

### Études d'interaction gène-environnement : défis et opportunités

Fruit d'un atelier de travail du 64<sup>e</sup> Congrès annuel de l'*American Society of Human Genetics*, cet article\* expose les difficultés de la recherche dans le domaine très prometteur des interactions entre facteurs de susceptibilité génétique et facteurs de risque environnementaux. Balayant les besoins des investigateurs, il montre que de nombreux outils déjà développés ou en cours d'acquisition peuvent être utilisés pour des études d'interaction. La difficulté réside plus dans leur choix et la pertinence de leur assemblage.

*This article\* from a workshop at the 64th annual congress of the American Society of Human Genetics explains the difficulties of research in the promising field of interactions between genetic susceptibility and environmental risk factors. Exploring the needs of researchers, it shows that many tools exist or are currently being developed for use in interaction studies. The difficulty lies more in choosing the most appropriate tools for each study.*

Les études d'association pangénomique par génotypage à haut débit (*genome-wide association studies* – GWAS) ont mis en évidence des milliers de régions chromosomiques (locus) associées à des maladies multifactorielles. Cette approche « agnostique » (sans hypothèse *a priori*) n'est pas conçue pour identifier le variant génétique causal qui influence le risque de maladie et explique l'association (porté par l'un des gènes du locus ou par un gène distant régulant le niveau d'expression du variant génotypé), mais elle met sur la piste de nombreux gènes auparavant non suspectés d'être impliqués dans des maladies et dynamise la recherche d'interactions gène-environnement.

Les promesses portées par la découverte de nouvelles interactions sont nombreuses : avancées dans la compréhension des mécanismes biologiques conduisant à une maladie complexe, développement de meilleurs modèles prédictifs de ces maladies, identification de sous-groupes de populations à haut risque sur la base d'une combinaison de facteurs génétiques et environnementaux pour des actions préventives ciblées, etc., y compris la détection d'autres loci de susceptibilité génétique par une meilleure prise en compte des effets modificateurs de facteurs environnementaux dans les GWAS.

Toutes les branches de l'épidémiologie, et pas uniquement l'épidémiologie génétique, sont appelées à intégrer de plus en plus à leurs travaux des analyses d'interaction gène-environnement (GxE). Pourtant,

malgré l'enthousiasme pour ce champ de recherche, les histoires à succès restent rares. L'une des plus marquantes à ce jour est la découverte de l'interaction entre un variant de la N-acétyltransférase-2 (NAT2) et l'exposition au tabac dans le cancer de la vessie. En s'appuyant sur cet exemple, plusieurs exigences émergent pour des études GxE à grande échelle chez l'homme.

#### Besoins et moyens

Une étude conçue pour rechercher une interaction GxE doit disposer d'une puissance statistique nettement supérieure à

celle d'une étude visant à examiner uniquement les effets d'un facteur génétique ou environnemental. Tenant compte des comparaisons multiples, l'échantillon requis pour détecter des associations dans une étude GxE à l'échelle du génome peut atteindre une taille démesurée. Différentes méthodes permettant des gains de puissance statistique ont été développées en alternative aux tests d'interaction GxE traditionnels à partir d'un échantillon cas-témoins (test d'interaction à partir d'un échantillonnage cas-seul, méthodes bayésiennes, approche cas-parent, tests conjoints, approche en deux étapes avec filtrage), tandis que d'autres sont plus appropriées à des études



sur des variants génétiques rares. Quelques méthodes ont par ailleurs été élaborées pour des analyses quantitatives ou tenant compte de la variation d'une exposition dans le temps. Face au choix des méthodes et outils informatiques, il revient au chercheur de sélectionner les plus appropriés en fonction de considérations relatives à l'objet et au cadre de ses recherches. Les auteurs de l'article mettent en garde contre la tentation de multiplier les méthodes analytiques pour traiter les données : la probabilité d'obtenir des résultats significatifs s'en trouve augmentée, mais il s'agit clairement d'une stratégie biaisée qui favorise l'émergence de fausses associations.

Le besoin d'informations relatives à la fonction des gènes et aux voies métaboliques peut être partiellement comblé par des bases de données publiques comme *Encyclopedia of DNA Elements* (ENCODE), *Epigenomics Roadmap*, *Genotype-Tissue Expression* (GTEx) et *Cancer Genome Atlas* (TCGA), aucun projet ne fournissant toutefois d'information sur l'influence d'expositions environnementales. Les hypothèses d'interactions GxE peuvent également être alimentées par la science des « omiques » (métabolomique, protéomique, épigénomique, etc.) qui génère une abondante

matière pour la recherche, ainsi que par de nouveaux modèles *in vitro* ou murins (populations de souris enrichies en variants génétiques pour mimer la diversité des populations humaines). Ces modèles peuvent aussi être exploités lorsque les études de réplication dans des populations humaines sont particulièrement difficiles en raison de la rareté d'une maladie, d'un variant génétique ou d'une exposition, ou du fait du niveau de sophistication de la méthode analytique. Si la modélisation est très utile pour la validation et l'éclairage mécanistique, l'extrapolation à l'homme demeure toutefois délicate.

### La mesure de l'exposition

Les études GxE requièrent des méthodes de mesure de l'exposition de grande qualité, toute erreur de classement entravant la mise en évidence d'une interaction. Le versant « expositions environnementales » d'une étude GxE est le plus complexe des deux : outre la nature de l'exposition elle-même (physique, chimique, biologique ou psychosociale), il doit appréhender tous les facteurs susceptibles d'influencer le risque de maladie, dont la fenêtre temporelle et la durée de l'exposition, ainsi que, pour une substance, sa voie d'entrée et son

devenir dans l'organisme (distribution dans les tissus cibles, métabolisme, excrétion). De nouveaux moyens technologiques développés dans le champ de l'exposome peuvent servir à générer des données d'exposition externe (capteurs individuels, outils connectés, systèmes d'informations géographiques) et internes (technologies omiques). Lesquels sont les plus appropriés à la détection d'interactions GxE ? Les plus efficaces pour augmenter la puissance statistique ? Comment les intégrer à des études à large échelle ? L'acquisition de données peut-elle se concentrer sur un sous-échantillon (mesures répétées d'un biomarqueur, exploration omique approfondie) puis être extrapolée à une population plus vaste ? Ces questions et d'autres soulevées dans cet article restent à résoudre.

Laurence Nicolle-Mir

\*Mc Allister K, Mechanic LE<sup>1</sup>, Amos C, *et al.* Current challenges and new opportunities for gene-environment interaction studies of complex diseases. *Am J Epidemiol* 2017 ; 186 : 753-761. doi : 10.1093/aje/kwx227

<sup>1</sup> Genomic Epidemiology Branch, Epidemiology and Genomics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, États-Unis.