

## Marqueurs génétiques et maladies multifactorielles. Quelle stratégie adopter ?

Françoise  
Clerget-Darpoux  
Catherine Bonaïti-Pellié  
Josué Feingold

Les marqueurs génétiques n'ont pas seulement un intérêt dans les maladies monogéniques, ils peuvent également apporter de l'information sur le déterminisme des maladies multifactorielles. Pour ces dernières — qui ne sont pas dues à un défaut génétique rare mais à la combinaison défavorable de multiples facteurs génétiques et/ou liés à l'environnement — le but est de détecter ces facteurs, d'évaluer leur effet et leur interaction. Les gènes dont la fonction a un rapport avec la physiologie de la maladie sont de bons candidats comme facteurs potentiels de risque. De ce fait, il est bien plus pertinent de concentrer les efforts de recherche sur des marqueurs de ces gènes dits « gènes candidats », que sur des marqueurs pris aléatoirement sur le génome. Par ailleurs, il ne faut pas se limiter à l'information apportée par la liaison génétique, mais aussi utiliser celle fournie par les associations préférentielles entre les allèles du gène candidat et les allèles du marqueur.

### ADRESSE

F. Clerget-Darpoux : directeur de recherche à l'Inserm. C. Bonaïti-Pellié : directeur de recherche à l'Inserm. J. Feingold : directeur de recherche à l'Inserm et directeur de l'Inserm U. 155, château de Longchamp, bois de Boulogne 75016 Paris, France.

### TIRÉS A PART

J. Feingold

m/s n° 10 vol. 8, décembre 92

**G**âce aux nouveaux marqueurs génétiques, l'épidémiologie génétique offre aujourd'hui de nouvelles perspectives dans l'étude des maladies humaines. Ces marqueurs, issus des techniques de biologie moléculaire, apportent en effet de l'information sur la composante génétique des maladies. La nature de cette information sera différente selon que l'on étudie une maladie monogénique ou une maladie multifactorielle.

Pour les maladies monogéniques, en effet, l'étude de la ségrégation des marqueurs génétiques dans des familles de malades permet de localiser le gène impliqué. Il est possible d'entreprendre une étude systématique de tout le génome, étude dont on exclut progressivement des régions pour circonscrire finalement celle où se situe le gène incriminé. Cette recherche systématique sur le génome est parfaitement adaptée aux problèmes des maladies mendéliennes pour lesquelles, comme le souligne Edwards [1],

## RÉFÉRENCES

1. Edwards JH. Exclusion mapping. *J Med Genet* 1987 ; 24 : 539-43.
2. Morton NE. The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type. *Am J Hum Genet* 1956 ; 8 : 80-96.
3. Clerget-Darpoux F, Bonaïti-Pellié C, Hochez J. Effects of misspecifying genetic parameters in lod score analysis. *Biometrics* 1986 ; 42 : 393-9.
4. Clerget-Darpoux F, Bonaïti-Pellié C. Strategies based on marker information for the study of human diseases. *Ann Hum Genet* 1992 ; 56 : 147-55.
5. Morton NE. Sequential tests for the detection of linkage. *Am J Hum Genet* 1955 ; 7 : 277-318.
6. Newman B, Austin MA, Lee M, King MC. Inheritance of human breast cancer : evidence for autosomal dominant transmission in high risk families. *Proc Natl Acad Sci USA* 1988 ; 85 : 3044-8.
7. Claus EB, Risch N, Thompson WD. Genetic analysis of breast cancer in the cancer and steroid hormone study. *Am J Hum Genet* 1991 ; 48 : 232-42.
8. Iselius L, Slack J, Littler M, Morton NE. Genetic epidemiology of breast cancer in Britain. *Ann Hum Genet* 1991 ; 55 : 151-9.
9. Margaritte P, Bonaïti-Pellié C, King MC, Clerget-Darpoux F. Linkage of familial breast cancer to chromosome 17q21 may not be restricted to early-onset disease. *Am J Hum Genet* 1992 ; 50 : 1231-4.
10. McGuffin P, Huckle P. Simulation of Mendelism revisited : the recessive gene for attending medical school. *Am J Hum Genet* 1990 ; 46 : 994-9.
11. Weitkamp LR, Stancer HC, Persad E, Flood C, Guttormsen S. Depressive disorders and HLA : a gene on chromosome 6 that can affect behavior. *N Engl J Med* 1981 ; 305 : 1301.
12. Egeland JA, Gerhard DS, Pauls DL, et al. Bipolar affective disorders linked to DNA markers on chromosome 11. *Nature* 1987 ; 325 : 783-7.
13. Sherrington R, Brynjolfsson J, Pettersson H, et al. Localization of a susceptibility locus for schizophrenia on chromosome 5. *Nature* 1988 ; 336 : 164-7.

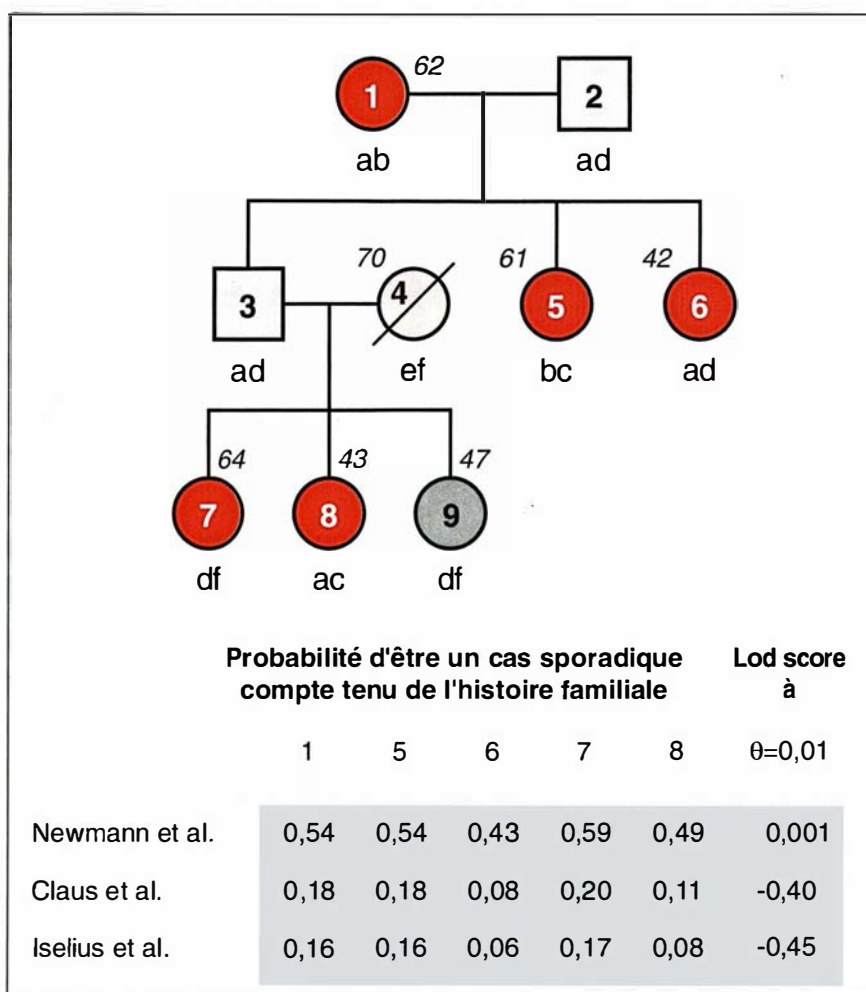


Figure 1. **Variation des valeurs de lod score en fonction de la correspondance phenotype-génotype.** Dans cette famille, comportant 5 femmes atteintes, on s'attend que la plupart des cas soient génétiques. En utilisant les différentes correspondances phenotype-génotype publiées dans la littérature à partir d'analyses de ségrégation [6-8] et suivant le principe que nous avons décrit par ailleurs [9], nous avons, connaissant l'état des différents membres de sa famille, calculé, pour chacune de ces femmes, la probabilité qu'elle soit en fait un cas sporadique. Ces probabilités varient beaucoup, par exemple de 8 % à 49 % pour la femme n° 8. Si l'on utilise de plus l'information apportée par la ségrégation d'un marqueur dans cette famille (allèles notés abcd), la valeur du lod score varie en parallèle. En italiques : âge de début pour les femmes atteintes, âge courant ou de décès pour les femmes non atteintes.

le gène se trouve nécessairement quelque part. Et de fait, elle a été utilisée avec succès pour localiser les gènes de maladies telles que la myopathie de Duchenne de Boulogne, la mucoviscidose, la dystrophie myotonique de Steinert... Notons cependant qu'une éventuelle hétérogénéité peut compliquer l'analyse. Des gènes situés à des endroits différents du génome (différents *loci*) peuvent être responsables de la même symptomatologie,

comme c'est le cas de l'elliptocytose [2].

Ce succès a engendré un grand enthousiasme chez les généticiens et il a pu laisser croire qu'on pourrait reconduire la même stratégie pour l'étude de toutes les maladies à composante familiale. Or, la plupart de ces maladies ont une étiologie complexe et elles impliquent de nombreux facteurs génétiques et environnementaux qui interagissent. Le plus plau-

sible, dans ces conditions, n'est pas qu'il existe un « gène de maladie » au sens d'une mutation rare délétère, mais plutôt des facteurs génétiques (allèles de susceptibilité) qui augmentent le risque de certains individus. Chaque facteur, pris isolément, peut être fréquent dans la population générale. C'est la combinaison de certains d'entre eux et leur interaction avec des facteurs de milieu qui induiront le processus pathogène. Il n'y a pas lieu de cataloguer comme pathologique chacun des facteurs, d'autant que ceux-ci peuvent se révéler protecteurs dans un autre contexte. C'est très probablement le cas des facteurs de la région HLA impliqués dans les maladies auto-immunes. Dans une telle situation, le but sera d'identifier les différents facteurs impliqués, leur interaction et les risques correspondants.

Utiliser alors une stratégie de recherche systématique de liaison génétique avec des marqueurs du génome peut se révéler totalement inefficace [3, 4], et il devient nécessaire, pour étudier ces maladies, de définir des stratégies mieux adaptées.

### **Problèmes posés par une stratégie de recherche systématique dans les maladies multifactorielles**

#### **Faible probabilité de détecter une liaison génétique**

La méthode la plus couramment utilisée pour tester une liaison génétique entre deux *loci* est la méthode des *lod scores* [5]. Elle permet de trancher en faveur de la liaison génétique (si la valeur du *lod score* est supérieure à 3) ou en faveur de l'indépendance génétique (si la valeur de *lod score* est inférieure à -2). Pour les valeurs de *lod score* intermédiaires, l'information est insuffisante pour conclure.

Il est important de noter que le calcul d'un *lod score* entre les *loci* d'un gène de susceptibilité et d'un marqueur génétique suppose la connaissance du modèle sous-jacent associant aux phénotypes les probabilités des génotypes (correspondance phénotype-génotype). La probabilité de détecter une liaison génétique dépend hautement des caractéristiques de ce modèle, en particulier de la fréquence

de l'allèle de susceptibilité, à la fois chez les sujets atteints et dans la population générale. Moins l'allèle de susceptibilité sera discriminant entre malades et non-malades — autrement dit, plus souvent les malades et non-malades auront le même génotype — et plus la probabilité de détecter la liaison sera faible. Cette ambiguïté dans la correspondance phénotype-génotype est constitutivement importante dans les maladies multifactorielles pour lesquelles la correspondance est même généralement inconnue. Dans ce cas, nous avons montré qu'une spécification incorrecte du modèle génétique provoquait une diminution, parfois très importante, du *lod score* et diminuait donc encore la puissance de la méthode [3, 4].

#### **Forte probabilité d'exclure faussement une liaison génétique**

Dans certains cas, la diminution du *lod score* peut être importante au point de conduire à exclure faussement la présence d'un facteur de risque dans la région du marqueur testé.

C'est ainsi que les premières analyses de liaison faites avec certaines maladies associées au système HLA rejetaient la présence d'un gène de susceptibilité dans la région elle-même. Elles situaient ce gène à une distance telle que les associations de ces maladies avec des antigènes HLA devenaient paradoxales. Ces études supposaient, pour la maladie étudiée, l'implication d'un allèle de susceptibilité rare. En fait, il semble maintenant clair que pour expliquer l'ensemble des données, cet allèle doit être fréquent dans la population générale.

On peut également illustrer le risque de fausse exclusion avec des maladies où l'on s'accorde pourtant sur l'existence de sous-entités mendéliennes. Pour le cancer du sein, par exemple, la plupart des analyses familiales [6-8] ont conclu à l'existence d'un mélange de formes génétiques à transmission autosomique dominante (cas génétiques) et de cas non transmis (cas sporadiques). Cependant, pour une femme atteinte à un âge donné, la probabilité d'être un cas génétique diffère d'une étude à l'autre. Or, nous avons montré que la spécification de cette probabilité influence les valeurs de *lod scores* (figure 1) et peut

conduire à exclure faussement la liaison dans certaines familles [9].

Soulignons qu'il n'existe pas nécessairement dans toutes les maladies à composante familiale une sous-entité mendélienne représentant une proportion non négligeable des cas. Le fait que certaines familles comportent de nombreux malades ne prouve nullement qu'existe un gène pathologique responsable. Des facteurs non génétiques peuvent induire une forte concentration familiale. Ainsi, l'existence de familles de médecins n'implique pas l'existence d'un gène déterminant cette profession ! Transmission génétique et transmission culturelle sont particulièrement difficiles à distinguer [10]. En résumé, il faut de solides arguments, d'ordre biologique ou résultant d'une bonne analyse des répartitions familiales, pour que l'existence d'une sous-entité mendélienne puisse être retenue, comme, par exemple, dans le cancer du sein, la sclérose latérale amyotrophique ou le syndrome de Beckwith-Wiedemann. Or, pour un grand nombre de maladies humaines, on ne dispose pas de tels arguments et la relation phénotype-génotype est totalement inconnue. Le risque de fausse exclusion pose alors un problème crucial et il se peut qu'au terme d'un travail long et coûteux, on en arrive même à exclure entièrement le génome !

#### **Et les faux positifs !**

En utilisant la méthode des *lod scores*, certaines études ont conclu à une liaison significative entre des maladies multifactorielles et des marqueurs génétiques pris au hasard sur le génome : un gène pour la psychose maniaco-dépressive dans HLA sur le chromosome 6 [11], puis sur le chromosome 11 [12] ; un gène de la schizophrénie sur le chromosome 5 [13]. Que sont devenus ces gènes ? Pourquoi ces résultats n'ont-ils pas pu être reproduits, malgré de nombreuses tentatives ? Certains ont même été infirmés par leurs auteurs [14]. On peut s'interroger sur l'obtention de résultats positifs dans ces maladies, pour lesquelles la mise en évidence d'une liaison génétique était *a priori* très peu probable. Nous avons montré que de tels résultats (faux positifs) pouvaient être la conséquence

## RÉFÉRENCES

14. Kelsoe JR, Ginns EI, Egeland JA, *et al.* Re-evaluation of the linkage relationship between chromosome 11p loci and the gene for bipolar affective disorder in the old order Amish. *Nature* 1989 ; 342 : 238-43.
  15. Clerget-Darpoux F, Babron MC, Bonaïti-Pellié C. Assessing the effect of multiple linkage tests in complex diseases. *Gen Epidemiol* 1990 ; 7 : 245-53.
  16. Weeks DE, Lehner T, Squires-Wheeler E, Kaufmann C, Ott J. Measuring the inflation of the lod score due to its maximization over model parameter values in human linkage analysis. *Gen Epidemiol* 1990 ; 7 : 237-43.
  17. Thomson G. Investigation of the mode of inheritance of the HLA associated disease by the method of antigen genotype frequencies among diseased individuals. *Tissue Antigens* 1983 ; 21 : 81-104.
  18. Risch N. A general model for disease-marker association. *Ann Hum Genet* 1983 ; 47 : 245-52.
  19. Thomson G, Bodmer W. The genetics of HLA and disease associations. In : Christiansen FB, Fenchel T, eds. *Measuring Selection in Natural Populations*. Berlin : Springer Verlag, 1977 : 545-64.
  20. Clerget-Darpoux F, Babron MC, Prum B, Lathrop GM, Deschamps I, Hors J. A new method to test genetic models in HLA associated diseases : the MASC method. *Ann Hum Genet* 1988 ; 52 : 247-58.
  21. Svejgaard A, Platz P, Ryder LP. Insulin-dependent diabetes mellitus : joint results of the 8th workshop study. In : Terasaki PI, ed. *Histocompatibility Testing 1980*. Los Angeles : UCLA Tissue typing laboratory 1980 ; 638-56.
  22. Cudworth AG, Woodrow JC. Evidence for HLA linked genes in « juvenile » diabetes mellitus. *Br Med J* 1975 ; 3 : 133-5.
  23. Clerget-Darpoux F, Babron MC, Deschamps I, Hors J. Complementation and maternal effect in insulin dependent diabetes. *Am J Hum Genet* 1991 ; 49 : 42-9.
  24. Bell GI, Horita S, Karam JH. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes* 1984 ; 33 : 176-83.
  25. Julier C, Hyer RN, Davies J. *et al.* Insulin-IGF2 region on chromosome 11p encodes a gene implicated in HLA-DR4-dependent diabetes susceptibility. *Nature* 1991 ; 354 : 155-9.
  26. Billiard M, Seignalet J. Extraordinary association between HLA-DR2 and narcolepsy. *Lancet* 1985 ; 1 : 226-7.
- d'analyses multiples [15]. En effet, dans une maladie où le mode de transmission est inconnu et où, de surcroît, le spectre clinique est souvent large, il est classique d'effectuer les calculs de *lod scores* sous différentes hypothèses génétiques et sous différents critères de classification. Si, de plus, de nombreux laboratoires travaillent indépendamment avec de nombreux marqueurs, la probabilité pour qu'une de ces équipes obtienne par hasard un résultat positif n'est pas négligeable. La valeur de *lod score* obtenue est alors difficilement interprétable, car on contrôle mal le nombre de tests qui ont été réellement effectués. Notons d'ailleurs que certaines règles simples, proposées dans la littérature [16] pour tenir compte des tests multiples, ne sont applicables qu'aux situations particulières envisagées par leurs auteurs et ne sont en aucun cas généralisables.
- En définitive, la stratégie de recherche systématique sur le génome avec la méthode des *lod scores* a de sérieux inconvénients : elle amène parfois à de faux positifs, mais surtout elle a peu de chance de mettre en évidence un facteur de risque génétique et même toutes les chances d'exclure sa présence. C'est pourquoi, il nous semble plus approprié de privilégier une autre stratégie, celle qui recourt à des marqueurs de gènes candidats.

### La stratégie alternative du gène candidat

#### Implication du gène candidat : détection, modélisation, évaluation

On appelle « gène candidat » un gène dont la fonction suggère qu'il joue un rôle dans la maladie étudiée. Par extension, on définira une « région candidate » comme contenant potentiellement un gène candidat. C'est le cas par exemple de la région HLA pour des maladies en relation avec le système immunitaire. On peut également définir des régions candidates par homologie à celles qui sont impliquées chez les animaux (souris, rat drosophile...). Il est naturel de rechercher si un gène candidat joue un rôle dans le processus pathogénique et s'il constitue un facteur de risque pour la maladie ; plus précisément si, parmi les différentes expressions de ce gène (allèles), l'une

d'entre elles augmente le risque de maladie (allèle de susceptibilité). En général, les différentes expressions ne sont pas directement mesurables et on peut chercher à disposer, sur le gène candidat, de marqueurs qui nous serviront d'indicateurs. Plus la corrélation entre les différentes formes d'expression du gène candidat et celles du marqueur est forte — autrement dit, plus le déséquilibre de liaison entre les allèles du marqueur et les allèles du gène candidat est grand —, meilleure sera l'information apportée par le marqueur sur le rôle du gène candidat.

Il est alors clair que la question posée n'est plus de localiser un gène pathologique. Elle est d'abord de confirmer que le gène candidat joue un rôle (détection), puis dans l'affirmative de préciser son mode d'action (modélisation) et enfin d'estimer le risque attribuable à ce gène (évaluation), comme on cherche à le faire pour n'importe quel facteur de risque en épidémiologie.

#### Information apportée par un marqueur de gène candidat

A cet effet, un marqueur du gène candidat peut apporter de l'information à deux niveaux : à celui de la population s'il existe un déséquilibre de liaison entre ses allèles et ceux du gène candidat, et au niveau familial si les allèles du marqueur et des allèles du gène candidat ne se transmettent pas de façon indépendante. Au niveau de la population, on évaluera le déséquilibre en comparant les distributions des allèles du marqueur chez les malades et dans un groupe témoin (test d'association). Au niveau familial, on testera l'indépendance de transmission des allèles (test de liaison génétique). Ces deux types d'information serviront, d'une part, à mettre en évidence le rôle éventuel du gène candidat et, d'autre part, à tester des modèles génétiques. Parmi les méthodes proposées, certaines utilisent l'information du marqueur au niveau de la population [17, 18], d'autres au niveau familial [19]. La méthode MASC [20] que nous avons proposée utilise les deux informations (voir annexe). Il est important de noter que ces informations ne sont pas redondantes et que leur prise en compte simultanée peut augmenter

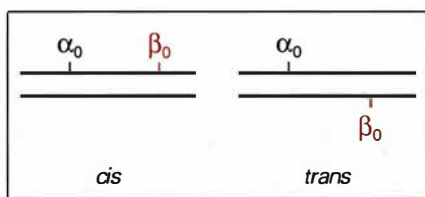


Figure 2. **Modèle de complémentation.** Susceptibilité due à la présence de deux allèles  $\alpha_0$  et  $\beta_0$  situés en deux emplacements très proches l'un de l'autre, soit sur le même chromosome de la paire (position cis), soit chacun sur un chromosome différent de la paire (position trans). Dans le premier cas, la susceptibilité a été héritée d'un seul parent, mimant une transmission dominante ; alors que dans le deuxième cas, chaque parent contribue à la susceptibilité, mimant une transmission récessive.

énormément non seulement la capacité à détecter le rôle du gène candidat, mais aussi celle à discriminer entre modèles génétiques. Par ailleurs, la méthode permet d'évaluer le risque pour la parenté d'un malade en fonction de toute l'information disponible.

### L'exemple HLA - diabète insulino-dépendant

Le cas du marqueur HLA pour le diabète insulino-dépendant (DID) constitue une bonne illustration de l'apport d'information d'un marqueur de gène candidat sur une maladie multifactorielle. Dans un premier temps, ce marqueur a permis de confirmer que les diabètes de types I et II étaient des entités différentes. En effet, seule la forme de type I présente une association avec certains antigènes du système HLA [21] : 95 % des malades possèdent des antigènes DR3 ou DR4, contre 20 % dans la population générale. De surcroît, les frères et sœurs atteints de DID sont beaucoup plus souvent identiques pour le marqueur HLA que ne le voudrait le hasard, témoignant d'une transmission non indépendante du marqueur et de la maladie [22]. Cela a permis d'établir qu'un facteur génétique, localisé dans la région HLA, était impliqué dans l'étiologie du DID. Au vu des données familiales, un mode récessif a d'abord été retenu pour ce facteur [19], mais il est apparu ultérieu-

rement que les données observées au niveau de la population ne s'ajustaient pas à un tel modèle [17]. La méthode MASC a alors permis de montrer que l'on pouvait expliquer l'ensemble des données par la présence chez les malades d'un hétérodimère formé par la complémentation de deux séquences spécifiques situées très proches l'une de l'autre, en position cis ou trans, dans la région HLA [23] (figure 2). L'une des séquences serait toujours associée au DR3 mais non au DR4 et *vice versa* pour l'autre séquence. Ce n'est clairement pas le cas des séquences qui ont été proposées au niveau moléculaire, à savoir l'absence d'acide aspartique en position 57 sur la chaîne DQ $\beta$  et la présence d'arginine en position 52 sur la chaîne DQ $\alpha$ . MASC a donc ici rejeté un modèle proposé au niveau moléculaire [23]. De quoi noter au passage que si l'épidémiologie génétique bénéficie grandement des outils de la génétique moléculaire, elle peut en retour lui fournir de nouvelles hypothèses de recherche ! Par ailleurs, nous avons observé que parmi les malades DR3, cet antigène provenait plus souvent de la mère que du père, ce qu'on pouvait expliquer par un effet maternel [23] ou une empreinte

parentale. Enfin, il existe d'autres facteurs familiaux, donc potentiellement génétiques, impliqués dans le DID [23]. Cela a conduit à étudier d'autres marqueurs de gènes candidats, et à montrer un effet possible de l'insuline [24] en interaction avec HLA [25]. Nous avons récemment étendu la méthode MASC pour prendre en compte l'information simultanée de marqueurs de deux gènes candidats.

Nous avons aussi calculé le risque de récurrence chez les germains (frères ou sœurs) du patient en utilisant l'information HLA disponible (figure 3). Les informations pertinentes pour affiner la prédiction de risque, qui est ici d'environ 5 % sans information préalable, sont : le génotype HLA du patient (risque élevé si ce patient est DR3 DR4), le nombre d'haplotypes en commun du patient et du germain considéré, la présence ou non de l'antigène DR3 chez la mère, le statut des parents (malades ou non).

### Perspectives dans les maladies multifactorielles

Les marqueurs génétiques peuvent apporter de l'information sur la com-

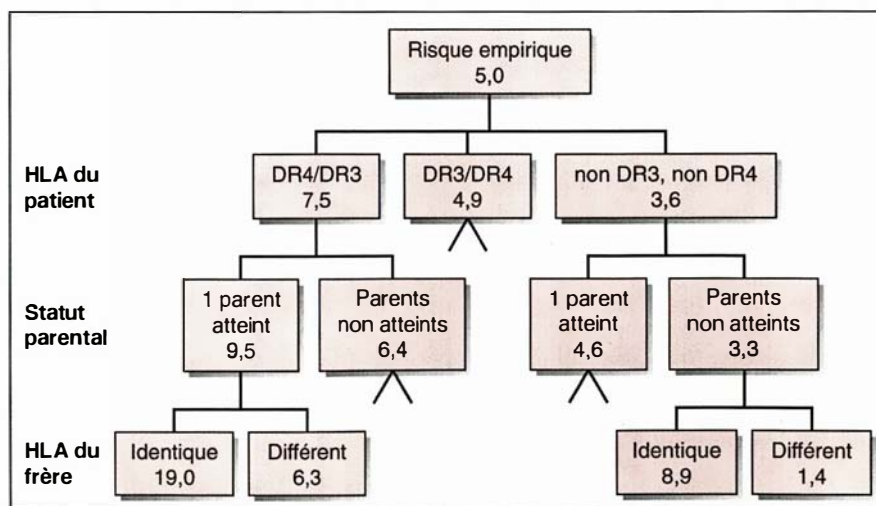


Figure 3. **Risque de diabète insulino-dépendant pour un frère d'un malade selon l'information disponible.** Le risque est calculé en tenant successivement compte du 1) génotype HLA du malade au locus DR (dans l'ordre allèle paternel, allèle maternel) ; 2) statut des parents vis-à-vis de la maladie ; 3) fait que le frère considéré est identique ou non au malade pour le marqueur HLA. Notons que le risque correspondant au génotype DR4 DR3 est plus élevé lorsque le DR3 provient de la mère.

posante génétique d'une maladie complexe, comme on vient de le voir sur l'exemple du diabète insulino-dépendant. Cette information permet en particulier de déterminer d'éventuels sous-groupes d'individus à risque différent pour un gène candidat donné. Notons qu'on dispose alors d'une puissance *a priori* plus importante pour détecter le rôle potentiel d'un autre gène candidat et son interaction avec le premier. On peut de la sorte espérer détecter progressivement des facteurs génétiques impliqués et évaluer les risques correspondants. En outre, l'identification de sous-groupes à risques génétiques différents peut faciliter la détection d'éventuels facteurs de milieu. La narcolepsie est un bon exemple d'interaction entre facteur(s) génétique(s) et environnemental(aux) (ces derniers restant aujourd'hui inconnus). On sait que la quasi-totalité des malades possèdent l'antigène DR2 du système HLA [26]. Comparer l'exposition de ces malades à celle des témoins DR2, ayant donc le même facteur de risque génétique, facilitera la mise en évidence des facteurs de milieu impliqués. On observera bien évidemment une plus grande différence que si l'on choisit les témoins dans la population générale.

Mettre en évidence des facteurs génétiques dans les maladies multifactorielles n'est certes pas une finalité en soi, puisque chercher à les modifier relève clairement du non-sens. Cependant, une bonne connaissance du processus pathogénique augure mieux d'une mise au point de thérapeutiques adaptées. La connaissance de facteurs génétiques de susceptibilité facilite par ailleurs la détection de facteurs de risque de milieu.

En définitive, les marqueurs génétiques peuvent être utilisés aussi bien pour l'étude des maladies monogéniques que pour celles des maladies multifactorielles, mais les problèmes posés, et par conséquent les démarches à adopter, sont fondamentalement différents dans les deux cas. Dans le premier cas, il s'agit de localiser un gène dont on connaît les caractéristiques (dominance, fréquence, pénétrance) mais non la fonction. Dans le second, on part d'un gène dont la localisation et la fonction sont connues et l'on cherche à savoir s'il est impliqué dans

la maladie et comment. Les stratégies à adopter dans ces deux situations demandent des efforts différents, tant au niveau du choix des individus que des marqueurs à étudier. Dans la première situation, on focalisera sur des généalogies informatives pour la liaison génétique et on étudiera des marqueurs régulièrement espacés sur le génome. Dans la seconde, au contraire, on voudra disposer d'un échantillon important de malades non apparentés, même si les données sont plus restreintes au niveau familial. En ce qui concerne les marqueurs, on les étudiera sur des gènes candidats en recherchant des associations avec la maladie et un maximum de polymorphisme. Notons que dans certaines maladies, on ne dispose actuellement que d'un nombre limité de gènes candidats. Il nous semble important, dans le cadre des maladies multifactorielles, que la recherche s'oriente vers la mise en évidence de gènes candidats en

accordant toute leur place aux études de physiologie fondamentale.

Une grande confusion règne aujourd'hui sur la façon d'étudier les maladies multifactorielles. Certes, la complexité de ces maladies explique la difficulté à s'accorder sur des stratégies d'étude. Mais il semble aussi que le succès grandissant de la génétique moléculaire dans l'étude des maladies monogéniques ait tendance à ériger en dogmes les stratégies de recherche aléatoire sur le génome. La cartographie génétique représente actuellement un défi majeur pour les généticiens, et on assiste à un rapide enrichissement des cartes publiées. Voir cependant sur les cartes se côtoyer des maladies monogéniques et multifactorielles ajoute à la confusion. Avant de s'engager sur des études longues et coûteuses, il faudrait mieux s'accorder sur les questions fondamentales : que cherchons-nous et pourquoi ? Peut-être saurions-nous alors dire comment ■

## Annexe

### Méthode MASC (marker association segregation chi-square)

Pour une maladie donnée, la méthode MASC [20] permet d'étudier si certains gènes candidats sont facteurs de risque. Plus précisément elle vise à :

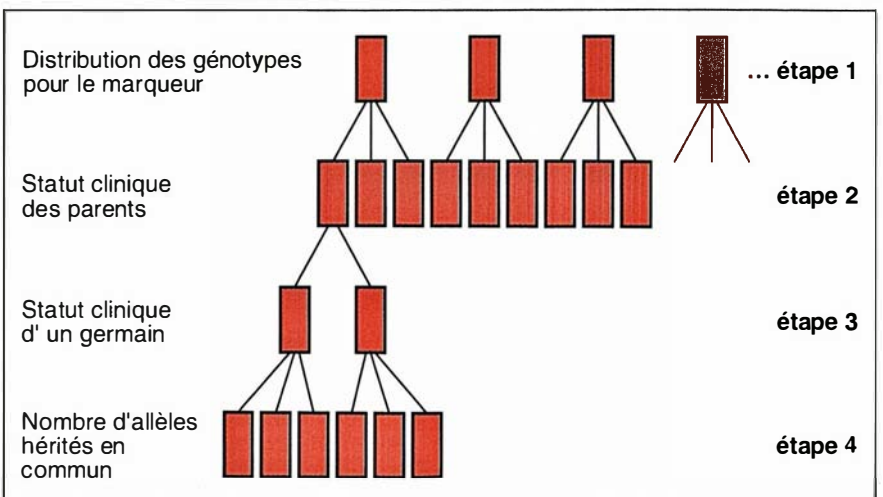
- mettre en évidence leur effet ;
- modéliser leur rôle et interaction ;
- évaluer le risque attribuable.

Elle utilise l'information fournie par un marqueur du gène candidat au niveau population et au niveau familial. Pour ce faire, un échantillon de malades est classé suivant le schéma indiqué. Dans une première étape les catégories correspondent aux différents génotypes pour le marqueur. A partir de chacune des catégories ainsi définies, les malades sont à nouveau

classés suivant le statut clinique de leurs parents (2, 1 ou 0 parents atteints), puis suivant le statut d'un de leurs germains (frère ou sœur) et en dernier lieu suivant le nombre d'allèles hérités en commun avec ce germain (2, 1 ou 0).

Un programme informatique permet de :

- calculer sous différents modèles génétiques les nombres attendus dans chacune des catégories ;
- tester ces modèles en comparant les données observées à celles attendues (test de  $\chi^2$ ) et de retenir le modèle ayant la meilleure adéquation ;
- calculer les risques pour les apparentés d'un malade à partir du modèle retenu.



## Summary

### Genetic markers and multifactorial diseases. Which strategy to adopt ?

Genetic markers provide information on the determinism of multifactorial diseases. However the question addressed in those diseases is fundamentally different from the one in monogenic diseases. In the latter case a mutated gene is known to be responsible for the disease and the aim is to identify the primary defect and the involved protein. Thus, genetic markers are used to localize the corresponding DNA sequence. In this context, a systematic search for linkage with random markers is an efficient strategy. For multifactorial diseases — which are not due to a rare mutation, but to the unfavourable combination of several genetic and/or environmental factors — the aim is to detect the risk factors involved and to evaluate their effect and interaction. Genes which are functionally related to the disease are then good candidates for risk factors. Focusing on markers of these genes appears to be a more appropriate strategy than using information on random genetic markers. In addition, the marker information does not have to be restricted to the one provided by linkage, but extended to the one on preferential associations between the alleles of the candidate gene and the marker alleles.

#### **Le Professeur Claude GRISCELLI**

Lauréat 1992  
du Prix de la Recherche  
de la Fondation Athena-Institut  
de France

Claude GRISCELLI, Chef du département de pédiatrie et de l'Unité d'immunologie et d'hématologie de l'Hôpital Necker-Enfants Malades à Paris, a été primé par la Fondation Athena-Institut de France pour ses travaux sur les déficits immunitaires de l'enfant.