

La carte génétique de l'homme enfin sur orbite... de microsatellites

En dépit du nombre impressionnant de cartes génétiques de chromosomes humains proposées ces dernières années, la carte génétique de l'homme souffrait encore récemment de plusieurs défauts majeurs : (1) les cartes de liaison étaient surtout fondées sur la localisation de RFLP (*restriction fragment length polymorphisms*) qui s'avèrent d'une utilité limitée dans la cartographie des maladies génétiques. Ces RFLP, pour la plupart bi-alléliques, ont permis de cartographier les affections génétiques les plus fréquentes, mais sont insuffisants dans le cas de maladies plus rares ou hétérogènes sur le plan génétique ; (2) certaines régions chromosomiques demeurent encore mal couvertes et il faut souvent plusieurs années pour définir un intervalle génétique minimal fondé sur la définition des points de recombinaison à partir d'une localisation primaire ; (3) la mise en évidence des *loci* à effet mineur dans les maladies multifactorielles, repose sur une carte génétique à haute densité. Il y a donc à la fois nécessité de développer un grand nombre de marqueurs génétiques et de disposer de marqueurs hautement informatifs.

En raison de la faible informativité des marqueurs bi-alléliques on s'oriente de plus en plus systématiquement vers la recherche de marqueurs multi-alléliques. De courtes répétitions en tandem (10 à 30 fois) de séquences nucléotidiques très simples (di-, tri-, tétranucléotides, etc.) sont disséminées en plusieurs dizaines de milliers d'emplacements distincts du génome des mammifères. Ces motifs, encore appelés microsatellites ou STR (*simple* ou *short tandem repeats*), sont fréquemment multi-alléliques : les variations correspondent à des changements du nombre des répétitions à l'intérieur d'un motif (*m/s n° 6, vol. 6, p. 576*). Ces changements de taille de quelques paires de bases sont mis en évidence après amplification par PCR du *locus* considéré et séparation des fragments

sur gel d'acrylamide dénaturant. Cette méthode de détection de *loci* polymorphes [1, 2] est cependant lourde : elle nécessite l'acquisition de données de séquence des régions flanquant les motifs polymorphes et requiert l'utilisation de systèmes analytiques contraignants (marquages radioactifs, gels d'acrylamide dénaturants). Malgré ces inconvénients, l'utilisation généralisée de ce type de marqueurs en cartographie génétique est en train de modifier radicalement l'état des cartes génétiques chez les mammifères.

Ainsi chez l'homme, deux cartes de liaison génétique de l'ensemble du génome viennent d'être publiées à moins d'un mois d'intervalle. Une première série de travaux patronnée par le NIH et le CEPH a permis d'intégrer dans des cartes surtout constituées d'anciens RFLP bi-alléliques, 340 marqueurs microsatellites dont la moitié dépasse le seuil d'hétérozygotie de 70 % [3]. Une nouvelle carte génétique de deuxième génération entièrement constituée de marqueurs microsatellites dont les trois quarts ont une hétérozygotie supérieure à 70 % vient d'être proposée par le laboratoire Généthon [4]. Les distances génétiques couvertes par cette dernière sont notablement plus faibles que celles publiées par le NIH/CEPH, mais plus proches des valeurs attendues théoriquement. Ces différences sont principalement dues aux erreurs de génotypage assez courantes et difficilement contrôlables lorsque les cartes sont établies à partir de données disparates comme c'est le cas de la carte NIH/CEPH. Les erreurs ont pour effet d'augmenter artificiellement le nombre de recombinaisons et donc les distances génétiques.

La très grande majorité des cartes génétiques est établie à partir des génotypes des ADN de familles de référence rassemblées par le CEPH (*m/s n° 3, vol. 6, p. 286*). L'utilisation de ces ADN de référence permet

d'étudier les mêmes méioses dans des laboratoires différents (une centaine actuellement) et de préciser les quelques 50 à 60 événements de recombinaison d'un génome diploïde à partir des résultats provenant de l'ensemble de la communauté des cartographes et rassemblés dans une base de données commune maintenue par le CEPH, les deux cartes mentionnées ici ont été établies à partir des génotypes des familles du CEPH.

Le niveau de résolution de ces cartes dépend du nombre de méioses informatives analysées. En raison de l'important déchet de génotypages non informatifs des marqueurs bi-alléliques, les collaborateurs du CEPH s'étaient engagés à analyser un total de 40 familles. Lorsque le taux d'hétérozygotes observés augmente, on peut obtenir une résolution de même niveau en établissant le génotype de moins de familles. Le génotypage de huit familles du CEPH a permis à l'équipe de Généthon d'ordonner 550 marqueurs avec un rapport de vraisemblance supérieur à 1 000, dépassant souvent celui obtenu avec 40 familles pour des marqueurs bi-alléliques. En outre, l'analyse d'un nombre plus important de familles ne permet pas des gains de résolution importants, notamment pour l'ordonnement de marqueurs séparés par de très faibles distances de recombinaison. Il sera sans doute plus aisé d'ordonner ces marqueurs sur une carte physique (*m/s n° 8, vol. 8, p. 881*).

Le projet de cartographie génétique de Généthon repose sur quelques options destinées à mener de manière intensive un ensemble de manipulations propres aux laboratoires de recherche : clonage, criblage de banques, séquençage, analyse informatique de séquences, amplification par PCR, génotypage, analyse et traitement des données de génotypage. Parmi les nombreux motifs microsatellites existants, le motif CA est de loin le plus fréquent. C'est aussi celui dont les caractéristiques sont

les mieux connues. Il n'existe pas à ce jour d'étude systématique permettant de connaître avec précision le degré de polymorphisme des autres motifs microsatellites. Pour ces diverses raisons, seul le motif CA a été retenu pour l'établissement de cette carte.

Afin de réduire le nombre de gels de séquence, une technique en « multiplex » dérivée de la méthode de séquençage de Church et Kieffer-Higgins [5] a été utilisée pour le déterminisme du génotype. Les produits de PCR de huit à 16 marqueurs différents d'un même individu ont été mélangés, mis à co-migrer sur une seule piste de gel, puis transférés sur membrane. Les produits d'amplification ont ensuite été détectés par hybridation à des sondes non radioactives correspondant à une des amorces allongée puis couplée à la peroxydase. Cependant, pour pouvoir démarrer ce projet le plus rapidement possible, aucune automatisation *ad hoc* n'a été développée et les manipulations sont pour l'essentiel menées manuellement avec l'aide d'automates du commerce. En raison du volume des données à traiter il a néanmoins été indispensable d'automatiser l'enchaînement des opérations de calcul de *linkage* et de construction des cartes.

L'établissement de cette carte génétique n'est que dans sa première phase. Une carte au niveau de résolution obtenu est déjà d'une utilité importante dans la cartographie primaire de maladies génétiques à transmission mendélienne, puisqu'elle met à la disposition des généticiens, pour des études de *linkage* ; un ensemble de marqueurs régulièrement espacés et permettant de couvrir la quasi-totalité du génome. Il s'est donc avéré indispensable de publier cette première carte partielle. Celle-ci, d'une résolution moyenne de 6 cM présente encore une douzaine de régions de plus de 20 cM dépourvues de microsatellites. Certains des intervalles non couverts ont aussi été observés sur les autres cartes. Il n'est donc pas exclu que ces régions soient associées à une « recombino-génécité » particulièrement élevée ayant pour effet d'amplifier les distances génétiques par rapport aux distances physiques. Cela a déjà été observé pour des régions subtélomériques et, de fait, certaines font partie des intervalles mal couverts.

L'établissement d'une carte à 2 000 marqueurs est en cours. La plupart des grands intervalles ont déjà été réduits par l'analyse génotypique de nouveaux marqueurs obtenus au hasard. Mais il est possible que l'obtention d'une carte homogène à haute densité nécessite de faire appel à des approches ciblées ou à l'application de stratégies particulières permettant l'isolement préférentiel de marqueurs de régions mal couverts. Certaines de ces stratégies sont en cours d'évaluation.

A côté d'une carte génétique plus dense, il devient impératif d'établir un lien entre cette carte génétique et les segments contigus de clones de YAC chevauchants obtenus par la méthode d'analyse des *fingerprints* récemment décrite [6] (*m/s n° 8, vol. 8, p. 881*). Cela constituera la première version d'une carte intégrée qui devrait avoir un impact considérable dans la cartographie des maladies génétiques et servir de point de départ à une collection de segments ordonnés et chevauchants d'ADN cloné représentant l'ensemble du génome humain ■

RÉFÉRENCES

1. Weber JL, May PE. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 1989 ; 44 : 388-96.
2. Litt M, Luty JA. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet* 1989 ; 11 : 397-401.
3. NIH/CEPH Collaborative Mapping Group. A comprehensive genetic linkage map of the human genome. *Science* 1992 ; 258 : 67-86.
4. Weissenbach J, Gyapay G, Dib G, *et al.* A second-generation linkage map of the human genome. *Nature* 1992 ; 359 : 794-801.
5. Church GM, Kieffer-Higgins S. Multiplex DNA sequencing. *Science* 1988 ; 240 : 185-8.
6. Bellanné-Chantelot C, Lacroix B, Ougen P, *et al.* Mapping the whole genome by fingerprinting yeast artificial chromosomes. *Cell* 1992 ; 70 : 1059-68.

Jean Weissenbach

Cnrs URA 1445, Institut Pasteur, 25, rue du Docteur Roux, 75724 Paris Cedex 15, France.

TIRÉS A PART

J. Weissenbach.