

# LES NOUVELLES TECHNIQUES de séquençage

Le développement des technologies de séquençage de nouvelle génération ou *NGS* (*Next Generation Sequencing*) durant les 10 dernières années est une révolution technologique sans précédent. Alors que le séquençage d'un seul génome humain aura nécessité 10 ans de collaborations internationales de 1993 à 2003 et 2,7 milliards de dollars, il est désormais possible, en l'espace de quelques semaines, voire de quelques jours, pour des coûts inférieurs à 1000 €, de séquencer l'intégralité de la région codante des 23 000 gènes humains, l'exome, représentant 34 millions de paires de bases de l'ADN.

Le séquençage de génomes humains entiers dans un but médical (3 milliards de paires de bases) a déjà été réalisé par quelques équipes aux États-Unis, au Royaume-Uni, aux Pays-Bas et en France. Cette révolution technologique est à l'origine d'une révolution médicale : la médecine de précision ou génomique dont l'objectif est d'optimiser le diagnostic, la prévention et le traitement des maladies humaines en fonction des variations génétiques individuelles.





## SÉQUENCER L'ADN

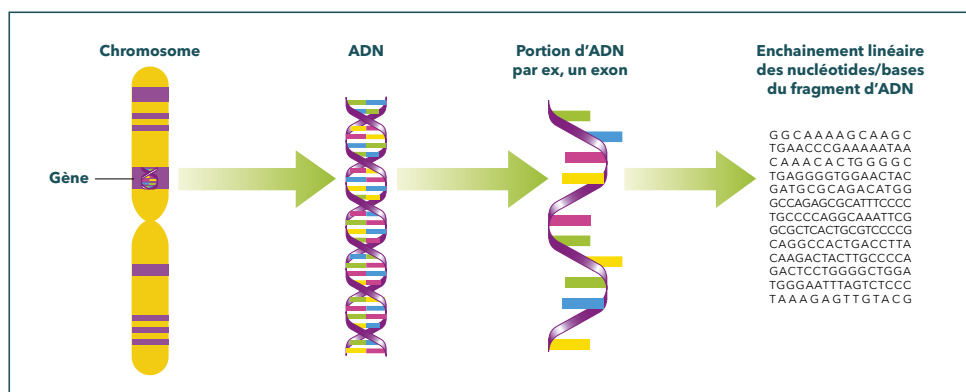
Le séquençage de l'ADN est une technique qui permet de lire l'enchaînement des nucléotides (ou bases) d'un fragment voire de la totalité de l'ADN, la molécule constituant nos gènes. Différentes approches technologiques ont été développées au cours du temps.

### Pourquoi séquençer l'ADN ?

Lorsque l'on recherche l'anomalie génétique en cause dans une maladie génétique, le séquençage de l'ADN permet d'identifier les variations dans certains gènes. Ces variations peuvent ensuite être reliées (ou non) à l'apparition de la maladie génétique, par comparaison avec des variations génétiques connues.

### Accéder à un diagnostic génétique

On dit d'un diagnostic de maladie génétique qu'il est précis lorsque l'anomalie génétique en cause dans cette maladie a pu être identifiée par un test génétique. Il précise le diagnostic clinique. Identifier les variations du message dans un gène, par séquençage de son ADN, peut permettre d'accéder au diagnostic génétique. L'analyse de l'ADN d'un patient est effectuée à partir d'un échantillon de ses cellules sanguines, recueillies lors d'une prise de sang.



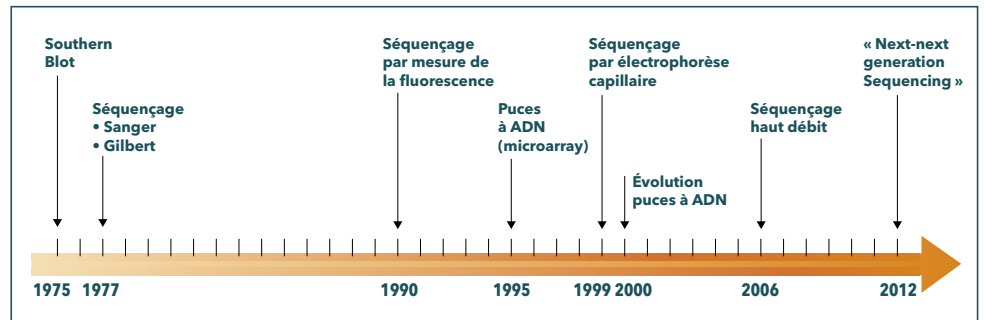
### ADN, nucléotide, gène, exome, séquençage, variants... De quoi parle-t-on ?

- **Génome.** Le génome humain est l'ensemble du matériel génétique d'un individu porté par l'ADN.
- **ADN.** L'ADN est le support moléculaire de l'information génétique. Il est formé par l'enchaînement de 4 nucléotides : ces 4 molécules sont constituées chacune par une base différente, A, T, C, G. L'enchaînement des nucléotides de l'ADN constitue sa séquence. Séquençer l'ADN signifie lire l'enchaînement linéaire des nucléotides d'un fragment d'ADN donné.
- **Gènes.** Notre génome contient environ 23 000 gènes répartis sur les chromosomes. Les gènes sont présents en double exemplaire, un sur chaque chromosome homologue. Ces versions peuvent être identiques ou différentes : ce sont les **allèles** (ou différentes versions d'un gène). Les gènes sont des portions d'ADN permettant la fabrication de protéines (la dystrophine est une protéine). Dans chaque gène, les nucléotides s'enchaînent d'une façon particulière, aboutissant à une information spécifique.
- **Exons.** Les régions codantes des gènes sont appelées les exons. C'est à partir des informations contenues dans les exons que la cellule fabrique les protéines : on dit qu'un gène « code » une protéine.
- **Exome.** C'est l'ensemble des régions codantes de notre génome (des 23 000 gènes). Elles représentent 34 millions de paires de bases de l'ADN.
- **Introns.** Les régions non codantes des gènes sont appelées les introns. Elles sont situées entre les régions codantes. Elles ne permettent pas de produire des protéines mais ont d'autres fonctions.
- **Variations dans l'ADN.** Environ 0,5 % du génome diffère d'un individu à l'autre, ce qui fait des millions de variations possibles. Elles peuvent apparaître dans les séquences codantes des gènes, les exons, ou dans d'autres régions.
  - Certaines de ces variations sont courantes dans la population (fréquence supérieure à 1 %) et ne sont pas pathogènes : on parle de **polymorphisme**. Elles rendent chaque individu unique.
  - D'autres variations dites "**anomalies génétiques ou mutations**" sont beaucoup plus rares et sont la cause de maladies.
  - Il existe aussi beaucoup de variations de l'ADN dont on ignore si elles sont ou non responsables d'une maladie : on parle alors de **variants génétiques**.

# LES DIFFÉRENTES TECHNIQUES de séquençage

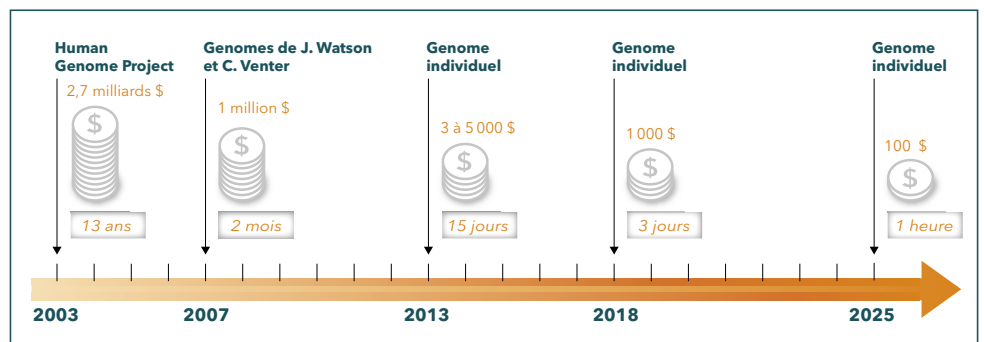
Selon la situation du patient et les besoins diagnostiques, différentes techniques de séquençage peuvent être utilisées. Elles permettent, soit de cibler l'analyse génétique lorsque l'on a déjà des informations assez précises sur le gène en cause, soit de rechercher des anomalies dans plusieurs gènes à la fois (panel de gènes) lorsque l'information n'est pas assez précise.

En 40 ans, les techniques de séquençage sont devenues plus performantes et moins coûteuses.



## De plus en plus d'informations obtenues en un temps record

Les techniques de séquençage de l'ADN se sont considérablement développées. Aujourd'hui, il est possible d'analyser de grandes portions d'ADN en même temps. Entre 2003 et 2018, le développement des techniques de séquençage aura permis de réduire le temps et le coût du séquençage de l'ADN de notre génome.



## Séquençage classique : la méthode de Sanger

Le séquençage classique fait référence à la méthode de Sanger, où l'on séquence l'ADN gène par gène. La méthode de Sanger (du nom de son inventeur en 1977) a été la première technique de séquençage de l'ADN.

- Avant l'implantation des techniques de séquençage de nouvelle génération (NGS) dans les laboratoires médicaux de génétique, la stratégie du diagnostic moléculaire des maladies rares d'origine génétique était itérative. On procédait à l'analyse gène après gène, et leur exploration était hiérarchisée en suivant des arbres décisionnels qui prenaient en compte les données cliniques, paracliniques et génétiques. Les analyses étaient basées, dans la majorité des cas, sur l'amplification par PCR de chaque exon du gène ou des gènes connus, suivie de leur séquençage selon la méthode de Sanger.

- Cette méthode de séquençage est encore utilisée aujourd'hui et pertinente lorsque les informations cliniques sont assez précises pour orienter la recherche vers un tout petit nombre de gènes (1 à 3) à séquençer et/ou pour confirmer des résultats obtenus avec les techniques de séquençage nouvelle génération (NGS) lorsque la mutation et le gène ont été identifiés.

Séquencer l'ADN ..... 2

Les différentes techniques de séquençage ..... 3

Les questions posées par le séquençage de nouvelle génération ..... 5

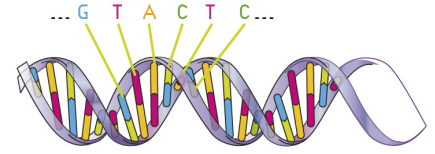
**Le séquençage par la méthode de Sanger**

Le principe de cette méthode de séquençage est de lire l'ADN par petits fragments. À partir d'une amorce qui se fixe sur l'ADN à séquencer, des brins d'ADN complémentaires sont fabriqués (polymérisation par PCR). Les arrêts aléatoires de la polymérisation à différents endroits de la séquence d'ADN permettent ensuite de lire cette séquence, nucléotide par nucléotide.

**Lire l'ADN : le séquençage**

L'ADN est formé de brins complémentaires composés de 4 bases différentes : A, T, C, G.

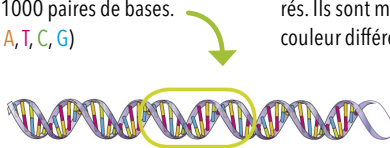
Séquencer un fragment d'ADN, consiste à déterminer l'ordre des bases A, T, C, et G qui le constituent.



**1 - Ingrédients de la réaction**

Principe de séquençage de l'ADN par la méthode de Sanger (1977) modernisée :

- Échantillon d'ADN avec la portion à séquencer de 100 à 1000 paires de bases.
- Nucléotides (A, T, C, G)
- Polymérase
- Amorce



**• En plus de la PCR :**

Les didésoxynucléotides (bases-stop) qui empêcheront l'addition d'autres bases une fois incorporés. Ils sont marqués par un agent fluorescent, une couleur différente pour chaque base.



**PCR : Polymerase Chain Reaction**

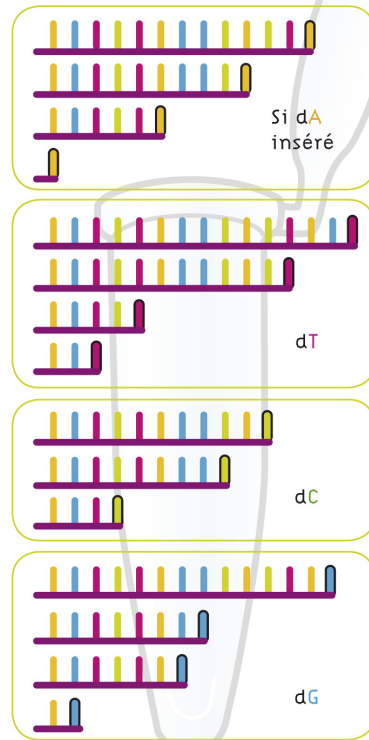
Ou réaction de polymérisation en chaîne. Cette technique permet de copier rapidement des séquences précises d'ADN en de très nombreux exemplaires. Ces copies se font à partir d'une **amorce**, un morceau d'ADN simple brin complémentaire d'une petite zone de l'ADN à séquencer, et grâce à une enzyme, l'**ADN polymérase** et des **nucléotides A, T, C, G**.

**Le génome humain**

Pour séquencer le génome humain, les 3,2 milliards de paires de bases qu'il contient ont été découpées en fragments d'environ 500 à 800 paires de base. Après avoir séquencé chaque fragment, le travail a consisté à les ordonner. Le Consortium international pour le séquençage du génome humain a réuni 20 centres de séquençage dans six pays pendant plus de 10 ans (de 1990 à 2003).

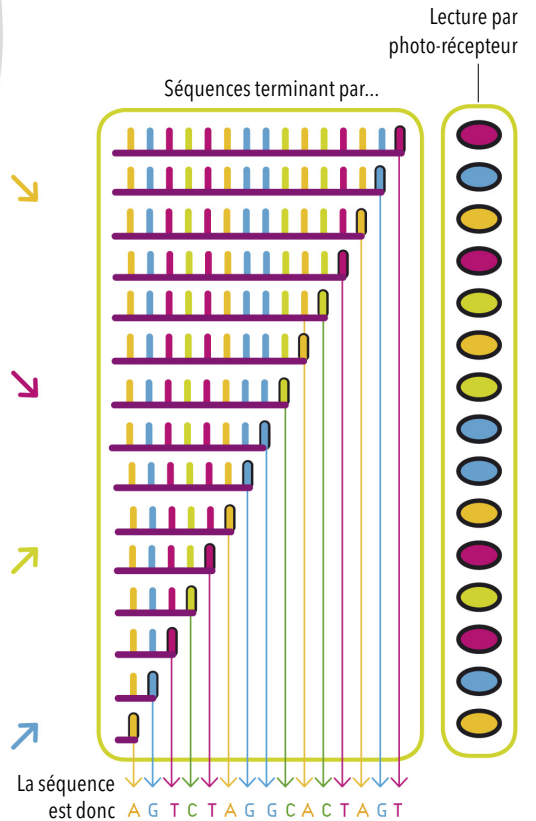
**2 - Réaction de polymérisation**

Principe de la PCR. Si une base-stop est insérée, la synthèse du fragment est interrompue. À la fin des cycles, le mélange contient donc des fragments de tailles différentes.



**3 - Électrophorèse :**

Migration des fragments en fonction de leur taille. Pour connaître le nom de la base-stop : détection de fluorescence (si jaune, c'est un dA...).



Limite de la technique : entre 600 et 1000 paires de bases séquencées par réaction.

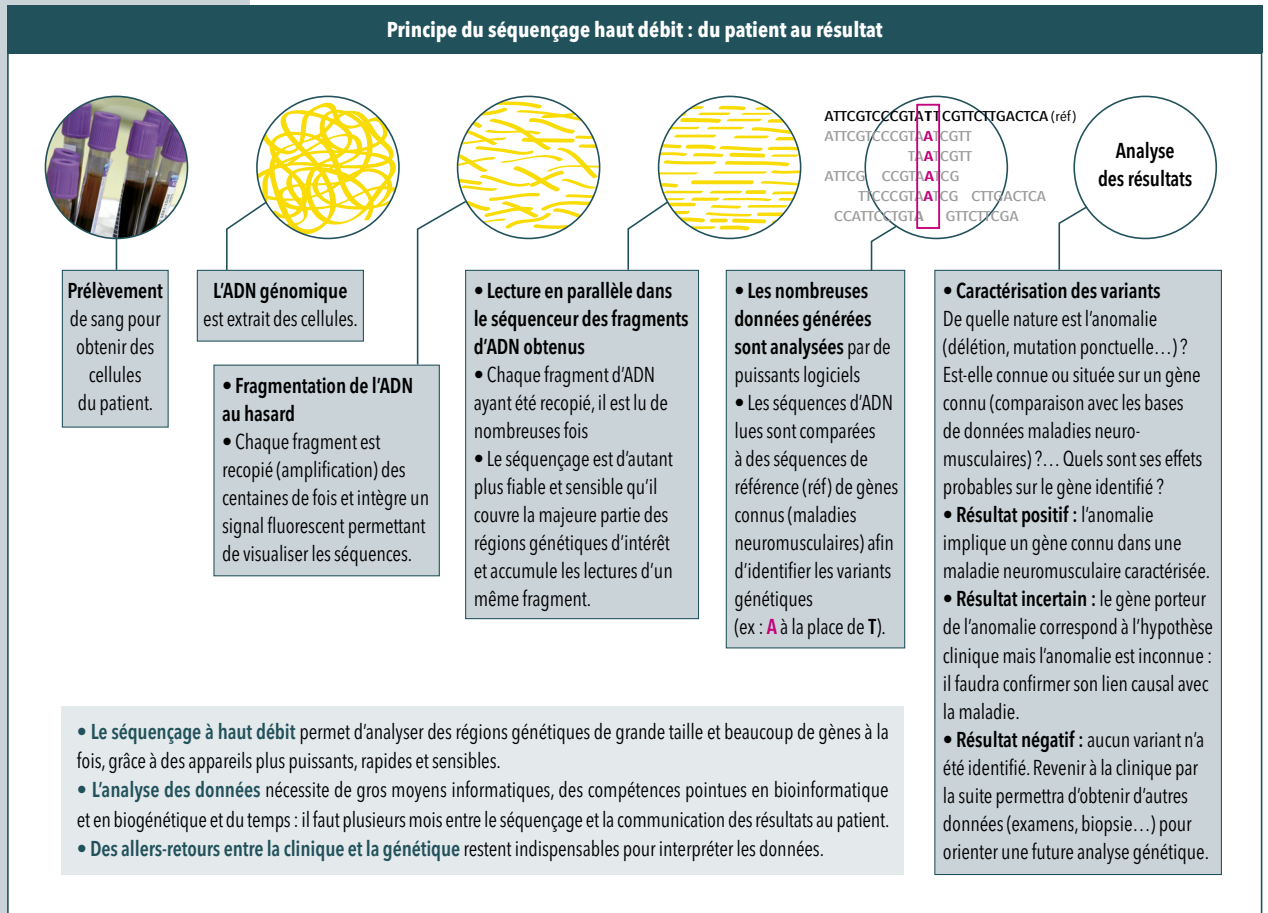
Séquencer l'ADN ..... 2

Les différentes techniques de séquençage ..... 3

Les questions posées par le séquençage de nouvelle génération ..... 5

## Depuis 2007 : les nouvelles techniques de séquençage (NGS)

Le terme « Nouvelle génération de séquençage à haut-débit » (**Next generation sequencing, NGS**) regroupe l'ensemble des technologies ou plateformes de séquençage développées depuis 2007 par quelques sociétés de biotechnologies.



## LES QUESTIONS POSÉES par le séquençage de nouvelle génération

**Le séquençage de nouvelle génération (NGS) permet d'accéder à un grand nombre de données génétiques. Ce qui nécessite au préalable de cibler ce qui doit être analysé afin d'être le plus pertinent possible.**

### Que séquence-t-on : panels de gènes, exome, génome ?

L'apport du séquençage de nouvelle génération (NGS) au diagnostic des maladies rares est considérable. Le NGS peut être utilisé selon 3 strates de complexité croissante :

1. analyse ciblée d'un **panel de gènes** (*Targeted Sequencing*), c'est-à-dire un groupe de gènes choisis pour leur implication dans une maladie ou un groupe de maladies ;
2. **exome** (*Whole Exome Sequencing*) ;
3. **génom**e entier (*Whole Genome Sequencing*).

Dans chaque situation, la qualité de l'évaluation clinique préalable reste primordiale pour définir ce qui va être analysé en NGS.

La puissance du NGS peut être d'abord utilisée pour réaliser l'analyse ciblée de gènes connus impliqués dans une maladie génétique.

Séquencer l'ADN ..... 2

Les différentes techniques de séquençage ..... 3

Les questions posées par le séquençage de nouvelle génération ..... 5

### Choisir l'approche dans les maladies neuromusculaires

En fonction de ce que l'on connaît de la maladie du patient, on pourra choisir entre ces approches :

- 1. méthode de Sanger**, lorsque la maladie est bien caractérisée avec peu de gènes impliqués ;
- 2. panel de gènes**, lorsque la maladie est caractérisée et que de nombreux gènes sont connus. Le panel pourra varier entre un panel ciblé sur certaines maladies neuromusculaires (ex LGMD, CMT) et un panel large (avec plusieurs centaines de gènes différents).
- 3. exome**, lorsque le phénotype de la maladie est caractérisé mais que les gènes majeurs connus ont été exclus ; dans ce cas-là, effectuer un prélèvement pour une analyse génétique des apparentés augmente de façon importante les chances d'obtenir un résultat.

Aujourd'hui, le diagnostic moléculaire des maladies liées à un seul gène (monogéniques) repose sur le NGS. Le NGS permet l'analyse simultanée en parallèle de tous les gènes déjà incriminés dans une maladie héréditaire donnée et regroupés dans un panel de gènes.

#### Une (r)évolution technologique dans les laboratoires français

Une première évolution technologique est actuellement en cours en France dans les laboratoires médicaux de génétique moléculaire. En effet, la DGOS (Direction générale de l'offre de soins), dans le cadre du 2<sup>e</sup> Plan national maladies rares, a octroyé aux laboratoires des financements destinés à l'acquisition de plateformes de NGS de moyen débit, ce qui permet progressivement d'augmenter le nombre de patients analysés et d'analyser un panel de gènes impliqués dans la maladie, et ainsi de réduire les délais et les coûts des analyses tout en améliorant les performances du diagnostic moléculaire.

### Analyse comparative des différentes méthodes de séquençage

Cible d'ADN	Taille de la région d'ADN*	Nombre de Gènes	Avantages	Inconvénients
<b>Cible petite</b>	10 Kb (10 000)	≈ 1 à 3	• Rapide, peu coûteux	• Non exhaustif
<b>Petit panel</b>	250 Kb (250 000)	≈ 20	• Rapide • Expertise de gènes • Bonne couverture de gènes • Peu coûteux	• Non exhaustif
<b>Grand panel</b>	1 Mb (1 000 000)	≈ 100-300	• Meilleure efficacité diagnostique • Bonne couverture de gènes • Découvertes fortuites ( <i>Incidental findings</i> ) limités à la pathologie neuromusculaire	• Interprétation plus longue et plus complexe • Pas d'expertise pour certains gènes • Besoin de mettre à jour régulièrement le panel (nouveaux gènes)
<b>Exome « clinique »</b>	12 Mb (12 000 000)	≈ 5 000	• Plus ciblé que l'exome entier donc diminution du risque de découverte de variants de signification mal connue	• On ne trouve que là où l'on cherche
<b>Exome entier (Whole Exome Sequencing)</b>	30 Mb (30 000 000) = toutes les parties codantes des gènes	≈ 20 000	• Exhaustivité • Pas de mise au point nécessaire par rapport à un nouveau panel • Augmente beaucoup les chances d'identifier de nouveaux gènes associés à une pathologie	• Coût ++ • Interprétation plus complexe • Risque de découvertes fortuites ( <i>Incidental findings</i> ) • Possibilité de découverte de variants de signification mal connue • Couverture moindre • Matériel : séquenceur, ressources informatiques
<b>Génome (Whole Genome Sequencing)</b>	3 GB (3 000 000 000) = tout l'ADN	≈ 20 000	• Le plus complet : séquences codantes et non codantes (inclut les séquences régulatrices)	• Coût ++++ • Interprétation délicate • Aujourd'hui, plus utilisé en recherche qu'en clinique, mais pourrait changer

\* En bases (b) (une base = un nucléotide d'ADN, A, T, C ou G).

**Les découvertes fortuites (*incidental findings*) : que dire au patient ?**

Les techniques d'étude globale du génome utilisées sont susceptibles de déceler des anomalies non recherchées (par exemple la perte d'un gène impliqué dans une prédisposition génétique aux cancers, alors que l'analyse est prescrite pour chercher la cause d'un retard mental) : c'est ce que l'on appelle des découvertes fortuites ou *incidental findings*.

Ces découvertes peuvent poser des questions éthiques quant à la communication ou non de leurs résultats (sans lien avec la recherche diagnostique en cours) à la personne concernée.

En amont de la recherche génétique, le patient doit donc être informé du risque que de tels éléments soient découverts : cela doit faire l'objet d'un consentement écrit de sa part.

**Des grandes quantités de données à analyser**

Une conséquence importante de cette révolution technologique est l'accroissement de la quantité de données de séquençage à exploiter à chaque analyse de NGS et donc la nécessité de renforcer les infrastructures informatiques. Cela augmente aussi les besoins humains, bio-informatiques, ainsi que le temps nécessaire à l'analyse de toutes ces données. Le délai pour obtenir un résultat de test génétique peut être long.

**La difficulté d'interprétation d'un variant rare**

Compte-tenu de la variabilité du génome humain, le grand défi de la génétique médicale à l'ère du NGS n'est plus la détection des variations du génome mais leur interprétation.

L'interprétation médicale d'un variant rare détecté chez un patient repose sur un faisceau d'arguments incluant :

1. la prédiction bio-informatique de l'effet biologique du variant en utilisant plusieurs algorithmes ;
2. la confrontation avec le tableau clinique du patient pour s'assurer que la voie biologique impactée par cette variation est cohérente avec la présentation clinique ;
3. les analyses de transmission de la maladie au sein de la famille : comparer l'ADN du patient avec celui de certains de ses apparentés (sains ou malades) permet de différencier les mutations pathogènes et les variants non pathogènes ;
4. l'analyse statistique des résultats avec des tests adaptés aux variations rares pour démontrer un enrichissement significatif de ces variations chez les patients par rapport à la population générale, ce qui nécessite de connaître la fréquence des allèles rares dans les différents groupes de population.

Séquencer l'ADN .....	2
Les différentes techniques de séquençage .....	3
Les questions posées par le séquençage de nouvelle génération.....	5

## EN SAVOIR +

**www.afm-telethon.fr**

Site internet de l'AFM-Téléthon.

**www.myobase.org**

Base documentaire sur les maladies neuromusculaires et le handicap moteur éditée par le service documentation de l'AFM-Téléthon.

**www.filnemus.fr**

Plan Un diagnostic pour chacun :  
<http://www.filnemus.fr/menu-filiere/evenements/actualites/article/news/la-plan-un-diagnostic-pour-chacun/>

Cette Fiche Technique Savoir & Comprendre a été rédigée dans le cadre de l'action prioritaire « Un diagnostic pour chacun » et validée par le Comité de pilotage AFM-Téléthon/Filnemus du Plan Diagnostic.



Association reconnue d'utilité publique

1, rue de l'Internationale - BP 59 - 91002 Évry cedex  
 Tél. : 33 (0)1 69 47 28 28 - Fax : 33 (0)1 60 77 12 16  
 Siège social : AFM - Institut de Myologie  
 47-83, boulevard de l'Hôpital - 75651 Paris cedex 13  
[www.afm-telethon.fr](http://www.afm-telethon.fr)