

Faculté des sciences de la nature et de la vie
Département de Biologie Physico-Chimique
Filière : Sciences Biologique
Spécialité : Génétique Fondamentale et Appliquée

UFR Sciences et Techniques
Département de Biologie
Mention : Biologie Santé
Parcours : CDG BioTher

Mémoire de fin de cycle
En vue de l'obtention du diplôme

Master

Thème

*Caractérisation de l'effet des variations de séquence sur l'épissage
de l'ARN du gène APC impliqué dans le syndrome FAP et la
prédisposition héréditaire au cancer colorectal*

Unité d'accueil : Inserm UMR 1245

« Génétique médicale et fonctionnelle du cancer et des maladies
neuropsychiatriques »

Directeur de l'unité : Pr Gaël Nicolas

Groupe « ARN et Maladies génétiques », Responsable : Dr Alexandra MARTINS
Faculté de Médecine et de Pharmacie, Université de Rouen

Présenté par : Aziza BOUKHEDAD

Soutenu le : 15 septembre 2022

Devant le Jury, composé de :

- M me D. ATMANI
- M me N. DEBBACHE.
- M me N. CHERAFT

Professeur

MCA

MCB

Présidente

Encadreur

Examineur

Remerciements

Tout d'abord je remercie le **bon dieu** "Allah", le tout Puissant, et le Miséricordieux, d'avoir toujours été là pour moi, de m'avoir donné les capacités et le courage à surmonter les obstacles que j'ai rencontrés durant ce parcours et pour achever ce travail, tout en prenant du plaisir et en faisant de belles rencontres. Et pour tout on lui rend grâce.

J'adresse mes remerciements également aux personnes qui m'ont aidé dans la réalisation de ce projet

Je présente mes sincères remerciements à l'équipe pédagogique pour leur collaboration à cette mobilité d'échange Erasmus+ notamment Mr Ali Aksas, Mlle Karima Slimani, et Mme Karima Ayouni, Mme Valérie Perrot, Mlle Anna Pilate, et Mr Olivier Boulin.

Je tiens à remercier le Professeur Gaëlle Nicola de m'avoir accueillie au sein de son unité de recherche Inserm U1245.

J'aimerais tout particulièrement remercier Mme Alexandra Martins pour sa confiance qui m'a permis de faire partie de son équipe, mais aussi pour sa gentillesse et ses conseils. Qu'elle sache qu'elle a contribué d'une grande partie à la réalisation de mon projet professionnel, en m'ouvrant les portes vers le monde de la recherche.

Je souhaiterais également présenter mes chaleureuses remerciements à ma promotrice Mme Nadjat Debbache, d'avoir toujours été à l'écoute, pour ses conseils, sa disponibilité, son soutien, ses encouragements, sa patience et sa bienveillance. Qu'elle trouve ici l'expression de ma profonde gratitude, car dieu m'a orienté vers elle.

Je remercie notamment Laëtitia Meulemans, pour tout ce qu'elle a pu m'apprendre durant mon stage, que ce soit les techniques, le savoir-faire et la réalisation des protocoles. Merci bien pour son temps consacré aux explications et manipulations, sa patience, ses conseils précieux, ses orientations et sa sympathie.

Mes remerciements vont également à tous les membres du laboratoire pour leur bonne humeur, et leur aide au quotidien.

Je tiens aussi à remercier tous les membres du jury, pour leurs temps et attention consacré à mon mémoire.

Et enfin, un grand merci à tous mes amis et tous les membres de ma famille que ce soit en Algérie ou à l'étrangers qui m'ont soutenu tout le long de ce parcours, que dieu vous récompense, et vous bénisse.

Dédicace

Avant toute chose je dois remercier dieu pour tous les biens qui m'a procuré ; la santé, la gratitude, la famille, les amis, les ambitions, la joie de vie, et surtout de m'avoir permis de sentir son existence et vouloir apprendre à le connaître.

Je dédie ce travail à mes très chers parents :

Je ne saurais pas expliquer ce que je ressens, les mots ne sont pas assez pour exprimer ma gratitude envers vous, je suis la plus chanceuse d'être votre fille, vous avez fait tout pour que je réussisse, devenir chercheur et m'épanouir. Je vous aime très forts.

***Ma mère** Nouara ; tu m'as toujours soutenu dans les moments les plus difficiles de ma vie, tu m'as supporté et tu m'as accepté comme je suis. Tes conseils, tes mots doux, et ton sourire radieux, gravés dans mon cœur sont ma source de force pour affronter les difficultés de la vie.*

***Mon père** Moussa ; tu es mon héros, et mon exemple. A travers nos conversations et sorties très chaleureuses et enrichissantes tu m'as transmis des valeurs et principes indispensables au fondement de ma vie. Tu as toujours cru en moi, Tu m'as encouragé, tu m'as appris à avoir confiance en moi, à être moi-même et surtout tu m'as guidé vers le bon chemin.*

***Ma grand-mère** Fatima un exemple parfait du courage et la persévérance, tu es toujours rayonnante malgré les circonstances. Tu étais toujours là pour maintenir et soutenir notre grande famille, et tu m'as tellement aidé à réussir, et arrivée là où je suis. Que dieu te garde pour nous tous.*

*A mon **frère** Youcef et ma **sœur** Sarah, avec qui j'ai partagé mon enfance et mes secrets. On a vécu ensemble des moments inoubliables.*

*A tous mes **oncles**, mes **tantes**, mes **cousins** et mes **cousines** avec leurs femmes, époux et enfants. Et à tous mes **amies**, notamment pour cette année mes tentes Hayette et Mabrouka, mes cousines Amina, Amel, Fafounii, Anissa et Nassima, mes cousins Samir, Yacine, Ghilas, Karim et Salim, mes amies Clémance, Souad, Souhila, Somia, Frédérique, Lamia, Mina et Wissam.*

A tous ceux qui m'ont aidé de loin ou de près.

Que dieu vous bénisse et illumine votre vie comme vous m'avez illuminé la mienne.

Aziza

Résumé

Aujourd'hui l'un des enjeux majeurs en génétique médicale n'est plus la détection des variations dans le génome des patients mais leur interprétation biologique et clinique afin de mieux prendre en charge des patients et leurs apparentées pour les maladies monogéniques à transmission mendélienne. Dans le cas d'*APC*, un gène suppresseur de tumeur se manifeste par le syndrome FAP, et prédispose au cancer colorectal, 9563 variations sont répertoriées dans la base de donnée ClinVar et l'impact clinique de 58% d'entre elles est inconnu.

Cette étude a été réalisée dans le but de caractériser l'effet de certaines variations exoniques susceptibles d'affecter les ESR, et variations IVS -1+1+2 des sites 3'ss 5'ss d'épissage, sur le pré-ARNm du gène *APC*, afin de contribuer à leur classification selon les recommandations ACMG.

La méthode *in silico* (bases de données et logiciels de prédictions bio-informatique) nous a servie pour la sélection des variations et les exons d'intérêts 5, 6, et 11. De plus, l'effet de ces variations a été étudié par le test indicateur d'anomalie d'épissage minigène.

La fiabilité a été confirmée de certains prédicteurs bio-informatique sur l'effet de certaines variations notamment SpliceAI mais des améliorations sont primordiales pour des interprétations cliniques. Ce jeu de données a également permis de constater que l'interprétation des variations ne peut pas être généralisée par rapport au même type et la localisation. Chaque variation doit être étudiée séparément dans son contexte, car certaines variations +2TC n'ont aucun effet sur l'épissage, donc des analyses supplémentaires seront nécessaires afin de confirmer les résultats obtenus.

Mots clés : APC, VUS, minigène pCAS2, outils bio-informatiques, classification ACMG

Abstract

Today, one of the major challenges in medical genetics is no longer the detection of variations in the genome of patients but their biological and clinical interpretation in order to better manage patients and their relatives for monogenic diseases with Mendelian transmission. In the case of APC, a tumor suppressor gene manifesting as FAP syndrome, and predisposing to colorectal cancer, 9563 variations are listed in the ClinVar database and the clinical impact of 58% of them is unknown.

This study was carried out with the aim of characterizing the effect of certain exonic variations likely to affect the ESRs, and IVS -1+1+2 variations of the 3'ss 5'ss splice sites, on the pre-mRNA of the APC gene, in order to contribute to their classification according to ACMG recommendations.

The in silico method (databases and bioinformatics prediction software) was used to select the variations and the exons of interest 5, 6, and 11. In addition, the effect of these variations was examined by the minigene splicing anomaly indicator test.

The reliability has been revealed of certain bioinformatics predictors on the effect of certain variations, in particular SpliceAI, but improvements are essential for clinical interpretations. This dataset also showed that the interpretation of variations cannot be generalized with respect to the same type and location. Each variation must be studied separately in its context, because some +2TC variations have no effect on splicing, so additional analyzes will be necessary in order to confirm the results obtained.

Keywords: APC, VUS, pCAS2 minigene, bioinformatics tools, ACMG classification

Sommaire

I- INTRODUCTION	1
II-SYNTHESE BIBLIOGRAPHIQUE.....	2-18
I-1. Le cancer.....	2
I-2. Les mutations	2
I-3. Les variations géniques	5
I-4. L'épissage	11
I-5. APC	13
III- MATÉRIELS ET MÉTHODES	19-31
II-1. Matériel	19
II-1.1 In silico.....	19
II-1.2 Test fonctionnel d'épissage (minigène pCAS2)	21
II-2. Méthode.....	22
II-2.1 Analyses <i>in silico</i>	20
II-2.1 Test fonctionnel (minigène).....	23
III- RÉSULTATS	32-48
III-1 Analyses in silico.....	32
III-1.1 Sélection des variations.....	32
III-1.2 Bases de données.....	34
III-1.3 Prédictions bio-informatiques.....	35
III-2. Test fonctionnel minigène.....	40
III-2.1 RT-PCR semi quantitative.....	40
III-2.2 RT-PCR Fluo quantitative.....	45
IV- DISCUSSION	49-53
V- CONCLUSION	54-55
VI- RÉFÉRENCES BIBLIOGRAPHIQUES	

Liste des tableaux

Tableau 1 : Indication de la séquence de référence.....	4
Tableau 2 : Les codes qui représentent les variations de séquences sur l'ADN (Huret.J.L, 2008).....	6
Tableau 3 : La taille en pb et AA des exons de APC avec la taille en pb des introns en amont et aval.(d'après Alamut)	14
Tableau 4 : les variations d'APC sélectionnées pour le test fonctionnel.....	32
Tableau 5 : Les informations collectées sur les variations sélectionnées	34
Tableau 6: Tableau récapitulatif des prédictions Splice AI Look up sur les scores et les positions des sites concernés (accepteur et donneur) pour chaque variation d'intérêt.....	38
Tableau 7: Le score le Δ tESRseq des variations exoniques d'intérêts à base du logiciel HexoSplice	40
Tableau 8: Résultats issus de l'analyse des alignements après séquençage	44

Liste des figures

Figure 1 : séquences de référence d'après HGVS.....	5
Figure 2 : Règles de nomenclature des variants, nomenclature officielle de la HGVS. Modifiée à partir (Escande.F, 2015).....	7
Figure 3 : organisation des critères de classification des variations en tant que pathogène ou bénigne selon les recommandations ACMG (Richards.S e. a., 2015)	9
Figure 4 : Représentation schématique des signaux et protéines impliqués dans le mécanisme d'épissage.....	11
Figure 5 : Représentation schématique sur les défauts d'épissage engendrés par les variations de séquences.....	12
Figure 6 : Représentation schématique dimensionnelle de la composition du gène APC en exons et introns, tirée à partir du serveur Alamut visual	13
Figure 7: Représentation schématique des exons et introns du gène APC, qui met en avant la notion In frame et Out of frame tirée à partir de UMD-APC	14
Figure 8: Corrélation phénotype-génotype du gène APC.....	17
Figure 9: Présentation schématique de l'endroit de fixation des amorces sur le vecteur pCAS2	22
Figure 10: Principe du test fonctionnel indicateur d'anomalies d'épissage, basé sur l'utilisation du minigène pCAS2	23
Figure 11: La taille des inserts amplifiés par PCR par le biais des amorces InFusion	24
Figure 12: Migration des produits PCR sur gel d'agarose à 1.5%.....	25
Figure 13: Représentation schématique des sites de digestion sur pCAS2 par BamHI et MluI, la migration sur gel d'agarose du pCAS2 digéré et non digéré	25
Figure 14: Représentation schématique du clonage de l'insert dans le vecteur pCAS2 digéré par recombinaison homologue	26
Figure 15: Principe de la mutagenèse dirigée : PCRI et PCRII	27
Figure 16: La migration sur gel d'agarose à 1.5% des produits PCRI	28
Figure 17: La migration sur gel d'agarose des produits de PCRII.....	28
Figure 18: Courbe d'amplification d'un fragment cible par PCR en temps réel	31
Figure 19: Description de la localisation des variations sélectionnées dans l'exon 5, 6 et 11 d'APC, et leurs régions introniques flanquantes.	33

Figure 20 : Diagramme qui représente la force des sites d'épissage 5'ss et 3'ss pour chaque exon d'APC tirées à partir de SSF-L et MaxEnt dans Alamut	35
Figure 21: Courbe qui représente la force des sites accepteurs 3' et donneurs 5' constitutifs du gène APC basés sur les logiciels de prédiction SSF-L et MaxEnt tirés à partir de l'interface Alamut	36
Figure 22: Diagramme qui représente le score percentile constitutif des sites donneur 5' et accepteur 3' d'APC tirés à partir de MaxEntScan SROOGLE.....	37
Figure 23: Histogramme qui représente la force des sites sélectionnés en état sauvage WT ou muté (variant) tirées par HSF.....	38
Figure 24: l'effet sur l'épissage des 12 variations, sélectionnées dans l'exon 5, 6, 11 de APC, analysées à l'aide du test fonctionnel minigène.....	43
Figure 25: Séparation des produits RT-PCR fluo (profils d'épissage) par électrophorèse sur gel d'agarose 2%	46
Figure 26: Quantification relative, par RT-PCR fluorescente, des transcrits produits à partir des minigènes « APC exon 5, 6 et 11 » sauvage et mutés (variations +1 et +2)	47

Liste des abréviations

ADNg : *Acide DésoxyriboNucléique génomique*
ADNc : *Acide DésoxyriboNucléique complémentaire*
ARNm : *Acide RiboNucléique messenger*
APC: *adenomatous polyposis coli*
BET : *Bromure d'EThylum*
CMV : *CytoMégaloVirus*
dNTP : *desoxyNucleotide TriPhosphate*
DSB : *Double-Strand Break*
ESE : *Exonic Splicing Enhancer*
ESR : *Exonic Splicing Regulatory elements*
ESRseq : *Exonic Splicing Regulator Sequence*
ESS : *Exonic Splicing Silencer*
FAP: *Familial adenomatous polyposis*
FL : *Full-length (« pleine-longueur »)*
ISE : *Intronic Splicing Enhancer*
ISS : *Intronic Splicing Silencer*
IVS: *intervening sequence (intron)*
MES : *MaxEntScan*
PCR : *Polymerase Chain Reaction*
pré-ARNm : *Acide RiboNucléique pré-messenger*
RT-PCR : *Reverse Transcription-Polymerase Chain Reaction*
SR : *Serine aRginine-rich*
SNV : *Single Nucleotide Variant*
VSI : *Variation de Signification Inconnue*
WT : *Wild Type*
5'/3'ss : *5'/3' splice site*
 Δ ESRseq : *total ESRseq score change*

INTRODUCTION

I. Introduction

La génétique humaine a pu relever plusieurs défis ces dernières années, telle que la détection des variations dans le génome des patients et la mise en évidence de la très grande variabilité interindividuelle du génome humain, et ce grâce à la révolution technologique et la mise à disposition de nouvelles techniques de séquençage de l'ADN à haut débit. Mais l'interprétation biologique et clinique de ces variations reste un défi majeur en génétique médicale (Richards.S et al, 2015).

Les individus qui sont porteurs d'une altération causale au niveau constitutionnelle (germinale) présentent un risque plus élevé de développer un cancer à un âge précoce par rapport à la population générale. Il est donc essentiel d'identifier et de caractériser les gènes de prédisposition aux cancers et de classer les différentes variations identifiées dans ces gènes afin d'optimiser la prise en charge des patients et de leurs apparentés (Gaildrat.P, 2021).

Selon l'institut national du cancer, les recherches en oncogénétique ont permis d'identifier plus de 100 gènes de prédisposition à certains cancers (Rahman.N, 2014), tel que le gène *APC* (*Adenomatous Polyposis Coli*), un gène suppresseur de tumeur (Polakis.P, 1995).

La protéine APC exerce plusieurs fonctions au sein de la cellule : elle régule la signalisation WNT en régulant la stabilité et la localisation de la β catenine, l'adhésion et la migration cellulaire, l'assemblage des microtubules et la ségrégation des chromosomes (Fodde.R, 2002). Par conséquent des altérations génétiques dans des régions spécifiques de ce gène peuvent engendrer la perte des fonctions d'APC essentielles au bon fonctionnement cellulaire. Elles se manifestent cliniquement par le syndrome FAP (Familial Adenomatous Polyposis). C'est un syndrome à transmission autosomique dominante, caractérisé par la survenue de polypes dans le colon et le rectum à un âge précoce et prédispose au CCR (Cancer ColoRectal) (Leoz.M.R, 2015).

Aujourd'hui, 9563 variations nucléotidiques d'APC sont répertoriées dans la base de données ClinVar (ClinVar APC). Parmi ces variations, 5613 (58%) ont été classées comme des variations à signification inconnue (IVS : leur impact biologique et clinique reste à déterminer).

Le but de ce projet de recherche est de caractériser les effets des variations à l'origine des substitutions introniques localisées en position -1, +1, +2 des exons, 5, 6, et 11 d'APC et également une substitution exonique de ces exons sur l'épissage du pré ARNm d'APC. En utilisant le test fonctionnel Minigène et les prédictions bio-informatiques, et ce afin de contribuer à la classification de ses variations selon les recommandations ACMG.

SYNTHESE

BIBLIOGRAPHIQUE

II. Synthèse bibliographique

II.1 Le cancer

II.1.1 Définition

Le cancer englobe un groupe hétérogène de maladies, qui se manifeste par le développement anarchique des cellules anormales au sein de l'organisme. Ce dérèglement au niveau cellulaire est lié aux changements qui surviennent dans le génome des cellules. Ces mutations sont la conséquence des interactions entre plusieurs facteurs de risque : facteurs externes (environnementaux, mode de vie, infections), facteurs internes (facteurs génétiques, dérèglement des hormones et système immunitaire..) (Le Cancer, 2016).

Le cancer est considéré comme un problème majeur de santé publique, car il est l'une des principales causes de mortalité dans le monde à l'origine de près de 10 millions de décès en 2020 (Ferlay.J, 2020) d'où la prévention, l'innovation du diagnostique et thérapeutique sont une nécessité au niveau mondiale car plusieurs cancers peuvent être guéris dans le cas où le diagnostic est fait à des stades primaires suivi d'un traitement efficace (Cancer, 2022).

II.1.2 Classification

Le cancer est classé selon son origine en deux groupes :

II.1.2.1 Sporadiques : dit aussi somatiques, représentent 85 à 90%, ils sont caractérisés par la présence des altérations génétiques au niveau de la tumeur, accumulées tout au long de la vie (vieillessement, facteurs environnementaux).

II.1.2 .2 Génétiques : représentent 5 à 10%, caractérisé par la présence de l'altération causale dès la naissance, soit en constitutionnelle (pré-zygotique) héritée par les parents, ou mosaïque (post zygotique) variation de novo .

II.2 Les mutations

II.2.1 Définition

Elles sont le résultat d'un processus biologique qui conduit à la modification de la séquence d'un génome, elle résulte de l'action d'agents physiques, chimiques, ou d'erreurs de réplifications (Kebir.S, 2015-2016).

II.2.2 Classification

Les mutations peuvent être classées selon différents critères

II.2.2.1. Selon la localisation cellulaire et chromosomique: (Kebir.S, 2015-2016)

- Mutation germinale (constitutionnelle) : mutation présente dans les cellules germinales avant la fécondation (cellules souches d'un gamète), elle est héritée par les parents et sera transmissible à la descendance.
- Mutation somatique (acquise) : elle apparue dans une cellule somatique, peut être à l'origine d'un clone cellulaire porteurs de cette mutation, et les mutations ne seront pas transmissible à la descendance.
- Mutation autosomique : survient dans les gènes situés sur les autosomes.
- Mutation liée au sexe : survient dans des gènes situés sur les chromosomes X/Y.

II.2.2.2 Selon leurs effets biologiques

La conséquence d'une mutation peut être très variable. La nature de la séquence touchée, l'ampleur de la modification génétique, le contexte biologique, génétique, environnemental, sont autant d'éléments qui déterminent les effets de ces changements (Bodart.J.F, 2007).

Mutation perte de fonction, Mutation hypomorphe, Mutation hypermorphe, Mutation gain de fonction, Mutation conditionnelle, Mutation létale, Mutation suppressive, Mutation neutre (Chellat-Rezgoune, 2021) (Kebir.S, 2015-2016).

II.2.2.3 Selon la taille du changement

Les mutations chromosomiques, et les mutations géniques (ponctuelle), (Chellat-Rezgoune, 2021).

II.2.2.4 Selon la nature de changement moléculaire : (Chellat-Rezgoune, 2021)

Sont des mutations ponctuelles (variation de séquence = variant allélique) : concerne le changement de une ou quelques positions adjacentes de l'ADN, il existe 2 grands types :

Les substitutions (mutation silencieuse, faux sens, non-sens), et les Indel : insertion ou délétion, décalage de cadre de lecture : frameshift.

II.2.3. L'impact phénotypique des variations génétiques

Les mutations sont la source de la diversité entre individus, mais elles sont aussi à l'origine des maladies génétiques monogéniques, et des prédispositions génétiques aux maladies multifactorielles. Les conséquences de toute mutation dépendent de son effet fonctionnel, qui

peut être neutre, conduit à l'amélioration d'une fonction (diversité, évolution) ou à son altération (effet pathogène) (Kebir.S, 2015-2016).

II.3 Les variations de séquences

II.3.1 Terminologie

Le terme « mutation » et « polymorphisme » MAF>1%, ont créé des confusions, car ils réfèrent à une classification par rapport à l'effet pathogène et bénin respectivement. Afin de palier à ce problème il a été recommandé de remplacer ces deux termes par « variant » (Richards.S, et al, 2015) qui réfère à tout changement de la séquence d'ADN de référence, sans préjuger de sa conséquence fonctionnelle.

Le changement de la séquence d'ADN peut concerner un seul nucléotide (SNV), 2 à 50pb (Indels), ou bien des variations structurales >50pb (duplication en tandem, duplication intrinsèque, inversion, translocation et variation du nombre de copie CNV) (Gaildrat.P, 2021).

II.3.2 Nomenclature

L'uniformisation de la nomenclature des variations est définie et maintenue par le consortium Human Genome Variation Society (HGVS) en établissant des recommandations à base des critères standardisées au niveau international afin d'assurer une description claire des variations de séquence, pouvoir les partager mais aussi les utiliser en aval pour des diagnostics moléculaires (Den Dunnen.J.T, 2016).

II.3.2.1 La séquence de référence

Tableau 1 : Indication de la séquence de référence.

	Codant	c.	NM_
ADN	Génomique	g.	NG_
	Mitochondrial	m.	
ARN		r.	
Protéine		p.	NP_

La description des variations doit toujours être faite au niveau de l'ADN en lien avec une séquence de référence qui peut être :

- Une séquence génomique, qui est symbolisée par le préfixe « g. » précède la description du variant. Elle est précise car elle fait référence au chromosome et à la version du génome (GRCh37/hg19 et GRCh38/hg38), utilisée par la bio-informatique pour le traitement des résultats de séquençage à haut débit.
- Une séquence codante du transcrit du gène étudié (ADNc), précède la description du variant par le préfixe « c. ». Il faut bien préciser la séquence de référence choisie car un même gène peut avoir plusieurs transcrits, et le choix se portera généralement sur le transcrit principal, qui est souvent le plus abondant dans les cellules.

Elles proviennent de la base de données RefSeqGene (Reference Sequences Gene), du national center of Biotechnology Information (NCBI) <http://www.ncbi.nlm.nih.gov/refseq/rsg/>

- Une séquence protéique, qui décrit les variants par le préfixe « p. », mais pour des raisons de précision, les variants doivent être décrits au niveau de l'ADN, car un même AA peut être le résultat de différents variants nucléotidiques. (Escande.F, 2015).

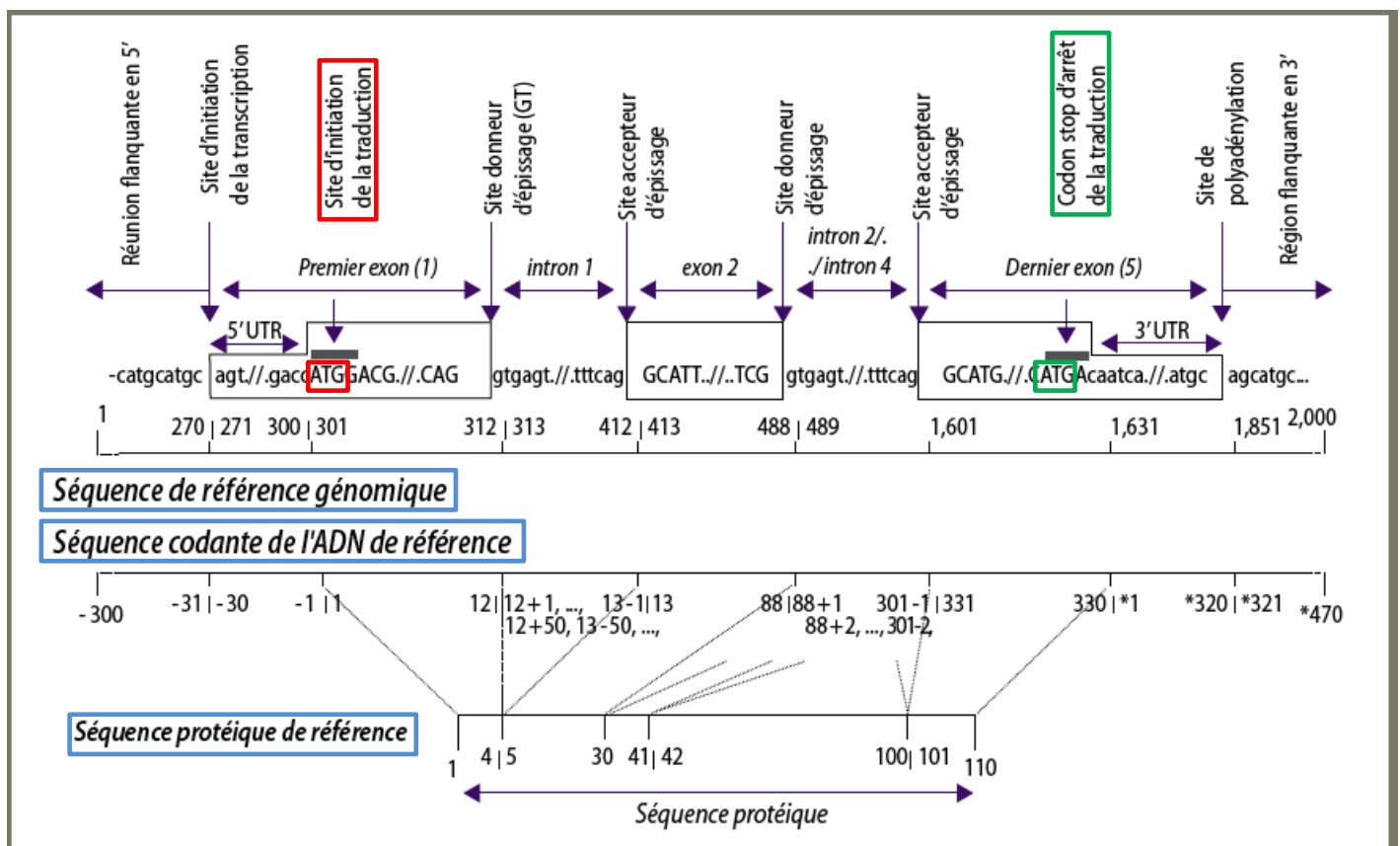


Figure 1 : Séquences de référence d'après HGVS (Escande.F, 2015).

Sur la séquence gnomique :

- Désigne les nucléotides par rapport à leur position sur le chromosome

Sur l'ADNc :

- le premier chiffre 1 est toujours référé au premier nucléotide A, du codon de l'initiation ATG
- La partie en amont du codon d'initiation de la traduction (5'UTR) est notée -1, -2, ect
- Les nucléotides situés au début des introns (IVS : InterVening Sequence) seront numérotés : le chiffre du dernier nucléotide de l'exon +1, +2, ect. Et ceux situés à la fin de l'intron : le chiffre du premier nucléotide de l'exon en aval -1, -2, ect.
- Le dernier nucléotide du codon de terminaison de la traduction portera le dernier chiffre et la séquence en aval (3'UTR) sera numérotée *1, *2, ect

Sur séquence protéique :

- Le premier codon de traduction est numéroté 1

II.3.3 Les codes utilisés pour décrire les variations:

→ Au niveau de l'ADN :

Tableau 2 : Les codes qui représentent les variations de séquences sur l'ADN (Huret.J.L, 2008)

Substitution	>	Inversion	Inv	Translocation	t
Etendue	-	Conversion	Con	Mosaïque	,
Délétion	del	Extension	Ext	Allèle	[]
Insertion	ins	Codon stop	X	Incertain	()
		Décalage de		Autres	
Duplication	dup	cadre de lecture	fsX	changements sur l'allèle	;

- Exemples de nomenclatures de variations nucléotidiques :

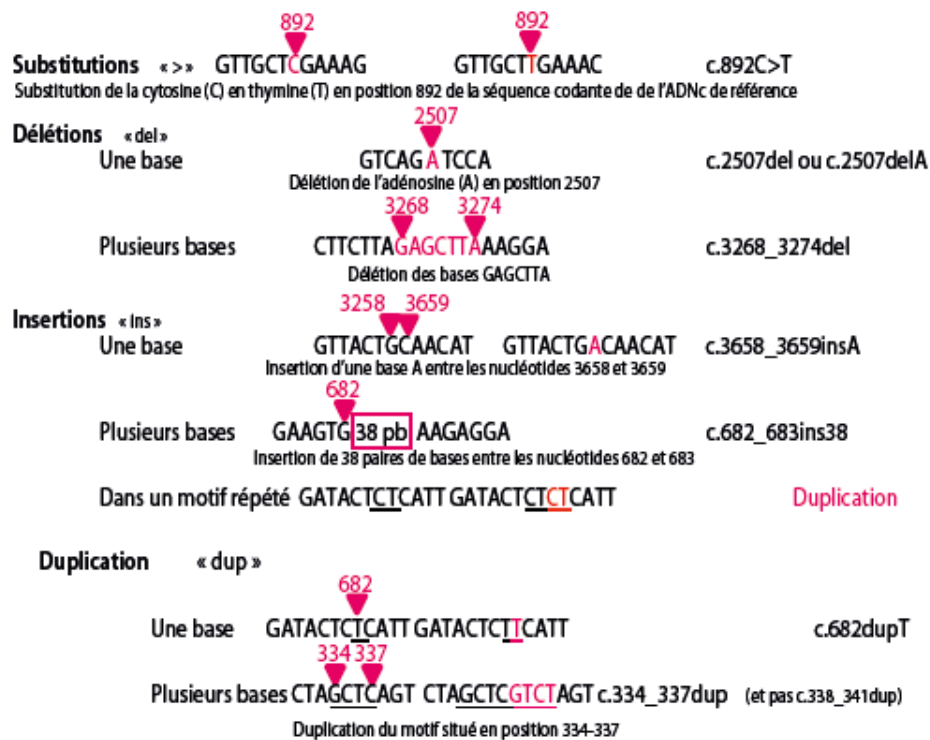


Figure 2 : Règles de nomenclature des variants, nomenclature officielle de la HGVS. Modifiée à partir (Escande.F, 2015).

II.3.4 L'effet des variations géniques au niveau moléculaire

Ces variations géniques peuvent modifier plusieurs processus moléculaires tel que :

II.3.4 .1 La régulation de la transcription (éléments cis-regulateurs : séquence de l'ADN ou des facteurs de transcription se lient) ce qui affecte le taux de l'ARN produit.

II.3.4 .2 L'épissage (les ESR, ISR, le site accepteur 5' et donneur 3', site d'embranchement) ce qui affecte la maturation des ARNm : élimination des introns et assemblage des exons.

- Destruction d'un site naturel d'épissage (5'/3' ss)
- Création d'un nouveau site d'épissage (5'/3' ss)
- Altération d'un ou plusieurs éléments de régulation (SRE, e.g. ESE/ESS)
- Altération des sites de branchement

II.3.4 .3 La traduction (la conformation de l'ARNm en trois dimensions : l'endroit de liaison des facteurs d'initiation, ou bien un décalage de cadre de lecture : un codon stop prématuré)

Ce qui affecte la quantité 'dégradation par NMD' et la qualité de la séquence nucléotidique respectivement.

II.3.4 .4 La conformation spatiale de la protéine (changement de la nature des acides aminés, protéine tronquée, protéine instable) ce qui affecte la fonction de la protéine en question (activité réduite, nulle, ou continue)

II.3.5 Classification selon les recommandations ACMG (American College of Medical Genetics): (Richards.S, et al, 2015)

Les variations localisées dans des gènes impliqués dans des maladies mendéliennes sont classées en 5 catégories :

classe	Type de variation
5	Pathogène
4	Probablement pathogène
3	Variation à signification inconnue (VUS)
2	probablement bénigne
1	Bénigne

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very strong
Population data	MAF is too high for disorder BA1/BS1 OR observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in <i>trans</i> with a dominant variant BP2 Observed in <i>cis</i> with a pathogenic variant BP2		For recessive disorders, detected in <i>trans</i> with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

Figure 3 : organisation des critères de classification des variations en tant que pathogène ou bénigne selon les recommandations ACMG (Richards.S, et al, 2015)

BS= benign strong/ BP= benign supporting/ FH= family history/ LOF= loss of function/ MAF= minor allele frequency/ path= pathogenic/ PM= pathogenic moderate/ PP= pathogenic supporting/ PS=pathogenic strong/ PVS= pathogenic very strong.

D'après l'ACMG et l'AMP (Association for Molecular Pathology), 6 aspects doivent être pris en compte pour pouvoir interpréter les variations de séquences afin de les classer :

II.3.5.1 La fréquence allélique dans la population générale :

La fréquence allélique, MAF (Minor Allele Frequency) de la variation dans la population générale doit être compatible avec la prévalence de la maladie. (Manolio.T.A, 2009)

Les maladies mendéliennes sont généralement rare par conséquent la MAF de la variation en question dans la population générale doit être <1%, car si la MAF est ≥1% la variation sera ignorée et sera considérée comme polymorphisme.

La MAF, peut être consultée à travers les différentes bases de données : dbSNP, gnomAD...

II.3.5.2 La fréquence dans les cohortes des malades :

Comparaison entre des cohortes des cas vs témoin, sur la présence ou l'absence de la variation en question. Ces cohortes doivent être bien choisies en tenant compte de : taille, l'origine, le phénotype, la pénétrance (complète ou partielle) et l'âge d'apparition (précoce ou tardé).

II.3.5.3 La co-ségrégation de la variation dans les familles des patients :

L'analyse de l'arbre généalogique de plusieurs patients en prenant compte de la pénétrance et l'âge d'apparition

II.3.5.4 Prédiction sur la base du type de variation :

Il y a des types de variations qui sont prédites perte de fonction du fait leurs nature :

Variation nonsense, et les variations qui crée un décalage de cadre de lecture (indel non multiple de 3), car elles conduisent à l'apparition des codons stops prématurés, et ces derniers ont un effet perte de fonction de la protéine. De ce fait, si ces variations sont localisées dans des gènes impliqués dans une maladie monogénique, elles sont classés pathogènes.

Les IVS±1/2 sont généralement classées pathogènes car elle conduise à un saut d'exon, délétion de quelques nucléotides exonique ou bien une rétention intronique.

II.3.5.5 Prédiction bio-informatiques de l'impact biologique de la variation :

Ses logiciels de prédictions bio-informatique sont interpellés en cas des variations faux-sens (changement d'AA) de signification inconnue (VUS) qui sont susceptible d'engendrer des défauts d'épissage.

Ces logiciels sont basés sur le degré de conservation de l'AA d'une part et d'une autre part sur les caractéristiques biochimiques de l'AA.

Ces prédictions bio-informatiques ont une fiabilité variable mais sont utilisées à fin de sélectionner des variations pour les tests fonctionnels (Jinchen.L, 2018).

(ESRseq scores, MMssplice, MaxEntScan, Human Splice Finder, Splice AI...)

II.3.5.6 Données fonctionnelles :

La grande majorité des variations affectent l'épissage et ça engendre généralement des

ARNm aberrants à cause d'un frameshift de plus souvent.

Parmi les tests fonctionnels indicateurs de défaut d'épissage :

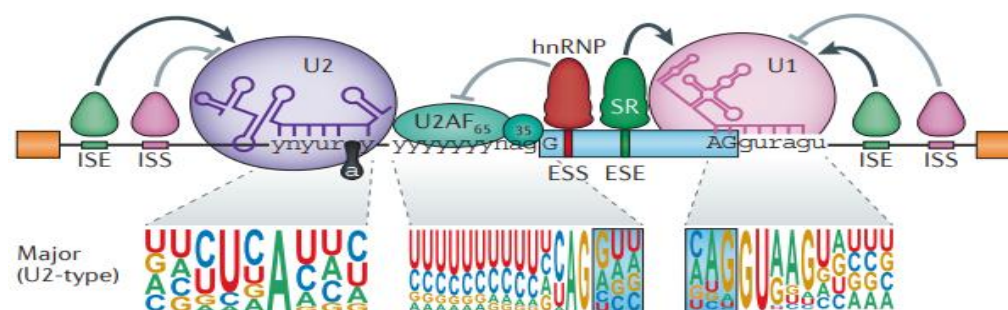
Analyse par RT-PCR sur ARN du sang ou lignée lymphoblastoïde de patient, en cas de non disponibilité d'un échantillon de patient, on peut utiliser le test basé sur l'utilisation de minigène.

Plusieurs tests ont été développés pour évaluer la fonctionnalité de d'une protéine, soit dans son contexte biologique complet ou sur une caractéristique fonctionnelle spécifique.

II.4 L'épissage :

II.4.1 Définition

C'est un mécanisme de régulation post-transcriptionnel, et il est considéré comme une étape clés pour la maturation des pré-ARNm, permettant l'excision des introns et l'assemblage des exons.



(Scotti.M.M, 2016)

Figure 4 : Représentation schématique des signaux et protéines impliqués dans le mécanisme d'épissage

Ce processus met en jeu, un complexe moléculaire dynamique le spliceosome, qui est constitué de 5 ribonucléoprotéines nucléaires (les snRNP U1, U2, U4-U6) et plus de 200 protéines. (Scotti.M.M, 2016)

Cette machinerie, qui a comme rôle, d'une part la catalyse des réactions de trans-estérification, et d'autre part le maintien de la structure du transcrit, reconnaît des endroits spécifiques sur la séquence du pré-ARNm, on distingue :

- les sites d'épissage : le site donneur 5' ss, le site accepteur 3'ss, et le site de branchement (région riche en pyrimidines) permettent la reconnaissance des jonctions exon/intron.
- les éléments de régulation de l'épissage : les éléments répresseurs exoniques (ESS, Exonic Splicing Silencer) ou introniques (ISS, Intronic Splicing Silencer) qui lient les protéines hnRNP (heterogeneous nuclear RiboNucleo Proteins), et les éléments stimulateurs exoniques (ESE, Exonic Splicing Enhancer) ou introniques (ISE, Intronic Splicing Enhancer) qui constituent des sites de fixation pour les protéines SR (Serine/Arginine-rich), permettant ainsi la modulation et le recrutement du spliceosome. (Scotti.M.M, 2016)

Les séquences du 5'ss et 3'ss sont très conservées notamment sur les di-nucléotides situés aux extrémités 5' et 3' des introns : IVS -1+2 GT et IVS+1+2 AG.

II.4.2 Les défauts d'épissage induits par des variations de séquence :

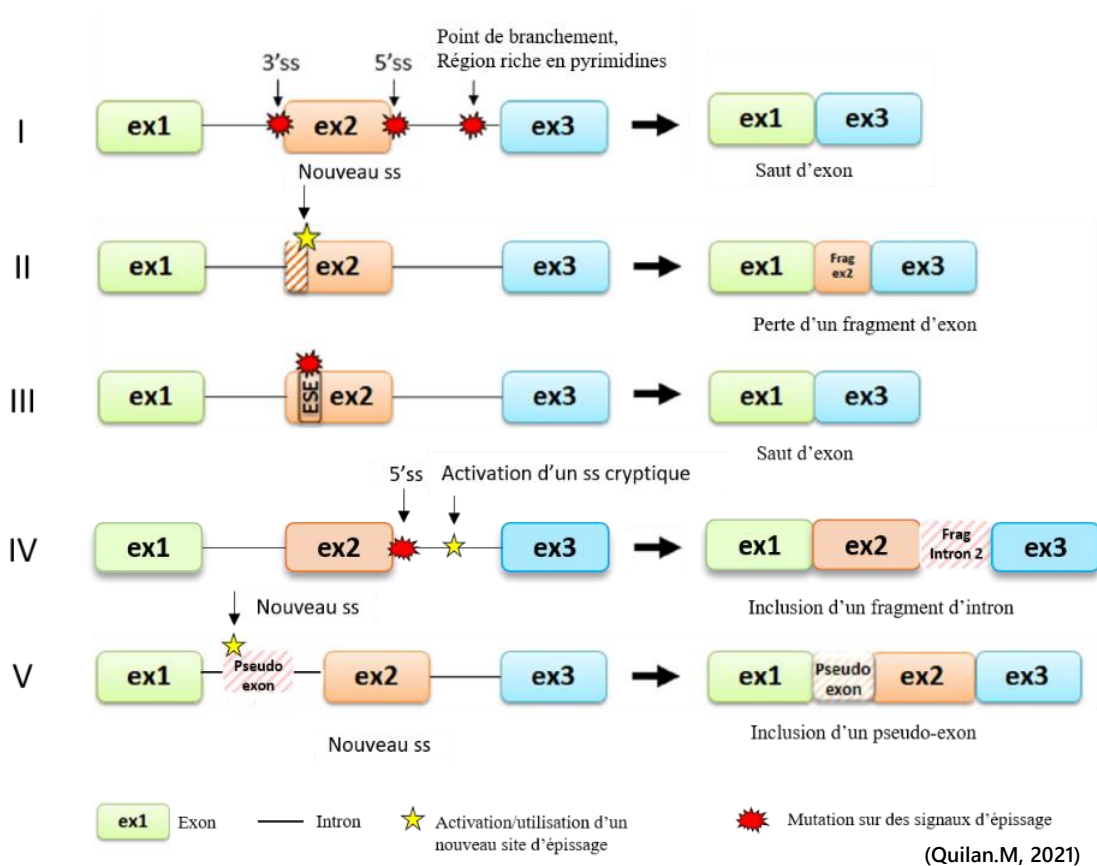


Figure 5 : Représentation schématique sur les défauts d'épissage engendrés par les variations de séquences

Selon (Wimmer et al, 2007) dans (Abramowicz et Gos., 2018)

Les mutations qui altèrent l'épissage peuvent être divisées en 5 catégories :

(I)Mutations qui affectent les signaux principaux d'épissage. (II)Mutations qui créent des nouveaux sites d'épissages. (III)Les mutations qui affectent les éléments de régulations d'épissage. (IV)Les mutations qui activent des sites cryptiques. (V)Les mutation introniques.

II.4.3 Les approches expérimentales de détection des anomalies d'épissage: (Meulemans.L, 2021)

II.4.3.1 Analyses expérimentales dans un contexte artificiel :

- Test indicateur d'anomalies d'épissage basé sur l'utilisation de minigènes
- Test indicateur d'anomalies d'épissage basé sur l'utilisation de transgènes ou de modifications du génome

II.4.3.2 Analyses expérimentales à partir du matériel biologique des patients:

- Analyse des profils d'épissage des transcrits par RT-PCR
- Analyse de l'expression allélique
- Analyses des défauts d'épissage par la technique de RNA-seq

II.5 APC (adenomatous polyposis coli)

II.5.1 Définition :

C'est un gène suppresseur de tumeur, localisé sur le chromosome 5 brin q en position 22.2, il est composé de 16 exons (-1...15), et sa protéine comporte 2846 acides aminés (AA) de 310KDA (Leoz.M.R, 2015)

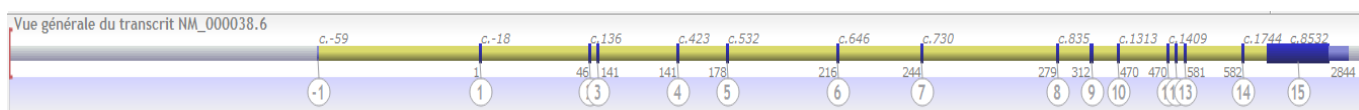


Figure 6 : Représentation schématique dimensionnelle de la composition du gène APC en exons et introns, tirée à partir du serveur Alamut visual ([NM_000038.6 Hg19](#))

Tableau 3 : La taille en pb et AA des exons de APC avec la taille en pb des introns en amont et aval.(d'après Alamut)

Exon	In frame or out of frame	Taille (pb)	Taille (AA)	Intron en avant	Intron en arrière
-1	5'UTR	41	0		16947
1	5'UTR+Start	153	45	16947	11300
2	OUT	85	29	11300	778
3	OUT	202	68	778	8238
4	OUT	109	37	8238	5052
5	IN	114	38	5052	11542
6	IN	84	28	11542	8749
7	IN	105	35	8749	14111
8	IN	99	33	14111	3372
9	OUT	379	127	3372	2551
10	IN	96	33	2551	5116
11	OUT	140	47	5116	681
12	IN	78	26	681	849
13	IN	117	39	849	5978
14	OUT	215	72	5978	2387
15	Stop	6574	2192	2387	

NB :

La notion de In frame et Out of frame

Nombre de pb constituant l'exon /3, si le nombre est entier (multiple de 3) donc l'exon est dit In frame, si ce n'ai pas un nombre entier l'exon est Out of frame

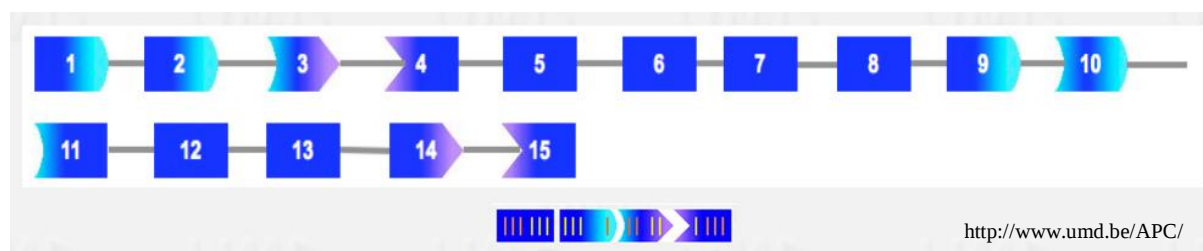


Figure 7: Représentation schématique des exons et introns du gène APC, qui met en avant la notion In frame et Out of frame tirée à partir de UMD-APC

La protéine APC est multifonctionnelle et se trouve sous plusieurs isoformes dans la cellule, elle contient des domaines et motifs AA qui lui permettent de s'oligomériser et interagir avec plusieurs protéines (Leoz.M.R, 2015). L'activité d'APC est corrélée avec son niveau de phosphorylation (Bio-informatique)

II.5.2 Les rôles intracellulaire de la protéine APC :

Le gène *APC* code pour une protéine multifonctionnelle, qui peut participer à plusieurs processus cellulaire : la transduction du signal (voie WNT), l'assemblage des microtubules et la ségrégation des chromosomes, l'adhésion et la migration cellulaire. (Fodde.R, 2002)

II.5.2.1 La régulation négative de la voie WNT: cette régulation se fait à travers la dégradation de la β caténine par la protéine APC associée à un complexe destructif : Axin/Axin2, GSK-3b et CK1 dans le cytoplasme (Neo.O, 2021). En effet en absence de l'activité d'APC, la protéine β caténine s'accumule et se transloque dans le noyau, et interagit avec d'autres facteurs de transcription (FT) permettant ainsi l'expression de plusieurs oncogènes impliqués dans la : prolifération, migration, progression et apoptose. (Leoz.M.R , 2015)

II.5.2.2 La stabilisation des microtubules : la protéine APC s'associe avec d'autres protéines et exerce:

- La régulation de la duplication des centrosomes par l'interaction entre les centrosomes et les tubulines.
- La fixation correcte du fuseau mitotique aux kinétochores des chromosomes. (Fodde.R, 2002)

En effet ces deux fonctions assurent l'assemblage et la stabilité des microtubules ce qui conduit à la stabilisation des chromosomes. En effet une APC inactive implique un défaut de ségrégation des chromosomes. (Leoz.M.R, 2015)

II.5.2.3 Le contrôle l'adhésion cellulaire :

La protéine APC régule la β caténine, et cette dernière non seulement elle est l'un des modulateurs de la voie WNT mais elle est aussi un composant essentiel pour l'adhérence des jonctions. Elle lie entre les Ecadherines et α caténine, et entre les actines et leurs protéines associées (Ben-Ze'ev.A, 1998)

II.5.2.4 la régulation du cycle cellulaire :

La protéine APC intervient principalement durant la transition G1/S et G2/M. (He.T.C, 1998), (Rubinfeld.B, 1996). La régulation au niveau du G1/S se fait grâce à l'activation de la cycline D1 par la β caténine sous forme inhibée (Heinen.C.D, 2002). La phase G2/M, c'est l'état de phosphorylation d'APC qui joue un rôle dans cette transition, en effet APC non phosphorylée inhibe la progression de G2 en M, mais dans son état phosphorylé elle permet la progression de la mitose. (Neo.O, 2021)

II.5.2.5 le rôle dans le mécanisme de réparation BER :

La protéine APC est aussi impliquée dans le mécanisme de régulation BER (base excision repair) (Jaiswal.A.S, 2008). Le BER est impliqué dans la réparation des petites lésions d'ADN (Liu.Y, 2007), en cas des grandes lésions, la transcription d'APC est stimulée, cette dernière forme un complexe avec les modulateurs de ce mécanisme tel que Fen-1 et pol-b, ce qui permet le blocage du processus de BER. L'ADN non réparé induit le déclenchement du mécanisme d'apoptose. (Jaiswal.A.S, 2008)

II.5.3 Les manifestations phénotypiques d'une perte de fonction d'APC

II.5.3.1 Syndrome FAP : (familial adenomatous polyposis)

II.5.3.1.1 Définitions

C'est une maladie génétique à transmission autosomique dominante (Polakis.P, 1995), généralement causée par des mutations germinales au niveau du gène APC, mais 25% sont des mutations de novo. Caractérisé par la formation précoce de nombre variable de polypes adénomes colorectal multiple (Leoz.M.R, 2015)

NB :

Polypes : c'est une excroissance qui se développe sur la muqueuse du côlon, on distingue deux types : polype bénins qui ne présentent pas un risque de développement, et des polypes adénomateux, qui sont susceptible de se développer en tumeur cancéreuse. (Médiste, 2018)

Adénome : est une dysplasie, une tumeur néoplasique bénigne, qui peut se développer en tumeur néoplasique maligne, (ex : CCR : cancer colorectal) (cancer)

Les patients avec le syndrome FAP, développent d'autres manifestations en plus des polypes gastriques de duodénales : cancer de la thyroïde, cerveau et l'estomac, hypertrophie congénitale, kystes épidermoïdes (Leoz.M.R, 2015)

II.5.3.1.2 Phénotypes de FAP :

Selon le nombre de polyposes et l'âge de leurs survenus, on distingue

- FAP Classique : centaine de milliers de polypes tout au long du colon et rectum à l'âge de 15 ans, et CCR inévitable à l'âge moyen 35ans
- FAP Atténué (AFAP) : 10 à 1000 polypes, qui se manifestent à un âge un peu tardif, et le risque de développer le CCR est moindre, une forme d'imitation d'une forme sporadique (évolution bénigne de la maladie) (Leoz.M.R, 2015)

II.5.3.1.3 Corrélation entre phénotype FAP et son génotype

Il existe une association entre les manifestations cliniques et la localisation des mutations germinales du gène APC

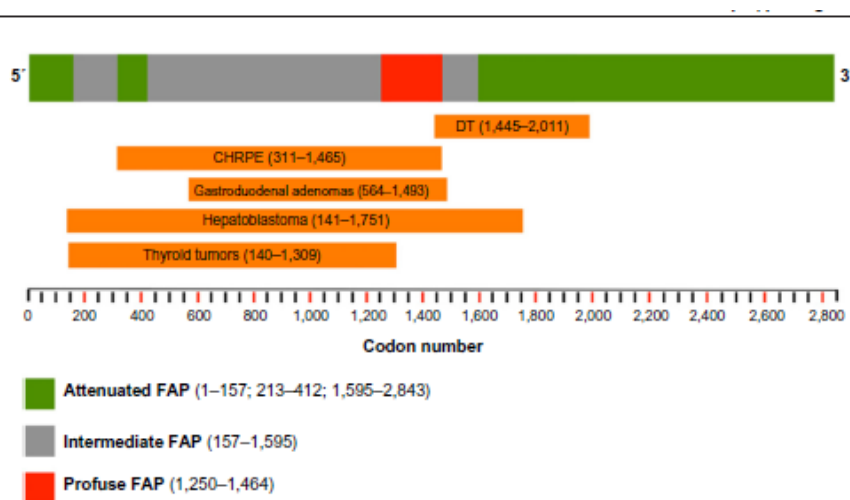


Figure 8: Corrélation phénotype-génotype du gène APC (Leoz.M.R, 2015)

DT : tumeur desmoïde, CHRRE : hypertrophie congénitale de l'épithélium pigmentaire rétinien, FAP : familial adenomatous polyposis,

Plus de 1100 mutations probablement pathogènes ont été répertoriées, on trouve majoritairement :

- Mutations tronquantes : 28% non-sens, 10% insertion, 46% délétions, 3% faux sens et 13% grosses altérations : délétion des exons (défaut d'épissage), ou duplication.
- Les mutations faux sens sont rare mais 60 sont déjà classées potentiellement

pathogène

Selon cette corrélation on distingue :

- Polypose agressive (profuse FAP) présenté en rouge : entre les codons 1250 et 1464, principalement 0309
- AFAP (Attenuated FAP) présenté en vert : généralement les extrémités 5' (avant le codon 157) et 3' (après le codon 1595), en plus régions d'épissage alternatif de l'exon 9 (213-4012)
- Phénotype intermédiaire de FAP (classique) présenté en gris : le reste du gène ; entre les codons 157 et 1595 sauf le 1309. (Nieuwenhuis.M.H, 2006)

Les manifestations extra coloniales sont associées également à des mutations spécifiques de gène APC, généralement situées avant le codon 1400 sauf pour le DT : tumeur desmoid en aval de 1400.

II.5.3.2 Cancer colorectal

Les patients avec le syndrome FAP, présente un risque majeur au développement d'un cancer colorectal à un âge précoce, d'où l'intérêt de préventions par des approches et interventions endoscopiques et chirurgicales. (Leoz.M.R, 2015)

Le gène *APC* est un gène suppresseur de tumeur car sa perte de fonction est impliquée dans l'initiation et la promotion du cancer : (Leoz.M.R, 2015)

- L'initiation du cancer :

La perte de fonction d'APC fait acquérir un avantage fort de sélection positive aux cellules colorectales, par la perte de régulation nucléaire de β catenin, ce qui imite une activation constitutive de la voie WNT, et promouvoir ainsi l'initiation à l'amplification clonale

- La progression tumorale :

Les mutations dans le domaine C-ter de APC, permet l'apparition des réarrangements chromosomiques (CIN), former ainsi de nouvelles mutations au niveau des gènes suppresseurs de tumeur et les oncogènes, ce qui confères aux cellules le caractère maligne et métastases.

II.5.3.3 Manifestations extracolorectal,

Les mutations au niveau d'APC sont aussi liées au cancer gastrique, du sein, des poumons et l cancer cervical (Neo.O, 2021)

MATERIEL ET METHODE

II. Matériels et méthodes

II.1 Matériels

II.1.1 *In silico*

Les Bases de données :

L'interface où les variations sont répertoriées, ont été consultées soit à partir de l'interface Alamut Visual (version 2.15, Feb 2020) ou bien directement sur leur interface web

gnomAD (Genome Aggregation Database) : une base de donnée qui comprend le séquençage d'exome de 13216 individus, et séquençage du génome de 15496 individus de populations différentes, permettant ainsi d'estimer la fréquence allélique pour chaque variation (Karczewski.k, 2017)

ClinVar : archive publique, qui rapporte la relation entre phénotype et variations humaine (interprétation clinique des variations) basé sur les échantillons des patients hébergés par NCBI (National Center for Biothechnology information), et financé par NIH (National institutes of Health) (Landrum.M, 2013).

dbSNP (Single Nucleotide Polymorphism database) : archive publique, développé par NCBI dans le but d'assembler des variations génétiques inter et intra-espèces, incluant: substitutions d'une seule nucléotide, répétitions micro-satellites, petites indel (Sligielski.E M, 2000).

UMD (Universal Mutation Database) : c'est une base de donnée développée en France par INSERM-UMR_S910, dédiée à la collection des mutations retrouvées dans le génome humain associées à des maladies génétiques, afin d'établir la corrélation phénotype-génotype. (Beroud, 1992) <http://www.umd.be/>

UMD-APC : fournie des informations publiées ou non, sur des variations germinales du gène APC trouvé par l'analyse des tests génétiques établies par des laboratoires de diagnostic français. Actuellement 3717 variations ont été répertoriées. (Laboratories, 2014) <http://umd.be/APC/gene.shtml>

Cosmic (Catalogue Of Somatic Mutations In Cancer) : est une base de données hébergée par l'institut Wellcome Sanger, dans le cadre du Cancer Genome Project. Elle regroupe les mutations somatiques acquises dans les cancers humains. (Cancer Sanger)

Les Outils bio-informatiques :

Algorithmes de prédictions de l'impact de ses variations sur le mécanisme d'épissage.

SpliceAI Lookup : outil basé sur l'utilisation de l'intelligence artificielle (réseau neuronal artificiel profond) pour prédire si à chaque position d'un l'ARNpré-messager arbitraire, est un site donneur, accepteur ou ni l'un ni l'autre.

Il prédit, le gain et la perte de la force d'épissage à plusieurs centaines de paires de bases pour une variation donnée, et les potentiels sites cryptiques qui peuvent être engendrés. (Jaganathan.K, 2019).

MMsplice (modular modeling of splicing) : les modules de cet outil se base sur la notation des sites d'exons, d'intron et d'épissge, par un réseau de neurones. Ces modules sont combinées pour prédire les effets des variations sur : le saut d'exon, le choix du site d'épissage, l'efficacité de l'épissage et la pathogénicité (Cheng.J, 2019).

SROOGLE : est un serveur Web qui met à la disposition du biologiste la séquence du signal d'épissage, et les score dans un format intégré, visuel et facilement interprétable. Le navigateur graphique met en avant les 4 signaux d'épissage et leurs scores associés, en se basant sur neuf algorithmes différents. L'interface offre également la possibilité d'examiner l'effet des mutations ponctuelles à n'importe quelle position donnée (Schwartz.S H. A., 2009).

MaxEnt (Maximum Entropy) : c'est un modèle de prédiction pour des motifs de séquences données en tant que sites d'épissage donneurs ou accepteurs, basé sur le principe d'entropie maximale (MEP) (Yeo.G, 2004).

HSF (Human Splicing Finder): c'est un outil bio-informatique en ligne, qui combine 12 algorithmes différents pour identifier les signaux d'épissage : les dites accepteurs 3'ss et donneurs 5'ss, les séquences auxiliaires d'épissage (ESE / ESS) et les sites de branchement au niveau de génome humain. Cet outil est également conçu pour prédire théoriquement l'impact des mutations (indel, SNP) sur ces signaux. (Desmet.F.O, 2009).

HExoSplice : c'est une interface web, désignée pour prédire l'effet des SNVs (single nucleotide variations) sur des éléments régulateurs de l'épissage SREs (splicing regulatory elements), incluant les ESEs (exonic splicing enhancers) et les ESSs (exonic splicing silencers). Cet outil est principalement basé sur l'évaluation quantitative de tous les hexamères de l'ARN comme des potentiels ESRseq (exonic splicing regulatory sequences) (Lefebvre.A, 2020).

I.1.2 Test fonctionnel

Cellules HeLa : décongelées (azote -180°C) le 24/01/22 : commandés dans (ATCC)

C'est une lignée cellulaire immortelle cultivée *in vitro*, elle a été obtenue à partir d'un cancer cervical affectant une femme Afro-américaine Henrietta Lacks sans consentement, utilisées dans le domaine de la recherche (Jorden.BR, 2013).

Le vecteur pCAS2 : préalablement construit dans le laboratoire. Il est composé d'un gène de résistance à la Carbénicilline, 2 exons (A et B) séparés par une région intronique présentant deux sites de restriction BamHI et MluI. L'expression de ce minigène est sous le contrôle d'un promoteur CMV (CytoMégaloVirus) (Gaildrat.P, et al, 2010)

ADNg témoin d'un patient : référence T403 LM070318 (263,5 ng/μL)

Bactéries DH5 alpha : présente à -80°C (Invitrogen). Sont des bactéries *Escherichia coli* compétentes, génétiquement modifiées en introduisant 3 mutations : recA1 faciliter l'insertion du plasmide, endA augmente le rendement, lacZΔM15 est utilisé pour la sélection des clones par rapport à leurs couleurs bleue/blanc (DH5α Competent Cells).

Enzymes de restriction : BamHI et MLUI

Les KITs utilisés :

CloneAmp HIFI PCR Primix (ADN polymérase HIFI + dNTP + tampon de l'enzyme), permet l'amplification des fragments ADN.

Clonage In-fusion : permet le clonage des produits PCR avec le vecteur préalablement digéré, par la recombinaison homologue de 15pb (ozyme).

Nucleospin plasmid DNA purification: Miniprep, permet l'extraction et la purification des plasmides à partir des bactéries (Macherey-Nagel).

NucleoSpin® Gel and PCR Clean-Up : extraction d'ARN, extraction des acides nucléiques, du gel d'agarose (Macherey-Nagel)

Les amorces (IDT) : sont des brins courts d'ADN qui s'hybrident avec complémentarité de bases servent comme point de départ à l'amplification de l'ADN d'intérêt, *in vivo* et *in vitro*, par la liaison d'une ADN polymérase. (How to make primers for PCR)

In vitro, elles peuvent servir : la PCR, la mutagenèse dirigée, le séquençage, les analyses ciblées...

Les différentes amorces utilisées dans notre recherche sont présentées dans l'annexe 1.

- ➔ Amorces RT-PCR (pCASKOIF/ pCAS-2R)
- ➔ Amorces pCASseq F/R
- ➔ Amorces infusion (Bam F/MLU R)
- ➔ Amorces Mutagenèse dirigée (-1, exonique, +1, +2)

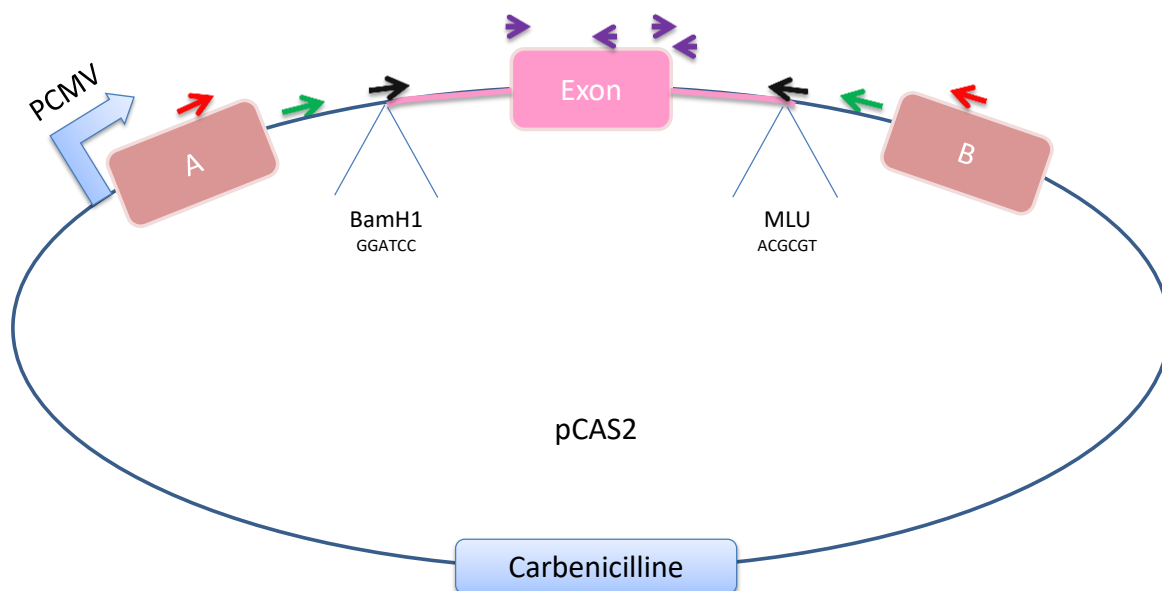


Figure 9: Présentation schématique de l'endroit de fixation des amorces sur le vecteur pCAS2

II.2 Méthodes

Pour répondre à la problématique, nous avons opté pour deux approches complémentaires : les analyses *in silico* et l'analyse du test fonctionnel minigène.

II.2.1 Les analyses *in silico*

Ces analyses ont été basées sur l'interrogation de différentes bases de données, et interfaces de prédictions bio-informatiques afin de sélectionner les variations à analyser par le test minigène d'une part, et d'autre part les résultats ont été comparés avec les différents algorithmes de prédiction bio-informatique.

II.2.2 Test fonctionnel minigène

- Le principe

C'est un test fonctionnel indicateur d'anomalies de l'épissage, dont l'effet de la variation sur l'épissage est étudié dans un contexte génomique et cellulaire artificiel dans le but d'établir un lien direct de causalité entre la variation et son effet sur l'épissage. (Meulemans.L, 2017)

Ce test a été conçu comme une alternative à l'ARN de patients, essayant de mimer le contexte physiologique, ou étudier dans un contexte général de certaines variations qui n'ont pas été répertoriées dans les bases de données.

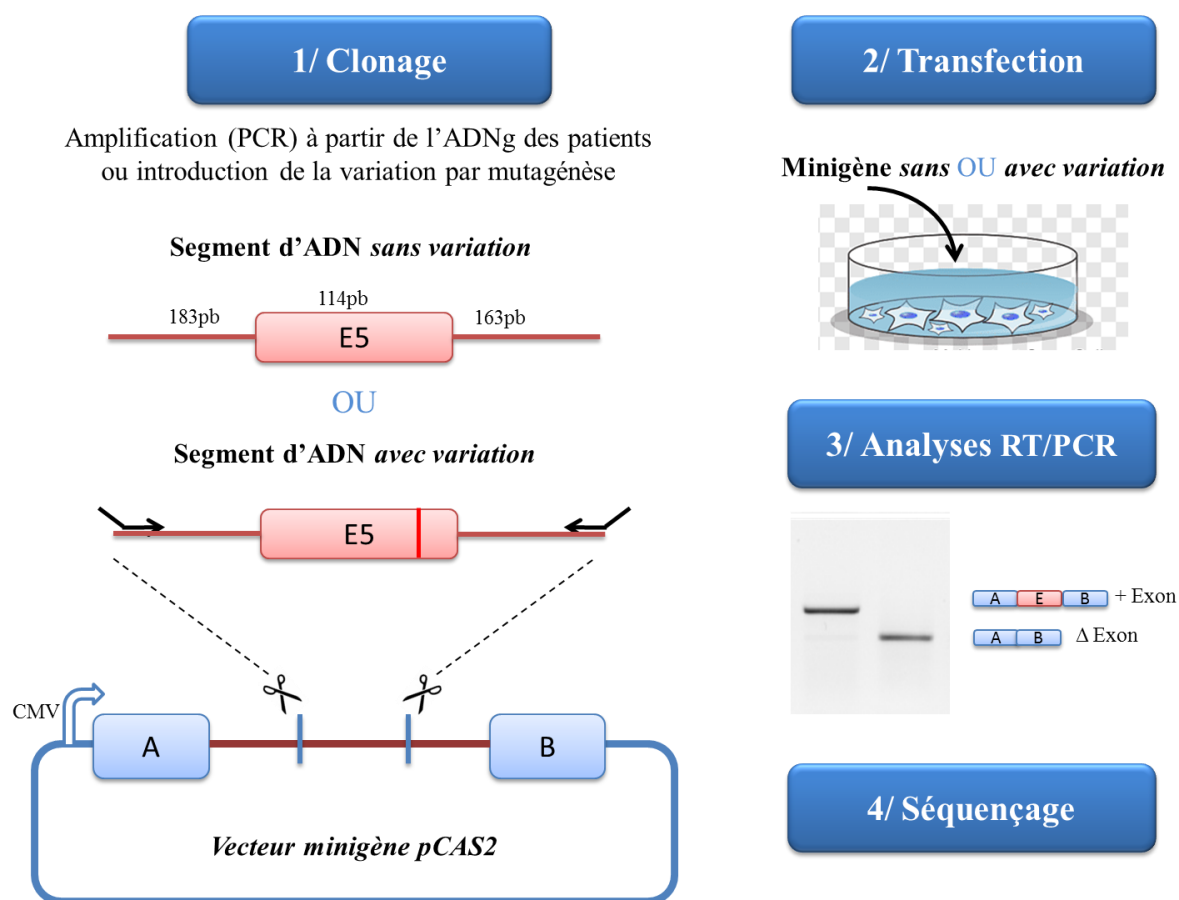
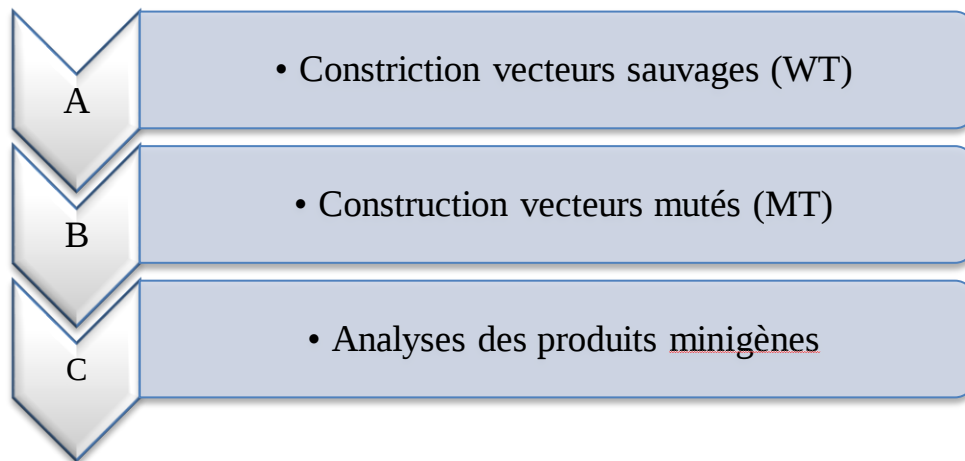


Figure 10: Principe du test fonctionnel indicateur d'anomalies d'épissage, basé sur l'utilisation du minigène pCAS2

Les étapes de manipulation



A/ Construction vecteur sauvage (WT)

A.1 Amplification par PCR les parties d'intérêts de APC (exon 5, 6 et 11 et leurs introns).

- à partir d'un ADNg témoin (T403 LM070318) préalablement récupéré au laboratoire à 263.5ng/ul.
- des amorces spécifiques « BamHI/MluI infu » qui s'hybrident dans les introns de part et d'autre de chaque exon.
- Premix : cloneAmp HIFI PCR qui contient tous les acteurs nécessaires pour la réaction de PCR : enzyme polymérase, les dNTP, le tampon de l'enzyme.

La réaction d'amplification se fait dans un thermocycleur : 98°C pendant 10s (dénaturation des brins d'ADN) puis 55°C pendant 5s (hybridation des amorces sur les brins d'intérêts) par la suite 72°C pendant 5s (la température optimale pour l'enzyme HIFI : l'élongation) ce cycle doit se répéter 29 fois.

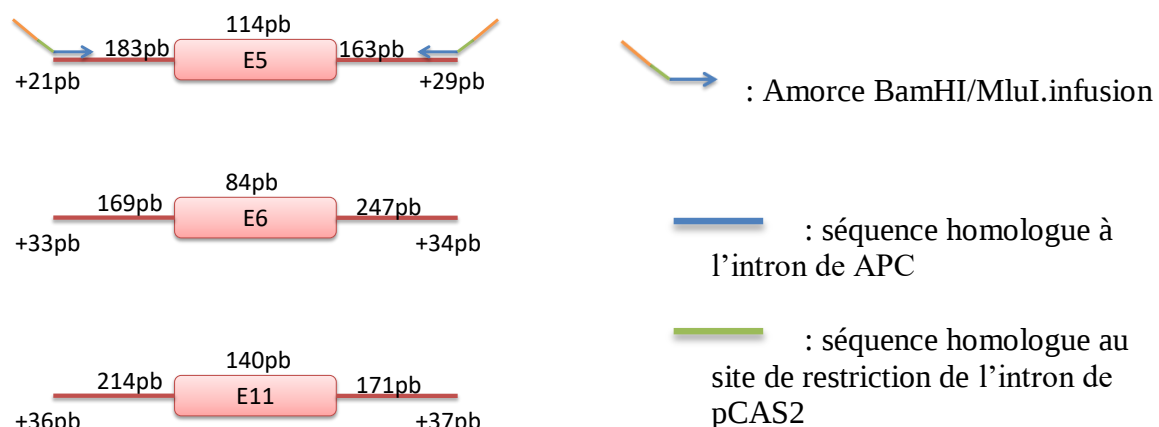


Figure 11: La taille des inserts amplifiés par PCR par le biais des amorces InFusion

Les produits PCR ont été migrés sur un gel d'agarose à 1,5%, pour s'assurer que

les fragments d'intérêts sont présents, ces derniers ont été purifiés et récupérés grâce au kit *Macherey Nagel PCR and Gel Clean Up*.

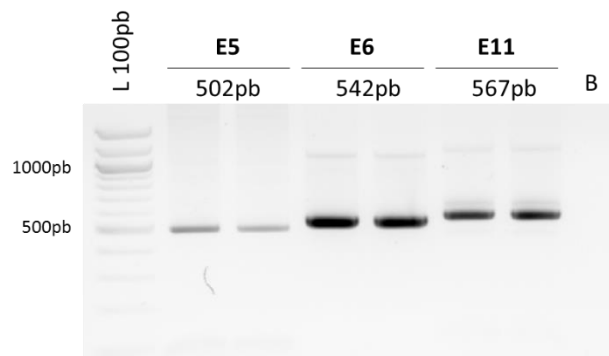


Figure 12: Migration des produits PCR sur gel d'agarose à 1.5%

A. 2 Digestion du vecteur pCAS2 :

- Le vecteur pCAS2 préalablement construit au laboratoire (LM160119) à une concentration (623ng/ul) a été utilisé.
- Les enzymes de restriction BamHI et MluI et le tampon de ces enzymes ont été utilisés.

La réaction de digestion se fait dans un thermocycleur pendant 2h à 37°C, puis les bandes ont été migrées dans un gel d'agarose à 1.5%.

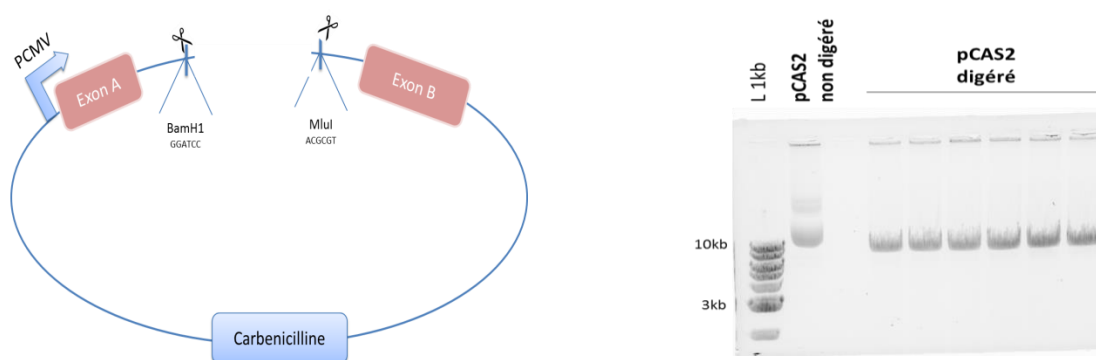


Figure 13: Représentation schématique des sites de digestion sur pCAS2 par BamHI et MluI, la migration sur gel d'agarose du pCAS2 digéré et non digéré

A. 3 Clonage InFusion des inserts :

Le clonage se fait par recombinaison homologue à 50°C pendant 15min entre le vecteur pCAS2 digéré et les inserts (produits PCR) à l'aide d'un prémix In Fusion.

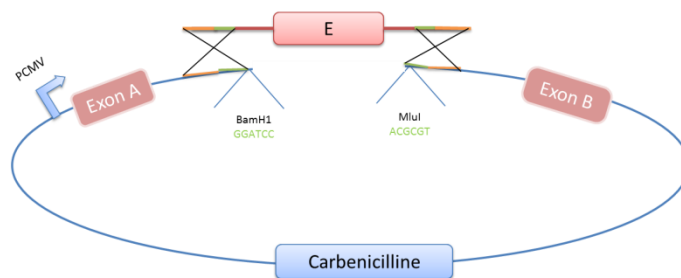


Figure 14: Représentation schématique du clonage de l'insert dans le vecteur pCAS2 digéré par recombinaison homologue

A. 4 Transformation bactérienne

- Les bactéries compétentes DH5α ont été transformées par choc thermique (30mn sur glace => 45s à 42°C => 2mn sur glace).
- Les bactéries ont été ensuite amplifiées dans 1ml de LB (milieu de culture liquide) pendant 1h à 37°C sous agitation.
- Les bactéries ont été coulées sur une boîte qui contient LB+CB (milieu de culture solide + antibiotique carbenicilline) afin de laisser pousser uniquement les bactéries transformées par les vecteurs qui contiennent les inserts d'intérêt.
- Après une nuit d'incubation, les vecteurs ont été récupérés (miniprep) en utilisant le kit Nucleospin plasmid DNA purification, Macherey-Nagel.

A. 5 Séquençage des plasmides

On a passé par cette étape pour s'assurer que les vecteurs d'intérêt contiennent les bons inserts et qu'ils sont WT et aucune mutation n'a été engendrée durant les manipulations.

La réaction de séquence : La partie flanquante entre exon A et B a été amplifiée, en utilisant les amorces pCASseqF (Forward) et pCASseqR (Reverse), le kit Big Dye (dNTP fluorescente terminateurs, enzyme polymérase) et le vecteur pCAS2.

- La réaction se fait dans un thermocycleur à 95°C pendant 2mn (dénaturer le plasmide)

puis 40 cycles de 96°C/10s, 50°C/5s, 60°C/2mn (amplification de fragments d'intérêts), 12°C infinie.

- Les produits de PCR sont ensuite placés dans un séquenceur, et les résultats ont été comparés avec une séquence de référence par un logiciel d'alignement en ligne MultAlin : <http://multalin.toulouse.inra.fr/multalin/>

B /Construction des vecteurs Mutés

B. 1 Mutagénèse dirigée

C'est une méthode basée sur 2 réactions de RCR successives en utilisant des amorces de mutagénèse dirigée (amorces qui contiennent la mutation d'intérêt), permet d'avoir des fragments d'ADN dont chacun porte une seule mutation.

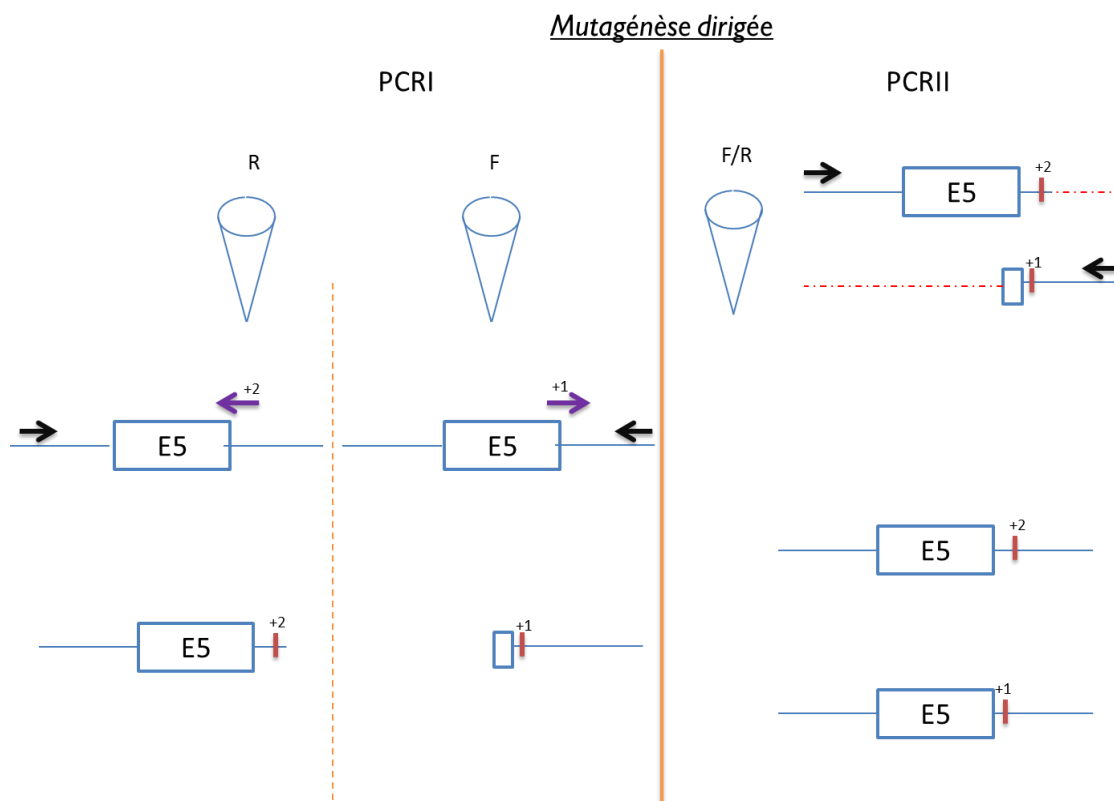


Figure 15: Principe de la mutagénèse dirigée : PCRI et PCR II

La PCRI :

On a utilisé comme matrice les vecteurs WT (plasmide pCAS2+l'insertWT), les amorces InFusion, et les amorces de mutagenèse dirigée.

La réaction d'amplification se fait comme expliquer dans l'étape (A.1 Amplification par PCR les partie d'intérêts)

Les produit de cette PCRI sont des fragments incomplets des inserts mais contiennent la mutation d'intérêt.

Pour déterminer la taille des fragments, ces derniers ont été migrés sur un gel d'agarose à 1.5%

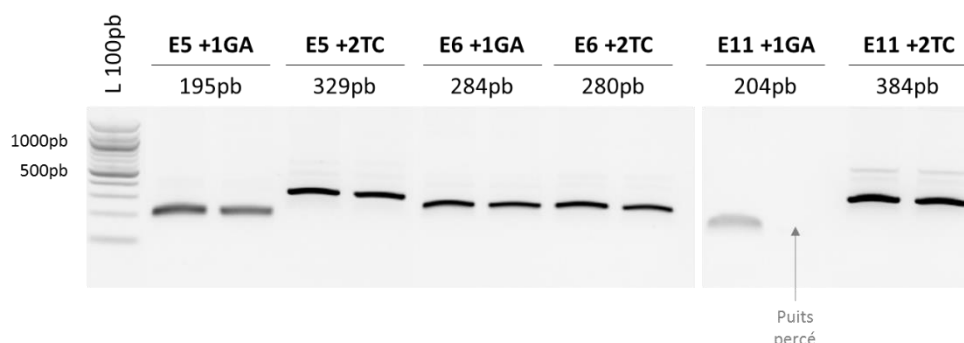


Figure 16: La migration sur gel d'agarose à 1.5% des produits PCRI

La PCR II :

Les matrices de cette réaction sont les produits de la PCRI, qui vont s'hybrider par complémentarité de bases afin d'avoir des fragments complets, puis dans les prochains cycles, ce sont les amorces InFusion qui vont prendre le relai pour assurer l'amplification

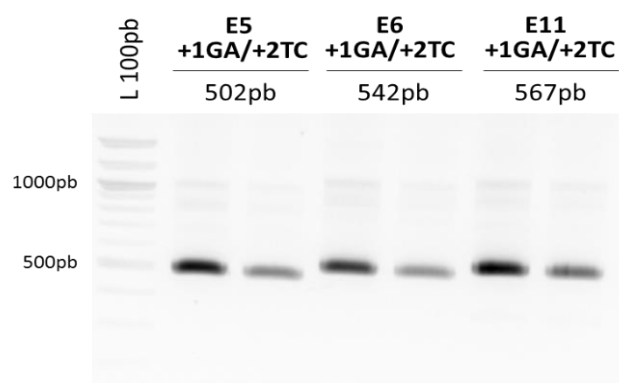


Figure 17: La migration sur gel d'agarose des produits de PCRII

NB :

Pour la construction du vecteurs mutés (pCAS2+insertMuté) les étapes de de la construction des vecteurs sauvage (pCAS2+insertWT) ont été suivies, sauf l'étape de la digestion du vecteur car la quantité de vecteur préalablement digérés est suffisante.

Il s'agit du clonage InFusion des inserts mutés dans le vecteur pCAS2, la transformation bactérienne et le séquençage des plasmides.

C /Analyse des produits minigènes :

Pour pouvoir analyser le profil d'épissage de ces minigènes WT et mutés, ces derniers ont été transfectés dans des cellules Héla, ensuite les ARNm ont été extraits.

- Dans un premier temps, les cellules Héla ont étéensemencées dans des plaques de culture de 12 puits (60 000 cellules/puits)

(Les cellules Héla ont été entretenues dans des flasques à 37°C, puis le milieu de culture a été aspiré et la trypsine a été ajoutée. Les flasques ont été ensuite incubées dans un incubateur à 37°C pendant 5min, les cellules ont été décollées et mises dans un milieu de culture DMEM

Les cellules ont été comptées pour avoir une concentration de 60 000 cellules/puits, de ce fait, les cellules ont été ajoutées en goutte à goutte dans chaque puits, et la plaque de culture de cellules a été mise dans un incubateur.)

- Dans un second temps, les cellules ont été transfectées après 24h d'incubation et incubées à 37°C (correspondant à une confluence de 75%) avec les vecteurs WT et mutés à l'aide du réactif Fugene (il crée des pores au niveau de la membrane des cellules)
- Les ARN totaux ont été extraits à partir des cellules Héla après une nuit d'incubation (correspondant à une confluence de 98%) en utilisant le kit : Macherey-Nagel Nucleospin RNA Isolation

Afin d'analyser le profil d'épissage des ARNm (issus de la transcription des vecteurs sous le contrôle du promoteur CMV) deux méthodes complémentaires ont été sélectionnées.

C. 1/ RT-PCR classique suivie par le séquençage

But : cette analyse est qualitative, en effet, les produits RT-PCR ont été d'abord migrés pour visualiser les bondes (chaque bonde correspond à un épissage unique) ce qui nous permet de savoir le nombre de profils d'épissage engendrés pour chaque type de mutation. Ensuite chaque bonde est séquencée afin d'identifier les fragments.

C. 1.1. RT-PCR :

Pour pouvoir convertir les ARN d'intérêts en ADNc, nous avons utilisé les ARN Totaux, deux amorces pCASKOI-F et pCAS2-R (complémentaire à une séquence dans l'exon A et B respectivement), l'enzyme One Step, tampon de l'enzyme, dNTP.

- La réaction sur le thermocycleur est faite comme suit : 50°C/30mn, 95°C/15mn puis

40 fois le cycle (94°C/30s, 57°C/30s, 72°C/1mn) 72°C/1mn, 12°C à l'infini

- Les produits de la RT-PCR ont été migrés sur gel d'agarose à 2% pendant 45mn à 130V (short migration). La migration a été encore reprise pendant 45 mn (long migration).

C. 1.2. Le séquençage :

Les ADNc issus de la purification sur gel d'agarose, l'amorce pCASKOI-F (pour séquencer en Forward) et la pCAS2-R (pour séquencer en reverse), Big Dye ont été utilisés.

- La réaction est réalisée dans un thermocycleur à 95°C pendant 2mn (dénaturer le plasmide) suivie de 40 cycles de 96°C/10s, 50°C/5s, 60°C/2mn (amplification de fragments d'intérêts), 12°C infinie
- Les produits de la réaction de séquence sont ensuite placés dans un séquenceur, et les résultats ont été comparés avec une séquence de référence par un logiciel d'alignement en ligne MultAlin : <http://multalin.toulouse.inra.fr/multalin/>

C. 2. RT-PCR fluo suivie par Gene Scan :

But : cette analyse est quantitative, c'est-à-dire elle nous permet de quantifier chaque type de fragment (profil d'épissage) issu du même type de vecteur, dans des conditions dénaturantes

C. 2.1. RT-PCR fluo :

La réaction : c'est une RT-PCR classique, l'amorce pCASKOI-F* fluo a été utilisée, pour cela des gradients de cycles ont été appliqués pour choisir le cycle optimal pour chaque exon (5, 6 et 11).

Le cycle optimal doit être inclus dans la phase 2 : croissance exponentielle

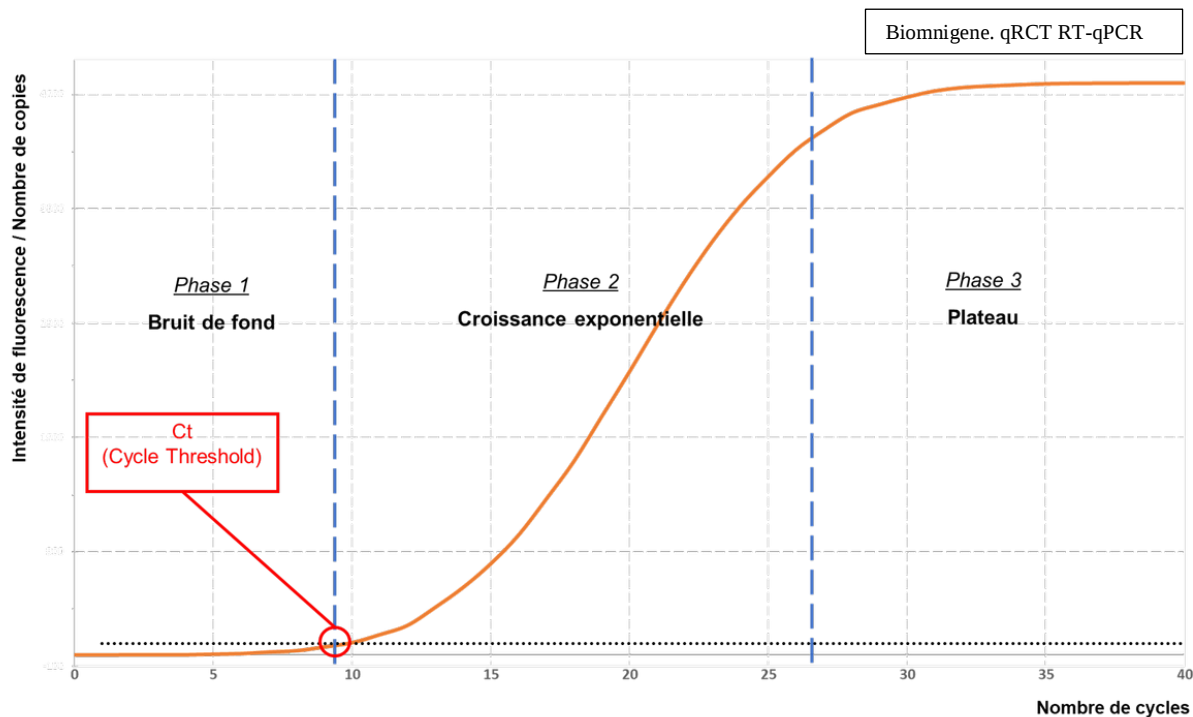


Figure 18: Courbe d'amplification d'un fragment cible par PCR en temps réel

Le principe du gradient du cycle: la réaction est arrêtée chaque deux cycles, et les produits de la RT-PCR ont été migrés sur un gel d'agarose à 2%, les bandes sont ensuite quantifiées par un logiciel. Les données obtenues sont utilisées pour dessiner une courbes d'amplification, ce qui a permis de choisir le cycle optimal pour chaque exon.

- Une fois le cycle optimal est choisi : 32 cycles pour E5 et 26 cycles pour E6 et E11, la réaction RT-PCR fluo est répétée avec le cycle optimal et les produits de cette réaction ont été utilisé d'une part pour une migration et d'une autre part pour le Gene Scan

C. 2.2. Gene Scan :

Le milieu réactionnel contient : ADNc dilué (à 1/100 pour E6 et E11, et à 1/10 pour E5), marqueur GS 500 ROX(fragments de références fluo joue le rôle d'un leader), Formamide (dénature l'ADNc)

- La dénaturation a été lancé dans le thermocycleur pendant 2mn à 95°C
- Le produit de cette dénaturation est lu sur le séquenceur, les résultats bruts engendrés ont été analysés par le logiciel Gene-Scan Analysis.

RESULTATS

III. Résultats

III. 1 Analyse *in silico*

III.1. 1 La sélection des variations à analyser par test fonctionnel :

Les résultats des analyses *in silico* ont permis de choisir les mutations à analyser par le test fonctionnel minigène,

Les critères :

- Les variations susceptibles de ne pas affecter le processus d'épissage sur la base des prédictions bioinformatiques, principalement les IVS +1, +2,
- Les mutations qui ont été précédemment classées pathogènes dans au moins une base de donnée,
- Des variations qui affectent le processus d'épissage comme témoins

Tableau 4 : les variations d'APC sélectionnées pour le test fonctionnel

	E5	E6	E11
SET #2	c.532-1GA	c.646-1GA	c.1409-1GA
	c.602 AT	c.688 CA	c.1532 GT
SET #1	c.645+1GA	c.729+1GA	c.1548+1GA
	c.645+2TC	c.729+2TC	c.1542+2TC



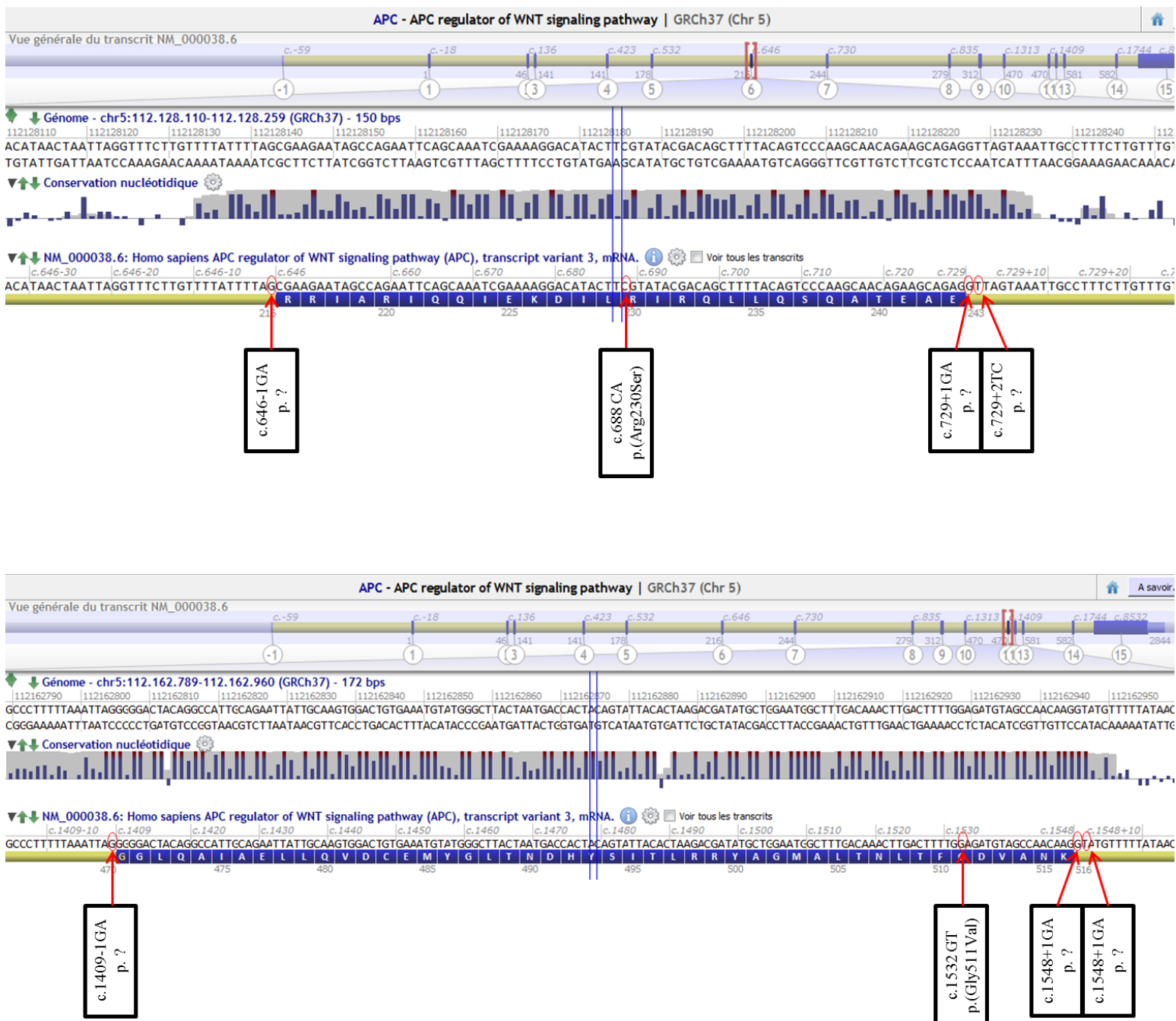


Figure 19: Description de la localisation des variations sélectionnées dans l'exon 5, 6 et 11 d'APC, et leurs régions introniques flanquantes.

Dans le panel du haut, la localisation de l'exon 5, 6 et 11 est présentée dans le contexte du gène APC (16 exons : rectangles bleus ; 15 introns : lignes jaunes). Dans le panel du bas, la distribution des variations sélectionnées dans le cadre de ce projet est indiquée par rapport à la séquence nucléotidique et la séquence protéique de chaque exon. La nomenclature des changements nucléotidiques est indiquée en coordonnées « c. » selon les recommandations du consortium international HGVS (*Human Genome Variation Society*) sur la base de la numérotation des nucléotides dans la séquence codante du transcrit de référence APC (NM_000038.6). Les conséquences prédites de ces variations au niveau protéique sont

également indiquées en coordonnées « p. ». Ces variations sont celles qui ont finalement fait l'objet d'une analyse fonctionnelle dans le test minigène. Les images présentées ici correspondent à des représentations modifiées extraites de l'interface Alamut Visual 2.15 (Interactive Biosoftware, Rouen).

III.1. 2 Bases de données

Tableau 5 : Les informations collectées sur les variations sélectionnées

Exon	Variation nucléotidique	Localisation	Nature de variation	Type de signal d'épissage	ClinVar	UMD	gnomAD	dbSNP	Cosmic
E5 In frame	c.532-1 G>A	Intron 4	Naturelle	3'ss (IVS-1G>A)	Pathogène	causal	N/R	In ClinVar	Pathogène
	c.602A>T	Exon 5	Artificiel	ESR	N/R	N/R	N/R	N/R	N/R
	c.645+1G>A	Intron 5	Naturelle	5'ss (IVS+1G>A)	Pathogène	Pathogène	N/R	p/pathogène	Pathogène
	c.645+2T>C	Intron 5	Artificiel	5'ss (IVS+2T>C)	N/R	N/R	N/R	N/R	N/R
E6 In frame	c.646-1G>A	Intron 5	Naturelle	3'ss (IVS-1G>A)	Pathogène	causal	N/R	N/R	Pathogène, VUS
	c.688C>A	Exon 6	Artificiel	ESR	N/R	N/R	N/R	N/R	N/R
	c.729+1G>A	Intron 6	Artificiel	5'ss (IVS+1G>A)	N/R	N/R	N/R	N/R	N/R
	c.729+2T>C	Intron 6	Artificiel	5'ss (IVS+2T>C)	N/R	N/R	N/R	N/R	N/R
E11 Out of frame	c.1409-1G>A	intron 10	Naturelle	3'ss (IVS-1G>A)	Pathogène	causal	N/R	N/R	Pathogène
	c.1532G>T	exon 11	Artificiel	ESR	N/R	N/R	N/R	N/R	N/R
	c.1548+1G>A	Intron 11	Naturelle	5'ss (IVS+1G>A)	Pathogène	Pathogène	N/R	N/R	Pathogène
	c.1548+2T>C	Intron 11	Naturelle	5'ss (IVS+2T>C)	p/pathogène	N/R	N/R	p/pathogène	N/R

N/R : non répertorié

p/pathogène : probablement pathogène

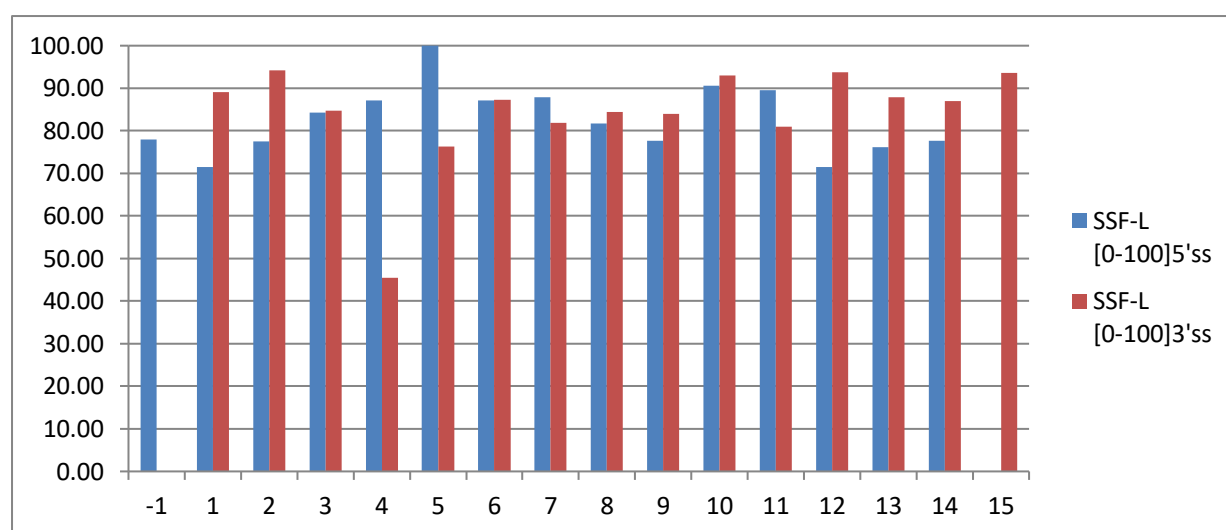
Pour les trois exons 5, 6 et 11, 4 variations ont été choisies: une exonique susceptible d'affecter les ESR (éléments régulateur d'épissage), les IVS-1GA, qui ont été déjà répertoriés dans les bases de données comme pathogènes, et les IVS +1GA et +2TC car ces positions sont très conservés et un changement conduit automatiquement à une altération du site d'épissage.

III.1.3. Les prédictions bio-informatique :

III.1.3.1 SSF-L et MaxEnt :

Les forces des sites donneurs 5' et accepteurs 3' tirés à partir de l'interface Alamut

SSF-L



MaxEnt

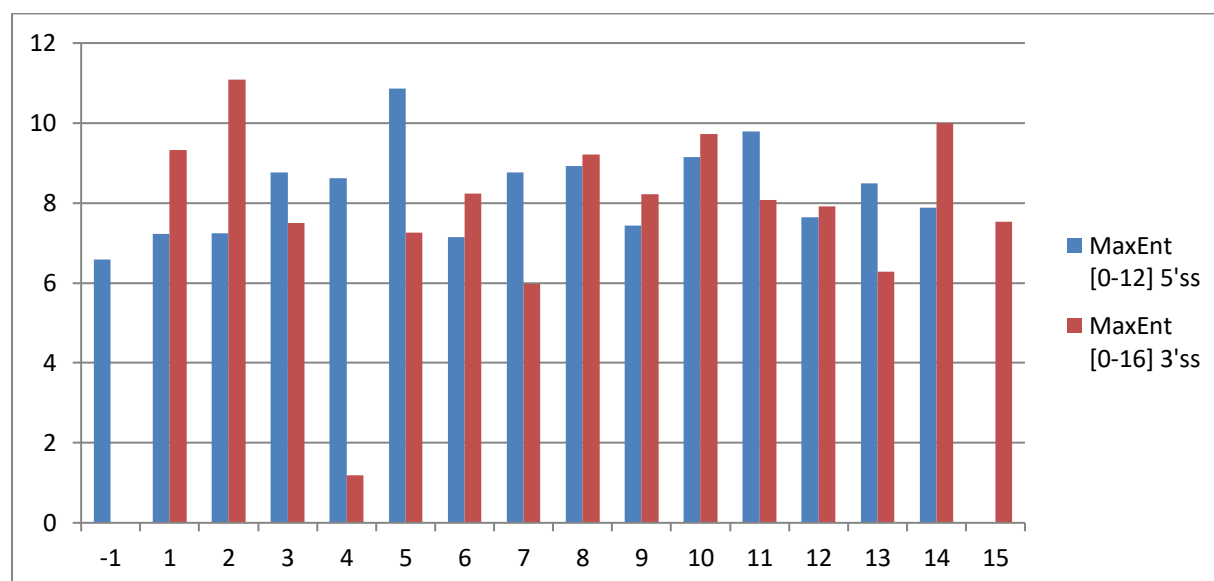


Figure 20 : Diagramme qui représente la force des sites d'épissage 5'ss et 3'ss pour

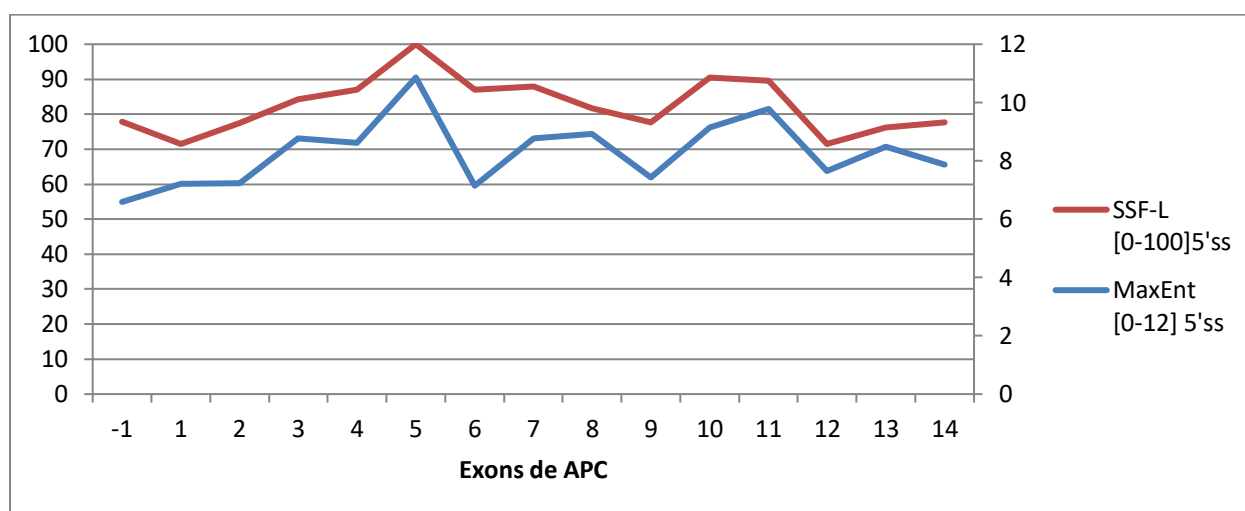
chaque exon d'APC tirées à partir de SSF-L et MaxEnt dans Alamut

D'après MaxEnt et SSF l'exon 5 a le site donneur 5'ss le plus fort parmi tous les exons de APC, et l'exon 2 pour le site accepteur 3'ss.

On remarque également que l'exon 4 a un site accepteur 3'ss considérablement faible par rapport au autres.

Les exons d'intérêts 5, 6, et 11 ont des sites d'épissage 3'ss et 5'ss généralement forts.

5'ss



3'ss

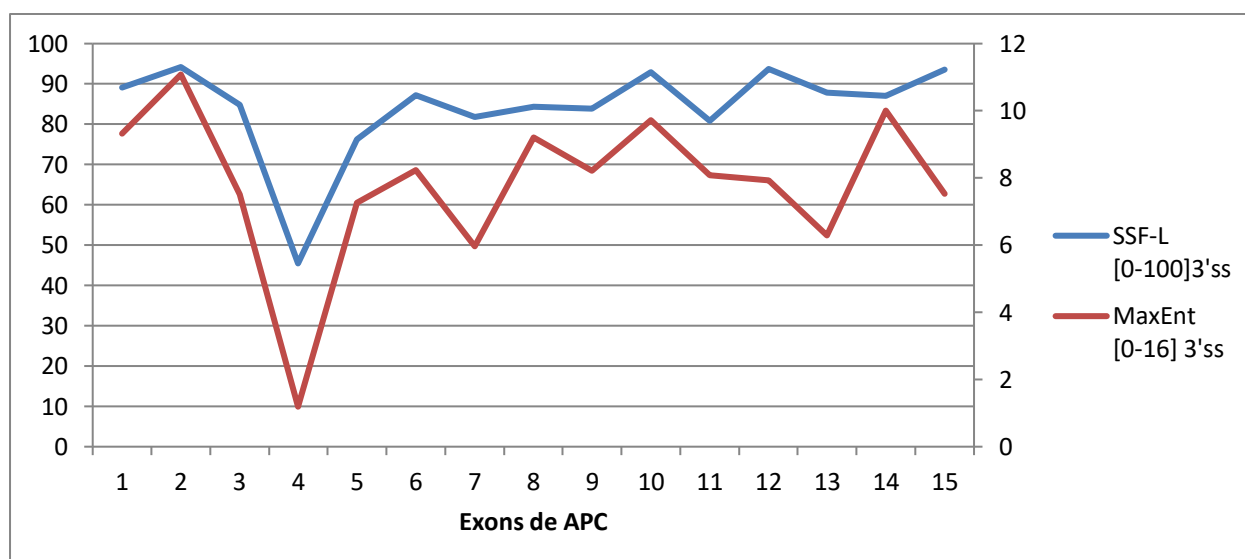


Figure 21: Courbe qui représente la force des sites accepteurs 3' et donneurs 5' constitutifs du gène APC basés sur les logiciels de prédiction SSF-L et MaxEnt tirés à partir de l'interface Alamut

En comparant les prédictions bio-informatiques sur la force des sites d'épissages de tous les exons d'APC tirées à partir des logiciels MaxEnt et SSF, il a été remarqué qu'elles sont en concordance.

III.1.3.2. SROOGLE:

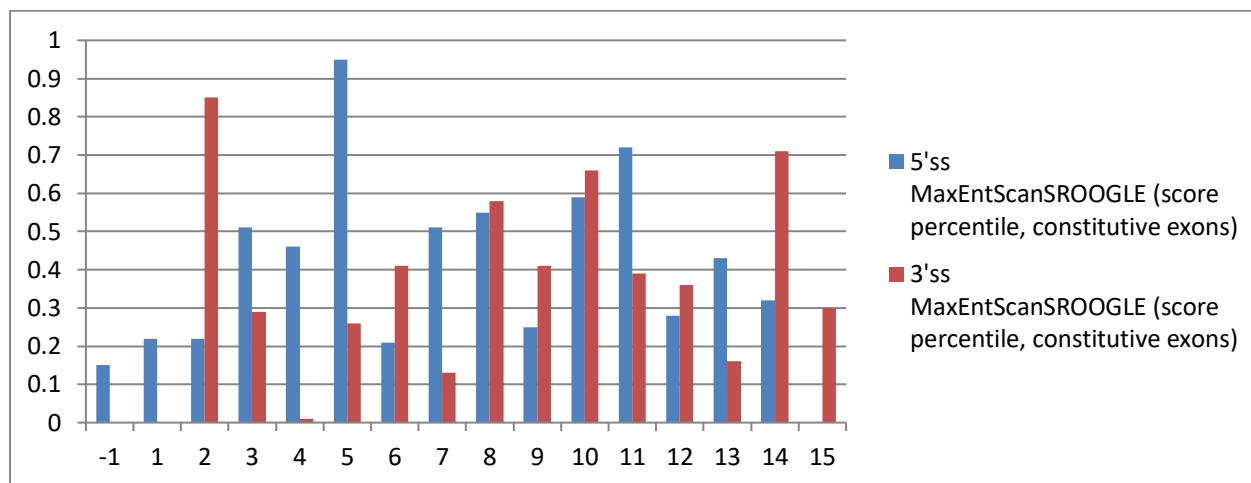


Figure 22: Diagramme qui représente le score percentile constitutif des sites donneur 5' et accepteur 3' d'APC tirés à partir de MaxEntScan SROOGLE

Contrairement aux deux logiciels précédents (MaxEnt/SSF) qui prédisent la force des sites d'épissage en fonction des exons du gène concerné, SROOGLE donne le ratio de ces forces par rapport à tous les sites d'épissage qui existent dans le génome humain.

Cela nous a permis de visualiser d'une manière globale la force des sites d'épissage des exons d'intérêts :

- Il a été remarqué que le site 5'ss de l'exon5 est très fort avec une valeur de 0.95, contrairement à son site accepteur 3'ss qui est le plus faible parmi les exons d'intérêt avec une valeur de 0.26.
- Les sites 3'ss de l'exon 6 et 11 ont la même force d'épissage avec une valeur de 0.41 et 0.39 respectivement qui sont en dessous de la moyenne.
- Le site 5'ss de l'exon 6 est considérablement faible avec une valeur de 0.21 contrairement à celui de l'exon 11 qui est en dessus de la moyenne avec une valeur de 0.72.

III.1.3.3. HSF :

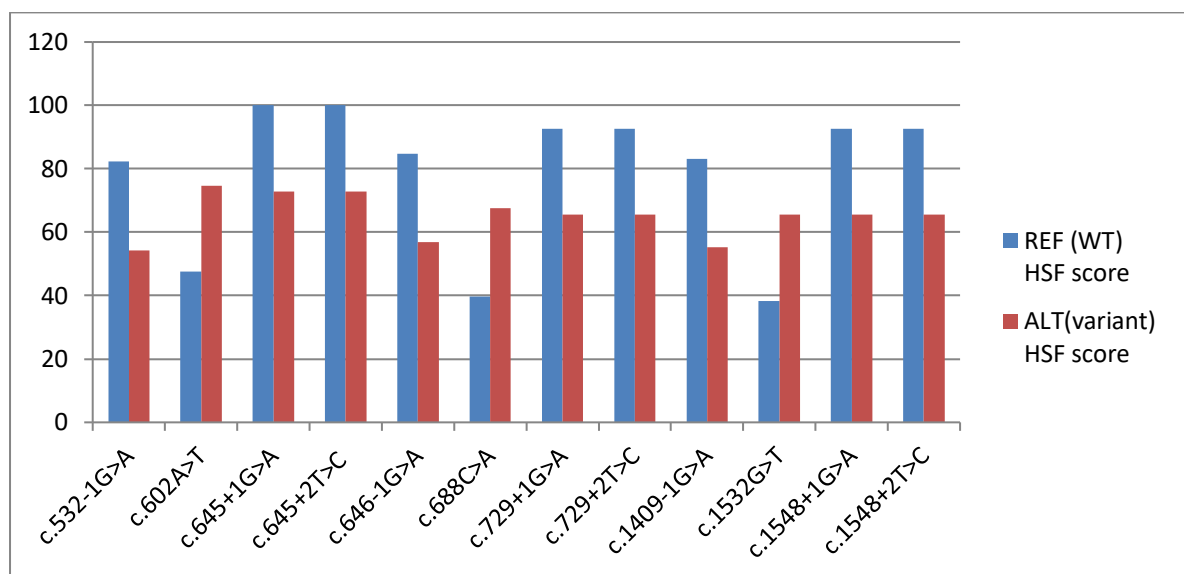


Figure 23: Histogramme qui représente la force des sites sélectionnés en état sauvage WT ou muté (variant) tirées par HSF

HSF a prédit la force des sites d'épissage à l'état sauvage (REF / WT) et en cas d'inclusion des variations d'intérêts (ALT / variant) :

Il a été remarqué que les variations IVS -1GA / +1GA / +2TC pour les exons 5, 6 et 11 conduisent toutes à la diminution de la force des sites d'épissage, contrairement aux variations exoniques qui affectent positivement les ESR, car les sites deviennent plus forts.

III.1.3.4. Splice AI look up

Tableau 6: Tableau récapitulatif des prédictions Splice AI Look up sur les scores et les positions des sites concernés (accepteur et donneur) pour chaque variation d'intérêt

Exon	Localisation variation	A/L/S	A/L/P	A/G/S	A/G/P		D/L/S	D/L/P	D/G/S	D/G/P
E5	c.532-1G>A	1	1	0,89	17		0	/	0	/
	c.602A>T	0,02	-54	0			0	/	0	/
	c.645+1G>A	0,02	-114	0	/		0,97	-1	0,39	-33
	c.645+2T>C	0,01	-99	0	/		0,05	-2	0,08	-34
E6	c.646-1G>A	1	1	0,85	11		0,03	-142	0	/
	c.688C>A	0,06	-37	0,01	396		0,02	-185	0,01	489
	c.729+1G>A	0,19	-48	0	/		0,99	-1	0,38	3
	c.729+2T>C	0,11	-80	0	/		0,4	-2	0,25	2
E11	c.1409-1G>A	1	1	0,99	2		0	/	0	/

c.1532G>T	0	0	0	0		0	0	0	0
c.1548+1G>A	0,04	-140	0	/		0,99	-1	0,94	37
c.1548+2T>C	0,02	-141	0	/		0,31	-2	0,65	36

A : acceptor L : loss S : score P : position G : gain D : donor

Splice AI prédit pour chaque variation donnée le gain ou la perte de la force des sites donneurs et accepteurs constitutifs, ainsi que la probabilité de création de nouveaux sites adjacents dû aux variations.

NB : Entre 0 et 0.2 le score est insignifiant / plus le score se rapproche de 1 plus le changement de la force d'épissage est important dans les deux sens.

- Il a été remarqué que toutes les variations IVS-1GA =1 conduisent à la perte du site accepteur des exons 5, 6 et 11, et en parallèle elles créent de nouveaux sites accepteurs à 17pb, 11pb, 2pb de la variation, avec des scores : 0.89 / 0.85 / 0.99 respectivement.
- Toutes les variations exoniques sensées affecter les ESR sont sans effet, car leurs scores varient entre 0 et 0.06
- Les IVS+1GA n'ont pas un effet sur les sites accepteurs (score < 0.2), en revanche, elles conduisent d'une part à la destruction des sites donneurs constitutifs avec des scores 0.97 / 0.99 / 0.99 pour l'exon 5, 6, et 11 respectivement. Et d'une autre part à la création de nouveaux sites donneur à la position -33pb, 3pb, et 37 des variations, avec des scores : 0.39 / 0.38 / 0.94
- Les IVS+2CT n'ont pas d'effet significatif sur les sites accepteur (score < 0.2), mais Splice AI prédit une possibilité de la perte des sites donneurs pour l'exon 6 et 11 avec des score 0.4 / 0.31 respectivement et à la création de nouveaux sites à 2pb et 37pb de la variation avec des scores de 0.25 et 0.65 respectivement. Aussi, il a été trouvé que le site donneur de l'exon 5 est insensible à cette variation avec un score de 0.05 qui est largement < à 0.2

III.1.3.4. HexoSplice :

Tableau 7: Le score le Δ tESRseq des variations exoniques d'intérêts à base du logiciel HexoSplice

APC exon	Variant c.nomenclature	Variant p.nomenclature	HexoSplice score change	Total ESRseq
Exon 5	c.602A>T	p.(Glu201Val)	-5,5746	
Exon 6	c.688C>A	p.(Arg230Ser)	-4,9384	
Exon 11	c.1532G>T	p.(Gly511Val)	-3,0097	

Prédictions pour les variations exoniques susceptibles d'altérer les éléments de régulation exonique Enhancer ou Silencer, en effet une à une modification potentielle des éléments de régulation est considérée si le Δ tESRseq < -0,70, et on remarque que nos 3 variations ont des valeurs très négatives donc probablement, elles vont affecter le processus d'épissage.

III. 2 Test fonctionnel Minigène

Le test fonctionnel basé sur l'utilisation de minigène a été choisi afin d'évaluer l'impact des variations sur l'épissage des exons 5, 6, et 11 de gène APC. Les profils d'épissage des transcrits produits à partir des minigènes avec des inserts sauvage et mutés, après transfection des cellules HÉLa ont été étudiés à l'aide de la technique de RT-PCR, suivie d'une analyse par électrophorèse et d'un séquençage des produits correspondants.

De plus, une RT-PCR Fluorescente a été mis à profit afin de quantifier avec précision les effets sur l'épissage (Meulemans.L, Identification fonctionnelle des variations génétiques à l'origine d'une modification de l'épissage sans perte d'activité du produit du gène : le modèle de l'exon 12 du gène BRCA2, 2017)

III.2.1 RT-PCR semi quantitative :

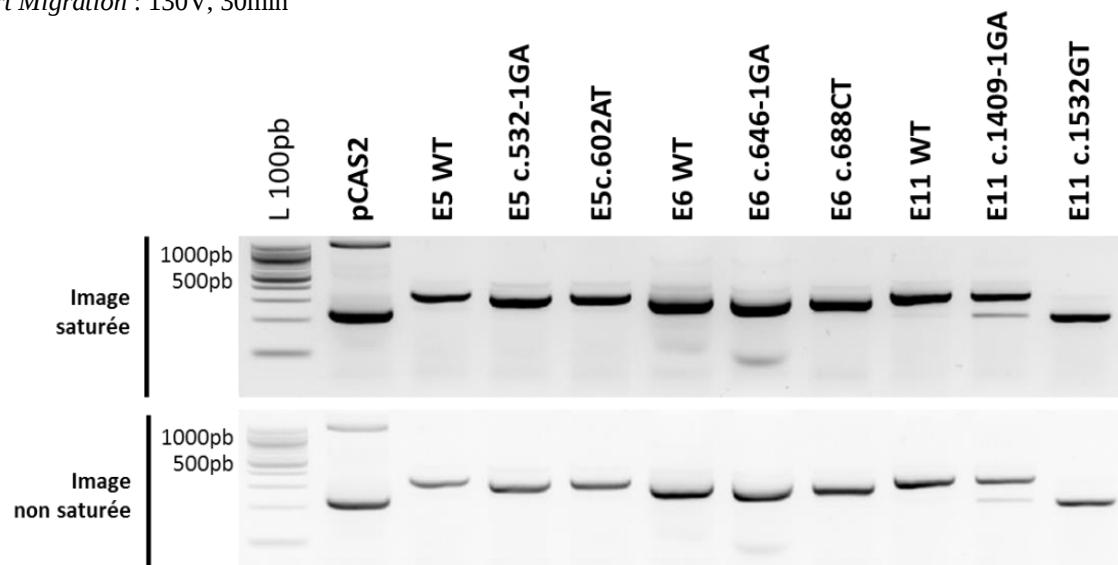
Le vecteur minigène sans insert (pCAS2 vide) et les vecteurs minigènes avec l'insert « APC Exon 5, 6, et 11 », dans le contexte sauvage ou muté, ont été transfectés dans les cellules HÉla.

Après extraction des ARN totaux, les profils d'épissage des transcrits produits à partir de ces différents minigènes ont été analysés par RT-PCR à l'aide des amorces pCAS-KOI-F et pCAS-2R localisées, respectivement, dans les exons A et B du minigène pCAS2.

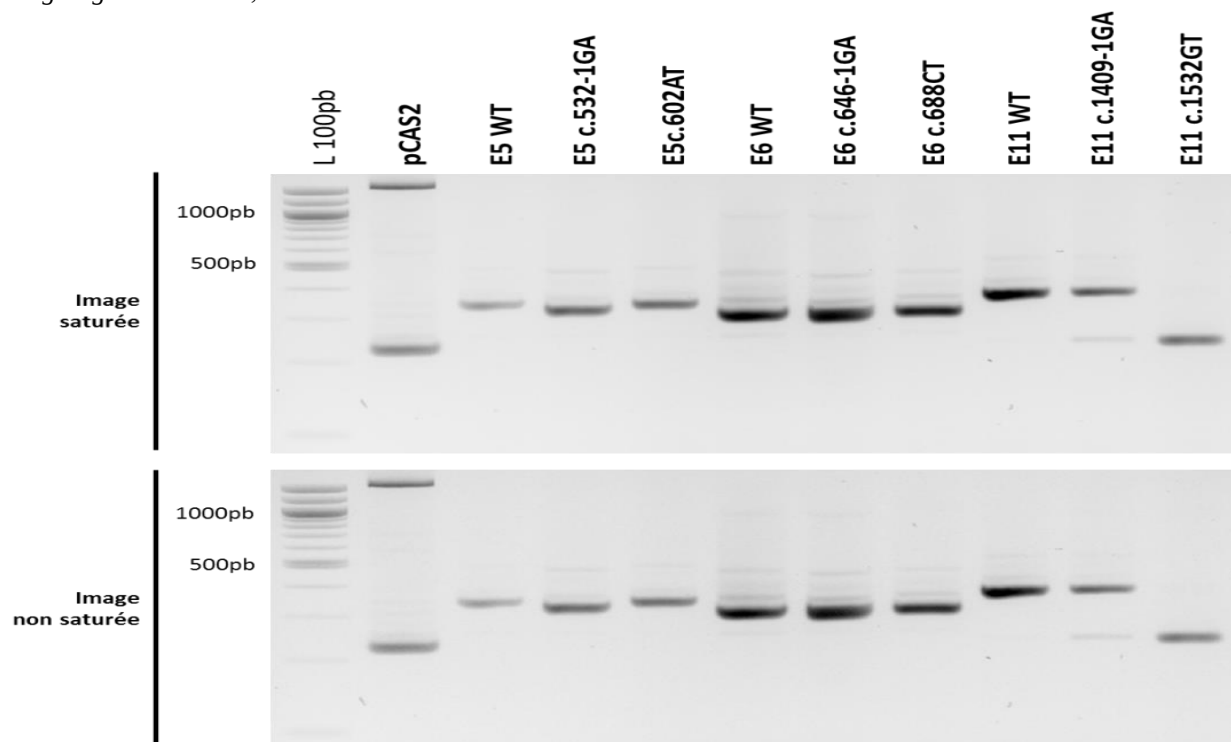
III.2.1.1 Migration :

SET#2

Short Migration : 130V, 30min

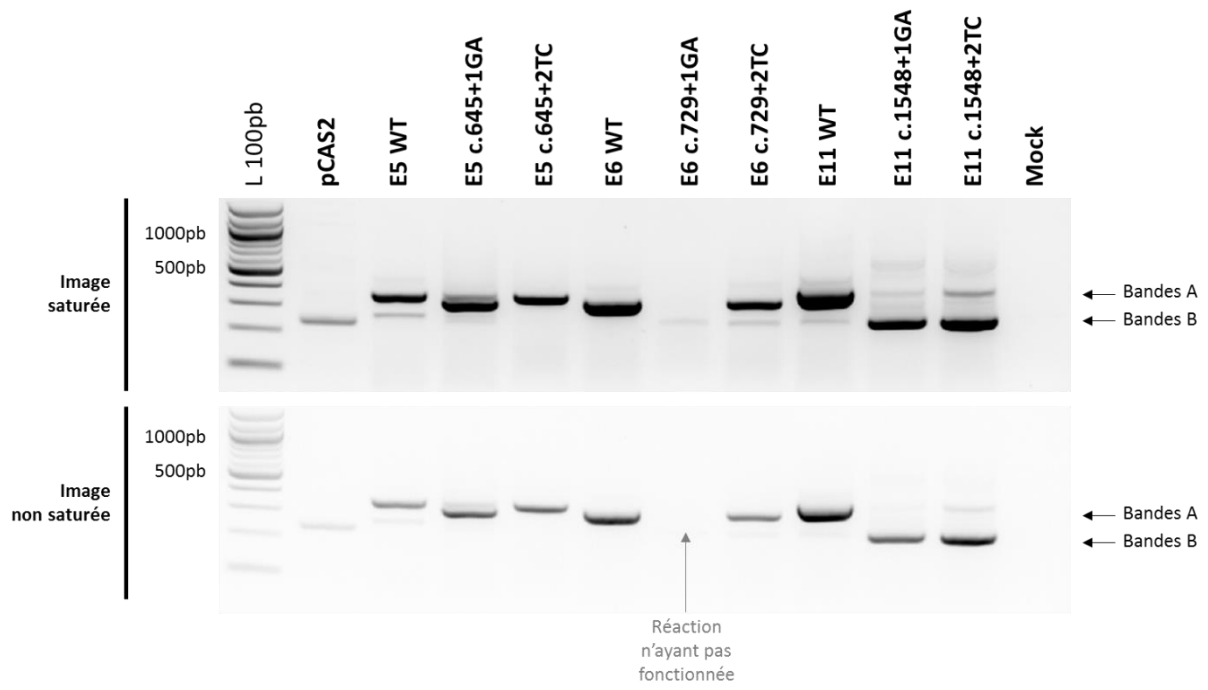


Long Migration : 120V, 30min

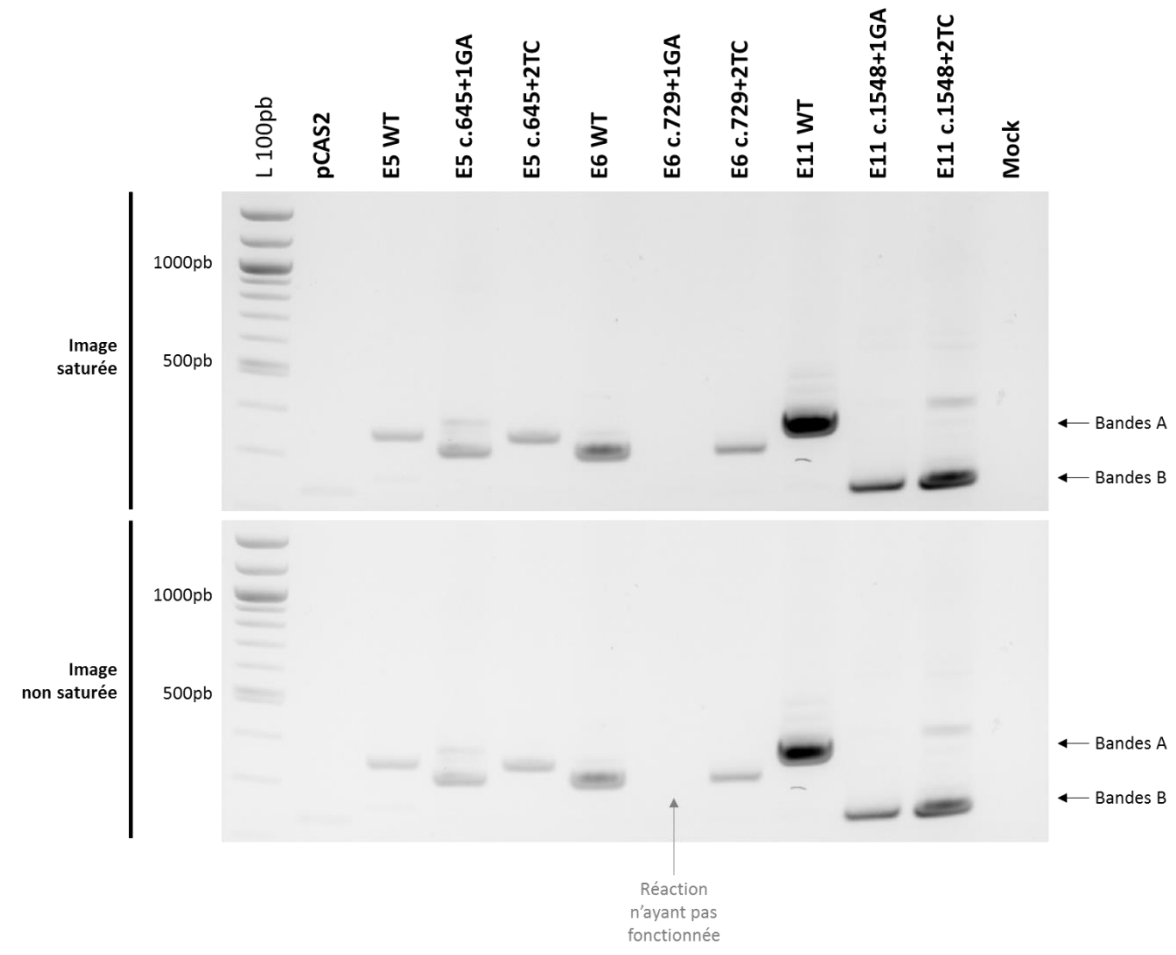


SET#1

Short Migration : 130V, 45min



Long Migration : 130V, 45min



Deuxième réaction pour avoir le profil de la c.729+1GA :

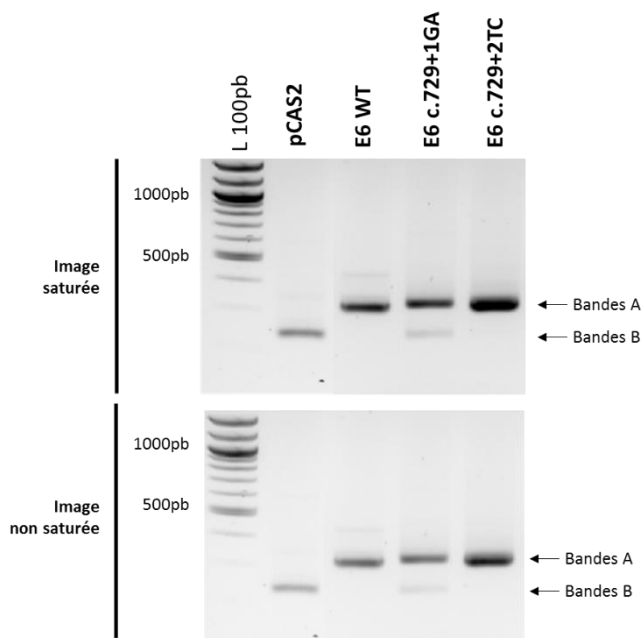


Figure 24: l'effet sur l'épissage des 12 variations, sélectionnées dans l'exon 5, 6, 11 de APC, analysées à l'aide du test fonctionnel minigène

Cette figure présente l'électrophorèse sur gel d'agarose 2% (m/v), en présence de BET, des produits de RT-PCR, en conditions saturante et non saturantes d'exposition aux rayons ultra-violets.

III.2.1.2 Séquençage :

Tableau 8: Résultats issus de l'analyse des alignements après séquençage

Type du fragment analysé		Bande 1	Bande 2	Conclusions
pCAS2	vide	Séquence Mitochondriale	EA+EB	Contamination avec ADNg
pCAS2+ Exon 5	E5 WT	E5 FL	/	Exon5
	c.532-1GA	EA+E5-16nuc+EB	/	Délétion de 16 premiers nucléotides de l'exon 5
	c.602AT	EA+E5(FL)+EB	/	Aucun effet
	c.645+1GA	EA+E5-32nuc+EB	EA+E5-32nuc+EB	Délétion de 32 derniers nucléotides de l'exon 5
	c.645+2TC	E5 FL	/	Aucun effet
pCAS2+ Exon 6	E6 WT	E6 FL	/	Exon6
	c.646-1GA	EA+E6-5nuc+EB	EA+E6-5nuc+EB	Délétion de 5 premiers nucléotides de l'exon 6
	c.688CA	EA+E6(FL)+EB	EA+E6+EB	Aucun effet
	c.729+1GA	EA+E6+4nuc+EB	EA+EB (Δ E6)	*Rétention de 4 premiers nucléotides intronique *Délétion de l'exon 6
	c.729+2TC	EA+E6(FL)+EB	/	Aucun effet
pCAS2+ Exon 11	E11 WT	EA+E11(FL)+EB	/	Exon11
	c.1409-1GA	EA+E11-1nuc+EB	EA+EB (Δ E11)	*Délétion du premier nucléotide exonique *Délétion de l'exon 11
	c.1532GT	EA+EB (Δ E11)	/	Délétion de l'exon 11
	c.1548+1GA	Hétéroduplexe	EA+EB (Δ E11)	*Délétion de l'exon 11 *Rétention intronique ?
	c.1548+2TC	Hétéroduplexe	EA+EB (Δ E11)	*Délétion de l'exon 11 *Rétention intronique ?

Le pCAS2 devait donner une seule bande correspondant à l'exon A et B, par contre deux bandes ont été obtenues. Le séquençage a révélé que la séquence nucléotidique du fragment en haut qui après alignement par Blast.ncbi (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) elle

s'aligne parfaitement avec une séquence d'ADN mitochondriale, ce qui nous permet de conclure que l'extrait est contaminé par ADN génomique.

- La bonde apparente pour le vecteur pCAS2 « APCexon5WT », correspond à l'exon A, l'exon 5 complet(FT), et l'exon B, c'est le cas également pour le produit des deux variations : c.602AT et IVS c.645+2TC (les sites d'épissages 5'ss et 3'ss sont restés insensibles à ces deux variations) par contre la variation IVS c.532-1GA, conduit à une délétion de 16 premiers nucléotides de l'exon 5, et la variation c.645+1GA engendre une délétion de 32 premiers nucléotides de cet exon, c-à-d, elles altèrent le site d'épissage 3'ss constitutif et créent un nouveau site accepteur 3'ss à 17pb et 33pb du site WT.
- Le produit de la variation c.729+2TC, et c.688CA correspond à celui issu du vecteur pCAS2 « APCexon6WT », ce qui implique qu'une substitution de T en C en position +2 pour l'exon6 d'APC et C en A en position 688, n'ont aucun effet sur l'épissage. En revanche la variation c.646-1GA provoque la délétion de 5 nucléotides de l'exon6, et la c.729+1GA, génère deux fragments de taille différentes (séparés en 2 bondes), l'un correspond à un saut total de l'exon6 $\Delta E6$, et l'autre à une rétention de 4 premiers nucléotides de l'intron 6.
- Toutes les variations sélectionnées pour l'exon11 (IVS -1GA / +1GA / +2CT, et la c.1532GT) conduisent au saut de cet exon $\Delta E11$, en plus de ça la variation IVS-1GA génère aussi une délétion du premier nucléotide de l'exon, et les IVS+1 et +2 ont donné aussi une bande, qu'après séquençage a révélé un hétéroduplexe, ce qui nous laisse supposer en se basant sur Splice AI, que cet hétéroduplexe se compose d'un saut total d'exon $\Delta E11$ et une rétention intronique (cette supposition a été testée par RT-PCR Fluo).

III.2.2 RT-PCR **Fluo** quantitative :

Après une deuxième transfection des cellules Hela (n=2) avec le vecteur minigène sans insert (pCAS2 vide) et les vecteurs minigènes avec l'insert « APC Exon 5, 6, et 11 », dans le contexte sauvage ou muté, les ARN totaux ont été extraits et convertis en ADNc par une RT-PCR fluo avec l'amorce pCASKOI-F fluo et pCAS2-R localisées, respectivement, dans les exons A et B du minigène pCAS2.

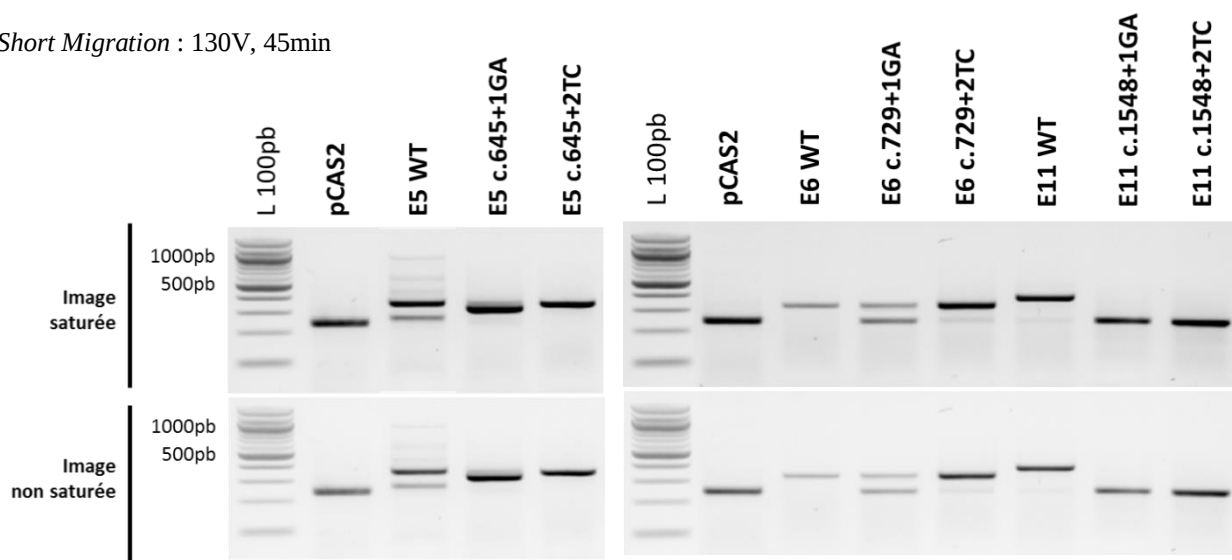
Cette réaction de RT-PCR fluo a été optimisée par le bon choix des cycles : 32 cycles pour E5 et 26 cycles pour E6 et E11 d'après les gradients de cycles.

NB :

Cette manipulation a été uniquement réalisée pour le SET#1 : E5, E6, E11 : WT/ IVS+1GA / IVS+2CT, du fait de problèmes techniques rencontrés lors des manipulations

III.2.2.1 La migration :

Short Migration : 130V, 45min



Long Migration : 130V, 45min

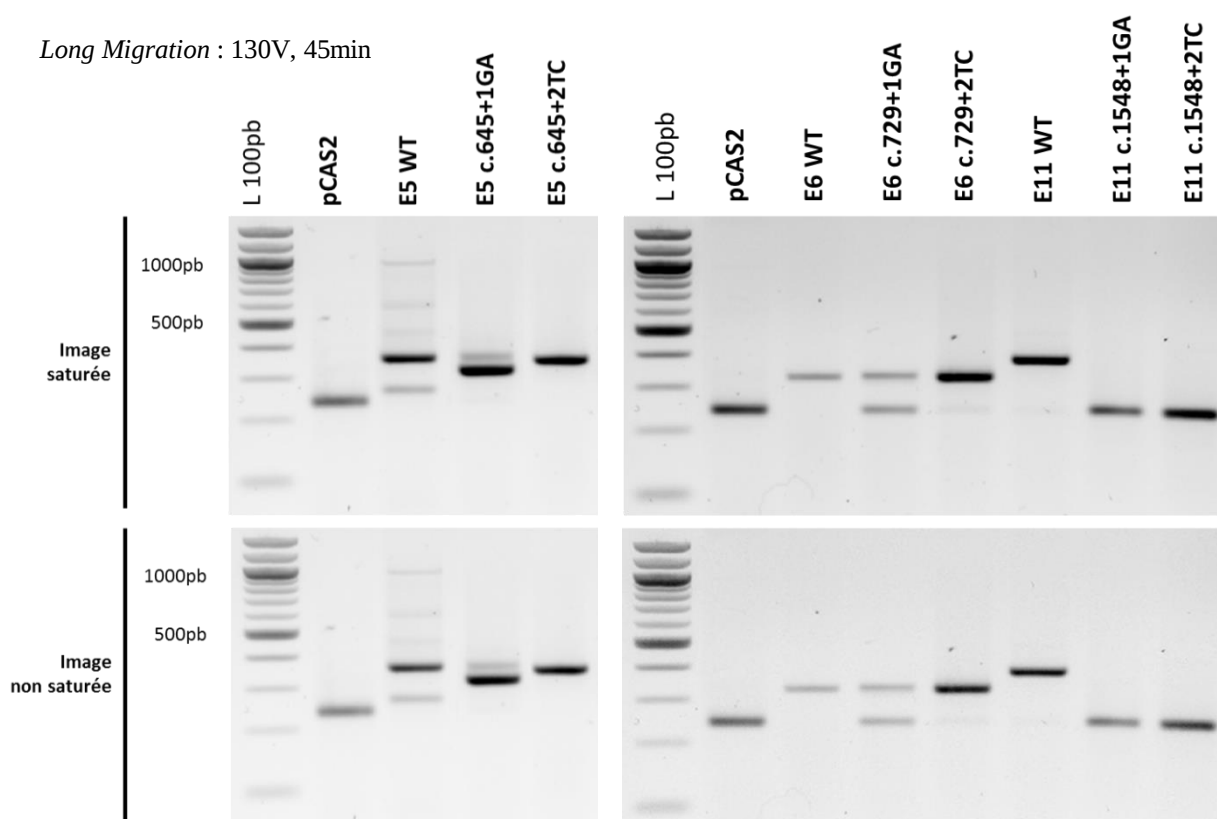


Figure 25: Séparation des produits RT-PCR fluo (profils d'épissage) par électrophorèse sur gel d'agarose 2%

III.2.2.2 Gene Scan:

Après dénaturation des produits RT-PCR Fluo, ils ont été séparés selon leur taille par migration sur séquenceur, les taux relatifs (%) des différents transcrits calculés sur la base de l'aire des pics de fluorescence obtenus sont représentés sous forme d'histogramme.

	E5 WT	c.645+1GA	c.645+2TC
+E5 349pb	67	5	100
ΔE5q32 317pb	0	91	0
? 256pb	31	0	0
ΔE5 235pb	2	4	0

	E6 WT	c.729+1GA	c.729+2TC
+E6 319pb	100	0	88
ΔE6p4 323pb	0	33	4
ΔE6 235pb	0	67	8

	E5 WT	c.645+1GA	c.645+2TC
+E11 374pb	95	0	0
ΔE11 335pb	5	100	100

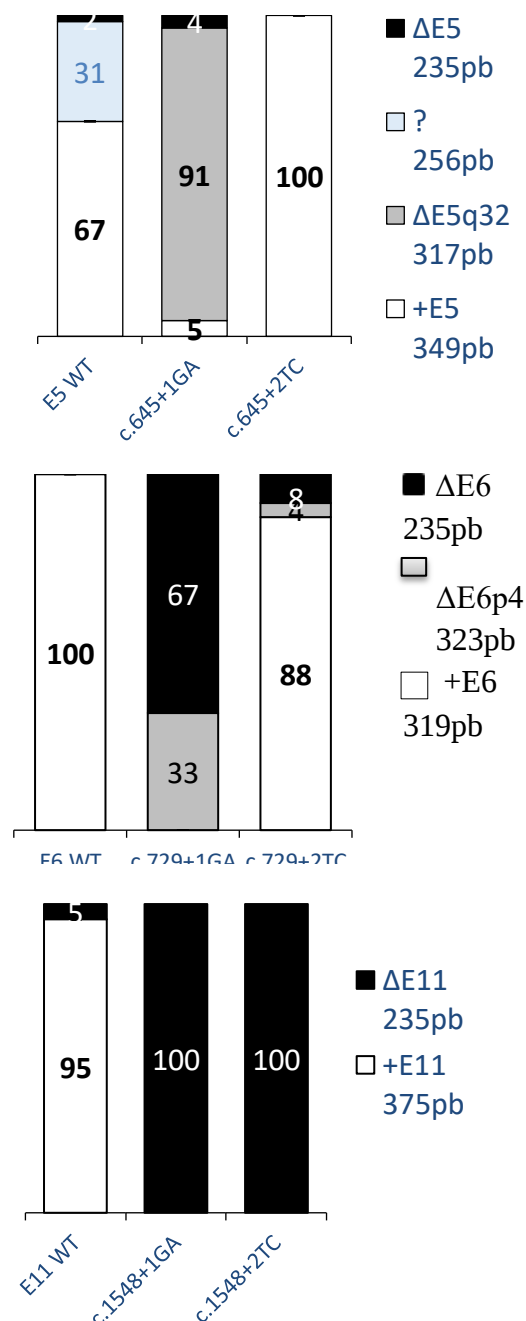


Figure 26: Quantification relative, par RT-PCR fluorescente, des transcrits produits à partir des minigènes « APC exon 5, 6 et 11 » sauvage et mutés (variations +1 et +2)

Analyse de l'effet des variations sur l'exon 5 :

- Le Gene Scan a révélé d'une manière intéressante que le vecteur pCAS2 « APC E5WT » a généré trois différents fragments visibles sur RT-PCR Fluorescente mais non visibles sur gel : 67% des transcrits sont de taille 317pb qui correspond à l'inclusion normale de l'exon5, 2% uniquement ont été obtenus, de saut total de l'exon 5 (235pb), et 31% d'un transcrit de taille 256pb qui a été préalablement séquencé lors de du séquençage des produits RT-PCR qui correspond à une séquence mitochondriale.
- La variation c.645+1GA est responsable d'un défaut majeur, car le transcrit majoritaire 91% correspond à un saut totale d'exon 5, et uniquement 5% avec inclusion normale de l'exon. Et les 4% qui reste sont des transcrits qui ont une délétion de 32 nucléotide du début de cet exon.

Analyse de l'effet des variations sur l'exon 6 :

- Tous les transcrits du vecteur WT incluent l'exon 6 complet (100%)
- Les transcrits majoritaires de la variation c.729+1GA (67%) ont un saut total de l'exon6 et les 33% qui restent incluent l'exon6 suivi de 4 nucléotides introniques (rétention intronique $\Delta E6p4$) donc elle induit un défaut partiel de l'épissage, tandis que la variation c.729+2TC a un effet plus ou moins prononcé avec 88% de l'inclusion normale de l'exon 6 et seulement 8% du saut total de l'exon et 4% correspond à une rétention intronique de 4 nucléotides

Analyse de l'effet des variations sur l'exon 11:

- A l'état sauvage cet exon est à 95% inclus dans les transcrit et absent dans les 5% qui reste (saut d'exon)
- Les deux mutations IVS+1GA et IVS+2TC concernant cet exon, sont responsable d'un défaut majeur d'épissage, (+exon11 = 0%)

DISCUSSION

IV. Discussion

L'interprétation des variations trouvées dans des gènes impliqués dans les maladies héréditaires à transmission mendéliennes, et le cancer, présentent un défi majeur dans le domaine de la génétique médicale (Richars.S, 2015). Dans le cas du gène *APC*, qui est un gène suppresseur de tumeur, responsable du syndrome FAP avec une pénétrance complète et prédispose au cancer colorectal et à d'autres manifestations extra-colorectales (Neo.O, 2021). Le gène *APC* présente un large spectre mutationnel, en effet 9563 des variations nucléotidiques sont répertoriées dans la base de données (ClinVar) et 58% parmi elles sont considérées comme des variations à signification inconnue, qui ne peuvent pas être interprétées dans le cadre médicale.

Dans ce projet de recherche nous nous sommes intéressés aux effets de certaines variations localisées dans les jonctions exon/intron (IVS-1 +1 +2), exoniques et sur le processus d'épissage du pré ARNm d'*APC*, en utilisant deux techniques complémentaires :

Dans un premier temps, la technique d'analyses *in silico* (interrogation des bases de données, et les prédictions bio-informatiques) a été sélectionnée, afin de stratifier les variations. Ces outils permettent la prédiction de la force de chaque site d'épissage constitutif ou cryptique dans le contexte sauvage, et dans le contexte muté (en présence des variations). Ces prédictions, indiquent la diminution ou la destruction de la force des sites d'épissage constitutif, et la création ou l'activation de sites cryptiques à distance des jonctions naturelles exons/introns (Hartmann.L, 2008), (Houdayer.C, 2012).

Les résultats *in silico* seuls sont insuffisants pour l'identification précise de l'effet des variations sur la nature des transcrits générés d'un point de vue qualitatif et quantitatif. En effet ils ne sont pas mis à profit pour une interprétation biologique et clinique, mais ils sont très utilisés pour réaliser une première stratification des variations à analyser dans les tests fonctionnels d'épissage. (Meulemans.L, 2021). Ainsi, Les variations sélectionnées ont fait l'objet d'une étude expérimentale, par un test fonctionnel (minigène) indicateur d'anomalies d'épissage, dans un contexte génomique et cellulaire artificiel. Cette techniques est largement utilisée et validée par plusieurs laboratoires de recherche et de diagnostic moléculaire (Baralle.M, 2003) (Cooper.TA., 2005) (Kishore.S, 2008) (Gaildrat.P K. M., 2010) (Lin.J.H T. B., 2019) (Lin.J.H W. Z., 2021).

Les critères utilisés pour la sélection des variations sont basés sur l'interrogation des logiciels de prédictions sur toutes les jonctions exon/intron (IVS+1 +2). De façon intéressante, les IVS+2 T>C des exons 5, 6 et 11 d'*APC*, ont un score dans le logiciel Splice AI < 0.5 ce qui

prédit que cette variation n'affecte pas le site naturel donneur d'épissage 5'ss, or que les variations introniques localisées au niveau des positions invariables AG/GT (-2,-1/+1,+2) des sites d'épissage sont parmi les variations systématiquement classées pathogènes (Baralle.D L. B., 2009).

La fiabilité de Splice AI a été confirmée par plusieurs travaux indépendants : l'équipe de Chen Jian-Min, en 2020 a mené une étude qui avait pour but l'évaluation de la performance de l'outil SpliceAI basé sur l'intelligence artificielle (deep learning), en comparant ses scores avec les 160 variations dont les effets fonctionnels sur l'épissage ont été confirmés dans des études antérieures basées sur des analyses expérimentales (Chen.J.M, 2020) (Wai HA, 2020) et (Saint-Martin.C, 2021).

En fonction des résultats obtenus, les exons 5, 6 et 11 ont été sélectionnés pour IVS+2 T>C, les mêmes exons ont été maintenus pour sélectionner des IVS+1 G>A, car elles sont répertoriées dans les bases de données comme pathogènes à l'exception de la +1 G>A de l'exon 6 (non répertoriée). A ce stade deux mutations dans chacun des exons 5, 6 et 11 ont été sélectionnées (SET 1#).

Durant la phase expérimentale pour ce SET 1#, des analyses *in silico* ont été menées, dans le but de visualiser la force d'épissage naturelle de tous les exons d'APC, en interrogeant des logiciels de prédictions bio-informatique notamment MaxEntScan et SSF-L dans l'interface Alamut Visual. Ces deux logiciels ont été approuvés avec un autre logiciel HSF, dans des études qui ont visé à vérifier la concordance des résultats des données expérimentales sur les différents gènes avec les prédictions émises de ces logiciels (Sharp.A, 2004) (Tournier.I, 2008) (Sharma.N, 2014) (Théry.J.C, 2011) (Houdayer.C, 2012).

A partir des scores issus de MaxEntScan et SSF-L, qui concordent, il a été enregistré que la force des sites d'épissage en 5' et 3' des exons d'intérêts sont dans la moyenne par rapport à tous les exons d'APC à l'exception du site 5'ss de l'exon 5 qui a la valeur la plus forte. Une étude plus poussée a été réalisée afin d'avoir une vision globale sur la force d'épissage des jonctions exon/intron du gène APC par rapport à toutes les jonctions du génome humain en sollicitant le prédicteur bio-informatique SROOGLE (Schwartz.S H. A., 2009), utilisé dans plusieurs études (De Boer.M, 2017) (Thakur.P.K, 2019) (Crespo.C, 2020).

Dans un second temps, une variation IVS-1 a été retenue pour chacun des exons 5, 6 et 11, et les substitutions -1 G>A ont été sélectionnées, car elles étaient répertoriées dans les bases de données comme pathogènes, à l'exception de l'exon 6 (non répertorié). De plus, une substitution exonique pour chacun des exons à la base des prédictions du logiciel de

HexoSplice a été choisie, basé sur des évaluations expérimentales et donne un score ESRseq. Si ce dernier est positif, la séquence sera prédite comme ESE (favorise l'inclusion de l'exon), si la valeur est négative donc c'est un ESS (contribue à l'exclusion de l'exon). Dans le cas où le score est nul donc l'élément est neutre (Ke.S, 2011).

Les résultats dérivés de cette étude a mené au développement du score $\Delta tESRseq$, qui est obtenu par le calcul de la différence entre la somme des scores ESRseq dans le contexte muté et sauvage, si la valeur est $< -0,70$, donc l'élément de régulation est affecté par la variation (Di Giacomo.D, 2013). Le choix est porté sur la c.602A>T / c.688C>A / c.1532G>T pour l'exon 5, 6, et 11 respectivement, car elles ont des scores $\Delta tESRseq$ très négatifs. Ces trois variations sont artificielles car elles ne sont pas trouvées dans les bases de données. (SET 2#)

L'outil HSF a renseigné les forces des sites d'épissage dans le contexte sauvage et en présence de la variation, il a été constaté que les variations IVS-1+1+2 induisent à la diminution de la force des sites d'épissage, contrairement aux variations exoniques d'intérêts qui favorisent l'augmentation de la force des sites d'épissage.

Ces prédictions pour quelques variations ont pu être démontré expérimentalement, en effet les variations exoniques c.602AT et c.688CA de l'exon 5 et 6 respectivement, n'ont aucun effet sur la qualité du transcrit car une seule bande a été obtenue lors de la migration sur gel, et le séquençage a révélé l'inclusion totale de l'exon 5 et 6 (l'absence d'effets sur l'épissage est prédite par Splice AI). Or que la variation c.1532GT localisé au niveau de l'exon 11 est responsable de la destruction du site d'épissage, induit le saut total de l'exon qui n'est pas en phase ce qui engendre un décalage de cadre de lecture et l'apparition d'un codon stop prématuré dans l'exon 12 à la position c.1605, ce qui confère à cette variation le caractère pathogène.

Il est à noter que les différents logiciels utilisés dans notre étude n'ont pas pu prévenir l'effet de cette variation.

La RT-PCR qualitative suivie par le séquençage a pu révéler de façon impressionnante, des cas où des sites cryptiques (détectés par MaxEntScan et SSF-L sur l'interface Alamut visual) ont été activés du côté 5' des exons après destruction des sites constitutif 3'ss par les variations, notamment les IVS-1GA des exons 5, 6, 11 respectivement.

*La c.532-1GA conduit à la délétion de 16 premiers nucléotides de l'exon5, et ça engendre un décalage de cadre de lecture et l'apparition d'un codon stop prématuré au niveau de l'exon 5 à la position c.551.

*La c.646-1GA conduit à la production d'un transcrit aberrant avec une délétion de 5 premiers nucléotides de l'exon 6, induisant un décalage de cadre de lecture et l'apparition d'un codon stop prématuré dans l'exon 7 à la position c.750.

*La c.1409-1GA engendre une délétion du 1^{er} nucléotide de l'exon 11, induit un frameshift, et l'apparition d'un codon stop prématuré au niveau de l'exon 11 à la position c.1490.

Tandis que le GeneScan a révélé que les variations IVS+1GA engendrent un défaut majeur d'épissage, en effet :

*La c.645+1GA conduit seulement à 4% de l'inclusion normale de l'exon, 5% au saut total en phase de l'exon (pas de décalage de cadre de lecture) et à 91% des transcrits aberrants avec une délétion de 32 derniers nucléotides de l'exon 5 par l'activation d'un site cryptique, et ça engendre un décalage de cadre de lecture et un codon stop prématuré sur l'exon 7 à la position c.750.

*La c.729+1GA responsable d'un saut partiel en phase de l'exon 6 ($\Delta Ex6 = 67\%$) et 33% qui reste ont la rétention de 4 premiers nucléotides de l'intron 6 par la création d'un nouveau site cryptique d'épissage, et ça induit à un décalage de cadre de lecture et la survenue d'un codon stop prématuré sur l'exon 7 à la position c.750.

*La c.1548+1GA responsable de la destruction du site 5' d'épissage, induit le saut total de l'exon 11, et ça engendre un décalage de cadre de lecture et la survenue d'un codon stop prématuré sur l'exon 10 à la position c.1605

En revanche pas toutes les variations IVS+2TC, ont un effet néfaste sur l'épissage, en effet :

*La c.645+2TC n'as aucun effet sur l'épissage de l'exon 5 (100% inclusion de l'exon).

*La c.729+2TC, conduit à 3 types de transcrits, le majoritaire 88%, à l'inclusion normal de l'exon 6, 8% du saut de l'exon, et 4% correspondent à la rétention des 4 premiers nucléotides de l'intron6.

Seulement la c.1548+2TC est responsable d'un défaut majeur d'épissage avec un saut total de l'exon ($\Delta Ex11 = 100\%$), mais ça ne concorde pas avec les résultats de la RT-PCR qualitative, qui après séquençage a révélé un hétéroduplexe.

Il est à noter que ces résultats expérimentaux, ont montré de façon surprenante, que les profils d'épissage visualisés sur gel ne sont pas toujours les mêmes dans des expériences indépendantes pour les mêmes variations, et les résultats GeneScan des vecteurs pCAS2 APC E5 et E11 WT (contexte sauvage) engendrent en grande partie des transcrits FL (inclusion de

l'exon) mais aussi en faible concentration des transcrits avec le saut de l'exon, or tous les transcrits de pCAS2 E6 WT sont FL, et ça ne concorde pas avec les prédictions bio-informatiques sur la force des sites d'épissage constitutifs, car SSF-L, MaxEntScan et SROOGLE ont appréhendé que nos exons d'intérêts 5, 6, et 11 ont des sites forts notamment celui de l'exon 5. Cela nous laisse suggérer que le test minigène ne reproduit pas forcément les événements présents dans le contexte physiologique car nous n'avons pas l'intégralité du gène (uniquement l'exon d'intérêt et ses régions introniques flonquantes), et la cellule utilisée pour la transfection HÉLa est cancéreuse (stress cellulaire). En effet la fiabilité de cette technique est récemment testée et mis en cause par (Lin.J.H W. Z., 2021), qui ont conclu que cette approche est limitée pour certains types de variations, car ses résultats ne concordent pas avec les résultats obtenus à partir des échantillons de patients, mais aussi entre eux dans des expériences indépendantes.

D'une manière paradoxale, cette technique Minigène est largement utilisée dans le but d'établir un lien direct de causalité entre la variation et les événements aberrants d'épissage car l'analyse des minigènes sauvages et mutés est réalisée en parallèle ce qui rend l'interprétation des données plus aisée (Cooper.T.A, 2005). Dans la bibliographie plusieurs études ont démontré la pertinence de cette approche pour la détection des défauts d'épissage induit par les variations. Ces études ont rapporté que la majorité des résultats obtenus dans le contexte des minigènes présentaient une très grande concordance avec les données établies à partir du matériel biologique de patients. Ainsi, ils peuvent être utilisées pour les interprétations biologiques et cliniques (Tournier.I, 2008) (Gaildrat.P K. D., 2012) (van der Klift.H.M, 2015) (Soukariéh.O, 2016) (Baralle.D B. , 2017) (Tubeuf.H, 2020).

CONCLUSION

V. Conclusion

Lors de cette étude, par défaut de temps dédié à ce mémoire qui est de 2 mois et les problèmes techniques rencontrés lors des manipulations, seulement 2 transfections pour chaque variation du SET1# (variation IVS+1+2) ont été réalisées, et une seule transfection pour le SET2# (IVS-1 et exonique), or qu'on devait faire 3 transfections pour affirmer les résultats obtenus. La technique GeneScan (RT-PCR Fluo quantitative) est uniquement réalisée pour le SET1#. Les deux autres types de variation SET2# ont été analysées uniquement par une RT-PCR classique (qualitative).

Les résultats obtenus ne peuvent pas être utilisés pour des interprétations cliniques, qui exigent une confirmation par d'autres études complémentaires.

Une étude comparative ne pourra pas être réalisée, car selon nos recherches les variations choisies dans cette étude, n'ont pas fait objet d'études antérieures.

En revanche les résultats obtenus ont confirmé d'une part la pertinence du logiciel de prédiction bio-informatique SpliceAI qui prédit le degré du changement de la force du site d'épissage en présence de la variation mais aussi l'éventuelle création ou activation des sites cryptiques en amont et en aval pour certaines variations mais pas la totalité, ce qui a été déjà rapporté par (Chen.J.M, 2020). De plus, ces données soulignent l'importance de tester expérimentalement l'effet de toutes les variations qui existent sur l'épissage que ce soit en position -2-1/+1+2 ou exonique, afin de déterminer leurs conséquences réelles sans généraliser par rapport au type et la localisation des variations. Il a été constaté dans cette étude que les conséquences des variations ne sont pas les mêmes si le type de la variation et sa localisation par rapport à l'exon est le même.

A présent, aucun outils bio-informatique n'est capable de prédire la qualité et la quantité des transcrits engendrés en présence et en absence de la variation qui concordent à 100% avec le contexte physiologique.

Une étude plus poussée est nécessaire pour démontrer le caractère pathogène ou non de ces variations, en analysant des données génétiques, clinique, et familial des patients porteurs de ses variations.

Les logiciels bio-informatique doivent être performant afin que les chercheurs et cliniciens puissent compter sur leurs prédictions pour une interprétation clinique et biologique solide, en effet il est quasiment impossible de passer toutes les variations à signification inconnus (VUS) par des analyses expérimentales. Ils peuvent notamment se baser et s'investir plus sur

l'intelligence artificielle car à présent, uniquement SpliceAI rapporte des prédictions proches des conséquences réelles sur l'effet des variations sur l'épissage.

RÉFÉRENCES

BIBLIOGRAPHIQUES

VI. Références bibliographiques :

Le Cancer. (2016, 11 09). Récupéré sur Fondation Contre le Cancer:

<https://www.cancer.be/le-cancer/quest-ce-que-le-cancer>

Cancer. (2022). Consulté le 09 2022, sur Organisation Mondiale de la Santé:

<https://www.who.int/fr/news-room/fact-sheets/detail/cancer>

Baralle,D, B. (2017). RNA splicing in human disease and in the clinic. *Clinical Science*., 355-368.

Baralle,D, L. B. (2009). Missed threads. The impact of pre-mRNA splicing defects on clinical practice. *EMBO Rep*, 10, 810-816.

Baralle.M, B. D. (2003). Identification of a mutation that perturbs NF1 agene splicing using genomic DNA samples and a minigene assay. 40. *J Med Genet*., 40, 220–2.

Ben-Ze'ev.A, G. (1998). Differential molecular interactions of beta-catenin and plakoglobin in adhesion, signaling and cancer. *Current Opinion in Cell Biology*, 629-639.

Beroud, M. C. (1992). *The UMD website*. Consulté le 02-05 2022, sur <http://www.umd.be/>

bio-informatique, B. e. (s.d.). *P25054 · APC_HUMAN*. Consulté le 08 30, 2022, sur UniProt: <https://www.uniprot.org/uniprotkb/P25054/entry#function>

Bodart.J.F, B. B. (2007). Uniciel. Lille, France.

cancer, L. l. (s.d.). *Polypes et adénomes colorectaux*. Consulté le 03 2022, sur Association pour le dépistage du cancer colorectal en Alsace (ADECA): https://www.adeca68.fr/le_cancer_colorectal/histoire_naturelle/polypes_et_adenomes_colorectaux.178.html

Chellat-Rezgoune, D. (2021). *Cours génotoxicologie*. Constantine: University of Constantine 1.

Chen.J.M, L. M. (2020). The Experimentally Obtained Functional Impact Assessments of 5' Splice Site GT>GC Variants Differ Markedly from Those Predicted. *Current Genomics*, 21, 56-66.

- Cheng.J, e. a. (2019). MMSplice : modular modeling improves the predictions of genetic variant effects on splicing. *Genome Biology*.
- Cooper.T.A. (2005). Use of minigene systems to dissect alternative splicing elements. *Methods*, 331-40.
- Crespo.C, E. O. (2020). Molecular analysis of GALT gene in Argentinian population: Correlation with enzyme activity and characterization of a novel Duarte-like allele. *Molecular Genetics and Metabolism Reports*, 100695.
- De Boer.M, L. G. (2017). Mutation in an exonic splicing enhancer site causing chronic granulomatous disease. *Blood Cells, Molecules, and Diseases*, 66, 50-57.
- Den Dunnen.J.T, e. a. (2016). HGVS Recommendations for the Description of Sequence Variant: 2016 Update. *Human Mutation*.
- Desmet.F.O, H. L. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nuclein Acids Research*, 1-14.
- DH5 α Competent Cells*. (s.d.). Consulté le 05 2022, sur ThermoFisher SCIENTIFIC: <https://www.thermofisher.com/fr/fr/home/life-science/cloning/competent-cells-for-transformation/competent-cells-strains/dh5a-competent-cells.html>
- Di Giacomo.D, G. A. (2013). Functional Analysis of a Large set of BRCA2 exon 7 Variants Highlights the Predictive Value of Hexamer Scores in Detecting Alterations of Exonic Splicing Regulatory Elements. *Human Mutation*, 1547-57.
- Escande.F, R. (2015, juillet-août-septembre). Nomenclature des variations de séquence en génétique : application en génétique somatique. *Correspondances en Onco-Théranostic*, pp. Vol. IV - n° 3 119-124.
- Ferlay.J, E. L. (2020, 10). *Global Cancer Observatory: Cancer Today*. Consulté le 09 2022, sur International Agency for Research on Cancer: <https://gco.iarc.fr/today>
- Fodde.R. (2002). The APC gene in colorectal cancer. *European Journal of Cancer*, 867-871.
- Gaildrat.P. (2021). *Méthodes d'évaluation du caractère pathogène des variations génétiques*. Rouen: Université de Rouen, Normandie.

- Gaildrat.P, K. D. (2012). Multiple sequence variants of BRCA2 exon 7 alter splicing regulation. *J Med Genet*, 609-17.
- Gaildrat.P, K. M. (2010). Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. *Methods Mol Biol*, 653, 249-5.
- Hartmann.L. (2008). Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? *Front Biosci.*, 3252.
- He.T.C, S. A. (1998). Identification of c-MYC as a target of the APC pathway. *Science*, 1509-12.
- Heinen.C.D, G. C. (2002). The APC tumor suppressor controls entry into S-phase through its ability to regulate the cyclin D/RB pathway. *Gastroenterology*, 751-63.
- Houdayer.C, C.-M. K. (2012). Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Hum Mutat*, 1228-38.
- How to make primers for PCR.* (s.d.). Consulté le 05 2022, sur Sawakinome:
<https://fr.sawakinome.com/articles/science/how-to-make-primers-for-pcr.html>
- Huret.J.L. (2008, 10 01). *Atlas of Genetics and Cytogenetics in Oncology and Haematology*. Consulté le 06 08, 2022, sur Nomenclature pour la description des mutations et autres variations de séquence:
<http://atlasgeneticsoncology.org/teaching/30103/nomenclature-pour-la-description-des-mutations-et-autres-variations-de-s-eacute;quence>
- Jaganathan.K, e. a. (2019). Prediction Splicing from Primary Sequence with Deep Learning. *Cell*, 535-548.
- Jaiswal.A.S, N. (2008). A novel function of adenomatous polyposis coli (APC) in regulating DNA repair. *Cancer Lett*, 272-80.
- Jinchen.L, e. a. (2018). Performance evaluation of pathogenicity-computation methods for missense variants. *Nucleic Acide Research*.
- Jorden.BR. (2013). Chronique génomique Henrietta Lacks et les cellules Héla, es leçons d'une hisoire. *Médecine/Sciences*.

- Karczewski.k. (2017, 02 27). *MacArthur Lab*. Consulté le 02-05 2022, sur <https://macarthurlab.org/2017/02/27/the-genome-aggregation-database-gnomad/>
- Ke.S, S. K. (2011). Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Research*, 21, 1360-74.
- Kebir.S, D. (2015-2016). *Cours de génétique* . ANNABA: Université BADJI MOKHTAR-ANNABA, Faculté de médecine .
- Kishore.S, K. S. (2008). Rapid generation of splicing reporters with pSpliceExpress. *Gene*, 427, 104–10.
- Laboratories, D. (2014, 10). *The UMD-APC mutations database* . Consulté le 02-05 2022, sur <http://umd.be/APC/gene.shtml>
- Landrum.M, L. R. (2013). ClinVar. *The NCBI Handbook [Internet]. 2nd edition. Bethesda (MD): National*.
- Lefebvre.A, M. L. (2020). *HExoSplice: a new software based on overlapping hexamer scores for prediction and stratification of exonic variants altering splicing regulation of human genes*. Consulté le 08 22, 2022, sur [bioinfo.univ-rouen.fr: http://bioinfo.univ-rouen.fr/HExoSplice/](http://bioinfo.univ-rouen.fr/bioinfo.univ-rouen.fr/HExoSplice/)
- Leoz.M.R, e. a. (2015). The genetic basis of familial adenomatous polyposis and its implications for clinical practice and risk management . *The application of Clinical Genetics*.
- Lin.J.H, T. B. (2019). First estimate of the scale of canonical 5' splice site GT>GC variants capable of generating wild-type transcripts. *Human Mutation*, 1856-0873.
- Lin.J.H, W. Z. (2021). Splicing Outcomes of 50 Splice Site GT>GC Variants That Generate Wild-Type Transcripts Differ Significantly Between Full-Length and Minigene Splicing Assays. *Frontiers in Genetics*, 701652.
- Liu.Y, P. B. (2007). Coordination of steps in single-nucleotide base excision repair mediated by apurinic/apyrimidinic endonuclease 1 and DNA polymerase β . *J Biol Chem*, 13532-41.
- Manolio.T.A, e. a. (2009). Finding the Missing Heritability of Complex Diseases. *Nature*.

- Médiste, L. R. (2018, 04 03). *Cancer du côlon : qu'est-ce que le polype adénomateux ?*
Consulté le 03 2022, sur Médisite: <https://www.medisite.fr/cancer-colon-rectum-cancer-du-colon-quest-ce-que-le-polype-adenomateux.2789084.38943.html>
- Meulemans.L. (2017). *Identification fonctionnelle des variations génétiques à l'origine d'une modification de l'épissage sans perte d'activité du produit du gène : le modèle de l'exon 12 du gène BRCA2*. Rouen: Université de Rouen-Caen, France.
- Meulemans.L. (2021). *Caractérisation fonctionnelle de variations splicéogéniques à l'origine d'anomalies d'épissage en phase dans des gènes de prédisposition aux cancers : implications en génétique médicale*. Rouen: Université de Rouen Normandie.
- NCBI.NLM.NIH. (s.d.). Consulté le 09 09, 2022, sur ClinVar APC:
<https://www.ncbi.nlm.nih.gov/clinvar/?term=APC%5Bgene%5D&redir=gene>
- Neo.O, F. R. (2021). Adenomatous polyposis coli in cancer and therapeutic implications. *Oncology Reviews (pageeress)*, 524.
- Nieuwenhuis.M.H, V. (2006, 07 06). Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): A review of the literature. *Oncology Hematology*, 61, 153-161.
- Polakis.P. (1995). Mutations in the AK gene and their implications for pro&in structure and function. *Current Opinion in Genetics & Development* , 66-71.
- Rahman.N. (2014). Realizing the promise of cancer predisposition genes. *Nature*, 302-308.
- Richards.S, A. B. (2015, May). Standardsand Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5), 405-424.
- Rubinfeld.B, A. P. (1996). Binding of GSK3 β to the APC- β -catenin complex and regulation of complex assembly. *Science*, 1023-6.
- Saint-Martin.C, C.-L. M. (2021). Functional characterization of ABCC8 variants of unknown significance based on bioinformatics predictions, splicing assays, and protein analyses: Benefits for the accurate diagnosis of congenital hyperinsulinism. *Human Mutation*, 24164.

- Schwartz.S, H. A. (2009). SROOGLE : webserver for integrative, user-friendly visualization of splicing signals. *Nucleic Acids Research*, 189-192.
- Scotti.M.M, S. (2016). RNA mis-splicing in disease. *Nature Reviews Genetics*, 19-32.
- Sharma.N, S. R. (2014). Experimental Assessment of Splicing Variants Using Expression Minigenes and Comparison with In Silico Predictions. *Human Mutation*, 1249-59.
- Sharp.A, P. L. (2004). RNA analysis reveals splicing mutations and loss of expression defects in MLH1 and BRCA1. *Hum Mutat*, 272-272.
- Sligielski.E M, e. a. (2000). dbSNP : a database of single nucleotide polymorphisms. *Nucleic Acids Research*, 352-355.
- Soukariéh.O, G. H. (2016). Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. *PLOS Genet*, 1005756.
- Thakur.P.K, R. O. (2019). Bioinformatics Approaches for Studying Alternative Splicing. *Life Science*, 2, 221-234.
- Théry.J.C, K. G. (2011). Contribution of bioinformatics predictions and functional splicing assays to the interpretation of unclassified variants of the BRCA genes. *Eur J Hum Genet*, 1052-8.
- Tournier.I, V. M. (2008). A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. *Hum Muta*, 1412-24.
- Tubeuf.H, C. S. (2020). Calibration of Pathogenicity Due to Variant-Induced Leaky Splicing Defects by Using BRCA2 Exon 3 as a Model System. *Cancer Res*, 1811-29.
- van der Klift.H.M, J. v. (2015). Splicing analysis for exonic and intronic mismatch repair gene variants associated with Lynch syndrome confirms high concordance between minigene assays and patient RNA analyses. *Mol Genet Genomic Med*, 327-45.
- Wai HA, L. J. (2020). Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance. *Genet Med*(22), 1005-14.
- Yeo.G, B. (2004). Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *Journal of Computational Biology*, 2-3.

ANNEXES

VII. Les annexes :

1/ Présentation des différentes amorces utilisées au cours de cette étude

Nom de l'amorce	Type de l'amorce	Localisation de l'amorce	Séquence nucléotidique	Taille fragment amplifié	Utilisation
APC.ex5.BamHI.InFus-F	sens	Intron 4 d'APC	AGGCTAAGAAGTGCA <u>GGATCC</u> AGATTGAGTC TGACACC	502	PCR et mutagenèse dirigé
APC.ex5.MluI.InFus-R	Anti-sens	Intron 5 d'APC	AGGGGTCAAAACAAG <u>ACGCGT</u> GATTATGTTGC TCAGCAGCCATG		
APC.Ex6.InFus.BamHI-F	sens	Intron 5 d'APC	AGGCTAAGAAGTGCA <u>GGATCC</u> GAACTGACCC CAATTTGTTATTAAAGG	542	
APC.Ex6.InFus.MluI-R	Anti-sens	Intron 6 d'APC	AGGGGTCAAAACAAG <u>ACGCGT</u> TGTAAACTGA CAGCTAAAGTAAGGTATC		
APC.Ex11.InFus.BamHI-F	sens	Intron10 d'APC	AGGCTAAGAAGTGCA <u>GGATCC</u> GGCATAAAAT GGAATAATTGTCAGTTGTAC	567	
APC.Ex11.InFus.MluI-R	Anti-sens	Intron11 d'APC	AGGGGTCAAAACAAG <u>ACGCGT</u> GGCATAAGAC AAACAAATGAGTAAAGATAAG		
Exon 5					
APCE5-c.532-1GA-F	sens	Intron 4-Exon 5	TGTGGTTTTTA <u>A</u> TTTTTCCTTACAA	317	Mutagenèse dirigé
APCE5c.602AT-R	Anti-sens	Exon 5	TAGTTGTTCT <u>A</u> CCATCGCAACT	285	
APCE5-c.645+1GA-F	sens	Exon5-Intron5	AACGAGCACAG <u>A</u> TAAGTTACTT	203	
APCE5-c.645+2TC-R	Anti-sens	Exon5-Intron5	AGTAACTT <u>G</u> CCTGTGCTCGT	328	
Exon 6					
APCE6-c.646-1GA-F	sens	Intron5-Exon6	GTTTTATTTTA <u>A</u> CGAAGAATAGC	377	Mutagenèse dirigé
APCE6c.688CA-R	Anti-sens	Exon 6	TGTCGTATACT <u>A</u> AAGTATGTCC	255	
APCE6-	sens	Exon6-	ACAGAAGCAGAG <u>A</u> TTAGTAAAT	293	

c.729+1GA-F		Intron6			
APCE6- c.729+2TC-R	Anti- sens	Exon6- Intron6	ATTTACTA GC CTCTGCTTCTG	296	
Exon 11					
APCE11- c.1409-1GA-F	sens	Intron10- Exon11	TTTTAAATTA A GGGGACTACAG	359	
APCE11c.1532 GT-R	Anti- sens	Exon11	GGCTACATCT A CAAAAGTCAA	384	Mutagenèse dirigé
APCE11- c.1548+1GA-F	sens	Exon11Intro n11	GTAGCCAACAAG A TATGTTTTT	220	
APCE11- c.1548+2TC-R	Anti- sens	Exon11Intro n11	AAAAACAT GC CTTGTTGGCTA	400	
pCAS2					
pCAS-seq-F	sens	Intron du pCAS2	GGGGTCAATAGCAGTGAGAG		Séquençage de l’ADN plasmidique
pCAS-seq-R	Anti- sens	Intron du pCAS2	GCTCCATTTCACAGGTAGAGA		
pCAS-KOI-F	sens	Exon A du pCAS2	TGACGTCGCCGCCCATCAC		Analyse du profil d’épissage par RT-PCR et séquençage des produits RT-PCR
pCAS-2R	Anti- sens	Exon B du pCAS2	ATTGTTGTTGAGTTGGTTGTC		

Les nucléotides en :

Rouge : ont été modifiés par rapport à la séquence WT, pour introduire la variation d'intérêt par le biais de la technique de mutagenèse dirigée.

Noir : complémentaire à la partie exonique

Gris : complémentaire à la partie intronique de l'exon

Noir : complémentaire à la partie exonique du pCAS2 (exon A et B)

Gris : complémentaire à la partie intronique du pCAS2

Noir : site de restriction BamHI et MluI

2/ Les séquences référence de APC-insert (transmises par Meulemans Laëtitia)

2.1/ E5

>APCE5-MinipCAS2-ref

ATCAAAGCAAGAGTTTGTGTAGAAAAGAGCTGTGGGGAGGAGAGGTTTGGCTGCTGTGACACAG
ACGGTGGTCCCAGGCTAGGGCCTGTGAAGACAGGGAAGAATGGTAGCAAACATACGTGGGTCTGG
GTG **GGGGTCAATAGCAGTGAGAG**GTCCCAGTGGTCTGGGAGGTCATGGACTAGAAAATAAGGCTAA
GAAGTGCA **GGATCC**AGATTGAGTCTGACACCTATAATCAAATTTAAACACTCCTTGGAGTAAAAA
ATAATTTTCTCATGCACCATGACTGACGTATTTGCTTATTCATTTTCTTTATTGGTTCTTATATGCTTT
TTTGCTTTTACTGATTAACGTAAATACAAGATATTGATACTTTTTTATTATTTGTGGTTTTAGTTTTTC
CTTACAAACAGATATGACCAGAAGGCAATTGGAATATGAAGCAAGGCAAATCAGAGTTGCGATGG
AAGAACAACCTAGGTACCTGCCAGGATATGGAAAAACGAGCACAGGTAAGTTACTTGTTTCTAAGT
GATAAAACAGCGAAGAGCTATTAGGAATAAAATGAATTACAGCTCTGTTAATATTGATTAAATTTT
ATTAAGACATAAGGCTGTGTTTATTTTGGCTCTATTTCAAATAAGATTTATCATGGCTGCTGAGC
AACATAATC **ACGCGT**CTTGTTTTGACCCCTAGCTGTTTCTCCATTTCTGTGAGACCCTGGACTAGT
TAATATGTGTCGGTATGAGTT **TCTCTACCTGTGAAATGGAGC**TAATGCGTGGTTTCTCAGTGGAGG
ATCTGTTGACATTGGGGCTGGATAATTCATTGCTGAGGGGCTACCCTGTGCATTTTAGGATGTTTTG
CAGTATATCTCTAGCCTCTACGCACTAGATTCCGGTAGCATCTCCCTGCTGTCTCCCTGTGACAAC
CAAAAATGTCTCTG

pCASseq-F

pCASseq-R

>APCE5-ADNc-ref

TGACGTCGCCGCCATCACGCCTCCAGGCTGACCCTGCTGACCCTCCTGCTGCTGCTGGCTGGG
TTTTCTTACAAACAGATATGACCAGAAGGCAATTGGAATATGAAGCAAGGCAAATCAGAGTTGC
GATGGAAGAACAACCTAGGTACCTGCCAGGATATGGAAAAACGAGCACAG **GATAGAGCCTCCTCAA**
ATCCAAATGCTACCAGCTCCAGCAGCCAAGATCCAGAGAGTTTGCAAGACAGAGGCGAAGGGAAG
GTGCAACAACAGTTATCTCCAAGATGCTATTCGTTGAACCCATCCTGGAGGTTTCCAGCTTGCC **GA**
CAACCAACTCAACAACCAAT

pCASKOI-F

pCAS-2R

>APCE6-Mini-Ref

ATCAAAGCAAGAGTTTGTGTTAGAAAAGAGCTGTGGGGAGGAGAGGTTTGGCTGCTGTGACACAG
ACGGTGGTCCCAGGCTAGGGCCTGTGAAGACAGGGAAGAATGGTAGCAAACATACGTGGGTCTGG
GTG**GGGGTCAATAGCAGTGAGAG**GTCCCAGTGGTCTGGGAGGTCATGGACTAGAAATAAGGCTAA
GAAGTGCA**GGATCC**GAACTGACCCCAATTTGTTATTAAAGGGTGAATATATTTATATGTCTAGCTT
TTTAAATGAGAATGATTTGACATAACCCTGAGCTTTTAAGTGGTAGCCATAGTATGATTATTTCTAT
TAATATTATTAATAAAAAACATAACTAATTAGGTTTCTTGTTTTATTTTAGCGAAGAATAGCCAGAAT
TCAGCAAATCGAAAAGGACATACTTCGTATACGACAGCTTTTACAGTCCCAAGCAACAGAAGCAG
AGGTTAGTAAATTGCCTTTCTTGTTTGTGGGTATAAAAAATAGGTAGTTATTCTGAGAAAAGAAAAC
ATGTATAATTTAATGTGACACCATTGAAATATAGATGTTCTTTTCAGAGAATTTAAATACCGTAATTT
TTTTCGTGAAATTAAATTATCAAAGATTTGGACTATTTTGATTTTATCTAACTTTTAGGCAGTTAAA
ATTTATAAAACTGTAAATATAGATACCTTACTTTAGCTGTCAGTTTACA**ACGCGT**CTTGTTTTGACC
CCTAGCTGTTTCTCCCATTTCTGTGAGACCCTGGACTAGTTAATATGTGTCGGTATGAGTT**TCTCTA**
CCTGTGAAATGGAGCTAATGCGTGTTTTCTCAGTGAGGATCTGTTGACATTGGGGCTGGATAATT
CATTGCTGAGGGGCTACCCTGTGCATTTTAGGATGTTTTGCAGTATATCTCTAGCCTCTACGCACTA
GATTCCGGTAGCATCTTCCCTGCTGTCTCCCTGTGACAACCAAAAATGTCTCTG

>APCE6-ADNc-Ref

TGACGTCGCCGCCATCACGCCTCCAGGCTGACCCTGCTGACCCTCCTGCTGCTGCTGCTGGCTGGG
CGAAGAATAGCCAGAATTCAGCAAATCGAAAAGGACATACTTCGTATACGACAGCTTTTACAGTCC
CAAGCAACAGAAGCAGAG**GATAGAGCCTCCTCAAATCCAAATGCTACCAGCTCCAGCAGCCAAGA**
TCCAGAGAGTTTGCAAGACAGAGGCGAAGGGAAGGTCGCAACAACAGTTATCTCCAAGATGCTAT
TCGTTGAACCCATCCTGGAGGTTTCCAGCTTGCC**GACAACCAACTCAACAACCAAT**

2.3/ E11

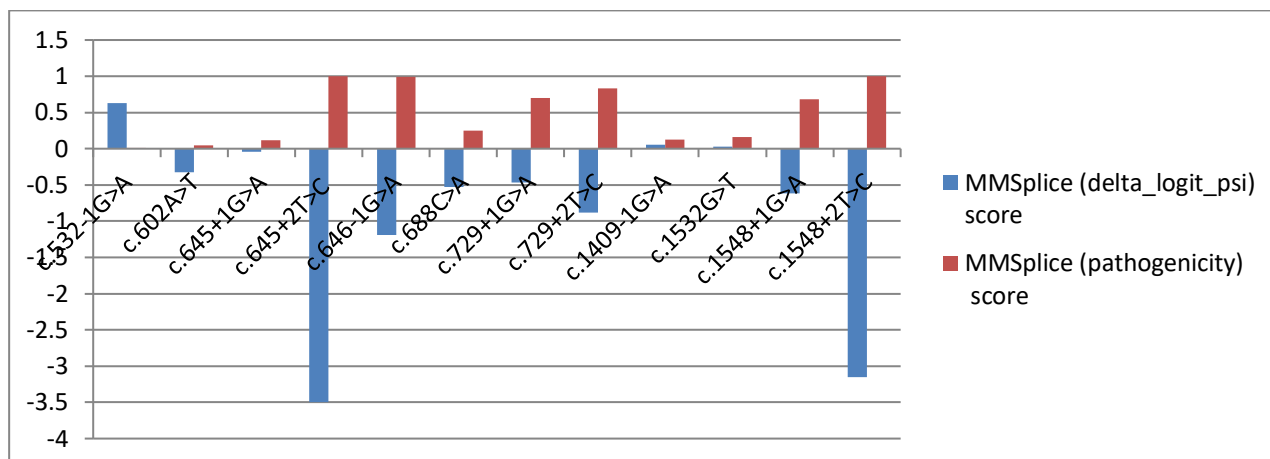
>APCE11-Mini-Ref

ATCAAAGCAAGAGTTTGTGTAGAAAAGAGCTGTGGGGAGGAGAGGTTTGGCTGCTGTGACACAG
ACGGTGGTCCCAGGCTAGGGCCTGTGAAGACAGGGAAGAATGGTAGCAAACATACGTGGGTCTGG
GTG**GGGGTCAATAGCAGTGAGAG**GTCCCAGTGGTCTGGGAGGTCATGGACTAGAAATAAGGCTAA
GAAGTGCA**GGATCC**GGCATAAAATGGAATAATTGTCAGTTGTACTTTATAAATATATTATACAGAA
GTTCTTTATAACAGTTTTTTGTAGCTTATAATTCTAAAGGCAAATTTAAACCATATATTCTCATTGATT
GAGTTTTTTTTTTCCTAGTATTTAAGTTACCAACTTGGTACCAGTTTGTTTTATTTTAGATGATTGTCT
TTTTCTCTTGCCCTTTTTAAATTAGGGGGACTACAGGCCATTGCAGAATTATTGCAAGTGGACTGT
GAAATGTATGGGCTTACTAATGACCACTACAGTATTACACTAAGACGATATGCTGGAATGGCTTTG
ACAAACTTGACTTTTGGAGATGTAGCCAACAAGGTATGTTTTTATAACATGTATTTCTTAAGATAGC
TCAGGTATGAGTTAATTTACTTTCATACAAATACATTTTACTGATTTTCTTTTTTTTCACTCTCCTCAT
TAAACAATGACTGATAAAAACCTGTGCTTCACATTCGCTTATCTTTACTCATTTGTTTGTCTTATGCC
ACGCGTCTTGTTTTGACCCCTAGCTGTTTCTCCCATTTCTGTGAGACCCTGGACTAGTTAATATGTG
TCGGTATGAGTT**TCTCTACCTGTGAAATGGAGC**TAATGCGTGGTTTCTCAGTGAGGATCTGTTGAC
ATTGGGGCTGGATAATTCATTGCTGAGGGGCTACCCTGTGCATTTTAGGATGTTTTGCAGTATATCT
CTAGCCTCTACGCACTAGATTCCGGTAGCATCTTCCCTGCTGTCTCCCTGTGACAACCAAAAATGTC
TCTG

>APCE11-ADNc-Ref

TGACGTCGCCGCCATCACGCCTCCAGGCTGACCCTGCTGACCCTCCTGCTGCTGCTGGCTGGG
GGGGACTACAGGCCATTGCAGAATTATTGCAAGTGGACTGTGAAATGTATGGGCTTACTAATGACC
ACTACAGTATTACACTAAGACGATATGCTGGAATGGCTTTGACAACTTGACTTTTGGAGATGTAG
CCAACAAG**GATAGAGCCTCCTCAAATCCAAATGCTACCAGCTCCAGCAGCCAAGATCCAGAGAGT**
TTGCAAGACAGAGGCGAAGGGAAGGTCGCAACAACAGTTATCTCCAAGATGCTATTCGTTGAACC
CATCCTGGAGGTTTCCAGCTTGCC**GACAACCAACTCAACAACCAAT**

3/ Prédictions MM Splice :



Histogramme qui représente les scores de delta-splicing-efficiency et les scores de pathogénicité des variations d'intérêts (Cheng J N. C., 2019)

4/ Prédictions HEXplorer et HAL sur les variations exoniques

APC exon	Variant c.nomenclature	Variant p.nomenclature	HEXplorer Delta HZEI	HAL DELTA_PSI WT=95%
Exon 5	c.602A>T	p.(Glu201Val)	-121.6	-90.7
Exon 6	c.688C>A	p.(Arg230Ser)	-88.16	-84.4
Exon 11	c.1532G>T	p.(Gly511Val)	-122.16	-89.1

5/ Les positions en g. c. et p. du coté 3' et 5' de chaque exon d'APC

Exon	g. 3'	g. 5'	c. 3'	c. 5'	p. 3'	p. 5'
-1	112 073 582	112 073 582	-59	-19	0	0
1	112 090 570	112 090 722	-18	135	1	45
2	112 102 023	112 102 107	136	220	46	74
3	112 102 886	112 103 087	221	422	74	141
4	112 111 326	112 111 434	423	531	141	177
5	112 116 487	112 116 600	532	645	178	215
6	112 128 143	112 128 226	646	729	216	243
7	112 136 976	112 137 080	730	834	244	278
8	112 151 192	112 151 290	835	933	279	311
9	112 154 663	112 155 041	934	1312	312	438
10	112 157 593	112 157 688	1313	1408	438	470
11	112 162 805	112 162 944	1409	1548	470	516
12	112 163 626	112 163 703	1549	1626	517	542
13	112 164 553	112 164 669	1627	1743	543	581
14	112 170 648	112 170 862	1744	1958	582	653
15	112 173 250	112 179 823	1959	8532	653	2844