

# Chroniques génomiques

## Maladie de Crohn et GWAS, d'analyses en méta-analyses

Bertrand Jordan



Marseille-Nice Génomole,  
case 901,  
Parc scientifique de Luminy,  
13288 Marseille Cedex 9, France.  
[brjordan@club-internet.fr](mailto:brjordan@club-internet.fr)

Les études d'association génétique à grande échelle (*genome-wide association studies*, GWAS) se poursuivent de plus belle : il s'en est pratiqué aujourd'hui plus d'un millier et, à lire les sommaires de *Nature Genetics*, on pourrait être tenté de rebaptiser cette revue *Nature GWAS*... Compte tenu du nombre de travaux publiés, mais aussi d'une méthodologie maintenant bien établie et de critères statistiques précis et homogènes, il s'avère possible de combiner plusieurs études portant sur la même affection afin d'en accroître la puissance statistique et de mettre ainsi en évidence des locus qui ne pouvaient être identifiés de manière concluante dans les analyses originelles. Je vais commenter ici les enseignements d'une méta-analyse très récente portant sur la maladie de Crohn qui montre la puissance mais aussi les limites de cette approche.

Rappelons tout d'abord que la maladie de Crohn est une affection inflammatoire chronique touchant différents segments du tube digestif, maladie grave entraînant toutes sortes de troubles, sans traitement autre que symptomatique et pouvant déboucher sur une chirurgie mutilante. Elle est relativement fréquente, avec une prévalence de deux à quatre pour mille, et présente une composante génétique importante : son héritabilité au sens strict a été évaluée à 50 % [1]. C'est donc, à l'évidence, une cible intéressante pour des études génétiques susceptibles d'éclairer son étiologie, d'améliorer le diagnostic et, à terme, de favoriser la mise au point de traitements efficaces. Les premiers travaux, « pré-GWAS », ont permis d'identifier de manière concluante le gène *NOD2*<sup>1</sup>,

d'obtenir des indications assez solides pour un deuxième locus de susceptibilité appelé *IBD5* (*inflammatory bowel disease 5*) situé en 5q13, et ont montré que des études beaucoup plus approfondies allaient être nécessaires, car ces deux premières entités ne rendaient compte que de 10 % environ de l'héritabilité constatée [2].

### Les GWAS et les méta-GWAS entrent en scène

Avec les nouvelles possibilités de génotypage à grande échelle offertes par les puces à ADN *Illumina* ou *Affymetrix* et la codification progressive du mode opératoire des études d'association génome entier, la maladie de Crohn a naturellement été soumise à de telles analyses. Un premier bilan publié en 2008 [3] et résumant les résultats de 7 GWAS fait état d'une dizaine de locus solidement établis et indique les hypothèses que l'on peut formuler sur la pathogenèse de l'affection. Bien que ce point ne soit pas discuté explicitement dans la revue citée, il semble pourtant que ces 10 locus soient loin de rendre compte de la forte composante génétique observée pour cette affection. Une méta-analyse des études présentant le plus fort effectif, publiée la même année [4], confirme ce soupçon : bien que l'on arrive alors à 32 locus identifiés de manière statistiquement significative, ils ne rendent compte à eux tous que de 20 % de l'héritabilité observée. Une nouvelle méta-analyse vient de paraître et annonce maintenant un total de 71 locus : c'est cet article, récemment paru dans *Nature Genetics* [5], qui fait l'objet de la présente chronique.

Les chiffres sont impressionnants : la publication, qui comporte 91 auteurs provenant de 69 laboratoires,

<sup>1</sup> Le gène *NOD2* ou *CARD15*, codant pour la protéine NOD2, fait partie d'un système de reconnaissance des pathogènes (voir [2]).

rassemble et réanalyse les données de 6 GWAS portant au total sur plus de 6 000 malades et 15 000 témoins et ayant examiné chez chacun d'eux près d'un million de Snip (*single nucleotide polymorphism*). Grâce à la puissance statistique accrue que permet ce regroupement, l'étude identifie 30 nouveaux locus, en confirme 9 publiés entre temps et porte ainsi le total à 71. Un rêve - ou un cauchemar ? - de généticien : ces locus sont certes repérés de manière rigoureuse, comme ceux des étapes précédentes qui ont à chaque fois été retrouvés lors des analyses ultérieures, mais que peuvent-ils - compte tenu de leur nombre - nous apprendre sur la maladie et sur les mécanismes de l'hérédité ?

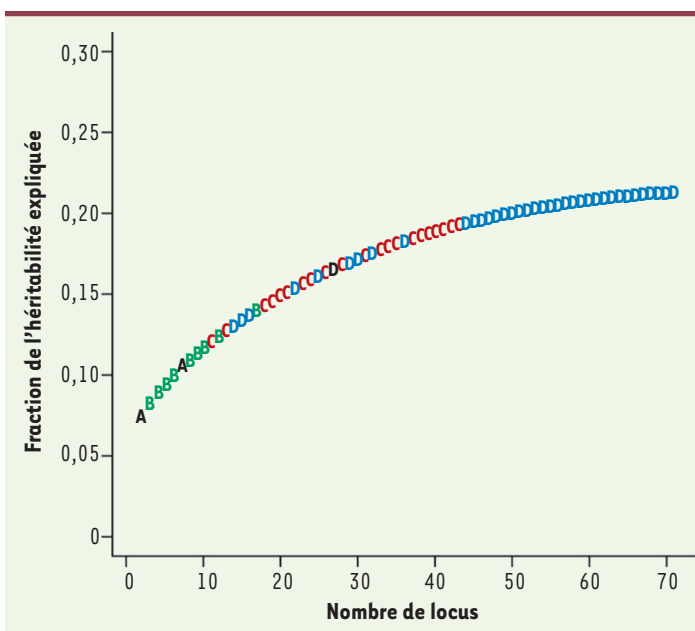
### Avancées et limites de l'interprétation fonctionnelle portant sur 71 locus

Notons d'abord que les gènes eux-mêmes ne sont pas toujours connus : certaines des régions désignées par les GWAS ou méta-GWAS en contiennent plusieurs, et il s'agit aussi de définir lesquels parmi eux sont les meilleurs candidats. Quoiqu'il en soit, le nombre même de locus identifiés confirme bien la complexité de l'étiologie de la maladie et indique que l'interprétation des résultats ne sera pas simple. Au-delà des considérations que l'on peut développer individuellement sur chacun de ces gènes, sur leur fonction et leur implication possible dans la pathogenèse de l'affection - considérations qu'il n'est évidemment pas possible de présenter ici -, les auteurs font une tentative intéressante pour déceler des relations fonctionnelles entre ces différentes entités. On peut en effet espérer simplifier l'interprétation s'il s'avère que plusieurs gènes sont impliqués dans une même voie métabolique afin, en somme, de passer de 71 locus à une dizaine de voies impliquant chacun 5 à 10 gènes. Pour évaluer globalement les relations fonctionnelles, les auteurs utilisent l'outil informatique GRAIL (*gene relationships across implicated*

*loci*<sup>2</sup>) qui détecte les relations entre gènes grâce à la fréquence de leurs cocitations dans les *abstracts* présents sur *PubMed*. Ils trouvent ainsi des liaisons significatives<sup>3</sup> entre une bonne vingtaine d'entre eux. Ils emploient aussi les données d'expression ainsi que celles - encore préliminaires - du *1 000 genomes project*<sup>4</sup> pour affiner leur choix et s'approcher d'interprétations fonctionnelles. Sans, encore une fois, entrer dans les détails, on peut noter que beaucoup des gènes ainsi désignés sont déjà connus pour leur implication dans d'autres maladies inflammatoires ou d'étiologie auto-immune et que les mutations en cause, lorsqu'elles sont connues, touchent plutôt des zones impliquées dans la régulation que des régions codantes.

### Que devient l'héritabilité manquante ?

Sur un plan plus général, l'un des intérêts de cet article est de montrer ce que l'on peut attendre des GWAS pour résoudre le problème persistant de l'« héritabilité manquante » [6], à savoir le fait que l'ensemble des locus identifiés ne rend compte que d'une partie de l'héritabilité pour chaque maladie complexe étudiée. Rappelons que la première méta-analyse portant sur la maladie de Crohn, publiée en 2008, arrivait à 32 locus et 20 % seulement de l'héritabilité. Où en sommes-nous maintenant avec nos 71 locus ? Naïvement, on pourrait espérer qu'en ayant plus que doublé le nombre de gènes nous allons arriver à 40 %, ou au moins à 30-35 %. La réalité est cruelle : la nouvelle valeur est de 23,2 % seulement ! En d'autres termes, les nombreux nouveaux locus détectés ont chacun un effet extrêmement faible - quoique



<sup>2</sup> <http://www.broadinstitute.org/mpg/grail/>

<sup>3</sup> Par rapport à l'analyse de données simulées dans lesquelles les gènes sont choisis au hasard.

<sup>4</sup> <http://www.1000genomes.org/>

**Figure 1. Contributions des 71 locus identifiés à l'héritabilité de la maladie de Crohn.** En abscisse, nombre de locus, en ordonnée, fraction cumulative de l'héritabilité. Les locus sont rangés dans l'ordre de leur contribution à l'héritabilité, qui reproduit à peu près l'ordre chronologique de leur découverte. Voir le texte pour plus de détails. **A.** Locus détectés par les études pré-GWAS (deux en tout). **B.** Locus supplémentaires (9) détectés par les premières études GWAS. **C.** Locus supplémentaires (21) détectés par la première méta-GWAS. **D.** Locus supplémentaires (39) détectés par la deuxième méta-GWAS (adapté de la figure 2 de [5]).



démonstrable et statistiquement significatif – et, à eux tous, n'expliquent que 3,2 % de l'influence génétique. Cela apparaît clairement sur la *Figure 1* qui présente l'ensemble de ces 71 locus, rangés par ordre d'influence sur l'héritabilité, ordre qui se trouve être à peu près celui de leur découverte. Ce n'est naturellement pas un hasard : les gènes qui ont l'effet le plus important sont aussi les plus faciles à mettre en évidence. On voit que parmi les premiers se trouvent les deux locus pré-GWAS (*NOD2* et *IBD5*), suivis et entourés par les 9 qui découlent des GWAS de première génération. Viennent ensuite (toujours dans l'ordre d'influence décroissante) les 21 de la première méta-analyse, et enfin (avec un certain mélange au début) les 39 ajoutés lors de la publication de l'article discuté dans cette chronique. Au vu de cette courbe assez décourageante, on voit que l'information supplémentaire apportée est fort limitée et que l'intérêt de ces nouveaux locus en termes de diagnostic et même d'interprétation de l'affection risque de s'avérer très faible. Notons que dans l'article de Franke *et al.* [5] comme dans les analyses précédentes, rien n'indique une interaction entre ces différents déterminants qui pourrait augmenter leur effet combiné. En somme, le mystère de l'héritabilité manquante subsiste. L'article montre même une extrapolation à 20 000 allèles communs indépendants (!) de la tendance observée qui aboutirait tout au plus à expliquer 50 % de l'héritabilité !

### Que faire ?

Bien naturellement, les auteurs proposent des hypothèses pour expliquer cette discordance. Curieusement, ils mentionnent à peine la contribution de mutations rares non détectées par les analyses pratiquées à l'aide de puces, qui rend vraisemblablement compte d'une bonne partie de l'héritabilité manquante [6]. Ils préfèrent invoquer l'idée que les Snip seraient des indicateurs imparfaits (*imperfect proxies*) pour des mutations causales dans les locus impliqués et que, de ce fait, ils sous-estimeraient le risque relatif lié à certains allèles.

Ils se réfèrent pour cela à un travail récent sur l'héritabilité de la taille [7], selon lequel l'héritabilité manquante serait une sorte d'artéfact lié aux méthodes statistiques employées. Voilà une hypothèse fort intéressante, dont il va falloir suivre attentivement le devenir : si ces derniers auteurs ont raison – leurs conclusions semblent très discutées –, cela signifierait que le monde de la génétique humaine s'angoisse depuis deux ou trois années sur un faux problème. En attendant, et de manière plus orthodoxe, l'article évoque l'intérêt de mener des travaux sur des patients regroupés en sous-phénotypes (ce que certains appellent endophénotypes) afin de détecter des associations génétiques plus fortes et plus significatives. Quoiqu'il en soit, il me semble que GWAS et méta-GWAS, sous leur forme actuelle, ont trouvé leurs limites : il existe certainement de meilleurs investissements en recherche que d'ajouter à grand-peine (et grands frais) quelques points supplémentaires sur la quasi-asymptote de la courbe (*Figure 1*) reproduite dans cette chronique. ♦

### Crohn's disease: from GWAS to meta-GWAS

#### CONFLIT D'INTÉRÊTS


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

#### RÉFÉRENCES

1. Tysk C, Lindberg E, Järnerot G, Flodérus-Myrhed B. Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut* 1988 ; 29 : 990-6.
2. Schreiber S, Rosenstiel P, Albrecht M, *et al.* Genetics of Crohn disease, an archetypal inflammatory barrier disease. *Nat Rev Genet* 2005 ; 6 : 376-88.
3. Mathew CG. New links to the pathogenesis of Crohn disease provided by genome-wide association scans. *Nat Rev Genet* 2008 ; 9 : 9-14.
4. Barrett JC, Hansoul S, Nicolae DL, *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 2008 ; 40 : 955-62.
5. Franke A, McGovern DPB, Barrett JC, *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 2010 ; 42 : 1118-25.
6. Jordan B. À la recherche de l'héritabilité perdue... *Med Sci (Paris)* 2010 ; 26 : 541-3.
7. Yang J, Benyamin B, McEvoy BP, *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010 ; 42 : 565-9.

#### TIRÉS À PART

B. Jordan



**Tarifs d'abonnement M/S - 2011**

**Abonnez-vous**

**à Médecine/Sciences**

**> Grâce à m/s, vivez en direct les progrès des sciences biologiques et médicales**

---

**Bulletin d'abonnement**

**page 253 dans ce numéro de m/s**

