

Chroniques génomiques

Nos défauts génétiques cachés

Bertrand Jordan



Depuis que l'on a constaté que beaucoup des maladies héréditaires qui nous affectent sont transmises sur le mode autosomique récessif, on a cherché à savoir combien de telles altérations porte chacun de nous dans son génome. Ces altérations sont prêtes à se révéler en cas de procréation avec un(e) partenaire portant une mutation dans le même gène et entraînent, dans ce cas, un risque de 25 % de voir naître un enfant atteint d'une affection grave, pour laquelle il n'existe souvent aucune thérapie efficace. Depuis Herman Müller, en 1950 [1], jusqu'à des travaux récents examinant plus de 23 000 personnes [3], en passant par des analyses théoriques [2], les estimations ont varié de quelques mutations pathogènes récessives cachées, jusqu'à une dizaine et même une centaine par personne. Mais les plus sérieux de ces travaux étaient ciblés sur un nombre relativement restreint de gènes (au mieux une centaine) et ne pouvaient donc offrir une vision globale de la question. C'est pourquoi l'article que je vais discuter ici [4] est important, dans la mesure où il examine tous les variants détectés par le séquençage des exomes de deux populations, et élimine ainsi nombre de biais dans l'étude.

Une étude de grande ampleur : premiers résultats

Ce travail, mené par un ensemble d'équipes européennes dans lesquelles dominent les laboratoires des Pays-Bas et d'Israël, a porté sur deux cohortes, l'une hollandaise, l'autre estonienne. Après élimination des personnes apparentées et de celles dont l'ascendance n'est pas majoritairement européenne, leur effectif respectif est de 4 120 et 2 327 individus. Les exomes de toutes ces personnes ont été séquençés à haute redondance (couverture 75X)¹ puis analysés pour détecter tous les variants par rapport au génome humain de référence (la séquence humaine « consensus » réguliè-



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

rement actualisée²). L'étude a ensuite été ciblée sur les 1 929 gènes répertoriés dans la base de données OMIM (*On-line Mendelian Inheritance in Man*)³ comme impliqués dans des affections héréditaires de type autosomique récessif. Les phénotypes associés à chacun de ces gènes (lorsqu'ils sont mutés) ont été examinés manuellement pour définir ceux qui sont associés à des phénotypes « sévères »⁴ : cela concerne 1 119 gènes sur les 1 929 étudiés. Après divers contrôles de qualité, environ 91 000 variants sont trouvés dans les séquences des exons concernés de la cohorte hollandaise et 46 000 dans l'estonienne.

Vient alors une étape essentielle, à savoir définir lesquels de ces variants sont pathogènes. On sait en effet que beaucoup de mutations affectant les exons n'ont pas d'effet majeur sur la protéine correspondante, soit qu'ils correspondent à des mutations silencieuses (un codon différent pour le même acide aminé), soit qu'ils concernent une région non critique de la protéine. Les auteurs ont utilisé une série d'approches pour sélectionner les seuls variants pathogènes : prise en compte de la nature des variants (mutations non-sens ou décalant le cadre de lecture), exploitation de bases de données spécialisées comme ClinVar⁵ ou VKGL (*Vereniging Klinisch Genetische Laboratoriumdiagnostiek*)⁶, qui répertorient les mutations pathogènes connues, classification de l'ACMG (*American College of Medical Genetics*). Ils éli-

² voir https://en.wikipedia.org/wiki/Reference_genome

³ www.ncbi.nlm.nih.gov/omim

⁴ Une maladie sévère est définie comme « une affection pour laquelle un couple 'moyen' ferait en sorte d'éviter la naissance d'un enfant atteint »

⁵ www.ncbi.nlm.nih.gov/clinvar/

⁶ www.vkgl.nl/

¹ Dans ce cas, la séquence d'ADN étudiée est lue 75 fois, afin de limiter les erreurs ou « trous » de lecture.

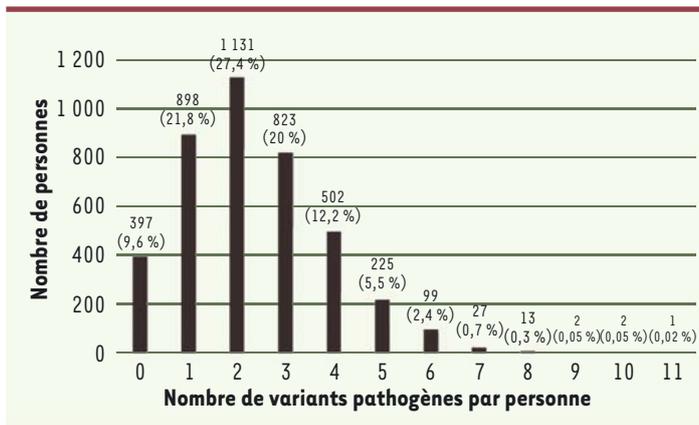


Figure 1. Nombre de variants pathogènes (à l'état hétérozygote) par personne, pour la cohorte hollandaise (extrait partiel et remanié de la figure S2 du supplément de [4]).

minent ainsi plus de 95 % des variations détectées et, après diverses vérifications, arrivent finalement à 3 734 mutations pathogènes pour la cohorte hollandaise et 1 664 pour la cohorte estonienne. La différence des deux nombres tient au fait que la grande majorité de ces mutations sont très rares et ne sont trouvées qu'une seule fois dans l'échantillon – donc plus l'échantillon est grand, plus on détecte de mutations différentes. Pour la même raison, le nombre de mutations communes aux deux populations est assez faible : 373. On peut alors (enfin !) évaluer le nombre moyen de mutations pathogènes récessives par personne : il est de l'ordre de 2 si l'on envisage le jeu complet de gènes (1 929), de 1 à 1,5 si l'on se limite aux phénotypes sévères (1 119). Ce ne sont là que des valeurs moyennes, et le chiffre individuel va de zéro (pour quelques centaines de personnes) à plus de 5 (pour quelques dizaines) (Figure 1). Comme nous le verrons plus loin, ces valeurs sont sans doute sous-évaluées ; retenons toutefois que chacun d'entre nous porte au moins une ou deux mutations récessives pouvant aboutir à une affection grave dans sa descendance, en cas de procréation avec une personne portant une mutation délétère dans le même gène.

Fréquence par gène et nombre de couples à risque

À ce stade, il est aussi possible d'étudier la fréquence des mutations (dans la population considérée), en fonction des gènes envisagés. Le résultat est clair : pour la très grande majorité des 1 929 gènes étudiés, les mutations sont rares, avec une fréquence dans l'échantillon inférieure à 0,5 % (soit une vingtaine de cas où le gène envisagé est muté parmi les 4 120 personnes de la population hollandaise). C'est le cas pour 1 863 gènes sur les 1 929 analysés ; seuls 66 gènes portent une mutation chez un plus grand nombre d'individus, dont une dizaine (y compris, par exemple, le gène *CFTR* [*cystic fibrosis transmembrane conductance regulator*], impliqué dans la mucoviscidose) pour lesquels la fréquence approche ou dépasse 5 % (Figure 2).

En résumé, chaque individu porte une ou plusieurs mutations potentiellement pathogènes, mais les gènes en cause sont généralement différents d'une personne à l'autre, sauf pour quelques dizaines de gènes dont la

mutation est relativement fréquente dans la population. Ce résultat a des conséquences cliniques importantes : on pratique de plus en plus fréquemment, notamment aux États-Unis, un diagnostic pré-conceptionnel (*carrier screening*) pour lequel les deux partenaires d'un couple envisageant d'avoir un enfant se font tester pour savoir si cet enfant risque d'être atteint d'une maladie due à un allèle récessif présent chez les deux personnes [5, 6] (→).

(→) Voir la Chronique bioéthique de B. Baertschi, m/s n° 1, janvier 2019, page 72

Une question se pose alors : combien de gènes différents faut-il tester ? Compte tenu des possibilités techniques actuelles, on pourrait vouloir définir un ensemble très large : une des entreprises commerciales proposant ce type de test, *Invitae*⁷, offre par exemple un test portant sur 47 gènes et un autre qui, lui, en examine 289 (à un prix différent, bien sûr)⁸. Mais les données représentées dans la Figure 2 montrent que cet élargissement est futile : les gènes dont les mutations sont très peu fréquentes (la très grande majorité) ont des chances infinitésimales de se retrouver mutés chez les deux partenaires, et un ensemble de moins de 100 gènes permettra de détecter l'essentiel des couples à risque. L'estimation de la proportion de couples à risque peut être faite à partir de ces résultats. Cela passe par une simulation de tous les appariements possibles entre les membres de chacune des deux populations étudiées, et un comptage des cas où il s'avère que les deux partenaires portent une mutation dans le même gène. Selon qu'on se limite ou non aux gènes pour lesquels le phénotype est sévère, on trouve une fréquence de couples à risque qui va de 0,8 à 1,5 %. En admettant que les deux populations étudiées sont représentatives, on arrive donc à la conclusion qu'au moins 1 % des couples européens sont à risque, c'est-à-dire qu'ils peuvent avoir un enfant atteint d'une affection génétique grave, alors qu'eux-mêmes sont indemnes. La comparaison de cette valeur avec les données de génétique clinique ne contredit pas cette estimation [4]. Plus de 90 % de ces couples à risque sont porteurs de mutations touchant un petit nombre de gènes, 70 pour la population hollandaise dans cette simulation.

Le risque de la consanguinité

On peut aussi examiner la fréquence des couples à risque par rapport aux différents phénotypes, en

⁷ www.invitae.com/en/individuals/reproductive-genetic-testing/carrier-screening/

⁸ Une autre entreprise, FullGenomics (siège près de Pasadena en Californie), teste 277 gènes (2 405 mutations testées, correspondant à 348 maladies, autosomiques récessives) pour la PMA en Espagne.

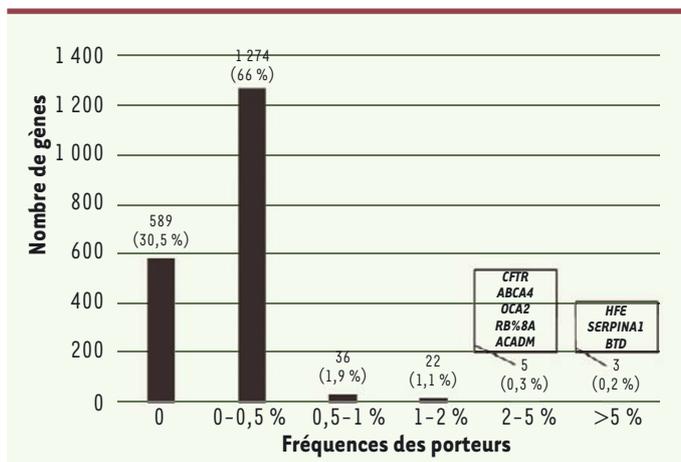


Figure 2. Fréquences, dans la cohorte hollandaise, des porteurs hétérozygotes d'allèles pathogènes, selon le gène envisagé. On voit que, pour la grande majorité des gènes (1863 sur 1929), la fréquence de ces porteurs est très faible (< 0,5 %) ou nulle, sauf pour un petit nombre de gènes (partie droite de la figure), parmi lesquels est présent le gène *CFTR* impliqué dans la mucoviscidose. Les gènes pour lesquels la fréquence indiquée est nulle n'ont été trouvés sous forme mutée que dans la cohorte estonienne (extrait partiel et remanié de la figure S2 du supplément de [4]).

regroupant les gènes selon l'organe ou la fonction que la mutation affecte. La *Figure 3A* montre le nombre total de couples à risque (pour 100 000 personnes), soit environ 1,5 %, et sa répartition selon les phénotypes considérés. On voit notamment que les maladies métaboliques sont les plus représentées : avec les affections de la vue, elles regroupent 79 % des couples à risque. Mais ce qui précède suppose que les individus de la population ne sont pas apparentés, et le calcul donne des résultats très différents pour des cousins au premier degré (« cousins germains ») : l'on a alors une augmentation considérable du nombre des couples à risque, qui représentent maintenant 25 % de la population. Le risque de révélation d'une maladie génétique récessive est donc très fortement augmenté chez ces personnes (16 fois environ, 24 895 couples à risque sur 100 000 pour la cohorte hollandaise au lieu de 1554). La répartition des phénotypes change également de manière très sensible, comme le montrent la *Figure 3B* (fréquence des couples à risque selon les affections) et la *Figure 3C* (rapport des fréquences de couples à risque entre couples consanguins et non consanguins selon les affections).

On note, en particulier, la forte augmentation de la fréquence de couples à risque pour des déficiences intellectuelles et des anomalies du squelette en cas de consanguinité. Cela s'explique sans doute par le fait que ces affections sont en fait très multigéniques [7], mettant en jeu bien d'autres gènes que celui que l'on examine : deux personnes apparentées ont plus de chances de porter de larges segments de chromosome identiques et donc des allèles également identiques pour d'autres gènes impliqués dans le phénotype. Les auteurs évoquent également la possibilité que la présence d'une mutation à l'état hétérozygote dans cette catégorie de gènes diminue légèrement le succès

reproducteur (*fitness*) de l'individu, ce qui tendrait à réduire la fréquence de ces hétérozygotes dans la population et à exacerber l'effet de la consanguinité.

Quoi qu'il en soit, la consanguinité est bien un important facteur de risque, surtout si elle est étroite : pour les cousins au second degré (cousins issus de germains partageant les mêmes arrière-grands-parents), l'augmentation globale de la fréquence des couples à risque n'est plus que d'un facteur 3 environ, et elle est imperceptible pour les cousins au troisième degré. Et cette fois, les gènes impliqués sont nombreux : il faut 749 gènes pour rendre compte de 90 % des couples à risque. Cette différence est importante sur le plan clinique : elle indique qu'un test préconceptionnel examinant un petit nombre de gènes, suffisant pour la population générale, ne donnera pas des résultats fiables si les deux partenaires sont apparentés.

Des conclusions importantes... et assez solides

Ce travail apporte des informations importantes sur un sujet qui a fait par le passé l'objet de publications parfois contradictoires. Il montre d'abord que nous possédons presque tous, dans notre génome, une ou plusieurs mutations récessives délétères susceptibles d'aboutir à un enfant affecté en cas de procréation avec une personne portant une mutation dans le même gène. Contrairement aux études précédentes, qui s'intéressaient à un jeu limité de gènes, cette conclusion est fondée sur l'analyse de l'ensemble de l'exome de milliers de sujets européens. On peut certes penser qu'étendre l'étude à l'analyse des génomes entiers aurait permis de détecter quelques mutations supplémentaires (dans les introns, ou dans des séquences de régulation) : le résultat obtenu ici est certainement un peu sous-évalué de ce fait. Naturellement, séquencer le génome entier (trois gigabases) pour des milliers d'individus est une entreprise bien plus longue et onéreuse que de se limiter à leur exome (une vingtaine de mégabases). En tout état de cause, le séquençage systématique qui est en cours pour les ADN de banques comme la *UK Biobank*⁹ va bientôt per-

⁹ www.ukbiobank.ac.uk

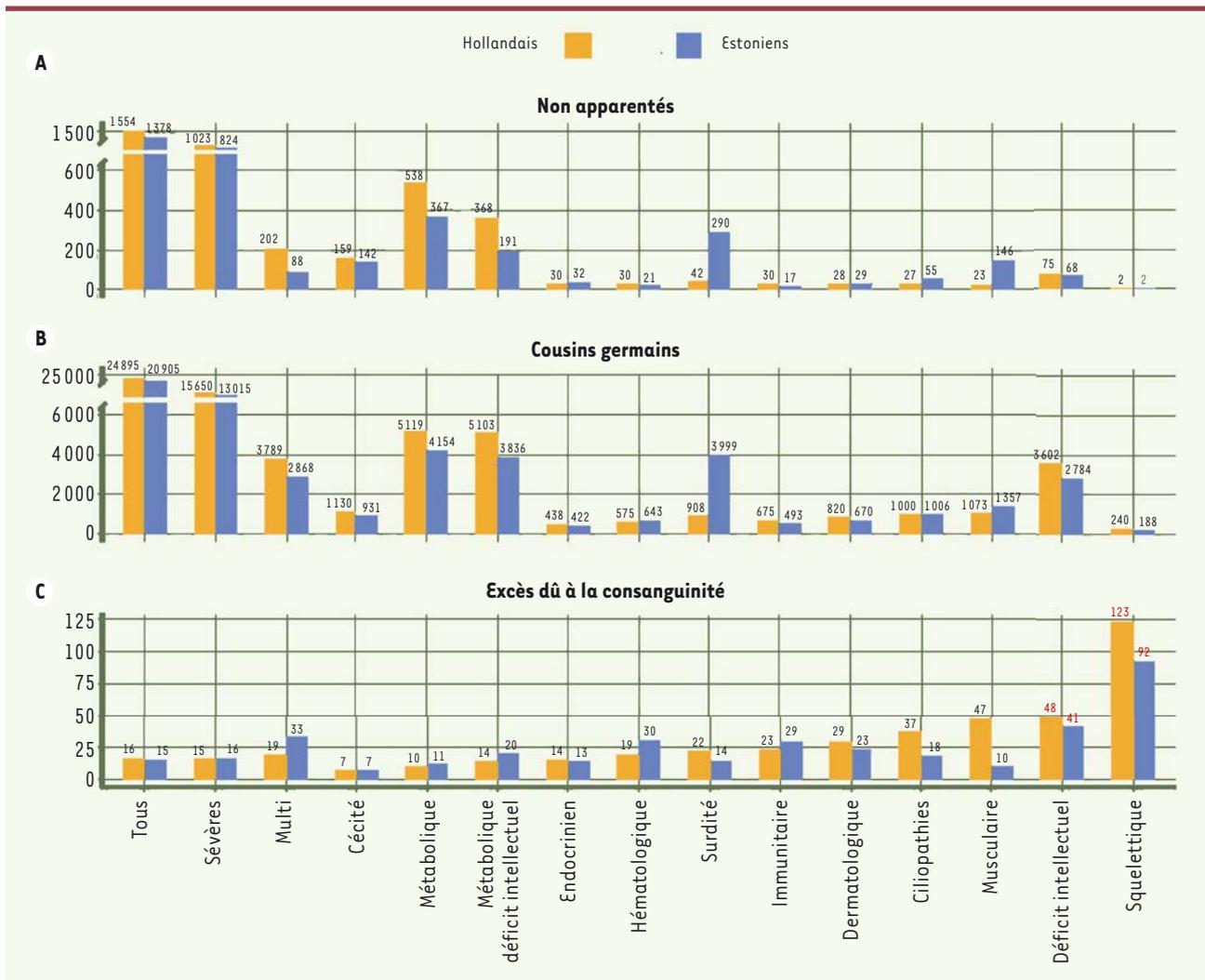


Figure 3. Fréquence des couples à risque (pour 100 000 couples) en fonction du type de pathologie. Résultats pour les deux cohortes étudiées, en tenant compte des 1 929 gènes définis. En haut (A) : répartition pour un ensemble d'individus non apparentés ; au milieu (B) : pour des cousins au premier degré (noter le changement de l'échelle verticale), montrant une forte augmentation du risque et un changement de la répartition selon les maladies. En bas (C) : augmentation du risque dû à la consanguinité ; noter que l'augmentation est particulièrement forte pour les déficits intellectuels et les troubles squelettiques. Multi : multi-systèmes (extrait remanié de la figure 3 de [4]).

mettre des analyses qui préciseront ce point. En tous cas, ces données permettent d'évaluer la fréquence de couples à risque¹⁰ au sein de la population européenne à 1,5 % environ.

Un deuxième élément important qu'apporte cette étude est que la très grande majorité des mutations délétères détectées sont rares, et que le gène en cause n'est muté que chez un, ou un tout petit nombre d'individus au sein d'une cohorte de plusieurs milliers de personnes. Seuls quelques dizaines de gènes sont fréquemment mutés au sein

d'une telle cohorte : cela implique que lors d'une analyse visant à détecter des couples à risque, il est inutile d'élargir le jeu de gènes examiné au-delà de la centaine.

Dernier point, qui vient d'ailleurs moduler quelque peu le précédent : l'impact de la consanguinité est considérable, multipliant par seize le nombre de couples à risque. Et la consanguinité modifie considérablement la répartition des phénotypes affectés, mettant en jeu un nombre de gènes plus important et imposant cette fois l'examen d'un grand nombre de gènes (près d'un millier) pour atteindre un bon taux de détection des

¹⁰ Rappelons que pour un couple à risque, la probabilité d'avoir un enfant atteint est de 25 %. Le fait de connaître ce risque permet de mettre en place des mesures préventives, diagnostic prénatal ou diagnostic préimplantatoire.

