

Chroniques génomiques

Du bon usage des exomes

Bertrand Jordan



Marseille-Nice Génopole,
case 901,
Parc scientifique de Luminy,
13288 Marseille Cedex 9, France.
brjordan@club-internet.fr

Les limites des GWAS

Comme nous l'avons déjà évoqué à plusieurs reprises, l'étude des maladies fréquentes et complexes (diabète, arthrite rhumatoïde, maladie de Crohn, etc.) par association pangénomique (GWAS, *genome-wide association studies*) se heurte actuellement à un sérieux problème, celui de l'« héritabilité manquante » [1, 2]. Les associations dûment démontrées par GWAS s'avèrent correspondre à des risques relatifs faibles (typiquement de 1,2 à 1,5 au maximum) et, cumulées, ne rendent compte que d'une petite partie de l'héritabilité observée pour chacune de ces affections. Les insertions, duplications et délétions, dont la découverte est relativement récente, rendent sans doute compte d'une fraction notable de l'influence génétique ; mais il semble que le rôle le plus important soit joué par les variants rares, correspondant aux snip dont l'allèle mineur a une fréquence très inférieure à 1 %. Les puces à ADN employées pour les études GWAS sont, par construction, limitées aux variants fréquents : il faudrait sinon qu'elles examinent non plus 500 000 ou un million de snip dans le génome, mais des centaines de millions voire des milliards, ce qui est techniquement hors de portée.

Du coup, il ne reste plus guère que le séquençage pour détecter ces variants rares : on imagine que, dans les affections complexes, chacun d'eux exerce une influence notable sur le risque de maladie sans, bien sûr, en être le déterminant principal - sinon on retomberait dans le cas des maladies monogéniques et mendéliennes. Mais, malgré les fulgurants progrès enregistrés au cours des deux ou trois dernières années, une étude GWAS fondée sur la séquence ne constitue

pas une perspective réaliste : le séquençage complet d'un ADN humain coûte au minimum quelques milliers d'euros [3], et il faudrait pratiquer cette analyse sur des milliers de patients et de témoins pour arriver à une puissance statistique suffisante. De plus, comme il s'agit de détecter des variants rares, la qualité de cette séquence doit être excellente : il ne s'agit pas de confondre erreurs de séquençage et variants vrais ! La multiplicité généralement adoptée¹, de l'ordre de trente, risque fort d'être insuffisante, et il faudrait aller jusqu'à cent ou deux cents, ce qui augmente encore le coût. Les premiers résultats du *1 000 genomes project* [4] sont à la fois encourageants et décevants, dans la mesure où la détection systématique d'allèles réellement rares reste hors de portée : on se trouve donc face à une impasse, peut-être temporaire mais néanmoins extrêmement frustrante.

La montée en puissance des exomes

La tactique des exomes [5], sur laquelle je m'étais montré assez sceptique par le passé [3], semble, en fait, avoir le vent en poupe. Elle consiste en un séquençage ciblé sur les séquences exprimées, grâce à la sélection de ces dernières dans l'ADN total du prélèvement à l'aide soit de puces à ADN contenant l'ensemble des séquences exprimées, soit d'une hybridation en solution avec les mêmes séquences munies d'une étiquette permettant la sélection des complexes formés après incubation. Ces méthodes fonctionnent très correctement et permettent de récupérer, avant séquençage,

¹ Nombre de « passages », nombre de fois (en moyenne) où est lue chaque base.

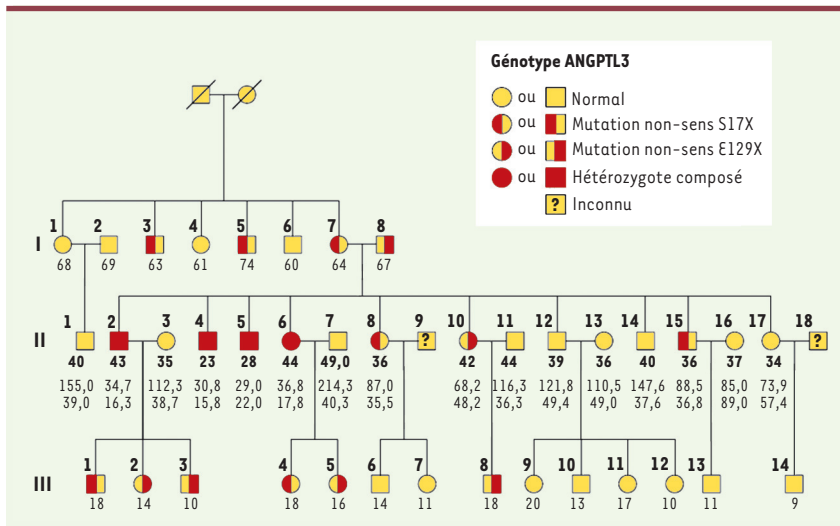


Figure 1. Arbre généalogique de la famille sur laquelle a porté l'étude de Musunuru et al. [6]. Le statut de chaque personne du point de vue du gène *ANGPTL3* a été déterminé pour 3 générations (I, II et III). On révèle ainsi de nombreux hétérozygotes (voir légende pour les génotypes), qui présentent en général des taux bas de LDL, ainsi que quatre homozygotes composés, les deux (II-4 et II-5) dont l'exome a été séquencé et les deux autres patients déjà repérés (II-2 et II-6). Les chiffres sous chaque symbole indiquent l'âge au moment de l'étude, puis, plus bas, la concentration en cholestérol LDL et HDL (en milligrammes par 100 ml) (extrait remanié de la figure 1 de [6]).

un échantillon d'ADN qui est majoritairement constitué de séquences exprimées - lesquelles, rappelons-le, constituent moins de 1% de l'ADN de départ. Cette tactique diminue considérablement le coût du séquençage, le ramenant à quelques centaines d'euros par personne, et autorise donc des études à grande échelle. Notons que, ce faisant, on se prive de toute information sur les séquences non exprimées, qui peuvent renfermer des éléments de régulations impliqués dans telle ou telle pathologie - mais c'est pour le moment le prix à payer afin d'arriver à un projet réalisable.

L'hypolipidémie, ou l'exploitation intelligente d'un cas limite

Un tout récent article du *New England Journal of Medicine* [6] montre comment cette approche peut élucider un déterminant génétique d'une affection complexe. Il s'agit dans ce cas de l'hypolipidémie familiale, qui se manifeste par un taux de cholestérol LDL (*low-density lipoprotein*) très bas entraînant différentes manifestations cliniques (stéatose hépatique, malabsorption des lipides, etc.). Cette affection présente un caractère héréditaire marqué, et une partie des cas a pu être reliée à des mutations dans le gène *APOB* qui code pour l'apolipoprotéine B, impliquée dans la sécrétion de LDL dans le foie. Mais dans certaines familles affectées ce gène est apparemment fonctionnel, et aucun locus n'a été désigné de manière concluante par les analyses génétiques. La stratégie suivie par les auteurs a consisté à se focaliser, dans une grande famille présentant quatre cas d'hypolipidémie, sur les deux cas les plus extrêmes, deux garçons d'une fratrie de douze enfants (Figure 1), et à analyser en détail l'ensemble de leurs deux exomes. L'idée, bien sûr, est que s'il existe quelque part dans le génome un variant causal, celui-ci doit être le même chez ces deux patients - et ne doit pas, *a priori*, être retrouvé chez des témoins normaux. Notons qu'il s'agit bien d'une approche globale, ne faisant aucune hypothèse sur la nature ou la localisation du ou des gènes recherchés, et non d'une tactique de type « gène candidat » dont on connaît les faiblesses.

Dix-huit mille variants - et un gène identifié !

Les auteurs ont donc isolé, par une technique d'hybridation en solution [7], les exomes (165 000 exons représentant 16 000 gènes) des deux patients ainsi que ceux de soixante témoins non apparentés mais provenant d'une population comparable. Ces exomes ont été séquencés de manière très exhaustive en accumulant six gigabases de séquence pour chacun d'eux, ce qui correspond à une multiplicité élevée : chaque base a été lue deux cents fois, ce qui réduit le taux d'erreurs à une valeur très faible. Dix-huit mille variants ont été repérés en comparant cette soixantaine de séquences : le problème était alors de déterminer lesquels étaient susceptibles d'être impliqués dans l'affection. C'est notamment à cela qu'ont servi les soixante exomes témoins : en éliminant les variants qui y sont trouvés, plus ceux qui figurent dans les bases de données génomiques, il ne restait plus que trois cents variants. Et un seul d'entre eux est présent chez les deux frères ! Ce variant désigne donc très vraisemblablement le gène qui joue un rôle central dans ces deux cas - et son examen le confirme : il s'agit du gène *ANGPTL3* (*angiopoietin-like 3*), dont le produit est exprimé principalement dans le foie et est connu pour réguler le niveau des triglycérides chez la souris : différentes indications permettent de lui attribuer également un rôle dans la régulation du niveau de cholestérol - fonction qui n'était pas connue chez l'homme, mais qui apparaît hautement vraisemblable.

Chez ces deux patients, la séquence du gène présente en fait deux mutations non-sens. Une étude a alors été effectuée sur l'ensemble de la famille, qui comporte trois générations, avec, chez de nombreux individus, des anomalies plus ou moins marquées au niveau des LDL.



Elle révèle (Figure 1) plus de dix hétérozygotes : cela indique que les deux mutations repérées se trouvent en fait sur deux exemplaires différents du gène, les deux cas extrêmes étudiés correspondant à des homozygotes composés. Les deux autres patients ont le même génotype ; quant aux hétérozygotes, ils ont généralement des niveaux de LDL bas (la normale est d'environ 130 mg/100 ml), ce qui confirme encore le rôle causal des variants de ce gène dans cette famille (Figure 1).

Ce très joli travail d'investigation scientifique a des conséquences pratiques : le diagnostic génétique pratiqué dans cette famille (et d'autres ayant des caractéristiques proches) permet une détection précoce des hétérozygotes (et bien sûr des homozygotes) et la mise en oeuvre de mesures palliatives ; de plus, la protéine ANGPTL3 étant sécrétée par le foie et présente dans le sang, cela suggère une approche thérapeutique nouvelle pour diminuer le niveau de cholestérol LDL lorsqu'il est excessif.

Séquencer - mais à bon escient !

Depuis quelques mois, nous voyons paraître les résultats de séquençages globaux pratiqués dans des situations pathologiques - mais ceux-ci ont parfois apporté plus de confusion que de clarté : l'analyse de cellules tumorales révèle littéralement des millions de différences par rapport au tissu normal, et il s'avère fort difficile de démêler lesquelles sont importantes du point de vue fonctionnel [8]. L'exemple rapporté ici montre comment le choix de cas extrêmes au sein d'une affection, l'emploi raisonné du séquençage (focalisation sur l'exome,

multiplicité très élevée pour assurer une excellente qualité) et l'importance accordée aux témoins (qui ont ici permis d'éliminer près de 90 % des différences observées) autorise des conclusions solides et ayant une réelle utilité clinique. On voit ainsi que l'intelligence et le professionnalisme dans la conception des stratégies expérimentales gardent toute leur importance, y compris pour ces approches à grande échelle que certains voudraient qualifier de pur exercice technologique. ♦

Making good use of exome sequencing

CONFLIT D'INTÉRÊTS

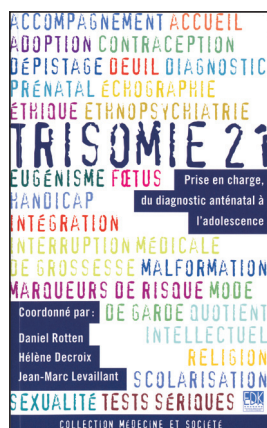
L'auteur déclare n'avoir aucun conflit d'intérêts concernant les données publiées dans cet article.

RÉFÉRENCES

1. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009 ; 461 : 747-53.
2. Jordan B. À la recherche de l'héritabilité perdue. *Med Sci (Paris)* 2010 ; 26 : 541-3.
3. Jordan B. Séquence : le juste prix... *Med Sci (Paris)* 2009 ; 25 : 988-9.
4. 1000 genomes project consortium, Durbin RM, Abecasis GR, Altshuler DL, et al. A map of human genome variation from population-scale sequencing. *Nature* 2010 ; 467 : 1061-73.
5. Maher B. Exome sequencing takes centre stage in cancer profiling. *Nature* 2009 ; 459 : 146-7.
6. Musunuru K, Pirruccello JP, Do R, et al. Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N Engl J Med* 2010 ; 363 : 2220-7.
7. Gnirke A, Melnikov A, Maguire J, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 2009 ; 27 : 182-9.
8. Pleasance ED, Cheetham RK, Stephens PJ, et al. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 2010 ; 463 : 191-6.

TIRÉS À PART

B. Jordan



ISBN : 2-84254-105-7 248 pages

Bon de commande

À retourner à EDK, 2, rue Troyon - 92316 Sèvres Cedex
Tél. : 01 55 64 13 93 - Fax : 01 55 64 13 94 - E-mail : edk@edk.fr

NOM : Prénom :

Adresse :

Code postal : Ville :

Pays :

Fonction :

Je souhaite recevoir l'ouvrage **Trisomie 21** : 15 € + 3 € de port = **18 € TTC**

en exemplaire, soit un total de €

Par chèque, à l'ordre de EDK

Par carte bancaire : Visa Eurocard/Mastercard

Carte n° | | | | | | | | | | | | | | | | | | | | | |

Signature :

Date d'expiration : | | | | | |

N° de contrôle au dos de la carte : | | | |



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

Chaque mois, avec les articles de référence de M/S

Chaque jour, sur www.medecinesciences.org



Médecine/Sciences

est indexé dans PubMed/Medline

Current Contents, série Life Sciences
EMBASE/Excerpta Medica
PASCAL
CABS
BIOSIS

- > Des articles rédigés par des médecins et des chercheurs reconnus sur la scène internationale qui posent avec rigueur les bases des débats scientifiques.
- > Des synthèses, éditoriaux, dossiers techniques et analyses toujours replacés dans leur contexte pour que l'information soit la plus exacte, intelligible et objective.
- > La dimension humaine privilégiée, avec l'analyse des retombées diagnostiques, thérapeutiques, la prévention et l'éthique liées aux nouvelles avancées.

- > Un panorama clair et concis de l'actualité scientifique : des nouvelles, des brèves, des données chiffrées, des repères et perspectives pour qu'aucun fait significatif ne vous échappe.



Tarifs d'abonnement M/S - 2010 Mensuel - 10 numéros/an

Abonnez-vous à Médecine/Sciences

Mon règlement :

Par mail edk@edk.fr

Uniquement pour les paiements par carte bancaire

Par fax en envoyant ce bulletin au 01 55 64 13 94

Uniquement pour les paiements par carte bancaire

N°

Date d'expiration Signature :

N° de contrôle au dos de la carte

Par chèque à l'ordre de Médecine/Sciences, en envoyant ce bulletin à :

Éditions EDK

2, rue Troyon

92316 Sèvres Cedex, France

Pour recevoir une facture, cochez cette case

Je souhaite m'abonner à M/S :

Nom : Prénom :

Adresse :

Code postal Ville :

Pays :

E-mail-obligatoire :

Je choisis l'abonnement :

	Particuliers		Institutions		Étudiants*		Enseignants*	
	Papier + Électronique	Électronique seul	Papier + Électronique	Électronique seul	Papier + Électronique	Électronique seul	Papier + Électronique	Électronique seul
France	<input type="checkbox"/> 185 €	<input type="checkbox"/> 125 €	<input type="checkbox"/> 415 €	<input type="checkbox"/> Sur devis	<input type="checkbox"/> 95 €	<input type="checkbox"/> 75 €	<input type="checkbox"/> 120 €	<input type="checkbox"/> 100 €
UE + autres	<input type="checkbox"/> 245 €	<input type="checkbox"/> 125 €	<input type="checkbox"/> 505 €	<input type="checkbox"/> Sur devis	<input type="checkbox"/> 135 €	<input type="checkbox"/> 75 €	<input type="checkbox"/> 205 €	<input type="checkbox"/> 100 €

* Joindre un justificatif