

QUELLE VALEUR ACCORDER AUX PREDICTIONS DE RISQUES POUR LES MALADIES MULTIFACTORIELLES ?

Il y a trente ans, la révolution de la biologie moléculaire a permis l'identification progressive de centaines de gènes impliqués dans des maladies monogéniques, ouvrant la voie à leur diagnostic moléculaire, à la compréhension de leur physiopathologie et à la recherche de nouvelles thérapies. Cette démarche ouvrait aussi la voie au projet Génome Humain qui aboutit, en Avril 2003, à la publication d'une séquence complète de notre génome. Ces succès remarquables ont soulevé beaucoup d'espoirs. On a pensé, en particulier, que l'identification des facteurs génétiques impliqués dans la physio-pathologie des maladies dites multifactorielles, mettant en jeu génétique et environnement, permettrait de prédire à une personne ses risques à partir de l'analyse de son génome («médecine personnalisée» ou «médecine prédictive»).

Nous assistons présentement sur internet à une explosion du nombre de firmes commerciales, soutenues par des scientifiques et des organismes de renom, qui proposent des «tests génétiques en libre accès» et de la «médecine personnalisée» pour des maladies multifactorielles telles que le diabète, l'asthme, des maladies cardio-vasculaires, des maladies auto-immunes, des maladies psychiatriques ou neuro-dégénératives.

Or un certain nombre de concepts souvent mal compris, voire mal définis, mènent à une interprétation contestable des résultats fournis par ces compagnies. Dans le texte annexé il est précisé ce que sont les maladies multifactorielles et les limites de validité des outils développés pour analyser leur composante génétique. Les études d'association pangénomique (GWAS : Genome Wide Association Studies) sont aujourd'hui un des outils les plus utilisés pour l'analyse des maladies multifactorielles et c'est à partir de leurs résultats que les sites internet et de nombreuses publications proposent une évaluation des risques de développer ces maladies. Le problème que nous soulevons concerne l'interprétation des risques ainsi calculés.

Dans la plupart des cas, les causes des maladies multifactorielles et leur mode d'action restent encore inconnus. Face à cet inconnu, l'interprétation des résultats présuppose que les facteurs génétiques sont nombreux, avec des effets petits, indépendants, cumulatifs, et sans interaction avec des facteurs du milieu. Si cette hypothèse était vérifiée, on pourrait imaginer que les études pan-génomiques finiraient par répertorier tous les facteurs génétiques impliqués dans chaque maladie ce qui permettrait alors de caractériser les personnes en fonction de leur risque, comparé à celui d'une population de référence. Or pour la majorité des maladies multifactorielles, cette hypothèse n'est pas justifiée. En effet le risque de développer de telles maladies dépend certes de facteurs génétiques mais aussi de facteurs du milieu auxquels on peut être exposé ou non et qui peuvent être essentiels dans le déclenchement de la maladie.

Ainsi, pour la lèpre, les différences génétiques qui ont été décrites dans la prédisposition à cette maladie n'interviennent qu'en présence du bacille de Hansen. Autre exemple, une personne porteuse de tous les facteurs de risques génétiques associés au diabète de type 2 court en fait moins de risque d'être atteinte si elle est mince et sans antécédents familiaux qu'une personne sans aucun de ces facteurs génétiques mais obèse avec des antécédents familiaux. Quelle signification accorder à un calcul de risque sur la base d'études génomiques si les facteurs du milieu, et leur interaction avec les facteurs génétiques, sont encore mal connus, inconnus ou non pris en compte ?

En conclusion :

- Si les études pangénomiques apportent une contribution essentielle à la connaissance scientifique, l'utilisation exclusive de l'information qui en résulte est dénuée de sens en matière de prédiction de santé. Elle conduit à une perception erronée du risque encouru par les individus.
- Il est du devoir de la communauté scientifique de ne pas servir d'alibi en matière de prédictions individuelles de risque pour les maladies multifactorielles à partir de la seule information génomique.
- Il importe d'informer avec rigueur et clarté sur l'évolution des connaissances en génomique et de débattre sur les limites de leurs applications.

Signatures :

- **SFGH (Société Française de Génétique Humaine)**
- **FFGH (Fédération Française de Génétique Humaine)**
- **ACLF (Association des Cytogénéticiens de Langue Française)**
- **ANPGM (Association Nationale des Praticiens de Génétique Moléculaire)**
- **AFGCCG (Association Française de Génétique Clinique et de Conseil Génétique)**
- **AFCG (Association Française des Conseillers en Génétique)**
- **SOFFOET (Société française de foetopathologie)**
- **CPEGM (Collège des Praticiens et Enseignants de Génétique Médicale)**
- **SFG (Société Française de Génétique)**
- **ADELFA (Association des Epidémiologistes de Langue française)**

Annexe

Argumentaire scientifique du texte co-signé par les sociétés savantes et les associations professionnelles de génétique et de génétique humaine

1) LES MALADIES MULTIFACTORIELLES ET LES PETITS EFFETS INDEPENDANTS

La définition d'une maladie multifactorielle est tellement peu claire qu'on lui substitue d'autres qualificatifs comme «maladie complexe» ou «maladie commune». Si le terme «complexe» n'est pas faux, il n'est pas vraiment spécifique d'une maladie multifactorielle ; quant au mot « commun », il est inapproprié puisque la fréquence de ces maladies est extrêmement variable, certaines ayant des fréquences inférieures à 1‰ (autisme, sclérose en plaques...) et d'autres supérieures à 10% (diabète de type 2 chez les plus de 65 ans, maladie d'Alzheimer chez les plus de 80 ans).

Les analyses génétiques supposent, en général, qu'une maladie multifactorielle est due à des facteurs génétiques nombreux aux effets petits, indépendants et cumulatifs (la maladie s'exprimant au-delà d'un seuil) et à des effets du milieu. Ce modèle (dit polygénique) ainsi que le concept d'héritabilité ont été introduits par Fischer en 1918. Dans les espèces animales ou végétales, pour lesquelles on peut contrôler croisements et environnement, ce modèle est fort utile pour sélectionner sur certains traits quantitatifs. Ce modèle, présenté comme alternatif au modèle monogénique où la cause primordiale de la maladie est la mutation d'un gène, pose comme hypothèse que les facteurs génétiques n'interagissent ni entre eux, ni avec les facteurs d'environnement. Dans les maladies humaines, cette hypothèse est pourtant clairement fautive pour les maladies qui ne s'expriment que dans un environnement précis comme la lèpre ou la maladie cœliaque, et peu réaliste aussi pour les maladies multifactorielles pour lesquelles on connaît mal (voire pas du tout) les facteurs d'environnement. La remise en question du modèle polygénique pour une maladie multifactorielle amène à s'interroger sur les valeurs de risque et d'héritabilité qui lui sont attribuées

2) MALADIES MULTIFACTORIELLES ET HERITABILITE

L'héritabilité d'un trait mesure, sous un modèle donné, la contribution de la variabilité génétique à la variabilité totale du trait au sein d'une population. Son estimation suppose, dans la population étudiée, une absence d'interaction et de corrélation entre facteurs génétiques et environnementaux.

Idée n°1: l'héritabilité d'une maladie n'a de validité que si les hypothèses faites sont correctes.

Pour les maladies multifactorielles, on ne connaît ni le nombre de facteurs impliqués, ni l'importance de leur effet individuel ni comment interagissent les facteurs génétiques et les facteurs d'environnement.

Idée n°2: La mesure d'une héritabilité, à supposer qu'elle soit légitime, n'est pas une mesure de l'importance comparée des facteurs génétiques et des facteurs d'environnement dans la physiopathologie de la maladie. La confusion entre contribution des variations et contribution des facteurs dans l'interprétation de l'héritabilité était dénoncée bien avant les études génomiques (voir Lewontin, 1974).

Une forte héritabilité n'implique pas que les facteurs génétiques sont importants et que l'environnement ne joue qu'un rôle réduit dans la physiopathologie de la maladie. Elle peut en effet signifier qu'il y a peu de variabilité, pour la population concernée, dans l'exposition aux facteurs d'environnement impliqués dans la maladie. Lorsque les membres d'un village sont tous exposés au bacille de la lèpre, la différence entre ceux qui développeront la lèpre et ceux qui ne la développeront pas va s'expliquer par des différences génétiques. Il n'en reste pas moins que l'élément déterminant est la présence du bacille. De la même façon pour la maladie cœliaque (ou intolérance au gluten), tout le monde est exposé au gluten et seule une petite proportion de la population est atteinte. Les facteurs de risque génétique apparaissent ici comme essentiels et pourtant le gluten est un élément causal déterminant puisque sa suppression dans l'alimentation suffit à guérir un malade.

Les utilisations et interprétations erronées de l'héritabilité sont patentées sur le site internet de certaines compagnies proposant des tests génétiques. Ainsi peut-on lire sur le site de 23andme que les facteurs du milieu jouent un rôle plus important que les facteurs génétiques dans le diabète de type 2 car son héritabilité n'est que de 26%. De la même manière, on y affirme de façon non fondée, que les risques de maladie cœliaque ou de dégénérescence maculaire liée à l'âge sont majoritairement attribuables aux facteurs génétiques.

Idée n°3: les limites apportées à la notion d'héritabilité et à son interprétation valent également pour celle, très à la mode, «d'héritabilité manquante» (Maher et al, Nature, 2009). Selon les promoteurs de ce concept, l'héritabilité manquante permettrait de montrer le «chemin qui reste à parcourir» pour identifier l'ensemble des facteurs génétiques de risque. Ainsi, un article récent de Nature (Manolio et al 2009) rapporte dans un tableau la proportion d'héritabilité expliquée pour plusieurs traits ou maladies. Par exemple, 5 facteurs de risque génétique associés à la dégénérescence maculaire liée à l'âge (DMLA) expliqueraient 50% de son héritabilité et les auteurs de conclure qu'il faut poursuivre le travail pour identifier les 50% de facteurs de risque génétique restants, alors que l'information manquante n'est pas mesurable.

3) MALADIES MULTIFACTORIELLES ET CONCENTRATION FAMILIALE,

On observe pour les maladies multifactorielles un risque plus élevé pour les apparentés d'un malade que pour quelqu'un pris au hasard de la population générale. Le rapport entre le risque d'être atteint pour le germain (frère ou sœur) d'un malade et le risque d'être atteint pour une personne prise au hasard dans la population générale (risque relatif noté λ_s) est souvent utilisé comme argument en faveur de l'existence de facteurs génétiques et même pour en mesurer l'importance.

- Idée n°4: Le risque relatif λ_s est très variable d'une maladie à l'autre. Par exemple, il est estimé entre 20 et 40 pour la sclérose en plaques (Ebers et al, 1995) et à environ 1.8 pour le diabète de type 2 (Weijnen et al, 2002). A un très grand risque relatif λ_s peut correspondre un petit risque d'être atteint (et vice-versa). Dans le cas de la sclérose en plaques, le risque relatif d'un germain est élevé (20 à 40) mais son risque d'être atteint n'est que de 2 à 3%, car la fréquence de la maladie dans la population générale est entre 1/1000 et 1/2000. Inversement, dans le diabète de type 2, le risque relatif d'un germain est faible (seulement 1,8) mais son risque d'être atteint à 55 ans est de 23% car il est déjà en population générale égal à 13% à cet âge-là.
- Idée n°5: un risque relatif élevé ne traduit pas nécessairement des effets génétiques importants dans la physiopathologie. Une partie de l'augmentation du risque pour un apparenté, voire la totalité, peut s'expliquer par des facteurs d'environnement partagés par les membres d'une même famille. L'augmentation de l'obésité observée pour les frères et sœurs d'un enfant obèse est peut être due au partage de facteurs génétiques, mais très certainement aussi au fait qu'ils partagent les mêmes habitudes alimentaires.

4) MALADIES MULTIFACTORIELLES ET CALCUL DE RISQUE

Les études d'association pangénomique (GWAS : genome wide association studies) sont aujourd'hui un des outils les plus utilisés pour l'analyse des maladies multifactorielles ; c'est à partir de leurs résultats que les sites internet et de nombreuses publications calculent les risques de développer ces maladies.

Le génome humain varie d'un individu à un autre en un grand nombre de points (plusieurs dizaines de milliers par chromosome), variations également appelées polymorphismes, parmi lesquels des SNP (Single Nucleotide Polymorphisms) qui se présentent sous deux formes possibles qu'on désignera par A et a.

Les études d'association cherchent à déterminer les SNPs pour lesquels les fréquences de A et a sont différentes chez les malades et les non malades (témoins) ; l'importance de cette différence est souvent mesurée par le rapport entre le risque d'être atteint quand on est porteur d'une des formes et le risque d'être atteint quand on n'en est pas porteur (Odd-Ratio ou « OR »). Un SNP, qui montre une différence significative (ou un OR $\neq 1$), signale qu'il existe dans son voisinage génomique un ou plusieurs facteur(s) génétique(s) impliqué(s) dans le processus physio-pathologique.

- Idée n°6: L' OR est un signal qui ne rend compte qu'imparfaitement de la complexité des facteurs génétiques avoisinants, impliqués dans le processus physio-pathologique. Autrement dit, la valeur de l'OR peut être très supérieure si elle est calculée directement sur les facteurs génétiques jouant un rôle dans la maladie. Ceci est bien illustré dans la maladie coeliaque pour laquelle la valeur d' OR du SNP le plus fortement associé dans la région HLA est de 7 (Van Heel et al, 2007) alors qu'il est de 25 lorsqu'on prend en compte simultanément le polymorphisme de deux gènes de la région HLA (Margaritte-Jeannin et al, 2004)

D'une façon plus générale, un long chemin reste à parcourir pour aller de la détection d'un signal donné par un OR significatif aux variations génétiques réellement impliquées dans le processus pathologique. A ce sujet il convient de rappeler le peu d'avancées dans la compréhension des mécanismes sous-tendant les associations observées depuis plus de 35 ans entre des antigènes HLA et un grand nombre de maladies (diabète de type 1, sclérose en plaques ...).

- Idée n°7: Les calculs portant sur tout un ensemble de SNPs associés à une maladie sont fondés à nouveau sur l'hypothèse qu'il n'y a pas d'interactions entre les facteurs génétiques. Or celles-ci existent évidemment dans les processus complexes qui sous-tendent par exemple les maladies auto-immunes ou du métabolisme.

- Idée n°8: A un OR élevé peut correspondre un risque faible d'être atteint

Par exemple, si on considère la maladie cœliaque en Europe, à un OR de 25 pour les porteurs des polymorphismes fonctionnels du système HLA correspond un risque d'être atteint de seulement 1%. En effet le risque est proportionnel à la fréquence de la maladie -environ 1/300 en Europe- mais aussi inversement proportionnel à la fréquence des polymorphismes HLA impliqués laquelle est élevée (environ 25% de porteurs dans la population européenne).

Idée n°9: hétérogénéité étiologique des maladies dites multifactorielles

Les études d'association pangénomiques ont permis de mettre en évidence, pour un bon nombre de maladies multifactorielles, des signaux de faible amplitude (faible valeur d'OR pour les SNPs associés)

grâce à des échantillons de malades et de témoins de tailles impressionnantes (plusieurs milliers). Ceci implique qu'on impose très peu de contrainte dans l'échantillonnage en termes d'homogénéité clinique et environnementale. En conséquence, il est très vraisemblable que l'hétérogénéité étiologique y soit très forte. Ainsi l'analyse de 14.586 cas de diabète de type 2 a très certainement porté à la fois sur des malades n'ayant aucun problème pondéral et sur des malades en surpoids. L'analyse de Weijmen et al (diabetes, 2002) suggère pourtant qu'ils pourraient correspondre à des étiologies différentes. De la même façon, les études pangénomiques portant sur des cancers mélangent vraisemblablement des formes monogéniques et multifactorielles.

- Idée n° 10: importance de l'information manquante

L'information fournie par les SNPs associées dans les études pangénomiques peut être très pauvre au regard d'autres types d'information (information familiale, information environnementale...) (Bourgain et al, 2007, Clerget-Darpoux & Elston, 2007).

Contrairement à ce que laisse penser ceux qui mesurent « une héritabilité manquante », il n'existe aucun moyen de mesurer l'importance et la nature de l'information manquante. Un risque peut être totalement changé par l'apport d'une information nouvelle. Une personne ayant les formes les plus à risque pour les SNPs associés au diabète de type 2 a en fait moins de risque de développer cette maladie si elle est mince et sans aucun antécédent familial qu'un individu n'ayant pas les formes à risque de ces SNPs, mais obèse ou avec des antécédents familiaux. De la même façon, sur la base de l'analyse de son génome, quelqu'un peut être déclaré à risque élevé d'être atteint de la lèpre alors qu'il ne court aucun risque s'il n'est pas exposé au bacille.

Ce texte est le résultat du travail de réflexion et de synthèse d'un comité mis en place par la SFGH (Société française de génétique humaine) comprenant :

- Alain Bernheim (directeur de recherche INSERM, IGR Villejuif, vice président de la SFGH)
- Catherine Bourgain (chargée de recherche INSERM)
- Anne Cambon-Thomsen (directrice de recherche CNRS, Toulouse)
- Françoise Clerget (directeur de recherche, INSERM)
- Pierre Darlu (directeur de recherche CNRS)
- Marc Fellous (professeur émérite, université Paris VII-Diderot)
- Josué Feingold (directeur de recherche émérite INSERM)
- Gérard Huber (psychanalyste, écrivain, président du club « santé-solidarité_prospective 2100 »)
- Jean-Claude Kaplan (professeur émérite, université Paris V-Descartes)
- Jean-Louis Serre (professeur, université de Versailles, président de la SFGH)

- **Références:**

1. **Fisher:** The correlation between relatives on the supposition of mendelian inheritance. Transactions of the Royal Society of Edinburg, 1918, 52, 399-433.
2. **Lewontin:** The analysis of variance and the analysis of causes Am J Hum Genet, 26,400-401
3. **Bomprezzi R, Kovanen PE Martin R:** New approaches to investigating heterogeneity in complex traits. J Med Genet 2003;40 553-559
4. **Maher B:** Personal genomes : the case of the missing heritability. Nature 456, 18-21 (2008)
5. **Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM:** Finding the missing heritability of complex diseases. Nature. 2009 Oct 8;461(7265):747-53
6. **Ebers, G. C., Sadovnick, A. D. & Risch, N. (1995)** A genetic basis for familial aggregation in multiple sclerosis. Nature 377:150-151.
7. **Weijnen CF, Rich SS, Meigs JH, Krolewski AS, Warram JH:** Risk of diabetes in siblings of index cases with type 2 diabetes : implications for genetic studies. Diabetes, 2002, 19:41-50.
8. **DA van Heel, L Franke, K A Hunt, R Gwilliam, A Zernakova, M Inouye, M C Wapenaar, M Barnardo, G Bethel, G Holmes, C Feighery, D Jewell, D Kelleher, P Kumar, S Travis, J Walters, D S Sanders, P Howdle, J Swift, R J Playford, W M McLaren, M L Mearin, C J Mulder, R McManus, R McGinnis, L R Cardon, P Deloukas, and C Wijmenga :** A genome-wide association study for celiac disease identifies risk variants in the region harboring *IL2* and *IL21*. *Nature Genetics* 39, 827 - 829 (2007)
9. **P Margaritte-Jeannin, M-C Babron, M Bourgey, A Louka, F Clot, S Pecopo, I Coto, J-P Hugot, H Ascher, L Sollid, L Greco, F Clerget-Darpoux :** A study of the European Genetic cluster on Coeliac disease Tissue antigens, 2004, 63, 562-567
10. **Bourgain C, Genin E, Cox N, Clerget-Darpoux F:** Are genome-wide association studies all that we need to dissect the genetic component of complex human diseases ? Eur J Hum Genet, 2007,15(3): 260-263.
11. **Clerget-Darpoux F, Elston RC:** Are linkage analysis and the collection of family data dead? Prospects for family studies in the age of genome-wide association. Hum Hered 2007, 64: 91-96.