

Dissection génétique des maladies à hérédité complexe

Dominique Campion

Dans les maladies à hérédité complexe, la corrélation entre génotype et phénotype n'est pas directe et n'obéit pas aux lois de Mendel. Les gènes dits de susceptibilité ne sont ni indispensables ni suffisants pour causer une maladie. Celle-ci n'apparaît que lorsqu'un seuil de susceptibilité est franchi, qui peut être atteint par une action conjointe de plusieurs facteurs de risque. Les analyses de liaison utilisant des méthodes dépendantes d'un modèle de transmission sont inadaptées pour la recherche des localisations chromosomiques de ces facteurs alors que les méthodes indépendantes de ce modèle manquent de puissance et fournissent des localisations imprécises. Les études d'association utilisant des polymorphismes bi-alléliques risquent d'engendrer un grand nombre de faux positifs, compte tenu du nombre de tests réalisés lors d'un criblage systématique du génome. Ces études trouvent l'essentiel de leur intérêt lorsqu'elles s'appliquent à des « gènes candidats ». Afin de caractériser ces gènes, il est indispensable de connaître les voies biologiques impliquées. Pour définir ces voies, la mise en évidence de sous-entités mendéliennes et l'utilisation de phénotypes intermédiaires sont des étapes nécessaires. Cette approche, illustrée par des exemples concernant les maladies neuro-psychiatriques, conduit à mieux formuler certains concepts de l'hérédité en pathologie humaine.

ADRESSE

D. Campion: Inserm EMI 9906, Faculté de médecine, 22, boulevard Gambetta, 76183 Rouen Cedex, France.

E-mail: dominique.campion@univ-rouen.fr

m/s n° 11, vol. 17, novembre 2001

Au cours de ces vingt dernières années, près d'un millier de gènes impliqués dans des maladies mendéliennes ont été localisés et un grand nombre d'entre eux a été cloné. Toutefois, outre ces maladies à transmission mendélienne dont la prévalence est faible, une compo-

sante génétique est également présente dans des affections très fréquentes, telles que les maladies cardiovasculaires, le diabète, les maladies inflammatoires ou les maladies neuropsychiatriques. Dans ces maladies, qui sont dites « complexes » ou « multifactorielles », la récurrence chez les apparentés d'un patient est en géné-

ral assez faible et ne répond pas aux lois de Mendel. Alors que dans une maladie mendélienne, une mutation rare, au niveau d'un seul gène, est nécessaire et suffisante pour faire apparaître le phénotype morbide, les maladies à hérédité complexe impliquent très souvent la présence simultanée de nombreuses variations génétiques, affectant des gènes différents, et ayant chacune un effet modeste. Ce concept n'est pas exactement superposable à celui de maladie polygénique. Les modèles polygéniques utilisés en analyse de ségrégation impliquent en effet une contribution strictement additive de plusieurs gènes, au détriment d'interactions épistatiques plus complexes. Pour cette raison, il est donc préférable de parler de déterminis ne « multigénique » plutôt que « polygénique ». Chacun de ces variants génétiques, considéré isolément, n'est ni indispensable ni suffisant pour entraîner la maladie. Au contraire, celle-ci n'apparaît que lorsqu'un certain « seuil » de susceptibilité est dépassé [1] et ce seuil, dans un modèle de portée très générale, peut être atteint par l'action conjointe de facteurs génétiques (résultant de diverses combinaisons alléliques entre différents locus) et de facteurs environnementaux. Dans ces circonstances, la fréquence de chaque allèle conférant un risque pour la maladie peut donc ne s'avérer que très légèrement augmentée dans un groupe de malades comparé à un groupe de témoins. De plus, lorsque l'on considère la co-ségrégation de tel allèle particulier et de la maladie au sein d'une généalogie, le type de transmission est très différent de celui observé dans le cas

Tableau I. Quelques facteurs de risque génétique.

Maladie	Gènes
Maladie d'Alzheimer	Apolipoprotéine E
Hypertension artérielle	Angiotensinogène
Diabète insulino-dépendant	- Insuline
	- HLA
Diabète non-insulino dépendant	Calpaïne 10
Emphysème	$\alpha 1$ antitrypsine
Thrombose veineuse	Facteur V Leiden
Hypertriglycéridémie	Lipoprotéine lipase
Sclérose en plaques	HLA
Polyarthrite rhumatoïde	HLA
Spondylarthrite ankylosante	HLA
Maladie cœliaque	HLA

des maladies mendéliennes. Alors que, dans celles-ci, la co-ségrégation entre le génotype pathologique et la maladie est parfaite (figure 1), il n'en va pas de même en ce qui concerne les facteurs de risque. Il existe d'une part des individus portant le génotype « à risque » au locus considéré, et non atteints (cela correspond au concept de pénétrance incomplète de ce génotype) et, d'autre part, des individus non porteurs de ce génotype qui sont atteints (il s'agit alors de phénotopies). En conséquence, dans le cas des maladies à hérédité complexe, la correspondance génotype/phénotype est, pour un gène donné, très mauvaise (figure 2). Dès lors, il n'est pas étonnant que si peu de facteurs de risque génétiques aient été caractérisés, comparativement à l'extraordinaire moisson de mutants rares identifiés dans les maladies mendéliennes. Comme on peut le constater sur le *Tableau I*, les maladies dans lesquelles des facteurs de risque génétique ont été reconnus avec certitude sont encore fort peu nombreuses.

À la recherche de méthodes

Les méthodes utilisées pour identifier les gènes impliqués dans les maladies humaines sont de deux types : (1) les études de liaison génétique fondées sur la co-ségrégation de la maladie et des allèles à un locus marqueur au sein de généalogies ; (2) et les études d'association allélique cherchant à mettre en évidence une différence de fréquence allélique au niveau d'un tel locus entre des sujets non apparentés, respectivement atteints et non atteints (pour revue, voir [2]).

Méthode des lod-score

Grâce à la méthode des lod-score, un grand nombre de gènes impliqués

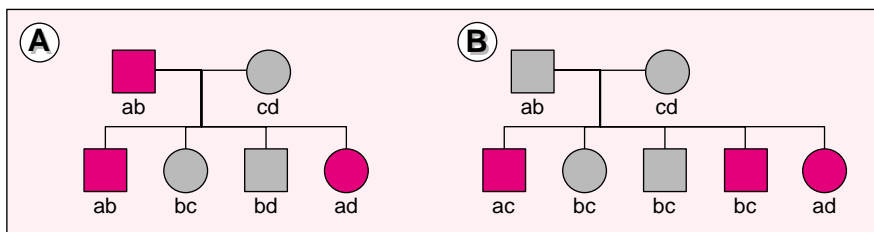


Figure 1. Maladie mendélienne et agrégation familiale due à des facteurs de risque. A. Transmission autosomique dominante d'une mutation rare (allèle a). B. Ségrégation d'un facteur de risque (allèle a), le père transmetteur n'est pas atteint (pénétrance incomplète), le fils ayant un génotype bc est atteint (phénotopie).

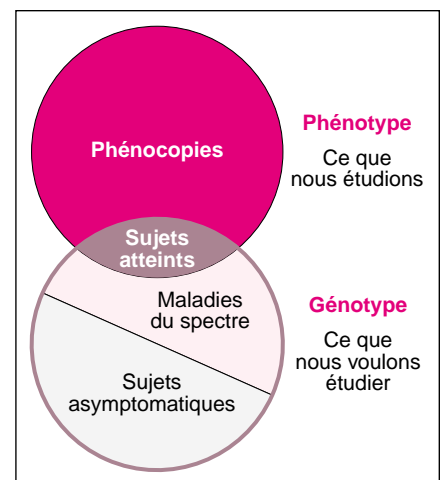


Figure 2. Non-concordance entre phénotype et génotype dans le cas d'un facteur de risque génétique.

dans des maladies mendéliennes a pu être cartographié. Cette méthode d'analyse de liaison requiert la connaissance du mode de transmission au locus maladie et la spécification de paramètres concernant cette transmission (fréquence de l'allèle pathologique, pénétrances des divers génotypes). Lorsque ces paramètres sont spécifiés de manière incorrecte, il en résulte une perte de puissance pour détecter une liaison existante [3]. Dans le cas des maladies à hérédité complexe, la méconnaissance de ces paramètres conduit généralement à multiplier les modèles et le nombre de tests réalisés, ce qui a pour conséquence la détection de liaisons inexistantes, du seul fait du hasard. La méthode des lod-score est donc particulièrement inadaptée puisqu'elle produit d'une part des faux négatifs, liés à la mauvaise spécification des paramètres génétiques, et d'autre part des faux positifs, liés aux tests multiples.

Méthode des paires de germains

La méthode des paires de germains atteints (*affected sib pairs*) est une méthode d'analyse de liaison « indépendante du modèle ». Son but est de montrer que la transmission d'un segment chromosomique, dans des paires de sujets atteints provenant de fratries, n'est pas compatible avec une ségrégation aléatoire, mais qu'au contraire, ces sujets héritent d'un même segment chromosomique parental plus souvent que ne voudrait le hasard. Pour un locus marqueur donné, elle cherche à estimer, au sein de paires de patients, la proportion d'allèles identiques par descendance (*identical by descent, IBD*), c'est-à-dire provenant d'un même allèle parental. Si le marqueur ségrège indépendamment de la maladie, la proportion de paires de germains atteints partageant 0,1 ou 2 allèles IBD au locus marqueur est respectivement 1/4, 1/2, 1/4. Si ce locus marqueur est lié au locus maladie, deux frères atteints auront reçu plus de 50 % d'allèles IBD.

Cette méthode indépendante du modèle ne requiert pas la spécification *a priori* d'un modèle de transmission pour la maladie. Cependant, la puissance de la méthode pour détecter une liaison dépend, elle, du

modèle sous-jacent. Puisque cette méthode est fondée sur l'étude de sujets atteints, elle s'affranchit du problème posé par les pénétrances incomplètes. En revanche, la présence de phénotopies est susceptible d'interférer avec la détection de la liaison. Ce facteur est peu important lorsque l'on étudie une maladie due à la présence d'un allèle mutant rare, il peut devenir très gênant lorsque l'allèle impliqué sur le gène étudié est un allèle fréquent contribuant faiblement à la susceptibilité dans un cadre multigénique. La puissance de la méthode des paires de germains pour détecter la liaison entre une maladie et un locus de susceptibilité dépend de la contribution propre de ce locus à l'augmentation du risque chez cette classe d'apparentés, par comparaison avec la prévalence de la maladie dans la population générale (*sibling recurrence risk ratio λ_s*) [4]. De façon générale, cette méthode est peu puissante, elle requiert, pour des $\lambda_s < 1,5$ (ce qui correspond le plus souvent aux valeurs attendues pour les facteurs de risque qui nous intéressent), des effectifs considérables pour détecter une liaison. Ainsi, il a été calculé que la mise en évidence du locus de l'Apolipoprotéine E (APO E) comme facteur de risque dans la maladie d'Alzheimer ($\lambda_s = 2$), nécessiterait un effectif minimal de plus de 250 paires de germains, en utilisant des marqueurs complètement informatifs et en supposant l'absence de recombinaison. En se plaçant dans des conditions plus réalistes d'un taux modéré de recombinaisons et d'une informativité réduite des marqueurs, ce nombre déjà élevé doit en réalité être multiplié par un facteur deux ou trois [5]. Notons que l'identification du facteur de risque APOE 4 est généralement possible à partir 50 malades et 50 témoins lors d'une étude d'association.

Enfin, la méthode des paires de germains n'est généralement pas capable de localiser un gène avec précision. A titre d'exemple, la mise en évidence récente d'un nouveau facteur de risque putatif pour la maladie d'Alzheimer, dans une étude qui a nécessité plus de 400 paires de germains, a certes abouti à une localisation, mais dans une région de 50 mégabases sur le bras long du chromosome 10 [6].

Études d'association allélique

Les limites inhérentes à l'analyse de liaison dans la recherche des facteurs de risque génétique amènent de plus en plus à promouvoir des stratégies alternatives. La recherche d'une association allélique entre un marqueur génétique et une maladie est alors préconisée. Le principe en est fort simple : un allèle A1 est dit associé à une maladie s'il est présent plus fréquemment parmi des malades non apparentés que parmi les témoins. Pour qu'une telle étude soit valide, témoins et malades doivent provenir d'une même population panmictique*. Cette condition est très difficile à respecter : une association retrouvée lors d'une étude cas/témoins peut relever uniquement d'un effet de stratification, c'est-à-dire d'une différence de fréquence allélique au locus considéré entre la population d'où proviennent les malades et celle d'où proviennent les témoins (*m/s 1998, n° 11, p. 1301*). Il est possible de s'affranchir de ce problème en choisissant des témoins au sein même des familles des malades. C'est le principe du TDT (*transmission disequilibrium test*) [7], dans lequel un allèle transmis par un parent à un enfant atteint est apparié avec le second allèle non transmis par ce parent. La robustesse du TDT vis-à-vis des biais de stratification se paie toutefois par une diminution de la puissance à détecter l'association. Enfin, quel que soit le test utilisé, la question des faux positifs engendrés par les tests multiples demeure : tester n locus avec k allèles revient à réaliser environ $n(k-1)$ tests indépendants et la signification de chaque test doit, en toute rigueur, être divisée par ce facteur (correction de Bonferroni). Par exemple, une valeur de signification nominale de $p = 0,0005$ est requise pour atteindre un simple taux de 5 % de faux positifs lorsque l'on teste indépendamment 100 polymorphismes bi-alléliques sur des gènes différents.

* *Panmixie* : état d'un groupe de population humaine où les unions entre individus sont liées au seul hasard.

Association allélique et déséquilibre de liaison

Une association allélique positive a deux significations possibles: (1) soit l'allèle A1 (associé) constitue directement un facteur causal pour la maladie, et dans ce cas l'association sera le plus souvent retrouvée dans l'ensemble des populations étudiées, quelle que soit leur origine ethnique (il peut toutefois exister des exceptions à cette règle); (2) soit cet allèle A1 ne joue en lui-même aucun rôle causal, mais se trouve en déséquilibre de liaison (DL) avec un allèle causal B1.

Un déséquilibre de liaison entre deux allèles situés sur des locus différents s'observe chaque fois que la fréquence des gamètes porteurs de ces deux allèles diffère de celle résultant du simple produit de leurs fréquences alléliques [8]. Il existe plusieurs mécanismes susceptibles de produire un DL, l'un d'entre eux (celui qui nous intéresse ici) étant fondé sur la proximité physique des deux locus sur un même chromosome. Dans ce dernier cas, l'existence du DL s'interprète par le fait que, dans la population considérée, la plupart des individus portant « l'allèle à risque » B1, portent également l'allèle A1 sur un même segment chromosomique et que ce « déséquilibre gamétique » se maintient au fil des générations, en dépit des recombinaisons et des nouvelles mutations.

Un tel déséquilibre peut résulter de divers mécanismes. Dans certains cas, le DL s'est constitué au sein d'une petite population « fondatrice » d'où est issue la population actuelle. La possibilité d'identifier un « allèle à risque » B1 par DL avec des marqueurs proches [9] dépend alors en premier lieu de la fréquence de cet allèle. Plus celui-ci est commun, plus la probabilité qu'il ait été associé à des contextes chromosomiques différents dans la population ancestrale croît, moins il devient possible de le détecter par DL. Ainsi, même une population remarquablement isolée, telle la population finlandaise, qui semble avoir évolué à partir d'une petite population fondatrice d'environ un millier d'individus il y a 100 générations, et qui est bien adaptée pour l'identification par DL de

gènes mendéliens rares, pourrait se révéler inutilisable pour l'étude de facteurs de risque fréquents [10]. De fait, Kruglyak [11] a montré que, lorsque la fréquence d'un « allèle à risque » se situait au-delà de 0,10, la taille de la population fondatrice d'où provient le DL initial devait être inférieure à 100 individus pour permettre d'identifier cet allèle par DL. Outre la taille de la population ancestrale, l'histoire ultérieure de cette population est évidemment un autre facteur à prendre à compte. On a parfois considéré que les brassages de population constituaient un facteur défavorable au maintien d'un DL et que celui-ci était donc maximal dans des populations isolées. En effet, un « allèle morbide » B1 peut montrer un DL avec l'allèle A1 dans une population, avec l'allèle A2 dans une seconde, et aucune association avec chacun de ces allèles dans une troisième population résultant du mélange des deux premières. C'est ce que l'on observait lorsque le diagnostic de la drépanocytose était réalisé au moyen d'un marqueur flanquant le locus du gène de β -globine, marqueur en DL avec la mutation causale: dans une partie du bassin méditerranéen, le DL concernait l'allèle A1, dans l'autre l'allèle A2, et dans des villes comme Naples où les populations s'étaient mélangées, plus aucune association n'était retrouvée. Les populations résultant de multiples brassages ethniques seraient donc mal adaptées pour l'étude du DL [10]. Toutefois, les « flux géniques continus » provoqués par les brassages de populations n'ont pas toujours cet effet égalisateur et, au contraire, dans certaines circonstances, ils concourent à créer et à maintenir un DL entre locus physiquement proches dans la population mélangée. Ceci peut être mis à profit pour réaliser des études d'association [12]. La question du choix des populations à partir desquelles mener les études fondées sur le DL est donc très discutée. Alors que les isolats étaient initialement considérés comme la panacée, la tendance actuelle est plutôt à l'utilisation de populations continentales de structure hétérogène [13].

Au total, les études d'association, malgré leurs nombreuses limites, sont certainement une approche

utile pour aborder l'étude des facteurs de risque génétique. Au regard de la très faible distance physique au niveau de laquelle un DL est susceptible de perdurer au fil des générations sur la molécule d'ADN, on considèrerait classiquement qu'elles ne s'appliquaient qu'à l'étude de « gènes candidats », et non dans le cadre d'un criblage systématique, ce qui limitait d'autant leur emploi. Cette restriction fait actuellement l'objet d'une controverse sur laquelle nous reviendrons ultérieurement.

Le succès de la dernière décennie: APO E et maladie d'Alzheimer

Compte tenu de toutes ces difficultés méthodologiques, on ne s'étonnera guère que la moisson de facteurs de risque récoltée dans les pathologies multifactorielles ait été particulièrement maigre ces dernières années, surtout comparée aux énormes efforts déployés, particulièrement dans le cadre des pathologies neuropsychiatriques. Un succès indéniable a pourtant marqué cette période, il s'agit de l'identification de l'allèle E4 du gène de l'apolipoprotéine E (*APOE*) comme facteur de risque dans la maladie d'Alzheimer. Revenons sur l'histoire de cette découverte. Contrairement à ce qui a très souvent été écrit, l'attention n'a pas initialement été attirée sur ce locus par une analyse de liaison mais, par la mise en évidence d'une association: en 1987, l'équipe de Gérard Schellenberg, à Seattle [14] rapporte l'existence d'une association allélique entre la maladie d'Alzheimer et une série de polymorphismes bialléliques situés dans le gène *APO CII* (au sein du même complexe de gènes qu'*APO E*, et séparé de celui-ci par seulement 40 kb). Ce travail aboutit en 1995 à l'identification du gène *APO E* comme déterminant premier de l'association allélique [15, 16].

Ce n'est que dans un second temps que les analyses de liaison [17], réalisées de manière répétée sur cette région du génome, parvinrent à établir l'existence d'une liaison génétique. Elles donnèrent d'ailleurs une localisation fautive pour le gène de susceptibilité, puisque, dans chaque étude publiée, le complexe *APO*

CII/APO E était fermement exclu en analyse de *lod-scores* bipoint, et le locus putatif était situé à une distance de plusieurs centimorgans en analyse multipoint. Cette fausse localisation fut probablement due à une erreur de paramétrage, puisque ces études supposaient l'existence d'un allèle morbide rare (1 %) transmis de façon dominante, alors que le véritable modèle, on le sait à présent [18], est celui d'un allèle « à risque » fréquent (15 %) transmis de façon semi-dominante avec des courbes de pénétrances différentes pour l'homozygote et l'hétérozygote. Enfin, comme l'ont souligné à juste titre Risch et Merikangas [5], on sait que la réalisation d'une étude de liaison avec des marqueurs associés à la maladie accroît la chance de détecter cette liaison. Or, c'est l'utilisation du marqueur APO E, ou de marqueurs en DL avec lui, qui ont permis la détection de la liaison dans ces études, ce qui autorise Risch et Merikangas à conclure: «*Thus, in reality, the "linkage" discovery on chromosome 19 was actually based on an association between marker loci and the disease*»*.

Reste une question: pourquoi s'être intéressé au gène APO CII? En quoi un gène impliqué dans le transport des lipides pouvait-il, au regard des connaissances de l'époque, être considéré comme « gène candidat » dans la maladie d'Alzheimer? Schellenberg *et al.*, dans leur travail princeps [14], ont étudié deux locus, celui de l'APO CII et celui d'un facteur du complément, situé sur le chromosome 6. Pour ce qui concerne l'APO CII, voici comment ils justifient leur choix: «*The APO CII locus was chosen because of availability of a restriction fragment length polymorphism (RFLP) probe and appropriate control population data*»**. Ainsi, cette remarquable découverte fut l'œuvre du pur hasard... et ne saurait en soi servir de modèle de stratégie pour l'avenir.

* Ainsi, en réalité, la découverte d'une « liaison » sur le chromosome 19 fut de fait fondée sur une association entre des sites marqueurs et la maladie.

** Le locus APO CII fut choisi à cause de la disponibilité d'une sonde pour un polymorphisme de restriction, et de l'existence de données témoins adaptées.

Le recours aux SNP: un leurre ?

L'accent mis à l'heure actuelle sur les études d'association implique de recourir aux marqueurs bialléliques, rebaptisés SNP (*single nucleotide polymorphisms*) [19]. De tels marqueurs sont extrêmement fréquents au sein du génome, à la fois dans les régions codantes et non codantes. Une étude récente [20] en a identifié 560 au niveau de 106 gènes candidats pour des maladies cardiovasculaires, endocriniennes ou neuropsychiatriques (dont 392 dans les régions codantes, et le reste au niveau des promoteurs). A noter que la fréquence de l'allèle minoritaire dépasse 5 % dans près de 50 % des cas, ce qui montre que ces allèles n'exercent pas d'effets délétères fortement contre-sélectionnés, contrairement aux mutations impliquées dans les maladies mendéliennes, généralement maintenues en dessous d'une fréquence de 1 % par la sélection.

Les possibilités actuelles de génotypage de milliers de SNP différents ont conduit Risch et Merikangas [21] à proposer d'étudier systématiquement l'ensemble des SNP des promoteurs et des régions codantes de « gènes candidats » au cours d'études d'association. Puis Collins *et al.* [22] ont suggéré d'utiliser des SNP espacés tous les 6 kilobases (distance sensée représenter la « zone minimale de DL stable entre deux marqueurs » au sein de la population humaine) [11] pour rechercher systématiquement des marqueurs associés à des maladies par DL. Ceci requiert de typer 500 000 SNP différents au cours d'un criblage génomique. En dépit de sa lourdeur, cette proposition a reçu un accueil enthousiaste, malgré certaines mises en garde. Il a notamment été objecté qu'une telle stratégie se heurterait (outre sa lourdeur technologique) à des difficultés redoutables qui tiennent à la grande hétérogénéité du DL au sein du génome et à sa non linéarité en fonction de la distance entre marqueurs [23, 24].

Mais à supposer ces problèmes résolus, ce type d'analyse nécessitera une masse énorme de tests, qu'il faudra ensuite absolument corriger. Le parti pris d'utiliser des critères « relâchés » pour corriger les tests multiples dans le cadre des analyses de liaison [25] a

conduit à engendrer une grande quantité de faux positifs: il n'est plus guère, aujourd'hui, de région chromosomique pour laquelle, dans une maladie comme la schizophrénie, une « liaison suggestive » n'ait été publiée. Que ferons-nous demain avec 1 000 ou 2 000 SNP supposés se trouver « en DL » avec la maladie ? D'un autre côté, l'utilisation de critères stringents (c'est-à-dire correction de Bonferroni) pour juger de la pertinence d'un résultat risque fort de s'avérer trop brutale et, à l'inverse, de produire des faux négatifs. Enfin, le critère de réplification sera particulièrement difficile à obtenir puisque le DL est un phénomène hautement dépendant de la population étudiée. Dans ces conditions, il est plus que probable qu'en dépit de l'énorme effort envisagé, la recherche de DL au moyen de marqueurs aléatoires conduira au même fiasco que les tentatives de criblage systématique par liaison génétique. On ne résoud pas un problème qui est fondamentalement d'ordre méthodologique par la seule inflation technologique. Il est permis de penser que les études d'association utilisant les SNP ne donneront des résultats appréciables que si elles sont appliquées à des « gènes candidats ». Si nous voulons progresser, nous devons donc disposer d'un moyen pour caractériser ces gènes.

À la découverte de nouveaux concepts

Du complexe au simple: les sous-entités mendéliennes

En ce qui concerne les maladies neuro-psychiatriques, certains ont proposé de considérer l'ensemble des gènes exprimés dans le système nerveux comme « gènes candidats », ce qui revient évidemment à vider le concept de tout sens. La focalisation sur des ensembles plus ciblés semble malheureusement impossible, puisque, dans un grand nombre de ces maladies, les mécanismes physiopathologiques restent inconnus ou ne sont décrits que de façon très fragmentaire. Une observation importante doit toutefois être prise en compte, qui incite à plus d'optimisme. Dans plusieurs cas de maladies à hérédité complexe, des sous-entités mendéliennes, elles-mêmes

Tableau II. Sous-entités mendéliennes au sein de pathologies à hérédité complexe.

Pathologie	Fréquence de la sous-entité mendélienne	Hétérogénéité génétique
DNID	5 %	– MODY 1 → HNF – 4 α – MODY 2 → Glucokinase – MODY 3 → HNF – 1 α – MODY 4 → IPF – 1
Hypercholestérolémie	4 %	– Récepteur LDL – Apolipoprotéine B – FH3 (1p32)
Cancer du sein	8 %	– BRCA 1 – BRCA 2
Maladie d'Alzheimer	2 %	– APP – Préséniline 1 – Préséniline 2
Maladie de Parkinson	?	– Alpha synucléine – Parkine – UCH – L1
Démence fronto-temporale	?	– Tau – ?
Sclérose latérale amyotrophique	10 %	– SOD – 1 – ?

MODY : maturity onset diabetes of young; *DNID* : diabète non insulino-dépendant; *HNF* : hepatocyte nuclear factor; *IPF* : insulín promoter factor; *BRCA* : breast cancer; *APP* : amyloid precursor protein; *FH3* : locus 3 de l'hypercholestérolémie familiale, gène encore inconnu; *SOD* : superoxyde dismutase; *UCH - L1* : ubiquitine carboxy-terminal hydrolase L1.

hétérogènes génétiquement, ont été identifiées (Tableau II).

Quel est donc le rapport entre ces formes mendéliennes, qui ne représentent qu'une petite fraction des malades, et la majorité des cas répondant à un déterminisme non mendélien ?

Mutations et polymorphismes

Pendant longtemps, il a été admis plus ou moins implicitement que maladies mendéliennes et multigéniques concernaient deux types de caractères biologiques radicalement différents : il existerait des traits phénotypiques qui seraient à sous-déterminisme strictement monogénique, alors que d'autres seraient multigéniques. En fait, même à un niveau phénotypique très élémentaire, considérer que l'information contenue dans un gène suffit à contrôler la production d'une protéine constitue déjà une simplification abusive. Dès que l'on quitte le niveau protéique pour s'intéresser à des phénotypes plus intégrés en terme de fonctionnement cellulaire, le déterminisme

strictement monogénique d'un caractère biologique a encore moins de sens. Les gènes mutés dans les maladies mendéliennes n'agissent pas de façon « isolée » mais agissent au sein de *voies biologiques* et contribuent à ce titre à l'expression de caractères biologiques quantitatifs, dont le déterminisme ultime est multigénique. L'apparition dans une famille d'une maladie « monogénique », c'est-à-dire d'une variation phénotypique qualitative due à la mutation d'un seul gène, signe la survenue d'une perturbation majeure au niveau de l'un des chaînons d'une telle voie biologique. Les mutations à l'origine de maladies mendéliennes, qu'elles soient transmises de façon dominante ou récessive, ont toujours de profondes conséquences fonctionnelles, mesurables en termes de perte ou de gain de fonction pour la protéine considérée.

Les variations phénotypiques à déterminisme multigénique, que l'on observe couramment au sein des populations, ne résultent pas d'altérations fonctionnelles aussi sévères des gènes impliqués. Elles sont le fait,

non de mutations délétères, mais de polymorphismes de l'ADN, qui modifient légèrement les propriétés des gènes, sans vraiment altérer leur fonction. La distinction entre l'effet phénotypique d'une mutation et l'effet d'un polymorphisme peut être objectivée de façon plus rigoureuse : pour un locus contribuant à l'expression d'un trait quantitatif, Risch a proposé de quantifier l'effet phénotypique associé à toute variation allélique (A,a) au moyen d'une variable appelée « déplacement », qui correspond à la différence – en déviations standards – pour la valeur moyenne du trait, entre individus portant les génotypes AA et aa (figure 3).

Seuil pathologique et voie biologique

L'accumulation de petits effets sur une voie biologique se traduit par une variation phénotypique quantitative, généralement non pathogène. Pour que ces petites variations continues parviennent à produire des différences phénotypiques qualitatives (des maladies), il faut de nouveau

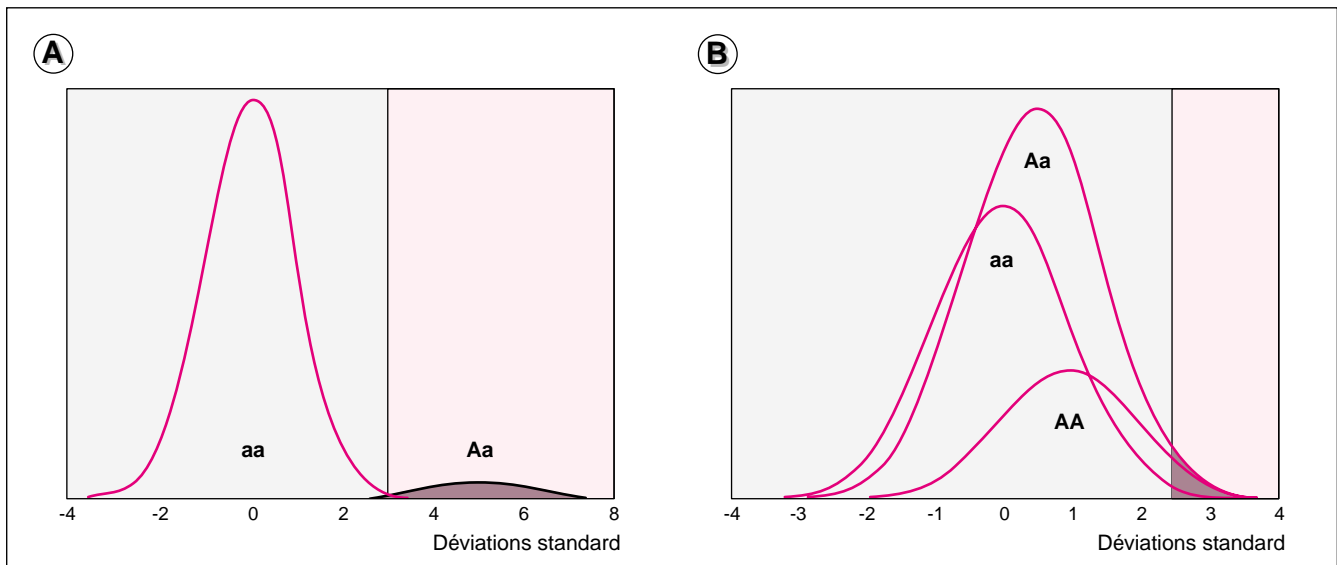


Figure 3. **Exemples d'hérédité mendélienne et non mendélienne.** **A.** Locus mendélien avec transmission dominante et fréquence de l'allèle pathologique A: $p = 0,00275$; déplacement $t = 5$ ds. La maladie survient après le seuil de 3 ds. Le risque morbide pour les hétérozygotes (Aa) est de 98%. Pour les homozygotes (aa), il est de 0,13%. Prévalence de la maladie: 0,67%. **B.** Locus non mendélien avec fréquence de l'allèle à risque A: $p = 0,40$ et déplacement $t = 0,5$ ds, pour chaque allèle A (déplacement total $t = 1$). La maladie survient au-dessus du seuil de 2,5 ds. Le risque morbide pour les homozygotes (AA) est 6,7%, pour les hétérozygotes (Aa), il est de 2,3%, et pour les homozygotes (aa), il est de 0,62%. Prévalence de la maladie, 2,4% (d'après [35]).

faire appel à la notion de seuil pathologique, à partir duquel le quantitatif se mue en qualitatif. Dans de nombreux cas, ce seuil n'est vraisemblablement atteint que lorsque plusieurs facteurs de risque situés sur une même voie biologique sont simultanément présents chez un individu. Ces polymorphismes interférant peu avec la capacité reproductive de leurs porteurs, ne sont pas contre-sélectionnés. Leur fréquence est souvent relativement élevée, ce qui permet la formation de combinaisons inter-alléliques « à risque » chez un nombre important d'individus. De telles combinaisons défavorables, interférant avec des facteurs environnementaux, finissent par exercer un effet pathologique qui, généralement, ne s'exprime qu'au cours du vieillissement. Dans ces maladies, le dépassement du seuil de susceptibilité est souvent dépendant de l'âge.

Maladies monogéniques et multigéniques: un continuum

Plus on connaît dans le détail les conséquences des variations de l'ADN, plus il s'avère qu'entre le polymorphisme « innocent » et la mutation délétère, il existe en fait un conti-

num de variations génétiques responsables de toute une gamme de phénotypes plus ou moins pathologiques. Même dans les maladies mendéliennes, des critères quantitatifs sont utilisés pour apprécier les relations phénotypes-génotypes. Ainsi, il est proposé de définir des « seuils critiques » au-delà desquels la fonction de la protéine mutante échappe à toute possibilité de contrôle par les autres variables biologiques avec lesquelles elle interagit [28]. Ces cas, relativement rares, sont ceux du mendélisme strict, où la présence de la mutation est la condition nécessaire et suffisante pour faire apparaître le trait morbide. En deçà de ce seuil, se profile la cohorte sans cesse croissante des maladies « monogéniques » à pénétrance incomplète et à expressivité variable. Dans ces cas, l'effet biologique résultant de la présence de protéines mutantes est partiellement sensible à l'action modulatrice d'autres facteurs génétiques et à celle de facteurs environnementaux. Plus le rôle de ces facteurs « modificateurs » s'affirme, plus on glisse doucement de la maladie « monogénique » à la maladie « multigénique », plus les traits « mendéliens » deviennent eux-mêmes des traits complexes [29].

Hétérogénéité génétique des maladies mendéliennes

Si les maladies mendéliennes (et les sous-entités mendéliennes, dans le cadre des maladies à hérédité complexe) sont généralement hétérogènes génétiquement, c'est parce qu'il existe de très nombreux gènes au sein d'une voie biologique, dont la mutation isolée est suffisante pour dérégler l'ensemble de cette voie. Toutes ces altérations, impliquant des gènes différents, seront finalement responsables de phénotypes morbides semblables. On dénombre, par exemple, plus d'une vingtaine de gènes responsables de rétinite pigmentaire, maladie caractérisée par une dégénérescence des cellules photoréceptrices de la rétine. Qu'elles soient autosomiques (dominantes ou récessives) ou liées à l'X, ces rétinites sont dues à des mutations de gènes codant pour des protéines impliquées dans la transduction du signal lumineux au niveau rétinien. Ces mutations perturbent donc une même voie biologique, ce qui explique qu'elles produisent une même maladie. Dans ces conditions, on peut penser que l'étude des sous-entités mendéliennes offre l'opportunité, non d'identifier des catégories particulières

de gènes qui établiraient de façon isolée un contrôle « monogénique » sur des traits biologiques complexes, mais de disséquer ces voies biologiques et d'en identifier un certain nombre de composants majeurs. Il faut cependant quelque peu nuancer cette conclusion, car l'expérience montre que des mutations drastiques affectant certains gènes ne sont pas viables, et ne produisent en conséquence aucun phénotype observable, alors que, dans d'autres cas, l'inactivation d'un gène est susceptible d'être complétée par l'effet d'autres gènes, et passe également inaperçue.

QTL ou mutagenèse ?

En dépit de ces réserves, il est probable qu'une telle stratégie, permettant la dissection génétique « de bas en haut » de traits quantitatifs constituera la méthode de choix pour préciser la structure d'ensemble de ces traits. Elle constitue une alternative à la recherche de locus impliqués dans de tels traits biologiques « de haut en bas », au moyen de la seule information apportée par les polymorphismes fonctionnels présents au niveau de ces divers locus. Une telle approche, fondée sur la recherche de QTL (*quantitative traits loci*) est actuellement la méthode de choix [28] pour caractériser la contribution de divers locus à la variation d'un trait quantitatif. En pratique, il s'avère que la localisation, par une telle approche, d'un locus responsable d'une part modeste (10 à 15 %) de la variance sur un trait continu est une tâche particulièrement ardue. D'où le recours à la mutagenèse plutôt qu'aux QTL pour identifier ces locus chez les mammifères [29].

L'inactivation de gènes chez l'animal

De fait, une aide importante à la dissection des voies biologiques impliquées dans les pathologies à hérédité complexe peut venir de mutagenèses réalisées chez l'animal. La puissance de ce type d'approche est illustrée par un exemple récent [30]. La narcolepsie fait partie de ces nombreuses maladies dont la prédisposition est influencée par HLA. Il est établi depuis près de 15 ans que 90 % des sujets atteints sont DQB1*0602 positifs, sans que la

raison d'une telle association soit connue. Par ailleurs, des études génétiques menées ces dernières années chez l'animal ont permis de montrer que l'inactivation de gènes codant pour un neuropeptide (Hcr, hypocretine), ou pour ses récepteurs, reproduisait les symptômes de la narcolepsie. Dans un cas très sévère de narcolepsie chez un sujet DQB1*0602 négatif, une mutation transmise de façon dominante a été observée sur le gène codant pour le récepteur Hcr-2. Ce récepteur appartient à un système de neurotransmission, qui concerne environ 20 000 neurones sur les 20 milliards du cerveau. Des études réalisées sur des cerveaux de patients atteints de forme sporadique de narcolepsie ont ensuite montré une réduction de 85 à 95 % du nombre de ces cellules, vraisemblablement du fait d'une réaction auto-immune très spécifique, sur un terrain HLA particulier. Ainsi, l'inactivation chez l'animal d'un gène impliqué dans une forme mendélienne rare peut éclairer du même coup la physiopathologie des formes courantes.

Gène mendélien et/ou facteur de risque

Dès lors qu'une voie biologique joue un rôle dans le déterminisme d'une maladie, il devient légitime de considérer les polymorphismes de l'ensemble des gènes constituant cette voie comme des « facteurs de risque » potentiels. Il est, de plus, très tentant de supposer qu'un même gène peut se comporter soit comme un gène mendélien, soit comme un facteur de risque pour la maladie, en fonction des variations géniques qu'il porte (mutations ou polymorphismes). Cette possibilité théorique est validée par l'exemple suivant : des mutations faux sens du gène *tau* (MAPT, *microtubule associated protein tau*), transmises de façon autosomique dominante avec une pénétrance quasi complète à 65 ans, sont responsables de certaines démences fronto-temporales. Ces mutations ont deux conséquences, elles diminuent la liaison de tau aux microtubules ou favorisent la production d'isoformes particulières, en interférant avec l'épissage. Outre ces mutations rares, il existe sur le gène *tau* un haplotype comprenant une douzaine de polymorphismes en DL

complet. Ces polymorphismes se retrouvent au niveau du promoteur, dans les séquences exoniques (mutations traductionnellement silencieuses) et dans les introns. L'haplotype majoritaire, présent à 80 % dans la population générale, est un facteur de risque pour la paralysie-supranucléaire-progressive (maladie neurodégénérative en général sporadique, atteignant le sujet âgé, associant des troubles de l'oculo-motricité avec un syndrome extra-pyramidal et évoluant vers une démence frontale), où sa fréquence dépasse 95 %. On ne connaît pas encore, au sein de l'haplotype considéré, le polymorphisme particulier jouant un rôle causal dans cette augmentation de risque, mais il est probable qu'il se trouve au niveau d'un régulateur d'épissage (pour revue, voir [31]).

Du complexe au simple : les endophénotypes

Dans certaines maladies à hérédité complexe, on ne retrouve pas de sous-entité mendélienne. En d'autres termes, il ne semble pas exister, au niveau des voies biologiques concernées, de gène dont la mutation isolée soit suffisante pour faire apparaître le phénotype morbide. Ce type de phénotype résulte vraisemblablement de variations multiples au sein d'architectures génétiques particulièrement « tamponnées », caractérisées par l'interaction de nombreuses voies biologiques.

Dans plusieurs maladies psychiatriques, dont la schizophrénie, on cherche donc à identifier des phénotypes intermédiaires (que l'on appelle des endophénotypes), obéissant à un déterminisme génétique plus lisible. Il s'agit de traits neurophysiologiques élémentaires, témoignant de dysfonctionnements neuroaux relativement simples, stables dans le temps, fortement associés à la maladie, mais que l'on peut retrouver également chez des apparentés sains, et qui sont susceptibles de jouer un rôle dans les désordres cognitifs prédisposant à la maladie. Ces traits ne sont qu'associés à la maladie. Par rapport à un tel trait élémentaire, présent chez un individu donné, l'apparition du phénotype schizophrénique est donc modulée par la présence de facteurs

additionnels, tels la présence concomitante d'autres endophénotypes est l'intervention de facteurs de milieu. Les endophénotypes ne sont rien d'autre que les pièces d'un puzzle qui, reconstitué, correspondra au « terrain biologique » sur lequel se développe la maladie.

Les traits biologiques auxquels se rapportent les endophénotypes sont généralement des traits quantitatifs, dont le déterminisme est le plus souvent multigénique. Ces traits se caractérisent par une variation continue au sein de la population. Il existe cependant, pour chacun d'eux, des individus dont le phénotype s'écarte franchement de la norme quantitative autour de laquelle fluctue la population générale. Si l'on étudie les apparentés de tels sujets, on constate que ce phénotype anormal semble transmis de façon mendélienne. Certaines variations génétiques concernant des gènes impliqués dans l'expression de ces traits paraissent donc perturber suffisamment les voies biologiques sous-jacentes pour faire apparaître, dans les familles considérées, un phénotype discret transmis selon les lois de Mendel. C'est précisément ce qui nous manquait lorsque nous nous intéressions au phénotype « schizophrénique ». Tirant parti de ce fait, il devient possible d'employer la puissance de la génétique mendélienne pour tenter d'identifier un à un les gènes impliqués dans ces voies biologiques.

Dissection génétique d'un comportement : l'exemple de la schizophrénie

Deux endophénotypes associés avec la schizophrénie ont été caractérisés ces dernières années, ce sont les troubles des mouvements de poursuite oculaire lents d'une part, les troubles du filtrage sensoriel d'autre part. Considérons ce dernier caractère : lorsqu'une personne normale est soumise consécutivement à deux stimulus auditifs identiques, l'activité neuronale provoquée par le second stimulus, que l'on mesure au niveau de l'amplitude de l'onde P50 des potentiels évoqués auditifs (P50 signifie positif 50 ms, il s'agit donc d'une onde cérébrale positive qui apparaît 50 ms après le stimulus) diminue au moins de moitié par rapport à celle

enregistrée lors du premier stimulus. Il y a habitude, un mécanisme biologique intervient pour modifier le traitement du second signal. Lorsqu'un dysfonctionnement se produit au niveau d'un tel mécanisme élémentaire de filtrage de l'information sensorielle, il est probable que les modules cognitifs traitant ces signaux à un niveau plus élevé sont inondés d'informations et se trouvent donc dans une situation plus difficile pour intégrer correctement ces messages. La symptomatologie psychotique pourrait, pour partie au moins, être liée à un tel traitement défectueux de l'information au niveau neuronal.

Chez environ 90 % des schizophrènes, la diminution d'amplitude de l'onde P50 lors du second stimulus est très faible ou ne s'observe pas du tout, alors qu'une telle anomalie ne concerne que 10 % à 15 % des individus de la population générale. Le fait que près de la moitié des apparentés au premier degré (schizophrènes ou non) d'un sujet présentant un tel trouble soient également porteurs de l'anomalie, suggère un mode de transmission autosomique dominant pour ce phénotype anormal, ou du moins un effet de gène majeur [32]. Toutefois, en l'absence d'analyse de ségrégation, ce point n'est pas encore établi avec certitude. On peut reproduire ce phénotype anormal par des manipulations pharmacologiques du système cholinergique (impliqué dans les mécanismes de focalisation de l'attention au niveau cortical). L'une des cibles moléculaires qui, lorsqu'elle est bloquée par un antagoniste spécifique, permet de mimer le phénotype, est la sous-unité $\alpha 7$ du récepteur nicotinique à l'acétylcholine. Ce récepteur est particulièrement exprimé au niveau des neurones thalamiques et des interneurons gabaergiques présents dans l'hippocampe. Il est donc envisageable que des variations de séquence au niveau du gène codant pour cette sous-unité du récepteur puissent être à la base du trouble électrophysiologique constaté chez les malades et leurs apparentés porteurs de l'endophénotype [33].

La ségrégation familiale de l'endophénotype semble compatible avec une transmission autosomique dominante. Toutefois, on ne peut exclure

qu'une telle homogénéité de mode de transmission cache une hétérogénéité génétique sous-jacente et que, donc, le gène impliqué ne soit pas identique dans toutes les familles dans lesquelles le trait ségrège. Quoi qu'il en soit, si l'implication de cette voie biologique se vérifie, la conclusion sera qu'un trouble des mécanismes attentionnels médiés par le système cholinergique, transmis de façon dominante au sein de certaines familles, représente un des facteurs génétiques prédisposant à la schizophrénie.

Conclusions

Dans la détection de facteurs de risque, l'emploi des méthodes traditionnelles s'avère d'un très bas rendement [34]. Le recours au déséquilibre de liaison avec criblage systématique de SNP, actuellement considéré comme la panacée, est probablement une voie sans avenir. En dépit de quelques mises en garde [35], l'engouement dénué de critique vis-à-vis de cette approche rappelle singulièrement celui qui régnait il y a 15 ans, lorsqu'on nous annonçait qu'aucune maladie à hérédité complexe ne résisterait à une étude de liaison au moyen de 300 marqueurs judicieusement placés sur le génome. Plutôt que de compter aveuglément sur les progrès des méthodes de génotypage à haut débit, la mise en œuvre de stratégies fondées sur la caractérisation préalable de « gènes candidats » est probablement la voie la plus prometteuse. Dans ce cadre, l'étude des sous-entités mendéliennes et des phénotypes intermédiaires (endophénotypes) représente une étape décisive, car elle permet de rassembler des connaissances sur les voies biologiques impliquées, ce qui, à terme, offrira l'opportunité d'un meilleur ciblage des gènes « candidats » au rôle de facteur de risque. Une fois ces gènes ciblés de façon rationnelle, les études d'association alléliques utilisant les SNP pourront être utilisées avec profit ■

Remerciements

L'auteur remercie F. Clerget-Darpoux pour ses commentaires sur ce manuscrit.

RÉFÉRENCES

1. Falconer D. The inheritance of liability to certain diseases. *Ann Hum Genet* 1965; 29: 51-76.
2. Lander ES, Schork NJ. Genetic dissection of complex traits. *Science* 1994; 265: 2037-46.
3. Clerget-Darpoux F, Bonaïti-Pellié C, Hochez J. Effects of misspecifying genetic parameters in lod score analysis. *Biometrics* 1986; 42: 393-9.
4. Risch N. Linkage strategies for genetically complex traits. *Am J Hum Genet* 1990; 46: 229-41.
5. Risch N, Merikangas K. Genetic analysis of complex diseases. *Science* 1997; 275: 1327-30.
6. Myers A, Holmans P, Marshall H, et al. Susceptibility locus for Alzheimer's disease on chromosome 10. *Science* 2000; 290: 2304-5.
7. Spielman RS, Ewens WJ. A sibship based test for linkage in the presence of association: the sib transmission/disequilibrium test. *Am J Hum Genet* 1998; 62: 450-8.
8. Feingold J. Le déséquilibre de liaison. *Med Sci* 1991; 7: 161-8.
9. Laan M, Pääbo S. Mapping genes by drift-generated linkage disequilibrium. *Am J Hum Genet* 1998; 63: 654-6.
10. Wright A, Carothers A, Pirastu M. Population choice in mapping genes for complex diseases. *Nat Genet* 1999; 23: 397-403.
11. Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 1999; 22: 139-44.
12. Pfaff C, Parra E, Bonilla C, et al. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *Am J Hum Genet* 2001; 68: 198-207.
13. Boehnke M. A look at linkage disequilibrium. *Nat Genet* 2000; 25: 246-7.
14. Schellenberg GD, Deeb SS, Boehnke M, et al. Association of an apolipoprotein CII allele with familial dementia of the Alzheimer type. *J Neurogenet* 1987; 4: 97-108.
15. Yu C, Payami H, Olson JM, et al. The apolipoprotein E/C1/CII gene cluster and late-onset Alzheimer disease. *Am J Hum Genet* 1994; 54: 631-42.
16. Strittmatter WJ, Roses AD. Apolipoprotein E and Alzheimer's disease. *Proc Natl Acad Sci USA* 1995; 92: 4725-7.
17. Pericak-Vance MA, Bebout JL, Gaskell PC, et al. Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. *Am J Hum Genet* 1991; 48: 1034-50.
18. Bickeboller H, Campion D, Brice A, et al. Apolipoprotein E and Alzheimer disease: genotype-specific risks by age and sex. *Am J Hum Genet* 1997; 60: 439-46.
19. Gray IC, Campbell DA, Spurr NK. Single nucleotide polymorphisms as tools in human genetics. *Hum Mol Genet* 2000; 9: 2403-8.
20. Cargill M, Altshuler D, Ireland J, et al. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 1999; 22: 231-8.
21. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996; 273: 1516-7.
22. Collins FS, Guyer MS, Chakravarti A. Variations on a theme: cataloging human DNA sequence variation. *Science* 1997; 278: 1580-1.
23. Jorde LB. Linkage disequilibrium as a gene mapping tool. *Am J Hum Genet* 1994; 52: 362-74.
24. Martin RE, Lai EH, Gilbert J, et al. SNPing away at complex diseases: analysis of single-nucleotide polymorphisms around APOE in Alzheimer's disease. *Am J Hum Genet* 2000; 67: 383-94.
25. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995; 11: 241-7.
26. Dipple KM, McCabe ERB. Phenotypes of patients with «simple» mendelian disorders are complex traits: thresholds, modifiers, and systems dynamics. *Am J Hum Genet* 2000; 66: 1729-35.
27. Estivill X. Complexity in a monogenic disease. *Nat Genet* 1996; 12: 348-50.
28. Lander E, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 1989; 121: 185-99.
29. Nadeau J, Frankel W. The roads from phenotypic variation to gene discovery: mutagenesis versus QTL. *Nat Genet* 2000; 25: 381-4.
30. Peyron C, Faraco J, Rogers W, et al. A mutation in a case of early onset narcolepsy and a generalized absence of hypocretin peptides in human narcoleptic brains. *Nat Med* 2000; 6: 991-7.
31. Heutink P. Untangling tau-related dementia. *Hum Mol Genet* 2000; 9: 979-86.
32. Siegel C, Waldo M, Mizner G, Adler L, Freedman R. Deficit in sensory gating in schizophrenic patients and their relatives. *Arch Gen Psychiatry* 1984; 41: 607-12.
33. Leonard S, Adams C, Breese CR, et al. Nicotinic receptor function in schizophrenia. *Schizophrenia Bull* 1996; 22: 431-45.
34. Risch N. Searching for genetic determinants in the new millennium. *Nature* 2000; 405: 847-56.
35. Weiss KM, Terwilliger JD. How many diseases does it take to map a gene with SNP? *Nat Genet* 2000; 26: 151-7.

Summary

Genetic dissection of complex diseases

Non mendelian diseases are characterized by complex correspondence between genotypes and phenotypes. Often, the phenotype is determined by the effect of several genes, combined or not with environmental effects. Each susceptibility gene, considered separately, is neither essential nor sufficient to produce the disease. Rather, liability to the disease is a quantitative variable, and the morbid phenotype appears when a liability threshold is reached by the combination of effects of several risk factors. Methods used to map susceptibility genes are reviewed. It is concluded that model dependent linkage studies (e.g. lod score method) are unsuited. Model free linkage methods (affected sib pairs) are more robust but are lacking power to detect risk factors with a small effect and do not allow a precise localisation. Single Nucleotide Polymorphisms (SNP) based association studies are likely to generate a very large amount of false positive results, given the number of tests realized in a whole genome scan. SNP based association studies are more likely to give firm results when applied to « candidate genes ». Use of specific strategies allowing to characterize « candidate genes » is thus strongly advocated. It is suggested that focusing on sub-mendelian entities and intermediate phenotypes (endophenotypes) are a necessary step to define the biological pathways in which susceptibility genes are likely to be found. This approach is illustrated with examples from neuropsychiatric diseases.

TIRÉS À PART

D. Campion.