Les fausses promesses des scores polygéniques en pratique clinique

Résumé de la conférence du Dr Françoise Clerget-Darpoux par Michel Favre[1]

Introduction

Cette conférence a été organisée par le GIS Autisme lors du webinaire Pierre Royer « Psychiatrie et Génétique » qui s'est tenu le 23 mars 2021. Bien que les données scientifiques présentées lors de cette conférence ne soient pas directement impliquées dans l'autisme, elles représentent un bon exemple de ce qu'on peut attendre des données génétiques dans l'interprétation de nombreuses maladies, notamment les maladies psychiatriques. Dans le cas de l'autisme, les données rapportées pourraient être utilisées pour favoriser un meilleur diagnostic et conduire à un programme d'accompagnement adéquat. En préalable au résumé de la conférence du Dr Clerget-Darpoux, un bref rappel des connaissances sur la génétique de l'autisme est présenté afin de permettre une meilleure compréhension du webinaire.

Génétique et autisme

Depuis plus de 40 ans il est largement admis que la constitution génétique de l'hôte est impliquée dans l'origine de l'autisme. En effet, en tenant compte de critères stricts de définition de l'autisme, de nombreuses études ont montré que lorsqu'un enfant était autiste dans 98% son jumeau monozygote l'était aussi alors que ce risque était d'environ 50% dans le cas d'un jumeau dizygote. Depuis ces premières études, plus de 300 mutations dans de nombreux gènes ont été rapportées. La plupart de ces gènes sont impliqués dans la formation du système nerveux en modulant la formation des synapses (synaptogénèse) ou dans la régulation des gènes mis en jeu dans les voies essentielles de signalisation cellulaires comme l'ont montré les travaux de l'équipe d'Eric Courchesne aux États-Unis.

Dans une minorité de cas d'autisme on peut mettre en évidence une mutation dans un seul gène inhibant ou modifiant son activité (hypothèse monogénique). Ainsi par exemple des mutations dans les gènes des neuroligines 3 et 4, Shank 3, NRX3, contactine… ont été identifiées. Il faut cependant noter que très souvent la fonction précise du gène impliqué n'est pas ou mal connue. En revanche, dans la majorité des cas, la situation est beaucoup plus complexe pouvant mettre en jeu différents allèles des gènes (polymorphisme).

Un allèle (abréviation d'allélomorphe) est une version variable d'un même gène, c'est-à-dire une forme différente qui peut être distinguée par des variations de sa séquence nucléotidique. Il existe généralement quelques allèles connus pour chaque gène, certains gènes (par exemple ceux du complexe majeur d'histocompatibilité, CMH) possèdent plusieurs dizaines d'allèles. C'est ainsi qu'au sein d'une même espèce, le génome d'un individu est différent de celui d'un autre individu, c'est le polymorphisme génétique. Il peut donc exister dans les populations naturelles plusieurs séquences différentes d'ADN pour un même gène.

Pris séparément le polymorphisme d'un gène n'a pas d'effet mais la combinaison de plusieurs allèles de quelques gènes augmente le risque d'être atteint d'autisme. L'extrême variabilité des troubles autistiques et le grand nombre de formes alléliques des gènes rendent l'interprétation des études génétiques d'association très difficiles et nécessitent l'analyse de plusieurs milliers de cas ou de familles.

Devant la complexité de ces études, plusieurs travaux ont été consacrés à des modèles relativement plus simples et faciles d'accès comme la schizophrénie ou le cancer du sein car ces maladies s'adressent à des populations de patients homogènes. L'analyse

1. Président de Pro Aid Autism.Mail : michel.favre21@sfr.fr

ARAPI – Licence CC v.4.0 Int.
Attribution – Pas d'Utilisation Commerciale – Pas de Modification

génétique de ces deux maladies ont fait l'objet d'un exposé du Dr Françoise Clerget-Darpoux dans le cadre du séminaire Pierre Royer qui s'est tenu en mars dernier consacré au score de risques polygéniques en psychiatrie. Cette intervention avait pour titre **les fausses promesses des scores polygéniques en pratique clinique**. Il s'agit d'un sujet brûlant encore très controversé et qui suscite de nombreuses questions et interrogations. Néanmoins, ces scores peuvent se révéler très utiles pour les cliniciens dans des études familiales. De plus, les études génétiques à grande échelle menées sur plusieurs milliers de familles sont actuellement en cours. Elles pourraient permettre de dégager des critères très utiles pour le suivi et le devenir des patients.

En préalable à la communication du Dr Clerget-Darpoux, le modérateur de la session, le professeur Boris Chaumette, a fait un certain nombre de rappels et donné les bases pour l'étude de l'association des maladies avec des gènes.

Il a rappelé que ces scores polygéniques sont basés sur une étude d'association pan génomique (en anglais *Genome-Wide Association Study* ou GWAS) qui correspondent à une analyse de nombreuses variations génétiques chez de nombreux individus, afin d'étudier leurs corrélations avec des traits phénotypiques (symptômes et apparence). Ainsi des fréquences de polymorphismes génétiques entre des cas et des individus contrôles sont comparées ce qui permet de dériver des valeurs de p statistique et un ODD RATIO ou rapport des risques. Ces valeurs correspondant à un risque relatif souvent utilisé en épidémiologie notamment pour les maladies psychiatriques. Ces scores polygéniques sont très utilisés par des laboratoires privés pour prédire un risque de développer une maladie. Ces données, très utilisées à l'étranger, sont interdites en France d'autant plus que les individus ayant sollicité ces laboratoires privés peuvent interpréter ces résultats à titre personnel et en déduire leur propre atteinte à une maladie psychiatrique. Il s'agit donc d'un sujet très critiquable et discuté quant à son intérêt en clinique. En conclusion à sa courte introduction, Boris Chaumette a donné la parole au Docteur Françoise Clerget-Darpoux afin qu'elle puisse présenter et discuter les données disponibles sur les scores polygéniques et leur utilité en clinique. Elle possède une double formation en mathématique et en génétique.

Association maladies et complexe majeur d'histocompatibilité

Dans un premier temps, le Dr Françoise Clerget-Darpoux a rappelé que depuis plus de 50 ans il était envisagé que l'information génétique pourrait prédire et prévenir les maladies. C'est par exemple le cas de l'association des allèles des gènes HLA localisés dans le complexe majeur d'histocompatibilité (CMH) avec plusieurs maladies. Historiquement, le professeur Jean Dausset reçut le Prix Nobel de Médecine en 1980 pour sa découverte du CMH. Selon lui l'identification de certains allèles pouvait être d'une très grande utilité pour la mise en place de traitements et éventuellement d'éradiquer les maladies. Cependant ces études d'associations des allèles des gènes HLA et maladies ont révélé le très grand fossé existant entre association génétique et explication fonctionnelle. De plus, il est toujours nécessaire en plus des données génétiques de tenir compte des données physiopathologiques et de l'information familiale, en d'autres termes des antécédents familiaux. Depuis lors, les scores polygéniques qui reflètent le risque d'un individu ont conduit à définir un risque absolu faible de développer une maladie. Celui-ci est calculé dans les études cas-témoins par l'estimation des porteurs d'un facteur de risque par rapport aux individus normaux. Les ODD RATIO constituent une estimation des risques relatifs. L'étude de la sclérose en plaque (SEP), constitue un bon exemple pour illustrer ces données d'association entre allèles des gènes HLA et une maladie auto-immune inflammatoire qui attaque le système nerveux central. Ainsi les personnes qui sont porteuse de l'allèle DR15 sont 4 fois plus à risque d'avoir une SEP que les autres individus. Cependant, 20% des individus de la population générale sont DR15. Ainsi le risque absolu d'avoir une SEP pour les porteurs de l'allèle DR15 n'est que de 1/1000, ce qui constitue une donnée de peu d'intérêt pour les cliniciens. Ainsi, l'ensemble des études consacrées au rôle du CMH dans la survenue de la plupart des maladies associées aux génotypes HLA a permis de conclure que l'identification de l'allèle ne permet pas de prédire des risques en population générale.

Il existe cependant une exception représentée par une mutation dans le gène HFE qui est localisé dans la région chromosomique des gènes HLA. Cette mu-

ARAPI – Licence CC v.4.0 Int.
Attribution – Pas d'Utilisation Commerciale – Pas de Modification

tation permet de prédire une hémochromatose, caractérisée par une hyperabsorption intestinale de fer entraînant une destruction des organes.

Cartographie génétique et maladies monogéniques ou multifactorielles

Depuis ces premiers travaux sur le CMH, de nombreuses études génétiques ont été menées dans des familles à l'aide du génotypage des microsatellites ou du polymorphisme nucléotidique encore appelé SNPs ou « single nucleotide polymorphisms ». Ces travaux ont permis d'identifier des gènes candidats dans des maladies monogéniques comme la SEP, certaines formes de la maladie d'Alzheimer ou le cancer du sein avec les gènes BRCA1 et BRCA2. Ces études familiales ont par ailleurs apporté beaucoup d'informations sur le risque relatif lié au génotype de certains SNPs des gènes associés à une maladie. C'est le cas par exemple dans la SEP de 2 SNPs du gène IL2RA ou de 3 SNPs dans le gène PTPN22. Cependant, ces études de génotypage se sont révélées peu informatives dans le cas des maladies génétiques complexes et multifactorielles.
En revanche, l'arrivée au début du 21ème siècle des techniques de génotypage et de séquençage à haut débit ont permis d'apporter un jour nouveau aux études génétiques. Ceci avait conduit le directeur du *« National Institute of Health »* Francis Collins à prédire qu'un diagnostic génétique serait disponible pour toutes les maladies dans les 10 ans à venir. Malheureusement les années ultérieures n'ont pas permis de concrétiser ce rêve.
Il faut aussi noter que dès les années 2006-2007, certaines compagnies de vente en ligne de tests génétiques ont proposé de fournir des scores de risque polygéniques et de donner une prédiction du futur médical des individus. La communauté scientifique française s'est fortement élevée contre la vente de ces tests. A la même époque, une étude a proposé d'utiliser les GWAS (*genome-wide association study*) comme outil pour prédire le risque de développer une maladie (*WRAY NR, GODDARD ME and WISSCHER, Genome Research 2007, prediction of individual risk to disease from genome-wide association studies*). Il était mentionné dans cet article que les cliniciens pourraient utiliser ces données pour faire du diagnostic et du pronostic et éventuellement traiter les maladies ce qui est toujours l'objet d'une controverse. En effet, une mauvaise interprétation de ces données génétiques peut conduire à des faux positifs. Au cours de son exposé, le docteur Clerget-Darpoux s'est donc intéressée à donner plusieurs éléments pour comprendre la problématique des scores de risque polygéniques.

Scores de risque polygénique (SRP)

En premier lieu, ces scores peuvent être utilisés selon un modèle polygénique additif qui fait appel à l'existence d'une variable quantitative mesurant une prédisposition due à la fois à des facteurs génétiques et des facteurs environnementaux ayant de petits effets additifs, cumulatifs et indépendants les uns des autres. Ce modèle permet de prédire un seuil au-dessus duquel les individus sont atteints. Ceci suppose une homogénéité étiologique des facteurs. Ce modèle fait appel aux SNPs associés à une maladie et permet de calculer son héritabilité. Il est mentionné que ce calcul conduit à de nombreuses erreurs d'interprétation.

L'héritabilité est définie comme le rapport de la variance génétique et de la variance phénotypique. Il est rappelé que la variance est une mesure du degré de dispersion d'un ensemble de données. La variance (symbolisée par S^2) et l'écart-type (la racine carrée de la variance, symbolisée par S) sont les mesures de dispersion les plus couramment utilisées. On la calcule en prenant la moyenne de l'écart au carré de chaque nombre par rapport à la moyenne d'un ensemble de données. Ainsi par exemple pour les nombres 1, 2 et 3, la moyenne est 2, et, la variance correspond à [(1 - 2)2 + (2 - 2)2 + (3 - 2)2] ÷ 3 soit 0,667.
Ainsi l'héritabilité correspond à un différentiel sur le risque entre les malades et les sujets sains. En d'autres termes elle permet de définir l'importance de la variabilité génétique en fonction de la présence de la maladie dans la population générale. Ainsi, les scores de risque relatifs comparent les porteurs d'allèles à risque par rapport à ceux qui ne le portent pas. Dans son exposé le Docteur Clerget-Darpoux a mentionné que les scores de risque polygéniques reviennent à décomposer la variance génétique en petits éléments en tenant compte du fait que certains SNPs ou variants génétiques sont plus fréquemment associés à la maladie étudiée. Ainsi, d'après l'article de WRAY et al précédemment cité, la meilleure

ARAPI – Licence CC v.4.0 Int.
Attribution – Pas d'Utilisation Commerciale – Pas de Modification

valeur capable de prédire une maladie donnée sera obtenue en agrégeant l'ensemble des variants expliquant la variabilité génétique et donc l'héritabilité. Aussi depuis 10 ans, il apparait essentiel pour toute maladie et notamment les maladies psychiatriques, de trouver un maximum de SNPs associés. Des logiciels ont ainsi été développés permettant de calculer le score de risque polygénique d'une personne en connaissant le génotype des SNPs et la variabilité autour de ce marqueur génétique. Ceci a conduit à l'essor de très nombreuses applications pour l'étude des maladies multifactorielles afin de pouvoir prédire le SRP de chaque individu et fournir des mesures préventives pour la clinique.

La question posée lors de la conférence a été de savoir si le modèle polygénique additif est adéquat pour calculer les SRP, en d'autres termes que l'architecture relativement simple du modèle peut expliquer toutes les formes des maladies complexes. Ceci a conduit le Dr Clerget-Darpoux à évoquer un syndrome de *«garbage-in garbage-out (GIGO)»* compte tenu de la qualité des données et des hypothèses disponibles. Le garbage-in garbage-out est un concept informatique selon lequel des données d'entrée défectueuses ou absurdes produisent des sorties absurdes ou "déchets". Il est bien connu que des mauvaises informations sont synonymes de mauvaises conclusions et que pire encore l'absence d'information est synonyme d'absence de conclusions.

De plus il est clair qu'association ne signifie pas obligatoirement un rôle causal dans la maladie étudiée mais pourrait correspondre seulement à une conséquence. Ainsi une association peut refléter une stratification environnementale liée à des éléments sociaux-culturels. Elle pourrait également traduire seulement le rôle indirect et non causal de variants génétiques. Par exemple, il a été ainsi montré que des SNPs associés au cancer du sein peuvent correspondre à des variants génétiques spécifiques de l'indice de masse corporelle et ne jouer aucun rôle dans ce type de cancer. En outre il ne faut pas oublier le rôle de gènes ayant des effets pléiotropiques, voire ubiquitaires tels que les gènes mis en jeu dans la transcription ou dans la régulation des voies de signalisation cellulaires. Ceci pourrait être le cas pour le rôle de ces gènes dans les maladies psychiatriques et dans l'autisme. Ainsi, dérégulation ou modification de l'expression de gènes ne pourrait être que la conséquence de la maladie et non sa cause essentielle.

Dans le même ordre d'idée, un autre modèle dit « omnigénique » a été proposé par Boyle, Li et Pritchard (Cell, 2017) qui pourrait montrer que probablement de nombreux gènes et donc notre génome jouent un rôle dans les maladies.

Il est rappelé dans la conférence que les SRP n'expliquent pas tout mais pourraient être considérés comme un facteur de prédiction complémentaire à côté de facteurs cliniques, familiaux ou environnementaux. Dans le cas de l'autisme, des mutations « de novo » de pénétrance incomplète pourraient se rajouter aux SRP et pourraient apporter un éclairage supplémentaire sur ce handicap. Ainsi se dégage l'idée que les SRP constituent un élément important et vont apporter de l'information aux facteurs mentionnés ci-dessus.
Un accent tout particulier a été mis sur le fait que les SRP ont des valeurs locales et temporelles. En effet, ceci signifie que pour une maladie donnée les variances phénotypique, génétique et environnementale varient à la fois dans le temps et dans l'espace. Il convient de souligner que les SRP tout comme l'héritabilité, n'ont pas une valeur intrinsèque absolue. En effet, ils diffèrent en fonction des populations étudiées et de leurs caractéristiques. Ainsi par exemple dans une publication de Mostafavi (Elife, 2020) il est rapporté que pour une population considérée homogène il faut tenir compte de l'âge, du sexe et du statut socio-économique. Dans cette analyse, il a été montré que les SRP varient avec ces différentes variables.

SRP associés à des maladies psychiatriques

Pour illustrer ses propos en psychiatrie, le Dr Clerget-Darpoux a présenté comme exemple la schizophrénie et le trouble bipolaire. En effets de nombreux SNPs communs sont retrouvés dans ces deux maladies (travaux du Consortium International sur la schizophrénie parus dans la revue Nature en 2009). Dans des publications ultérieures a été développé le double outil SRP schizophrénie, le SRP trouble bipolaire et l'identification des SNPs associés comme prédicteur de la maladie (*Nature* 2014, *Nature Genetics* 2019). Le développement de psychoses a constitué un élément important dans ces études. L'apparition de troubles de l'humeur a été également consi-

ARAPI – Licence CC v.4.0 Int.
Attribution – Pas d'Utilisation Commerciale – Pas de Modification

déré comme prédicteur pour les personnes à risque de développer des psychoses (Jonas et al., *Translational Psychiatry*, 2019).

Une étude plus approfondie de la schizophrénie sur une population scandinave s'est intéressée à la prédiction des SRP compte tenu du contexte familial. Il a été montré que les SRP sont plus élevés chez les patients avec des antécédents familiaux (Luy et al, Psychol Med, 2018).

SRP associés au cancer du sein

De nos jours, il convient de noter que les scores de risque polygéniques sont déjà utilisés dans le traitement personnalisé des cancers. Dans le cas du cancer du sein, le Docteur Clerget Darpoux a fait mention d'une étude parue en 2015 (*Journal of Natl Cancer Inst* 2015) portant sur 33 000 femmes avec un cancer et 33 000 individus témoins d'origine européenne. Dans cette étude est rapportée le nombre de variants nucléotidiques des SNPs qui peuvent correspondre à des allèles de gènes à risque pour le cancer du sein. Les SRP ont inclus 77 SNPs spécifiques de gènes associés à ce type de cancer. Il convient de remarquer que les 77 SNPs correspondent à 154 allèles dont certains sont présents chez les patients et les sujets normaux. Il est clair cependant qu'aucun individu ne présente les 154 allèles.

L'analyse comparative des allèles chez les femmes avec cancer du sein et les individus normaux a montré l'existence de 5 allèles de risque chez les femmes avec cancer. Il a été trouvé que les femmes avec le SPR le plus élevé sont 3 fois plus à risque de développer un cancer du sein.

Conclusions

Plusieurs données de la littérature ont mis l'accent ces dernières années sur l'interprétation erronée des scores de risque polygéniques. C'est le cas par exemple de l'article de Wald N and Old R *"the illusion of disease risk prediction"*. Comme cela a été mentionné au cours du séminaire, il convient de rappeler que les SRP sont des sommes de variance et, par exemple, dans le cas des études basées sur des SNPs, on cumule les risques relatifs et non les risques absolus. On considère que les SRP mesurent la combinaison d'allèles positivement associés à une maladie comme le cancer du sein versus le génotype sans allèle à risque. Ainsi, dans une population homogène et dans un environnement bien défini, les SRP vont permettre de classifier les individus comme étant plus ou moins à risque de développer un cancer du sein. En revanche, les SRP ne donnent en aucun cas le risque absolu de développer la maladie. En effet, un génotype de SNPs à risque peut être détecté chez des individus indemnes de maladie ou absent chez les patients. Il faut donc remarquer que, comme pour de nombreux tests biologiques, le danger est représenté par les faux positifs et les faux négatifs. Ceci souligne le risque d'une mauvaise interprétation des tests génétiques disponibles sur internet.

Dans l'autisme, il faut remarquer que la très grande diversité phénotypique des cas rend difficile l'identification d'allèles de risque. On est donc loin de pouvoir prédire uniquement sur des bases génétiques la survenue d'autisme dans la population générale. Cependant, des études génétiques menées à grande échelle sur plusieurs milliers de familles en France et dans le monde sont actuellement en cours. Elles pourraient apporter un éclairage nouveau quant au suivi et devenir des patients.

En conclusion, les SRP ne sont en aucun cas un outil de prédiction en population générale et il convient de manipuler avec beaucoup de prudence et esprit critique les scores polygéniques en pratique clinique. Ils peuvent seulement apporter au clinicien un complément dans l'interprétation du cas clinique étudié. Il faut également remarquer que beaucoup de maladies génétiques ne sont pas héritées mais résultent de mutations de novo. Le séquençage à haut débit devrait permettre d'identifier ces mutations. Dans l'autisme, ceci limite grandement l'utilisation des SRP et il convient de rechercher ces mutations de novo dans des gènes reconnus jouer un rôle dans ce handicap. En outre, il est nécessaire en plus des données génétiques de combiner des données familiales, environnementales et fonctionnelles. Ainsi, une meilleure connaissance de la fonction et de l'activité des gènes dans différents contextes devrait fournir une meilleure compréhension des maladies génétiques.

■

ARAPI – Licence CC v.4.0 Int.
Attribution – Pas d'Utilisation Commerciale – Pas de Modification