

## Mécanismes et conséquences des mutations

Nadine Hanna, Béatrice Parfait,  
Dominique Vidaud, Michel Vidaud

► L'identification des mutations à l'origine de maladies génétiques chez l'homme a pris ces dernières années un essor considérable. Il est devenu possible d'établir le spectre des mutations délétères pour une maladie génétique donnée, et des bases de données internationales sont aujourd'hui accessibles *via* le réseau Internet. Le diagnostic génotypique des maladies héréditaires occupe actuellement une place prépondérante en matière de conseil génétique et de diagnostic prénatal. La connaissance du type de mutation délétère et des mécanismes en cause est essentielle pour déterminer la stratégie de diagnostic moléculaire adaptée à chaque situation. Cet article a pour objectif de présenter les différents types de mutations responsables de maladies génétiques (substitutions nucléotidiques, délétions ou insertions de petite taille, mutations dynamiques, grands remaniements...) et de récapituler les connaissances actuelles concernant les mécanismes moléculaires à l'origine de ces mutations. Leurs conséquences sur l'expression du gène (transcription et maturation du transcrit) et sur la fonction de la protéine sont également abordées dans cet article. ◀



Laboratoire de Génétique Moléculaire EA 3618, Université René Descartes Paris 5, Faculté des Sciences Pharmaceutiques et Biologiques, 4, avenue de l'Observatoire, 75006 Paris, France.  
[michel.vidaud@univ-paris5.fr](mailto:michel.vidaud@univ-paris5.fr)

*of Medicine* de Cardiff [2] : en avril 2005, elle compte 52 165 entrées dans 2 034 gènes (<http://www.hgmd.org>). L'analyse des données fournies par cette base, bien que non exhaustive, permet d'observer que les substitutions nucléotidiques et les microdélétions constituent 85 % des mutations responsables de maladies héréditaires chez l'homme (Figure 1) ; quant aux autres types de mutations, il s'agit d'insertions, duplications, inversions, mutations dynamiques, grands remaniements ou réarrangements complexes. Certains des mécanismes à l'origine de l'apparition de ces mutations sont aujourd'hui élucidés. Ces anomalies moléculaires peuvent résider non seulement dans les régions codantes des gènes, mais également dans les régions non codantes, au niveau des promoteurs, au sein des introns, voire à distance du gène, et leurs conséquences sont variées.

### Nature des mutations

Il est possible de distinguer 3 grandes classes de mutations : les substitutions nucléotidiques, les insertions/délétions de quelques nucléotides et les remaniements géniques de grande taille. Des mécanismes mutationnels plus rares seront également présentés.

### Nomenclature des principales mutations délétères

Afin d'apporter davantage de précisions sur la nature et les conséquences des variations de séquences nucléotidiques ou protéiques impliquées dans les maladies

La liste des mutations responsables de maladies héréditaires chez l'homme ne cesse de croître. Deux principales bases de données regroupent ces nombreuses mutations : la première correspond à la version actuellement en ligne du catalogue des maladies héréditaires chez l'homme, OMIM (*Online mendelian inheritance in man*), créé dans les années 60 par Victor McKusick et ses collaborateurs à l'Université John Hopkins de Baltimore [1]. En juin 2005, un grand nombre de mutations associées à des maladies génétiques sont répertoriées au sein de 1 755 gènes sur un total de 10 202 actuellement identifiés (<http://www.ncbi.nlm.nih.gov/omim>). La seconde base de données, HGMD (*Human gene mutation database*), a été créée par David Cooper et ses collaborateurs à l'Université *Wales College*

Article reçu le 25 avril 2005, accepté le 17 juin 2005.

génétiques, la nomenclature a récemment été modifiée.

Les principaux changements concernent le remplacement du code à une lettre des acides aminés par un code à trois lettres, ainsi que l'obligation de préciser la nature de la séquence modifiée par une lettre minuscule : g. pour une séquence d'ADN génomique, c. pour une séquence d'ADN complémentaire (la numérotation débutant (+ 1) sur le codon d'initiation de la traduction), p. pour une séquence protéique, m. pour une séquence d'ADN mitochondrial et r. pour une séquence d'ARN. Il est également indispensable de numéroter la mutation en se référant à une séquence extraite des bases de données, en précisant son numéro d'accès et sa version (exemple : NM\_004006.1) (*Encadré 1*).

### Remplacements d'une base : substitutions nucléotidiques

Elles constituent près de 70 % des mutations : on distingue les transitions (remplacement d'une base pyrimidique (C ou T) ou purique (A ou G) par une autre base de même nature) et les transversions (remplacement d'une base purique par une base pyrimidique, ou inversement). Les substitutions nucléotidiques peuvent être induites par des agents environnementaux mutagènes (substances chimiques, rayonnements...) ou par le métabolisme endogène, mais également être la conséquence d'erreurs spontanées intervenant lors de la réplication de l'ADN, et non détectées par les systèmes de réparation.

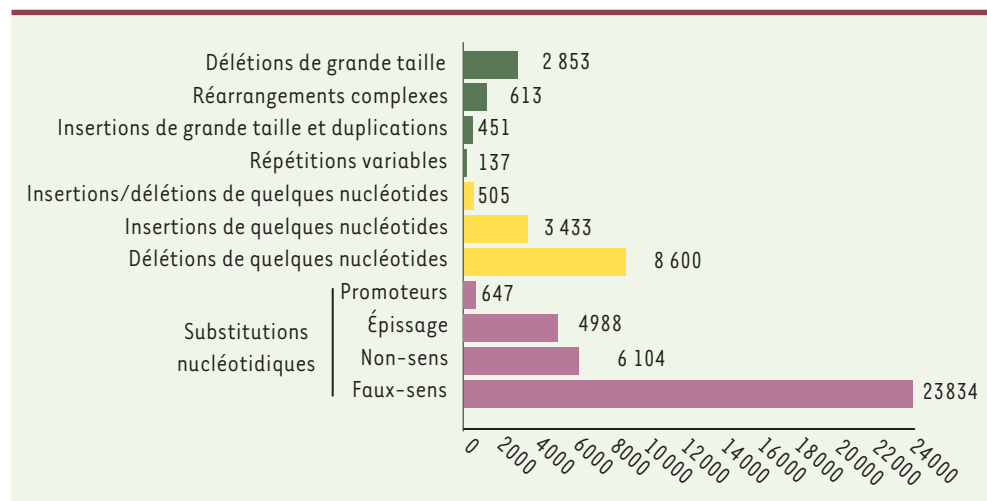
Les transitions sont en moyenne deux fois plus fréquentes que les transversions, et celles qui concernent le dinucléotide CpG représentent à elles seules plus de 20 % des substitutions nucléotidiques responsables de maladies génétiques [3]. La méthylation de l'ADN intervient préférentiellement en position 5 des cytosines précédant les guanines (dinucléotides CpG). Or les cytosines et les cytosines méthylées subissent avec une certaine fréquence des événements de désamination qui les transforment respectivement en uraciles et en thymines. À la différence du mésappariement U:G, très efficacement réparé par un complexe enzymatique spécifique (*uracile DNA glycosylase*), la réparation du

mésappariement T:G est beaucoup plus problématique, puisque la thymine est un composant normal de l'ADN. Par conséquent, en raison du caractère palindromique du dinucléotide CpG et de sa méthylation, la réparation des mésappariements T:G peut conduire, malgré l'existence de thymine glycosylases spécifiques, à des transitions CpG vers TpG ou CpA. Les dimères CpG représentent donc des points chauds de mutation chez l'homme, même s'il existe de grandes variations locales dans le taux de mutation de chaque dimère CpG.

### Délétions ou insertions de quelques nucléotides

Les délétions ou insertions d'un ou plusieurs nucléotides (moins de 20) sont, après les substitutions nucléotidiques, les anomalies de séquences nucléotidiques les plus fréquentes, représentant près de 25 % des anomalies répertoriées dans la base de données HGMD. Des délétions ou insertions non multiples de trois bases entraînent au niveau des séquences codantes un décalage du cadre de lecture (*frame shift*) qui aboutit à l'apparition d'un codon Stop prématuré et à l'éventuelle présence d'une protéine incomplète, la plupart du temps non fonctionnelle. Dans les cas de délétions ou insertions de trois bases ou d'un multiple de trois bases, la protéine codée présentera une délétion ou une insertion d'un ou plusieurs acides aminés, avec des conséquences fonctionnelles variables.

Ces anomalies moléculaires surviennent souvent au niveau de courtes répétitions en tandem, très probablement par un mécanisme de glissement (*slippage*) de l'ADN polymérase en raison de l'appariement décalé de séquences répétées lors de la réplication de l'ADN. Selon la façon dont le mésappariement est résolu, ce dérapage peut être à l'origine d'insertions ou de délétions d'un ou plusieurs motifs répétés.



**Figure 1. Répartition des différents types de mutations identifiées dans des gènes humains à l'origine de maladies génétiques.** Données extraites de la base de données HGMD le 20 avril 2005 (<http://www.hgmd.org>).

### Mutations dynamiques

La grande majorité des maladies liées à des mutations dynamiques impliquent des répétitions de triplets [4]. Il est possible de distinguer les répétitions qui affectent la région codante des gènes, où l'amplification de codons CAG (glutamine) reste modérée, et celles qui affectent les régions non codantes des gènes, où la taille de l'élément amplifié peut atteindre



## 1. NOUVELLE NOMENCLATURE DES MUTATIONS DÉLÉTÈRES : QUELQUES EXEMPLES

D'après les *Recommandations pour la description des variations de séquence*, dont la dernière modification, en date du mois de mai 2005, est accessible sur le site de l'HGV (*human genome variation society*, [www.hgvs.org](http://www.hgvs.org)).

### Substitution nucléotidique exonique et intronique

M82814 : c.4868G>A : remplacement de la guanine par une adénine en position 4 868 de la séquence codante (cds) du gène *NF1*.

M82814 : c.5750+332A>C : remplacement de l'adénine par une cytosine au niveau de la 332<sup>e</sup> base de l'intron qui suit immédiatement la position 5 750 qui correspond à la dernière base de l'exon (intron 30 du gène *NF1*).

M82814 : c.2991-1G>T : remplacement de la guanine par une thymine au niveau de la 1<sup>re</sup> base de l'intron précédant immédiatement la position 2 991 qui correspond à la première base de l'exon (intron 17 du gène *NF1*).

### Délétion/insertion nucléotidique

Les délétions sont identifiées par « del », précédé de la position du premier et du dernier nucléotide délétés et suivi de la nature des bases délétées. Les insertions sont identifiées par « ins », précédé de la position des nucléotides encadrant les bases insérées. Il est possible ensuite d'indiquer la nature ou uniquement le nombre de bases insérées :

M82814 : c.6700\_6715delATTAGCAAACGAGTGT : délétion des nucléotides 6 700 à 6 715 de la séquence codante (cds) du gène *NF1*.

M82814 : c.2033\_2034insT : insertion d'une thymine entre les nucléotides 2 033 et 2 034 de la séquence codante (cds) du gène *NF1*.

### Cas particuliers

Les duplications d'un ou plusieurs nucléotides sont identifiées par « dup », précédé de la position du (des) nucléotide(s) dupliqué(s) :

M82814 : c.7712\_7718dupATTACG : duplication de la séquence ATTACG comprise entre les nucléotides 7 712 et 7 718 de la séquence codante (cds) du gène *NF1*.

### Insertions/délétions

M82814 : c.2677\_2681delAAATinsC : délétion des nucléotides 2 677 à 2 681 du gène *NF1* et insertion d'une cytosine entre les nucléotides 2 676 et 2 677 de la séquence codante (cds) du gène *NF1*.

### Conséquences au niveau protéique

AAA59924 : p.Arg2237X : mutation non-sens remplaçant le codon arginine par un codon stop en position 2 237 de la neurofibromine.

AAA59924 : p.Leu300Pro : mutation faux-sens remplaçant l'acide aminé leucine par un acide aminé proline en position 300 de la neurofibromine.

AAA59924 : p.Gln514fs : décalage du cadre de lecture à partir de l'acide aminé glutamine en position 514 de la neurofibromine.

10 kb. Il existe quelques cas où le motif nucléotidique répété est supérieur à trois paires de bases. L'épilepsie myoclonique autosomique récessive d'Unverricht-Lundborg est associée à une amplification d'un motif de 12 paires de bases (CCCCGCCCGCG) en amont du site d'initiation de la transcription du gène *CSTB* [5]. L'ataxie spinocérébelleuse de type 10 est due à une amplification du motif (ATTCT) localisé dans l'intron 9 du gène *E46L* [6]. Enfin, dans le cas de la dystrophie myotonique de type 2, le motif nucléotidique répété correspond à une séquence de 4 paires de bases (CCTG) dans le premier intron du gène *ZNF9* [7].

Les maladies dues à des mutations dynamiques sont nombreuses et pour la plupart des maladies neurodégénératives. Elles sont caractérisées par le phénomène d'anticipation : au cours des générations successives, l'âge d'apparition de la maladie est de plus en plus précoce et s'accompagne d'une accentuation de la gravité, en raison de l'amplification du nombre de motifs répétés au cours des générations successives. Il est généralement admis qu'au-delà d'un certain seuil de longueur (environ 50 répétitions), les répétitions formeraient des structures anormales, de type épingles à cheveux ou triple hélice, perturbant la réplication et favorisant les dérapages ou glissements réplicatifs. La deuxième hypothèse, non exclusive, pour expliquer ce phénomène d'instabilité des trinucleotides implique le mécanisme de réparation des mésappariements (MMR) produits lors de la réplication.

### Expansions de type poly-alanine

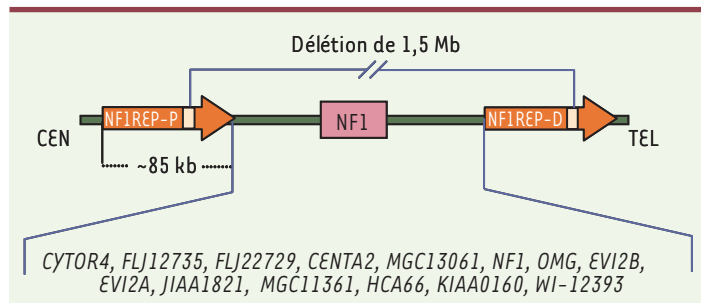
Des répétitions d'alanine sont observées de manière naturelle au sein de différents facteurs de transcription et autres protéines de localisation nucléaire ; toutefois, la longueur maximale de 20 alanines successives n'est jamais dépassée, quelles que soient les espèces, signant l'existence d'une contrainte au cours de l'évolution. Au cours de ces dernières années, des expansions de trinucleotides codant l'alanine ont été décrites au sein de plusieurs gènes codant des facteurs de transcription (*FOXL2*, *ZIC2*, *PHOX2B*...), à l'origine de pathologies du développement, ainsi qu'au sein de *PABPN1*, un gène codant une protéine nucléaire de liaison à la queue polyA, impliquée dans la maturation des transcrits [8]. Contrairement aux répétitions de glutamine, les expansions de type polyalanine sont le plus souvent codées par des répétitions imparfaites de triplets (code génétique dégénéré) et sont stables en méiose comme en mitose, présentant un faible degré de polymorphisme. Le « dérapage réplicatif » vraisemblablement à l'origine des expansions de glutamine ne permet donc pas d'expliquer ces expansions d'alanines, qui seraient en fait dues à des recombinaisons inégales entre chromosomes homologues lors de la méiose [9]. Les conséquences de ces expansions semblent être liées à des anomalies de repliement, ainsi qu'à l'agrégation et la dégradation des protéines concernées.

### Réarrangements génomiques : délétions, duplications, inversions et translocations

Les réarrangements génomiques résultent en général, mais non obligatoirement, d'événements de recombinaison homologue non allélique (NAHR, *non allelic homologous recombination*) intra- ou extrachromosomique, impliquant des séquences très semblables mais non néces-

sairement identiques, comme les séquences répétées dispersées Alu. Récemment, l'étude de réarrangements récurrents et la connaissance de la séquence complète du génome humain ont permis d'identifier une catégorie particulière de séquence répétée, appelée duplication segmentaire ou LCR (*low-copy repeats*). Ces séquences, qui représentent environ 5 % du génome humain, correspondent à des doublements de 10 à 500 kb survenus généralement il y a moins de 25 millions d'années [10,11]. Les deux copies sont quasiment identiques (plus de 95 % de similitude), mais non alléliques, et sont situées soit sur des chromosomes distincts, en particulier dans les régions péricentromériques ou subtélomériques, soit sur un même chromosome, à distance variable (faible dans le cas des gènes alpha de globine ou de l'intron 22 du facteur VIII de la coagulation, très importante (> 1 Mb) dans le syndrome Williams). Ces duplications constituent des structures hautement recombinogènes, favorisant les réarrangements génomiques par recombinaison inégale au moment de la méiose, avec comme conséquences des délétions, inversions, duplications ou translocations. Ces réarrangements sont observés de façon récurrente dans certaines maladies monogéniques lorsqu'un seul gène est impliqué dans le réarrangement :  $\alpha$  thalassémies, hémophilie A sévère, neurofibromatose de type 1 (Figure 2), maladie de Charcot-Marie-Tooth type 1A/neuropathie tomaculaire...). Dans le cas des syndromes dits « des gènes contigus », plusieurs gènes sont impliqués dans le réarrangement : syndromes de Di-George, de Smith-Magenis, de Rubinstein-Taybi...

D'autres remaniements géniques, non récurrents, sont la conséquence d'événements de recombinaison non homologue ou illégitime entre séquences qui ne présentent pas, ou très peu, d'homologie de séquence. C'est le cas de la plupart des délétions du gène codant pour la dystrophine. Si les mécanismes moléculaires de la recombinaison illégitime sont mal connus, un certain nombre de motifs pourraient constituer des points chauds de recombinaison : séquences alternées de purines-pyrimidines, régions MAR (*matrix-associated regions*) riches en nucléotides A/T, sites de clivage des topo-isomérases ou, encore, séquences palindromiques [12-13].



**Figure 2.** Séquences répétées encadrant le locus *NF1* et impliquées dans la délétion de 1,5 Mb retrouvée chez environ 7 % des individus atteints de neurofibromatose de type 1. Une quinzaine de gènes représentés ici sont délétés de façon concomitante avec le gène *NF1*. Les rectangles blancs au sein des séquences REP (séquences répétées) symbolisent les points chauds de cassure.

## Mécanismes rares : insertions d'éléments transposables, conversions géniques

### Insertion d'éléments transposables

Les éléments transposables représentent environ la moitié du génome humain et sont divisés en deux types : les transposons (environ 3 % du génome humain), qui se mobilisent sous forme d'ADN selon un mécanisme de type « couper/coller » gouverné par une transposase, et les rétrotransposons (plus de 40 % du génome humain), qui utilisent un intermédiaire ARN lors de leur mobilisation selon un mécanisme de type copier/coller gouverné en particulier par une transcriptase inverse. Le déplacement des éléments transposables d'un point à un autre du génome peut engendrer des mutations soit directement par leur insertion à côté ou à l'intérieur d'un gène, ce qui peut en altérer la séquence codante et/ou l'expression, soit indirectement par la recombinaison homologue entre deux copies d'éléments transposables de séquences similaires et situés à des positions non orthologues, ce qui peut amener à des délétions ou des duplications (voir ci-dessus).

Chez l'homme, les insertions responsables de maladies génétiques concernent exclusivement les rétrotransposons sans LTR (*long terminal repeat*) autonomes (LINE) et non autonomes (SINE) [10]. Depuis les premières descriptions par les groupes de H.H. Kazazian en 1988 [14] et de F. Collins en 1990 [15], une cinquantaine d'insertions *de novo* de séquences LINE-1 (seul élément actif au sein des LINE) et de séquences Alu (SINE) ont été rapportées. Parmi les 500 000 séquences LINE-1 et le million de séquences Alu du génome humain, seule une faible proportion est encore active (moins de 100 séquences LINE-1 et quelques milliers de séquences Alu), et la fréquence des événements de transposition a été estimée chez l'homme à 1 événement pour 50 à 200 cellules germinales [16]. Dans tous les cas, les insertions d'éléments transposables sont des événements rares, dont l'identification a néanmoins permis d'enrichir les connaissances des rétrotransposons chez l'homme. Enfin, il est à noter qu'aucun cas d'insertion *de novo* de rétrotransposons à LTR ou HERV (*human endogenous retrovirus*) n'a été rapporté à ce jour.

### Conversion génique

La conversion génique est un mécanisme de recombinaison qui, contrairement au *crossing-over*, correspond à un transfert unidirectionnel d'information génétique [17]. Elle conduit au remplacement d'une séquence d'ADN par une autre, apparentée non allélique (conversion interlocus) ou allélique (conversion interallélique). Ce remplacement peut entraîner dans le gène receveur une série de changements nucléotidiques répartis sur une région assez courte (Figure 3). Dans certains cas,

la séquence donneuse est un pseudogène, inactivé par l'accumulation de mutations, et le transfert d'une de ces mutations inactive le gène receveur : c'est le cas de certaines mutations du gène de la stéroïde 21 hydroxylase dans l'hyperplasie surrénalienne, ou du gène de la  $\beta$  glucocérébrosidase dans la maladie de Gaucher.

### Mécanismes épigénétiques des mutations

Dans certaines circonstances, l'expression phénotypique de la maladie dépend du sexe du parent qui transmet l'anomalie moléculaire. Cette différence traduit le phénomène d'empreinte génomique parentale, une modification épigénétique, temporaire et réversible du génome nucléaire. Cette empreinte apposée durant la gamétogenèse conduit à l'expression différentielle du matériel génétique en fonction de son origine parentale. Chez l'homme, l'empreinte génomique a été impliquée dans un nombre croissant d'anomalies du développement, de maladies héréditaires et de cancers, une soixantaine de gènes soumis à empreinte étant identifiés à ce jour. Les pathologies liées à l'empreinte sont dues à une modification, pour un locus ou un gène donné, de l'équilibre entre allèle paternel et allèle maternel. Ce déséquilibre peut résulter d'une disomie uniparentale, c'est-à-dire de la présence dans une cellule diploïde de 2 séquences homologues héritées d'un seul parent, voire de 2 chromosomes entiers. Chez

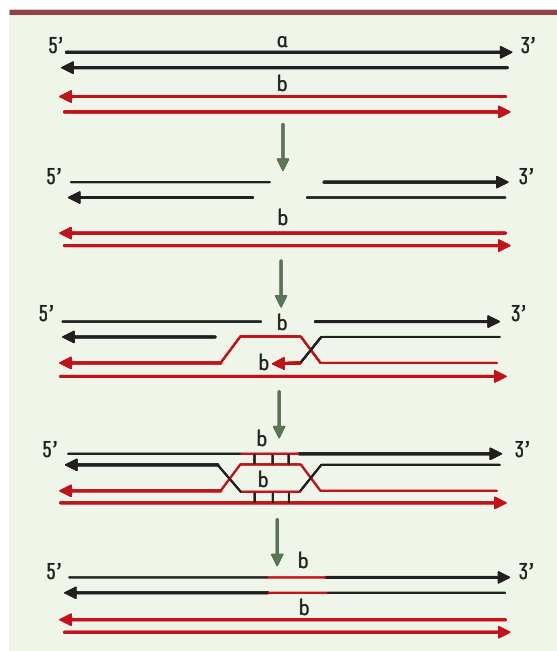
l'homme, la première observation de ce type date de 1988 : chez un enfant à caryotype normal, la présence en double copie d'un chromosome 7 d'origine maternelle porteur d'une mutation p. Phe508del du gène *CFTR* était associée à une mucoviscidose et une petite taille, suggérant l'existence au niveau du chromosome 7 d'un gène impliqué dans la croissance et soumis à empreinte, le gène maternel n'étant pas exprimé [18]. Cet exemple illustre les deux types d'effets délétères de la disomie uniparentale : transmission d'une maladie récessive par homozygotie (mucoviscidose) et absence d'expression d'un gène par empreinte parentale (petite taille).

L'un des meilleurs exemples est aujourd'hui fourni par les syndromes de Prader-Willi et d'Angelman, liés à une disomie maternelle et paternelle du chromosome 15, respectivement. Un déséquilibre de l'empreinte peut également être observé malgré une transmission biparentale, de petites délétions emportant, dans certains cas, les « centres de l'empreinte », éléments intervenant en *cis* dans la régulation de l'empreinte. Enfin, il a également été rapporté chez certaines femmes des cas de môles hydatiformes récurrents d'origine biparentale, conséquence de défauts globaux d'empreinte de la lignée germinale femelle, les allèles maternels des gènes présentant alors un profil épigénétique paternel. Cette impossibilité d'imprimer l'empreinte maternelle, phénomène baptisé *immaculate misconception* par Surani [19], doit donc être le fait de mutations dans un gène essentiel à l'empreinte maternelle, encore inconnu, mais localisé par analyse de liaison au sein d'une région de 1,1Mb sur le chromosome 19q13.4 [20].

Il est aujourd'hui clairement établi que la méthylation de l'ADN ainsi que les modifications post-traductionnelles des histones participent à l'établissement de l'expression mono-allélique des gènes soumis à empreinte. Cependant, la nature précise de la marque primaire de l'empreinte et son devenir pendant le développement restent en partie mystérieux ; leur identification devrait permettre d'envisager une meilleure compréhension des maladies liées à des modifications épigénétiques.

### Mutations *de novo*, mosaïques germinales et mosaïques somatiques

Un petit nombre de mutations dites *de novo*, ou néomutations, surviennent dans la lignée germinale, au cours des divisions mitotiques durant la spermatogenèse ou l'ovogenèse, ou pendant la méiose elle-même. Pour certaines maladies, la fréquence d'apparition de mutations *de novo* peut être très importante (plus de 30 % des cas pour la myopathie de Duchenne ou l'hémophilie A, 50 % des cas pour la neurofibromatose de type I, 90 % des cas pour l'achondroplasie). Le nombre de divisions cellulaires intervenant au cours de la formation des gamètes est très différent chez l'homme et la femme : au cours de l'ovogenèse, chaque gamète est le résultat de 22 divisions mitotiques et d'une division méiotique, survenant au cours de la vie fœtale et s'arrêtant au moment de la naissance. À l'opposé, la spermatogenèse se déroule de façon continue de la puberté à la fin de la vie, chaque spermatozoïde résultant donc, en moyenne, de 200 divisions, ce nombre augmentant avec l'âge. En toute logique, on s'attend à ce que



**Figure 3. Conversion génique entre deux chromosomes homologues.** Les séquences des brins du duplex accepteur ont été modifiées au profit des séquences donneuses.

la fréquence des néomutations paternelles augmente avec l'âge et soit supérieure à celle des néomutations maternelles : ce phénomène est effectivement observé, exacerbé encore par le fait que la méthylation des cytosines des dimères CpG est globalement plus importante dans les cellules germinales mâles que dans les cellules germinales femelles, favorisant ainsi les transversions de type C>T ou G>A. Il a ainsi été montré que les mutations de *FGFR2* (récepteur du facteur de croissance fibroblastique de type 2), responsables du syndrome d'Apert, sont systématiquement d'origine paternelle et associées à un âge paternel élevé. De récents travaux ont révélé que l'incidence élevée de la mutation majoritaire c.755C>G/p.Ser252Trp est très vraisemblablement la conséquence d'un avantage prolifératif conféré aux spermatogonies à travers la voie du FGF [21]. Ce concept de sélection positive médiée par la protéine mutée, familier dans les domaines de la cancérogenèse, de l'immunité ou des micro-organismes, semble ainsi pouvoir s'appliquer également aux mutations germinales survenant chez les vertébrés, et pourrait être à l'origine d'autres pathologies génétiques d'origine paternelle comme l'achondroplasie (mutations du gène *FGFR3*) ou la néoplasie endocrinienne multiple de type 2 (mutations du gène *RET*). Ce mécanisme illustre aussi un exemple de conflit d'évolution, où une mutation délétère pour l'organisme peut se révéler avantageuse dans le contexte cellulaire des testicules.

Une maladie due à une mutation *de novo* ne récidive habituellement pas dans la fratrie. Cependant, il a été décrit le cas de parents de phénotype normal ayant plus d'un enfant atteint. Pour expliquer ces observations, il faut supposer que l'un des parents est porteur d'une mosaïque germinale, et qu'il existe en fait un clone de cellules germinales porteur de la mutation. Les mosaïques somatiques sont quant à elles dues à des mutations survenues après la fécondation, à un stade plus ou moins tardif de l'embryogenèse. Leurs conséquences cliniques dépendent de

la nature de la mutation, du gène altéré et du tissu concerné. Elles constituent très probablement l'une des causes de l'hétérogénéité d'expression clinique des maladies héréditaires, et représentent un piège diagnostique à la fois clinique et moléculaire, l'anomalie nucléotidique ne concernant qu'une proportion limitée des cellules analysées.

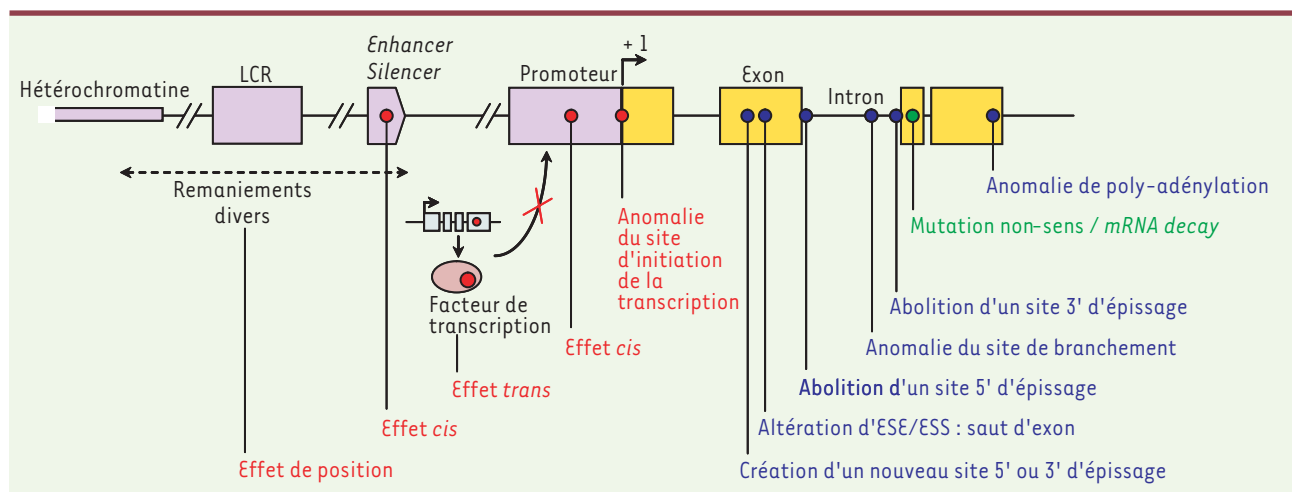
### Conséquences des mutations délétères sur l'expression du gène

En raison de leur forte contribution au spectre mutationnel, les anomalies d'expression génique prennent une place sans cesse grandissante en clinique, tant du point de vue de leur fréquence que de leur variété. Les différents types de mutations rencontrés sont récapitulés dans la Figure 4.

#### Mutations affectant la régulation de la transcription

##### Anomalies de régulation en cis

Une mutation délétère localisée dans le promoteur d'un gène peut, par plusieurs mécanismes, affecter l'expression de ce gène. Les méthodes mises en œuvre pour explorer précisément les conséquences de ces anomalies moléculaires reposent sur des expériences d'expression *in vitro* de gènes rapporteurs. Tout type de mutation peut être rencontré (mutations ponctuelles, délétions ou insertions de plus ou moins grande taille, inversions...) et de nombreux exemples sont décrits en pathologie humaine. La mutation peut altérer un site consensus de liaison d'un ou



**Figure 4. Localisation des différents types de mutations pouvant affecter l'expression d'un gène.** Les exons du gène sont représentés par des rectangles jaunes, les introns positionnés entre les exons. Les séquences impliquées dans la régulation de l'expression du gène sont représentées par des rectangles mauves. Les différents mécanismes moléculaires à l'origine d'une altération de la transcription du gène, de sa maturation ou de sa stabilité figurent en rouge, bleu et vert, respectivement. LCR : locus control region ; ESE : exonic splicing enhancers ; ESS : exonic splicing silencers.



plusieurs facteurs de transcription, et diminuer ainsi, voire abolir l'expression génique. La mutation du promoteur peut également altérer sa structure, entraînant la modification négative ou positive de l'expression du gène par l'intermédiaire de structures secondaires.

En dehors des régions promotrices, les mutations peuvent altérer d'autres éléments agissant en *cis*, pouvant se situer à plusieurs centaines, voire plusieurs milliers de paires de bases en 5' ou 3' du gène. Ces séquences régulatrices de courte taille, *enhancers* ou *silencers*, sont capables de fixer des facteurs de transcription, souvent tissu-spécifiques, et entraînent respectivement l'activation ou l'inhibition de la transcription en participant à l'assemblage de la machinerie transcriptionnelle.

Lorsque l'unité transcriptionnelle et le promoteur d'un gène sont intacts, son expression peut également être altérée par un « effet de position », modifiant son environnement chromatinien habituel. Cet effet de position peut être la conséquence de remaniements tels que des translocations, délétions, insertions ou inversions. Le réarrangement peut séparer le gène d'éléments régulateurs distants, de type *enhancer* ou *silencer*, ou bien, au contraire, placer à proximité d'un gène l'*enhancer* ou le *silencer* d'un autre gène.

Un effet de position s'observe également lors de la délétion de LCR (*locus control region*). Présents en différents endroits du génome, ces éléments de grande taille peuvent exercer leur effet, à de très grandes distances, sur l'expression de *clusters* entiers de gènes. Ils semblent définir des domaines dans lesquels la structure chromatinienne permet aux facteurs de transcription d'accéder à leurs sites de fixation sur l'ADN, entraînant l'activation de la transcription. À titre d'exemple, la délétion du LCR placé à plus de 60 kb en amont du *cluster*  $\beta$ -globine est une des causes de  $\beta$ -thalassémie.

Enfin, un cas particulier d'effet de position, dénommé variéation, résulte de la juxtaposition d'un gène avec une région d'hétérochromatine (ADN très condensé, riche en séquences répétées et hautement méthylé), la conséquence directe étant une extinction de l'expression du gène en question.

L'implication d'anomalies moléculaires à l'origine d'effets de position en pathologie humaine est probablement sous-estimée à ce jour, faute de critères simples permettant de déterminer leur caractère causal.

#### **Inhibition de l'expression d'un gène par un ARN antisens**

Une étude récente décrit un mécanisme moléculaire original à l'origine d'une  $\alpha$ -thalassémie liée à une délétion de 18 kb sur le chromosome 16, au sein du *cluster* HBA1, HBQ1 et HBA2, codant les sous-unités  $\alpha 1$ ,  $\theta$  et  $\alpha 2$  de la globine [22]. La conséquence de cette délétion est non

seulement la délétion des gènes HBA1 et HBQ1, mais aussi la juxtaposition de HBA2 en antisens en aval de LUC7L, un gène exprimé de manière importante et ubiquitaire. Un transcrite de fusion LUC7L (sens)-HBA2 (antisens) est alors produit, sous le contrôle du promoteur de LUC7L. Ce transcrite antisens pour HBA2 est à l'origine d'une extinction de l'expression du gène HBA2 et d'une modification épigénétique de son expression par méthylation de l'îlot CpG localisé dans le promoteur de HBA2.

#### **Anomalies de régulation en trans**

Des mutations de gènes codant des facteurs de transcription peuvent également être responsables d'anomalies de la transcription. La plupart de ces mutations sont létales, ou associées à des anomalies du développement. Un exemple intéressant concerne les mutations du gène *MECP2* impliquées dans le syndrome de Rett, un grave désordre neurologique lié à l'X. La protéine MecP2 (*methyl-CpG-binding protein*) se lie spécifiquement aux îlots CpG méthylés des promoteurs de nombreux gènes, et agit comme un répresseur transcriptionnel. Les mutations de *MECP2* conduisent ainsi à l'abolition de cette régulation négative [23].

#### **Situations complexes**

La dystrophie myotonique de Steinert (DM1) est une maladie neuromusculaire multisystémique caractérisée par une myotonie et une détérioration des fonctions neuromusculaires. L'anomalie moléculaire en cause dans cette maladie a été identifiée : il s'agit de l'amplification d'un trinucléotide CTG dans la région 3' non codante du gène *DMPK*. Différentes études ont été menées pour comprendre les répercussions de cette amplification de triplets au niveau de la physiopathologie de la maladie [24] : la répétition CTG pourrait agir en *cis* et altérer l'expression du gène *DMPK* lui-même, mais également d'autres gènes présents dans la région de *DMPK*, tels *SIX5* et *DMWD* (Figure 5). L'étude de la transcription de *DMPK* montre que le taux de transcrite primaire est comparable dans les cellules musculaires de sujets DM1 et les cellules musculaires normales, mais que le taux d'ARNm mature est diminué. Par ailleurs, la transcription de *SIX5* est clairement diminuée chez les patients. Ainsi, la perte de fonction de plusieurs gènes permet d'envisager DM1 comme un syndrome de « gènes contigus ».

Cependant, l'effet *cis* de l'amplification de CTG n'explique pas à lui seul tous les symptômes observés chez les patients. Les transcrits de *DMPK* porteurs de grandes amplifications CUG exercent également un effet *trans*-dominant. Ils sont séquestrés dans le noyau sous forme de focus. En présence de molécules d'ARN *DMPK* porteuses de grandes amplifications CUG, on observe une modification de la régulation de la répartition des CUG-BP, protéines d'épissage ayant la capacité de se fixer aux CUG. Si aucune colocalisation entre les protéines CUG et les foyers d'ARN *DMPK* mutés n'a encore pu être mise en évidence, on sait que les conséquences des altérations de la régulation des CUG-BP sont des anomalies de l'épissage de nombreux transcrits (*IR*, codant le récepteur de l'insuline, *CLCN1*, codant un canal chlore spécifique du muscle, *TNNT2*, codant la troponine T cardiaque et *MAPT*, codant la protéine tau associée aux microtubules au niveau du système nerveux central) : ces anomalies pourraient expliquer les atteintes multisystémiques rencontrées au cours de la maladie, ainsi que son caractère dominant. Une autre pro-

téine d'épissage capable de se lier aux CUG, appelée *Muscleblind*, voit également sa régulation modifiée et se colocalise avec les foyers d'ARN porteurs de grandes amplifications (Figure 5).

L'identification de l'amplification du motif CCGT au sein du gène *ZNF9*, anomalie moléculaire en cause dans la dystrophie myotonique de type II, a permis de mieux comprendre les bases moléculaires de la DM1, puisque l'on observe également une séquestration nucléaire en foci des transcrits *ZNF9* et une colocalisation de *Muscleblind* associées à des anomalies de l'épissage d'autres transcrits, suggérant fortement que les nombreux symptômes en commun dans les deux formes de la maladie DM1 et DM2 sont dus à l'effet *trans*-dominant du motif nucléotidique amplifié [7].

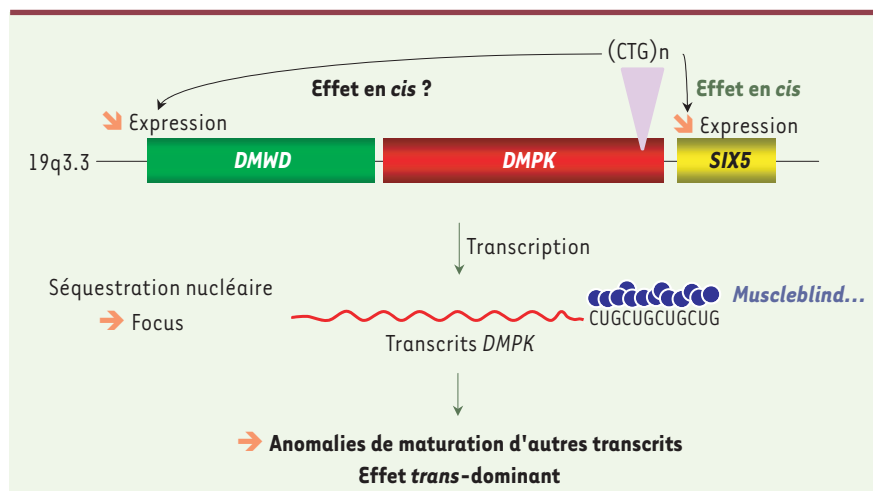
Un autre exemple de situation complexe concerne la dystrophie facio-scapulo-humérale (FSHD), une myopathie transmise sur un mode autosomique dominant dont l'incidence est de 1/20 000 naissances. Sur le plan moléculaire, elle est essentiellement associée à des contractions de la région subtélomérique D4Z4 du chromosome 4q [25], mais le mécanisme à l'origine de la maladie est encore aujourd'hui l'objet d'une recherche active soulevant plusieurs hypothèses [26]. Les répétitions D4Z4 sont situées dans une région d'hétérochromatine. La présence d'aucun gène n'ayant pu être incontestablement démontrée dans la région des répétitions ou à leur proximité immédiate, un effet de variéation a été évoqué; différents gènes, situés à des distances variables de D4Z4, ont été étudiés, dont l'expression pourrait subir cet effet de variéation: une augmentation de l'expression des gènes *FRG2*, *FRG1* et *ANTI* inversement proportionnelle au nombre de répétitions D4Z4 a ainsi été retrouvée dans le tissu musculaire de patients présentant une FSHD. Cette constatation a entraîné la découverte d'un complexe multiprotéique, capable

de se lier à une séquence spécifique contenue dans chaque répétition et impliqué dans la répression transcriptionnelle [27]. Cependant, plusieurs études indépendantes n'ont pu confirmer l'augmentation d'expression de ces 3 gènes, laissant la place à d'autres modèles [28, 29]. Parmi les hypothèses avancées, la formation de boucles internes à la région D4Z4 pourrait prévenir l'interaction de D4Z4 avec des gènes en *cis* à de grandes distances; la contraction observée dans la FSHD empêcherait la formation de ces boucles, entraînant une expression inappropriée de certains gènes [29]. Le modèle le plus récent, par ailleurs compatible avec le précédent, est celui d'une altération de l'organisation nucléaire, la région D4Z4 du chromosome 4 intervenant dans l'interaction avec la lamine nucléaire [30]. Cela pourrait notamment altérer le recrutement de facteurs de transcription et de facteurs modifiant la structure de la chromatine, et perturber ainsi l'expression d'un ou plusieurs gènes localisés ou non à proximité de la région D4Z4. Le développement de modèles cellulaires et animaux devrait sans doute fournir un outil important pour progresser dans la compréhension de cette maladie.

#### Altération de la maturation de l'ARN pré-messager en ARNm

Les mutations affectant les sites d'épissage représentent 10 % des mutations rapportées dans la base de données HGMD. Ce chiffre est cependant sous-estimé, car seules les mutations affectant les sites donneurs et

accepteurs d'épissage sont prises en compte. Une revue récente présente les différents mécanismes à l'origine d'anomalies de l'épissage [31]. Des substitutions nucléotidiques, des délétions, des insertions... peuvent abolir l'épissage d'un exon en modifiant les séquences consensus des sites donneurs et accepteurs d'épissage, mais également le site de branchement intronique. Des mutations hors de ces sites particuliers peuvent également dévoiler des sites cryptiques d'épissage, conduisant alors à l'épissage partiel d'exon. Récemment, la description de nouvelles séquences consensus exoniques, les ESE (*exonic splicing enhancers*) et les ESS (*exonic splicing silencers*), favorisant ou réprimant l'épissage de l'exon qui les contient, a permis de comprendre pourquoi des modifications de la séquence codante d'un gène peuvent être responsables de l'épissage complet de l'exon où elles sont positionnées [32].



**Figure 5. Effets en cis et en trans de l'amplification du triplet CTG dans la région 3' non codante du gène DMPK.** La répétition CTG agit probablement en *cis* en altérant non seulement l'expression du gène *DMPK*, mais aussi celle de *SIX5* et *DMWD*, présents dans la région de *DMPK*. Les transcrits de *DMPK* porteurs de grandes amplifications CUG exercent probablement aussi un effet *trans*-dominant. Ils sont séquestrés dans le noyau sous forme de foci; or la présence de transcrits porteurs de grandes amplifications CUG modifie la répartition des CUG-BP, protéines d'épissage ayant la capacité de se fixer aux CUG, avec pour conséquences générales des anomalies de l'épissage de nombreux transcrits. Par ailleurs, la régulation de *Muscleblind*, un autre type de protéine d'épissage capable de se lier aux CUG, et se colocalisant avec les foyers d'ARN porteurs de grandes amplifications, est également modifiée.



Une mutation du signal de poly-adénylation peut également perturber la maturation de l'extrémité 3' de l'ARN et être responsable d'une diminution de la stabilité du transcrit. Le premier exemple en pathologie humaine a été décrit en 1983 : il s'agit d'une mutation ponctuelle dans le signal de poly-adénylation du gène codant l' $\alpha$ 2-globine (allèle  $\alpha$ T-Saudi), ayant comme conséquence une diminution du niveau d'ARNm *HBA2* dans les cellules érythroïdes [33]. Depuis, plusieurs mutations de ce même motif ont été décrites, à l'origine d'autres cas d' $\alpha$ -thalassémies essentiellement observés dans les populations du bassin méditerranéen.

### Altération de la stabilité de l'ARNm : mRNA decay

La durée de vie des ARNm varie de quelques minutes à plusieurs heures, influençant donc la quantité de protéine synthétisée en aval. Deux mécanismes de dégradation des ARN messagers ont été identifiés : la dégradation constitutive NSD (*non stop decay*) et le NMD (*nonsense mediated mRNA decay*), aujourd'hui bien documenté [34, 35]. Le NMD est un processus permettant de dégrader sélectivement les ARNm portant un codon Stop prématuré, empêchant ainsi la traduction d'une protéine tronquée, potentiellement délétère par un effet dominant négatif. Le codon Stop prématuré peut être produit par substitution nucléotidique ou induit par décalage du cadre de lecture lors d'anomalies de l'épissage, et doit être positionné à plus de 50–55 pb en amont d'une jonction exon/exon pour que l'ARNm concerné soit dégradé. Le NMD met en jeu la fixation d'un complexe EJC (*exon junction complex*) composé d'une douzaine de protéines, qui se positionne sur 20–24 nt en amont d'une jonction exon/exon. Les ARNm ne seront donc pas dégradés par NMD lorsque la mutation non-sens est localisée dans le dernier exon du gène, étant donné l'absence de jonction exon/exon en aval. L'importance de ce phénomène en pathologie humaine a récemment été illustrée par l'étude de deux neurocristopathies liées à des mutations tronquantes du même facteur de transcription SOX10. En effet, lorsque la mutation survient dans le dernier exon du gène, l'ARNm échappe au mécanisme de NMD et conduit à la synthèse d'une protéine exerçant un effet dominant négatif, entraînant un phénotype sévère PCWH (*peripheral demyelinating neuropathy, central dysmyelinating leukodystrophy, Waardenburg syndrome and Hirschsprung disease*). À l'inverse, lorsque la mutation est située au sein d'exons internes, le NMD entraîne une réduction de l'ARNm synthétisé, prévenant l'effet dominant négatif de la protéine tronquée et conduisant à une neurocristopathie beaucoup moins sévère (WS4), qui combine uniquement syndrome de Waardenburg et maladie de Hirschsprung [36]. Le NMD pourrait ainsi contribuer à expliquer les différences phénotypiques existant entre plusieurs syndromes alléliques, c'est-à-dire associés à des mutations d'un même gène. Si le NMD

est un mécanisme probablement bénéfique pour la cellule, il entraîne un problème majeur lorsque le diagnostic moléculaire est réalisé sur le(s) transcrit(s) du gène, car l'anomalie moléculaire est absente *in vitro*.

### Conséquences des mutations sur la fonction de la protéine

Les mutations non-sens aboutissant à la formation de l'un des trois codons Stop UAA, UAG ou UGA, les mutations responsables d'un décalage du cadre de lecture (y compris les anomalies de l'épissage) et les mutations du codon d'initiation de la traduction entraînent généralement l'absence de formation d'une quelconque protéine, ou la formation d'une protéine tronquée dont l'activité est nulle ou très réduite. En revanche, les conséquences des mutations faux-sens ou des délétions/insertions en phase (y compris les anomalies de l'épissage) sont moins certaines. Ces mutations sont responsables d'un changement de la séquence protéique et peuvent notamment affecter la stabilité, l'adressage intracellulaire, la maturation de la protéine, son assemblage dans une structure multimérique, les sites importants pour l'activité enzymatique ou encore les interactions fonctionnelles avec des ligands et d'autres protéines. Mais le changement de la séquence de la protéine peut également être sans conséquence sur sa fonction. Ainsi, lorsqu'un changement nucléotidique de type faux-sens ou une délétion/insertion en phase surviennent au sein de la séquence codante d'un gène, un faisceau d'arguments est nécessaire pour déterminer les conséquences de cette anomalie dans la pathologie (*Encadré 2*).

Dans la majorité des cas, les stratégies d'identification des mutations à l'origine des maladies sont fondées sur l'étude de l'ADN génomique des sujets atteints, tandis que les conséquences des mutations au niveau des transcrits des gènes sont rarement explorées. Pourtant, des mutations identifiées *a priori* comme non-sens ou faux-sens par une approche ADN, de même que des délétions ou insertions, peuvent en réalité affecter la transcription, l'épissage ou encore la stabilité du transcrit du gène. Ces anomalies moléculaires sont probablement sous-estimées et ne peuvent être mises en évidence que par l'étude de l'ARN.

Dans le même ordre d'idées, certaines mutations faux-sens retentissent par des effets inattendus sur la maturation post-traductionnelle de la protéine. C'est le cas, notamment, des gains de glycosylation : l'exemple récent d'une mutation particulière du gène *IFNGR2* montre de manière univoque que la création d'un nouveau site de glycosylation est responsable du phénotype de susceptibilité mendélienne aux mycobactéries peu virulentes, et l'étude *in silico* des mutations pouvant altérer des sites de glycosylation au sein d'autres protéines suggère que ce type d'anomalie pourrait être responsable de nombreuses autres maladies génétiques [37].

### Perte de fonction

Dans cette situation, il faut distinguer les allèles amorphes ou nuls, responsables de la perte totale d'expression de la protéine (ex :  $\beta^0$  thalassémies) ou de la synthèse d'une protéine totalement inactive, et les allèles hypomorphes responsables de la perte partielle d'expression de la protéine (ex :  $\beta^+$  thalassémies) ou de la synthèse d'une protéine partiellement inactive. Les mutations par perte de fonction sont retrouvées dans la majorité des maladies récessives, qui nécessitent pour se manifester une

## 2. MUTATION OU POLYMORPHISME ?

Une des difficultés les plus importantes en génétique consiste à distinguer une mutation (variation d'ADN délétère) d'un polymorphisme. Plusieurs éléments doivent être analysés afin de mettre en évidence le caractère causal d'une mutation dans une pathologie.

- Il convient tout d'abord d'étudier si la mutation est située au sein d'une région essentielle à la fonction de la protéine, dans la mesure où ces régions sont connues. Dans le cas contraire, la localisation au sein d'une séquence très conservée orthologue et/ou paralogue est un argument important.
- La modification importante de la polarité et de la charge de l'acide aminé concerné, son encombrement stérique... peuvent également être des facteurs à prendre en compte.
- La vérification de l'absence de cette modification chez un grand nombre de sujets témoins est systématique, mais n'exclut pas la possibilité d'un polymorphisme rare.
- Enfin, l'apparition d'une mutation *de novo* chez un sujet atteint possédant des parents sains est un critère majeur, très en faveur de l'implication de cette mutation dans la pathologie.
- L'étude couplée du gène et de son transcrit peut permettre de mettre en évidence une altération de la transcription ou de l'épissage.
- Dans la mesure du possible, l'étude des conséquences de la mutation sur la fonction de la protéine devra être réalisée *in vitro*.
- L'analyse *in vivo* par l'étude de modèles animaux permettra enfin de signer définitivement le caractère causal de la mutation.

atteinte des deux allèles. Elles sont également retrouvées dans les maladies dominantes par haplo-insuffisance, comme l'hypercholestérolémie familiale dont la forme hétérozygote est beaucoup moins sévère que la forme homozygote, la perte de 50 % de l'activité du gène étant suffisante pour entraîner un phénotype clinique [38].

Une très grande hétérogénéité de mutations est quasiment la règle (hors effet fondateur) dans les maladies par perte de fonction. Les cas extrêmes d'hétérogénéité des mutations sont illustrés par la plupart des maladies sévères liées à l'X, ou par les maladies autosomiques dominantes affectant l'efficacité reproductrice des patients. Dans ces cas, les mutations sont perdues en quelques générations et on ne retrouve que des mutations « privées », spécifiques de chaque famille. Le spectre de ces mutations reflète alors directement la sensibilité du gène à différents mécanismes mutationnels.

La fréquence des événements mutationnels entraînant une perte de fonction dépend en partie de la taille de la cible (taille de la séquence codante, nombre d'exons, nombre d'acides aminés importants pour la fonction de la protéine) et de l'existence de points (ou régions) chauds de mutation. Dans certains cas, la perte totale de fonction est létale, et seules sont admises les pertes partielles : c'est notamment le cas pour le déficit en G6PD (glucose 6-phosphate déshydrogénase), où il n'existe que des mutations faux-sens hypomorphes, maintenant une activité enzymatique résiduelle [39].

### Effet dominant négatif

Cette situation correspond à la frontière entre les mutations entraînant une perte de fonction et celles entraînant un gain de fonction. On parle d'effet dominant négatif lorsque la protéine codée par le gène muté, dénommé allèle antimorphe, non seulement perd sa fonction, mais interfère aussi avec la fonction de l'allèle normal chez les hétérozygotes. De telles mutations caractérisent les gènes codant les protéines de structure, ou capables de former des homo- ou des hétérodimères : ces mutations entraînent des modifications conformationnelles qui affectent la fonction de la protéine normale. C'est le cas de certaines mutations responsables de l'ostéogénèse imparfaite, qui touchent les gènes codant les chaînes  $\alpha 1$  (*COL1A1*) et  $\alpha 2$  (*COL1A2*) du collagène de type I : une mutation affectant la structure

d'une chaîne de collagène (délétion d'un ou de plusieurs exons) exerce à l'état hétérozygote un effet plus délétère (dominant) qu'une perte totale d'expression (mutations récessives), car elle empêche une association correcte de la sous-unité raccourcie aux chaînes normales [40].

### Gain de fonction

Dans cette situation, plus rare, il faut distinguer les allèles hypermorphes des allèles néomorphes. On qualifie d'hypermorphe un allèle qui correspond à une surexpression du gène sauvage ou qui code une forme hyperactive de son produit. Il est exceptionnel que les maladies héréditaires par gain de fonction soit liées à une surexpression : un exemple rare est celui de l'augmentation de la quantité de protéine liée à une duplication génique associée à la neuropathie de Charcot Marie Tooth de type 1A, où la surexpression de *PMP22* provoque une myélinogenèse anormale.

La synthèse d'une protéine hyperactive est observée avec différentes mutations du récepteur FGFR3, à l'origine de trois chondrodysplasies : l'hypochondroplasie, l'achondroplasie, et le nanisme thanatophore (du phénotype le moins sévère au plus sévère). La fonction normale de FGFR3 est de réguler l'ossification, avec comme conséquence un ralentissement de la croissance. L'expression des différentes mutations chez le xénope démontre que les mutants associés à ces chondrodysplasies présentent une augmentation de leur activité tyrosine kinase par rapport à la protéine sauvage, indépendamment du ligand, entraînant une élévation du niveau global de phosphotyrosines [41].

La dénomination d'allèle néomorphe est attribuée à un allèle codant pour une protéine dont la fonction est différente de celle du produit sauvage. Ce phénomène est assez fréquemment observé dans les mutations somatiques associées aux cancers, mais beaucoup plus rarement dans les maladies



héréditaires. L'exemple classique est le cas de la mutation faux-sens p. Met358Arg du site actif de l' $\alpha$ 1-antitrypsine ( $\alpha$ 1-AT), qui normalement inhibe l'élastase leucocytaire; le variant Pittsburg perd ses propriétés anti-élastase pour devenir un puissant inhibiteur des facteurs de coagulation de type sérine protéase, et plus particulièrement de la thrombine [42]. En conséquence, les patients porteurs de cette mutation présentent un syndrome hémorragique. Un cas particulier de gain de fonction par acquisition de propriétés toxiques concerne la maladie de Huntington, affection neurodégénérative de progression rapide; le mécanisme pathogénique, par ailleurs commun aux maladies à expansion de polyglutamine, est lié à la constitution par les protéines altérées d'agrégats protéiques toxiques pour la cellule. Il en résulte une perte neuronale, touchant de manière prépondérante les neurones GABAergiques efférents du striatum [43].

### Situations complexes

#### L'exemple de RET

Ce gène code un récepteur de la membrane cellulaire. Lorsque son ligand, le GDNF (*glial cell derived neurotrophic factor*), se lie aux domaines extracellulaires, il induit la dimérisation du récepteur, qui transmet alors le signal à la cellule par l'intermédiaire de son domaine intracellulaire présentant une activité tyrosine kinase. Un grand nombre de mutations perte de fonction ont été rapportées, qui interfèrent avec la maturation post-traductionnelle de la protéine RET, une des causes de la maladie de Hirschsprung, maladie héréditaire autosomique dominante. Certaines mutations très spécifiques faux-sens sont pour leur part retrouvées dans des maladies totalement différentes, comme les cancers médullaires de la thyroïde familiaux. Ces mutations sont des mutations gain de fonction produisant un récepteur se liant de manière excessive au ligand, ou actif de manière constitutive et se dimérisant même en l'absence du ligand. De façon surprenante, certains patients présentent à la fois une maladie de Hirschsprung et un cancer médullaire de la thyroïde [44], ce qui illustre qu'une même mutation peut avoir différents effets (gain ou perte de fonction) dans différents types cellulaires, en fonction de l'expression du gène.

#### Allèle complexe

Un allèle complexe, ou allèle double-mutant, comporte deux anomalies moléculaires positionnées en *cis*. Avec plus de 1 300 mutations décrites dans le gène *CFTR*, la mucoviscidose est un bon modèle pour l'étude des allèles complexes [45]. Les études structure-fonction concernant ces allèles complexes ont montré que la coexistence de deux mutations faux-sens héritées en

*cis* peuvent aggraver fortement le phénotype *in vitro* et *in vivo* par rapport aux deux mutations considérées isolément. Ces observations ont des implications directes pour l'étude des relations phénotype-génotype et pour le diagnostic et le pronostic de la mucoviscidose. Leur fréquence étant probablement sous-estimée, ces allèles complexes participent certainement à la difficulté d'établir des corrélations phénotype-génotype, pour la mucoviscidose et d'autres pathologies autosomiques récessives.

Un autre exemple d'allèle complexe concerne le variant Antilles de la  $\beta$ -globine, qui combine en *cis* les mutations faux-sens p.Glu6Val et p.Val23Ile. Ce variant Antilles présente une solubilité plus faible que l'hémoglobine S (p.Glu6Val); sa présence à l'état hétérozygote est associée à la drépanocytose, alors que p.Glu6Val n'est associée à la maladie qu'à l'état homozygote [46].

#### Conséquences « à distance »

Dans certains cas, les mutations du gène entraînent des conséquences à distance sur l'expression ou la fonction d'autres gènes. C'est notamment le cas des mutations concernant des gènes impliqués dans le système de réparation des mésappariements de l'ADN (système MMR, *mismatch repair*) responsables du syndrome de prédisposition au cancer colorectal non polyposique. Les personnes prédisposées présentent une mutation germinale hétérozygote d'un gène codant une protéine du MMR. L'étape initiale de la progression tumorale est l'inactivation somatique de la copie normale du gène de prédisposition; l'acquisition d'un tel phénotype « mutateur » va alors favoriser l'accumulation de mutations au niveau de séquences répétitives de certains gènes (*tumor growth factor  $\beta$  receptor type II* et *Bax*, notamment), ouvrant une voie de cancérogenèse spécifique [47].

#### Conclusion

Au cours de ces 20 dernières années, la connaissance du génome humain ainsi que les progrès technologiques ont permis l'identification d'un nombre croissant de mutations à l'origine des maladies héréditaires humaines. Cependant, il est très rare que la totalité des mutations d'un gène pourtant impliqué de façon certaine dans une maladie mendélienne soit connue, ce qui suggère l'existence de nouveaux mécanismes à découvrir, portant notamment sur les séquences très conservées, mais non codantes, du génome humain [48].

À l'échelle d'une maladie, la connaissance du spectre des mutations permet de mieux comprendre la physiopathologie moléculaire (perte totale ou partielle de fonction, gain de fonction) et d'expliquer le mode de transmission récessif ou dominant. Néanmoins, le problème central à l'heure actuelle est d'analyser et de comprendre les relations entre génotype et phénotype. C'est pourquoi les bases de données informatisées spécifiques de locus morbides sont importantes à développer, comme par exemple celle concernant la mucoviscidose (<http://www.genet.sickkids.on.ca/cftr/>). Ces bases de données constituent un nouvel outil permettant d'étudier une maladie génétique donnée sous tous ses aspects: nosologiques, physiopathologiques, épidémiologiques, diagnostiques et thérapeutiques. ♦

## SUMMARY

### Mutations mechanisms and its consequences

The identification of mutations leading to human genetic diseases has grown into an intensive research field during the last few years. Through novel DNA analysis progress, it is now possible to determine the mutational spectrum for a given genetic disease and international databases are now available online. Genetic diagnosis of hereditary diseases has become an essential tool in genetic counselling and prenatal diagnosis. The knowledge of the deleterious mutation and the molecular associated mechanism is fundamental in order to devise the optimal molecular diagnosis strategy. This review aims to present the various mutation categories involved in genetic diseases (single base-pair substitutions, small deletions or insertions, dynamic mutations, gross DNA lesions...) and to summarize our current knowledge about the main molecular mechanisms responsible for these mutations. Their deleterious consequences on gene expression, including transcription and transcript maturation, and protein loss or gain of function are also discussed in this review. ♦

## REMERCIEMENTS

Les auteurs remercient Pascale Fanen, Geneviève Gourdon et Marc Jeanpierre pour leur aide.

## RÉFÉRENCES

1. Hamosh A, Scott AF, Amberger J, et al. Online Mendelian inheritance in man (OMIM). *Hum Mutat* 2000; 15 : 57-61.
2. Stenson PD, Ball EV, Mort M, et al. Human gene mutation database (HGMD) : 2003 update. *Hum Mutat* 2003; 21 : 577-81.
3. Krawczak M, Ball EV, Cooper DN. Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. *Am J Hum Genet* 1998; 63 : 474-88.
4. Brown LY, Brown SA. Alanine tracts : the expanding story of human illness and trinucleotide repeats. *Trends Genet* 2004; 20 : 51-8.
5. Lalioti MD, Scott HS, Buresi C, et al. Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. *Nature* 1997; 386 : 847-51.
6. Matsuura T, Yamagata T, Burgess DL, et al. Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. *Nat Genet* 2000; 26 : 191-4.
7. Liqouri CL, Ricker K, Moseley ML, et al. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* 2001; 293 : 816-7.
8. Albrecht A, Mundlos S. The other trinucleotide repeat : polyalanine expansion disorders. *Curr Opin Genet Dev* 2005; 15 : 285-93.
9. Warren ST. Polyalanine expansion in synpolydactyly might result from unequal crossing-over of HOXD13. *Science* 1997; 275 : 408-9.
10. Emanuel BS, Shaikh TH. Segmental duplications : an expanding role in genomic instability and disease. *Nat Rev Genet* 2001; 2 : 791-800.
11. Shaw CJ, Lupski JR. Implications of the human genome architecture for rearrangement-based disorders : the genomic basis of disease. *Hum Mol Genet* 2004; 13 : 57-64.
12. Abeyasinghe SS, Chuzhanova N, Krawczak M, et al. Translocation and gross deletion breakpoints in human inherited disease and cancer I : nucleotide composition and recombination-associated motifs. *Hum Mutat* 2003; 22 : 229-44.
13. Chuzhanova N, Abeyasinghe SS, Krawczak M, et al. Translocation and gross deletion breakpoints in human inherited disease and cancer II : potential involvement of repetitive sequence elements in secondary structure formation between DNA ends. *Hum Mutat* 2003; 22 : 245-51.
14. Kazazian HH Jr, Wrong C, Youssoufian H, et al. Haemophilia A resulting from *de novo* insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* 1988; 332 : 164-6.
15. Wallace MR, Andersen LB, Saulino AM, et al. A *de novo* Alu insertion results in neurofibromatosis type 1. *Nature* 1991; 353 : 864-6.
16. Deingerer PL, Batzer MA. Mammalian retroelements. *Genome Res* 2002; 12 : 1455-65.
17. Jeffreys AJ, Tamaki K, MacLeod A, et al. Complex gene conversion events in germline mutation at human minisatellites. *Nat Genet* 1994; 6 : 136-45.
18. Spence JE, Perciaccante RG, Greig GM, et al. Uniparental disomy as a mechanism for human genetic disease. *Am J Hum Genet* 1988; 42 : 217-26.
19. Surani MA. Genetics : immaculate misconception. *Nature* 2002; 416 : 491-3.
20. Fisher RA, Hodges MD, Newlands ES. Familial recurrent hydatidiform mole : a review. *J Reprod Med* 2004; 49 : 595-601.
21. Goriely A, McVean GA, van Pelt AM, et al. Gain-of-function amino acid substitutions drive positive selection of FGFR2 mutations in human spermatogonia. *Proc Natl Acad Sci USA* 2005; 102 : 6051-6.
22. Tufarelli C, Stanley JA, Garrick D, et al. Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat Genet* 2003; 34 : 157-65.
23. Weaving LS, Ellaway CJ, Gecz J, Christodoulou J. Rett syndrome : clinical review and genetic update. *J Med Genet* 2005; 42 : 1-7.
24. Ranum LP, Day JW. Myotonic dystrophy : RNA pathogenesis comes into focus. *Am J Hum Genet* 2004; 74 : 793-804.
25. Wijmenga C, Hewitt JE, Sandkuijl LA, et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet* 1992; 2 : 26-30.
26. van der Maarel SM, Frants RR. The D4Z4 repeat-mediated pathogenesis of facioscapulohumeral muscular dystrophy. *Am J Hum Genet* 2005; 76 : 375-86.
27. Gabellini D, Green MR, Tupler R. Inappropriate gene activation in FSHD : a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. *Cell* 2002; 110 : 339-48.
28. Winokur ST, Chen YW, Masny PS, et al. Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. *Hum Mol Genet* 2003; 12 : 2895-907.
29. Jiang G, Yang F, van Overveld PG, et al. Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. *Hum Mol Genet* 2003; 12 : 2909-21.
30. Masny PS, Bengtsson U, Chung SA, et al. Localization of 4q35.2 to the nuclear periphery : is FSHD a nuclear envelope disease? *Hum Mol Genet* 2004; 13 : 1857-71.
31. Houdayer C, Stoppa-Lyonnet D. Transcriptional abnormalities and genetic testing. *Med Sci (Paris)* 2005; 21 : 170-4.
32. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense : exonic mutations that affect splicing. *Nat Rev Genet* 2002; 3 : 285-98.
33. Higgs DR, Goodbourn SE, Lamb J, et al. Alpha-thalassaemia caused by a polyadenylation signal mutation. *Nature* 1983; 306 : 398-400.
34. Lejeune F, Maquat LE. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. *Curr Opin Cell Biol* 2005; 17 : 309-15.
35. Conti E, Izaurralde E. Nonsense-mediated mRNA decay : molecular insights and mechanistic variations across species. *Curr Opin Cell Biol* 2005; 17 : 316-25.
36. Inoue K, Khajavi M, Ohyama T, et al. Molecular mechanism for distinct neurological phenotypes conveyed by allelic truncating mutations. *Nat Genet* 2004; 36 : 361-9.
37. Vogt G, Chappier A, Yang K, et al. Gains of glycosylation comprise an unexpectedly large group of pathogenic mutations. *Nat Genet* 2005; 37 : 692-700.
38. Hobbs HH, Brown MS, Goldstein JL. Molecular genetics of the LDL receptor gene in familial hypercholesterolemia. *Hum Mutat* 1992; 1 : 445-66.
39. Notaro R, Afolayan A, Luzzatto L. Human mutations in glucose 6-phosphate dehydrogenase reflect evolutionary history. *FASEB J* 2000; 14 : 485-94.
40. Gajko-Galicka A. Mutations in type I collagen genes resulting in osteogenesis imperfecta in humans. *Acta Biochim Pol* 2002; 49 : 433-41.
41. Naski MC, Wang Q, Xu J, Ornitz DM. Graded activation of fibroblast growth factor receptor 3 by mutations causing achondroplasia and thanatophoric dysplasia. *Nat Genet* 1996; 13 : 233-7.
42. Owen MC, Brennan SO, Lewis JH, Carrell RW. Mutation of antitrypsin to antithrombin. alpha 1-antitrypsin Pittsburgh (358 Met leads to Arg), a fatal bleeding disorder. *N Engl J Med* 1983; 309 : 694-8.
43. Landles C, Bates GP. Huntingtin and the molecular pathogenesis of Huntington's disease. Fourth in molecular medicine review series. *EMBO Rep* 2004; 5 : 958-63.
44. Sasaki Y, Shimotake T, Go S, Iwai N. Total thyroidectomy for hereditary medullary thyroid carcinoma 12 years after correction of Hirschsprung's disease. *Eur J Surg* 2001; 167 : 467-9.
45. Clain J, Lehmann-Che J, Dugueperoux I, et al. Misprocessing of the CFTR protein leads to mild cystic fibrosis phenotype. *Hum Mutat* 2005; 25 : 360-71.
46. Monplaisir N, Merault G, Poyart C, et al. Hemoglobin S Antilles : a variant with lower solubility than hemoglobin S and producing sickle cell disease in heterozygotes. *Proc Natl Acad Sci USA* 1986; 83 : 9363-7.
47. Grady WM. Genomic instability and colon cancer. *Cancer Metastasis Rev* 2004; 23 : 11-27.
48. Dermizakis ET, Reymond A, Antonarakis SE. Conserved non-genic sequences - an unexpected feature of mammalian genomes. *Nat Rev Genet* 2005; 6 : 151-7.

TIRÉS À PART

M. Vidaud