

# Stratégies d'identification de gènes responsables des leucodystrophies de cause indéterminée

Imen Dorboz<sup>1-2</sup>, Florence Renaldo<sup>3</sup> Simon Samaan<sup>4</sup>, Monique Elmaleh<sup>5</sup>, Diana Rodriguez<sup>6</sup>, Odile Boespflug-Tanguy<sup>3</sup>

<sup>1</sup> INSERM U1141, PROTECT, Hôpital Robert Debré, Paris

<sup>2</sup> Association ELA;

<sup>3</sup> Service de Neuropédiatrie et maladies Métaboliques, centre de référence maladies rares pour les leucodystrophies « LEUKOFRANCE », Hôpital Robert Debré, Paris;

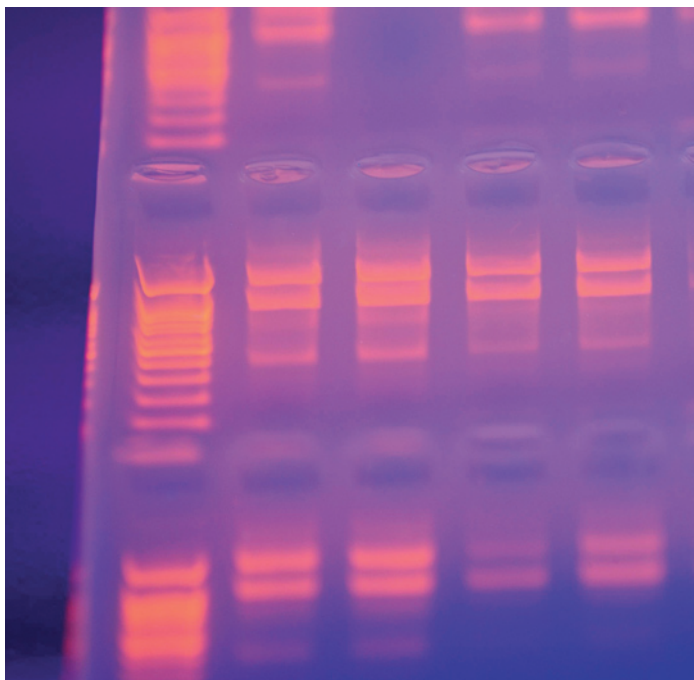
<sup>4</sup> Biologie Moléculaire, Département de Génétique, Hôpital Robert Debré, Paris;

<sup>5</sup> Radiologie Pédiatrique, Hôpital Robert Debré, Paris;

<sup>6</sup> Neurologie pédiatrique; Hôpital A. Trousseau, AP-HP, Paris.

Les leucodystrophies sont un groupe de **maladies génétiques rares**, qui affectent de **façon initiale et prédominante la substance blanche** du système nerveux central et son principal constituant, la **myéline**. Leur identification repose donc, sur la présence d'anomalies du signal de la substance blanche à **l'imagerie par résonance magnétique (IRM)** cérébrale, et/ou de la moelle épinière.

De nombreuses maladies acquises (liées par exemple, à une inflammation, une infection, une intoxication, une anoxie néonatale, une atteinte vasculaire...); ou génétiques, affectant primitivement la substance grise, les vaisseaux du cerveau ou d'autres organes que le cerveau, peuvent également donner des anomalies de la substance blanche à l'IRM. Ces maladies sont regroupées sous le terme de **leuco encéphalopathies**.



Seul un **faisceau d'arguments, basés sur l'aspect de l'IRM et un bilan clinique et biologique précis**, permettent d'orienter vers le diagnostic de leucodystrophies. Les caractéristiques des anomalies de la substance blanche sur les différentes séquences IRM (T1, T2, Flair) ont permis de distinguer des formes suggérant une destruction rapide de la myéline (**démýélinisation**), avec parfois un aspect œdémateux et/ou cavitaire de la substance blanche, et d'autres suggérant un défaut de son développement (**hypomyélinisation**). L'hétérogénéité clinique importante des leucodystrophies (âge de début, symptômes cliniques, évolution, y compris pour le même gène causal) et la grande diversité des anomalies génétiques, potentiellement responsables (près de 20 000 gènes), expliquent **les difficultés à mettre en évidence** le gène en cause. **Les formes sans anomalie génétique** identifiées sont dénommées **leucodystrophies de cause indéterminée (LDI)**. Lorsqu'un doute important persiste sur l'origine, primitivement dans la substance blanche ou génétique, de la maladie observée, le terme plus vague de **leuco encéphalopathie** est utilisé.

L'identification des gènes responsables des maladies génétiques a suivi l'évolution des techniques d'analyse du génome (biologie moléculaire). Jusqu'aux années 2000, l'identification des gènes en cause s'est faite en utilisant une **stratégie dite du gène candidat**. Cette approche avait pour but de tester un gène ou un groupe de gènes, du fait de leur implication suspectée dans la pathologie humaine étudiée. Pour orienter cette recherche, une localisation préalable du gène sur le génome était souvent nécessaire. Cette localisation génique nécessitait le prélèvement du plus grand nombre possible de familles atteintes de la même maladie, et au sein de chaque famille, d'individus atteints comme sains; afin de pouvoir déterminer statistiquement le lien entre des marqueurs de position du génome et la maladie (**étude de liaison génétique**). L'implication d'un gène situé dans la région génomique d'intérêt était ensuite basée sur les propriétés de la protéine produite à partir du gène, et/ou du fait de l'existence de modèles animaux de la maladie humaine. L'absence de gènes candidats nécessitait un travail long et coûteux de séquençage progressif de tous les gènes situés dans la région génomique d'intérêt (« marche sur le génome »).

Dans les leucodystrophies, ce travail a permis l'identification des gènes impliqués dans les formes les plus fréquentes. Ainsi:

- **Pour l'adrénoleucodystrophie liée à l'X (X-ALD)**, l'implication du gène ABCD1, responsable d'un transporteur des acides gras à très longues chaînes (ALDP) a été facilitée par l'existence d'une quantité anormalement élevée d'acides gras à très longues chaînes dans le sérum des garçons atteints.
- **Pour la maladie de Pelizaeus Merzbacher (PMD)**, le gène des protéolipoprotéines (PLP1) a été le gène candidat de choix; du fait: d'une part, de la fonction de la protéine PLP (constituant majeur de la myéline) et d'autre part, de l'existence de nombreux animaux mutés pour ce gène, mimant l'hypomyélinisation observée chez l'homme.
- **Pour la maladie d'Alexander (AxD)**, le gène GFAP, codant pour la protéine acide fibrillaire gliale, a été ciblé; du fait de la

production d'une lignée de souris avec une quantité élevée de gène GFAP humain, qui présentaient des lésions caractéristiques de la maladie humaine à l'analyse anatomique du cerveau (fibres de Rosenthal).

• **Pour le syndrome CACH (Childhood Ataxia with Central nervous system Hypomyelination) ou le syndrome d'Aicardi-Goutières (AGS)**, la localisation génique a été rendue difficile, du fait de l'existence de plusieurs gènes impliqués dans la même maladie (hétérogénéité génétique). La mise en évidence d'un gène principal a permis ensuite, l'identification rapide des autres par homologie de fonction (5 dans le syndrome CACH, 6 dans l'AGS).

Les progrès liés au séquençage du génome humain entier, ont permis le développement d'une **approche par recherche de l'anomalie génétique en cause sur l'ensemble du génome, en utilisant la technique de séquençage à haut débit**. C'est cette stratégie, actuellement utilisée pour les LDI que nous exposerons.

## Techniques de séquençage à haut débit (Next Generation Sequencing)

**Le séquençage de l'ADN** consiste à déterminer l'ordre d'enchaînement des lettres du code génétique A, T, G, C (nucléotides) pour un fragment d'ADN donné. **Le séquençage à haut débit** (Next Generation Sequencing ou NGS) est une technique qui permet un **séquençage aléatoire (sans ordre)** de courts fragments d'ADN, lus un grand nombre de fois, et couvrant l'intégralité du génome. Les fragments sont ordonnés automatiquement; en comparant la séquence d'enchaînement des lettres du code génétique obtenue, à la séquence de référence du génome humain obtenue à partir de l'ADN d'un individu sain dans les années 2000.

Le NGS permet donc, d'identifier **toutes les variations de séquence du code génétique de l'ADN de l'individu étudié, par rapport à la séquence de référence du génome humain**. La présence d'une variation (modification d'une lettre ou d'un enchaînement de lettres), peut être directement responsable de la maladie génétique étudiée; on parle alors de **variant pathogène**. La grande majorité des variations identifiées sont, cependant, liées à la diversité du génome (polymorphisme); que l'on observe d'un individu à l'autre, et qui fait que nous sommes tous différents et uniques sur le plan génomique. Ces variants sont dits **non pathogènes**, car leur présence n'est pas directement responsable d'une maladie génétique connue. Ces variants non pathogènes sont plus fréquemment retrouvés chez des individus appartenant au même groupe ethnique et, a fortiori, à la même famille.

La fiabilité du séquençage de chaque région est liée au nombre de lecture effectué pour cette région. Certaines régions, du fait de la configuration du génome ou de la qualité de l'ADN de l'individu, peuvent être, ainsi, mal couvertes par le NGS. Ces régions doivent être repérées, afin de déterminer celles ne permettant pas de conclure sur la présence ou non d'un variant. Ce point est très important, s'il s'agit d'une région contenant un ou plusieurs gènes importants pour la pathologie. Les techniques de NGS mises sur le marché, par différents concepteurs, varient en termes de performance, de temps de technique, et surtout évoluent très rapidement; ce qui peut expliquer qu'un résultat négatif avec une technique puisse devenir positif.

Le NGS peut permettre l'analyse d'un **groupe spécifique de gènes** (sous forme de panel), de **l'exome**, ou d'un **génomome entier**.

• **Séquençage d'un panel de gènes**: Cette technique permet d'obtenir, rapidement et simultanément, la séquence de plusieurs

gènes impliqués dans une maladie; (pouvant aller avec les nouvelles techniques jusqu'à une centaine), et ceci à moindres coûts. Cette approche, par panel, permet d'optimiser la couverture et l'analyse des gènes connus, pour être impliqués dans un groupe hétérogènes d'affections (testé initialement dans le retard mental). La technique et le traitement des données sont donc plus rapides.

• **Séquençage de l'exome**: **L'exome** est l'ensemble des régions du génome, dont le codage va permettre la fabrication d'une protéine (**régions codantes**). L'exome représente environ 5 % de l'ADN d'un individu; mais 85 % des mutations décrites dans des maladies génétiques humaines, sont localisées dans ces régions. Il s'adresse à l'ensemble des gènes, mais génère de nombreux variants à analyser. L'analyse de l'exome d'un individu génère, en effet, en moyenne 10000 variants. Des bases de données internationales permettent d'éliminer les variants non pathogènes connus et fréquents dans les différents groupes ethniques humains. Après « filtrage », le nombre de variants non référencés dans ces bases de données, et potentiellement causal de la pathologie, reste cependant élevé (autour de 2000).

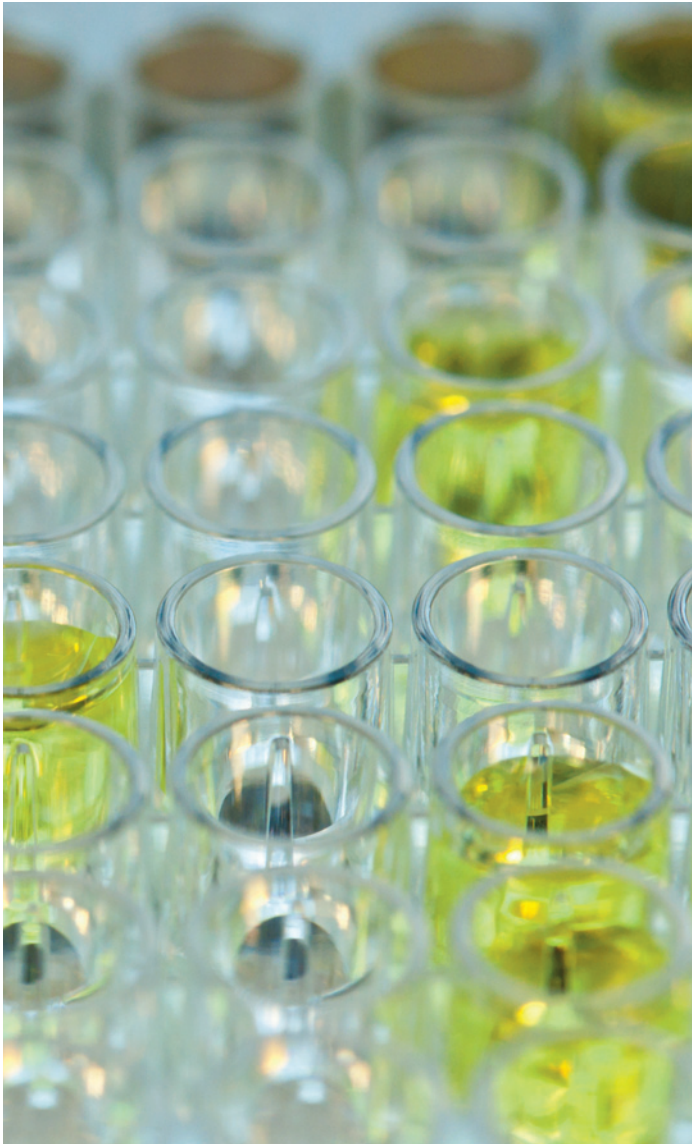
• **Séquençage du génome entier**: Le séquençage s'adresse à tout le génome, qu'il s'agisse de régions codantes ou non codantes; cette approche « génome entier » permet de mieux analyser les régions codantes, identifiant des variants non identifiés par l'approche « exome ». L'analyse des variants dans les régions non codantes est particulièrement complexe du fait de la très grande diversité des séquences de ces régions, entre individus, comparées aux régions codantes. Leur utilisation en diagnostic reste encore limitée.

Le travail le plus long reste donc la **validation d'un variant pour expliquer la maladie observée**. L'identification d'un variant non référencé comme non pathogène, chez un patient atteint d'une maladie génétique nécessite donc, toujours, de **vérifier si ce variant est présent ou non chez ses parents**. Si le variant est présent chez l'un des parents sains, son implication dans la maladie est peu probable. L'analyse d'autres individus sains de la famille est, parfois, nécessaire. Au cours de ce travail, un ou des variant(s) pathogène(s) connu(s), prédisposant à d'autres maladies (cancer, pathologies dégénératives) que celles observées, peuvent être mis en évidence. L'existence de ces variants « à risque de pathologie » ne sera révélée au patient, qu'en cas d'intérêt direct et certain pour le patient (surveillance, thérapeutique préventive, conseil génétique). Le consentement éclairé et signé par le patient (ou son représentant légal), obligatoire pour l'analyse moléculaire de l'ADN d'un individu, signale, de plus en plus souvent, cette éventualité; du fait de l'utilisation de plus en plus fréquente des techniques NGS pour le diagnostic moléculaire.

## Validation des variants pathogènes pour la maladie observée

Plusieurs paramètres sont pris en compte pour la confirmation de la responsabilité des variants identifiés dans la maladie:

- **la présence**, uniquement, chez les malades au sein de la famille (coségrégation) ou dans d'autres familles atteintes d'une maladie proche.
- l'absence dans une population non atteinte (témoin) de même origine ethnique, s'il s'agit d'une population particulière.
- la prédiction de son caractère pathogène, à partir d'outils bioinformatiques, prenant en compte différents paramètres (type de variant, conséquences au niveau de l'acide aminé composant la protéine; en particulier sa conservation entre les espèces, sa localisation dans un domaine fonctionnel de la protéine).



Au-delà, de l'utilisation des critères précédemment cités; le plus important, est de déterminer si le gène porteur du variant, peut expliquer le tableau clinique observé. Cela, nécessite des **interfaces multidisciplinaires régulières, entre les biologistes moléculaires responsables du NGS et les experts cliniciens de la pathologie** présentée par le patient. Lorsque le doute persiste, des études fonctionnelles sont indispensables pour conclure. Ces **études fonctionnelles** sont faciles à réaliser, lorsque les cellules affectées du patient sont accessibles; ce qui n'est pas le cas pour les cellules du cerveau. Des modèles cellulaires ou animaux sont alors indispensables.

Dans certains cas, le caractère causal des variants identifiés reste **impossible à déterminer**. La découverte d'un patient avec des signes cliniques proches, porteur du même variant, permet, parfois, de valider la pathogénicité du variant. Le développement de bases de données ouvertes à l'ensemble de la communauté scientifique, contenant ces variants **et les signes cliniques associés**, permettra d'accélérer cette recherche.

## Recherche de nouveaux gènes de leucodystrophies

Dans les LDI, la technique de NGS a été évaluée, en parallèle, avec une approche de panel de gènes et d'exome (financements « Push project » ELA).

### • Panel de gènes pour le diagnostic des leucodystrophies

En collaboration avec l'équipe de Jean Louis Mandel (Claire Redin, IGBMC, Illkirch); une puce NGS avec un panel de **68 gènes** de leucodystrophies, a été développée. L'analyse NGS de 169 patients LDI a permis de conclure, après une confrontation clinico-moléculaire, à la présence d'une variation pathogène responsable de la maladie dans **21 % des cas**; et d'identifier les gènes les plus fréquemment mutés dans les LDI.

Suite à cette collaboration, le laboratoire de biologie moléculaire de l'hôpital Robert Debré (Simon Saaman) a développé un panel contenant les 26 gènes, les plus fréquemment impliqués, avec une technique fiable et rapide (<1mois); applicable en pratique clinique. En 2015, l'analyse par cette stratégie, de 131 patients suspects de leucodystrophies, a permis l'identification de l'anomalie causale dans **26 % des cas**. Suite à l'analyse de l'exome (voir ci-dessous), un nouveau panel de 45 gènes, contenant également des gènes impliqués dans des leuco encéphalopathies génétiques (maladies primitivement métaboliques ou neuronales), est actuellement utilisé.

### • Analyse de l'exome pour le diagnostic des leucodystrophies

Sur les 70 familles analysées par cette technique, la mutation causale a été identifiée dans 50 % des cas (Imen Dorboz). Cette approche a permis d'identifier 3 sous-groupes: le 1er sous-groupe avec des mutations dans des gènes connus de leucodystrophies et avec un tableau clinique très atypique; un 2<sup>e</sup> sous-groupe avec des mutations dans des gènes déjà impliqués dans une autre maladie génétique du système nerveux central (leuco encéphalopathies génétiques); et un 3<sup>e</sup> groupe avec des mutations dans de nouveaux gènes.

## En conclusion

L'utilisation des techniques de séquençage à haut débit (NGS) dans les leucodystrophies permet:

- **l'identification plus rapide du variant causal (responsable) de la maladie**; afin d'influencer plus efficacement les décisions cliniques, en particulier, thérapeutiques et pour le conseil génétique (risque de survenue de la maladie dans la famille, diagnostic anténatal ou préimplantatoire).
- **l'identification de patients avec un tableau clinique atypique**; afin de mieux appréhender le spectre clinique des gènes impliqués dans des leucodystrophies ou des encéphalopathies connues.
- **l'identification de nouveaux gènes impliqués dans les leucodystrophies**; ce qui enrichit notre connaissance sur les mécanismes moléculaires à l'origine du dysfonctionnement de la substance blanche, et le développement de thérapeutiques ciblées.

Elles nécessitent cependant une confrontation étroite entre experts cliniciens et biologistes moléculaires, avant de conclure à l'implication d'une anomalie du génome dans la pathologie observée. L'absence d'anomalie identifiée chez 50 % des patients LDI, en utilisant l'exome, souligne l'hétérogénéité génétique des anomalies en cause. L'analyse NGS « génome entier » dans d'autres groupes de maladies génétiques, cliniquement hétérogènes, a montré qu'un nombre encore important de cas, reste sans anomalie génétique directement liée à la pathologie. La mise en commun des données obtenues (bases de données NGS et cliniques) au niveau international permettra, certainement, d'identifier des variants ou des associations de variants capables de moduler, sous l'effet de facteurs d'environnement (en particulier de stress), des gènes impliqués dans le développement, la stabilité et le fonctionnement de la substance blanche.