

Chroniques génomiques

Maladie de Huntington : enfin un mécanisme !

Bertrand Jordan

Une si longue attente...

L'identification, en 1993, du gène responsable (lorsqu'il est défectueux) de la maladie (anciennement : chorée) de Huntington [1] a été un des premiers grands succès de ce que l'on appelait à l'époque la « génétique inverse » [2] (→). **(→ Voir m/s n° 3, 1988, page 138)**

Bientôt rebaptisée « clonage positionnel », cette approche reposait sur la localisation approximative du gène recherché par une étude génétique au sein de familles présentant la maladie en cause. Suivait alors une (très laborieuse) analyse moléculaire de la région ainsi désignée pour faire le catalogue des gènes qu'elle contient puis, par diverses techniques, identifier celui dont la forme anormale provoque la maladie. Le séquençage intégral du génome humain, puis la mise au point des techniques de balayage du génome (*GWAS, genome-wide association studies*) [3] (→) et de très grands progrès dans l'obtention et l'interprétation des séquences d'ADN ont changé la donne et grandement accéléré les progrès ; mais, malgré l'identification précoce du gène et du défaut relevé chez les malades, le mécanisme pathogène de la maladie de Huntington est resté obscur [1] – jusqu'au très beau travail présenté dans un récent article de la revue *Cell* [4] et qui fait l'objet de cette chronique.

Rappelons les caractéristiques de la maladie : c'est une affection neurologique héréditaire, transmise sur le mode dominant et qui se manifeste à partir de l'âge adulte par des troubles moteurs, cognitifs et psychiatriques qui s'aggravent progressivement jusqu'au décès. Le striatum, région du cerveau particulièrement impliquée dans les fonctions motrices et cognitives, est fortement atrophié chez les malades. Le gène en cause, appelé *HTT (Huntingtin)* est situé sur le bras court du chromosome 4 et présente dans sa partie codante une répétition du triplet nucléotidique CAG qui entraîne la présence dans la protéine correspondante (appelée



Biologiste, généticien et immunologiste, Président d'Aprrogène (Association pour la promotion de la Génomique), Marseille, France. brjordan@orange.fr

huntingtine) d'une zone de poly glutamine (le triplet CAG codant pour la glutamine). L'anomalie consiste en un nombre anormalement élevé de triplets CAG : généralement de l'ordre d'une vingtaine, ce nombre est plus élevé dans la version pathogène, typiquement de l'ordre de 40 ou 50 mais pouvant atteindre 100. L'apparition des premiers symptômes est d'autant plus précoce que le nombre de triplets CAG est élevé. Depuis trente ans, on cherche à comprendre le mécanisme de l'affection : l'hypothèse la plus logique (et la plus explorée) envisage que la huntingtine anormale (présentant une région poly glutamine allongée) provoque directement ou non une toxicité cellulaire liée à l'accumulation de la protéine ou de ses produits de dégradation. Mais ces travaux n'ont pas apporté de résultat vraiment concluant, et des tentatives thérapeutiques (sur des modèles murins) cherchant à réduire la production de huntingtine pour ralentir l'évolution de la maladie n'ont pas abouti [1]. Des travaux récents [5] ont de plus montré qu'une mutation qui remplace un CAG par un CAA (mutation synonyme) dans la série de triplets, et ne modifie donc pas la structure de la huntingtine, empêche l'apparition de la maladie : c'est donc la structure du gène et non celle de la protéine qui cause l'affection.

Les CAG font des petits

Une information très importante est apportée par un travail déjà ancien mais peu remarqué à son époque : il s'agit de l'expansion au niveau somatique de la zone répétée au sein des cellules neuronales du striatum. Cette étude publiée en 2003 [6] observe une forte

Vignette (© Bertrand Jordan).

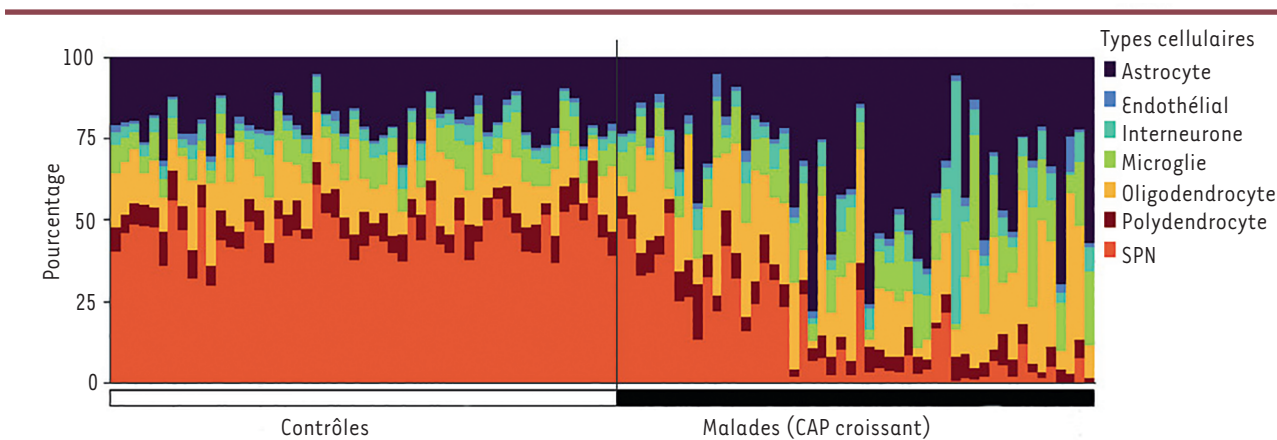


Figure 1. Proportion de différents types cellulaires dans le striatum de témoins et de malades. SPN, *striatal projection neurons*, neurones de projection striataux. Le score CAP indique la sévérité de l’atteinte. (Extrait partiel et modifié de la figure 1 de [4]).

expansion du nombre de triplets CAG (jusqu’à 1000) dans le gène *HTT* de cellules du striatum de malades. Il est connu que les séquences répétées peuvent s’amplifier dans les cellules somatiques (en l’absence de réplication) à la suite de réparations décalées intervenant après une transcription [7, 8] : la mise en évidence de telles amplifications est importante puisqu’elle suggère un mécanisme pathogène ne faisant pas intervenir la protéine huntingtine. Les travaux rapportés dans l’article de *Cell* [4], une étude de grande ampleur mettant en œuvre les techniques les plus récentes d’analyse au niveau de la cellule unique, documentent la réalité de ces amplifications, précisent leur cinétique et éclairent leurs conséquences sur le métabolisme cellulaire et sur la disparition des neurones du striatum.

Des neurones qui disparaissent

Les auteurs ont eu accès à une grande série (53) de cerveaux *post mortem* de malades et à une série équivalente (50) de témoins. À partir de tissu du striatum ils obtiennent une suspension de noyaux qui sont encapsulés dans des microgouttes. Après amplification PCR et incorporation de séquences « code-barre » (*barcodes*) le niveau d’expression de plusieurs milliers de gènes est mesuré pour chaque noyau [9]. Le profil d’expression permet d’identifier le type cellulaire correspondant [10], et ainsi de comparer l’abondance de ces différents types chez les malades et les témoins (Figure 1).

On voit que si les neurones de projection striataux (SPN, *striatal projection neurons*) constituent environ 50 % des cellules chez les témoins, leur proportion diminue fortement chez les malades et cela d’autant plus que leur atteinte est sévère. Le score CAP indiqué sur la figure est une mesure de cette sévérité et tient compte de l’âge du malade et du nombre de répétitions dans son gène *HTT*¹. Le niveau d’expression du gène *HTT* dans les cellules n’est pas lié au score CAP.

¹ Le score CAP est égal à l’âge du malade multiplié par le nombre de CAG hérités ($n(\text{CAG})_h$) moins 33,66. Une personne indemne ($n(\text{CAG})_h$ inférieur à 30) aura donc un score nul, un malade ($n(\text{CAG})_h = 45$) âgé de 50 ans un score de 510. Les premiers symptômes apparaissent généralement pour un score CAP de 300 à 350.

Au total, la disparition des neurones SPN se retrouve bien au niveau moléculaire et explique l’évolution pathologique ; reste à déterminer ce qui cause cette disparition. On soupçonne que l’amplification somatique des répétitions CAG [6] est en cause – reste à le démontrer.

Un tout de force très fructueux

Grâce à l’emploi de techniques de pointe, les auteurs ont réussi à mesurer *dans chaque cellule* le profil global d’expression (qui identifie la cellule et son état métabolique) et *en même temps* le nombre de triplets CAG présents dans son ADN (et dans le transcrit du gène *HTT*). Ils ont ainsi analysé plus de 500 000 cellules individuelles au cours de ce travail ! Le procédé employé est indiqué dans la Figure 2.

La méthode repose sur (1) l’incorporation d’un code-barre (*barcode*) aux acides nucléiques de chaque noyau avant, d’une part, (2) la mesure du profil d’expression comme précédemment, et d’autre part (3) une amplification ciblée sur la séquence transcrite du gène *HTT*, (4) une purification additionnelle ciblée sur la même séquence, (5) une sélection des molécules amplifiées les plus longues, et (6) un séquençage par une machine *Pacific Biosciences* qui peut lire de très longues séquences (jusqu’à plusieurs kilobases). On peut ainsi, grâce aux codes-barres, obtenir le profil global d’expression, identifier le type cellulaire, et mesurer la longueur de la séquence répétée (CAG) $_n$ présente dans le transcrit du gène *HTT* pour la même cellule. Les précautions prises pour la conservation des séquences longues et pour leur lecture limitent les artefacts d’amplification et assurent une détermination fiable du nombre de CAG. Ce travail – assez acrobatique du

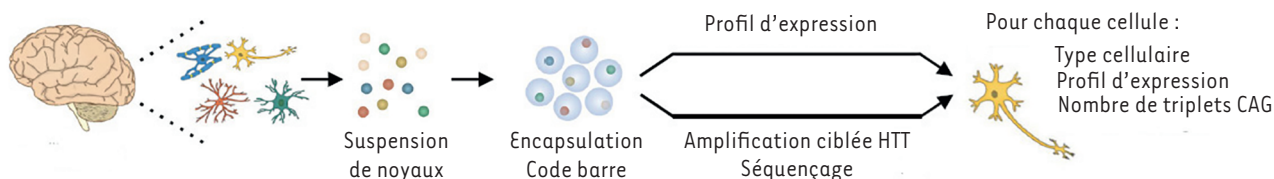


Figure 2. Représentation schématique de la méthode de détermination simultanée pour chaque cellule du type cellulaire, du profil d'expression et du nombre de triplets CAG (voir le texte). (Extrait partiel et modifié de la figure 2 de [4]).

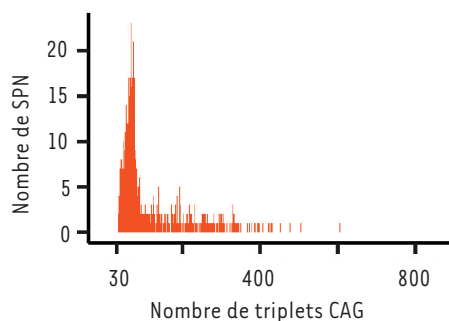


Figure 3. Longueur de la séquence (CAG)_n dans les neurones SPN d'un patient. (Extrait partiel et modifié de la figure 2 de [4]).

point de vue technique – a été réalisé sur six prélèvements provenant de malades et a débuté par l'analyse du segment (CAG)_n dans les différents types cellulaires présents dans le striatum (voir Figure 1). Il montre que seul l'allèle pathogène est amplifié (CAG_n>37), et seulement de manière significative dans les neurones SPN. L'amplification est variable d'une cellule à l'autre au sein du même prélèvement, et la très grande majorité des SPN montre une amplification modérée, 60 à 70 répétitions au lieu de 40 à 43 dans le gène transmis. Mais une importante minorité de neurones SPN porte une amplification plus importante, allant jusqu'à plusieurs centaines de répétitions (Figure 3).

Une évolution très lente, puis rapide et catastrophique

On peut alors examiner en détail le profil d'expression des gènes dans les différentes catégories de neurones SPN, en fonction du nombre de répétitions. Pour la grande majorité de ces cellules, qui présentent 60 à 70 répétitions (ce que les auteurs appellent stade A), ce profil n'est pas différent de celui trouvé chez les témoins. Pour les amplifications (minoritaires) ne dépassant pas 150 CAG (stade B) le profil reste

inchangé. Au-delà, par contre, entre 150 et 350 CAG, le profil change pour plus de 700 gènes (stade C). Pour les rares neurones portant plus de 350 CAG (cf. Figure 3), on observe l'activation d'une centaine de gènes jusque-là non exprimés, ce que les auteurs appellent la « crise de dérépression » (stade D). Parmi ces gènes nouvellement exprimés se trouvent *CDKN2A* (cyclin dependent kinase inhibitor 2A) et *CDKN2B* (cyclin dependent kinase inhibitor 2B) impliqués dans la sénescence et l'apoptose cellulaires ; leur expression entraîne la mort cellulaire des neurones SPN et leur élimination. Il est important de comprendre la temporalité de ces différentes phases : la phase A est la plus longue, 95 % des SPN à un moment donné se trouvent dans cette phase, et l'évolution est lente : environ une mutation (augmentation du nombre de CAG) par an. La phase B est plus rapide, mais les SPN sont encore fonctionnels (leur profil d'expression est inchangé) et la maladie reste asymptomatique. À partir de la phase C l'évolution devient rapide et les symptômes apparaissent ; quand la grande majorité des neurones SPN a disparu (voir Figure 1) l'affection devient sévère. L'évolution des SPN et leur entrée en phase B sont asynchrones, lentes et aléatoire ; ensuite l'évolution se précipite et parcourt les phases suivantes en quelques mois. Cette description repose sur un ensemble de simulations ainsi que sur des données biologiques, par exemple la manière dont des altérations du système de réparation de l'ADN (dans des modèles murins) accélèrent l'apparition des symptômes [8]. Ce mécanisme permet de comprendre ce qui était un grand mystère dans la pathogénèse de la maladie, le fait qu'elle se manifeste très tardivement bien qu'il s'agisse d'une affection génétique dominante ; il éclaire l'ensemble du processus tout en laissant ouvertes au moins deux questions : pourquoi l'expansion des triplets CAG se produit-elle essentiellement dans les neurones SPN et non dans les autres types cellulaires ? De quelle manière la présence d'une longue séquence (CAG)_n dans le gène *HTT* entraîne-t-elle des changements profonds de patrons d'expression ? Quoi qu'il en soit, il s'agit là d'un très beau travail, très solide et détaillé (nous n'en avons vu ici que les grandes lignes) et qui donne les premières clefs d'un mystère vieux de trente ans.

Des pistes pour Huntington – et pour d'autres affections

Au vu de cette nouvelle compréhension de la maladie de Huntington, on comprend l'échec des approches thérapeutiques tentées jusqu'ici :

elles visaient à réduire la production de la protéine huntingtine anormale, censée être responsable de la perte neuronale observée. On faisait fausse route puisque c'est la séquence (CAG)_n elle-même qui – par un mécanisme restant à préciser – déclenche les altérations transcriptionnelles qui aboutiront à la mort de la cellule. Il faut donc plutôt chercher à ralentir l'amplification de la séquence (CAG)_n, ce qui est envisageable en agissant sur les mécanismes de réparation de l'ADN [8] ; comme l'amplification de la séquence est très lente jusqu'à la phase C, on peut envisager que de tels traitements soient testés sur des personnes en début de phase symptomatique avec un résultat (positif ou négatif) assez rapide. Par ailleurs, il existe plus de 50 maladies génétiques liées à l'expansion d'une suite de trinucléotides [11] : mentionnons seulement le syndrome de l'X fragile et la myotonie dystrophique. Le mécanisme révélé par la présente étude [4] suggère à l'évidence un réexamen du processus de ces affections. En tous cas, nous avons ici un bel exemple d'un mystère apparemment insoluble résolu par un très beau travail utilisant toutes les ressources d'approches techniques novatrices. ♦

SUMMARY

Finally, an understanding of Huntington's disease!

A beautiful piece of work using extensive single-cell studies illuminates the mechanism of Huntington's disease: the somatic expansion of the (CAG)_n tract, very slow at first but accelerating once a critical repeat length is reached, drives extensive changes in gene expression in striatal neurons and eventually leads to cell death and atrophy of the striatum. This explains many puzzling features of the disease and may have important implications for possible therapy and for the understanding of other triplet repeat disorders. ♦

LIENS D'INTÉRÊT

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

RÉFÉRENCES

1. Gusella JF, Lee JM, and MacDonald ME. Huntington's disease: nearly four decades of human molecular genetics. *Hum. Mol. Genet.* 2021 ; 30 : R254–R263.
2. Jordan B. Grandeur et servitudes de la génétique inverse. *Med Sci (Paris)* 1988 ; 4 : 138–40.
3. Jordan B. Le déclin de l'empire des GWAS. *Med Sci (Paris)* 2009 ; 25 : 537–9
4. Handsaker RE, Kashin S, Reed NM, et al. Long somatic DNA-repeat expansion drives neurodegeneration in Huntington's disease, *Cell* 2025 ; 188 : 623–39.e19.
5. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. CAG repeat not polyglutamine length determines timing of Huntington's disease onset. *Cell* 2019 ; 178 : 887–900.e14.
6. Kennedy L, Evans E, Chen CM, et al. Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum. Mol. Genet.* 2003 ; 12 : 3359–67.
7. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. Identification of genetic factors that modify clinical onset of Huntington's disease. *Cell* 2015 ; 162 : 516–26.
8. Iyer RR, and Pluciennik A. DNA mismatch repair and its role in Huntington's disease. *J. Huntingtons Dis.* 2021 ; 10 : 75–94.
9. Macosko EZ., Basu A., Satija R., et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 2015 ; 161 : 1202–14.
10. Alquicira-Hernandez J, Sathe A., Ji HP, et al. scPred: accurate supervised method for cell-type classification from single-cell RNA-seq data. *Genome Biol.* 2019 ; 20 : 264.
11. Depienne C and Mandel JL. 30 years of repeat expansion disorders: what have we learned and what are the remaining challenges? *Am. J. Hum. Genet.* 2021 ; 108 : 764–85

TIRÉS À PART

B. Jordan

