

Bases moléculaires des mutations et Bases moléculaires du mode de transmission des maladies génétiques

Collège National des Enseignants et Praticiens de Génétique Médicale

Martin Krahn

Département de Génétique Médicale - Hôpital Timone Enfants, AP-HM,
Et INSERM UMR910 - Faculté de Médecine, Université de la Méditerranée
Marseille

Date de création du document 2010-2011

Table des matières

I Polymorphismes du génome	4
II Classification des lésions du génome.....	5
III Microlésions du génome.....	6
III.1 Généralités.....	6
III.2 Principaux types de microlésions et mécanismes de survenue.....	7
III.3 Conséquences des microlésions et impact sur les modes de transmission *.....	8
III.4 Conséquences des microlésions en fonction de leur localisation et de leur type.....	10
III.4.1 Conséquence des substitutions en séquence codante.....	11
III.4.2 Conséquence des insertions et/ou délétions de nucléotides en séquence codante... 	12
IV De la connaissance des types de microlésions à la conclusion de données mutationnelles : l'Interprétation de données mutationnelles.....	13

Note : Les notions importantes relatives aux modes de transmission des maladies génétiques sont indiquées par le symbole de l'étoile *.

INTRODUCTION

Le terme « mutation » désigne n'importe quel changement intervenu dans la séquence de l'ADN, sans préjuger de sa pathogénicité à l'échelle du gène ou du chromosome. On parle aussi de « variants ».

Il est important de souligner d'emblée qu'on attribue souvent à tort une connotation pathologique à ce terme de « mutation ». Mais les variations non pathogènes de l'ADN (appelées « polymorphismes ») sont par définition également des mutations.

Les mutations sont le moteur de l'évolution, et source de la diversité entre individus. Mais elles sont aussi à l'origine des maladies génétiques monogéniques et des prédispositions génétiques aux maladies polyfactorielles. Le caractère pathogène d'une mutation pourra être précisé en parlant de « mutation délétère » ou « mutation pathogène ».

La conséquence de toute mutation dépend de son effet fonctionnel, qui peut être neutre, conduire à l'amélioration d'une fonction (diversité, évolution) ou à l'altération d'une fonction (effet pathogène).

Dans une cellule vivante, l'ADN est en permanence exposé à différents types d'agression pouvant conduire à l'apparition de mutations. Il s'agit essentiellement d'agressions « exogènes » (radiations et agents génotoxiques de l'environnement), d'agressions endogènes (radicaux libres, ...), d'erreurs de réplication et d'accidents de recombinaison. La cellule possède une machinerie de réparation, qui corrige la plupart des anomalies. Mais un échappement au système de réparation est possible: c'est l'origine des mutations.

- *Mutations acquises :*

Une mutation apparue dans une cellule somatique d'un tissu est appelée « mutation somatique » ou « mutation acquise », puisqu'elle n'était pas présente initialement dans le génome de la cellule. Les mutations somatiques peuvent être à l'origine d'un clone cellulaire porteur de cette mutation, ne touchant qu'un seul ou quelques tissus, mais ne sont en revanche pas transmissibles à la descendance. Les mutations somatiques pathogènes sont notamment impliquées dans la formation de cellules tumorales.

- **Mutations constitutionnelles :**

Lorsqu'une mutation est présente ou survient avant la fécondation (soit nouvellement apparue, soit transmise de génération en génération), ou survient lors des premières divisions du zygote (donc nouvellement apparue), on parle de « mutation constitutionnelle ». Une mutation constitutionnelle sera présente dans toutes les cellules somatiques de l'individu, et également dans ses cellules germinales, donc transmissible à la descendance. Toute mutation nouvellement apparue est aussi appelée mutation « de novo » ou « néomutation ».

Certaines mutations surviennent lors de la méiose dans une cellule germinale, au niveau d'un gamète parental, et sont appelées « mutations germinales ». Les mutations germinales seront donc forcément présentes de façon « constitutionnelle » chez l'individu issu de ce gamète, qui sera donc porteur d'une mutation « de novo » ou « néomutation », non présente dans les cellules somatiques du parent qui lui a transmis cette mutation.

Les mutations constitutionnelles pathogènes, « de novo » ou transmises de génération en génération, sont à l'origine des maladies génétiques monogéniques et des maladies génétiques chromosomiques.

I POLYMORPHISMES DU GÉNOME

Comme cela a déjà été souligné précédemment, le terme de « mutation » désigne n'importe quel changement intervenu dans la séquence de l'ADN, sans préjuger de sa pathogénicité à l'échelle du gène ou du chromosome.

Avant de s'intéresser aux effets délétères des variations du génome, il faut rappeler l'importance des variations non pathogènes du génome, qui sont la base de la diversité entre les individus.

Ces variations non pathogènes du génome sont appelées « polymorphismes ». La notion de « polymorphisme » repose à la fois sur le caractère non pathogène de la variation de séquence, et la fréquence dans la population (>1% par définition). Un polymorphisme est une mutation et peut se situer en région codante ou non codante.

Différents types de polymorphismes ont été caractérisés, parmi lesquels les plus importants à retenir dans le contexte actuel sont :

- **Les SNPs (Single Nucleotide Polymorphisms):**

Il s'agit de polymorphismes de substitution au niveau d'un nucléotide (variation de séquence ponctuelle). Les SNPs sont très nombreux (>107 par génome humain) et répartis dans tout le génome (environ 1 SNP tous les 300 pb). Les SNPs sont référencés dans la base de données dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>).

- **Les CNVs (Copy Number Variations):**

Il s'agit de variation du nombre d'exemplaires contigus de grands segments génomiques (perte ou gain de fragments de quelques kb à plusieurs Mb). A ce jour, des CNVs ont été identifiés dans environ 15% du génome humain.

Les CNVs sont référencés dans la base de données Database for Genomic Variants : (<http://projects.tcag.ca/variation/>) ; avec plus de 65000 CNVs différents rapportés à ce jour).

- **Les polymorphismes de répétition:**

Il s'agit de séquences répétées en tandem de nombreuses fois, à partir de motifs de longueur variable (de quelques, à plusieurs centaines de paires de bases définissant en fonction de la taille les Microsatellites, Minisatellites, Satellites, et Mégasatellites).

II CLASSIFICATION DES LÉSIONS DU GÉNOME

On distingue les anomalies à l'échelle du chromosome appelées macrolésions et les anomalies à l'échelle du gène, appelées microlésions.

- *Macrolésions du génome*

Les anomalies chromosomiques sont classées en anomalies chromosomiques de nombre et anomalies chromosomiques de structure. Les anomalies de structure comportent différents types que nous n'allons pas détailler ici, par exemple les translocations, délétions, duplications de fragments chromosomiques, etc.

NB : les notions plus précises concernant les macrolésions du génome sont abordées plus en détail dans le chapitre intitulé « Types, fréquences et mécanismes de formation des anomalies chromosomiques » (D. Sanlaville).

- *Microlésions du génome*

A l'échelle du gène, les anomalies du génome sont surtout des substitutions appelées également mutations ponctuelles qui consistent en le remplacement d'un nucléotide par un autre. Il peut également s'agir de l'insertion et/ou de la délétion de quelques nucléotides et parfois jusqu'à quelques dizaines ou centaines de nucléotides.

- *Taille des lésions et méthodes d'étude*

A l'échelle du chromosome les anomalies sont recherchées par des approches qu'on appelle de génétique chromosomique (**cytogénétique**).

A l'échelle du gène, les anomalies sont recherchées par ce qu'on appelle les approches de **génétique moléculaire**.

Entre les deux, des techniques de **cytogénétique moléculaire** peuvent être utilisées. Il y a donc une différence de résolution des techniques ; en cytogénétique, il est possible de détecter des anomalies de quelques millions de paires de bases, alors qu'en génétique moléculaire, on recherchera des anomalies allant d'une seule paire de base à quelques milliers de paires de base. Il est à noter que le développement technologique actuel, et notamment les techniques d'analyses mutationnelles à haut débit, permettent de plus en plus d'élargir le champ de résolution.

III MICROLÉSIONS DU GÉNOME

III.1 GÉNÉRALITÉS

Les microlésions du génome constituent des anomalies à l'échelle du gène, en séquence codante ou non codante.

Il s'agit notamment :

- de substitutions, qui consistent en le remplacement d'un nucléotide par un autre
- d'insertions et/ou de délétions de 1 ou quelques nucléotides
- d'insertions et/ou délétions de quelques dizaines à centaines de nucléotides
- de mutations instables

NB : D'autres événements mutationnels plus rares existent, mais ne seront pas développés ici (inversions, mutagenèse induite par des éléments mobiles,...).

Un événement d'insertion consistant en la multiplication d'une séquence donnée est appelé « duplication » ($N=2$) ou « amplification » ($N>2$).

Le terme « mutation ponctuelle » est utilisé habituellement pour les microlésions touchant un ou quelques nucléotides (substitutions, insertions et/ou délétions de un ou quelques nucléotides).

Les microlésions peuvent être constitutionnelles (notamment dans le cadre des maladies

génétiques monogéniques) ou acquises (notamment impliquées dans la formation de cellules tumorales).

III.2 PRINCIPAUX TYPES DE MICROLÉSIONS ET MÉCANISMES DE SURVENUE

- *Substitutions*

Les substitutions constituent le remplacement d'un nucléotide par un autre nucléotide. Il s'agit du type le plus fréquent des microlésions, et globalement de loin le plus fréquent des remaniements affectant le génome, puisqu'elles représentent environ 70% des mutations. On distingue classiquement les transversions et les transitions.

Les *transitions* correspondent au remplacement d'une des purines (Adénine ou Guanine) par l'autre purine, ou d'une des pyrimidines (Cytosine ou Thymine) par l'autre pyrimidine.

Les *transversions* en revanche sont un changement d'une des pyrimidines en l'une des purines, ou le contraire, d'une des purines en l'une des pyrimidines.

Plusieurs mécanismes peuvent être en jeu dans la survenue des substitutions, et notamment des erreurs de réplication ayant échappé au système de réparation, des erreurs du système même de réparation, ou des perturbations biochimiques dues à des agents physiques ou chimiques exogènes ou produits par le métabolisme endogène.

- *Insertions et/ou de délétions de 1 ou quelques nucléotides*

Au cours du phénomène de réplication, des accidents de « dérapage réplicatif », impliquant les ADN polymérases, peuvent survenir, notamment au niveau de certaines séquences répétées. Ceci peut conduire à l'insertion (gain) et/ou à la délétion (perte) d'un ou de quelques nucléotides supplémentaires par rapport à la séquence initiale.

- *Insertions et/ou délétions de quelques dizaines à centaines de nucléotides*

Les microlésions de type insertion et/ou délétion de nucléotides peuvent concerner dans certains cas un grand nombre de nucléotides, de quelques dizaines à quelques centaines. Ces événements mutationnels peuvent impliquer des fragments, voire la totalité, d'un ou de plusieurs exons et/ou introns. Le mécanisme mutationnel est alors différent par rapport aux insertions et/ou délétions de un à quelques nucléotides, et fait suite à des réparations incomplètes de lésions de l'ADN, ou à des anomalies de recombinaison ou de réplication

- *Mutations instables*

Plusieurs notions importantes sont rattachées aux mutations instables : les notions d'**instabilité**, de **seuil**, de **prémuation** et **mutation complète**, et d'**anticipation**.

Certaines régions du génome présentent des répétitions de motifs de séquence d'ADN. Il peut s'agir de motifs dinucléotides (par exemple (TG)_n), trinucleotides (par exemple (CAG)_n), tétranucleotides, etc.

Ces répétitions peuvent être **instables**, c'est-à-dire avoir une tendance importante à la modification du nombre de répétitions du motif de base, au cours du phénomène de réplication. Cette modification est habituellement une expansion du nombre de répétitions, mais beaucoup plus rarement il peut également s'agir aussi d'une contraction. Elle survient surtout au cours de la réplication préméiotique (instabilité méiotique), donc lors de la transmission à la descendance. De plus, une instabilité mitotique peut exister. L'instabilité résulte d'un phénomène de dérapage répliatif (décrit ci-dessus), et peut concerner des régions codantes ou non codantes.

Le nombre de répétitions du motif de base est variable dans la population générale, mais se situe en dessous d'un **seuil**. En-dessous de ce seuil, la transmission de la répétition est stable de génération en génération. Par contre, au-delà de ce seuil, il y a instabilité et possibilité d'expansion du nombre de répétitions. Un nombre de répétitions modéré au-dessus du seuil constitue une « **prémutation** », avec une tendance à l'expansion, mais habituellement sans effet pathogène (phénotype habituellement normal, mais des exceptions existent selon les pathologies). Lorsque le nombre de répétitions dépasse une valeur limite au-dessus du seuil, entraînant l'apparition de la pathologie, on parle de « **mutation complète** ».

Les mutations instables sont impliquées notamment dans la survenue de certaines maladies neurodégénératives et neuromusculaires*.

On observe dans les maladies causées par des mutations instables un biais de transmission parentale des formes les plus sévères, et une augmentation au cours des générations successives du risque de développer la maladie, ou de la sévérité ou précocité des signes (phénomène « d'**anticipation** »).

III.3 CONSÉQUENCES DES MICROLÉSIONS ET IMPACT SUR LES MODES DE TRANSMISSION *

Les conséquences délétères des microlésions du génome dépendent essentiellement de leur **localisation** et de leur **type**.

Il est important de souligner que l'effet délétère d'une microlésion peut consister en un impact fonctionnel au niveau de l'ARN messenger et/ou de la protéine codée par le gène muté.

En effet, l'exemple le plus « classique » d'une microlésion est une anomalie de la séquence codante d'un gène, conduisant à une anomalie de la séquence en acides aminés de la protéine codée par ce gène, responsable d'une altération qualitative et/ou quantitative de la protéine. Mais les effets délétères possibles des microlésions sont multiples et complexes, dépassant largement la linéarité directe entre la séquence codante au niveau du gène et la séquence en acides aminés au niveau de la protéine.

Pour toute microlésion, il faut donc prendre en compte l'impact fonctionnel éventuel au niveau de l'ARN messager et/ou de la protéine codée. Il s'agit là d'une notion primordiale à retenir pour l'interprétation des données mutationnelles dans le cadre du diagnostic moléculaire.

Les conséquences délétères des microlésions sont classées en deux grandes catégories : la **perte de fonction** et le **gain de fonction**.

- ***Perte de fonction**** :

On désigne par **perte de fonction** un effet délétère du à la **diminution ou l'abolition de production de la protéine active, sur le plan quantitatif (niveau de synthèse de la protéine) et/ou qualitatif (fonctionnalité de la protéine)**. L'effet délétère de type perte de fonction se manifeste lorsque le niveau résiduel de protéine fonctionnelle passe en dessous d'un seuil, **et constitue la cause majoritaire des maladies récessives**. Lorsqu'une seule des deux copies d'un gène est mutée chez un individu, la synthèse d'une protéine normale par l'allèle non muté suffit habituellement pour maintenir la fonction cellulaire correspondante. Par contre, l'effet délétère se manifeste lorsque les deux copies du gène sont mutées.

Dans certains cas, la perte de fonction peut toutefois conduire à un effet délétère dominant. La présence d'une seule copie mutée du gène, à l'état hétérozygote, est alors suffisante pour franchir le seuil. Cette situation est appelée « **haploinsuffisance** », l'autre copie du gène n'étant pas suffisante pour compenser le déficit.

La perte de fonction peut résulter d'un effet délétère au niveau de l'ARN messager : par effet sur la régulation de la transcription, par altération de la maturation de l'ARNm (notamment l'épissage), par altération de la stabilité de l'ARNm entraînant sa destruction. Ceci va conduire à une diminution ou absence de production de la protéine active.

La perte de fonction peut également résulter d'un effet délétère au niveau de la protéine, qui peut être produite, mais instable et dégradée ; ou encore produite mais non fonctionnelle.

- ***Gain de fonction**** :

Le « gain de fonction » est un effet délétère du à l'acquisition d'une nouvelle fonction qui est délétère pour la cellule. Il s'agit de la cause majoritaire des maladies dominantes.

Le gain de fonction résulte habituellement d'un effet délétère au niveau de la protéine. Il peut s'agir tout d'abord d'un effet appelé « **dominant négatif** » : le produit protéique de l'allèle muté antagonise le produit de l'allèle normal (en particulier lorsque le produit du gène agit sous la forme de dimère ou polymère). Une autre possibilité est un effet « **toxique** », lorsque la mutation conduit à l'acquisition d'une nouvelle fonction cellulaire délétère, ou suite à un excès de fonctionnement.

L'effet délétère peut aussi se situer au niveau de l'ARN messenger, entraînant par exemple la séquestration d'ARNm mutés, ou la séquestration de protéines par des ARNm mutés. Mais les mécanismes exacts du gain de fonction délétère au niveau de l'ARNm restent encore mal élucidés à ce jour.

III.4 CONSÉQUENCES DES MICROLÉSIONS EN FONCTION DE LEUR LOCALISATION ET DE LEUR TYPE

La majorité des microlésions délétères est localisée en séquence codante, avec un effet direct sur la séquence en acides-aminés de la protéine correspondante. Plus rarement, des microlésions en région codante ou non-codante peuvent avoir un effet délétère sur la régulation de l'expression d'une protéine.

L'information génétique pour la synthèse des protéines est contenue majoritairement dans les exons. L'enchaînement en nucléotides de la séquence codante détermine, grâce au code génétique, l'enchaînement en acides-aminés de la protéine qui sera synthétisée. L'effet délétère des microlésions en séquence codante dépend alors essentiellement du *type* de microlésion (détaillé dans le paragraphe suivant). Mais le processus de synthèse protéique dépasse la simple linéarité entre la séquence codante génomique et la séquence en acides-aminés de la protéine correspondante, et fait intervenir des mécanismes complexes de régulation de l'expression. Les phénomènes d'organisation dynamique des loci génomiques, de transcription, d'expression et de maturation protéique sont ainsi contrôlés par des séquences régulatrices, qui peuvent être localisées hors de la séquence codante (introns, exons non-codants, régions intergéniques, ...), voire même directement dans la séquence codante. Toute microlésion altérant l'un de ces phénomènes peut avoir un effet délétère.

- *Microlésions en séquence codante*

NB : Par la linéarité du code génétique, une mutation en séquence codante génomique peut avoir un effet directement transposé au niveau de la séquence en acides-aminés de la protéine codée. Mais comme souligné auparavant, il ne faut pas oublier que toute mutation en séquence codante peut avoir un éventuel effet délétère sur l'ARN messager. **Pour toute microlésion, il faut donc prendre en compte l'impact fonctionnel éventuel au niveau de l'ARN messager et/ou de la protéine codée.**

III.4.1 Conséquence des substitutions en séquence codante

- Mutation de type « faux-sens »: le codon muté code un autre acide aminé*

La modification d'acide-aminé au niveau de la protéine peut être tolérée par la cellule sans conséquence délétère, ce qui explique que de nombreuses variations de séquence de type « faux-sens » n'ont pas d'effet pathogène, et constituent par ailleurs une part importante des polymorphismes (de type SNPs). Mais en fonction de la localisation de l'acide-aminé touché, les mutations faux-sens peuvent avoir des effets délétères (altération du repliement protéique, de la stabilité protéique, de domaines fonctionnels, de sites d'interaction avec d'autres protéines, etc.), de type perte de fonction ou gain de fonction.

- Mutation de type « non-sens »: le codon muté code un codon stop*

Ce type de mutation est généralement pathogène, responsable de la synthèse d'une protéine tronquée, qui sera instable et dégradée (effet perte de fonction), ou avec un effet dominant négatif (effet gain de fonction).

- Le cas particulier des mutations de type « isosémantique » ou « synonyme »: le codon muté code pour le même acide aminé

Le code génétique étant « dégénéré » (plusieurs codons pouvant coder un même acide-aminé), certaines substitutions au niveau de la séquence génomique ne modifient théoriquement pas la séquence en acides-aminés de la protéine correspondante, donc seraient sans effet pathogène. Ces mutations « isosémantiques » ont ainsi également été appelées « silencieuses ». Comme de nombreuses variations de type « faux-sens », les mutations « isosémantiques » constituent aussi une part importante des polymorphismes (de type SNPs). Cependant, depuis une dizaine d'années, il a été clairement démontré que certaines mutations « isosémantiques » peuvent avoir un effet délétère, résultant non d'une modification directe de la séquence protéique, mais d'un effet délétère de la mutation génomique sur un motif de séquence nucléotidique (par exemple effet sur un motif impliqué dans l'épissage, sur un motif impliqué dans la régulation du niveau d'expression,

etc.) Ce type de mutation est donc le meilleur exemple soulignant que pour toute microlésion, il faut prendre en compte l'impact fonctionnel non seulement au niveau de la protéine codée, mais aussi au niveau de l'ARN messager.

III.4.2 Conséquence des insertions et/ou délétions de nucléotides en séquence codante

La conséquence des insertions et/ou délétions de nucléotides dépend schématiquement de la conséquence sur le cadre de lecture, défini par la succession des codons constituant la séquence codante. Puisque chaque codon comporte trois nucléotides, deux situations sont possibles :

- **Les insertions et/ou délétions de multiples de trois nucléotides** , n'entraînant pas de décalage du cadre de lecture. La conséquence au niveau protéique pourra être un gain ou une perte en acides-aminés, avec éventuellement un changement d'acide-aminé par rapport à la séquence initiale, au niveau de la jonction résultant de l'insertion et/ou la délétion. Le retentissement fonctionnel est variable selon la localisation au niveau de la protéine : l'insertion et/ou la délétion de nouveau(x) acide(s)-aminé(s) peut être « tolérée », ou délétère. Un effet délétère important peut aussi résulter de la création d'un codon stop à la jonction résultant de l'insertion et/ou la délétion.

- **Les insertions et/ou délétions de non-multiples de trois nucléotides**, responsables d'un décalage du cadre de lecture, qui entraînera la survenue prématurée d'un codon stop (ou dans de rares cas un décalage du codon stop en aval). L'effet délétère sera donc semblable à l'effet des mutations non-sens: synthèse d'une protéine tronquée, qui sera instable et dégradée (effet perte de fonction), ou avec un effet dominant négatif (effet gain de fonction)*.

Il y a donc un retentissement fonctionnel sévère expliquant que ce type de mutation est généralement pathogène.

- ***Microlésions en séquence non-codante***

Les microlésions en séquence non-codante peuvent également être de type substitution ou délétions/insertions, de la même manière que les microlésions en séquence codante. Les séquences non-codantes (comportant notamment les introns et les régions intergéniques) constituent plus de 95% du génome, et ont longtemps été considérées comme non-fonctionnelles (« junk DNA » ou « ADN poubelle »). Mais il est aujourd'hui clairement établi que l'ADN non-codant comporte des séquences régulatrices essentielles pour l'expression des gènes.

En conséquence, des mutations en séquence non-codante peuvent avoir des effets délétères en altérant des séquences régulatrices. Il peut s'agir par exemple d'effets délétères sur la

régulation de la transcription (mutations du promoteur, d'Enhancers, de Silencers,...), de la maturation de l'ARN messager (et surtout l'épissage) ou de la stabilité de l'ARN messager. Encore une fois, l'effet délétère peut se situer au niveau de l'ARN messager (effet direct quantitatif ou qualitatif) ou au niveau protéique (effet indirect quantitatif ou qualitatif).

Il est à souligner que les techniques routinières de génétique moléculaire, et en particulier le séquençage direct (méthode de Sanger), imposent des limites en termes de taille de séquence analysable. Ceci constitue la principale raison pour laquelle **une analyse mutationnelle d'un gène porte le plus souvent essentiellement sur les exons du gène (analyse de la séquence codante), et les bornes introniques flanquantes (analyse des sites donneurs et accepteurs d'épissage). C'est en général dans ces régions que sont concentrées la majeure partie des mutations délétères.** Avec les techniques routinières, il serait trop lourd, long et coûteux d'analyser pour un gène donné la totalité du locus génomique correspondant, dont la taille est au moins dix fois plus grande que la séquence codante. Ainsi, une grande partie des régions non-codantes d'un gène donné n'est donc pas analysée en routine, ne permettant pas de mettre en évidence d'éventuelles mutations délétères dans ces régions.

L'avènement des techniques d'analyse moléculaire à haut débit, en particulier le séquençage à haut débit, devrait permettre d'être plus exhaustif.

IV DE LA CONNAISSANCE DES TYPES DE MICROLÉSIONS À LA CONCLUSION DE DONNÉES MUTATIONNELLES : L'INTERPRÉTATION DE DONNÉES MUTATIONNELLES

Qu'il s'agisse d'application à visée de recherche ou diagnostique, l'objectif des analyses mutationnelles est d'identifier des variations de séquence, et de conclure sur leur caractère pathogène ou non.

- *Nomenclature des mutations*

L'identification de variations de séquence dans un échantillon, en comparaison à une séquence de référence, nécessite dans un premier temps une description précise de ces variations. La « Human Genome Variation Society » a établi une nomenclature officielle internationale pour la description des données mutationnelles (<http://www.hgvs.org/mutnomen>). Cette nomenclature permet une description précise de variations de séquences d'un gène, sur le plan génomique, transcriptionnel (ARNm) et protéique.

De manière simplifiée, la règle consiste à décrire la localisation de la variation de séquence, par rapport à la séquence codante du gène, et le changement induit. Selon la technique de génétique moléculaire utilisée (analyse génomique ou transcriptionnelle), la description est faite par rapport à la séquence codante sur le plan génomique (indiqué par « c. ») ou sur le plan de l'ARN messager (indiqué par « r. »). En supplément, l'effet théorique attendu au niveau protéique est indiqué entre parenthèses, précédé de « p. ». La séquence de référence utilisée doit être indiquée.

Exemple : pour un patient porteur de deux mutations à l'état hétérozygote dans le gène de la dysferline (séquence de référence GenBank NM_003494.2) :

Exon 9 : c.895G>T (p.Gly299Trp) HTZ : il s'agit d'une substitution d'une guanine par une thymine dans l'exon 9 du gène de la dysferline, responsable théoriquement du remplacement d'un acide-aminé glycine par un acide-aminé tryptophane en position 299 de la séquence protéique, donc une mutation de type faux-sens.

Exon 18 : c.1617C>G (p.Tyr539*) HTZ : il s'agit d'une substitution d'une cytosine par une guanine dans l'exon 18 du gène de la dysferline, responsable théoriquement du remplacement d'un acide-aminé tyrosine par un codon STOP en position 539 de la séquence protéique, donc une mutation de type non-sens.

- **Interprétation de données mutationnelles**

La nomenclature constitue donc la première étape importante dans l'interprétation des données mutationnelles, en déterminant le type de mutation (faux-sens, non-sens, isosémantique, etc.). Comme cela a été détaillé plus haut, certains types de mutations ont un effet pathogène hautement probable (par exemple des mutations non-sens), alors que pour d'autres types la conclusion peut être plus difficile (par exemple des mutations faux-sens).

Après avoir précisément nommé la mutation, deux situations sont possibles : cette variation de séquence a déjà été décrite au préalable, ou non. Cette information sera obtenue par la consultation de bases de données mutationnelles. Un portail de la plupart des bases de données est disponible sur le site web de la « Human Genome Variation Society » (<http://www.hgvs.org>).

De nombreuses bases de données mutationnelles existent aujourd'hui, dont certaines collectent des données concernant la totalité du génome humain (*bases de données « centrales » ou « globales »*), et notamment des informations sur les polymorphismes alimentées par des projets de grande envergure de séquençage de génomes. Parmi les plus utilisées on trouve : UCSC Genome Browser (<http://genome.ucsc.edu>) ; Ensembl (<http://www.ensembl.org>) ; SNP database (<http://www.ncbi.nlm.nih.gov/SNP>) ; Human Gene Mutation Database (<http://www.hgmd.org>).

D'autres bases de données sont dédiées spécifiquement à un gène donné : ces **bases de données appelées « locus-spécifiques »** sont mises à jour par des spécialistes travaillant sur le gène concerné, ce qui permet une mise à jour très précise. Il n'en existe malheureusement pas encore pour tous les gènes.

Dans la démarche d'interprétation de données mutationnelles, la consultation de bases de données permet donc de vérifier si une variation de séquence a déjà été rapportée au préalable.

Si la variation de séquence a déjà été rapportée, les informations disponibles peuvent permettre de savoir si elle a un caractère délétère qui a déjà été confirmé au préalable chez d'autres patients, ou au contraire si elle a été identifiée sans effets pathologiques dans la population générale (polymorphisme). Ceci permet souvent de conclure sur le caractère pathogène ou non.

La situation est plus difficile pour les variations de séquence non rapportées au préalable, et la conclusion sur le caractère délétère ou non doit alors prendre en compte différents éléments. Il s'agit notamment du type de mutation (non-sens, faux-sens, etc.), de l'étude de la ségrégation de la variation de séquence à l'intérieur de la famille, et de la recherche de la variation dans une population de témoins sains. Dans certains cas, il est nécessaire de recourir à des tests fonctionnels (évaluation de l'épissage, fonctionnalité de la protéine, etc.), mais ceci est difficile en routine diagnostique. Une progression importante a été possible grâce au développement d'**outils bioinformatiques, permettant la modélisation et la prédiction de l'effet fonctionnel de variations de séquences**. Certains algorithmes permettent par exemple de prédire l'effet d'une mutation sur l'épissage, ou encore d'évaluer l'effet fonctionnel du remplacement d'un acide-aminé par un autre. Ces outils bioinformatiques deviennent de plus en plus performants, et apportent dorénavant une aide importante voire incontournable dans l'interprétation des données mutationnelles.

L'évaluation de ce faisceau d'éléments constitue souvent la difficulté de l'interprétation des données mutationnelles, et ne permet malheureusement pas toujours d'aboutir à une conclusion formelle. Avec l'avènement des techniques d'analyse moléculaire à haut débit, et la génération de plus en plus facile d'importantes quantités de données mutationnelles, l'interprétation risque de devenir un goulot d'étranglement et le développement d'outils bioinformatiques performants et adaptés est essentiel.

NB : des notions plus détaillées concernant les « Bases de données et outils bioinformatiques utiles en Génétique » sont abordées dans le chapitre rédigé par C. Bérout.

Pour en savoir plus...

L'ouvrage de référence recommandé aux lecteurs pour approfondir les notions abordées, est le livre « Biologie moléculaire et Médecine » de Jean-Claude Kaplan et Marc Delpech (Médecine Science, Flammarion).
