

Chroniques génomiques

Un cadeau empoisonné

Bertrand Jordan

On sait que la sévérité des infections par le coronavirus SARS-CoV-2 (*severe acute respiratory syndrome-coronavirus-2*), responsable de la COVID-19 (*coronavirus disease 2019*), est très variable, allant d'une absence quasiment totale de symptômes jusqu'à de graves problèmes respiratoires et à la mort. Les facteurs qui induisent une forme sévère de la maladie commencent à être élucidés : le plus connu est l'âge, mais différentes comorbidités, comme l'obésité ou l'hypertension, jouent aussi un rôle important. Même en tenant compte de ces facteurs, il reste d'importantes différences dans l'évolution de patients par ailleurs comparables, ce qui amène à s'interroger sur la possible intervention de facteurs génétiques influençant la réponse à l'infection. L'existence de tels facteurs a été démontrée pour différentes maladies infectieuses (hépatites, dengue, tuberculose, lèpre, etc.) [1] ; il est donc logique de les rechercher pour l'infection par le SARS-CoV-2. Comme nous allons le voir, un locus situé sur le chromosome 3 est effectivement impliqué dans la sévérité de l'infection [2] et, curieusement, l'allèle pathogène de ce locus semble provenir d'un lointain croisement avec un de nos cousins néandertaliens [3].

Un locus de susceptibilité

Deux études concordantes impliquent un locus situé sur le chromosome 3 pour le risque d'infection sévère. Il s'agit bien sûr de « balayages du génome » selon l'approche désormais classique dite GWAS (*genome-wide association study*). La première, effectuée par un large ensemble de laboratoires européens regroupés au sein du *Severe Covid-19 GWAS group*, a concerné 1 610 patients provenant d'Italie et d'Espagne et souffrant d'une infection sévère, et 2 205 témoins non infectés [2]. Une infection sévère est définie par l'existence de problèmes respiratoires ayant nécessité l'administration d'oxygène (avec ou sans intubation). La *Figure 1* montre le résultat de cette étude.

Le pic très net au niveau du chromosome 3 (plus précisément, en 3p21.31) correspond à un segment d'environ 50 kilobases (kb) contenant six gènes ; tous les



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

SNP (*single nucleotide polymorphism*) contenus dans cette région sont en très fort déséquilibre de liaison, indiquant que les mêmes allèles de ces SNP restent associés sans recombinaison (*Figure 2*).

Ce locus est associé à un risque relatif de 1,77, c'est-à-dire que les porteurs des allèles « pathogènes » (*risk alleles*) ont un risque augmenté de 77 % de présenter une forme sévère de l'infection, par rapport à la moyenne. Il est bien sûr tentant d'imaginer quel rôle peut jouer l'un ou l'autre des six gènes de ce locus dans l'apparition d'une forme sévère de la maladie : les auteurs évoquent les récepteurs de chimiokines CCR9 (*C-C motif chemokine receptor 9*) et CXCR6 (*C-X-C motif chemokine receptor 6*), ce dernier régulant la localisation des lymphocytes T CD8⁺ dans les voies respiratoires [4]. Par ailleurs le gène *SLC6A20* (*solute carrier family 6 member 20*) code le transporteur SIT1 (*signaling threshold regulating transmembrane adaptor 1*), qui interagit avec la protéine ACE2 (*angiotensin-converting enzyme 2*) qui est le récepteur cellulaire du SARS-CoV-2 [5]. Voilà donc des pistes intéressantes, qui font sûrement l'objet de recherches actives. Ces résultats sont confirmés par les données obtenues par la *COVID-19 Host Genetics Initiative* [6]. Il s'agit là d'un consortium mondial, dont l'objectif est d'« élucider le rôle de facteurs génétiques de l'hôte dans la susceptibilité et la sévérité » de la pandémie. L'article cité en référence [6] expose les raisons et les objectifs de cette initiative, mais beaucoup de données sont déjà disponibles sur un site dédié¹ qui présente notamment les résultats de GWAS effectuées dans différentes combinaisons (patients infectés versus témoins, hos-

¹ <https://www.covid19hg.org/results/>



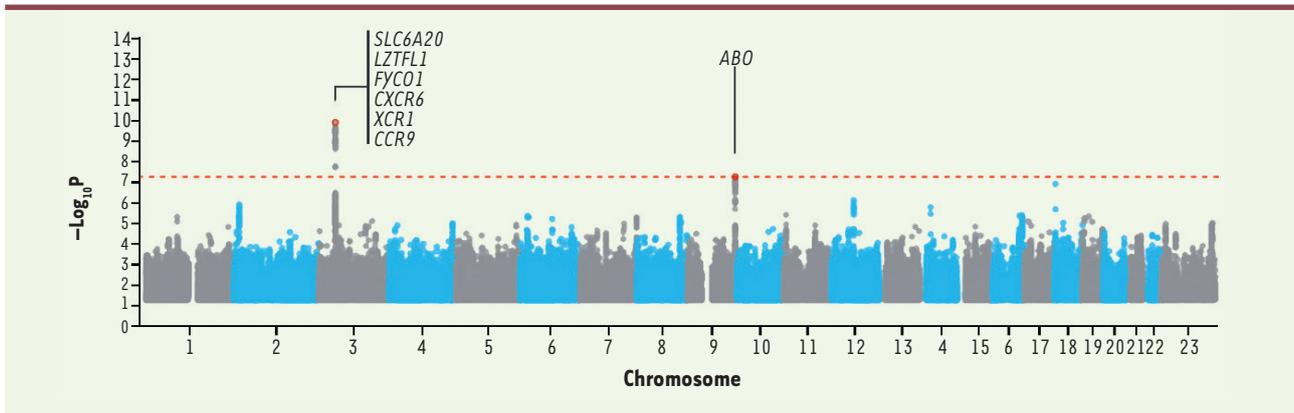


Figure 1. Manhattan plot de l'étude GWAS pour le risque d'infection sévère par le SARS-CoV-2 (severe acute respiratory syndrome-coronavirus-2). La ligne pointillée rouge indique le seuil de significativité. On voit un pic très net sur le chromosome 3 et un pic à la limite du seuil significatif sur le chromosome 9. Les gènes correspondants sont indiqués (extrait modifié de la figure 2 de [2]).

pitalisés *versus* témoins, cas très sévères (intubés) *versus* témoins, etc.). Les images varient un peu, mais le pic majeur de tous ces *Manhattan plots* se situe toujours en 3p, confirmant ainsi les données précédentes. Tout récemment, une nouvelle étude, encore à l'état de *preprint* [7], a encore confirmé le locus majeur sur le chromosome 3, tout en révélant plusieurs locus additionnels. Dans cette étude, qui porte sur des patients en état critique (dont 73 % ont été placés en réanimation), le risque relatif associé à l'allèle de risque est de 2,14. Parmi les autres locus (qui correspondent à des risques relatifs plus faibles), on note la présence d'un gène codant un récepteur de l'interféron et d'autres impliqués dans les inflammations pulmonaires. Toutes ces données demandent à être encore confirmées, mais elles montrent que les balayages du génome peuvent apporter des informations importantes sur les mécanismes à l'œuvre dans la réponse à l'infection et donc, à terme, ouvrir des pistes thérapeutiques.

Le cadeau d'un Néandertalien

L'article suivant dans cette histoire, est celui de Zeberg et Pääbo, qui a eu l'honneur rare d'une pré-publication dans la revue *Nature* [3]. Il faut dire que le titre est alléchant : « *Le principal facteur de risque pour la COVID-19 sévère est hérité des Néandertaliens* »². Publié le 30 septembre 2020 en tant que *Accelerated article preview* [3], cet article commence par reprendre la localisation du facteur de risque à partir des dernières données de la *Covid-19 Host Genetics Initiative* [6], en incluant dans l'analyse 3 199 patients « sévères »³ et 897 488 témoins. Dans ces conditions, on obtient un résultat très clair (*Figure 3*) : le seul locus significatif est (encore) situé en 3p. Le risque relatif pour les porteurs des allèles de risque est ici évalué à 1,6, et l'on retrouve la région d'environ 50 kb et la zone plus large d'environ 330 kb, comme dans la *Figure 2*.

Les données de séquence obtenues dans le cadre de *1 000 Genomes Project* [8] permettent immédiatement de déterminer la fréquence de l'allèle de risque⁴ de ce locus dans différentes populations humaines, et l'on constate que sa distribution est très inégale et qu'il est notamment absent en Afrique (*Figure 4*).

Cette distribution⁵, jointe à l'intérêt de longue date du groupe de Pääbo pour l'ADN de Néandertal, l'a conduit à examiner si l'allèle de risque de ce locus pouvait provenir de croisements anciens entre notre espèce et l'homme de Néandertal [9] (→).

(→) Voir la Chronique génomique de B. Jordan, m/s n° 4 avril 2020, page 421

De fait, la séquence d'un Néandertalien provenant de Croatie et remontant à environ 50 000 ans [10], est quasiment identique (11 sur 13 nucléotides) à l'allèle de risque pour le pic principal (*Figure 2*), et qui plus est à l'état homozygote ! Deux autres séquences de Néandertal contiennent aussi des allèles de risque, qui sont en revanche absents dans les séquences de Denisoviens, autres cousins d'*Homo sapiens*. Cette similitude ne prouve pas que l'allèle de risque nous ait été transmis par un Néandertalien : il pourrait provenir d'un ancêtre commun aux deux lignées. Mais la séparation entre Néandertal et *Homo sapiens* remonte à environ 550 000 ans, et, compte tenu du nombre de générations et de la fréquence des recombinaisons, on peut exclure qu'un segment de 50 kb ait été hérité intact d'un si lointain ancêtre. C'est donc bien un croisement avec un Néandertalien qui nous a apporté cet ensemble d'allèles. Une analyse phylogéné-

² *The major genetic risk factor for severe COVID-19 is inherited from Neanderthals.*

³ Il faut aller chercher dans les données supplémentaires pour voir qu'apparemment la définition de « sévère » correspond à « hospitalisé », un critère pas très sélectif.

⁴ Compte tenu du très fort déséquilibre de liaison entre la dizaine de SNP du locus de 50 kb en 3p21.31, on peut parler globalement d'allèle de risque pour l'ensemble de ces SNP.

⁵ Comme nous l'avons vu dans une chronique récente [9], l'ADN des Africains contient très peu de séquences de type Néandertal.

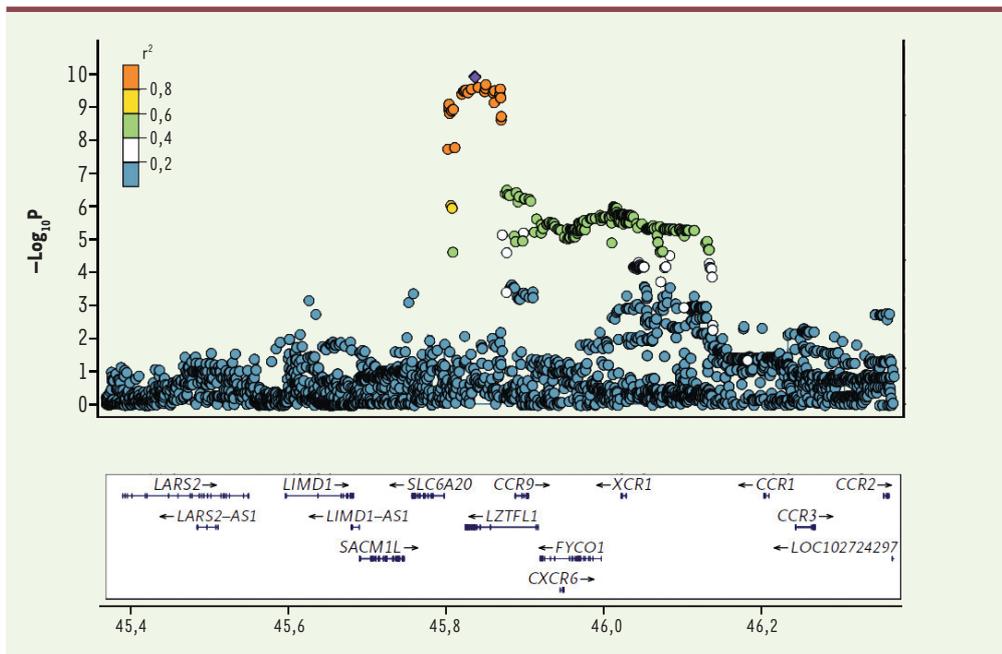


Figure 2. Détail de la région 3p21.31. Le pic principal (points orange) est en très fort déséquilibre de liaison (corrélation entre les allèles : $r^2 > 0,8$) et couvre environ 50 kb ; l'association avec la sévérité de l'infection s'étend au total sur environ 330 kb, avec un déséquilibre de liaison plus faible (r^2 entre 0,4 et 0,8). Les six gènes contenus dans cette région sont : *SLC6A20* (solute carrier family 6 member 20), *LZTFL1* (leucine zipper transcription factor like 1), *CXCR6* (C-X-C motif chemokine receptor 6), *FYCO1* (FYVE and coiled-coil

domain autophagy adaptor 1), *CCR9* (C-C motif chemokine receptor 9) et *XCR1* (X-C motif chemokine receptor 1). L'échelle en bas indique la position sur le chromosome, exprimée en mégabases (extrait partiel et modifié de la figure 3 de [2]).

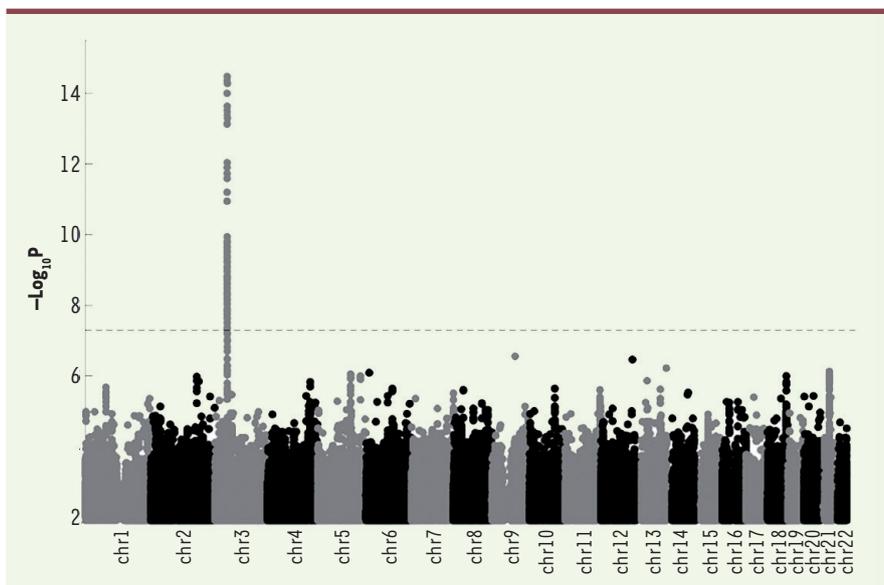


Figure 3. Manhattan plot obtenu par Zeberg et Pääbo [3] à partir des données de la Covid-19 Host Genome Initiative [6] (extrait de la figure 1 de [3]).

public. Il amène naturellement à s'interroger sur le rôle qu'aurait pu jouer cet allèle de risque chez les Néandertaliens. Présent à l'état homozygote dans la séquence de l'ADN de Néandertal de Croatie, pourrait-il avoir constitué un élément positif dans l'environnement de l'époque ? La poursuite des études sur l'ADN ancien donnera

tiques de cette région au sein des séquences répertoriées par le *1000 Genomes Project* montre que les séquences portant l'allèle de risque se regroupent bien avec les trois séquences de Néandertaliens, la plus proche étant celle qui provient de Croatie (Figure 5).

Une jolie histoire

On comprend que *Nature* ait souhaité mettre en avant cet article : ce résultat tout à fait inattendu est intrigant et suscite un vif intérêt, y compris dans le grand

sans doute quelques indications.

Ne perdons néanmoins pas de vue que ce facteur de risque, pour notable qu'il soit (de 1,6 à 2,14 selon les études), est bien moins important que d'autres éléments et notamment l'âge : la probabilité d'hospitalisation est multipliée par 5 et celle de décès par 90 lorsqu'on compare la tranche d'âge 65 à 74 ans à celle qui va de 18 à 29 ans [11]. Cela dit, les études génétiques et en particulier les dernières, plus détaillées [7], commencent à impliquer des gènes et à suggérer des mécanismes pathologiques qui peuvent avoir d'importantes conséquences sur les traitements de cette maladie. Le fait que le plus important allèle de risque génétique nous vienne de

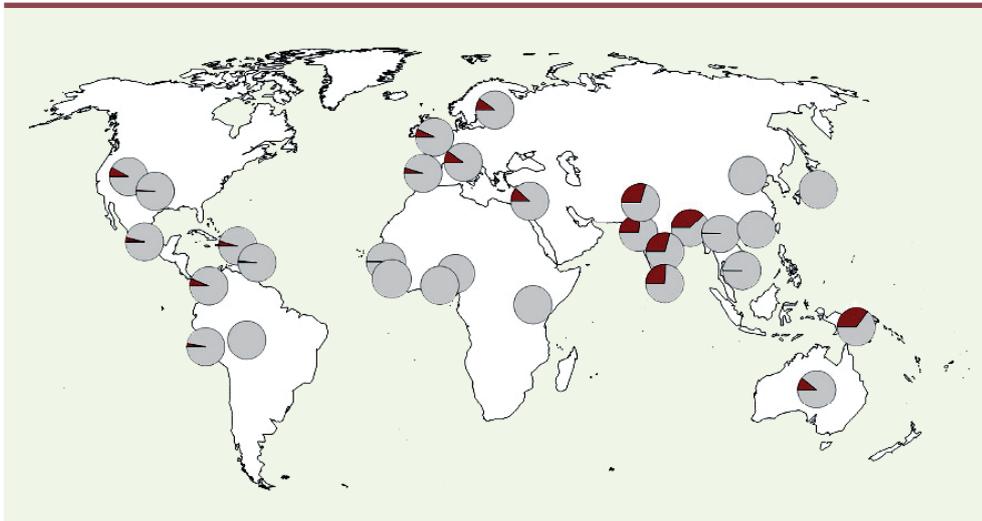


Figure 4. Distribution de l'allèle de risque pour une affection COVID-19 sévère (partie rouge des cercles) (extrait de la figure 3 de [3]).

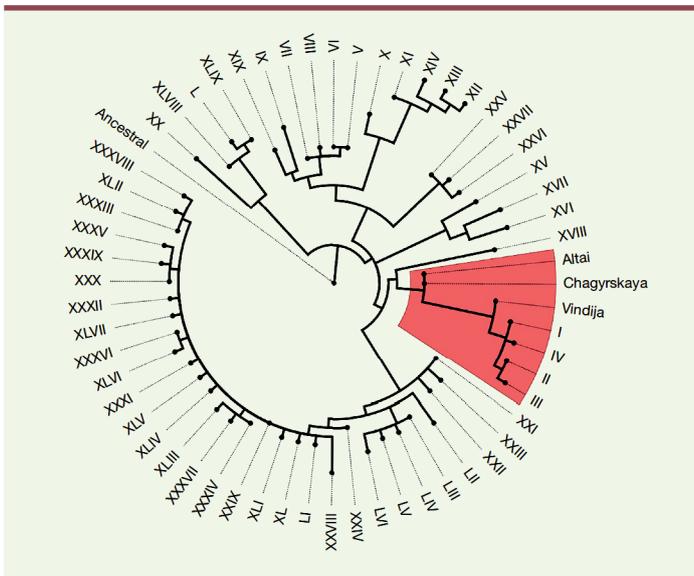


Figure 5. Analyse phylogénétique des ADN répertoriés dans le 1000 Genomes Project. La zone colorée indique les séquences portant l'allèle de risque en 3p21.31. Les trois séquences de Néandertaliens sont repérées par leurs noms, Vindija (la plus proche) étant celle qui provient de Croatie. Les chiffres romains désignent les différents sous-groupes qui apparaissent au sein des 1 000 génomes analysés (extrait remanié de la figure 2 de [3]).

Néandertal est intéressant mais relativement anecdotique ; l'élucidation des facteurs génétiques peut, elle, jouer un rôle important dans de futures avancées thérapeutiques. ♦

SUMMARY

A poisoned gift

GWAS analysis of severe Covid patients implicates a major locus on chromosome 3. The corresponding 50 kb segment appears to originate from Neanderthal/Sapiens crossings, raising interesting evolutionary questions. ♦

LIENS D'INTÉRÊT

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

RÉFÉRENCES

1. Chapman S, Hill A. Human genetic susceptibility to infectious disease. *Nat Rev Genet* 2012 ; 13 ; 175-88.
2. Ellinghaus D, Degenhardt F, Bujanda L, et al. Genomewide association study of severe covid-19 with respiratory failure. *N Engl J Med* 2020 ; 383 : 1522-34.
3. Zeberg H, Pääbo S. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* Sep 30. doi: 10.1038/s41586-020-2818-3.
4. Wein AN, McMaster SR, Takamura S, et al. CXCR6 regulates localization of tissue-resident memory CD8 T cells to the airways. *J Exp Med* 2019 ; 216 : 2748-62.
5. Kuba K, Imai Y, Ohto-Nakanishi T, Penninger JM. Trilogy of ACE2: a peptidase in the renin-angiotensin system, a SARS receptor, and a partner for amino acid transporters. *Pharmacol Ther* 2010 ; 128 : 119-28.
6. The COVID-19 Host genetics initiative. The COVID-19 host genetics initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur J Hum Genet* 2020 ; 28 : 715-8.
7. Pairo-Castineira E, Clohisey S, Klaric L, et al. Genetic mechanisms of critical illness in Covid-19. *MedRxiv* 2020.09.24.20200048.
8. 1 000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015 ; 526 : 68-74.
9. Jordan B. Neandertal et Afrique, le retour. *Med Sci (Paris)* 2020 ; 36 : 421-3.
10. Prüfer K, de Filippo C, Grote S, et al. A high-coverage Neandertal genome from Vindija cave in Croatia. *Science* 2017 ; 358 : 655-8.
11. COVID-19 Hospitalization and death by age, CDC. <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-age.html>

TIRÉS À PART

B. Jordan