

séance du lundi 27 septembre 1999

## ATOUTS ET ESPOIRS DE LA GENOMIQUE FRANÇAISE

*Pascal BRANDYS*

On ne peut commencer à faire un exposé sur le génome sans revenir à l'ADN. Lorsque nous avons créé notre société en 1989, nous étions encore obligés d'expliquer ce qu'était l'ADN. Depuis, évidemment, la connaissance sur l'ADN a beaucoup progressé, y compris dans le grand public depuis le film *Jurassic Park*, et nous disposons aujourd'hui incontestablement d'un sujet sur lequel nous pouvons évoquer les choses beaucoup plus en détail. Je rappellerai simplement que nous sommes tous constitués de dizaines de milliards de cellules et que, dans le noyau de chacune de ces cellules, nous avons une molécule porteuse du Code génétique, l'ADN, que l'on peut imaginer - c'est en général l'analogie que l'on fait - comme un collier de perles comprenant environ 3 milliards de perles différentes, 3 milliards d'éléments indépendants avec quatre couleurs, qui correspondent aux quatre bases fondamentales du Code génétique. Ce collier, qui mesure en vérité environ 1 m de long, est replié dans le noyau de chaque cellule. Le premier objectif de la génomique est de déterminer ce code fondamental. Le deuxième objectif, et sans doute le plus important, est d'étudier la diversité génétique. Nous sommes, bien entendu, tous différents les uns des autres et nous différons par des variations dans le Code génétique. Ces variations néanmoins sont tout à fait marginales. Environ une perle, un élément sur 1000, diffère d'un individu à l'autre. Et peut-être peut-on évoquer aujourd'hui la question, que l'on connaît moins, des différences entre espèces, ce qui conduira peut-être à un peu d'humilité. Les différences entre le Code génétique de l'homme et du chimpanzé sont de moins de 2 % ; nous partageons donc environ 98 % du Code génétique avec le chimpanzé. Même avec des animaux supérieurs comme la souris, environ 70 à 80 % du Code génétique sont communs et, dans le cas de la drosophile, nous avons plus de la moitié des gènes communs avec la mouche du vinaigre.

À quoi est dû le développement de la génomique et à quoi est due cette explosion ? Ce sont incontestablement deux innovations majeures, deux sauts technologiques qui ont permis finalement l'étude de l'ADN à grande échelle, au cours des dix dernières années. Ces deux innovations ont donné lieu chacune à un prix Nobel à leur auteur. La première, c'est la PCR, c'est-à-dire l'amplification en chaîne polymérase, due à Kary Mullis, qui permet, en utilisant les propriétés d'un enzyme qui réplique naturellement l'ADN, de répliquer de courts fragments d'ADN, de l'ordre de quelques centaines ou de quelques milliers de bases. La deuxième révolution est celle du séquençage de la lecture du Code génétique, qui permet, également à partir de courts fragments, de lire la couleur des perles du collier, c'est-à-dire une véritable lecture des fragments, ou lecture du Code génétique. Il faut aussi noter que ces deux inventeurs sont très différents puisque, dans le premier cas, il s'agit d'un jeune surfeur californien qui a inventé la PCR au sein d'une entreprise privée appelée Cetus. Cette société a déposé des brevets qui ont débouché ensuite sur de nombreuses applications et des licences dans le monde entier. Dans le deuxième cas, celui de l'invention de M. Sanger, il s'agit d'une invention qui a été réalisée au MRC en Angleterre, pur produit de la recherche publique britannique, qui, aujourd'hui, est dans le domaine public.

Il n'empêche que ces deux prix Nobel étaient complémentaires et c'est de cette complémentarité que vient aujourd'hui la possibilité d'étudier le génome dans son ensemble à très grande échelle. C'est ce que nous appelons la génomique. Ce travail d'analyse et d'étude de l'ensemble du génome humain est en fait un travail considérable, car, vous l'avez compris, il ne peut être mené que sûr des fragments et non pas sur l'ensemble du génome. On ne peut pas commencer à étudier le génome humain par une extrémité et finir par la fin, on est obligé de commencer par fragmenter le génome humain et ensuite étudier chacun de ces fragments de façon indépendante. Trois tâches sont donc essentielles : fragmentation d'abord, ensuite séquençage, ou lecture de ces différents éléments, et puis, enfin, reconstitution d'un gigantesque puzzle à la dimension de plusieurs millions d'éléments.

Ce travail extraordinairement complexe a nécessité des étapes intermédiaires. La première de ces étapes était la constitution de ce qu'on a appelé les cartes du génome humain, et la France a joué un rôle important dans ce domaine, puisque les premières cartes du génome humain ont été réalisées en 1993, par les Pr. Daniel Cohen et Jean Weissenbach. Ce travail continue aujourd'hui et est entrepris en parallèle dans de nombreux laboratoires publics et privés. On estime qu'à ce jour nous disposons d'environ 30 % de la séquence brute du génome humain, et, donc, la totalité de cette séquence sera probablement connue aux alentours de l'année 2002, au plus tard de l'année 2003. Il faut néanmoins bien comprendre que ce travail qui porte sur l'analyse brute du génome n'est qu'un début, car en fait cette séquence brute n'est que le point de départ d'un travail beaucoup plus complexe qui porte sur la structure elle-même du génome humain. Seule une fraction faible du génome code pour les molécules actives que sont les protéines, ces molécules qui réalisent les différentes fonctions et sont les travailleurs dans la cellule. On estime aujourd'hui qu'environ 5 % du génome est véritablement codant alors que les autres parties du génome humain sont, soit destinées à la régulation de l'expression des gènes, soit inactives ou encore ont des fonctions qui aujourd'hui nous demeurent inconnues. Il y a donc un autre travail à mener, extrêmement complexe, qui est celui du tri de ce code, ce véritable décryptage qui permettra de déterminer les régions codantes et les régions régulatrices du génome humain. Et c'est là d'ailleurs une nouvelle science qui va jouer un rôle fondamental, que l'on appelle la bio-informatique (encore un néologisme). L'objectif de la bio-informatique est de réaliser de façon automatique ce travail de tri entre les régions codantes, régulatrices, et les autres régions du génome humain. Ce travail qui, lui, s'écoulera sans doute sur une durée beaucoup plus longue nous permettra d'identifier véritablement la structure du génome, c'est-à-dire de connaître les quelque 100 000 gènes qui probablement le constituent. Ce travail d'analyse de la structure est beaucoup moins avancé puisqu'on estime qu'aujourd'hui on connaît moins de 10 % de ces 100 000 gènes. Beaucoup de ces connaissances sont dans le domaine public, mais il existe aussi des gènes qui ont été identifiés et qui sont encore aujourd'hui dans des bases de données privées, ou qui font l'objet aujourd'hui de dépôts de brevets, donc qui sont des données encore confidentielles, puisque toutes ces données n'ont pas encore été publiées. .

Troisième étape, qui probablement sera la plus importante: celle des applications de ces gènes bien évidemment, en particulier applications à la santé humaine qui font l'objet aujourd'hui d'une intense compétition internationale. Ces applications sont extrêmement importantes. Pourquoi ? Parce que la plupart des maladies ont une composante génétique. C'est un concept qui est bien connu pour les maladies génétiques pures, comme la mucoviscidose ou les myopathies, mais ces maladies génétiques pures représentent un cas extrême où il y a une relation biunivoque entre une erreur du Code génétique, une mutation, et la maladie. La mutation provoque la maladie dans tous les cas. Pour ce qui concerne les maladies communes, le phénomène est beaucoup plus complexe. Plusieurs gènes concourent au développement de la maladie, et il y a aussi, bien entendu,

l'influence de nombreux facteurs environnementaux qui jouent un rôle clé. À ce propos, il est sans doute intéressant et important de rappeler que le débat sur le rôle respectif de l'hérédité et de l'environnement est très probablement un faux débat. On pourrait le comparer à celui de savoir si c'est la longueur ou la largeur du rectangle qui est la plus importante pour la surface. Il s'agit en fait du produit des deux. Il n'en reste pas moins que, pour ces maladies, nous sommes aujourd'hui en mesure d'identifier des gènes défectueux qui sont à l'origine du développement des maladies. L'identification de ces gènes, dits de susceptibilité, est aujourd'hui bien avancée pour un certain nombre de maladies. C'est déjà par exemple le cas pour le cancer du sein, le cancer de la prostate, la maladie d'Alzheimer, pour lesquels on connaît un ou plusieurs gènes de susceptibilité. En revanche, on est dans l'incapacité aujourd'hui d'en identifier les facteurs environnementaux. La première clé dont on va donc disposer par rapport à ces causes fondamentales sera certainement un facteur génétique. Il y a bien entendu quelques exceptions, on peut citer par exemple celui du mélanome en Australie. On sait que la fréquence du cancer de la peau en Australie est beaucoup plus élevée que dans les autres pays du monde. Cela est dû au fait qu'il existe un gène de susceptibilité dans la population anglaise qui a ensuite été transporté en Australie, ce qui, à cause de l'exposition particulière aux rayonnements ultraviolets du soleil, donne un taux très élevé de cancers de la peau sur une population particulièrement sensible. Dans certaines situations de ce genre, exceptionnelles, on est capable d'identifier les facteurs environnementaux et les facteurs génétiques. Mais, dans la plupart des cas, les facteurs environnementaux restent absolument inconnus.

Cette approche, qui consiste à identifier les gènes de susceptibilité, a fait l'objet de travaux originaux en France, et c'est là que se situe le positionnement unique des équipes de recherche françaises. C'est, en particulier, depuis la création du Centre d'étude du polymorphisme humain, créé par le Pr. Jean Dausset, que les premiers travaux ont été menés pour identifier les variations entre le Code génétique des individus, qu'on appelle les polymorphismes, et leurs implications médicales. Les premières cartes du génome humain qui ont identifié ces variations ont été produites en France, comme je le rappelais, et, en fait, notre société est aujourd'hui activement impliquée dans l'identification systématique de ces polymorphismes et leur rôle sur le développement de plusieurs maladies.

Ces études se conduisent par analyse statistique. Nous sommes aujourd'hui en mesure, dans les laboratoires qui travaillent à grande échelle, de comparer des milliers de marqueurs entre malades et contrôles. Ces études statistiques impliquent la comparaison de centaines de malades et de centaines de contrôles. Pour certains marqueurs, nous allons observer des variations qui sont associées à certaines de ces maladies, et nous savons que les gènes de susceptibilité se situent au voisinage de ces marqueurs variables. La première étape est donc celle de la localisation du gène le long du génome humain. Il faut ensuite passer à une deuxième étape qui consiste à identifier le véritable gène et sa mutation causale, ce qui constitue un long travail. Ces travaux, qui jusqu'à présent nécessitaient jusqu'à une dizaine d'années, peuvent être aujourd'hui menés sur des périodes beaucoup plus courtes de quelques années. Aujourd'hui, notre société conduit des travaux sur plusieurs maladies, en particulier le cancer de la prostate, l'obésité mais aussi, ce qui est plus surprenant pour les non-initiés, des maladies psychiatriques, comme par exemple la schizophrénie. Nous avons identifié plusieurs gènes de susceptibilité, en particulier dans le cas du cancer de la prostate et de l'obésité.

L'autre caractéristique de notre activité est l'industrialisation de ces procédés. Les sociétés de génomique, dans le deuxième sens du mot, c'est-à-dire l'industrie de la génomique, mènent ces études à grande échelle dans des laboratoires qui sont automatisés comme notre laboratoire d'Évry. Aujourd'hui cette approche est complètement intégrée. Ces études à grande échelle peuvent être menées depuis les banques d'ADN, que nous

constituons en provenance de patients ou de contrôles normaux, jusqu'aux ordinateurs qui, finalement, analysent l'information à grande échelle. Ces équipes sont aujourd'hui des équipes très importantes, puisque, par exemple, sur notre site du génopole d'Évry, plus de 350 personnes se consacrent à cette activité, dont 60 sur la seule bio-informatique ; elles travaillent uniquement avec des ordinateurs. Au fur et à mesure des progrès de la génomique, nous déplaçons progressivement du monde de la biologie vers le monde de l'informatique.

La grande révolution est essentiellement la réduction du temps nécessaire à l'identification de ces gènes, environ deux ans aujourd'hui, contre dix ans auparavant. La conséquence en est que la concurrence internationale est particulièrement vive, en particulier pour les gènes majeurs qui sont à l'origine des grandes maladies, dont, je vous rappelle, moins de 10 % sont connus. On peut estimer, et c'est une information importante, qu'environ 300 gènes majeurs de susceptibilité sont responsables d'environ 90 % des maladies courantes. Il y a donc une course importante pour l'identification de ces 300 gènes majeurs qui sont des clés fondamentales par rapport aux applications, dans l'industrie pharmaceutique en particulier. Notre industrie est donc en fait essentiellement une industrie d'intermédiation. L'objectif est de découvrir des gènes d'intérêt médical, de déposer des brevets sur leurs applications et de les licencier à l'industrie pharmaceutique. Aujourd'hui, par exemple, Genset possède sept accords avec de grandes sociétés pharmaceutiques, qui recouvrent pratiquement tous les grands domaines d'application thérapeutique, depuis le vaccin jusqu'au médicament.

Les conséquences de la génomique sont sans doute considérables pour l'industrie pharmaceutique. Aujourd'hui, le coût moyen de développement d'un médicament est d'environ 4 milliards de francs. Ce coût considérable a deux causes: la première est que la recherche dans l'industrie pharmaceutique est encore essentiellement une recherche hasardeuse, où un très grand nombre de composés sont testés, jusqu'à ce qu'on obtienne un composé actif qui ensuite va pouvoir être développé au cours d'essais cliniques. Les essais cliniques sont longs et coûteux, avec des probabilités de succès relativement faibles : moins de 5 % des composés qui rentrent en essais cliniques pourront être finalement commercialisés. Les gènes de susceptibilité vont permettre une modification radicale de ces procédés et ces développements. Les cibles spécifiques permettront, d'une part, de s'attaquer à la cause plutôt qu'aux effets des maladies. D'autre part, ces cibles permettront aux sociétés de développer des médicaments avec de meilleures chances de succès, plus efficaces et présentant moins d'effets secondaires. Cela va donc s'accompagner d'une réduction des coûts et de meilleures chances de succès. Les premiers gènes que nous avons identifiés dans le cadre de notre programme avec le laboratoire pharmaceutique Sanofi-Synthélabo sur le cancer de la prostate sont aujourd'hui utilisés par notre partenaire qui recherche des molécules à partir de ces découvertes.

Mais, finalement, la révolution sera certainement le développement de traitements sur mesure, ce que l'on appelle la pharmaco-génomique. Le développement de traitements sur mesure est un enjeu important pour les sociétés pharmaceutiques puisque, en connaissant l'origine véritable des maladies, on pourra finalement proposer un traitement adapté. Une véritable définition génétique des maladies avec des traitements appropriés va devenir possible. Il est bien évident qu'il ne s'agit là que d'une continuation dans la définition des maladies. Si l'on remonte au XIX<sup>e</sup> siècle, on parlait alors de pneumonie. On désigne aujourd'hui des affections beaucoup plus précises comme la tuberculose, l'asthme. Notre connaissance et notre définition de ces maladies à partir de facteurs mieux identifiés sont plus précises. De la même façon, on se dirige vers une connaissance plus précise des maladies, une véritable classification des maladies, à partir de facteurs génétiques qui aujourd'hui nous restent encore inconnus. Un exemple des applications de la génomique est celui des maladies psychiatriques. Il s'agit d'un domaine où beaucoup de progrès

s'accomplissent, s'agissant de la neurobiologie; on découvre pratiquement toutes les semaines de nouveaux récepteurs, de nouveaux médiateurs, mais on est dans l'impossibilité d'établir une relation entre ces mécanismes biologiques et les pathologies concernées. La génomique constitue, ici, pratiquement, un des seuls espoirs pour identifier les mécanismes qui sont la cause de ces maladies psychiatriques, et la génomique permettra incontestablement une nouvelle définition de ces maladies avec une connaissance précise de leur cause génétique.

Peut-être, en matière de conclusion, pouvons-nous nous livrer à un petit exercice à la mode en cette fin de millénaire, qui est celui de la prévision. Et pouvons-nous nous poser la question de savoir quels seront les traitements du futur issus de cette industrie de la génomique et de cette recherche en génomique. Il est très probable que ces traitements du futur seront menés en plusieurs étapes. Grâce à la connaissance de ces gènes de susceptibilité, nous aurons certainement la possibilité d'identifier très tôt, voire même à la naissance, un certain nombre de risques majeurs pour chaque individu et de suivre ces risques tout au long de la vie. L'identification de ces risques nous permettra, grâce à un certain nombre de marqueurs biologiques, de suivre la progression asymptomatique de certaines maladies avec des diagnostics appropriés. Il est très probable et même certain que nous pourrons mettre en place et proposer des traitements préventifs, avant que les symptômes n'apparaissent, à partir de ces marqueurs biologiques. La médecine préventive deviendra donc très certainement une réalité. Et, une fois les symptômes apparus, il est clair que, grâce à ces informations, nous pourrons proposer des traitements sur mesure, adaptés à chaque individu, qui seront donc des traitements sur mesure postsymptomatiques. L'apparition de ces traitements préventifs et de ces traitements sur mesure est déjà signalée aujourd'hui dans le domaine cardio-vasculaire car, par exemple, l'hypertension ou le taux de cholestérol sont aujourd'hui des marqueurs biologiques qui nous permettent d'identifier des risques et de mettre en place un certain nombre de traitements préventifs spécifiques. Il s'agit donc simplement d'une évolution naturelle à partir d'une information qui existe déjà dans le domaine biologique et qui va passer au stade de l'information génétique. Il y a aussi des bouleversements considérables médicaux autour de ces technologies et de leurs applications, avec des conséquences bien entendu éthiques, sociologiques et économiques, qui sont aujourd'hui très difficiles à mesurer. Mais on peut affirmer avec une quasi-certitude que ces applications se mettront en place parce que cette information sera relativement facile à obtenir. Aujourd'hui, nous savons qu'avec des technologies relativement faciles à mettre en œuvre, nous pourrons aisément disposer de cette information sur les risques majeurs et sur la progression des maladies dans les années qui viennent.