

Chroniques génomiques

Scores polygéniques et risque de cancer

Bertrand Jordan

Les limites des GWAS

Les études d'association génome entier (GWAS, *genome-wide association studies*), largement pratiquées depuis une quinzaine d'années, ont permis l'identification de nombreux locus influant sur la prédisposition à diverses affections multifactorielles et à certains cancers [1]. Les variants ainsi découverts sont généralement assez nombreux pour une affection donnée, et l'effet de chacun d'eux (l'augmentation du risque pour le porteur du « mauvais » allèle) est faible, de l'ordre de 20 à 30 % en général (risque relatif de 1,2 à 1,3). C'est tout à fait logique : ces « balayages du génome » sont effectués à l'aide de puces à ADN (ou *microarrays*) qui analysent environ 500 000 points sur chaque ADN étudié, un petit échantillonnage par rapport aux trois milliards de bases d'un génome. Et ces points ne sont pas pris au hasard : chacun d'eux correspond à un polymorphisme connu dans l'ADN humain, un SNP (*single nucleotide polymorphism*) pour lequel l'allèle mineur est présent à une fréquence d'au moins 5 %¹. Ce choix est nécessaire pour que les analyses détectent des différences entre des échantillons d'ADN provenant de différentes personnes, mais il a pour résultat que les GWAS ne détectent que des allèles assez fréquents (plus de 5 %) dans la population. Si le « mauvais » allèle était associé à un doublement ou un triplement du risque, il serait rapidement éliminé par la sélection naturelle. Il est donc normal, et même obligatoire, que les locus de risque révélés par les GWAS soient associés à une faible augmentation du risque. Reste alors une interrogation : peut-on combiner les résultats obtenus pour en tirer une information utile en clinique ? Nous avons déjà abordé cette question dans une récente *Chronique génomique* [1] (→) et une publication toute récente [2] en fournit une bonne illustration.

(→) Voir la Chronique
génomique de B. Jordan,
m/s n° 12, décembre 2018,
page 1116



UMR 7268 ADÉS, Aix-Marseille,
Université /EFS/CNRS ;
CoReBio PACA, case 901,
Parc scientifique de Luminy,
13288 Marseille Cedex 09,
France.
brjordan@orange.fr

Dix SNP pour le cancer de la thyroïde

Le cancer papillaire de la thyroïde, forme la plus fréquente de cette affection, présente une héritabilité notable [3, 4] et a fait l'objet de plusieurs études d'association génétique, identifiant des locus impliqués dans la vulnérabilité à ce cancer [5]. Les dix SNP les plus significatifs correspondent à des valeurs de risque relatif (RR) allant de 1,2 à 1,7 ; la fréquence de l'allèle de risque allant de 0,04 (RR 1,7) à 0,85 (RR 1,3). Ce n'est sans doute pas par hasard que l'allèle le moins fréquent est aussi celui qui correspond au risque relatif le plus élevé... Les auteurs ont établi un score de risque polygénique (PRS, *polygenic risk score*) combinant les effets de ces dix locus et l'ont testé sur trois échantillons de population, le premier (« Ohio ») comprenant 1 544 malades et 1 595 témoins nord-américains, le deuxième provenant de la banque d'ADN de *Decode Genetics* (Islande)² et comprenant 723 malades et 129 556 témoins, le troisième groupe enfin n'est autre que celui de la *UK Biobank*³ qui s'avère compter 534 malades pour 407 945 témoins. Pour chaque individu, la connaissance des allèles présents pour ces dix SNP permet de calculer un score qui est d'autant plus élevé que le risque de cancer de la thyroïde pour cette personne est fort. Les résultats sont exprimés en utilisant le formalisme de la courbe ROC (*receiver operating characteristic*) ou « fonction d'efficacité du

¹ Entre 5 % et 1 % on parle de variants peu fréquents, et au-dessous de 1 % de variants rares.

² <https://www.decode.com/>

³ <https://www.ukbiobank.ac.uk/>



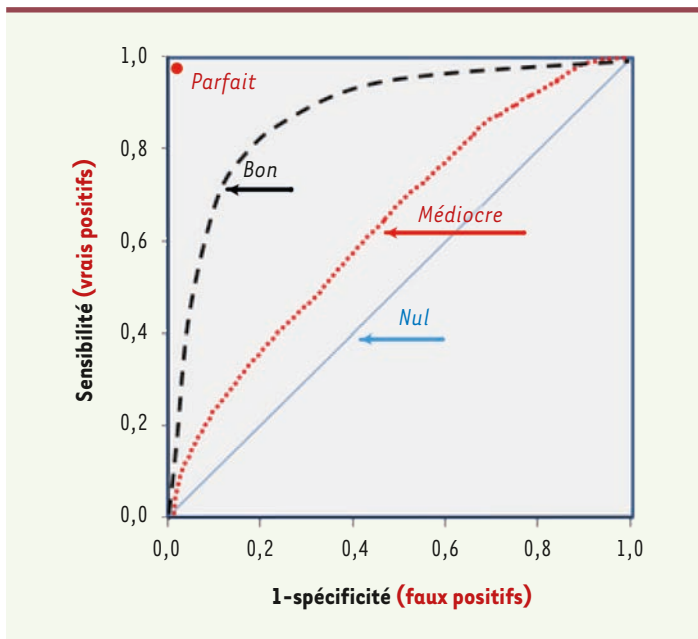


Figure 1. Courbes ROC et qualité des prédictions. Pour un diagnostic n'apportant aucune information (diagonale sur la figure), on obtient une AUC (aire sous la courbe) égale à 0,5. Pour un diagnostic parfait (sensibilité et spécificité simultanément égales à 1), la valeur de l'AUC serait de 1.

récepteur», qui n'est autre qu'une courbe sensibilité/spécificité, ou plus précisément sensibilité versus 1-spécificité permettant de juger les performances d'un système [6] (→).

(→) Voir la Chronique génomique de B. Jordan, m/s n° 3, mars 2012, page 325

En effet, sensibilité et spécificité d'un test ne sont pas indépendantes : si l'on abaisse le seuil de détection, on améliore la sensibilité de détection mais on augmente le taux de faux positifs et l'on perd donc en spécificité (Figure 1). La qualité d'un diagnostic est alors exprimée par la surface englobée par la courbe (AUC, *area under the curve*).

Il s'agit donc ici d'examiner la qualité de la prédiction du cancer de la thyroïde par le score polygénique (PRS) reposant sur les dix SNP majeurs. On ne s'attend pas à une valeur très élevée de l'AUC (0,8 ou 0,9) puisque cette affection n'est pas strictement héréditaire, mais il s'agit néanmoins de voir si le PRS apporte des informations utiles. La Figure 2 montre le résultat obtenu pour le score polygénique appliqué au groupe « Ohio » (les résultats sont similaires pour les deux autres groupes).

La courbe rouge correspond à l'évaluation du risque d'après les seules informations cliniques (âge, genre, origine ethnique et présence du même cancer dans la parenté de premier et deuxième degré). On voit que cette courbe rouge s'écarte (un peu) de la diagonale et indique donc que les facteurs cliniques ont une certaine valeur prédictive, qui reste modeste avec une AUC de 0,585 (rappelons qu'une AUC de 0,5 correspond à l'absence d'information). La courbe bleue indique les résultats si l'on ajoute le PRS aux facteurs cliniques : l'AUC monte alors à 0,692, une valeur déjà notable. Les auteurs ont également calculé un score polygénique prenant en compte l'ensemble de 592 000 SNP

mesurés, et l'ont combiné avec les PRS des dix SNP majeurs et les facteurs cliniques. On voit (courbe verte) que cela n'améliore pas le résultat : les deux courbes sont confondues et l'AUC passe de 0,692 à 0,693. Ce résultat est important : il signifie que les dix SNP majeurs rendent compte de la contribution de tous les variants fréquents (ceux que l'on détecte avec un *microarray*) dans l'ensemble du génome. En d'autres termes, inutile de refaire des GWAS sur ce modèle : il n'y a plus de locus (à polymorphisme fréquent) à trouver - du moins dans les populations majoritairement européennes qui ont été analysées. De ces dix SNP, un seul est localisé dans une séquence codante, les neuf autres dans des régions intergéniques ou dans des introns : l'effet des variants porte donc vraisemblablement sur la régulation de l'expression de gènes ou de groupes de gènes.

Une autre manière d'exprimer l'utilité clinique de ce PRS consiste à déterminer le risque relatif correspondant à chaque tranche de 10 % (décile⁴). Ceci est effectué en calculant, pour chaque individu, la valeur du PRS en fonction de son profil génétique, et en relevant la proportion de cancers de la thyroïde pour la population appartenant à un décile donné. Comme le montre la Figure 3, le risque relatif est presque sept fois plus élevé au dixième décile qu'au premier : cette différence est suffisamment importante pour envisager, par exemple, une stratégie de détection précoce centrée sur la population à haut risque.

Vers la généralisation des scores polygéniques ?

Rappelons que la quasi-totalité des SNP définissant le PRS est située dans des régions intergéniques ou dans des introns, indiquant que ces variants jouent sur la régulation de l'expression de gènes et non sur la structure d'une protéine. Notons aussi que ces SNP rendent compte de l'essentiel de l'influence génétique liée à des variants fréquents (fréquence de l'allèle mineur supérieure à 5 %), mais n'excluent pas l'existence de variants rares (non examinés par le *microarray*) et dont l'effet pourrait être notable. La tendance actuelle à caractériser l'ADN rassemblé dans les banques par séquençage intégral, comme cela a été entamé par la *UK Biobank*⁵ devrait à terme révéler de tels variants, à condition que l'effectif séquencé soit très important afin de permettre la détection d'allèles rares.

Cet article confirme néanmoins que les scores polygéniques permettent dès à présent une évaluation du

⁴ En statistique descriptive, un décile est chacune des 9 valeurs qui divisent un jeu de données, triées selon une relation d'ordre, en 10 parts égales, de sorte que chaque partie représente 1/10 de l'échantillon de population (Wikipedia).

⁵ <https://www.ukbiobank.ac.uk/wp-content/uploads/2019/09/WGS-FAQ-final-AT.pdf>

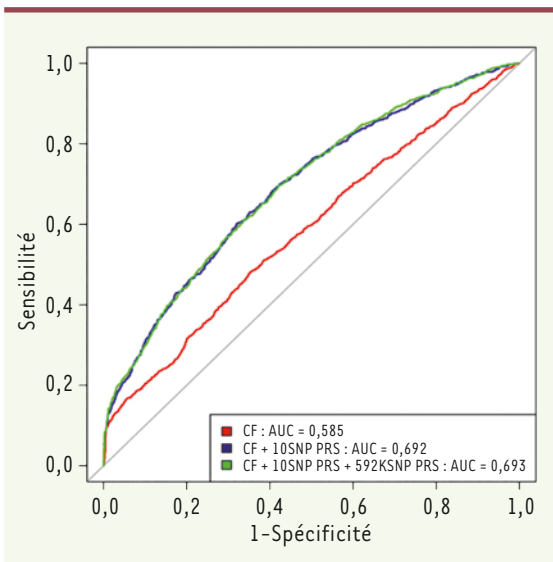


Figure 2. Courbe ROC pour le score polygénique (PRS) appliqué au groupe « Ohio ». CF : facteurs cliniques (copie partielle et modifiée de la figure 1 de [2]).

risque, même pour un cancer dont l'héritabilité est modeste, et sans nécessairement utiliser l'ensemble du profil génétique fourni par un *microarray* (coût de l'ordre de 50 €) ni recourir à un séquençage intégral (coût réel de l'ordre de 1 000 €). De nombreuses études récentes vont dans le même sens [7]. On peut donc penser que l'usage de ces scores va se généraliser, y compris pour des affections comme le diabète de type 2 ou la maladie coronarienne : plusieurs entreprises, dont 23andMe, se positionnent déjà sur ces créneaux [8]. Cela ne manquera pas de soulever quelques problèmes liés à une mauvaise compréhension de ce que l'on a assez improprement appelé la « médecine prédictive ». Il est nécessaire d'améliorer la perception de ce que représente un risque : une probabilité, éventuellement une vulnérabilité, mais certainement pas une prédiction et encore moins un destin. L'industrie du bien-être (*Wellness industry*) affiche, selon le *Global Wellness Institute*, un chiffre d'affaires mondial de 4 200 milliards de dollars, dont 575 pour la médecine préventive et personnalisée⁶ et promeut l'exploitation de profils génétiques personnalisés, chez l'adulte comme pour l'embryon [9] (→) : il ne faudrait pas que les réels progrès dans l'évaluation du risque aboutissent à transformer toute la population en « malades en puissance »... ♦

(→) Voir la Chronique génomique de B. Jordan, m/s n° 3, mars 2020, page 289

⁶ <https://globalwellnessinstitute.org/press-room/press-releases/wellness-now-a-4-2-trillion-global-industry/>

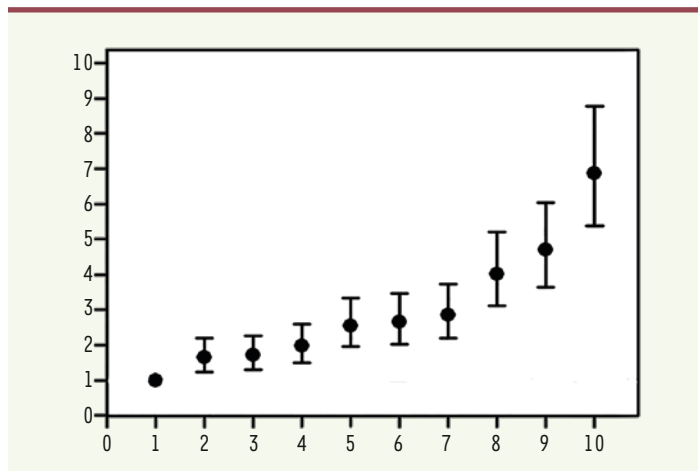


Figure 3. Risque relatif de cancer de la thyroïde (en ordonnée) en fonction des déciles du score polygénique (PRS) fondé sur 10 polymorphismes (SNP) (en abscisse). Les barres d'erreur indiquent l'incertitude statistique, variant notamment en fonction de l'effectif de chaque décile. Les valeurs sont normalisées par rapport au risque relatif trouvé pour le premier décile (extrait partiel et remanié de la figure 2 de [2]).

SUMMARY

Polygenic scores and cancer risk

Risk assessment for a cancer type with moderate heritability can be accurately performed using a relatively small number of SNPs detected by GWAS analyses to calculate a polygenic risk score (PRS) that has definite clinical utility. ♦

LIENS D'INTÉRÊT

L'auteur déclare n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

RÉFÉRENCES

- Jordan B. Balayage du génome et repérage des personnes à risque : des GWAS aux GPS. *Med Sci (Paris)* 2018 ; 34 : 1116-22.
- Liyanarachchi S, Gudmundsson J, Ferkingstad E, et al. Assessing thyroid cancer risk using polygenic risk scores. *Proc Natl Acad Sci USA* 2020 ; 117 : 5997-6002.
- Risch N. The genetic epidemiology of cancer: interpreting family and twin studies and their implications for molecular genetic approaches. *Cancer Epidemiol Biomarkers Prev* 2001 ; 10 : 733-41.
- Hińcza K, Kowalik A, Kowalska A. Current knowledge of germline genetic risk factors for the development of non-medullary thyroid cancer. *Genes (Basel)* 2019 ; 10. pii: E482.
- Figlioli G, Elisei R, Romei C, et al. A comprehensive meta-analysis of case-control association studies to evaluate polymorphisms associated with the risk of differentiated thyroid Carcinoma. *Cancer Epidemiol Biomarkers Prev* 2016 ; 25 : 700-13.
- Jordan B. Les tests génétiques grand public ont-ils une utilité clinique ? *Med Sci (Paris)* 2012 ; 28 : 325-8.
- Guochong Jia, Yingchang Lu, Wanqing Wen, et al. Evaluating the utility of polygenic risk scores in identifying high-risk individuals for eight common cancers. *JNCI Cancer Spectrum* 2020, pkaa021. <https://doi.org/10.1093/jncics/pkaa021>
- Regalado, A. 23andMe thinks polygenic risk scores are ready for the masses, but experts aren't so sure. *MIT Technology Review*, 8 mars 2019. <https://www.technologyreview.com/s/613095/23andme-thinks-polygenic-risk-scores-are-ready-for-the-masses-but-experts-arent-so-sure/>
- Jordan B. Scores polygéniques : vers l'embryon à la carte ? *Med Sci (Paris)* 2020 ; 36 : 289-91.

TIRÉS À PART

B. Jordan