

## Le projet «Génome humain» et l'Open Source

Ces dernières années, j'ai pu observer l'impact de la mise en ligne, en temps réel, des résultats de la recherche, en particulier dans le cadre du projet «Génome humain». J'y ai participé de manière un peu parallèle: ma spécialité concerne les profils d'expression des gènes grâce aux «puces à ADN»<sup>1</sup> et je suis arrivée au Généthon (en 1993) au moment où l'on commençait à faire le travail préalable au séquençage, c'est-à-dire les cartes génétiques.

Or le projet de séquençage du génome avait été conçu pour une diffusion publique et immédiate des résultats de la recherche. Les chercheurs mettaient en ligne leurs résultats dès qu'ils en obtenaient, en temps réel, dans une base de données ouverte. Ce sont des données descriptives: si l'on prend par exemple les microsatellites (courtes séquences d'ADN caractérisées par la répétition d'un motif) utilisés lors de la construction de la carte génétique, on peut comparer le fait de les rendre publics à la mise en place de panneaux de signalisation sur une route: ce n'est pas cela qui vous permet de circuler mais, sans eux, vous ne pouvez pas vous diriger! En ce qui concerne le séquençage d'un gène, une séquence n'a pas d'intérêt en soi, elle reste «figée» tant qu'on ne l'utilise pas, mais elle est en quelque sorte le matériel de base pour pouvoir ensuite faire de la recherche sur les maladies. La question se pose dans les mêmes termes, en matière de diffusion et de protection des données,

pour les gènes, pour les puces ADN ou pour toute autre biotechnologie.

Le fait de rendre publiques des cartes génétiques ou des données de séquençage n'empêche pas la reconnaissance de la paternité des découvertes en matière de décryptage du génome. Cette reconnaissance, qui passe toujours, selon le modèle scientifique traditionnel, par la publication et l'*Open Access*, n'en modifie pas sensiblement les formes et les modalités. Une fois la paternité établie, les données du génome n'ont d'intérêt que si les chercheurs les étudient et les utilisent dans leur propre projet de recherche.

Le consortium académique international qui travaillait dans le cadre du *Human Genome Project* sur le projet de séquençage demandait que soient immédiatement mis à disposition les résultats des recherches qu'il avait contribué à financer. Cependant, dès le début, la logique a été compliquée: en même temps que le consortium académique prescrivait la mise en ligne ouverte des résultats, une société privée, Celera Genomics, créée par Craig Venter, s'attaquait au décryptage du génome. Quoi qu'il en soit, le génome a été publié dans sa totalité en 2002-2003 et, aujourd'hui, on continue à tout mettre sur le Web.

En parallèle, se posait la question du brevetage du génome. Je considère pour ma part, tout comme la majorité des membres de la communauté scientifique, qu'il n'y a pas de brevet possible sur les séquences

elles-mêmes. Cela poserait la question de savoir à qui appartient le vivant. Ces données sur les gènes sont des ressources. Breveter aurait bloqué toute recherche ultérieure. Et pourtant, des acteurs privés ont breveté des séquences de gènes, malgré un tollé général.

La mise en ligne des résultats en temps réel a eu également un impact sur le besoin de standardisation dans la présentation des résultats. Par exemple, dans le domaine des puces ADN sur lequel j'ai beaucoup travaillé, il s'est mis assez vite en place un système destiné à faire que tout le monde présente les données avec un minimum d'informations associées (MIAME) dans un format standardisé, notamment la façon dont avaient été conduites les expériences et la manière dont les données avaient été analysées. Bientôt, de nombreuses revues décidèrent de n'accepter un article que si les données étaient déposées dans des bases ouvertes (la revue *Nature* la première).

Cette façon de travailler, en mettant en ligne les données et avec des techniques permettant de travailler à grande échelle (comme le permettent les puces ADN), a changé le métier. Les chercheurs ont dû changer de mentalité. Avant, chacun étudiait son gène ou sa protéine, et publiait sur son gène ou sur sa protéine. Il n'y avait pas de ressources collectives. On posait une hypothèse et on la vérifiait (*hypotheses driven*). Lorsque des tonnes de données ont été mises à la disposition des chercheurs, on est passé à un mode de fonctionnement que l'on pourrait quali-

fier de *resource driven*: on arrivait à des résultats en travaillant sur une masse de données et sans avoir fait d'hypothèse au départ. C'est le matériau qui fait émerger les hypothèses. Je l'ai bien ressenti quand on a produit les premiers résultats des puces ADN: les chercheurs se sont mis à exploiter les données, qui apportent des réponses à des questions qu'ils ne pouvaient pas avoir avant de consulter cet amas de données. C'est un véritable changement de perspective, lié à ces approches à grande échelle.

Bilan ? D'un côté, l'*Open Source* favorise l'accessibilité aux connaissances et aux ressources biologiques ; de l'autre, la privatisation gagne du terrain (dépôts de brevets sur les séquences, bases de données payantes ou protégées par le secret commercial, contrats d'accès exclusif aux données génétiques et médicales des populations, etc.). Le brevet sur le gène du cancer du sein est l'emblème des dérives d'une telle privatisation, mais il y a finalement peu d'exemples de ce type. Aujourd'hui, la mise à disposition des données est telle, que cette privatisation partielle ne pose guère de problèmes dans de nombreux domaines de la génétique. Les chercheurs continuent leurs travaux sans en être gênés.

Geneviève Piétu,  
Inserm et laboratoire I-Stem  
(propos recueillis par Cécile Méadel)

#### NOTE

1. Une puce à ADN est un capteur bio-informatique formé de molécules d'ADN disposées en rangées ordonnées sur une petite surface de verre, de silicium, ou de plastique. Cette biotechnologie permet d'analyser le niveau d'expression des

gènes dans une cellule, un tissu, un organe, un organisme, à un moment donné et dans un état donné par rapport à un échantillon de référence.