cet ADN se transmet inchangé de mère en fille, à de rares mutations près que l'on utilise pour définir des groupes d'individus portant les mêmes mutations. Le test permet ainsi de déterminer le groupe d'origine de la lignée maternelle. Certaines entreprises proposent des tests plus spécifiques en procédant à un séquençage des régions hypervariables de l'ADN mitochondrial, particulièrement informatives, ou à un séquençage complet de la molécule. Si ce test permet de remonter 20 à 100 générations (500 à 2 500 ans), il ne donne accès qu'à un seul ancêtre maternel sur le million d'ancêtres possibles d'un individu 20 générations plus tôt (en l'absence de consanguinité).

LE TEST DU CHROMOSOME Y

Il n'est accessible qu'aux hommes puisque seuls ces derniers ont un chromosome Y. Contrairement aux chromosomes non sexuels, la majorité (95 %) de l'ADN de ce chromosome ne subit pas de recombinaison (pas d'échange avec le chromosome homologue) lors de la formation des spermatozoïdes, aussi les mutations apparaissant sur cette molécule sont-elles transmises de père en fils. Quelques centaines de mutations ponctuelles du chromosome Y sont utilisées pour définir des groupes de mutations associés à différentes lignées masculines. Le test permet d'identifier le groupe d'origine de la lignée paternelle d'un individu, mais ne donne accès qu'à un seul de ses ancêtres (le père du père du père, etc.). Une femme peut demander à un membre masculin de sa famille de se faire tester (frère ou père) pour obtenir ces résultats.