

Point méthode : Exercices de séquence

Exemples d'exercices :

Le syndrome de Waardenburg est une maladie à transmission autosomique dominante due à une anomalie de migration des cellules de la crête neurale à destinée cutanée mélanocytaire et à destinée de l'oreille interne par atteinte du gène *PAX3*. Les principaux symptômes sont une surdité neurosensorielle (absence de cellules pigmentaires dans l'oreille interne), des anomalies pigmentaires (mèches blanches) et des particularités faciales : racine du nez fine, hypoplasie des ailes du nez, dystopie des canthi (coin interne des yeux trop écartés) et yeux de couleurs différentes.

Vous recevez en consultation Guillaume, un nouveau-né de 9 mois chez qui vous suspectez ce syndrome. Vous avez à votre disposition les séquences de référence d'ADNg et d'ADNc du gène *PAX3*.

Séquence 2

```

AGGTAAGCTCGTACCGTGATTCATGCGACAGGGGTAAGACCATCAGTAGTAGGGATAGTG 60
CCAAACCTCACTCACCCTGCCAATAAGGGATGAGGAACCCAAAAGGCGAACCGGGCCAG 120
      M R N P K G E P G Q 10
ACAACCCGGCGGTATCGCACTCAAAGCCGGGACACGACGCGTCACAGCCGAAAGAGGAA 180
T T R R Y R T Q S R D T T R H S R K E E 30
CCCCGGAGTGAAGACCTATGGGGCTGGATAAACTGCCGTGGCAACCGCCTTCAACAACC 240
P R S E D L W G W I K L P W Q P P S T T 50
CGAATACACCATAGCGATTTGACGCGGGCCTCGGAATACCGTTTCAGCAGGCGCTTGGT 300
R I H H S D L T R R L G I P F Q Q A L G 70
AAGGCCATCGGAATACCAGGTATCGTGTAAAGTAGCCGGCGAAAAGTCGGGGAGACAGCAG 360
K A I A N T R Y R V S S R R K S G R Q Q 90
CGGCTGCAGACATTATACCGCAACAACCTAAGGTGAGAGAACTCTTATGGGTTACAGCA 420
R L Q T L Y R N N T K V R E L L W V T A 110
ATCACATCCAAGACTGGCTATGCACGAAGCAACTCTTCAGTGTAAAATGTAGACCCTG 480
I T S K T G Y A R S N S S V L K C R P L 130
TATTTGGGATGCGGGTAGTAGATGAGTGCAGGGACTCCGAGGTCAAGTACATTACCTCT 540
Y L G C G * 135
CATAGGGGGC 550

```

Séquence 1

Les séquences exoniques sont en majuscule, les bordures introniques en minuscule.

```

AGGTAAGCTCGTACCGTGATTCATGCGACAGGGGTAAGACCATCAGTAGTAGGGATAGTG 60

CCAAACCTCACTCACCCTGCCAATAAGGGgtccttacctgaagaataagtgtcagccag 120

tgtaaccagATGAGGAACCCAAAAGGCGAACCGGGCCAGACAACCCGGCGGTATCGCACT
      M R N P K G E P G Q T T R R Y R T 17
CAAAGCCGGGACACGACGCGTCACAGCCGAAAGAGGAACCCGGAGTGAAGACCTATGG
      Q S R D T T R H S R K E E P R S E D L W 37
GGCTGGATAAAACTGCCGTGGCAACCGCCTTCAACAACCCGAATACgtggcacttcagga
      G W I K L P W Q P P S T T R I
ggcgccggaggggggatgttttctactattcgaggccgttcgttataactagtgtcggtt

cctagccgctataattgtctctttgccgactaatgagaacaaccagACCATAGCGATTG
      H H S D L 57
ACGCGCGCCTCGGAATACCGTTTCAGCAGGCGCTTGGTAAGGCCATCGCGAATACCAGG
      T R R L G I P F Q Q A L G K A I A N T R 77
TATCGTGTAAAGTAGCgtagggccgcacgcaagataaactgctagggaaaccggtttccac
      Y R V S S
gaccggtgcacgatttaatttcgccgacgtgatgacattccaggcagtgctctgccc

ggaccctctcgtgattggtagctggacatgcccttgaagataaacaagagcctgcc

tgtctaatgatctagCGGCGAAAGTCGGGGAGACAGCAGCGGCTGCAGACATTATACCGC
      R R K S G R Q Q R L Q T L Y R 97
ACAACACTAAGGTGAGAGAACTCgttaattgactacgcgttcctctagaccttacttga
      N N T K V R E L
ccggatacagtgctcttgacacagTTATGGGTTACAGCAATCACATCCAAGACTGGCTAT
      L W V T A I T S K T G Y 117
GCACGAAGCAACTCTTCAGTGTAAAATGTAGACCCTGTATTTGGGATGCGGGTAGTAG
      A R S N S S V L K C R P L Y L G C G * 135
ATGAGTGCAGGGACTCCGAGGTCAAGTACATTACCTCTCATAGGGGGCgttctagatca

cgttaccaccatattcagcagcatgacaccatctcgcgt

```

QCM 1

QCM 1 : À propos du gène *PAX3*, cochez la(les) réponse(s) vraie(s) :

- A. Ce gène contient 4 exons.
- B. La protéine obtenue fait 135 acides aminés.
- C. La délétion de l'exon 4 entraîne un décalage du cadre de lecture.
- D. La délétion de l'exon 3 entraîne un décalage du cadre de lecture.
- E. Une délétion des nucléotides g.895 à g.897 entraîne la production d'une protéine rallongée.

CORRECTION : B et D vraies.

D'abord, il faut déterminer quelle est la séquence d'ADNg et quelle est celle d'ADNc. L'ADNg est l'ADN retrouvé dans le noyau de nos cellules, avec des introns, des exons et des séquences régulatrices, alors que l'ADNc correspond à la rétrotranscription de l'ARNm après toutes les étapes de maturation, il n'y a donc plus les introns.

Ici, la séquence d'ADNg est donc celle qui contient les introns, donc la **séquence 1**, l'ADNc est donc la **séquence 2**.

Ensuite, il suffit de compter les exons dans la séquence 1, même ceux qui ne sont pas codants ! Il y en a donc **5 dans ce gène**.

La numérotation donnée dans la séquence 1 est celle des AA, le codon STOP ne comptant pas pour un AA, il y en a donc bien **135**.

Pour la délétion d'exon, il existe plusieurs méthodes pour savoir si il y aura un décalage du cadre de lecture ou non :

1. On crée un tableau ou on note la fin et le début de chaque exon :

	Exon 2	Exon 3	Exon 4	Exon 5
Début	ACCAT H H	CGGCGA R R	TTATGG L W
Fin	CGAATAC R I	AGTAGC S S	GAATC E T	...

Ici, on voit bien que si on enlève l'exon 3, on passe de CAC (H) en temps normal (fin exon2-début exon 3) à CCG puis GCG donc il y a bien un décalage.

Si on enlève l'exon 4, On passe de AGC puis CGG à AGC puis TTA et TGG mais il n'y a pas de décalage du cadre de lecture !! On enlève juste un exon.

2. On note le numéro de la première et de la dernière base de chaque exon en fonction de sa place dans le codon :

	Exon 2	Exon 3	Exon 4	Exon 5
Début	+2	+1	+1
Fin	+1	+3	+3

Dans l'exon 2, on finit par un C qui est en première place dans son codon (CAC) puis l'exon 3 commence par un A qui est le deuxième du codon (toujours CAC) et on continue comme ça. Pour qu'il n'y ait pas de décalage, il faut une suite 1-2-3-1, etc.

NDLR – Le + ici n'a pas d'importance.

Si on enlève l'exon 3, on passe d'un 1 à la fin de l'exon 2 à un 1 au début de l'exon 4 soit une suite 1-1 → Décalage.

Si on enlève l'exon 4, on passe d'un 3 à la fin de l'exon 3 à un 1 au début de l'exon 5 donc une suite 3-1 → Pas de décalage, on a juste enlevé tout ce qu'il y a au milieu.

Enfin, pour l'item 5, on se réfère à la séquence 1 (ADNg) car la numérotation est en g.895.

On remarque qu'il y a 20 AA par ligne, soit $20 \times 3 = 60$ nucléotides par ligne, on compte donc ainsi jusqu'à arriver au numéro 895. On tombe sur les 3 nucléotides du codon STOP. En supprimant le codon STOP, la traduction est donc censée continuer cependant attention, il y a un autre codon TAG directement après le premier. Il n'y a donc dans ce cas-là pas de rallongement de la protéine. Attention à bien regarder la suite de la séquence !

QCM 2

QCM 2 : Vous identifiez chez Guillaume 3 variants suspects :

- c.384 A > G ;
- c.234 T > C ;
- c. 304del.

- A. La région 5'UTR fait 90pb.
- B. Le variant c.384A>G est une mutation faux-sens.
- C. Le variant c.304del entraîne l'apparition d'un codon stop.
- D. Le variant c.234 T>C entraîne la création d'un AA pouvant subir une phosphorylation.
- E. Le variant C.304del entraîne la production d'une protéine de 102 AA.

CORRECTION : A,C et D vraies.

```

S D L T R R L G I P F Q Q A L G 70
ATACCAGGTA TCGTGTAAAGTAGCCGGCGAAAGTCGGGGAGACAGCAG 360
N T R Y R V S S R R K S G R Q Q 90
TATACCGCAACAACACTAAGGTGAGAGAACTCTTATGGGTTACAGCA 420
L Y R N N T K V R E L L W V T A 110
CTGGCTATGCACGAAGCAACTCTTCAGTGTAAAATGTAGACCCCTG 480
T G Y A R S N S S V L K C R P L 130
GGTAGTAGATGAGTGCAGGGACTCCGAGGTCAAGTACATTACCCTCT 540
G * 135
550

```

La numérotation en c. nous indique de nous pencher sur la séquence 2 (ADNc). Attention, la numérotation **commence au A du codon Start ATG** ! Tout ce qui est avant (qui correspond à la région 5'UTR) est **numéroté en c.-...** Ici, on compte donc 90 nucléotides avant le A du ATG, donc la région 5'UTR fait bien 90 nucléotides.

Comment trouver le bon nucléotide avec la numérotation en c ?

La numérotation donnée à droite de la séquence commence au niveau de ma région 5'UTR et non pas au A du codon ATG, il y a donc 90 nucléotides de différence entre les 2 numérotations. Il suffit alors de rajouter 90 à la numérotation en c donnée pour pouvoir par la suite utiliser celle qui nous est donnée.

Pour le variant c.384A>G, on cherche le nucléotide numéro $384+90 = 474$ dans la numérotation de droite, on tombe sur le A coloré en orange sur l'image ci-dessus. On procède de la même façon pour les 2 autres.

Le A du variant 1 appartient au codon AGA (R), si on le remplace par un G, on obtient AGG qui code aussi pour un R donc c'est un **variant synonyme**.

Le T du variant 2 appartient au codon TAT (Y), si on le remplace par un C, on obtient TAC qui code pour un T donc c'est un **variant faux-sens**.

Le c.304 du variant 3 correspond au G du codon GTG (V), si on le supprime, on décale le cadre de lecture et on entraîne l'apparition d'un TGA (STOP), c'est donc un **variant non-sens**. Ce codon STOP étant en position p.102, il entraîne la production d'une protéine de 101 AA car le codon STOP ne code pas pour une AA !

QCM 3

QCM 3 : Vous recevez quelques années plus tard le frère de Guillaume, lui aussi suspecté d'être atteint de la maladie, mais vous ne retrouvez pas les variants détectés chez Guillaume. Vous décidez alors d'étudier le gène *PAX2*, lui aussi responsable de ce syndrome. Vous trouvez alors un variant dans l'exon 6 du gène *PAX2* qui vous semble suspect.

Vous disposez de la séquence d'ADNc de l'exon 6 du gène *PAX 2* d'un sujet sain (S) et du frère de Guillaume (M). Le premier nucléotide donné ici est le nucléotide n°1428 de la séquence d'ADNc et le A du codon start est en position 61 de l'ADNc.

- A. La variation détectée peut s'écrire : c.1457C>T.
- B. La variation peut s'écrire : c.1397C>T.
- C. La variation peut s'écrire : p.465Q>*.
- D. La variation peut s'écrire : p.486S>L.
- E. La variation peut s'écrire : p.486S>*.

CORRECTION : B et C vraies.

S	AGGCTATTCGATAACGTACAGTGCTAGTCAGTTCAGTCATGATACGATAGGTCACATCTATTACCCTAGT
M	AGGCTATTCGATAACGTACAGTGCTAGTTAGTTCAGTCATGATACGATAGGTCACATCTATTACCCTAGT

Il y a plusieurs informations importantes dans l'énoncé. On dit ici que le A de début est le N°1428 de la totalité de la séquence d'ADNc, donc qui comprend les régions UTR ! 1428 **ne correspond donc pas à sa séquence en c.** On nous dit que le A de l'ATG se situe en position 61, donc les 60 nucléotides précédents forment la **région 5'UTR**. Pour savoir la numérotation en c. du premier nucléotide qui nous est donné, il suffit cette fois d'enlever les 60 nucléotides de la région 5'UTR. Donc $1428 - 60 = 1368$.

Le premier A qui nous est fourni se numérote **c.1368**.

Ensuite, il nous faut savoir quelle est la position du A dans son codon (**1^{ère}, 2^{ème} ou 3^{ème} place**).

Une règle simple est que si l'on divise la numérotation c. d'un nucléotide par 3 et que l'on obtient un chiffre entier, le nucléotide est en dernière place de son codon :

- Si on obtient un chiffre se terminant par « ,333 », le nucléotide est en 1^{ère} position du codon.
- Si on obtient un chiffre se terminant par « ,666 », le nucléotide est en 2^{ème} position du codon.

Ici, $1368 / 3 = 456$, c'est un chiffre entier, on sait donc que le A est le dernier nucléotide du codon 456 (il y a donc un codon chevauchant entre les exons 5 et 6). On peut donc maintenant continuer.

S	AGGCTATTCGATAACGTACAGTGCTAGTCAGTTCAGTCATGATACGATAGGTCACATCTATTACCCTAGT
M	AGGCTATTCGATAACGTACAGTGCTAGTTAGTTCAGTCATGATACGATAGGTCACATCTATTACCCTAGT

Notre mutation se situe donc en position c.1397 et concerne donc le C du codon CAG (en position 465) qui devient alors TAG (codon STOP).

Donc notre mutation s'écrit : **c.1397C>T ou p.465Q>***

Bon courage !!