

Génétique des Maladies Multifactorielles

Etat des lieux et enjeux

Toutes les maladies fréquentes (cancers, maladies cardiovasculaires, diabète, obésité, asthme, allergies, maladies inflammatoires, autoimmunes et infectieuses, maladies psychiatriques, etc...) ont une étiologie multifactorielle, c'est-à-dire qui implique à la fois des facteurs génétiques et de l'environnement, à des degrés variables, et des interactions entre ces facteurs. Ces maladies représentent un fardeau important en terme de santé publique (« health burden »), avec un impact socioéconomique important. Avec le vieillissement de la population, du fait de l'amélioration générale des diagnostics et traitements et de la réduction ou du contrôle de certains facteurs de risque environnementaux associés à une mort prématurée (principalement contrôle des infections), on s'attend à une augmentation de la charge de morbidité de ces maladies. La contribution des facteurs génétiques est fortement imbriquée avec celle d'autres facteurs, qui impliquent l'individu, l'environnement et la société, facteurs hautement variables mais aussi les plus modifiables, avec un impact important sur la prévalence des maladies, leur prévention et leur traitement. Avec la modification rapide actuelle des facteurs d'« environnement » (dans son concept large de facteurs non génétiques), le présent document, qui est centré sur l'aspect génétique, doit être considéré dans un contexte beaucoup plus général, incluant les modifications médicales et pharmacologiques, mais aussi environnementales et sociétales actuelles.

Le modèle généralement accepté et considéré quant au déterminisme génétique des maladies multifactorielles est un modèle polygénique impliquant un grand nombre de variants génétiques, chacun contribuant à une faible modulation du risque de la maladie (modèle : variant fréquent, risque génétique mineur). C'est sur la base de ce modèle qu'a été réalisé le projet international HapMap et les études d'association pan-génomiques utilisant des puces de SNPs, qui ont été mises en œuvre depuis 2006 pour de nombreuses pathologies (<http://www.genome.gov/gwastudies/>). Ces études nécessitent des populations d'étude de grande taille qui peuvent être obtenus notamment dans le cadre de consortiums internationaux et mobilisent des ressources de génotypage et de séquençage à grande échelle, ainsi que des moyens bioinformatiques et biostatistiques importants pour analyser les volumes de plus en plus importants de données. Ces études ont déjà fait la preuve de leur puissance, avec l'identification d'un nombre grandissant de variants génétiques associés aux maladies, qui permettent d'identifier de nouveaux gènes et des voies métaboliques et physio-pathologiques d'intérêt. Néanmoins, ces études ont des limites puisque les variants génétiques fréquents caractérisés à ce jour n'expliquent qu'une certaine part de la composante génétique des pathologies multifactorielles. Ces études permettent d'identifier (initialement) les SNPs associés aux maladies, et nécessitent d'être couplées à d'autres études (post-génomiques) pour identifier les gènes en cause et les mécanismes fonctionnels. Le risque morbide associé avec la plupart de ces variants génétiques est le plus souvent faible, ce qui ne permet pas leur utilisation en médecine prédictive. L'hypothèse alternative de variants rares conférant un risque élevé, n'est pas correctement testée par les puces de SNPs utilisées jusqu'à présent, et nécessite de nouvelles études spécifiques, en partie basées sur des approches de séquençage. En plus des variants rares associés à un risque génétique élevé, des variants de structure du génome (« copy number variants »), et des mécanismes épigénétiques pourraient également jouer un rôle important, et ont été peu explorés

jusqu'à présent. Ces considérations tendent à réduire la distance entre maladies ou traits purement monogéniques et maladies ou traits multifactoriels. Ces questions nouvelles laissent envisager une évolution rapide dans la nature et l'ordre de grandeur des explorations génétiques et génomiques, et auront des implications importantes quant aux stratégies à mettre en œuvre et à encourager pour les recherches en génétique sur ces maladies, ainsi que sur leurs applications potentielles.

Il y a actuellement des efforts importants pour des développements technologiques et en ressources, qui ont un impact direct sur les études des maladies multifactorielles (et les autres) ; ces évolutions sont constantes et progressent très rapidement, et nécessitent une grande réactivité. Ainsi, des nouvelles ressources sont disponibles, comme le « 1000 Genome Project » (<http://www.1000genomes.org/>), et des nouvelles technologies de séquençage sont développées, pour permettre d'identifier l'ensemble du spectre de variabilité génétique. Plusieurs technologies de séquençage pourront être utilisées, en fonction de la nature des études (questions biologiques) et leur échelle, leur performance n'étant pas un paramètre absolu, mais bien conditionné au problème. A ces technologies mises en œuvre à l'échelle du génome sont associées des études d'expression des gènes (transcriptomique), d'épigénomique, de protéomique, métabolomique et autres approches « omiques » pour étudier les aspects fonctionnels. L'intégration de ces approches couplées aux développements méthodologiques en bioinformatique et biostatistique permet de définir des systèmes biologiques complexes. En plus de ces études à grande échelle, les études ciblées sur des formes monogéniques de maladies multifactorielles, l'étude de modèles animaux ou cellulaires de gènes critiques, ainsi que d'autres études permettant de tester des hypothèses particulières (mode d'action des gènes, interactions gènes-environnement...) présentent un intérêt grandissant pour la compréhension des mécanismes physiopathologiques impliqués dans les maladies fréquentes.

Enjeux scientifiques importants

- *Impact de ces études* : Que devons nous attendre des études génétiques des maladies multifactorielles ? Les études d'association pan-génomiques constituent une première étape de découverte de variants génétiques associés aux maladies mais elles ne sont pas efficaces dans toutes les situations (en particulier : cas des variants rares), et doivent obligatoirement s'intégrer dans un ensemble plus large d'études à différents niveaux pour aboutir à l'identification des gènes en cause, leur validation fonctionnelle, la compréhension des mécanismes biologiques impliqués dans les processus normaux et pathologiques en rapport avec ces maladies, et des relations entre ces mécanismes et facteurs de l'environnement. Ainsi, par leur capacité à identifier des nouveaux locus, ces études génèrent un besoin accru dans toutes les approches à finalité fonctionnelle, tant systématiques et à l'échelle du génome, que ciblées sur des gènes et voies métaboliques particuliers. Les compétences en biologie, biochimie, physiologie, épidémiologie, bioinformatique et biostatistique, et leur intégration, sont indispensables pour exploiter au mieux les données de la génomique.
- *Approches et stratégies*: Mettre en œuvre différents types de stratégies pour l'étude des maladies multifactorielles, en particulier, favoriser la collection de nombreuses données phénotypiques de plus en plus fines et proches des mécanismes biologiques impliqués, de données sur les facteurs de l'environnement susceptibles d'interagir avec les gènes et aussi

renforcer les études sur la compréhension des mécanismes biologiques et sur les applications potentielles des découvertes génétiques (notion de médecine translationnelle), et toute stratégie qui pourrait couvrir des questions ou hypothèses qui ne sont pas bien couvertes par les approches actuelles.

- *Sélection des maladies et des traits à étudier plus particulièrement*: Faut-il favoriser certaines pathologies, et sur quels critères ? Définir des priorités ? Les priorités doivent être définies de manière concertée avec les autres instituts thématiques. L'impact potentiel de ces études en terme de santé publique pourrait intervenir dans ces choix, mais il reste souvent difficile à évaluer aujourd'hui. Si le potentiel prédictif des SNPs fréquents associés aux maladies multifactorielles, associés à un accroissement mineur de risque (qui sont la quasi-totalité de ceux qui ont été identifiés jusqu'à présent) est très limité et non utilisable pour cette fin, il faut souligner que cela pourrait être différent dans d'autres situations. En particulier, des études récentes indiqueraient que la réponse aux traitements ou à des facteurs d'environnement particuliers, seraient dans certains cas sous un contrôle génétique plus direct que la prédisposition aux maladies multifactorielles. Si ces observations étaient confirmées, cela pourrait avoir des implications importantes en vue d'interventions prédictives et thérapeutiques.
- *Multidisciplinarité* : Nécessité d'élargir le concept d'études génétiques, avec des modèles qui prennent en compte les autres facteurs impliqués (individus, environnement, société, modifications épigénétiques), et leur variabilité et évolutions récentes, et d'évaluer les conséquences en termes de santé publique et des mesures à prendre au niveau de la société.
- *Composantes éthiques et en santé publique* : Nécessité d'évaluer les bénéfices, mais aussi les risques potentiels, des applications des découvertes génétiques ; Nécessité de s'associer aux experts d'autres disciplines (écologie, pharmacologie, sociologie...) pour mesurer, prévoir et envisager les applications directes en terme de santé publique et de prise en charge médicale.

Enjeux technologiques et organisationnels

Les études génétiques des maladies multifactorielles sont très demandeuses technologiquement et pour leur gestion et traitement informatique et statistique, de par leur échelle (très grand nombre d'individus à étudier, pour le criblage et pour la confirmation des résultats; nécessité d'études génome entier dans la plupart des cas), et de par le type de technologie utilisée : génotypage, séquençage, sous-phénotypage et études basées sur la fonction des gènes telles que transcriptomique, protéomique, métabolomique, imagerie et explorations physiologiques et métaboliques. Ces technologies évoluent très rapidement et sont demandeuses en ressources humaines ayant des compétences d'un haut niveau de spécialisation aussi bien dans les technologies génomiques et post-génomiques qu'en bioinformatique, biostatistique et analyses de données. L'implantation de ces technologies et de ces compétences doit être maintenue de manière adaptée à la compétition internationale. Il faut aussi garder un équilibre entre le besoin de développer et d'installer les technologies les plus efficaces (hauts débits, technologies innovantes, développements technologiques de pointe) généralement les plus appropriées aux études à très grande échelle et le besoin de maintenir des études de plus petite échelle, qui peuvent être d'accès plus facile, pour explorer des mécanismes spécifiques et tester des hypothèses ciblées.

Points principaux :

- Rôle critique des plateformes nationales à haut débit (puces de SNPs, séquençage, épigénomique...) pour les projets à grande échelle et la capacité de développement technologique optimale, mais également des plateformes régionales pour mettre en oeuvre les projets ciblés de plus petite taille.
- Rôle majeur de la bioinformatique et de la biostatistique, besoins en moyens et en compétences, pour faire face à l'ampleur des données générées et pour développer des méthodologies statistiques et des modèles mathématiques capables d'intégrer différentes sources d'informations et de modéliser les mécanismes étiologiques complexes, ainsi que des méthodes d'interfaçage permettant de diffuser et partager les résultats, y compris avec les autres disciplines.
- Reconnaissance de l'importance de la dimension interdisciplinaire et de la diversité des approches, stratégies, méthodologies et technologies impliquées.

Analyse stratégique

- Forces :
 - o Le nombre, la diversité et la qualité des équipes, qui ont recruté ou ont la capacité de recruter de grandes collections de données avec une très bonne qualité des données médicales et phénotypiques individuelles, sur l'environnement et le suivi des patients.
 - o La présence de plateformes nationales à haut débit, de grande capacité et de grande qualité technologique, analytique et organisationnelle dans la génération et la gestion des données ; la présence d'un grand nombre de plateformes (de diverses spécificités) de moyenne échelle dans de nombreux centres de recherche.
 - o La présence d'équipes en génétique statistique reconnues au plan international mais en nombre insuffisant.
- Faiblesses :
 - o L'organisation n'est pas optimale : difficultés de maintenir des engagements, financements et personnels à long terme (précarité des contrats de recherche et des personnels, CDDs non renouvelables...), qui affectent l'efficacité générale des plateformes et des laboratoires.
 - o L'accès aux plateformes nationales est davantage orienté vers les projets à grande échelle ce qui pose le problème de la réalisation des projets à plus petite échelle ou atypiques. Ceci est à débattre avec le GIS IBISA qui coordonne les appels à proposition pour les travaux de génotypage et de séquençage à grande échelle et qui assure le financement des plateformes de génotypage et de séquençage à petite échelle.
 - o Décisions stratégiques insuffisamment basées sur le long terme et financement très insuffisant en France comparativement au niveau des autres pays (Royaume Uni, Japon, Etats-Unis).

- Goulots d'étranglement :
 - o Le coût important des études et la nécessité d'accès à des plateformes à haut débit dans des temps raisonnables et à plusieurs étapes des projets est un paramètre critique. Il est indispensable de trouver un mécanisme permettant de faciliter l'accès à ces plateformes, ainsi que le retour rapide en production.
 - o Le maintien de plateformes à haut débit, capables de se situer au sommet de la technologie génomique, est indispensable.
 - o Besoin de transferts technologiques et de formations au sein de plateformes de taille plus modeste et des laboratoires, pour l'utilisation et l'exploitation des données générées, ainsi que pour l'utilisation générale et l'intégration des données de la génomique, y compris par d'autres disciplines.
 - o Nécessité d'une meilleure coordination entre équipes pour la constitution de collections biologiques et de données de grande taille.
 - o Le manque de bioinformaticiens et biostatisticiens est actuellement critique en regard des données générées, et nécessite un effort particulier pour développer et maintenir les ressources humaines en bioinformatique et biostatistique.
 - o Applications en santé publique : en fonction de l'impact réel des avancées génétiques pour des applications diagnostiques et médicales, il faudra envisager rapidement les conséquences et mises en application, en concertation avec les autres experts (médecins, pharmaciens, écologistes, sociologues...) pour définir des stratégies importantes au niveau de la société. Il y a un risque d'une mauvaise compréhension du grand public (et des politiques) sur les conséquences des études génétiques, et leur traduction en terme de prédiction, prévention et thérapie, ainsi que des risques de dérives dans ces applications, qu'il est nécessaire d'anticiper. De telles dérives mercantiles existent déjà par exemple aux Etats-Unis (sociétés de génotypage proposant un diagnostic de risque de maladies multifactorielles, basées sur des SNPs fréquents, dont on connaît leur incapacité prédictive). Nous ne sommes pas à l'abri de telles dérives.

Priorités et actions proposées

- Organisation des plateformes de génomique en France : échelle, capacité, méthodes d'accès, financement des projets et des plateformes, action de formation et transferts technologiques. Cela concerne les plateformes de grande et moyenne échelle. Des discussions sont déjà en cours à ce sujet (GIS IBISA).
 - o Action possible: Les plateformes nationales se placent à la pointe technologique et en analyse, et sont à même de se maintenir à ce niveau d'excellence, sous réserve de financement adéquat. Elles sont actuellement davantage orientées vers les projets à très grande échelle et leur financement doit être en adéquation avec les niveaux internationaux pour maintenir une recherche de pointe au meilleur niveau mondial. Nous proposons que des plateformes régionales de plus petite taille puissent gérer des projets plus spécifiques en s'appuyant sur l'expertise des plateformes nationales. L'accès

à ces plateformes régionales se ferait sur appel d'offres et pourrait nécessiter une implication financière (au moins partielle) et en personnel des équipes demandeuses.

- Pathologies et traits à étudier, et priorités
 - o Les études génétiques des maladies multifactorielles doivent se faire en lien avec les autres instituts thématiques pour développer des collections de grande taille lorsque c'est nécessaire, et créer des réseaux d'équipes multidisciplinaires depuis le recueil jusqu'à l'analyse des données.
 - o Au vu de l'impact génétique fort qui commence à être mis en évidence sur certaines réponses à des médicaments ou facteurs environnementaux, et de l'augmentation rapide et récente de nombreuses maladies multifactorielles du fait de changements de facteurs de l'environnement, il apparaît important de susciter des études explorant la réponse aux facteurs d'environnement et/ou pharmacologiques, qui peuvent avoir des retombées en santé publique.

- Encourager les études après-gène, incluant des études à grande échelle et des études plus ciblées sur les gènes et voies physio-pathologiques :
 - o Etudes fonctionnelles basées sur des études grande échelle : transcriptome, protéome, épigénome, métabolome...
 - o Etudes fonctionnelles spécifiques visant à la compréhension moléculaire, médicale et conséquences cliniques, au moyen d'études directes chez l'homme, et dans des modèles humains (maladies monogéniques), animaux et cellulaires. La qualité des explorations cliniques et fonctionnelles est le premier facteur limitant dans les études des maladies multifactorielles

- Accroître substantiellement les compétences et les ressources en bioinformatique et biostatistique pour répondre à la quantité massive de données générées et pour les intégrer, et ainsi modéliser et étudier les systèmes biologiques complexes.
 - o Développer de manière importante la formation des jeunes à ces disciplines associant génétique, informatique et statistique par l'enseignement au niveau académique (universités, écoles d'ingénieurs...), en ciblant des bourses doctorales et post-doctorales dans ces disciplines de niveau acceptable pour des étrangers, et en favorisant l'organisation d'ateliers de niveau international.
 - o Ouvrir des appels d'offres (ANR et autres) spécifiques pour la génétique statistique et la bioinformatique avec la nécessité d'une pérennité de ces appels d'offre.
 - o Disposer des moyens nécessaires pour développer les réseaux entre équipes dans ces disciplines et faciliter le transfert des compétences.

- Encourager les études proposant des approches originales (éventuellement à risque) de maladies multifactorielles ou visant à leur compréhension, ou qui explorent des aspects qui ont été insuffisamment étudiés : études combinées génétiques/environnement, études de l'impact de modifications épigénétiques, études de sous-groupes sélectionnés de patients, études dans

d'autres populations... Il est nécessaire d'encourager la diversification des approches et l'exploration de concepts originaux et novateurs, même quand ils ne correspondent pas aux dogmes ou pratiques habituelles.

- Evaluer les conséquences des retombées attendues de ces recherches, et les mettre en œuvre dans un cadre contrôlé:
 - o Traduction des découvertes génétiques sur le plan de la santé publique : mise à disposition de nouvelles méthodes diagnostiques issues de la biologie à grande échelle, stratégies de prévention, traitement, information du public, des médecins, sociologues, grand public...; les considérations éthiques sont extrêmement critiques, pour éviter une mauvaise information, ou mauvaise utilisation des informations, et devront être débattues en anticipation avec les différents experts impliqués, en particulier avec l'Institut de Santé Publique, afin de définir des stratégies d'intervention appropriées et d'éviter des applications abusives.
 - o Les possibilités d'intervention pour le diagnostic, la prévention et le traitement seront évaluées dans le cadre d'essais cliniques pour évaluer le pronostic de la maladie et la réponse au traitement en fonction du profil génétique des sujets, ainsi que les aspects éthiques et sociétaux.