



Tutorat Lyon Est

Unité d'Enseignement 2

BANQUE DE QCM

2020-2021

2021-2022

2022-2023

Génome (séquences)

QUESTIONS et REPONSES

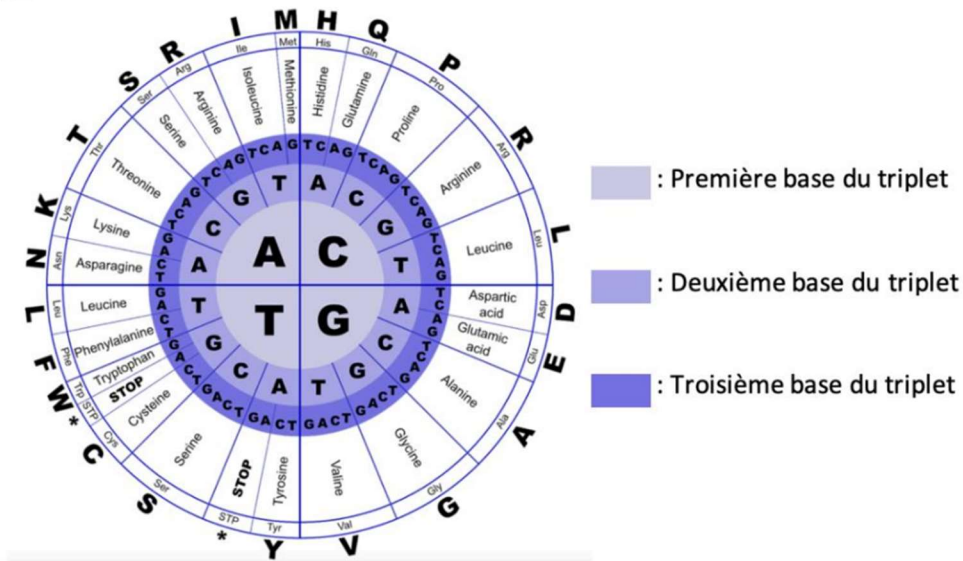
Question 1 – Séquence (***) :

Vous étudiez les modifications génomiques pouvant affecter la partie initiale d'un gène à l'aide de sa séquence nucléotidique d'ADNc (représenté ligne 1). Le début de la chaîne d'acide aminé correspondant à cette séquence est représenté en dessous (ligne 2). A partir de ces séquences et du diagramme représentant le code génétique ci-dessous, quelle(s) affirmation(s) est/ sont vraie(s) ?

GATCTCCCATGGGTCGATATAGCGCTA**A**TGG**G**CTT**C**ACA**G**T**C**AA**A**TTAG**C**CG**G**GTGG...

M A S Q V K F A G W ...

Code génétique



- A. L'ADN complémentaire d'un gène peut être obtenu par une étape de reverse transcriptase de l'ARNm mature de ce gène.
- B. Dans cette séquence, le +1 de l'ADN complémentaire correspond à une Guanine.
- C. Si on considère la mutation « c.9A>G », un acide aminé est modifié : il s'agit d'une modification non synonyme.
- D. Si on considère la mutation « c.20T>A », il s'agit d'une mutation NON-SENS.
- E. La mutation d'un seul acide nucléique sur l'ADNc peut être responsable de la formation d'une protéine non-fonctionnelle.

Ce type d'exercice, même s'il peut paraître difficile au premier abord, est classique à l'examen. Le but de cet exercice était de vous initier à l'étude d'une séquence grâce à la nomenclature apprise lors du premier cours sur le génome. Lors de l'examen, il faudra travailler sur des séquences bien plus longues mais vous aurez pour vous entraîner un ED de biologie moléculaire et un cours du soir proposé par le Tut'. Cet exercice nécessitait de connaître la numérotation de l'ADNc et d'avoir appris des notions importantes sur le code génétique et les SNV.

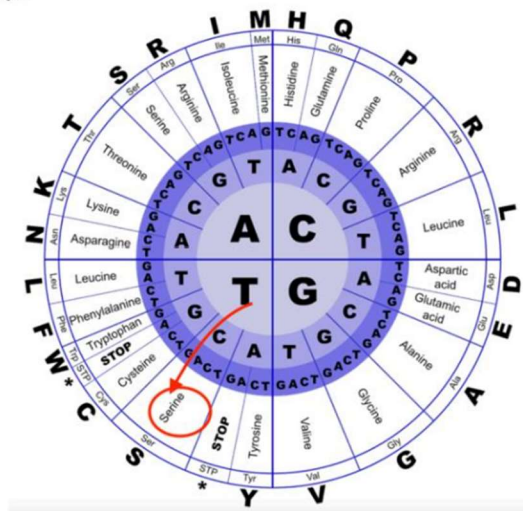
A VRAI L'ADNc est un ADN simple brin qui peut être obtenu par une étape de reverse transcriptase sur un ARNm mature d'un gène d'intérêt. Les introns et autres séquences épissées lors de la maturation de notre pré-ARNm sont donc absentes de la séquence de l'ADNc.

B FAUX Attention à la nomenclature ! Dans notre ADNc, le +1 correspond au +1 de la traduction, soit le A du premier codon ATG. Ici, notre +1 correspond donc à une Adénine.

+1 3 5 7 9 11 13 15 17 19 21 23 25 27 29
 GATCTCCCATGGGTCGATATAGCGCTA**ATGGCTTCACAAGTCAAATTAGCCGGGTGG...**
 M A S Q V K F A G W ...

C FAUX Grâce à la numérotation trouvé ci-dessus, on sait qu'une mutation c.9A>G correspond à un SNV en position 9 de notre ADNc avec une adénine qui devient une guanine. On se situe au niveau du dernier nucléotide d'un codon qui, initialement, code pour une Sérine (S). On regarde avec notre tableau du code génétique ce qu'entraîne cette modification génétique. On regarde d'abord au niveau du cadran inférieur gauche correspondant à notre premier acide nucléique, un T puis vers la ligne du C et enfin la sous-partie du G.

Code génétique



On constate que malgré le SNV, ce triplet nucléotidique code encore pour une sérine. Ce SNV entraîne donc une modification/mutation synonyme qui n'a pas de conséquence au niveau de la chaîne d'acide aminé car la Sérine en position 3 reste inchangée.

D VRAI Même raisonnement que pour la question précédente. Ici on se situe en +20 de notre séquence d'ADNc, une thymine devient une adénine. Notre triplet nucléotidique bleu clair TTA devient alors TAA. On s'intéresse à notre tableau du code génétique comme précédemment et on observe que ce triplet code un codon STOP, signal de fin de la traduction, il s'agit donc d'une mutation NON-SENS.

E VRAI Comme évoqué à la question précédente, un SNV peut entraîner la formation d'un codon STOP prématuré, cela peut entraîner la formation d'une protéine non fonctionnelle car incomplète, son activité biologique ne pourra alors pas avoir lieu.

Énoncé commun aux questions 1 à 5 (VOIR SEQUENCE TP53 EN ANNEXE):

Le gène TP53 se trouve sur le chromosome 17, et code pour un facteur de transcription p53 qui permet d'induire la réparation de l'ADN et réguler le cycle cellulaire. Dans une cellule saine, p53 active la transcription de nombreux gènes, dont MDM2 qui permet d'ubiquitiner p53 et donc dégrader p53 en continu, ce qui veut dire que dans une cellule saine p53 n'est que très peu présent. p53 régule donc son niveau d'expression protéique via MDM2. Lors d'un stress cellulaire (signal de dommage de l'ADN), ARF vient séquestrer MDM2 et empêche donc p53 d'être dégradé par MDM2, ce qui permet à p53 de pouvoir activer la transcription des gènes réparant l'ADN sans être dégradé. L'exon 5 code pour le domaine de transactivation, et l'exon 2 code pour le DBD (=DNA Binding Domain) de p53, chacun indispensable à la fonction de p53. En annexe se trouvent les séquences 1 et 2 du gène TP53 (bordures introniques en minuscules et exons en majuscules).

Question 1 – Séquence : (*)**

- A. La séquence 1 correspond à l'ADNg du gène TP53.
- B. Le gène possède 7 exons codants.
- C. Le 5'UTR possède 105 nucléotides.
- D. L'exon 1 est composé de 30 nucléotides.
- E. Ce gène code pour une protéine de 305 AA.

Pas de panique ! C'est la première fois que vous avez un exo type « problème de réflexion » avec une séquence. Au début, c'est compliqué à aborder mais avec le temps et les ED de biomol on gagne en automatisme et réflexion, puis cette colle sera un entraînement à refaire pour bien assimiler toutes les notions qu'on pourrait attendre de vous aux examens terminaux (certes cette colle est DIFFICILE, mais elle a pour but de vous adapter à tout type de question le jour j). Ne vous découragez pas ! Avec de l'entraînement ce sera dans votre poche ! ☺)

En premier lieu lorsque vous avez une séquence, c'est regarder si vous avez un exon non codant PUIS numérotez les exons (1/2/3/4...) ! Avec l'habitude de s'exercer ça viendra tout seul ne vous inquiétez pas, la première fois qu'on voit une séquence c'est fort logique qu'on ne comprenne rien. ☺

A FAUX La séquence ayant des minuscules (introns) est celle de l'ADNg. Donc ici, la séquence 1 correspond à l'ADNc (ARNm rétrotranscrit en ADN), et la séquence 2 correspond à l'ADNg.

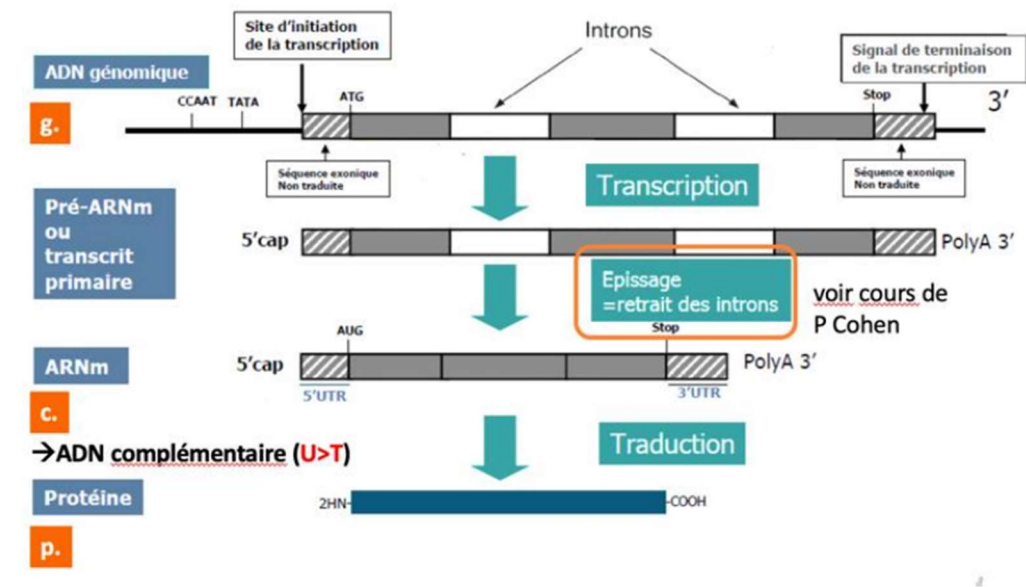
B FAUX Ce gène possède 7 exons, mais seulement 6 exons codants, car le premier exon est non codant (ne code pour aucun AA). Le premier exon, non codant donc, possède 30 nucléotides (en jaune) (**D VRAI**).


```

gatgtgggtgaaaattactacctatcaagtttctgttaccctcctgtaaatggtgta 60
agtgactgaccccagtggtatcctgtgacatgtaacagccaggagtgttaagcgttcaag 120
ACTCCACTCCACTTGGCTGTGATCACC AACgtctggtctgactggcttggagaaattcg 180
Ctgtattccttggatgcatgggttctggtctgactgacttggaaattccccgacctgacc 240
Cgccccagagaaaatccccagccagctttatagggcggcgggcggcctgcaagcccc 300
tcggcgtgcaattccttggatgcatggtagAGCGCCCCAGCGAGGAAGCAGCGCGCAGCC 360
CGCGGCCAGCGCACCCGAGCAGCGCCCGCAGCTCGTCCGCGCCATGTTCCAGGCGGCC 420
M F Q A A 5
GAGCGCCCCAGGAGTGGGCCATGGAGGGCCCCCGCAGCGGGCTGAAGAAGGAGCGGCTA 480
E R P Q E W A M E G P R D G L K K E R L 25
CTGGACGACCGCCAGCAGCGGCTGGACTCCATGAAAGACGAGGAGTACGAGCAGATG 540
L D D R H D S G L D S M K D E E Y E Q M 45
GTCAAGGAGCTGCAGGAGATCCGCCTCGAGCCGAGGAGGTGCCGCGGGCTCGGAGCCC 600
V K E L Q E I R L E P Q E V P R G S E P 65
TGGAAAGCAGCAGCTCACCGAGGACGGGGACTCgtgcctaatactacgatggaactga 660
W K Q Q L T E D G D S 75
ataaagtggcgggagccggggctgggggtctcggatcctaccctgcatcgcacatcgtc 720
ctgcatcgaccttaccgattcca (...) acccctggggtttttccctctcttccccacag 920
GTTCTGCACCTTGGCCATCATCCATGAAGAAAAGGCACTGACCATGGAAGTATCCGCCA 980
F L H L A I I H E E K A L T M E V I R Q 96
GGTGAAGGGAGACCTGGCCTTCCCAACTTCCAGAACAACCTGCAGCAGatactctaatt 1040
V K G D L A F L N F Q N N L Q Q 112
ccgcccgttgcctgcccgggttctctctgacc (...) acctgaatccttactacgat 1240
cgactacgttcaatccaacggccattggcttactttctctggtctctcttgcattcggtg 1300
actgactgacctgacctggagACTCCAAACCAGCCAGAAATGCTGAGGCACCTCTGGGA 1360
T P N Q P E I A E A L L G 125
GCTGGCTGTGATCCTGAGCTCCGAGACTTTCGAGGAAATACCCCCCTACACCTGCCTGT 1420
A G C D P E L R D F R G N T P L H L A C 145

```

C VRAI Pour calculer le 5'UTR, déjà il faut savoir à quoi il correspond, et pour cela je vous conseille VIVEMENT de bien visualiser ce schéma pour les séquences :



On voit donc que le 5'UTR est une séquence exonique (en majuscule) transcrite (présente dans ADNc) mais non traduite (ne code pour aucun AA), cette séquence se situe en amont du 1er ATG codant pour une méthionine, on regarde donc dans l'ADNc en amont du ATG :

```

ACTCCACTCCACTTGGCTGTGATCACCAACAGCGCCCGAGCGAGGAAGCAGCGCGCAGCC 60
CGCGGGCCAGCGCACCCGCAGCAGCGCCCGCAGCTCGTCCGCGCCATGTTCCAGGCGGCC 120
M F Q A A 5
GAGCGCCCCAGGAGTGGGCCCATGGAGGGCCCCGCGACGGGCTGAAGAAGGAGCGGCTA 180
E R P Q E W A M E G P R D G L K K E R L 25
CTGGACGACCGCCACGACAGCGGCCTGGACTCCATGAAAGACGAGGAGTACGAGCAGATG 240
L D D R H D S G L D S M K D E E Y E Q M 45
GTCAAGGAGCTGCAGGAGATCCGCCTCGAGCCGAGGAGGTGCCGCGGGCTCGGAGCCC 300
V K E L Q E I R L E P Q E V P R G S E P 65
TGGAAGCAGCAGCTCACCGAGGACGGGGACTCGTTCCTGCACTTGGCCATCATCCATGAA 360
W K Q Q L T E D G D S F L H L A I I H E 85
GAAAAGGCACTGACCATGGAAGTGATCCGCCAGGTGAAGGGAGACCTGGCCTTCTCAAC 420
E K A L T M E V I R Q V K G D L A F L N 105
TTCCAGAACAACCTGCAGCAGACTCCAAACCAGCCAGAAATTGCTGAGGCACTTCTGGGA 480

```

Le 5'UTR fait 105 nucléotides.

D VRAI Voir correction B ci-dessus.

E VRAI Pour cela on a juste à regarde la numérotation des AA (le dernier AA est le 305 ème).

```

TGGGGCCGCCAAGCACC CGGATACAGCAGCAGCTGGGCCAGCTGACACTAGAAAACCTT 900
W G R P S T R I Q Q L G Q L T L E N L 265
CAGATGCTGCCAGAGAGTGAGGATGAGGAGAGCTATGACACAGAGTCAGAGTTCCAGGAG 960
Q M L P E S E D E E S Y D T E S E F T E 285
TTCACAGAGGACGAGCTGCCCTATGATGACTGTGTGTTTGGAGGCCAGCGTCTGACGTTA 1020
F T E D E L P Y D D C V F G G Q R L T L 305
TGAGCGCAAAGGGGCTGAAAGAACATGGACTTGTATATTTGTACAAAAAAAAGTTTTAT 1080
*
TTTTCTTAAAAAAGCAAAAAAGCAAAAAAGCAAAAAAGCAAAAAAGCAAAAAAGCAAAAAAG 1140

```

Question 2 – Séquence back to back : (**)

- La mutation c.757A>G correspond à une mutation non sens.
- La mutation g.3069T>G correspond à une mutation faux sens.
- L'épissage alternatif de l'exon 4 conduit à un décalage du cadre de lecture lors de la traduction.
- L'épissage alternatif de l'exon 5 conduit à un décalage du cadre de lecture lors de la traduction.
- Le skipping simultané des exons 4 et 5 provoque un décalage du cadre de lecture lors de la traduction.

Pour la numérotation en c., il faut savoir que le c.1 correspond au A du premier ATG codant pour une méthionine, donc lorsque vous avez une mutation c.X, il faut rajouter ce qui est en amont du ATG sur l'ADNc, c'est-à-dire le 5'UTR, donc on fait X+5'UTR pour trouver la mutation grâce à la numérotation donnée sur la séquence (voir exemples ci-dessous pour mieux comprendre).

Pour la mutation en g., c'est plus simple car c'est la même numérotation que sur la séquence de l'ADNg, donc vous avez juste à chercher la mutation directement sur l'ADNg avec le numéro donné dans l'item.

L'épissage alternatif/exon skipping signifie un « saut » d'exon, c'est-à-dire que lors de la maturation de l'ARNm un exon va être épissé comme un intron (éliminé comme les introns), alors que souvent il ne devrait pas être éliminé (mais cela peut se passer aussi

de manière physiologique). Pour savoir si ces « sauts » d'exons provoquent un décalage du cadre de lecture, on va regarder si initialement il y a chevauchement d'un codon entre les exons, et si épissage alternatif il y a, si cela provoque un décalage du cadre de lecture (voir C/D/E ci-dessous pour mieux comprendre).

A FAUX FAUX sens, on nous donne la mutation c.757A>G, comme dis ci-dessus, je rajoute le 5'UTR pour ensuite pouvoir trouver directement ma mutation sur mon ADNc avec les numéros sur le côté, donc 757 + 105 (taille 5'UTR expliqué dans la question 5 C)) = **862**, je vais donc chercher au n°862 de mon ADNc (qui doit normalement correspondre à un A) : pour le trouver je regarde la numérotation sur le côté, en vert, puis je rajoute 22 pour arriver au 862 et j'arrive au premier A du codon ATA de l'isoleucine.

```

H G Y L G I V E L L V S L G A D V N A Q      205
GAGCCC TGTAATCACCTCGCAGTGGACCTGCAAAATCCTGACCTGGTGTCACTCCTGTTG      780
E P C N H L A V D L Q N P D L V S L L L      225
AAGTGTGGGCTGATGTCAACAGAGTTACCTACCAGGGCTATTCTCCCTACCAGCTCACC      840
K C G A D V N R V T Y Q G Y S P Y Q L T      245
TGGGGCCGCCAAGCACCCGGTACAGCAGCAGCTGGGCCAGCTGACACTAGAAAACCTT      900
W G R P S T R I Q Q Q L G Q L T L E N L      265
CAGATGCTGCCAGAGAGTGAGGATGAGGAGAGCTATGACACAGAGTCAGAGTTCACGGAG      960
Q M L P E S E D E E S Y D T E S E F T E      285
TTCACAGAGGACGAGCTGCCCTATGATGACTGTGTGTTTGGAGGCCAGCGTCTGACGTTA      1020
F T E D E L P Y D D C V F G G Q R L T L      305
TGAGCGCAAAGGGGCTGAAAGAACATGGACTTGTATATTTGTACAAAAAAAAGTTTTAT      1080
*

```

A modifié en G, donc ATA→GTA, ce qui code pour une Valine (pour cela je regarde ce que donne le codon GTA avec le code nucléotidique) et non une isoleucine comme avant la mutation, cette mutation change l'AA en un AA différent donc c'est une mutation Faux sens (non sens correspond à une mutation qui induit un codon stop prématuré).

B FAUX NON sens, on cherche 3069 sur l'ADNg, et normalement on devrait tomber sur un T (car g.3069T>G dans l'item ce qui montre que ce SNV change un T) :

```

ACACTAGAAAACCTTCAGATGCTGCCAGAGAGTGAGGATGAGGAGAGCTATGACACAGAG      2550
T L E N L Q M L P E S E D E E S Y D T E      280
TCAGAGTTCACGGAGTTCACAGAGGACGAGGtacttgatctaccgatcgatctagctaac      2610
S E F T E F T E D E      290
atgagtctgtgaaactccttcggcgctcttaactaat (...) atcgatcgacctatggcaac      3000
acttgcacgcattcctgactcatcatatgacgcttaacggctttttcccttgattcag      3060
CTGCCCTA GATGACTGTGTGTTTGGAGGCCAGCGTCTGACGTTATGAGCGCAAAGGGGC      3120
L P Y D D C V F G G Q R L T L *      305
TGAAAGAACATGGACTTGTATATTTGTACAAAAAAAAGTTTTATTTTCTAAAAAAGA      3180
AAAAAGAAGAAAAATTTAAAGGGTGTACTTATATCCACACTGCACACTGCCTGGCCCAA      3240
AACGTCTTAttgtgtaggatcagccctcattttggttctttgtgaactttttgtaggg      3300

```

T modifié en G donc TAT→TAG, donc le triplet ne code plus pour une tyrosine mais pour un stop prématuré ! (TAG, TAA, TGA sont les 3 codons stop, vous pouviez aussi le savoir grâce au code nucléotidique).

C VRAI Pour l'exon skipping de l'exon 4 : déjà on repère où est l'exon 4, ensuite on le barre (au crayon à papier, ou on le met entre parenthèses ou on s'imagine dans notre tête qu'il n'est plus là) :


```

ctgcatcgaccttacggattcca (...) acccctggggtttttccctctcttccccacag 920
GTTCTGCACTTGGCCATCATCCATGAAGAAAAGGCACTGACCATGGAAGTGATCCGCCA 980
  F L H L A I I H E E K A L T M E V I R Q 96
GGTGAAGGGAGACCTGGCCTTCCTCAACTTCCAGAACAACCTGCAGCAGgtactctaatt 1040
  V K G D L A F L N F Q N N L Q Q +3 112
ccgcgccgcttgccctggcccggttctctctgacc (...) acctgaatccttactacgat 1240
cgactacgttcaatccaacggccattggcttactttctctggtctctcttgcattcggtg 1300
actgactgacctgacctggag [REDACTED] 1360
      T P N Q P E I A E A L L G 125
[REDACTED] 1420
  A G C D P E L R D F R G N T P L H L A C 145
[REDACTED] 1480
  E Q G C L A S V G V L T Q S C T T P H L 165
[REDACTED] gtaacgtcatgcattccactgactgatgc 1540
  H S I L K A T N Y N 175
acatgtctgcctccctgcccaccccctcg (...) acatccaatcatgaccatac 1780
acctgactactaccacgatgactacgaggtggttatactttttccttttgttcttccag 1840
GCCACACGTGTCTACACTTAGCCTCTATCCATGGCTACCTGGGCATCGTGGAGCTTTTGG 1900
G+2H T C L H L A S I H G Y L G I V E L L 195
TGTCTTGGGTGCTGATGTCAATGCTCAGgtacttacgtcatgcaactgatgtcaatcqa 1960
V S L G A D V N A Q 205

```

Ensuite, on va regarder la fin de l'exon précédent l'exon épissé (le 4) et le début de l'exon suivant l'exon épissé, donc on regarde ici la fin de l'exon 3 et le début de l'exon 5. On voit que la fin de l'exon 3 est un codon non chevauchant (qui se termine sur le même exon où il commence) : le CAG (en jaune à côté du 1040). On note +3 car ça se termine par 3 nucléotides codant pour un AA. Ensuite on regarde la même chose mais au début de l'exon 5, on voit (en jaune) que l'exon commence par un codon incomplet (car chevauchant avec la fin de l'exon 4) GC (il manque le nucléotide précédant le GC qui se trouve sur l'exon 4 épissé), donc on note +2 (en gros on compte le nombre de nucléotides codant pour un AA qui sont sur le même exon, donc la plus part du temps c'est +3, mais si on a un codon chevauchant alors un exon codera par exemple pour un seul nt du codon, et l'exon suivant pour les deux nt restants du codon). Après cela, vous vérifiez que **le chevauchement des exons 3 et 5** (des exons bordant l'exon épissé, donc le précédent et le suivant) suit le cadre de lecture (multiple de 3), ici on a +3 et +2= 5, 5 n'étant pas un multiple de 3 alors cela va décaler le cadre de lecture : (nouvelle lecture des triplets, 3 par 3, à l'exon 5 on lira GCC puis ACA etc, ce qui codera pour différents AA qu'initialement et donnera probablement un STOP précoce) :

```

GGTGAAGGGAGACCTGGCCTTCCTCAACTTCCAGAACAACCTGCAGCAGgtactctaatt 1040
  V K G D L A F L N F Q N N L Q Q 112
ccgcgccgcttgccctggcccggttctctctgacc (...) acctgaatccttactacgat 1240
cgactacgttcaatccaacggccattggcttactttctctggtctctcttgcattcggtg 1300
actgactgacctgacctggag [REDACTED] 1360
      T P N Q P E I A E A L L G 125
[REDACTED] 1420
  A G C D P E L R D F R G N T P L H L A C 145
[REDACTED] 1480
  E Q G C L A S V G V L T Q S C T T P H L 165
[REDACTED] gtaacgtcatgcattccactgactgatgc 1540
  H S I L K A T N Y N 175
acatgtctgcctccctgcccaccccctcg (...) acatccaatcatgaccatac 1780
acctgactactaccacgatgactacgaggtggttatactttttccttttgttcttccag 1840
GCCACACGTGCTCTACACTTAGCCTCTATCCATGGCTACCTGGGCATCGTGGAGCTTTTGG 1900
G H T C L H L A S I H G Y L G I V E L L 195
TGTCTTGGGTGCTGATGTCAATGCTCAGgtacttacgtcatgcaactgatgtcaatcqa 1960

```

D VRAI Même chose que pour l'exon 4 mais avec l'exon 5 :


```

E Q G C L A S V G V L T Q S C T T P H L 165
CACTCCATCCTGAAGCTACCAACTACAATGtaaacgtcatgcattccactgactgatgc 1540
H S I L K A T N Y N 175
acatgtctgcctccctgccctgccccaccctcg(...)acatccaatcatgaccatac 1780
acctgactactaccacgatgactacgaggtggttatactttttccttttgttctccag 1840
G H T C L H L A S I H G Y L G I V E L L 195
gtacttacgtcatgcaactgatgtcaatcga 1960
V S L G A D V N A Q 205
attggtgcttctgccccgacgcactgagtcagg(...)gcaactgacctcatcactac 2250
catcactcaacggatcatcaccgatccaccataagcatctcaaattccttttggtttcat 2310
ctgtccattgacaagGAGCCCTGTAATCACCTCGCAGTGGACCTGCAAAATCCTGACCTG 2370
E P C N H L A V D L Q N P D L 220
GTGTCACCTCTGTTGAAGTGTGGGGCTGATGTCAACAGAGTTACCTACCAGGGCTATTCT 2430
V S L L L K C G A D V N R V T Y Q G Y S 240
CCCTACCAGCTCACCTGGGGCCGCCAAGCACCCGGATACAGCAGCAGCTGGGCCAGCTG 2490

```

On remarque que cela provoquera aussi un décalage du cadre de lecture, car $+1+3=4$ (différent d'un multiple de 3), et modifiera donc les AA finaux et provoquera sûrement un stop prématuré (ça fera AAT/GGA/GCC/CTG/TAA...).

E FAUX Cette fois ce sera un skipping des exons 4 et 5 simultanément et non l'un indépendamment de l'autre, c'est-à-dire qu'au lieu de sauter UN exon on va en sauter DEUX ! Donc on pourrait se dire que vu que le 4 et le 5 provoquent indépendamment un décalage, les deux simultanément en provoqueraient sûrement un aussi, mais non. Il faut vérifier de nouveau avec la même méthode en regarde si le cadre de lecture se suit entre les exons bordant le 4 et le 5, donc on va regarder la fin de l'exon 3 et le début de l'exon 6 :

```

ataaagtggcggagcccggggctggggatcgtcggatcctaccctgatcatgcacatcgtc 720
ctgcacgcaccttacgattcca(...)accctgggggtttttccctctcttccccacag 920
GTTCTGCACTTGGCCATCATCCATGAAGAAAAGGCACTGACCATGGAAGTGATCCGCCA 980
F L H L A I I H E E K A L T M E V I R Q 96
GGTGAAGGGAGACCTGGCCTTCCCAACTCCAGAACAACCTGCAGCAGgtactcgaatt 1040
V K G D L A F L N F Q N N L Q Q 112
ccgcgccgttgcctgccccgagttctctctgacc(...)acctgaatccttactacgat 1240
cgaactacgttcaatccaacggcattggcttactttctctggtctctcttgcattcgttg 1300
actgactgacctgacctggac 1360
T P N Q P E I A E A L L G 125
A G C D P E L R D F R G N T P L H L A C 1420
E O G C L A S V G V L T Q S C T T P H L 165
gtaacgtcatgcattccactgactgatgc 1540
H S I L K A T N Y N 175
acatgtctgcctccctgccctgccccaccctcg(...)acatccaatcatgaccatac 1780
acctgactactaccacgatgactacgaggtggttatactttttccttttgttctccag 1840
G H T C L H L A S I H G Y L G I V E L L 195
gtacttacgtcatgcaactgatgtcaatcga 1960
V S L G A D V N A Q 205
attggtgcttctgccccgacgcactgagtcagg(...)gcaactgacctcatcactac 2250
catcactcaacggatcatcaccgatccaccataagcatctcaaattccttttggtttcat 2310
ctgtccattgacaagGAGCCCTGTAATCACCTCGCAGTGGACCTGCAAAATCCTGACCTG 2370
E P C N H L A V D L Q N P D L 220
GTGTCACCTCTGTTGAAGTGTGGGGCTGATGTCAACAGAGTTACCTACCAGGGCTATTCT 2430
V S L L L K C G A D V N R V T Y Q G Y S 240
CCCTACCAGCTCACCTGGGGCCGCCAAGCACCCGGATACAGCAGCAGCTGGGCCAGCTG 2490
P Y Q L T W G R P S T R I Q Q Q L G Q L 260

```

$+3+3= 6$ (multiple de 3) donc le cadre de lecture est conservé !

Énoncé de la question 3 :

Chez une patiente de 20 ans, Soumaya, on suspecte un cancer de la médullaire de la thyroïde d'origine génétique. Vous savez que TP53 est souvent impliqué dans ce type de cancer, donc vous effectuez un séquençage Sanger. Un séquençage correspond à une technique permettant de déterminer l'ordre des nucléotides sur un brin d'ADN. Ici on a réalisé un Séquençage Sanger qui nous

a permis de connaître la séquence de nucléotides d'une partie du gène TP53 de Soumaya. Voici les résultats obtenus :

	Séquence commençant à partir du nucléotide g.936
Individu Sain	<u>5' CATCATCCATGAAGAAAAGGC 3'</u>
Patiente Soumaya	<u>5' CATCATCCTATGAAGAAAAGG 3'</u>

Question 3 – L'altération moléculaire retrouvée dans le séquençage Sanger chez Soumaya : (*)**

- A. Est un SNV.
- B. Se situe dans l'exon 3 du gène TP53.
- C. Provoque un décalage du cadre de lecture lors de la traduction et conduit à une protéine tronquée.
- D. Conduit probablement à une protéine p53 ne pouvant plus se lier à l'ADN.
- E. Conduit probablement à une protéine p53 ne pouvant plus lier d'autres protéines.

A FAUX On détecte une insertion d'un T. Or un SNV est une modification d'un nucléotide initialement présent. Ici, aucun nucléotide n'est modifié, il y a seulement une insertion d'un nucléotide en plus.

Individu Sain	<u>5' CATCATCCATGAAGAAAAGGC 3'</u>
Patiente Soumaya	<u>5' CATCATCC<u>T</u>ATGAAGAAAAGG 3'</u>

B VRAI On vous dit dans l'énoncé que notre séquence affichée ci-dessus commence au nt 936 de l'ADNg, on remarque que l'insertion se situe dans l'exon 3.

On a surligné en **vert** la séquence de l'individu sain.

Séquence 2:

```

gatgtggggtgaaaagtactacctgtcaaggtttggttaccctcctgtaaatggtgta 60
agtgactgaccccagtggtatcctgtgacatgtaacagccaggagtgttaagcgttcaag 120
ACTCCACTCCACTTGGCTGTGATCACCAACgtctggtctgactggcttgaggaaattcg 180
ctgtattccttgggtgcatggtttctggtctgactggcttgaaaattcccggagcctgacc 240
cggccccagagaaatccccagccagcgtttatagggcgccgcgggcgctgcaagagccc 300
tcggcgtgcaattccttgggtgcatggttagAGCGCCCAGCGAGGAAGCAGCGCGCAGCC 360
CGCGGCCAGCGCACCCGACGACGCGCCGCGAGCTCGTCCGCGCCATGTTCCAGGCGGCC 420
                                     M F Q A A
GAGCGCCCCAGGAGTGGGCCATGGAGGGCCCCGCGACGGGCTGAAGAAGGAGCGGCTA 480
E R P Q E W A M E G P R D G L K K E R L 25
CTGGACGACCCACGACAGCGGCTGGACTCCATGAAAGACGAGGAGTACGAGCAGATG 540
L D D R H D S G L D S M K D E E Y E Q M 45
GTCAAGGAGCTGCAGGAGATCCGCCTCGAGCCGAGGAGTGCCGCGGGCTCGGAGCCC 600
V K E L Q E I R L E P Q E V P R G S E P 65
TGAAGCAGCAGCTACCGAGGACGGGACTCgtgcctaatactgactacgtggaactgta 660
W K Q Q L T E D G D S 75
ataagtggcggaggccgggctgggggtcgtcggatcctaccctgatcatcgcatcgtc 720
ctgcatcgaccttacggattcca (...) acccctggggttttccctctcttccccacag 920
GTTCTGCACTTGGCCATCATCCATGAAGAAAAGGCACTGACCATGGAAGTGTATCCGCCA 980
P T H T A T T H P F K A T P M P V T D Q 96

```

On a surligné en **rouge** l'insertion d'un T dans la séquence de Soumaya.

Séquence 2:

```
gatgtggggtgaaaagtactacctgtcaaggtttgtgttacccctcctgtaaatgggtgta 60
agtgactgaccccagtggtatcctgtgacatgtaacagccagagtggttaagcgttcaag 120
ACTCCACTCCACTTGGCTGTGATCACC AACGtctggtctgactggcttggaggaaattcg 180
ctgtattccttgggtgcatggtttctggtctgactggcttggaaattccccgagcctgacc 240
ccgcccagagaaatccccagccagcgtttatagggcgccggcgccgctgcagagccc 300
tcggcgctgcaattccttgggtgcatggttagAGCGCCCCAGCGAGGAAGCAGCGCGCAGCC 360
CGCGGCCACGCGCACCCGACGAGCGCCCGCAGCTCGTCCGCGCCATGTTCCAGGCGGCC 420
M F Q A A 5
GAGCGCCCCAGGAGTGGGCCATGGAGGGCCCCCGCGACGGGCTGAAGAAGGAGCGGCTA 480
E R P Q E W A M E G P R D G L K K E R L 25
CTGGACGACCGCCACGACAGCGGCCTGGACTCCATGAAAGACGAGGAGTACGAGCAGATG 540
L D D R H D S G L D S M K D E E Y E Q M 45
GTCAAGGAGCTGCAGGAGATCCGCCTCGAGCCGAGGAGGTGCCGCGCGGCTCGGAGCCC 600
V K E L Q E I R L E P Q E V P R G S E P 65
TGAAGCAGCAGCTCACCGAGGACGGGGACTCgtgcctaatactacgatggaactgga 660
W K Q Q L T E D G D S 75
ataagtggcgaggccccggggtcggggtcgtcggatcctaccctgatcatcgcatcgtc 720
ctgcatcgaccttacgattcca (...) accctggggttttccctctcttccccacag 920
GTTCTGCACTTGGCCATCATCCATGAAGAAAAGGCCTGACCATGGAAGTGATCCGCCA 980
```

C VRAI L'insertion d'un nucléotide (donc d'un nombre de nucléotide qui n'est pas un multiple de 3) provoque un décalage du cadre de lecture. On observe alors un codon stop dans le nouveau cadre de lecture. Ce codon stop prématuré conduit à une protéine tronquée (plus petite).

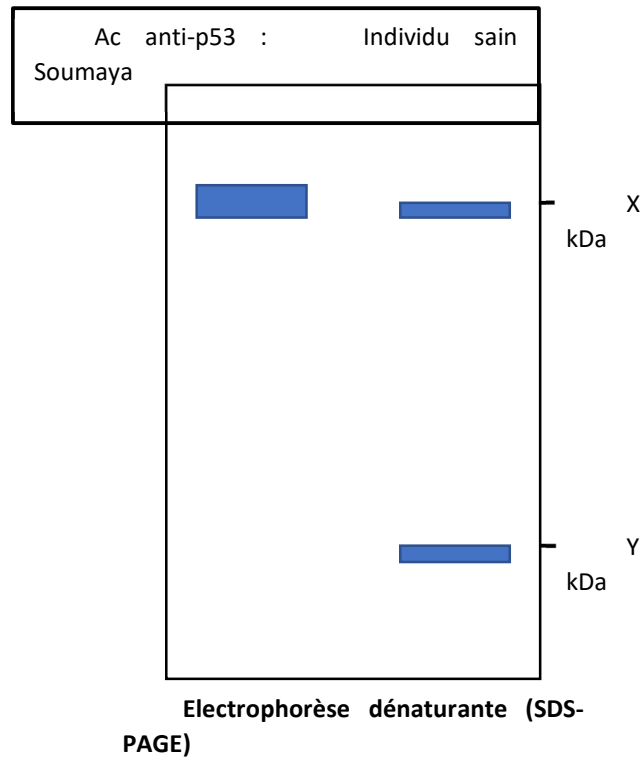
```
gatgtggggtgaaaagtactacctgtcaaggtttgtgttacccctcctgtaaatgggtgta 60
agtgactgaccccagtggtatcctgtgacatgtaacagccagagtggttaagcgttcaag 120
ACTCCACTCCACTTGGCTGTGATCACC AACGtctggtctgactggcttggaggaaattcg 180
ctgtattccttgggtgcatggtttctggtctgactggcttggaaattccccgagcctgacc 240
ccgcccagagaaatccccagccagcgtttatagggcgccggcgccgctgcagagccc 300
tcggcgctgcaattccttgggtgcatggttagAGCGCCCCAGCGAGGAAGCAGCGCGCAGCC 360
CGCGGCCACGCGCACCCGACGAGCGCCCGCAGCTCGTCCGCGCCATGTTCCAGGCGGCC 420
M F Q A A 5
GAGCGCCCCAGGAGTGGGCCATGGAGGGCCCCCGCGACGGGCTGAAGAAGGAGCGGCTA 480
E R P Q E W A M E G P R D G L K K E R L 25
CTGGACGACCGCCACGACAGCGGCCTGGACTCCATGAAAGACGAGGAGTACGAGCAGATG 540
L D D R H D S G L D S M K D E E Y E Q M 45
GTCAAGGAGCTGCAGGAGATCCGCCTCGAGCCGAGGAGGTGCCGCGCGGCTCGGAGCCC 600
V K E L Q E I R L E P Q E V P R G S E P 65
TGAAGCAGCAGCTCACCGAGGACGGGGACTCgtgcctaatactacgatggaactgga 660
W K Q Q L T E D G D S 75
ataagtggcgaggccccggggtcggggtcgtcggatcctaccctgatcatcgcatcgtc 720
ctgcatcgaccttacgattcca (...) accctggggttttccctctcttccccacag 920
GTTCTGCACTTGGCCATCATCCATGAAGAAAAGGCCTGACCATGGAAGTGATCCGCCA 980
F L H L A I I L * 83
GGTGAAGGGAGACCTGGCCTTCTCAACTTCCAGAACAACCTGCAGCAGgtactctaatt 1040
```

D FAUX Dans l'énoncé 1, on vous dit que l'exon 2 code pour le domaine de liaison à l'ADN (DBD) de p53. Sachant que l'insertion décale le cadre de lecture à partir de l'exon 3, l'exon 2 « sain » est donc conservé, il est donc fort probable que la protéine p53 de Soumaya puisse toujours lier l'ADN.

E VRAI D'après l'énoncé 1, on vous dit que l'exon 5 code pour le domaine de transactivation (qui permet de lier p53 à d'autres protéines). Sachant que l'insertion cause un codon STOP prématuré dans l'exon 3, alors les exons 4 à 7 ne codent plus pour des AA. Il est alors très probable que la protéine p53 de Soumaya ne puisse plus lier d'autres protéines.

Énoncé de la question 4 :

Pour vérifier les répercussions sur la protéine p53 de Soumaya, on a effectué un Western Blot avec des anticorps dirigés contre un seul épitope de p53.



Question 4 – A propos du Western Blot, on peut dire que : (*)**

- A. La protéine p53 saine a un poids moléculaire de 33,55kDa (à 0,5kDa près).
- B. La protéine p53 mutée de Soumaya a un poids moléculaire de 9,13kDa (à 0,5kDa près).
- C. L'altération retrouvée chez Soumaya est présente de manière hétérozygote.
- D. L'altération retrouvée chez Soumaya est présente de manière homozygote.
- E. Les anticorps dirigés contre p53 sont monoclonaux.

A VRAI La protéine saine de p53 possède 305 acides aminés et comme 1 AA = 0,11kDa alors la protéine p53 a un poids moléculaire de $0,11 \times 305 = 33,55$ kDa.

B VRAI La protéine p53 est tronquée à cause d'un codon stop prématuré, on regarde combien il y a d'acides aminés avant le codon STOP prématuré.


```

gatgtggggtgaaaagtactacctgtcaaggttgggttacccctcctgtaaatgggtga 60
agtgactgacccagtggtatcctgtgacatgtaacagccaggagtgttaagcgttcaag 120
ACTCCACTCCACTTGGCTGTGATCACCAACgctctggctgactggcttggaggaaattcg 180
ctgtattccttgggtgcatggtttctggctgactggcttggaaattccccgagcctgacc 240
ccgcccagagaaatccccagccagcgtttatagggcgccgaggcgctgcagagccc 300
tcggcctgcaattccttgggtgcatggttagAGCGCCCAGCGAGGAAGCAGCGCGCAGCC 360
CGCGGCCAGCGCACCCGAGCAGCGCCCGAGCTCGTCCGCGCCATGTTCCAGCGGGC 420
                                     M F Q A A 5
GAGCGCCCCAGGAGTGGGCTATGGAGGGCCCCCGCAGCGGCTGAAGAAGGAGCGGCTA 480
E R P Q E W A M E G P R D G L K K E R L 25
CTGGACGACCCACGACAGCGGCTGGACTCCATGAAAGACGAGGAGTACGAGCAGATG 540
L D D R H D S G L D S M K D E E Y E Q M 45
GTCAAGGAGCTGCAGGAGATCCGCCTCGAGCCGAGGAGGTGCCGCGGGCTCGGAGCCC 600
V K E L Q E I R L E P Q E V P R G S E P 65
TGGAAGCAGCAGCTCACCGAGGACGGGGACTCgtgcctaagtactacgatggaactgga 660
W K Q Q L T E D G D S 75
ataagtggcggagggccgggctgggggtcgtcggatcctaccctgatcatcgcatcgtc 720
ctgcatcgaccttacggattcca (...) acccctgggggtttttccctctcttccccacag 920
GTTCTGCACCTGGCCATCATCCATGAAGAAAAGGCACTGACCATGGAAGTGATCCGCCA 980
F L H L A I I L * 83
GGTGAAGGGAGACCTGGCCTCCTCAACTTCCAGAACAACCTGCAGCAGgtactctaatt 1040

```

On remarque alors que la protéine mutée possède 83 acides aminés et donc a un poids moléculaire de $83 \times 0,11 = 9,13$ kDa.

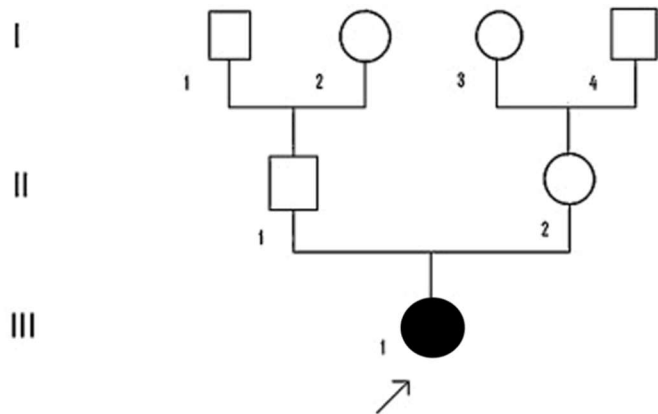
C VRAI Sur le Western Blot on remarque la présence de 2 bandes diminuées de moitié (par rapport à la bande de gauche) sur le profil de Soumaya dont une bande à la même hauteur que la bande du profil sain. Cela signifie qu'on retrouve des protéines saines chez Soumaya et des protéines mutées. C'est ce qu'on retrouve dans le cas de maladies hétérozygotes. Si la maladie avait été homozygote, on aurait seulement aperçu 1 bande sur le profil de Soumaya (celle de la protéine mutée/plus petite). On a un allèle du gène qui est muté et l'allèle de l'autre gène qui est non muté, ce qui fait que lors de la transcription/traduction, on aura deux versions différentes de la protéine p53.

D FAUX Cf. item C.

E VRAI Dans l'énoncé, on apprend que les anticorps utilisés pour le Western Blot sont dirigés contre UN SEUL épitope de p53, ils sont donc monoclonaux

Énoncé de la question 5 :

Après plusieurs analyses sur la médullaire de la thyroïde, nous pouvons confirmer que Soumaya est atteinte du cancer de la médullaire de la thyroïde à cause de l'altération évoquée dans les questions précédentes. Nous avons décidé d'établir un arbre généalogique de la famille de Soumaya pour prévenir d'éventuelles personnes à risque dans sa famille. La flèche indique Soumaya (en III.1).



Question 5 – Il s’agit d’une altération moléculaire : ()**

- A. Qui cause la maladie seulement si elle est présente sur les deux allèles (récessive).
- B. Qui cause la maladie si elle présente sur un allèle au minimum (dominante).
- C. Autosomique.
- D. Liée à l’X.
- E. Qui est apparue chez Soumaya sans que celle-ci provienne de ses parents (De novo).

Tout d’abord on sait que le gène TP53 se situe sur le chromosome 17, il s’agit donc d’une altération moléculaire autosomique (**C VRAI** et **D FAUX**). Grâce à l’exercice précédent, on sait que Soumaya est porteuse hétérozygote de l’altération, et dans l’énoncé on nous confirme qu’elle est malade à cause de cette altération. On comprend donc qu’il suffit d’un allèle muté pour être malade, donc c’est une maladie à expression dominante (notion que vous reverrez dans le chapitre anomalies du génome ne vous inquiétez pas) (**B VRAI, A FAUX**). Ensuite vous ne voyez qu’aucun des parents de Soumaya n’est atteint, alors que cette maladie est dominante (donc qu’il suffit d’un allèle muté pour être atteint), donc ses parents ne sont pas porteurs de l’altération. En gros, chacun transmet un allèle à Soumaya, si jamais l’altération provenait d’un des parents, cela reviendrait à dire que l’un des parents possède l’altération sur l’un de ses deux allèles et donc qu’il est malade (maladie dominante), ce n’est pas le cas ici (ils ne sont pas malades d’après la légende). Donc l’apparition de l’altération est survenue chez Soumaya au cours de sa vie (de novo). (**E VRAI**)

Énoncé commun aux questions 1 à 4 (VOIR SEQUENCES FBN1 EN ANNEXE):

Voici le gène FBN1. Ce gène code pour la fibrilline, une protéine représentant le principal composant des micro-fibrilles de la matrice extracellulaire du derme et qui participe à la formation des fibres élastiques. Des altérations génomiques de FBN1 sont impliquées dans le syndrome de Marfan. Vous recevez Paul, un jeune homme de 18 ans en consultation. Paul craint d'être atteint du syndrome de Marfan, en effet son père est atteint par ce syndrome. Vous savez par ailleurs que la mère de Paul n'est pas atteinte.

Vous étudiez alors le gène FBN1 de Paul et identifiez certains SNV caractéristiques.

Après étude de ceux-ci vous concluez que Paul est lui aussi atteint du syndrome de Marfan.

Les séquences 1 et 2 représente les séquences d'ADNg et d'ADNc du gène FBN1 de Paul. À l'aide de l'énoncé, de vos connaissances et du diagramme du code génétique en annexe, quelle(s) affirmation(s) est / sont vraie(s) ?

Question 1 - FBN1 (*):

Le dernier exon de ce gène est long (à 2 paires de bases près) de :

- A. 240 paires de bases.
- B. 245 paires de bases.
- C. 235 paires de bases.
- D. 277 paires de bases.
- E. 282 paires de bases.

Pour cette question, il faut bien comprendre que le dernier exon ne s'arrête pas nécessairement au codon STOP ou au dernier acide aminé mais qu'il comporte l'ensemble situé entre l'avant dernier intron et le dernier intron (en gros la dernière séquence continue de majuscule de la séquence d'ADNg).

```
ATCTCAGCGTCCAGACATGGATGTTACACACCCGATGTTCTACTCCTCGGACGGGTGGAT      269
 I S A S R H G C Y T P D V L L L G R V D
 CCGTCTGTATCTTGACGAGgtccaaacgccccggtgtagtagaaatgctcagtaccg
 R L C I L H E      276
 gcaaaccaacgactacaaaatgagaccgtcaatgtatgggaaatgcgcgtaccttccttg
 ttaccggggcatacaggtcagctcaccgtgcagCTTCCGACGAGAAGAGATGCGAGAATC
                               L P T R R D A R I      285
TGCGCATACCGCCACATAGCCATAGGTGCGACTCTTGCCGACGAGTACGCACCTGCTTTT
 C A Y R P H S H R S T L A D E Y A P A F      305
TGGCAGACTACCCAACGCCTGGATCAAGAgtgactcggtagcccttatattcggaagt
 W Q T T P T P G S R      315
ccactgcttacttaatcacgaaagcaaaaagcaatctccgatgctttgcgagcgatcagcg
 cgaatcattaccataaggcaccatcgctcctcaggcaaccaggggcccctagctcgaca
 gttgtgtgatgatattcctgcgcggtctocactctgttcagTTTACCAAAACCATC      320
                               F T K T I
CAGCCATACCTTGAGATGCGTTCGGTTGGGACCTGCATGCAGGCAGTGGTGTGGCTACCGG
 Q P Y L E M R R L G P A C R Q W C G Y R      340
GAGACAACCTGACCGACTGGCCAAGAATCGGTATCGCTACGGATACACTATGAAACGAGGA
 E T T D R L A K N R Y R Y G Y T M K R G      360
GTGCATCGGTCGCGGCACGGGAATTATACTGTAACGCGACGAACATCCAGTAGATACT
 V H R S A A R E L Y C N A T N I P V D T      380
AATCCCCCTGGTCTCTAGGAAAGCGCGGTCTCATGAACCTTACTAAGCTGCTCCCACC
 N P P W C S R K A R S H E P Y *
AGACGCAGCGAGACTACTTGATctgggaagcacataaagggatgtctggctctcccccat
```

Pour simplifier un calcul qui pourrait potentiellement être un peu laborieux, vous pouvez vous dire que chaque ligne pleine de bases contient 60 paires de bases. On a ici 4 lignes pleines soit 240 pbs. À cela on ajoute l'équivalent de 5 x 3 paires de bases pour la première ligne du dernier exon (car elle commence par 5 AA, codé chacun par un triplet de 3 paires de bases). Enfin, on dénombre 22 pbs sur la dernière ligne.

Le dernier exon de FBN1 contient donc $240 + 15 + 22 = 277$ pbs.

A FAUX

B FAUX

C FAUX

D VRAI

E FAUX

Question 2 - FBN1 (*) :

A propos du gène FBN1 :

- A. Contient une partie codante dans son premier exon.
- B. Contient 6 exons.
- C. Une délétion de l'exon 3 entraîne un décalage du cadre de lecture.
- D. Une délétion de l'exon 5 entraîne un décalage du cadre de lecture.
- E. Code pour une protéine de 395 acides nucléiques.

A FAUX Attention, le premier exon du gène FBN1 de Paul ne contient pas de partie codante. Pour différencier ces exons les uns des autres, il faut se servir de la séquence d'ADNg, qui permet de les distinguer puisqu'ils sont séparés de bordures introniques en minuscules. Rappelez-vous aussi que même si le codon STOP se situe toujours dans le dernier exon, le codon START ne se retrouve pas nécessairement dans le premier exon (et d'ailleurs de façon plus habituelle dans le deuxième exon). Ci-dessous le premier exon en bleu, le deuxième (codant la première Méthionine) en jaune.

```

GGCATTTTTGAGTATCACGCTATGTATTATAGCCGCAGACCGTTTGCTatcccgtaccaa
tggatatctgcacccaaaacatccgagtgctggagtgactctacatgaaAGAGACTATAAA
CAGAAAATATCCAGCAACGGGTATTTTGTAGTGATGTCCACTGACTTGGTAGATCGAGCA
                                     M S T D L V D R A      9
CAAGTGCCCGGGCGATCCTCCGTGCGAACTGGAATGACTTACCTCCCAGCTTCTGTGGCG
Q V P G R S S V R T G M T Y L P A S V A      29
ACACTTAAgtggaaacggtcggagggcacttgtcttcccctgggtcattacatcgggggg
T L K                                          32
ccgaaaaatccttgatatttataaatatgcccagtcctataagagactgaaaaggccgaca
ggcggacacagGACGTGGGAAGTCGCTGCAGACGCAGAACCCCGGGAAAGAGATGTTC
      T L G S R C R R R T P G K R C S      48
GGAATAACTGTGCTCTCAAAAAACGCAACAGGTGATGTAATTGCAACACCGTCTCACATT
G I T V L S K N A T G D V I A T P S H I      68

```

B VRAI Voir schéma ci-dessous.

GGCATTTTTGGAGTATCACGCTATGTATTATAGCCGAGACCGTTTGCTatcccgtaccaa
tggatctctgcaccccaaacatccgagtgctggagtgactctacatgaaAGAGACTATAAA
CAGAAAATATCCAGCAACGGGTATTTTGTAGTGTCCACTGACTTGGTAGATCGAGCA
M S T D L V D R A 9
CAAGTGCCCGGGCGATCCTCCGTGCGAAGTGAATGACTTACCTCCAGCTTCTGTGGCG
Q V P G R S S V R T G M T Y L P A S V A 29
ACACTTAAgtggaaccggctcggaggcacttgtcttcccctgggtcattacatcggggg
T L K 32
ccgaaaaatccttgatatttataaatatgccagtcctataagagactgaaaaggccgaca
ggcggacacagGACGTTGGGAAGTCGCTGCAGACGCAGAACCCTCCGGAAAGAGATGTTC
T L G S R C R R R T P G K R C S 48
GGAATAACTGTGCTCTCAAAAAACGCAACAGGTGATGTAATTGCAACACCGTCTCACATT
G I T V L S K N A T G D V I A T P S H I 68
CCCCCGTACATCTATCAAGCTTCGAAGCGCCAGTTCGGCATGGGTCAACACCTCAG
P P S H L S S F E A R Q F R M G Q Q P Q 88
ATAAGCGGACTGACCGACTCACACCCAGAGAACGGACCCTTCAGCTATGGGGCCGGTAC
I S G L T D S H P E N G P F S Y G A G Y 108
GGATTCGTCGTTAATCCACCCACTTAGACTATGCTTTTACCTCCTAAGGGAGTgtttgc
G F V V N P T H L D Y A F T S L R E 126
agaagtgcggtcggcttttatgtcgggctccaacaatgcgggccgaccacagTGTCTTCT
L S S 129
TGGGTCTGTTTTATGAACTCTCCCATTTGTGTGTATCACCGCTGTTGATGTGCAGTAC
W V C F M N S P H L C V S P A V D V Q Y 149
TGTTTGGCGCAGAAATATCAAAACCGTTTGGGCCTTCTACGCACTGTTTGGTATGTAGCG
C L A Q N Y Q T V W A F L R T V W Y V A 169
TGAAGCCTGGGCTGACGTACATGCAATCTAATAAGGGGAGAGCGCATCCCAAGGTAGC
W K P G L T Y M Q S N K G R A H P K V A 189
GTAATCGGAAACCGCGTTCGGTAAATAGCAGCCTTTGGGACTTCCAAGGTGTGGGGCTT
V I G N R R R L I A R F W D F Q G V G L 209
TCGGAGAGTCGTACTCCGAATCAGCCTTAGGCCAACCCTTCGTCCTCAAGGATGGATAGGT
S E S R T S E S A L G E P S S Q G W I G 229
TACATCATCGCCCTAGTCTACACATGTTTGTGACAAATCTGGGTGGATTGTGGATATA
Y I I A L V Y T C L M T N L G W I V D I 249
ATCTCAGCGTCCAGACATGGATGTTACACACCCGATGTTTCTACTCTCGGACGGGTGGAT
I S A S R H G C Y T P D V L L L G R V D 269
CGGCTCTGTATCTTGCACGAGgtccaaacgccccggtgttagtagaaatgctcagtaccg
R L C I L H E 276
gcaaaccaacgactacaaaatgagaccgtcaatgtatgggaaatgcgcgtaaccttcttctg
ttaccgggcatacaggtcagtcctaccgtgcagCTTCCGACGAGAAGAGATGCGAGAATC
L P T R R D A R I 285
TGCGCATAACGCCACATAGCCATAGGTCGACTCTTGCCGACGAGTACGCACTGCTTTT
C A Y R P H S H R S T L A D E Y A P A F 305
TGGCAGACTACACCAACGCCTGGATCAAGAgtgactcggtagcccttatattcggcaagt
W Q T T P T P G S R 315
ccactgcttacttaacacgaaagcaaaagcaatctccgatgctttgcgagcgatcagcg
cgaatcattaccataaggccaccatcgctcctcaggcaaccaggggcccctagctcgaca
gttgtgtgatatgatattcctgcgcggtctccactcttgttcagTTTACCAAAACCATC
F T K T I 320
CAGCCATACCTTGAGATGCGTCGGTGGGACCTGCATGCAGGCAGTGGTGTGGCTACCGG
Q P Y L E M R R L G P A C R Q W C G Y R 340
GAGACAACGACCGACTGGCCAAGAATCGGTATCGCTACGGATACACTATGAAACGAGGA
E T T D R L A K N R Y R Y G Y T M K R G 360
GTGCATCGGTCTGCGGCACGGGAATTATACTGTAACGCGACGAACATTCAGTAGATACT
V H R S A A R E L Y C N A T N I P V D T 380
AATCCCCCTGGTGTCTAGGAAAGCGCGGTCTCATGAACCTTACTAAGCTGCTCCCACC
N P P W C S R K A R S H E P Y *
AGACGCAGCGAGACTACTTGATctgggaagcacataaagggtgtctggtctccccat

C VRAI Il faut bien comprendre que souvent des **codons** (3 nucléotides) vont se retrouver sur la **jonction exon/exon**. C'est-à-dire que les premières bases d'un codon vont se trouver sur un exon et celles manquantes au début de l'exon suivant. Ainsi, après épissage des introns, le codon se retrouve assemblé et pourra être lu de manière normale. C'est quelque chose de tout à fait normal. Ainsi, lorsqu'on cherche à savoir si la délétion d'un ou plusieurs exon entraîne un décalage du cadre de lecture, il faut rechercher au niveau des débuts et fin de chaque exon où se situe la base sur le codon correspondant.

Prenons pour exemple l'exon 3 : on remarque que le premier acide aminé de l'exon 3 est une Thréonine (T) codé par le codon ACG sus-jacent (en ayant défini au préalable grâce

à la première méthionine que le codon codant l'acide aminé se situe sur la ligne du dessus, et que l'acide aminé est aligné sur la seconde base du codon).

On remarque néanmoins une base seule avant ce codon, il s'agit d'un G. Ce G n'est en réalité pas isolé mais correspond à la suite des deux dernières bases de l'exon 2, à savoir A et A. Après épissage, on lie les jonctions exons-exons et on retrouve un codon AAG qui code pour une Lysine (K) (représenté sous la deuxième base du codon, à la fin de l'exon 2).

Pour résoudre ce type d'exercice rapidement, on peut ainsi réaliser un tableau où l'on note sur quelle base d'un codon (première, deuxième ou troisième) commence / se termine chaque exon. Dans le cas du gène FBN1 de Paul on aurait un tableau comme-çeci.

Exon 2			Exon 3			Exon 4			Exon 5			Exon 6		
Début	/	fin	Début	/	fin	Début	/	fin	début	/	fin	début	/	fin
		+2	+3		+1	+2		+3	+1		+3	+1		

Ici, un +1 au niveau de la fin d'un exon signifie que la première base d'un codon se situe à la fin de cet exon et que les deux suivantes, logiquement, se situent sur l'exon suivant.

Ainsi, on remarque un enchaînement logique en cadre de lecture ou après épissage, les bases numéro 2 et 3 d'un codon se regroupe pour reformer un codon.

En étudiant ce tableau, on peut répondre aux questions, une excision de l'exon 3 entraîne bien un décalage du cadre de lecture car en faisant cela et après épissage, on passera de l'exon 2 à 4 et donc de la deuxième base d'un codon à la deuxième base d'un autre codon. La première base de l'exon 4 sera alors lu en fonction du cadre de lecture précédent, et sera lu comme la dernière base du codon précédent (et non plus comme la deuxième). On observera alors un décalage du cadre de lecture.

D FAUX Voir ci-dessus, si on délète l'exon 5 alors après épissage on se retrouvera avec une continuité entre les exons 4 et 6. L'exon 4 se termine en +3 tandis que l'exon 6 commence en +1, le cadre de lecture est ainsi respecté même après déletion de l'exon 4.

E FAUX Le gène FBN1 code pour une protéine de 395 acides AMINÉS.

Question 3 : SNV (*)

Soit le SNV c.405 C>G, l'acide aminé obtenu après mutation :

- A. Est impliqué dans des réactions de transaminations.
- B. Est un acide aminé essentiel.
- C. Peut subir une glycosylation.
- D. Peut subir une méthylation.
- E. Possède une charge positive au pH isoélectrique.

Pour étudier la mutation **c.405 C>G**, il faut d'abord retrouver où se situe celle-ci sur notre séquence. Pour cela, il convient de compter le nombre de paires de bases portées par notre région 5'UTR. Elle fait 93 pbs. Ensuite, on ajoute 405 à 93 pour retrouver la position de la mutation sur notre séquence grâce aux numéros du flanc droit. Il faut donc chercher en position 498. Notre codon AAC devient alors AAG et code donc pour une Lysine. On peut désormais répondre aux questions.

A FAUX

B VRAI

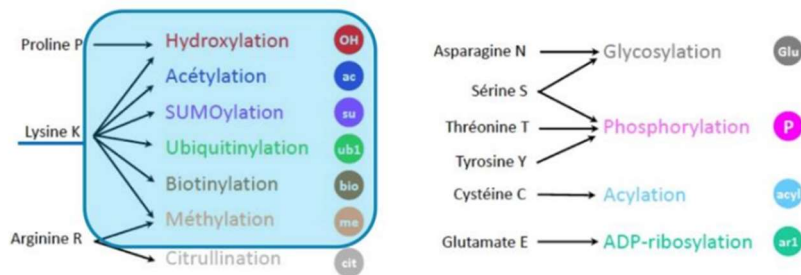
C FAUX

D VRAI

E FAUX

1 Acides aminés : Modifications

Voir cours
"Régulation de
l'expression
des gènes"



Lysine = AA très réactif

Question 4 : Transmission (**)

Voici un tableau regroupant les mutations de FBN1 portées par les parents de Paul et lui-même, comparé à la séquence de référence du gène FBN1. Ces mutations sont toutes présentes à l'état hétérozygote chez les individus porteurs de celles-ci.

À l'aide de ce tableau, de l'énoncé et des documents annexes, quelle(s) affirmation(s) est/ sont vraie(s) ?

Mère	c.325G>A	c.699A>C
Père	c.501T>G	c.697G>T
Paul	c.501T>G	c.699A>C

- A. Le variant c. 699C>A est probablement bénin.
- B. Le variant c.1188C>A est probablement bénin.
- C. Le variant c.325G>A dont est atteinte la mère de Paul entraîne la création d'un acide aminé basique.
- D. La transmission de cette maladie semble ici récessive.
- E. Le gène FBN1 pourrait être localisé sur le chromosome X.

A ANNULÉ Chez Paul, si l'on compare à la séquence de référence de FBN1, cette mutation est synonyme, un GCC devient un GCA et code aussi pour une Alanine. Par ailleurs, on constate que la mère de Paul, aussi porteuse de cette mutation n'est pas atteinte du syndrome. Tout cela nous laisse penser que cette mutation est probablement bénigne. (Erreur dans l'énoncé : le variant 699 devrait être c.699C>A pour la mère et Paul)

B FAUX Ici on constate que les deux individus porteurs de cette anomalie sont atteints (Paul et son père). Aussi on constate que cette anomalie code pour un codon STOP prématuré. De plus Paul ne possède que deux altérations moléculaires et on sait que la

seconde est SYNONYME et probablement bénigne. Cela nous fait entendre que cette mutation est sans doute pathogénique.

C VRAI Cette mutation entraîne la création d'un codon AGA, qui code pour une Arginine (Arg) , un acide aminé basique.

D FAUX On a montré jusqu'ici que la seule mutation probablement pathogénique portée par Paul est la mutation c.501T>G, On sait avec l'énoncé que toutes les mutations présentées ici sont hétérozygotes. La transmission semble donc ici dominante.

E FAUX Il suffit de constater que le père de Paul lui a transmis l'altération moléculaire c.501T>G, hors Paul étant un homme, son père lui a transmis un Y. La transmission est donc ici autosomique.

Énoncé commun aux questions 1 à 3 (VOIR SEQUENCE ATM EN ANNEXES):

ATM est une protéine se liant à l'ADN qui permet de détecter des dommages doubles brins de l'ADN. Des mutations de ATM ont été retrouvées dans l'ataxie télangiectasie (AT) ou syndrome de Louis-Bar. L'ataxie télangiectasie est une maladie rare, neurodégénérative et immunodépressive.

Vous avez à disposition en annexe les séquences d'ADNc et d'ADNg du gène atm codant pour la protéine ATM ainsi qu'un tableau du code génétique.

Question 1 – ATM (***) :

Vous étudiez différentes altérations moléculaires affectant le gène atm : c.372T>G (V1), c.334_474del (V2), c.604-1G>T (V3), c.712G>A (V4).

Parmi les propositions suivantes concernant le gène atm et ces variants laquelle/lesquelles est/sont vraie(s) ?

- A. Le gène ATM contient 5 exons codants et 1 non codant.
- B. La protéine ATM comporte 286 acides aminés.
- C. La séquence d'ADNc est la séquence 2.
- D. V1 entraîne la formation d'un codon STOP prématuré.
- E. V2 est une petite délétion.

A FAUX Le gène ATM contient 5 exons tous codants. En effet, le codon START ATG se situe dans le premier exon de notre séquence, le premier exon est donc bien codant. Ci-dessous les 5 exons (en majuscule) du gène.

Séquence 1

```
atgtgatgttctactcaggggaccgctcctactggagcacgggTATCCCCAGCCATAAGGTCGGTTC 66
ACCCATFACATACAAAGTCCGCTCGGGCTACAATAGGAGAGCATGAGTAAGGTGTTCCAGTGGGACC 132
M S K V F T W T 8
GGAGCATACTGGAGAACCCTCCGCCCTAAATTCGGCGCAAACCTCGCATGTGACTGGTACGACAGTAGT 198
G A Y W R T F R P N S R K L A C D W Y D S S 30
AATAAGACCCACCAgttgtgatacgttactagcagatagctctggggattaaggcgttgaggcaaaa 264
N K T H Q 35
atgagAGGCGGAACAGCTGGTGGCAAGTATCGGGCCTCCAATTACAGTAAGGAGATGATCGAGACA 330
G G T A G G K Y R A S N Y S K E M I E T 55
CGGAATATCAGCGCAAGTCTCATTCTTTACCGGGCGGCGCAAGGTATTAGATACCGACGTTAC 396
R N I S A S L I P L Y R A A Q G I R Y R R Y 77
AAACCCCTGGGTGAGTACAGAAGTCCAGATAGATATAAGCAGCCATCGTCTTAATATGACAGGAATCT 462
K P W V S T E V Q I D I S S H R S N M Q E S 99
ACCCgtttcggtaaatTTTcaataacttttacttttagcgtggccatacagTGGGCGTGGCTCCGACA 528
T L G V A P T 106
ATTATCCAAGAGTGTATATATAGACATTCGAGCCCAGGTGTGTGCTTACCTTATGTGGGTGGAAT 594
I I Q E C I Y R H S S P G V S L P Y V G V N 128
ACTCCCCCGCATACAACCTCCATGTCGACTCCAGATTTGAGGTTCCCTGACGCTCGACTTAAATGGC 660
T P P A Y N S M S T P D L R F L T L D L N G 150
CCCGACCGGTACGCTACCGGCTTGTACTACTTACCTAGTTTCGATCGGCGATTTGTTCAACTGT 726
P D R S A T G L L Y S Y L V S I G D L F N C 172
GGACGAAGGAAGATGATGTTGTGTGGGATTTATAGAGTCCCCAAGTTAGCTGAAACGCGTGGCGA 792
G R R K M M L C G I Y R V P K L A E T R W R 194
ACGGGACCGTCTGCTCAGTACgttttaggttctcccccaaatatcaactcgggtattataagcggg 858
T G P S A Q Y 201
agctcctagtagCATGGTCCGGGATATGTGACAAACCGACAACCTACCCCTTTTCGTGTGAGCTA 924
H G R G Y C Y K P T T Y P F S C E L 219
AGGCGGGATATTGTTGGTACAGGTACTCGACGGATACCACGAAACTTCATGCGGTTCAACCGGGG 990
R R G Y C W Y R Y S T D T T K L H A V Q P G 241
TCCGCCAGGGATCTCCgtacctttgagtcgaatatatgacagatctatcgtctatatttgaagt 1056
S A R D L L 247
cttcactcgtcaccacaacatgcccggcttcaaatggcagtcatacaccggacgcaagaat 1122

gtgggacccggtgtaaaagcttccagTCAAAGCCTAGCCTACCTCGCCAAAGCATAGCACCTTGACC 1188
Q S L A Y L A K H S T L T 260
SAGCCATGCGT TAGGTACATGTGGAGACCGTGTATGAGTGA TCGGCGGGTACCGGTTGAAACGACTG 1254
E P C V R Y M W R P C M S D R R V P G E R L 282
CAAGAGTCTAGACCTTCCGCCCTAAATTCGGCGCAAAG
Q E F *
```

B FAUX La protéine ATM comporte 285 acides aminés, le codon STOP marqué par un astérisque « * » ne code pas pour un acide aminé.

C VRAI On remarque que la séquence 2 représente la séquence d'ADNc car elle ne comporte que des séquences exoniques, en MAJUSCULE.

D VRAI Avant de répondre, il faut déterminer la taille de notre région 5'UTR. Ici, elle est assez facile à dénombrer puisqu'elle s'étend sur une ligne pleine seulement, dont 5'UTR = 66 pbs. On s'intéresse à la mutation c.372T>G, il faut donc regarder en position 438 sur la numérotation de notre séquence. La mutation porte sur la troisième base d'un codon TAT, qui devient alors un TAG. Il s'agit bien d'un codon STOP.

Séquence 2

```

TATCCCCAGCCATAAGGTCCGTTACCCATTACATACAAAGTGCCTGCGGTACAATAGGAGAGC 66
ATGAGTAAGGTGTTACAGTGGACGGGAGCATACTGGAGAACCTTCCGCCCTAATTCGCGCAAACCTC 132
M S K V F T W T G A Y W R T F R P N S R K L 22
GCATGTGACTGGTACGACAGTAGTAATAAGACCCACCAAGCGGAACAGCTGGTGGCAAGTATCGG 198
A C D W Y D S S N K T H Q G G T A G G K Y R 44
GCCTCCAATTACAGTAAGGAGATGATCGAGACACGGAATATCAGCGCAAGTCTCATTCTCTTTAC 264
A S N Y S K E M I E T R N I S A S L I P L Y 66
CGGGCGGCGCAAGGTATTAGATACCGACGTTACAAACCCCTGGGTGAGTACAGAAGTCCAGATAGAT 330
R A A Q G I R Y R R Y K P W V S T E V Q I D 88
ATAAGCAGCCATCGTCTAATATGCAGGAATCTACCCCTGGGCGTGGCTCCGACAATTATCCAAGAG 396
I S S H R S N M Q E S T L G V A P T I I Q E 110
TGTATATATAGACATTCGAGCCAGGTGTGTCGTTACCTTAATGGGGTGTGAATACTCCCGCCGCA 462
C I Y R H S S P G V S L P Y V G V N T P P A 132
TACAACTCCATGTCGACTCCAGATTTGAGGTTCTGACGCTCGACTTAAATGGCCCCGACCGGTCA 528
Y N S M S T P D L R F L T L D L N G P D R S 154
GCTACCGGCTTGTACTACTTACCTAGTTTCGATCGGCGATTGTTCAACTGTGGACGAAGGAAG 594
A T G L L Y S Y L V S I G D L F N C G R R K 176
ATGATGTTGTGGGATTTATAGAGTCCCAAGTTAGCTGAAACGCGCTGGCGAACGGGACCGTCT 660
M M L C G I Y R V P K L A E T R W R T G P S 198
GCTCAGTACCATGGTCCGGGATATTGCTACAAACCGACAACCTTACCCTTTTCGTGTGAGCTAAG 726
A Q Y H G R G Y C Y K P T T Y P F S C E L R 220
CGGGATATTGTTGGTACAGGTAAGTACGACGGATACCACGAAACTTCATGCGGTTCAACCGGGCTCC 792
R G Y C W Y R Y S T D T T K L H A V Q P G S 242
GCCAGGATCTCCTGCAAAGCCTAGCGTACCTCGCCAAGCATAGCACCTTGACCGAGCCATGCGTT 858
A R D L L Q S L A Y L A K H S T L T E P C V 264
AGGTACATGTGGAGACCGTGTATGATGATCGGCGGGTACCGGGTGAACGACTGCAAGAGTTCTAG 924
R Y M W R P C M S D R R V P G E R L Q E F *

```

Matthieu Lemesre
V1. c.372T>G. codon initial TAT → TAG → codon STOP

E FAUX La limite pour qualifier une insertion/ délétion de « petite » est de 50 paires de bases. Ici, notre délétion s'étend sur 141 paires de bases, il s'agit donc d'une grande délétion.

Question 2 – ATM (***) :

Vous étudiez différentes altérations moléculaires affectant le gène atm : c.372T>G (V1), c._334_474del (V2), c.604-1G>T (V3), c.712G>A (V4).

Parmi les propositions suivantes concernant le gène atm et ces variants laquelle/lesquelles est/sont vraie(s) ?

- A. V2 entraîne un décalage du cadre de lecture.
- B. V3 affecte un site donneur d'épissage.
- C. V3 peut entraîner une délétion de l'exon 4.
- D. V4 crée un site possible d'ubiquitinylation.
- E. V4 est une mutation SYONYME.

A FAUX Cette délétion, qui a lieu en plein dans l'exon 3 du gène ATM, affecte au total 141 paires de bases. Pour retrouver ce nombre, car il est ici impossible de compter avec ses doigts (à part en y passant 15 minutes), il faut se souvenir d'une petite astuce. Pour retrouver le nombre de paires de bases délétées, ici on peut faire (474-334)+1 = 141 (141

étant le multiple de 47 par 3). On délète donc 141 paires de bases, il n'y a donc pas de décalage du cadre de lecture mais une délétion de 47 acides aminés.

B FAUX Attention à la numérotation, une numérotation en c.X+1 ou c.X+2 correspond à un site donneur d'épissage tandis que la numérotation en c.X-1 ou c.X-2 correspond à un site accepteur d'épissage.

C VRAI V3 affecte un site accepteur d'épissage. Ce site accepteur d'épissage ne pourra donc plus être utilisé et il va donc se produire un épissage alternatif qui pourra, si un nouveau site accepteur est utilisé en aval de l'exon 4 (par exemple le site accepteur juste en amont de l'exon 5), entrainer potentiellement une délétion de l'exon 4.

D FAUX Cette mutation (voir sa localisation ci-dessous) touche la première base d'un codon GTT codant pour Valine et entraine la formation d'un codon ATT qui code pour une Isoleucine. C'est donc une mutation FAUX-SENS. Mais l'isoleucine n'est pas un site possible d'ubiquitinylation, ça aurait pu être le cas si l'acide aminé créé était une lysine.

Séquence 2

TATCCCAGCCATAAGGTCGGTTACCCATTACATACAAAGTGCCTGCGGCTACAATAGGAGAGC	66
ATGAGTAAGGTGTTACAGTGGACGGGAGCATACTGGAGAACCCTCCGCCCTAATTCGCGCAAACCTC	132
M S K V F T W T G A Y W R T F R P N S R K L	22
GCATGTGACTGGTACGACAGTAGTAATAAGACCCACCAAGGCGGAACAGCTGGTGGCAAGTATCGG	198
A C D W Y D S S N K T H Q G G T A G G K Y R	44
GCCTCCAATTACAGTAAGGAGATGATCGAGACACGGAATATCAGCGCAAGTCTCATTCTTTTAC	264
A S N Y S K E M I E T R N I S A S L I P L Y	66
CGGGCGGCGCAAGGTATTAGATACCGACGTTACAAACCCTGGGTGAGTACAGAAGTCCAGATAGAT	330
R A A Q G I R Y R R Y K P W V S T E V Q I D	88
ATAAGCAGCCATCGTCTAATATGCAGGAATCTACCCTGGGCGTGGCTCCGACAATTATCCAAGAG	396
I S S H R S N M Q E S T L G V A P T I I Q E	110
TGTATATATAGACATTCGAGCCCAGGTGTGTCGTTACCTTATCTGGGTGTAATACTCCCGCCGCA	462
C I Y R H S S P G V S L P Y V G V N T P P A	132
TACAACCTCCATGTCGACTCCAGATTTGAGGTTCCCTGACGCTCGACTTAAATGGCCCCGACCGGTCA	528
Y N S M S T P D L R F L T L D L N G P D R S	154
GCTACCGGCTTGCTATACTCTTACCTAGTTTCGATCGGCGATTGTTCAACTGTGGACGAAGGAAG	594
A T G L L Y S Y L V S I G D L F N C G R R K	176
ATGATGTTGTGGGATTTATAGAGTCCCAAGTTAGCTGAAAACGCGCTGGCGAACGGGACCGTCT	660
M M L C G I Y R V P K L A E T R W R T G P S	198
GCTCAGTACCATGGTCCGGGATATTGCTACAAACCGACAACCTACCCTTTTTCGTGTGAGCTAAGG	726
A Q Y H G R G Y C Y K P T T Y P F S C E L R	220
CGGGGATATTGTTGGTACAGTACTCGACGGATACCAGAACTTCATGCGGTTCAACCGGGCTCC	792
R G Y C W Y R Y S T D T T K L H A V Q P G S	242
GCCAGGGATCTCCTGCAAAGCCTAGCGTACCTCGCCAAGCATAGCACCTTGACCGAGCCATGCGTT	858
A R D L L Q S L A Y L A K H S T L T E P C V	264
AGGTACATGTGGAGACCGTGTAGTGTATCGGCGGGTACCGGGTGAACGACTGCAAGAGTCTAG	924
R Y M W R P C M S D R R V P G E R L Q E F *	
ACCTTCGCCCTAATTCGCGCAAAC	

Matthieu Lemesre
V1. c.372T>G. codon initial TAT → TAG → codon STOP

Matthieu Lemesre
V4. c.712G>A codon initial GTT (valine) → ATT (isoleucine)

E FAUX Voir question précédente, il s'agit d'une mutation FAUX-SENS.

Question 3 – ATM (*) :

Parmi les délétions d'exons possibles du gène atm suivantes, laquelle/ lesquelles entraine(nt) un décalage du cadre de lecture ?

- A. Délétion de l'exon 2.
- B. Délétion de l'exon 3.
- C. Délétion de l'exon 4.
- D. Délétion simultanée des exons 2, 3 et 4.
- E. Délétion simultanée des exons 3 et 4.

Pour résoudre cet exercice, soit on regarde directement notre séquence d'ADNg en repérant directement dans sa tête les délétions de quels exons entraineront ou pas un décalage du cadre de lecture. Deuxième possibilité, plus longue mais plus sûre : réaliser un tableau ! On observe sur la combienième base de chaque codon les exons

commencent-ils et se finissent ils et on remplit notre tableau en fonction. Je vous redonne ici un exemple, entre la fin de l'exon 1 et le début de l'exon 2 (voir schéma ci-dessous).

```

AATAAGACCCACCAggttgatgatacgttactagcgatagctctggggattaaggcgttgaggcaaaaag 264
  N  K  T  H  Q                                     35
atgagAGGCAGGGAACAGCTGGTGGCAAGTATCGGGCCTCCAATTACAGTAAGGAGATGATCGAGACA 330
  G  G  T  A  G  G  K  Y  R  A  S  N  Y  S  K  E  M  I  E  T  55
  
```

On constate que l'exon 1 se termine par un codon incomplete CA codant pour une glutamine, la dernière base de ce codon se situe au début de l'exon suivant, A.

L'exon 1 termine donc sur un +2 (deuxième base d'un codon) et l'exon 2 attaque sur un +3 (la dernière base de ce codon).

Exon 1			Exon 2			Exon 3			Exon 4			Exon 5		
début	/	fin	début	/	fin	début	/	fin	début	/	fin	début	/	fin
		+2	+3		+1	+2		+3	+1		+2	+3		

A VRAI En déléant l'exon 2, on passe donc d'un +2 à un +2 : il y aura donc décallage du cadre de lecture.

B VRAI En déléant l'exon 3, on passe donc d'un +1 à un +1 : il y aura donc décallage du cadre de lecture.

C VRAI En déléant l'exon 4, on passe donc d'un +3 à un +3 : il y aura donc décallage du cadre de lecture.

D FAUX En déléant simultanément les exons 2, 3 et 4, on passe donc d'un +2 à un +3 : il n'y aura donc pas de décallage du cadre de lecture.

E VRAI En déléant simultanément les exons 3 et 4, on passe donc d'un +1 à un +3 : il y aura donc décallage du cadre de lecture.

Énoncé commun aux questions 1 et 2 (VOIR SEQUENCE TYP EN ANNEXES):

Le gène TYR code pour la tyrosinase qui est impliqué notamment dans la synthèse de mélanine indispensable chez l'être humain pour la coloration des téguments (peau, cheveux, yeux). Cette synthèse passe par la formation de la dihydroxy-phénylalanine qui se produit à partir de la tyrosine grâce à l'enzyme codée par ce gène.

Il est localisé sur le locus 11q14.314.3 c'est-à-dire sur le bras long du chromosome 11. Les séquences de l'ADNc et de l'ADNg du gène TYR (qui code pour la protéine tyrosinase) sont présentes en annexe (exons en majuscules et bordures introniques en minuscules).

Question 1 – A propos des séquences ():**

- A. La séquence 1 correspond à l'ADNg du gène TYR.
- B. Le gène contient 5 exons qui seront transcrits et traduits.
- C. L'épissage alternatif de l'exon 3 conduit à une protéine tronquée.
- D. L'exon 5 commence en c.494.
- E. Ce gène code pour une protéine de 645 nucléotides.

A VRAI Effectivement, la séquence qui contient encore les introns est l'ADNg.

B FAUX Il contient 5 exons dont seulement 4 codant. Un exon codant sera transcrit et traduit.

C VRAI C'est vrai, question pas vraiment sympa mais mieux vaut bien s'entraîner maintenant ça n'en sera que plus simple ensuite. Tout d'abord quand on parle d'épissage d'un exon ou d'exon skipping (même signification), il faut regarder si le dernier exon précédent l'épissage et le premier suivant l'épissage sont composés d'un codon chevauchant. Pour cela, on regarde la séquence d'ADNg : On remarque que l'exon 2 se termine par 1 nt supplémentaire (une guanine qui est surligné en jaune sur la séquence ci-dessous) avant le site donneur GT de l'intron 2. En temps normal, après un épissage de l'intron 2, la guanine s'associe avec les 2 premiers nucléotides de l'exon 3 (A et G) après le site accepteur AG de l'intron 2 pour former de l'acide glutamique E. Cependant, comme l'exon 3 est épissé on s'intéresse aux premiers nucléotides de l'exon 4. On remarque qu'il ne s'agit pas d'un codon chevauchant puisque les 3 premiers nucléotides de l'exon 4 font tous partis du même codon (ATG qui codent pour une méthionine).

Comme il y a un nucléotide qui fait partie d'un codon chevauchant à la fin de l'exon 2 et qu'il n'y pas de codon chevauchant entre l'exon 2 et 3, il y alors un décalage du cadre de lecture d'1 nucléotide (s'il n'y avait aucun nt qui faisait partie de codons chevauchants ou 3 nts en faisait partie alors il n'y aura pas eu de décalage du cadre de lecture).

Maintenant, il s'agit de chercher si ce décalage provoque un codon stop (TGA) prématuré. Pour cela, il faut regarder les nouveaux codons formés après l'épissage de l'exon 3 ici en bleu et jaune. On remarque qu'un codon stop prématuré est formé à cause du frameshift (= décalage du cadre de lecture), la protéine est donc tronquée à cause de l'épissage alternatif.

Séquence 2 : ADNc

En rouge : l'exon 3 qui va être épissé

En **bleu** et **jaune** : les nouveaux codons qui vont être formés après le décalage du cadre de lecture.

En **vert** : le codon stop prématuré

```

AGAGTTGCGGGCTTCGCTCGGCTCGGGGTTCCCCACGATCTCTCACCTGTGTTTCGGCG      60
GAGTTGCGGGGTTCCCTCTGCTCCCAGGTCCTTGAGATCCCTGATCCCGCTCAGGGC      120
AGGGAATTCGGGGTTTCGCCCTCTGGGCTTCCCGCGATCCCAGAGCCCAGCGTTCCAAC      180
AGAATTGGGAGGGTTTATTGCGCTAGAGGAGTCTTCAGGATCTTCGCTCCCAGCCCTCCGA      240
GGTGGCCCTGAGAAGGCAGCAGGCCAGGAGGAGGAATTGGGTATCAGCCACCCCATCCC      300
ACTGCCAGTGCGGCCGAGCTGCTTGTCAAAGAGAGAACAATGGCAGTAACCCGTGCCT      360
CATGACTGAGTGCAGTGGCACCTCTCAGCCACCGCCAGTGTCCCACCCTGCAGC      420
M T E C S G T S Q P P P A S V P T T A A      20
TTCAGAGGGACGTATGGTCAATCCAGGATATTCCTGCTGTACCAGCAGAGGGCATGTGGA      480
S E G R M V I Q D I P A V T S R G H V E      40
GAAACACACCTGACCTGGTTTCAGACTCCACCTACTACAGCAGCTTCTACCAGCCGTCTCT      540
N T P D L V S D S T Y Y S S F Y Q P S L      60
GTTTCCTTATTACAACAATCTATACAACCTGCCCGCAGTACTCCATGGCCTTGGCTGTGA      600
F P Y Y N N L Y N C P Q Y S M A L A A D      80
TTCTGCTTCTGGGGAGGTGGGAAATCCCCTCGGGGATCCCCTGTGAAGAACAGCCTTCG      660
S A S G E V G N P L G G S P V K N S L R      100
GGGCCTCCCCGGACCTTATGTGCCTGGTCAGACAGGAAACCAGTGGCAGATGAAGAACAT      720
G L P G P Y V P G Q T G N Q W Q M K N M      120
GGAGAACCCTGCATGCAATGAGCTCCAGTACAGGATGCATTCTTACTACCCGCCCTCCCTC      780
E N R H A M S S Q Y R M H S Y Y P P P S      140
TTACCTGGGCCAGAGCGTGCCCAAGTTCTTCACTTTTCAGGATGCTCCCTCTTACCCGGA      840
Y L G Q S V P Q F F T F E D A P S Y P E      160
AGCCAGGGCGAGCGTATTCTCGCCGCCAGCAGTCAAGATTCGGCTTGGTTTCCCTCTC      900
A R A S V F S P P S S Q D S G L V S L S      180
GAGCAGCTCTCTATTAGTAACAAGAGCACAAAGGCAGTGCCTGAATGTGAGCCTGCGTC      960
S S S P I S N K S T K A V L E C E P A S      200
GGAGCCCAGCAGCTTACAGTCACTCCCCTCATCGAGGAGGACGAGTGAGCAGTGCCTGC      1020
E P S S F T V T P V I E E D E *      215
TGCCGATGGCGGTTCACTTGGAGTAACAGGCTTATTCCACTTCCATGGGGTTTGTAAAT      1080
ATTTTGCATTGACTCATACTATCTTAACTGTTGAGAACGATTTGGTTTATATTCCTTAT      1140
GAGTTTAGTCCAGAGGCTGTACACATTTGTAATACTTTAGGGTCCGTGACTACCATCTG      1200
CATGGTTTTAAGTGCTTTACTCACGGAGTTTAAATAATAGTGTTCATTTTTTTAATGACAC      1260
TGGTTTCATGTAGTTTTCAAGAAATAAAGAATTCATTCAAGTGAAGCCATTTGTGTGCC      1320
TCTAAATGAGTCATCTAATTAGATGTTACTTTTAGTTTTAAATGAAATCTTAGGTGCCT      1380
TAGGGG...      1386

```

D VRAI Je veux d'abord chercher où se trouve le nucléotide placé en c.494, pour cela je sais que je dois faire 494 + 5'UTR (car numérotation c. ne contient pas la 5'UTR donc on ne peut pas s'aider de la numérotation sur le côté de la séquence sans avoir additionner le 5'UTR), le 5'UTR fait 361 nucléotides : 5'UTR en marron, tout ce qui est en amont du 1er ATG codant pour une M dans l'ADNc)

```

AGAGTTGCGGGCTTCGCTCGGCTCGGGGTTCCCCACGATCTCTCACCCGTGTGTTTCGGCG      60
GAGTTGCGGGGGTTCCTCTGCTCCCGGGTCCCTGAGATCCCTGATCCCGCTCTCAGGGC      120
AGGGAATTCGGGGTTCGCCCTCTGGGCTTCCCCCGCATCCAGAGCCCGGTTCCAAC      180
AGAATTGGGAGGGTTATTGCGCTAGAGGAGTCTTCAGGATCTTCGCTCCCGCCCTCCGA      240
GGTGGCCCTGAGAAGGCAGCAGGCCAGGAGGAGGAATTGGGTATCAGCCACCCCATCCC      300
ACTGCCAGTGCGGCCGAGCTGCTTGTCAAAGAGAGAACAATGGCAGTAACCCGTGCCT      360
CATGACTGAGTGCAGTGGCACCTCTCAGCCACCGCCGGCCAGTGTCCCACCCTGCAGC      420
M T E C S G T S Q P P P A S V P T T A A      20
TTCAGAGGGACGTATGGTCATCCAGGATATTCCTGCTGTCACCAGCAGAGGGCATGTGGA      480
S E G R M V I Q D I P A V T S R G H V E      40
GAACACACCTGACCTGGTTTCAGACTCCACCTACTACAGCAGCTTCTACCAGCCGTCTCT      540
N T P D L V S D S T Y Y S S F Y Q P S L      60
GTTTCCTTATTACAACAATCTATAACAAGTCCCGCAGTACTCCATGGCCTTGGCTGCTGA      600
F P V V N N T V N C P O V S M A T A A D      80

```

Donc 494+361= 855, je cherche à le 855ème nt sur mon ADNc à l'aide de ma numérotation sur ma séquence, je tombe sur le T du GTA codant pour la valine 165 (surligné en **jaune**).

```

AGAGTTGCGGGCTTCGCTCGGCTCGGGGTTCCCCACGATCTCTCACCCGTGTGTTTCGGCG      60
GAGTTGCGGGGGTTCCTCTGCTCCCGGGTCCCTGAGATCCCTGATCCCGCTCTCAGGGC      120
AGGGAATTCGGGGTTCGCCCTCTGGGCTTCCCCCGCATCCAGAGCCCGGTTCCAAC      180
AGAATTGGGAGGGTTATTGCGCTAGAGGAGTCTTCAGGATCTTCGCTCCCGCCCTCCGA      240
GGTGGCCCTGAGAAGGCAGCAGGCCAGGAGGAGGAATTGGGTATCAGCCACCCCATCCC      300
ACTGCCAGTGCGGCCGAGCTGCTTGTCAAAGAGAGAACAATGGCAGTAACCCGTGCCT      360
CATGACTGAGTGCAGTGGCACCTCTCAGCCACCGCCGGCCAGTGTCCCACCCTGCAGC      420
M T E C S G T S Q P P P A S V P T T A A      20
TTCAGAGGGACGTATGGTCATCCAGGATATTCCTGCTGTCACCAGCAGAGGGCATGTGGA      480
S E G R M V I Q D I P A V T S R G H V E      40
GAACACACCTGACCTGGTTTCAGACTCCACCTACTACAGCAGCTTCTACCAGCCGTCTCT      540
N T P D L V S D S T Y Y S S F Y Q P S L      60
GTTTCCTTATTACAACAATCTATAACAAGTCCCGCAGTACTCCATGGCCTTGGCTGCTGA      600
F P Y Y N N L Y N C P Q Y S M A L A A D      80
TTCCTGCTTCGGGGAGGTGGGAAATCCCTCGGGGGATCCCTGTGAAGAACAGCCTTCG      660
S A S G E V G N P L G G S P V K N S L R      100
GGGCCCTCCCGGACCTTATGTGCTGGTCAGACAGGAAACCAGTGGCAGATGAAGAAT      720
G L P G P Y V P G Q T G N Q W Q M K N M      120
GGAGAACCAGCATGCAATGAGCTCCAGTACAGGATGCATTCTACTACCCGCTCCCTC      780
E N R H A M S S Q Y R M H S Y Y P P P S      140
TTACCTGGCCAGAGCGTCCCGAGTTCTTCACTTTTGAGGATGCTCCCTCTTACCCGGA      840
Y L G Q S V P Q Q F F T F E D A P S Y P E      160
AGCCAGGGCGAGCGTATTCTCGCCGCCAGCAGTCAAGATTCTGGCTTGGTTCCCTCTC      900
A R A S V F S P P S S Q D S G L V S L S      180
GAGCAGCTCTCCTATTAGTAACAAGAGCACAAAGGCAGTGCTTGAATGTGAGCCTCGCTC      960
S S S P I S N K S T K A V L E C E P A S      200
GGAGCCCAGCAGCTTACAGTCACTCCCGTATCGAGGAGGACGAGTGGCAGTGCCTGC      1020
E P S S F T V T P V I E E D E *      215
TGCCGATGGCGGTTCACTTGGAGTAACAGGCTTATCCACTTTCCATGGGGTTTGTTAAT      1080
ATTTTGCATTGACTCATACTATCTTAAGTGTGAGAACGTAATTTGGTTTATATTCCTTAT      1140
GAGTTTAGTCCAGAGGCTGTAACACATTTGTAATACTTTAGGGTCCGTGACTACCATCTG      1200
CATGGTTTAAGTGCCTTACTCACGGAGTTTAAATAATAGTGTTCATTTTTTAAATGACAC      1260
TGGTTTCATGTAGTTTCAAGAAATAAAGAATTCATTCAAGTGAAGCCATTTGTGTGCC      1320
TCTAAATGAGTCATCTAATTAGATGTTACTTTTAGTTTTAAATGAAATCTTAGGTGCCT      1380
TAGGGG (...)      1386

```

Pour savoir où se situe ce nt par rapport à un exon, je dois regarder sur mon ADNg (vu que c'est pas la même numérotation pour les nucléotides que sur l'ADNc, mais que celle des AA est la même je regarde la valine 165 sur l'ADNg) :


```

GTTTCCTTATTACAACAATCTATACAACTGCCCGCAGTACTCCATGGCCTTGGCTGCTGA
F P Y Y N N L Y N C P Q Y S M A L A A D      80
TTCTGCTTCTGGGGAGGTGGGAAATCCCCTCGGGGGATCCCTGTGAAGAACAGCCTTCG
S A S S G E V G N P L G G S P V K N S L R    100
GGGCCTCCCCGGACCTTATGTGCCTGGTCAGACAGGAAACCAGTGGCAGgtatgatatta
G L P G P Y V P G Q T G N Q W Q            116
attaccacagaaagt(...)tctgatgatatttcttcttttttcttaagcagATGAAGAACAT
M K N M                                     120
GGAGAACC GCCATGCAATGAGCTCCAGTACAGGATGCATCTTACTACCGCCTCCCTC
E N R H A M S S Q Y R M H S Y Y P P P S    140
TTACTTGGCCAGAGCGTGCCCCAGTTCCTCACCTTTGAGGATGCTCCCTCTTACCCGGA
Y L G Q S V P Q F F T F E D A P S Y P E    160
AGCCAGGGGAGCGataaatatctaatcacacagaactgactatgctc(...)ctttctctct
A R A S                                     164
cacctcactcgcagTATTCTCGCGCCAGCAGTCAAGATCTGGCTTGGTTTCCCTCTC
V F S P P S S Q D S G L V S L S           180
GAGCAGCTCTCCTATTAGTAACAAGAGCACAAAGGCAGTGTGAATGTGAGCCTGCGTC
S S S P I S N K S T K A V L E C E P A S    200
GGAGCCCAGCAGCTTCACAGTCACTCCCCTCATCGAGGAGGACGAGTGAAGCAGTGCCTGC
E P S S F T V T P V I E E D E *          215
TGCCGATGGCGGTTCACTTGGAGTAACAGGCTTATTCACCTTCCATGGGGTTTGTAAAT

```

Je remarque donc que mon T correspond bien au début de l'exon 5.

E FAUX Une protéine contient des AA et non des nucléotides ! Ce gène code donc pour une protéine de 215 acides aminés.

Enoncé question 2:

Plus de 330 variants du gène TYR sont actuellement décrits dans la base de données OMIM (Online Mendelian Inheritance in Man), parmi lesquels les variants suivants, dont l'activité enzymatique a été mesurée pour tester leur fonctionnalité.

Enzyme Tyrosinase	Variant
Sauvage	
Variant A	c.228C>T
Variant B	c.320T>A
Variant C	c.588T>A
Variant D	c.65-2A>T

Question 2 – A propos de ces variants (**):

- Le variant A correspond à une mutation non sens.
- Le variant B correspond à une mutation faux sens.
- Le variant C correspond à une mutation non sens.
- Le variant D affecte le site accepteur d'épissage de l'intron 3.
- Tous ces variants sont des variants de séquence.

A FAUX 228+361= 589, ce qui correspond au C du codon GCC codant pour l'alanine n*76, GCC-->GCT (voir code AA) ce qui donne aussi une alanine, donc c'est une mutation synonyme.

```

AGAGTTGCGGGCTTCGCTCGGCTCGGGTTCCCCACGATCTCTCACCCGTGTGTTTCGGCG      60
GAGTTGCGGGGGTTCCCTCTGCTCCCGGTCCCTGAGATCCCTGATCCCGCTCTCAGGGC      120
AGGGAATTCGGGGTTCGCCCTCTGGGCTTCCCGCGATCCCAGAGCCCGGTCCAAC      180
AGAATTGGGAGGGTTTATTGCGCTAGAGGAGTCTTCAGGATCTTCGCTCCCGCCCTCCGA      240
GGTGGCCCTGAGAAGGCAGCAGGCCAGGAGGAGGAATTGGGTATCAGCCACCCCATCCC      300
ACTGCCCAGTGGGGCCGAGCTGCTTGTCAAAGAGAGAACAATGGCAGTAACCCGTGCCT      360
CATGACTGAGTGCAGTGGCACCTCTCAGCCACCGCCGGCCAGTGTCCCACCACTGCAGC      420
M T E C S G T S Q P P P A S V P T T A A      20
TTCAGAGGGACGTATGGTCATCCAGGATATTCCTGCTGTCCAGCAGAGGGCATGTGGA      480
S E G R M V I Q D I P A V T S R G H V E      40
GAACACACCTGACCTGGTTTCAGACTCCACCTACTACAGCAGCTTCTACCAGCCGTCTCT      540
N T P D L V S D S T Y Y S S F Y Q P S L      60
GTTTCCTTATTACAACAATCTATACTGCCCAGTACTCCATGGCCTGGCTGCTGA      600
F P Y Y N N L Y N C P Q Y S M A L A A D      80
TTCTGCTTCTGGGGAGGTGGGAAATCCCTCGGGGGATCCCTGTGAAGAACAGCCTTCG      660
S A S G E V G N P L G G S P V K N S L R      100
GGGCCTCCCGGACCTTATGTGCCTGGTCAGACAGGAAACCAGTGGCAGATGAAGAACAT      720
G L P G P Y V P G Q T G N Q W Q M K N M      120
CGAGACCGGCATGCAATGACCTCCAGCTACAGCATGCAATTCCTACTACCCGGCTCCCTC

```

B VRAI 320+361= 681 qui correspond au T du GTG codant pour la valine 107, GTG→GAG, V→E, il y a donc un changement d'AA ce qui correspond à une mutation FAUX sens.

C VRAI 588+361= 949 qui correspond au T du codon TGT codant pour la cystéine 196, TGT→ TGA (codon stop), donc cela correspond à une mutation NON sens.

D FAUX 65+361= 426, qui correspond au A du codon GAG codant pour un glutamate :

```

AGAGTTGCGGGCTTCGCTCGGCTCGGGTTCCCCACGATCTCTCACCCGTGTGTTTCGGCG      60
GAGTTGCGGGGGTTCCCTCTGCTCCCGGTCCCTGAGATCCCTGATCCCGCTCTCAGGGC      120
AGGGAATTCGGGGTTCGCCCTCTGGGCTTCCCGCGATCCCAGAGCCCGGTCCAAC      180
AGAATTGGGAGGGTTTATTGCGCTAGAGGAGTCTTCAGGATCTTCGCTCCCGCCCTCCGA      240
GGTGGCCCTGAGAAGGCAGCAGGCCAGGAGGAGGAATTGGGTATCAGCCACCCCATCCC      300
ACTGCCCAGTGGGGCCGAGCTGCTTGTCAAAGAGAGAACAATGGCAGTAACCCGTGCCT      360
CATGACTGAGTGCAGTGGCACCTCTCAGCCACCGCCGGCCAGTGTCCCACCACTGCAGC      420
M T E C S G T S Q P P P A S V P T T A A      20
TTCAGAGGGACGTATGGTCATCCAGGATATTCCTGCTGTCCAGCAGAGGGCATGTGGA      480
S E G R M V I Q D I P A V T S R G H V E      40
GAACACACCTGACCTGGTTTCAGACTCCACCTACTACAGCAGCTTCTACCAGCCGTCTCT      540
N T P D L V S D S T Y Y S S F Y Q P S L      60
GTTTCCTTATTACAACAATCTATACTGCCCAGTACTCCATGGCCTGGCTGCTGA      600
F P Y Y N N L Y N C P Q Y S M A L A A D      80
TTCTGCTTCTGGGGAGGTGGGAAATCCCTCGGGGGATCCCTGTGAAGAACAGCCTTCG      660
S A S G E V G N P L G G S P V K N S L R      100
GGGCCTCCCGGACCTTATGTGCCTGGTCAGACAGGAAACCAGTGGCAGATGAAGAACAT      720
G L P G P Y V P G Q T G N Q W Q M K N M      120

```

On regarde ensuite où se situe cette mutation sur l'ADNg (on sait que ça touche le glutamate 22) :

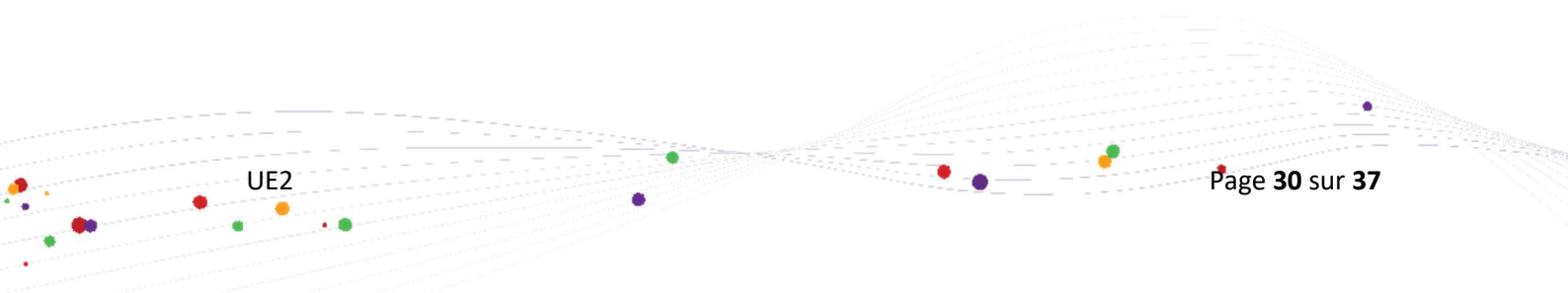
```

AGAGTTGCGGGCTTCGCTCGGCTCGGGTTCCCCACGATCTCTACCCTGTGTTTCGGCG
GAGTTGCGGGGGTTCCCTCTGCTCCCGGGTCCCTGAGATCCCTGATCCCCTCTCAGGGC
AGGGAATTCGGGGTTCGCCCTCTGGGCTTCCCCCGGATCCCAGAGCCCGGTCCAAC
AGAATTTGGGAGGGTTTATTGCGCTAGAGGAGTCTTCAGGATCTTCGCTCCCGCCCTCCGA
Ggtgaggetagaggggggattcgaacta (...) acaatatgactcattgtcgtatgctcca
qGTGGCCCTGAGAAGGCAGCAGGCCAGGAGGAGGAATTGGGTATCAGCCACCCCATCCC
ACTGCCAGTGCGCCGAGCTGCTTGTCAAAGAGAGAACAATGGCAGTAACCCGTGCCT
CATGACTGAGTGCAGTGGCACCTCTCAGCCACCGCCGCCAGTGTCCCCACCACATGCAGC
M T E C S G T S Q P P P A S V P T T A A 20
TTCAGgtaatctagagggactgaggttcac (...) tctagagatttttctttttcttccaa
S 21
tttagAGGGACGTATGGTCATCCAGGATATTCCTGCTGTCACCAGCAGAGGGCATGTGGA
E G R M V I Q D I P A V T S R G H V E 40
GAACACACCTGACCTGGTTTCAGACTCCACCTACTACAGCAGCTTCTACCAGCCGTCTCT
N T P D L V S D S T Y Y S S F Y Q P S L 60
GTTTCCTTATTACAACAATCTATACAACTGCCCGCAGTACTCCATGGCCTTGGCTGCTGA
F P Y Y N N L Y N C P Q Y S M A L A A D 80
TTCCTGCTTCTGGGGAGGTGGGAAATCCCTCGGGGGATCCCCTGTGAAGAACAGCCTTCG
S A S G E V G N P L G G S P V K N S L R 100
GGGCCCTCCCGGACCTTATGTGCCTGGTCAGACAGGAAACCAGTGGCAGgtatgatatta
G L P G P Y V P G Q T G N Q W Q 116
attaccacagagat (...) tctgatgatatttcttctttttcttaagcacATGAAGAACAT
M K N M 120

```

La mutation c.65-2 se situe donc en amont du A (du codon GAG codant pour la glutamate 22), cette mutation impacte donc le site accepteur de l'intron 2 et non 3 (compter quel intron cela touche).

E VRAI Tous ces variants sont des SNV et donc des variants de séquence (<50pb)



Énoncé commun aux questions 1 et 4 (VOIR SEQUENCES RET EN ANNEXE):

La maladie de Hirschsprung est une maladie intestinale due à une anomalie d'innervation intestinale. Elle est causée par un défaut de migration des cellules de la crête neurale à destinée entérique (il n'y a pas de plexus parasympathique donc pas de péristaltisme intestinal). Elle se traduit donc par un défaut de péristaltisme intestinal sur la partie distale, conduisant à des obstructions intestinales chez le nouveau-né. Le gène responsable de la pathologie est le gène *RET*. Vous recevez en consultation Clément, un nouveau-né de 9 mois chez qui on suspecte une maladie de Hirschsprung, vous apprenez également que le père est atteint de la pathologie. Sa mère est saine. Vous étudiez le gène *RET* de Clément et identifiez certains SNV intéressants. Après études, vous concluez que Clément est lui aussi atteint de la pathologie.

Vous disposez en annexe des séquences de référence d'ADNg et d'ADNc du gène *RET*.

Question 1 – RET :

Le dernier exon du gène *RET* est long de (à 2 paires de bases près) :

- A. 49 pb
- B. 52 pb
- C. 152 pb
- D. 165 pb
- E. 178 pb

Pour résoudre cet exercice, il faut se référer à la séquence de référence de l'ADNg du gène, donc la séquence 1 en annexe. Le dernier exon est celui qui contient le codon stop TAG. Cependant, attention, l'exon ne s'arrête pas au codon stop, il continue plus loin jusqu'à la position g.1055 (il y a 20 acides aminés par ligne donc 60 nucléotides). On repère également que le dernier exon commence en position g.891. L'exon mesure donc bien 165pb (on prend en compte le nucléotide numéro 891 et le numéro 1055).

Le piège ici était de s'arrêter au codon stop !

A FAUX

B FAUX

C FAUX

D VRAI

E FAUX

Question 2 – RET :

A propos du gène *RET*, cochez la(les) réponse(s) vraie(s) :

- A. Il contient 4 exons.
- B. Il contient une partie codante dans son premier exon.
- C. Il code pour une protéine de 162 acides aminés.
- D. Une délétion de l'exon 4 entraîne un décalage du cadre de lecture.
- E. Une délétion de l'exon 3 entraîne l'apparition d'un codon stop prématuré à la place d'une lysine en p.88.

Pour cette question, on se réfère également à la séquence d'ADNg du gène RET.

A FAUX Attention ! Il y a un premier exon qui ne contient pas de séquence codante au début du gène. Nous avons ensuite 4 exons contenant une séquence codante, le gène contient donc 5 exons et pas 4 !

B FAUX Cf Item A, le premier exon se situe sur les 47 premiers nucléotides du gène, il ne contient pas de partie codante, le codon d'initiation ATG se situe dans le 2^{ème} exon.

C VRAI D'après la nomenclature de la séquence, la protéine obtenue avec la séquence d'ADNg de référence du gène RET est bien constituée de 162 acides aminés.

D VRAI Ici, il faut tout d'abord repérer l'exon 4, c'est celui qui commence en g.647 et se termine en g.830. Ensuite, on considère que cet exon se retrouve éliminé lors de l'épissage. On s'intéresse alors à la fin de l'exon 3 et au début de l'exon 5. L'exon 3 se termine par un C qui n'appartient pas à un codon et qui se continue avec l'exon 4 (code pour une Glutamine), il y a donc dans la séquence de référence un chevauchement entre l'exon 3 et l'exon 4. Il en va de même pour la fin de l'exon 4 et le début de l'exon 5 où on retrouve une Sérine à cheval entre les 2 exons.

En éliminant l'exon 4, On se retrouve avec la création d'un codon CAC entre l'exon 3 et l'exon 5 qui code pour une Histidine puis un codon GAC (Acide Aspartique) etc. Cela correspond bien à un décalage du cadre de lecture par rapport au codon CGA de référence qui code pour une Arginine.

E FAUX Par le même mécanisme que l'item précédent, on se retrouve avec la création d'un codon AAC en début d'exon 4 qui code pour une Asparagine en position 85. Si on continue, on a bien l'apparition d'un codon stop TAG en position p.88. Cependant, attention, la lettre L correspond à une Leucine et non pas à une Lysine (K), l'item devient donc faux ! Attention à ne pas vous faire piéger sur les lettres des acides aminés !

Question 3 – SNV :

Vous retrouvez chez Clément le SNV c.231 C>G. A l'aide de vos connaissances, des séquences et du code génétique donnés en annexe, cochez la(les) réponse(s) vraie(s) concernant la mutation :

- A. Il s'agit d'une mutation faux-sens.
- B. Il s'agit d'une mutation non-sens.
- C. Elle entraîne la production d'une protéine tronquée.
- D. Elle entraîne l'apparition d'un acide aminé pouvant subir une phosphorylation.
- E. La protéine obtenue fait 77 acides aminés.

Il faut maintenant s'intéresser à la séquence d'ADNc du gène RET. Ce qu'il faut bien se souvenir, c'est que la numérotation de l'ADNc débute au niveau du A codon d'initiation ATG ! Le A devient donc le c.1 (qui est le 57^{ème} nucléotide sur la séquence 2), il faut ensuite compter 230 nucléotides pour retrouver la mutation identifiée.

On trouve alors une mutation sur le codon TAC de la Tyrosine en position p.77 qui devient alors TAG (qui est un codon stop !).

A FAUX Il s'agit d'une mutation non-sens puisqu'elle entraîne la formation d'un codon stop !

B VRAI Cf Item A.

C VRAI On a la formation d'un codon stop en position 77, on obtient donc bien à la suite de la traduction une protéine tronquée.

D FAUX Attention, si on se trompait en partant du premier nucléotide de la séquence 2, on tomberait sur un codon ACC codant pour une Thréonine qui devient AGC codant pour une Sérine qui peut bien subir une phosphorylation !

E FAUX Attention, le codon stop se retrouve alors en position 77, il ne compte pas comme un acide aminé, la protéine obtenue fait alors 76 acides aminés !

Question 4 – Transmission :

Vous étudiez les mutations retrouvées chez les parents de Clément, comparées à la séquence de référence du gène RET. Ces mutations sont toutes présentes à l'état hétérozygote chez les individus porteurs.

A l'aide de ce tableau, de l'énoncé et des données annexes, quelle(s) affirmation(s) est/sont vraie(s) ?

Père	c.231C>G	c.88A>G
Mère	c.105C>T	c.253T>C
Clément	c.231C>G	c.105C>T

- A. Le variant c.105C>T est probablement bénin.
- B. Le variant c.231C>G est probablement bénin.
- C. La transmission semble dominante.
- D. La mutation pathologique peut également être notée p.77Y>*.
- E. Le gène RET pourrait être localisé sur le chromosome X.

A VRAI D'après l'énoncé, on sait que la mère de Clément n'est pas atteinte, de plus, si on étudie ce variant sur la séquence d'ADNc du gène RET, on voit qu'elle entraîne la transformation d'un codon TTC en TTT qui code tous les deux pour une Phénylalanine. Avec ces données, on suppose que le variant est probablement bénin.

B FAUX On sait que le père est aussi atteint de la pathologie, on voit que ce variant est commun entre Clément et son père. De plus, on sait que tous les variants cités sont présents à l'état hétérozygote chez les porteurs, ce qui nous fait suspecter une transmission dominante. Avec ça, et les éléments de la question précédente, on suspecte ce variant d'être responsable de la pathologie.

C VRAI Cf Item B.

D VRAI En effet, d'après les données de la question 3 et l'étude des séquences, on remplace une Tyrosine par un codon stop en position 77.

E FAUX Ici, on sait que le père est atteint de la pathologie et que Clément en est également atteint. Clément étant un garçon, son père lui a forcément transmis son chromosome Y, le gène RET ne peut donc pas être sur le chromosome X puisque Clément est atteint.

Énoncé commun aux questions 1 à 3 (VOIR SEQUENCES FGFR3 EN ANNEXE) :

L'achondroplasie est la maladie osseuse constitutionnelle la plus fréquente avec une fréquence de 1/25 000 naissances vivantes. Cliniquement, on retrouve une petite taille, un raccourcissement des membres, une macrocéphalie, une hypotonie, ... Sur le plan génétique, elle est due à une mutation du gène *FGFR3* qui code pour un récepteur de croissance fibroblastique. Ce gène est aussi responsable de deux autres formes de nanisme : le nanisme thanatophore (souvent léta) et l'hypochondroplasie.

Après étude du gène *FGFR3* sur plusieurs individus malades, vous retrouvez quatre variants intéressants :

V1 : c.273T>A

V2 : c.243-1G>A

V3 : c.341del

V4 : c.449T>C

Vous disposez en annexe du code génétique et des séquences d'ADNg et d'ADNc de référence du gène *FGFR3*.

Question 1 – FGFR3 :

A propose du gène *FGFR3*, cochez la(les) réponse(s) vraie(s) :

- A. Ce gène contient 4 exons codants.
- B. Le 2^{ème} exon de ce gène fait 88pb.
- C. V1 est une mutation faux sens.
- D. V3 entraîne un décalage du cadre de lecture.
- E. V4 est une mutation faux-sens.

Ici, on s'intéresse à la séquence d'ADNg du gène. On compte bien 4 exons tous codants.

Z

Grâce à la numérotation des AA, on peut déterminer qu'il y a $23 \times 3 = 69$ nucléotides par ligne. On peut donc déduire la taille du 2^{ème} exon à 88pb.

Pour l'étude des variants, il faut se référer à la séquence d'ADNc. On déduit que la région 5'UTR fait 27pb donc pour nos variants :

V1 : $273 + 27 = 300$. On regarde le nucléotide en position 300, on tombe sur la transformation d'un TAT (Y) en un TAA (codon stop !), c'est donc une mutation NON-sens.

V2 : Il faut pour celui-là regarder les 2 séquences, le nucléotide en position $243 + 27 = 270$ de l'ADNc correspond au premier T de l'exon 3 sur la séquence d'ADNg. La variation en c.243-1 affecte donc le site accepteur d'épissage en transformant le AG en AA.

V3 : Le nucléotide en position $c.341+27=368$ de l'ADNc correspond au A du codon CAC codant pour une Histidine. Il y a délétion d'un nucléotide donc décalage du cadre de lecture !

V4 : Le nucléotide en position $c.449+27=476$ entraîne la transformation d'un ATT (I) en ACT (T) donc il s'agit bien d'une mutation faux-sens.

A VRAI

B VRAI

C FAUX V1 est une mutation non-sens !

D VRAI Délétion d'un seul nucléotide au milieu de la partie codante.

E VRAI Ce variant entraîne un changement d'AA donc il s'agit bien d'une mutation faux-sens.

Question 2 – Étude des variants :

A propos des quatre variants retrouvés dans le gène *FGFR3*, cochez la(les) réponse(s) vraie(s) :

- A. V4 entraîne l'apparition d'un acide aminé pouvant subir une phosphorylation.
- B. V2 entraîne la délétion de l'exon 4.
- C. V3 entraîne l'apparition d'un codon stop en position p.119.
- D. V1 entraîne la production d'une protéine de 91AA.
- E. V2 entraîne un décalage du cadre de lecture.

On étudie maintenant les conséquences de chaque variants :

Séquence 2

```

TTTTTAGCCTTGGCAGACCTCCAGATGAGATTGCCAGCATTGAGCTAGCGAGTCAGCGATATGCA 69
                    M R L P R I E L A S Q R Y A 14
TCACGGCGTTTCARAGCGTGGGAGTATGTGAACCAAGGCTCCGGACAGGACTATATACTTGGGTTGAT 138
S R A F K R R E Y V N Q G S G Q D Y I L G F D 37
CTGGCCCGGACAACCTGCAAACTCAACATTTATAGATTAAGGTTAGGAGTGAAGCCACTTTGCTTTC 207
L A P T T A N L N I Y R L K G R S E A T L L C 60
GGCTAGTCCCAATGAAACCTATGGACTTTGTTTGGGTAGCATCAGAAATCTCAACCTTTGTGAC 276
A V V P M K N L W T L F W V A S G I S T L C H 83
ACCTAGACTGCTCAAGTAATAAGGGAAGCGCGGGCTGGCCGAGCGGTTCCGGCGCCACGTGT 345
T Y T A R S K Y G K R A A W P E A F R A A T C 106
TGGTAACTGTTGATGGTGGCATAAGCAATACCGTAGTCCCTCAATTCAGCTCTGTATCTCGAG 414
S L T V D W W H I S H T V V P Q I Q L C Y L E 129
CGTTATGTGCAATGGCGTAGAAGGATAGACTGATAGAGATGGGTACGGGCTCCAAATCATCCA 483
R Y V S N G V E R D R L I E D G Y G P P I H P 152
ACACTTAGCCCTTCTCCAAGAGCTAGTAGGGCACCTGCA 524
T L Y A F S K S * 160
                    
```

```

GATCTGCCCCGACAACTGCAAACTCAACATTTATAGATTAAGGTTAGCGAaattgcaagtgttg
D L A P T T A N L N I Y R L K
gogcccgagactgtccccagagtgtggtctttgatctgacaaagcgcgagcctccatcgogggcgatt
gtttctgcgaccatgtcgtccatagattttgggcatgtttccagagTAGGAGTGAAGCCACTTTGCTT
                    G R S E A T L L 59
TGCGCGTAGTCCCAATGAAACCTATGGACTTTGTTTGGGTAGCATCAGAAATCTCAACCTTgtga
C A V V P M K N L W T L F W V A S G I S T L
atgtgggggtgcgagcatagacattatctccggttcaagttaggcatgaggtgcatgctaagTTGT
                    C 82
CACACTAGCTGCTCGAAGTAATATGGGAAGCGCGGGCTGGCCGAGCGGTTCCGGCGCCACAG
H T Y T A R S K Y G K R A A W P E A F R A A T C 105
TGTTGTTAACTGTTGATTGGTGGCACATAAGCAATACCGTAGTCCCTCAATTCAGCTCTGTATCTC
C S L T V D W W H I S H T V V P Q I Q L C Y L 128
GAGCGTTATGTGCAATGGCGTAGAAGGATAGACTGgttgacactagctgtgttcggttcggttaa
E R Y V S N G V E R D R L
                    
```

A VRAI V4 entraîne l'apparition d'une thréonine à la place d'une isoleucine. La thréonine peut bien subir une phosphorylation.

B FAUX V2 touche le site accepteur d'épissage de l'intron situé entre l'exon 2 et l'exon 3, l'épissage ne va donc pas s'arrêter à cet endroit et va continuer jusqu'au prochain site accepteur AG avant l'exon 4. On a donc une délétion de l'exon 3 avec V2.

C VRAI V3 entraîne un décalage du cadre de lecture qui aboutit quelques AA plus loin à l'apparition d'un codon stop en position p.119

D FAUX V1 : il entraîne la formation d'un codon stop TAA en position p.91 donc la formation d'une protéine de 90AA.

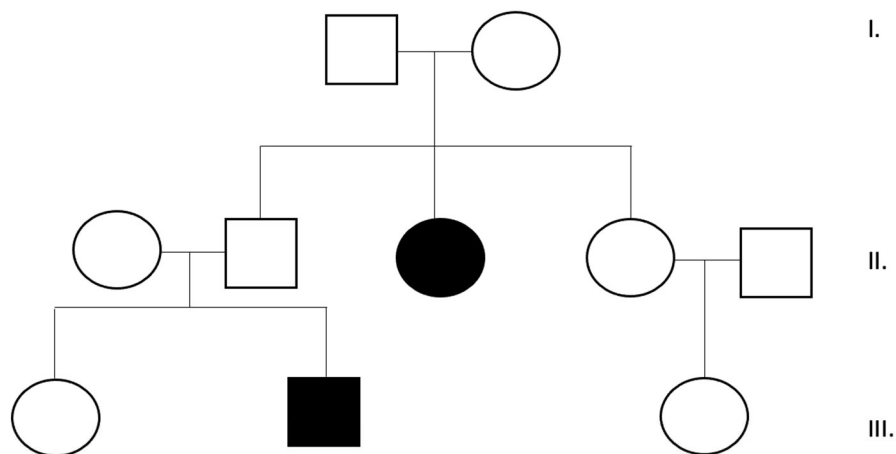
E VRAI On utilise la méthode pour savoir si la délétion d'un exon entraîne un décalage du cadre de lecture. On numérote donc le premier et le dernier nucléotide des exons qui nous intéressent en fonction de sa place dans le codon :

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

Si on enlève l'exon 3, on a un +2 en fin d'exon 2 et un +1 en début d'exon 4, il y a donc bien un décalage du cadre de lecture.

Question 3 – Transmission :

Vous recevez en consultation un nouveau-né chez qui vous découvrez une achondroplasie. Vous étudiez la transmission de cette pathologie dans sa famille, pour cela vous disposez d'un arbre phylogénétique de transmission ainsi que de l'ébauche d'un tableau regroupant les allèles portés par chaque individu de cette famille. Un allèle noté « A » est un allèle atteint de l'altération moléculaire responsable de l'achondroplasie, un allèle noté « a » est un allèle sain. D'après l'énoncé, l'arbre phylogénétique de transmission, l'ébauche de tableau de répartition allélique et vos connaissances, quelle(s) affirmation(s) est/ sont vraie(s) ? Le nouveau-né étudié est l'individu III.2



	I.1	I.2	II.1	III.2	III.3
Allèles	aa	aa	aa	aA	aa

- A. La transmission de la pathologie semble autosomique récessive.
- B. L'individu II.5 peut être hétérozygote aA.
- C. Il s'agit probablement de mutations de novo apparues dans cette famille.
- D. L'individu III.1 a un risque de transmettre un allèle muté à sa descendance.
- E. Les parents du nouveau-né ont un risque de 50% de transmettre la maladie en cas de prochaine grossesse.

Ici, par l'étude des allèles présents chez les individus de la première génération (tous les eux AA) et chez l'individu II.3 qui est atteint, on se rend compte que la pathologie chez cet individu est due à une mutation de novo. Grâce à l'étude du génotype du nouveau-né étudié, on se rend compte que la pathologie est dominante puisqu'il n'a qu'un seul allèle muté.

Les parents du nouveau-né étant sains, ils sont donc forcément homozygotes AA, donc la mutation présente chez ce nouveau-né est de novo également.

A FAUX La pathologie est dominante car il ne faut qu'un seul allèle malade pour être atteint (III.2 atteint est aA).

B FAUX L'individu II.5 étant sain, il est homozygote AA.

C VRAI

D FAUX L'individu III.1 (donc la sœur de notre cas) est saine, elle est donc homozygote AA, elle n'a aucun risque de transmettre la maladie à sa descendance.

E FAUX On sait que les mutations présentes dans la famille sont de novo, les parents du nouveau-né n'ont pas la mutation donc pas de risque de la transmettre !