

Vers une médecine personnalisée grâce à la recherche en génomique

Pierre Neuvial (2003), chercheur au laboratoire Statistique et Génome (CNRS, Université d'Evry)

Les techniques modernes d'expérimentation biologique permettent de sonder en détail la complexité du vivant. L'analyse statistique des données génomiques ainsi produites requiert des développements méthodologiques pointus. Ceux-ci doivent permettre d'améliorer la prise en charge thérapeutique des patients grâce à des traitements de plus en plus personnalisés.

Le traitement et la prévention des maladies humaines constituent des enjeux de santé publique fondamentaux qui nécessitent des progrès à la fois en recherche clinique (amélioration des diagnostics et des traitements) et en recherche fondamentale (compréhension des mécanismes biologiques des maladies). Ces progrès sont notamment rendus possibles aujourd'hui par le développement de nouvelles technologies d'acquisition de données moléculaires dites génomiques, qui permettent de mesurer rapidement et à grande échelle différents niveaux d'information génétique et épigénétique dans les cellules.

Cet article traite, au travers d'exemples en cancérologie, du rôle de la recherche en génomique dans le développement d'une médecine dite *personnalisée*, c'est-à-dire dont les traitements sont adaptés aux spécificités de chaque patient et de sa maladie.¹

La médecine personnalisée, un enjeu de santé publique majeur

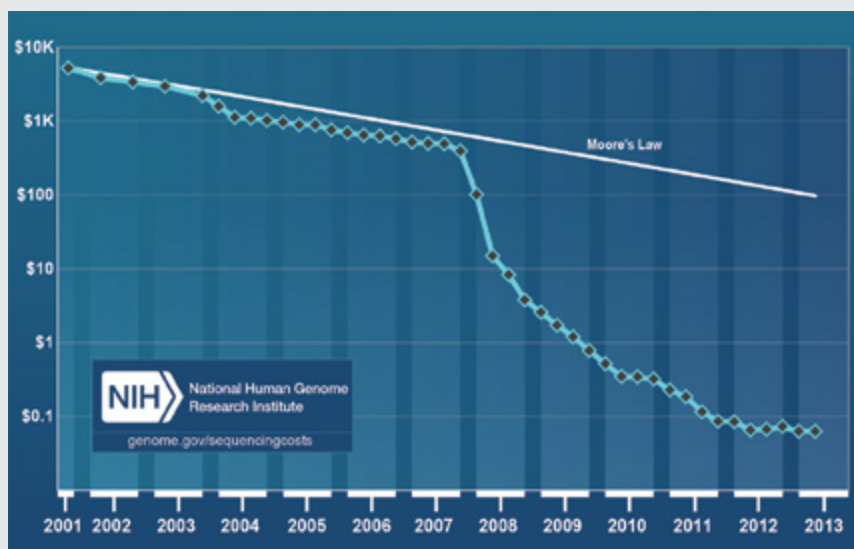
Le cancer du sein est la principale cause de décès par cancer chez les femmes dans le monde. La plupart des cas sont traités par chirurgie, suivie d'un traitement (dit adjuvant) par radiothérapie ou chimiothérapie qui a pour objectif de limiter le risque de rechute. La chimiothérapie adjuvante permet de diminuer considérablement (jusqu'à 50 %) le risque de rechute des cancers du sein infiltrants. Cependant, ce traitement est inefficace pour plus de la moitié des patientes, et il occasionne chez la plupart des effets secon-

naires importants (nausées, vomissements, chute de cheveux), dus au fait que les chimiothérapies ciblent et détruisent les cellules qui se divisent rapidement, ce qui est notamment le cas des cellules cancéreuses, mais aussi d'autres cellules du corps humain (comme celles des cheveux et de la flore intestinale). Cet exemple illustre un enjeu majeur de santé publique, non seulement en cancérologie mais pour toutes les maladies humaines que l'on ne sait pas encore parfaitement soigner : le développement d'une *médecine personnalisée*, c'est-à-dire d'une prise en charge médicale adaptée à la forme particulière de la maladie développée par chaque patient. Les bénéfices attendus de ce type de prise en charge sont double : une amélioration de l'efficacité des traitements, couplée à une amélioration de la qualité de vie des patients, par la diminution des effets secondaires.

Des exemples de telles thérapies ciblées existent déjà aujourd'hui, grâce aux progrès considérables réalisés durant les vingt dernières années dans la compréhension des mécanismes biologiques de formation des cancers. Ainsi, il a été établi en 1985 que certains cancers du sein sont caractérisés par une altération du gène HER2 dans les cellules tumorales. Un médicament permettant de réparer cette altération (notée HER2+), le Trastuzumab (Herceptin,®) est utilisé depuis le début des années 2000 pour traiter spécifiquement les patientes dont le cancer est HER2+, en remplacement ou en combinaison avec la chimiothérapie adjuvante classique.

Des altérations spécifiques ou des combinaisons d'altérations peuvent donc fortement influencer

¹ - Pour un traitement plus complet des enjeux économiques, éthiques et médicaux sous-tendus par la notion de médecine personnalisée, nous renvoyons à l'ouvrage de Paci et al (2013) : « Médecine personnalisée et cancer : Organiser et faciliter l'accès à l'innovation ».

Figure 1 : COÛT PAR MÉGABASE BRUTE DE SÉQUENCE ADN

la réponse aux thérapies anticancéreuses. Cependant, de telles thérapies ciblées sont encore relativement rares du fait de la complexité et de l'hétérogénéité des cancers, et car leur développement requiert plusieurs étapes difficiles avant même les essais cliniques (Garay & Gray, 2012) :

- identification d'une cible ou « biomarqueur » dont l'altération est associée à la progression tumorale dans un sous-groupe de patients,
- mise au point d'un traitement permettant de rétablir le fonctionnement normal de cette cible, ainsi que d'un test diagnostique dit « compagnon » permettant de prédire chez un patient donné l'efficacité du traitement.

Dans le cas du Trastuzumab, le médicament a été mis sur le marché en 2001, soit 16 ans après la découverte de l'altération de HER2. Pour de nombreuses maladies (ou sous-types de maladies), il n'existe toujours pas de biomarqueur connu.

Les données génomiques : un outil précieux mais complexe

Pour sonder cette complexité, les chercheurs disposent aujourd'hui de techniques d'expérimentation biologique qui permettent de mesurer simultanément des millions de paramètres biologiques. Le terme « génomique » désigne une discipline de la biologie qui étudie la structure et

le fonctionnement de l'ensemble du patrimoine génétique d'un individu (son génome), par opposition à la génétique classique, qui s'intéresse à un gène ou un groupe de gènes particulier. Cette discipline a émergé avec le projet de séquençage du génome humain (1993-2003). La rapidité des progrès technologiques est telle qu'aujourd'hui, c'est-à-dire dix ans après, il est possible de séquencer un génome entier en quelques heures et pour quelques milliers d'euros (voir la figure 1). Les technologies actuelles de la biologie moléculaire (protéomique, puces à ADN, imagerie, séquençage) permettent de caractériser à grande échelle différents niveaux d'information dans les cellules. De grands consortiums comme le Cancer Genome Atlas (TCGA) américain ou l'International Cancer Genome Consortium (ICGC) produisent et analysent de telles données génomiques à partir de dizaines de milliers d'échantillons tumoraux, provenant de plusieurs dizaines de types tumoraux différents. Ces consortiums produisent ainsi des téraoctets de données génomiques chaque jour.

L'exploitation de ces données requiert non seulement le développement d'infrastructures de transmission, de stockage, et d'accès à ces données, mais aussi le développement de méthodes statistiques d'analyse de ces données. Outre leur volume exceptionnel, une caractéristique essentielle des données génomiques est que le nombre de variables considérées (gènes ou *loci* le long du génome par exemple) peut varier entre quelques dizaines de milliers et des centaines de millions, et dépasse donc généralement de plusieurs ordres de grandeurs le nombre d'observations, qui n'excède généralement pas quelques centaines ; on parle de données *de grande dimension*. Par ailleurs, ces données sont également à la fois extrêmement *complexes*, car elles sont obtenues grâce à des expériences biologiques pointues, hétérogènes du fait de l'existence de différents niveaux d'information biologique (gènes, transcrits, protéines), et fortement *structurées* par des relations impliquant ces différentes entités biologiques (voies de signalisation, réseaux de régulation, réseaux métaboliques).

Défis statistiques de l'analyse de données génomiques

Dans ce contexte, un défi majeur pour les statisticiens est le développement d'outils qui reposent sur des modèles rigoureux sur le plan mathématique et décrivant correctement la réalité biologique, qui soient assortis d'algorithmes suffisamment efficaces pour traiter de gros volumes

de données, et enfin qui fournissent des résultats interprétables sur le plan biologique et clinique.

L'identification de nouveaux biomarqueurs (cibles thérapeutiques potentielles) nécessite de mettre en relation les variations moléculaires observées chez les patients avec les marqueurs diagnostiques dont on dispose, que sont généralement l'apparition ou la sévérité de la maladie. Dans les analyses les plus simples, les biomarqueurs peuvent être la mutation, la sur-expression, ou la méthylation d'un ou plusieurs gènes. Leur identification repose sur l'utilisation d'outils statistiques spécifiques aux données génomiques, qui suscitent de nombreux développements à la fois sur le plan de la modélisation et de l'analyse statistique, ainsi que sur le plan algorithmique. Les questions statistiques sous-jacentes sont l'objet d'une intense recherche : tests multiples ; régression, classification et sélection de modèles en grande dimension. Des analyses plus complexes visent à mettre en évidence des ensembles de gènes connus pour interagir en un « réseau de régulation », et qui sont altérés dans tel sous-type de cancer. Le rétablissement du fonctionnement normal de ce réseau peut alors permettre d'éliminer les cellules cancéreuses. Les approches statistiques utilisées combinent les données génomiques avec un *a priori* biologique modélisant l'état des connaissances sur le réseau de régulation.

Dans de nombreux cas, les approches visant à identifier directement un petit nombre de biomarqueurs fortement associés au développement d'une maladie sont rendues très difficiles par la grande dimension et la complexité des données génomiques à disposition. En particulier, il est rarement possible de mettre en évidence des relations causales à partir de ces observations. Une approche complémentaire appelée pharmacogénomique consiste à modéliser (*in vitro*) la réponse à des composés chimiques connus afin d'identifier des composés susceptibles d'agir sur les patients atteints par la maladie (ou plus fréquemment le sous-type de maladie) analysée et non sur les autres patients.

Perspectives médicales de la recherche en génomique

L'expression « médecine personnalisée » peut sembler redondante, car le propre de la médecine est d'identifier et de mettre en place un traitement adapté au couple patient / maladie. L'originalité des approches reposant sur la recherche en génomique est que l'examen clinique du patient peut être complété par la mesure d'un grand nombre de

paramètres biologiques quantitatifs, permettant de sonder plus finement les propriétés du patient, de sa maladie, ainsi que leurs interactions. Alors que les thérapies ciblées sont déjà une réalité clinique, notamment en cancérologie ainsi que pour le traitement du sida, l'extrême complexité du vivant révélée par les données génomiques pose de nouveaux défis aux biologistes et aux cliniciens ; les questions statistiques sous-jacentes stimulent des développements mathématiques pointus qui ont fait émerger une communauté scientifique dédiée très active.

Bien que le développement de nouveaux médicaments soit un processus relativement long (de l'ordre d'une dizaine d'années entre l'identification d'une cible thérapeutique et la mise sur le marché d'un médicament associé), les approches récentes reposant sur les données génomiques portent déjà leurs premiers fruits. Ainsi, de nouveaux mécanismes biologiques ont été identifiés qui expliquent la résistance de certains cancers du sein de type HER2+ au Trastuzumab, et qui ont permis de proposer un traitement plus efficace pour ces cancers.

L'impact économique du recours à de telles thérapies ciblées est important car les traitements anticancéreux sont onéreux, et car le gain d'efficacité des traitements ciblés ne concerne par définition qu'une fraction des malades. Cependant, l'analyse économique doit également prendre en compte le fait que les tests moléculaires développés pour guider l'application de ces thérapies peuvent éviter le recours à une chimiothérapie inefficace. Ceci permet non seulement de supprimer des dépenses inutiles mais également d'améliorer très sensiblement la qualité de vie de patients n'ayant pas à subir les effets secondaires d'un traitement lourd, si l'on est capable par ailleurs de prédire son inefficacité. ■

Références :

- Garay, Joseph P., and Joe W. Gray. «Omics and therapy—A basis for precision medicine» *Molecular Oncology* 6.2 (2012): 128-139.
- Paci, Angelo, et al. « Médecine personnalisée et cancer : Organiser et faciliter l'accès à l'innovation ».
- Les presses de l'Institut Gustave Roussy, 2013.