



HAL
open science

Cinétique mutationnelle et transcriptionnelle des cancers colorectaux microsatellites instables, sporadiques et héréditaires (i.e. le syndrome de Lynch) : de la crypte MMR-déficiente à l'adénocarcinome

Hugo Montémont

► **To cite this version:**

Hugo Montémont. Cinétique mutationnelle et transcriptionnelle des cancers colorectaux microsatellites instables, sporadiques et héréditaires (i.e. le syndrome de Lynch) : de la crypte MMR-déficiente à l'adénocarcinome. Génomique, Transcriptomique et Protéomique [q-bio.GN]. Sorbonne Université, 2023. Français. NNT : 2023SORUS603 . tel-04847303

HAL Id: tel-04847303

<https://theses.hal.science/tel-04847303v1>

Submitted on 19 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sorbonne Université

École Doctorale Physiologie, Physiopathologie et Thérapeutique (ED394)

Centre de Recherche Saint-Antoine UMR_S 938, INSERM

Équipe « Instabilité des microsatellites et cancers »

THÈSE DE DOCTORAT

Spécialité Génétique des Cancers

**Cinétique Mutationnelle et Transcriptionnelle des Cancers
Colorectaux Microsatellites Instables, Sporadiques et
Héréditaires (*i.e.* le syndrome de Lynch) : de la Crypte MMR-
déficiente à l'Adénocarcinome**

Par Hugo MONTÉMONT

Dirigée par le Docteur Ada COLLURA

Présentée et soutenue publiquement le 18 décembre 2023

Devant un jury composé de :

Pr. Franck BIELLE

Président du jury

Pr. Julie PANNEQUIN

Rapporteuse

Dr. Manuel RODRIGUES

Rapporteur

Pr. Pierre SAINTIGNY

Examineur

Dr. Rémy NICOLLE

Examineur

Dr. Ada COLLURA

Directrice de thèse

Pr. Alex DUVAL

Directeur du laboratoire

À mon grand-père Loulou,

À ma grand-mère Ninou,

À mon amie Brigitte,

Remerciements

Je tiens à remercier le professeur **Franck BIELLE** pour avoir accepté d'être le président de mon jury de thèse. Merci au professeur **Julie PANNEQUIN** et au docteur **Manuel RODRIGUES** qui m'ont fait l'honneur de juger mon travail et de le rapporter. Je remercie aussi le docteur **Pierre SAINTIGNY** ainsi que le docteur **Rémy NICOLLE** pour avoir accepté d'examiner mes travaux de recherche.

Je remercie grandement le professeur **Alex DUVAL** de m'avoir accueilli au sein de son équipe. Merci pour ta confiance, tes conseils et ta disponibilité.

Je tiens à remercier chaleureusement le docteur **Ada COLLURA** pour avoir dirigé mes travaux de thèse. Voilà presque 5 ans (depuis que je suis venu faire un stage en 3^{ème} année de licence) que l'on travaille ensemble et que tu encadres mes projets de recherche. Depuis, je n'ai plus quitté le labo, où j'ai pu développer ma passion pour la recherche translationnelle en cancérologie, et cela en grande partie grâce à toi. Lorsque tu m'as proposé ce super projet mêlant biologie, anatomopathologie et bio-informatique, tu as piqué ma curiosité. Pendant la thèse tu m'as donné une grande autonomie que j'ai beaucoup apprécié, tout en m'accompagnant dès que j'en ai eu besoin dans les moments difficiles. Je retiens aussi la confiance entière que tu m'as donné, tes nombreux conseils mais aussi nos multiples moments de rigolade qui ont rythmé mon épanouissement dans ma thèse. Notre travail ensemble a confirmé ma volonté de poursuivre une carrière de recherche académique, et pour cela je te suis entièrement reconnaissant et je te remercie infiniment pour tout !!

Je remercie la **Ligue Nationale Contre le Cancer** qui a financé mes quatre années de thèse.

Merci aux docteurs **Alexandre HOW-KIT** et **Sophie GARNIER** pour avoir participé à mes comités de suivis de thèse. Vos conseils ont été précieux pour faire mûrir mon projet de thèse.

J'adresse un immense merci à tous les membres du Centre de Recherche Saint-Antoine qui m'ont accompagné durant ces quatre dernières années :

Au professeur **Bruno FÈVE**, directeur du CRSA et au professeur **Xavier HOUARD**, directeur de l'École Doctorale 394. Merci pour votre disponibilité et votre soutien continuels à mes travaux de thèse, comme à mon investissement au sein du centre recherche. À l'équipe administrative et plus particulièrement à **Sandrine, Nicolas, Ibrahima, Olgine et Nadine** qui ont su répondre à mes questionnements et se sont souvent rendus disponibles.

À tous mes collègues de l'équipe MSI. Tout d'abord à **Vincent** avec qui je publie en co-premier auteur mon article de thèse ; merci de m'avoir transmis ton savoir en bio-informatique et ta maîtrise dans l'élaboration de figures (l'art de représenter simplement des résultats complexes). À **Pierre, Aurélie et Quentin**, je vous suis entièrement reconnaissant pour l'expertise et l'appui technique que vous m'avez apporté pendant ma thèse, notamment pour l'élaboration du protocole ILAM-NGS ; sans vous le déroulé de ma thèse aurait été tout autre. À **Florence, Magali et Sarah** qui m'ont formé et, plus que cela, m'ont trouvé une nouvelle passion pour l'anatomopathologie ; cette discipline complexe a été essentielle dans la réussite de ma thèse, et je tâcherais de continuer à l'utiliser dans la suite de mes projets de recherche. À mes supers co-thésards **Kathleen et Toky** !!! Les longues journées (semaines/mois/années) passées au labo ont été plus agréables grâce à vous, à votre gentillesse, à votre positivité, à vos mots forts pour me remonter le moral et à nos nombreux fous rires. Un immense merci pour tout cela, et je vous souhaite une grande réussite (dans le monde de la start-up nation) ! À la relève, en la personne de **Clément** qui anime le bureau depuis 1 an par son humour, sa positivité et sa bienveillance, ainsi que **Lisa, Léo, Cody, Amélia**, je vous souhaite une belle fin de thèse ! À **Philippe et Emmanuelle**, nos enseignants préférés ; merci pour les (très) longues discussions scientifiques mais pas que. À **Olivier, Aurélien, Diehl**, la team bio-info toujours disponible pour répondre à mes questions de novice mais aussi pour nos échanges philosophiques. Aux anciens stagiaires de M2 **Victoire, Alix, Marc-Antoine, Paula, Juliette, Clémentine et Assane** que j'ai eu le plaisir de rencontrer. À notre doyenne (^) **Marie-Annick** et sa gentillesse et bienveillance ultime ! Aux cliniciens sans qui, notre travail de chercheur serait bien plus compliqué. À mes anciens collègues **Martine, Coralie, Adem, Malorie, Romane** qui m'ont accueilli petit et m'ont vu grandir ; merci pour vos conseils qui continuent à forger mon projet professionnel.

Au **Club des Jeunes Chercheurs** et notamment **Laëtitia, Jeanne, Anna, Nawa, Rania, Kathleen, Julia** avec qui je me suis lancé dans cette très belle aventure et à **Lara, Tiphaine, Juliane, Carine, Alya, Fanny, Kenza, Pauline, Mathias, Sabrina, Allan** qui nous ont ensuite rejoints ; ainsi qu'au membres du CAS et notamment à **Nathalie, Marie-Laure, Anne** et **Véronique**. J'ai adoré m'investir pour le centre de recherche, dans l'animation scientifique et la valorisation des jeunes chercheurs. Cette expérience m'a permis de m'épanouir dans ma thèse et de trouver un équilibre dans ma vie au labo. J'espère que cela vous apportera la même chose et que vous continuerez à faire vivre le centre par vos activités scientifiques et sociales.

À **Paul-Patrice** pour tes mots encourageants tous les matins ! À **Hélène** et **Stéphane**, pour votre gentillesse et votre humour. À la team CRSAPEROOOO pour nos nombreux happy hour pour évacuer la pression. Et à tous mes autres potes du centre, **Ines, Pierre, Gabrielle, Sylvain...**

Je tiens également à remercier mes amis **Soufyan, Julien** – vous qui avez déjà brillamment soutenu votre thèse, je vais tâchez de marcher dans vos pas – et **Tristan**, pour les moments de détente, de jeux et de rire qui m'ont permis de m'aérer l'esprit quand j'en avais bien besoin. Aussi, j'adresse une pensée pleine d'émotions à mes amis de cœur **Nicolas, Christie** et **Fabien**. Il serait difficile d'énumérer tout ce que vous avez fait pour moi ou toutes les fois où vous avez été présents, tant votre amitié est inestimable : pour tout cela je vous dis merci. Enfin, je remercie tous mes autres amis qui m'ont accompagné et soutenu !

Je suis aussi très reconnaissant du soutien inconditionnel que **ma famille** m'a apporté tout au long de ses années de thèse, malgré la distance et les longues périodes sans se voir, ainsi que l'intérêt que vous avez toujours porté pour mon travail. Je pense notamment à ma **grand-mère Ninou**, à mes frères **Léo** et **Arthur** et à mes parents **Manu** et **Jos** (qui à leur manière ont participé à la rédaction de ce manuscrit).

Enfin, j'ai une pensée particulière à **mon grand-père Cric-Crac**, à **mon grand-père Loulou** et à **mon amie Brigitte** présents dans mon cœur. Je suis tellement triste de ne pas pouvoir partager avec vous ce moment unique qui clos un chapitre de ma vie, vous qui m'avez tant aidé à me construire personnellement et à me projeter dans mon avenir personnel et professionnel ; mais je sais que vous seriez fier et que vous continuerez à m'accompagner au gré de mes aventures.

Résumé

Le système MMR (pour *Mismatch Repair*) est un mécanisme de réparation des mésappariements de bases survenant pendant la réplication. La déficience du MMR induit une instabilité nucléotidique de l'ADN particulièrement importante dans les séquences répétées du génome (ou microsatellites). L'accumulation de nombreuses mutations entraîne la transformation des cellules et le développement de cancers appelés MSI (pour *Microsatellite Instability*). Les cancers MSI sont observés dans 10-15% des cancers colorectaux (CRC, *ColoRectal Cancer*) et ils sont caractérisés par un phénotype hypermutateur et une forte infiltration en cellules immunitaires. Dans la majorité des cas, les cancers MSI sont d'origine sporadique, et se développent suite à une modification épigénétique inactivatrice du système MMR. Cependant, dans 10% des cas, les cancers MSI sont d'origine héréditaire (mutation germinale du MMR) : c'est le syndrome de Lynch.

La tumorigenèse MSI colique se décrit en trois phases : (i) la perte d'activité du système MMR objectivée par l'apparition de cryptes MMR-déficientes pré-cancéreuses dans la muqueuse colique saine des patients atteints par le syndrome de Lynch exclusivement (initiation) ; (ii) l'accumulation de mutations qui induit la transformation en tumeur bénigne ou adénome ; (iii) la progression du cancer en tumeur maligne ou adénocarcinome. Malgré des années de recherches, les mutations clés (*drivers*) de la tumorigenèse MSI colique sporadique et héréditaire n'ont pas été formellement identifiées. L'avènement des technologies de séquençage haut-débit (de l'exome et/ou du transcriptome) offre de nouvelles opportunités pour décrire le paysage mutationnel et transcriptionnel de la tumorigenèse MSI colique, afin d'identifier les mutations et/ou les mécanismes *drivers*.

Durant mes travaux de thèse, j'ai étudié l'implication de l'instabilité microsatellitaire intronique dans l'initiation et la progression tumorale MSI colique. Nous avons démontré que cette instabilité est impliquée dans l'altération de l'épissage des ARN pré-messagers. Nous avons mis en avant une signature spécifique et *driver* de la tumorigenèse MSI colique, de 985 exons impliqués dans la prolifération, la différenciation et la survie cellulaire (Jonchère & Montémont et al., en révision dans le journal *Genome Biology*). Également, dans des travaux préliminaires, j'ai décrit l'instabilité microsatellitaire codante des phases précoces et identifié

des mutations potentiellement *drivers* de l'initiation et de la progression tumorale. Enfin, j'ai développé une approche technique pour étudier les lésions précoces pré-cancéreuses, appelée ILAM-NGS pour *Immunohistochemistry Laser-Assisted Microdissection for Next-Generation Sequencing*.

Abstract

The Mismatch Repair (MMR) is a repair mechanism for base mismatches occurring during replication. MMR deficiency leads to DNA nucleotide instability, particularly in repeated genome sequences (or microsatellites). The accumulation of numerous mutations leads to cell transformation and the development of cancers known as Microsatellite Instability (MSI). MSI cancers are observed in 10-15% of colorectal cancers (CRC) and are characterised by a hypermutator phenotype and a high infiltration of immune cells. In most cases, MSI cancers are sporadic in origin, developing because of an epigenetic modification that inactivates the MMR system. However, in 10% of cases, MSI cancers are hereditary (germline mutation of the MMR): this is Lynch syndrome.

Colonic MSI tumorigenesis can be described in three phases: (i) loss of MMR system activity observed with pre-cancerous MMR-deficient crypts in the healthy colonic mucosa of patients with Lynch syndrome only (initiation); (ii) accumulation of mutations that induce transformation into a benign tumour or adenoma; (iii) progression to a cancer stage called adenocarcinoma. Despite years of research, the driving mutations of MSI tumorigenesis have not been formally identified. The advent of high-throughput sequencing technologies (exome and/or transcriptome) offers new opportunities to describe the mutational and transcriptional landscape of colonic MSI tumorigenesis, in order to identify the driving mutations (and/or mechanisms).

During my thesis, I studied the involvement of intronic microsatellite instability in the initiation and progression of colonic MSI tumours. We demonstrated the role of non-coding mutations in the alteration of RNA splicing. We have highlighted a specific MSI-driven signature of 985 exons involved in proliferation, differentiation, and cell survival (Jonchère & Montémont et al., under review in the *Genome Biology* journal). In preliminary work, I have also described coding microsatellite instability and identified mutations that are potential drivers of tumour initiation and progression. Finally, I have developed a technical approach for studying early pre-cancerous lesions called ILAM-NGS (for Immunohistochemistry Laser-Assisted Microdissection for Next-Generation Sequencing).

Table des matières

REVUE BIBLIOGRAPHIQUE 1

Chapitre 1 – L'origine du phénotype MSI : l'inactivation du mécanisme de réparation des mésappariements, le système MMR 3

1. Le système MisMatch Repair (MMR)..... 3

1.1. Le mécanisme de réparation des mésappariements..... 4

1.2. Le système MMR et le remodelage de la chromatine 6

1.3. Le rôle du système MMR dans le processus apoptotique 8

2. Les conséquences de l'inactivation du MMR 11

2.1 Les mutations dans les séquences répétées : les microsatellites..... 11

2.1.1. Les microsatellites codants..... 12

2.1.2. Les microsatellites non-codants 12

2.2. Les mutations dans les séquences non-répétées 14

3. L'épissage des ARN pré-messagers 14

3.1. Le mécanisme d'épissage des ARN 15

3.2. La régulation de l'épissage des ARN 17

4. La surveillance et la dégradation des ARNm aberrants : le système NMD 19

5. La surveillance immunitaire et ses mécanismes d'échappement..... 23

5.1. La surveillance immunitaire anti-tumorale 23

5.2. Les mécanismes d'échappement de la surveillance immunitaire 25

5.2.1. L'altération de la présentation antigénique par le HLA 25

5.2.2. L'expression des inhibiteurs des points de contrôles immunitaires 26

5.2.3. Le rôle immunosuppresseur du microenvironnement..... 27

Chapitre 2 – Les caractéristiques cliniques des cancers MSI29

1. La classification moléculaire et les signatures mutationnelles des CRC MSI 29

2. Les cancers colorectaux MSI sporadiques 31

2.1. Caractéristiques génétiques et moléculaires..... 31

2.2. Valeur pronostic des CRC MSI sporadiques 32

3. Les cancers MSI héréditaires : le syndrome de Lynch 33

3.1. Caractéristiques épidémiologiques et prédispositions génétiques..... 33

3.2. Critères d'identification du syndrome de Lynch 35

3.3. Surveillance clinique..... 36

4. Le diagnostic des cancers MSI	36
4.1. Approches standards reconnues	37
4.1.1. L'immunohistochimie : évaluation de l'expression des protéines du MMR.....	37
4.1.2. La PCR : détermination du statut d'instabilité microsatellitaire.....	38
4.1.3. Diagnostic pathologique : sporadique vs syndrome de Lynch	41
4.2. Les approches émergentes pour le diagnostic MSI	42
4.2.1. Les outils bio-informatiques	43
4.2.2. ADN tumoral circulant	44
4.2.3. L'imagerie et l'intelligence artificielle pour diagnostiquer les tumeurs MSI	44
5. Les traitements des cancers colorectaux MSI	46
5.1. Les protocoles thérapeutiques conventionnels.....	47
5.1.1. La thérapie chirurgicale	47
5.1.2. La chimiothérapie	48
5.1.3. L'immunothérapie anti-ICK (ou ICI)	50
5.2. Les approches thérapeutiques innovantes.....	52
 Chapitre 3 – La tumorigenèse MSI colique : de la crypte MMR-déficiente à l'adénocarcinome	54
 1. Description anatomopathologique de la tumorigenèse MSI colique	54
1.1. L'initiation tumorale : les cryptes MMR-déficientes (syndrome de Lynch)	55
1.1.1. Généralités du rôle physiologique des cryptes saines dans l'homéostasie colique	55
1.1.2. Les cryptes MMR-déficientes : mise en évidence et caractéristiques histologiques ..	56
1.2. La progression tumorale MSI : de l'adénome à l'adénocarcinome	60
1.2.1. Les adénomes.....	61
1.2.2. Les adénocarcinomes	64
2. Le paysage mutationnel des cancers colorectaux MSI.....	66
2.1. Les voies de signalisation de l'homéostasie colique et les principaux gènes mutés associés	66
2.1.1. La voie de signalisation Wnt/ β -caténine	66
2.1.2. La voie de signalisation Notch	67
2.1.3. La voie de signalisation BMP/TGF- β	67
2.1.4. La voie de signalisation EGFR/MAPK	67
2.1.5. Les communications entre les voies de signalisation	68
2.2 Les modèles de sélection mutationnelle dans les cancers MSI.....	71
2.2.1. Sélection et classification des événements mutationnels selon la taille de répétition des microsatellites.....	72

2.2.2. Sélection et classification des évènements mutationnels selon les propriétés immunogènes.....	74
--	----

TRAVAUX DE RECHERCHE..... 77

“Massive Cis-Perturbation of Splicing Due To Microsatellite Instability In Colorectal Cancer”	78
---	----

Résultats Préliminaires : Analyse de l’instabilité microsatellitaire codante pour l’identification des mutations impliquées dans l’initiation et la progression tumorale des cancers colorectaux MSI.	171
--	-----

1. Analyse quantitative de l’instabilité microsatellitaire globale selon l’origine héréditaire <i>versus</i> sporadique des lésions.....	171
--	-----

2. Identification des mutations potentiellement <i>drivers</i> de la tumorigenèse MSI (syndrome de Lynch)	175
---	-----

3. Biomarqueurs distinctifs (potentiels) des formes héréditaires et sporadiques, au stade adénocarcinome	185
--	-----

CONCLUSION, DISCUSSION & PERSPECTIVES 188

Conclusion, discussion et perspectives	189
--	-----

ANNEXE 197

“Immunoguided Laser-Assisted Microdissection for Whole Exome and RNA-Sequencing: A Precise Technology to Explore Genomics and Transcriptomics Alteration in Early Stage of Cancer”	198
--	-----

BIBLIOGRAPHIE 201

Liste des Figures et Tables

REVUE BIBLIOGRAPHIQUE

Figure 1 Le mécanisme de réparation du système mismatch repair (MMR).	5
Figure 2 Le remodelage de la chromatine régule le système de réparation MMR.	7
Figure 3 La signalisation apoptotique MMR-dépendante de la réponse aux dommages à l'ADN.	9
Figure 4 La prise en charge des lésions O6-MeG induites par des agents alkylants, dans des contextes de présence ou absence des systèmes de réparation MGMT et MMR.	10
Figure 5 La distribution des microsatellites dans le génome humain.	13
Figure 6 Le mécanisme d'épissage des ARN pré-messagers et ses composants régulateurs.	16
Figure 7 Les évènements d'épissage alternatif les plus fréquemment observés dans les cancers.	19
Figure 8 La voie de signalisation EJC du système NMD.	21
Figure 9 Les différentes situations où un PTC déclenche (ou non) la dégradation des ARNm par le système NMD.	22
Figure 10 L'environnement tumoral immunitaire.	25
Figure 11 Les mécanismes d'évasion immunitaire des tumeurs.	28
Figure 12 Les outils conventionnels du diagnostic des cancers colorectaux MSI.	40
Figure 13 L'algorithme décisionnel pour la distinction entre les cancers colorectaux sporadiques et ceux associés au spectre du syndrome de Lynch.	42
Figure 14 La détermination du statut MSI grâce à l'imagerie et l'intelligence artificielle.	46
Figure 15 La classification TNM des cancers colorectaux.	47
Figure 16 La réponse à l'immunothérapie ICI (pembrolizumab) en fonction du statut MMR.	50
Figure 17 L'organisation des cryptes de la muqueuse colique.	55
Figure 18 Les cryptes MMR-déficientes de la muqueuse colique saine.	57
Figure 19 L'expansion clonale des cellules souches de la niche et la division des cryptes amplifient la tumorigenèse dans un contexte MMR-déficient.	59
Figure 20 La séquence adénome-carcinome.	61

Figure 21 La classification histologique des adénomes selon leur niveau de dysplasie. ____	62
Figure 22 Lésions adénomateuses conventionnelles et sessiles festonnées. _____	64
Figure 23 La classification histologique des adénocarcinomes. _____	65
Figure 24 Les voies de signalisation de l'homéostasie intestinale fréquemment mutées dans la carcinogenèse colorectale. _____	70
Figure 25 Le modèle de sélection des mutations au cours de la tumorigenèse MSI selon la taille de répétition des microsatellites. _____	74
Figure 26 Le modèle de sélection des mutations au cours de la tumorigenèse MSI selon des propriétés immunogènes (et pro-tumorales) des mutations. _____	76
Table 1 Critères d'Amsterdam II et de Bethesda révisés. _____	35

TRAVAUX DE RECHERCHE

Figure 1 La cohorte d'étude des évènements de la cinétique mutationnels de la tumorigenèse MSI. _____	172
Figure 2 L'instabilité microsatellitaire globale dans les différentes lésions des cancers MSI sporadiques (en rouge) versus celle associée au syndrome de Lynch (bleu). _____	173
Figure 3 Distribution du nombre de microsatellites instables codants (bleu) et non-codants (orange) par patient selon le type de lésion. _____	174
Figure 4 Représentation des mutations partagées à travers la tumorigenèse MSI du syndrome de Lynch, et des voies de signalisation associées. _____	176
Figure 5 Identification des mutations potentiellement <i>drivers</i> de la tumorigenèse MSI colique associée au syndrome de Lynch. _____	178
Figure 6 Évaluation de la pression de sélection subie par les mutations potentiellement <i>drivers</i> de la tumorigenèse MSI associée au syndrome de Lynch. Modèle de sélection par la taille de répétition des microsatellites dans les cryptes MMR-déficientes (A) et dans les adénocarcinomes (B). _____	181
Figure 7 Représentation en réseau des gènes mutés et potentiellement <i>drivers</i> de la tumorigenèse MSI associée au syndrome de Lynch. _____	183
Figure 8 Détermination <i>in silico</i> de la conséquence des mutations sur la fonction des gènes. _____	184

Figure 9 Représentation des mutations partagées et spécifiques des cancers sporadiques <i>versus</i> du syndrome de Lynch, et des voies de signalisation associées, au stade adénocarcinome.	186
Figure 10 Détermination des mutations partagées et spécifiques dans les adénocarcinomes sporadiques et associées au syndrome de Lynch.	187
Table Liste des gènes identifiés comme potentiellement drivers de la tumorigenèse MSI associé au syndrome de Lynch.	179

Liste des abréviations

ADN : acide désoxyribonucléique

ARN : acide ribonucléique

ADP/ATP : adenosine di/triphosphate

ARNm : acide ribonucléique messenger

BMP : *Bone Morphogenic Protein*

BP : point de branchement

CRC : *colorectal cancer*

CIMP : *CpG island methylator phenotype*

CIN : *chromosomal instability*

CMMRD : *constitutional mismatch repair deficiency*

cMS : *coding microsatellite*

CMS : *consensus molecular phenotype*

ConA : *conventional adenoma*

CTL : *cytotoxic T lymphocyte*

CTLA-4 : *cytotoxic T lymphocyte antigen-4*

ECM : *extra-cellular matrix*

EJC : *exon junction complex*

ES : *exon skipping*

FOLFOX : 5-FU, leucovirin, irinotécan

FSP : *frameshift peptide*

HLA : *human leucocyte antigen*

HE : hématoxyline et éosine

indels : *insertion & deletion*

ICK : *Immune checkpoint*

ICI : *immune checkpoint inhibitor*

INCa : Institut National du Cancer

IHC : *immunohistochemistry*

LS : *lynch syndrome*

MF : *mutational frequency*

MHC : *major histocompatibility complex*

MMR : *mismatch repair*

MSI : *microsatellite instability*

MSS : *microsatellite stability*

ncMS : *non-coding microsatellite*

NGS : *next-generation sequencing*

NMD : *non-mediated decay*

pb : paire de bases

PCR : *polymerase chain reaction*

PD-1 : *programmed death-1 receptor*

PTC : *premature termination codon*

pval : p-value

PY : *polypyrimidine tract*

SNV : *single nucleotide variation*

snRNP : *small nuclear ribonucleoprotein*

SP : *sporadic*

SSA : *sessile serrated adenoma*

U2AF1/2 : *U2 small nuclear RNA auxiliary factor 1/2*

TGF- β : *Transforming Growth Factor β*

UTR : *untranslated region*

UMI : *unique molecular identifier*

VAF : *variant allele frequency*

AVANT-PROPOS

J'ai réalisé ma thèse, sous la direction du Dr. Ada COLLURA, dans le laboratoire « Instabilité des Microsatellites (MSI) et Cancer » dirigé par le Pr. Alex DUVAL, au Centre de Recherche Saint-Antoine (CRSA). Mon équipe d'accueil développe ses recherches sur la physiopathologie tumorale MSI afin de mieux comprendre la dynamique mutationnelle de la tumorigenèse MSI, de développer de nouveaux outils (diagnostic et/ou pronostic) et de proposer des cibles thérapeutiques afin d'améliorer la prise en charge des patients atteints d'un cancer de type MSI. L'équipe bénéficie au niveau de son site d'implantation à l'Hôpital Saint-Antoine, d'une interface clinique exceptionnelle pour développer ses recherches. Notre équipe fait partie du consortium Siric Curamus, l'un des huit sites de recherche intégrés sur le cancer (SIRIC CANCER) labellisés par l'INCA, le Ministère de la Solidarité et de la Santé, l'INSERM et l'ITMO Cancer Aviesan (Alliance nationale pour les sciences de la vie et de la santé). Aussi, la réalisation de ces travaux de recherche est rendue possible grâce à la collaboration de chercheurs en biologie, de bio-informaticiens et de cliniciens (oncologues, chirurgiens digestifs, pathologistes) au sein de notre équipe.

Pendant mes quatre ans de thèse, je me suis intéressé à l'identification de certaines mutations microsatellitaires pour comprendre leur pertinence physiopathologique dans l'initiation et la progression tumorale MSI sporadique et/ou associée au syndrome de Lynch (forme héréditaire de prédisposition aux cancers MSI).

La revue bibliographique que je présente dans ce manuscrit est composée de quatre chapitres. Dans un premier chapitre, je décrirai le système de réparation des mésappariements de base, le Mismatch Repair (MMR) dont l'inactivation, à l'origine de l'instabilité nucléotidique et plus particulièrement microsatellitaire, entraîne l'accumulation des mutations qui induisent la transformation cellulaire. Également, j'aborderai brièvement les mécanismes de régulation de l'expression des ARN (épissage des ARN et contrôle des ARN mutés *via* le système NMD) ainsi que la surveillance immunitaire anti-tumorale. Le chapitre 2 sera dédié aux caractéristiques cliniques et génétiques des cancers MSI ainsi que des outils actuellement utilisés pour diagnostiquer et traiter ces cancers. Le chapitre 3 portera sur la description

anatomopathologique et génomique de la tumorigenèse MSI colique, de la crypte MMR-déficiente à l'adénocarcinome.

Les résultats de ma thèse sont présentés sous forme d'un article, en révision dans le journal *Genome Biology* (Article 1). Également, des analyses complémentaires seront présentées sous la forme de résultats préliminaires. En annexe, je présente un résumé du protocole technique que j'ai développé pour l'étude des cryptes MMR-déficientes.

Une discussion générale sur mes travaux et leurs perspectives qui en découlent conclura ce manuscrit.

REVUE BIBLIOGRAPHIQUE

Chapitre 1 – L'origine du phénotype MSI : l'inactivation du mécanisme de réparation des mésappariements, le système MMR

L'instabilité génomique, qu'elle soit chromosomique ou nucléotidique, est une caractéristique commune à tous les cancers sporadiques comme héréditaires (Hanahan, 2022). Dans la majorité des cancers, l'instabilité est chromosomique (CIN, *Chromosomal Instability*) : elle se traduit par des réarrangements structuraux des chromosomes ainsi que des gains ou des pertes de grandes régions d'ADN (Weeden et al., 2023). La CIN est responsable du développement des cancers dits non-MSI (MSS, *MicroSatellite Stable*) (Lengauer, W.Vogelstein, 1998). L'instabilité nucléotidique est, quant à elle, caractérisée par des mutations ponctuelles, qui résultent d'erreurs répliquatives de l'ADN polymérase – malgré son activité de relecture (endonucléase 3'-5'), elle n'est pas complètement fidèle (Ganai and Johansson, 2016). Dans un contexte physiologiquement sain, ces erreurs sont prises en charge par les systèmes de réparation de l'ADN comme le MisMatch Repair (MMR) qui répare les mésappariements. Les cancers MSI sont caractérisés par une instabilité nucléotidique consécutive à l'inactivation du système MMR qui induit l'accumulation de nombreuses mutations dans l'ADN : on parle de phénotype hypermutateur (Ionov et al., 1993; Thibodeau et al., 1993). Les mutations se situent principalement dans les microsatellites – ce sont des séquences répétées en tandem d'un motif de 1 à 5 nucléotides, les cibles privilégiées (*hotspots*) des erreurs répliquatives en raison de la nature répétée de ces séquences (Levinson and Gutman, 1987; Tautz, 1989).

1. Le système MisMatch Repair (MMR)

Le mécanisme de réparation des mésappariements de bases (MMR) intervient durant la réplication et il est nécessaire pour préserver l'intégrité du génome. En effet, son activité réduit le taux d'erreur répliquative de 10^{-6} - 10^{-7} (fidélité de l'ADN polymérase) à 10^{-9} - 10^{-10} (Ganai and Johansson, 2016).

Dans cette partie, je décrirai le fonctionnement de ce système de réparation et de sa régulation ainsi que, plus brièvement, son rôle dans la signalisation apoptotique de la réponse aux dommages à l'ADN (DDR, *DNA Damage Response*).

1.1. Le mécanisme de réparation des mésappariements

Le système de réparation MMR est très conservé au cours de l'évolution (Groothuizen and Sixma, 2016). Chez les eucaryotes supérieurs, il est constitué de deux complexes protéiques, les hétérodimères MutS α (MSH2, MSH6) et sa variante MutS β (MSH2, MSH3) et MutL α (MLH1, PMS2) et sa variante MutL β (MLH1, PMS1) (Huang and Li, 2020). Le complexe MutS α reconnaît les mésappariements de base et les petites boucles d'insertion/délétion (IDL, *Insertion Deletion Loop*) alors que le complexe MutS β prend en charge exclusivement les grandes IDL (> 2 pb) (Reyes et al., 2015). Les protéines MSH6 et MSH3 entrent en compétition pour se fixer à la protéine MSH2 ; étant donné la meilleure affinité de MSH2 pour MSH6, le complexe MutS α est 10 fois plus représenté que le MutS β dans la cellule (Marra et al., 1998). De la même manière, MutL α est le principal complexe formé (90%) – le rôle biologique de MutL β n'est pas encore élucidé (Kadyrov et al., 2006).

Le processus de réparation du système MMR se déroule en plusieurs phases (Figure 1). L'hétérodimère MutS reconnaît et se fixe au mésappariement/IDL entraînant un changement conformationnel qui libère une molécule d'ADP (*Adenosine DiPhosphate*) et introduit une molécule d'ATP (*Adenosine TriPhosphate*). La forme active du complexe (MutS-ATP) en conformation « pince coulissante » (*sliding clamp*), glisse le long de l'ADN à la recherche d'un signal d'identification du brin néosynthétisé et permet aussi de recruter MutL α (Kadyrov et al., 2006). Chez les eucaryotes, c'est le chargement du complexe mono-trimérique PCNA (*Proliferating Cell Nuclear protein*) sur l'ADN par le complexe RFC (*Replication Factor-C*) dans une orientation spécifique, qui permet l'identification du brin néosynthétisé (Pluciennik and Modrich, 2007; Reyes et al., 2015). Également, le PCNA active le complexe MutL *via* une interaction forte avec la protéine MLH1, induisant l'excision du mésappariement par l'activité endonucléase (sens 3'-5') de PMS2 (Genschel et al., 2017). L'exonucléase EXO1 peut assister PMS2 dans cette excision *via* son activité exonucléase dans le sens 5'-3' ; son rôle est indispensable dans la signalisation de la réponse aux dommages de l'ADN à travers l'activité du MMR (Dzantiev et al., 2004; Kadyrov et al., 2006). Le complexe RPA se fixe sur l'ADN nouvellement simple brin afin de le maintenir sous cette forme monocaténaire. Enfin, l'ADN polymérase re-synthétise le brin hydrolysé et l'ADN ligase I finalise le processus par la ligation du fragment réparé (Guo et al., 2006).

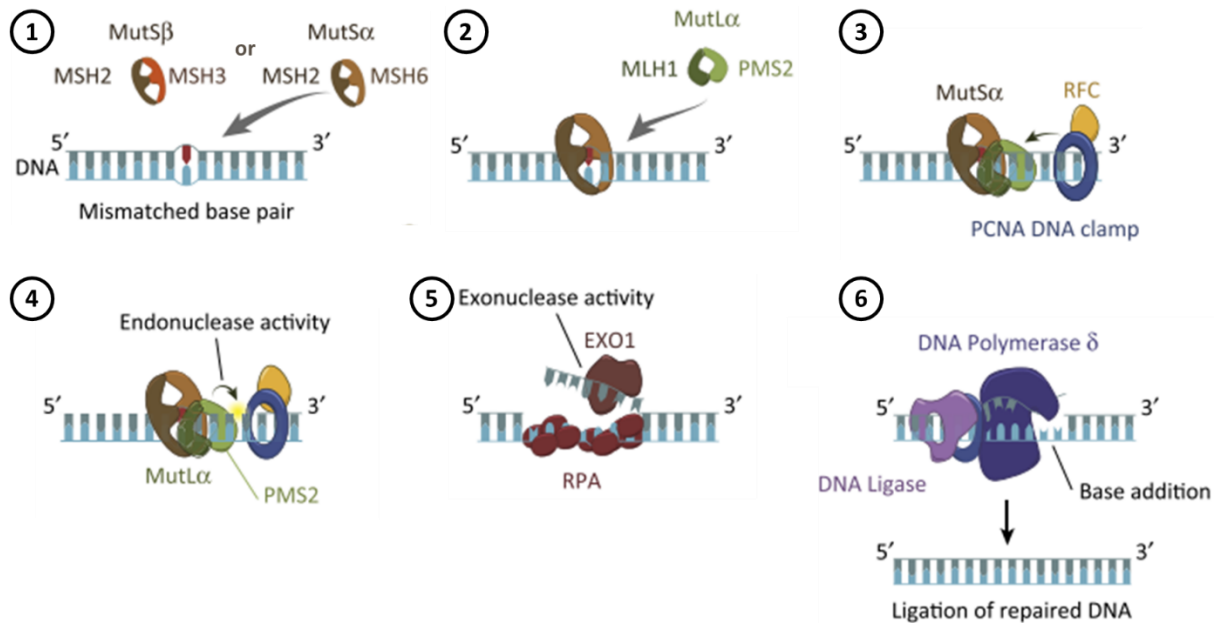


Figure 1 | Le mécanisme de réparation du système mismatch repair (MMR).

(1) Reconnaissance du mésappariement MutS. **(2)** Recrutement de MutL. **(3)** Chargement du PCNA sur l'ADN *via* RFC. **(4)** Activation de l'activité endonucléase de PMS2. **(5)** Excision du mésappariement par EXO1. **(6)** Resynthèse de l'ADN et ligation (d'après Li and Martin, 2016).

La reconnaissance du mésappariement par MutS serait associée à la flexibilité intrinsèque de l'ADN (structure tridimensionnelle) consécutive à la lésion, ainsi qu'aux séquences nucléotidiques qui la flanquent (Mazurek et al., 2002). De ce fait, l'efficacité de la réparation est dépendante de la propension du mésappariement à engendrer une distorsion (G-T : faible distorsion, bien réparée ; C-C : forte distorsion, moins bien réparée) (Hsieh and Yamane, 2008). Également, le changement de conformation de MutS en *sliding clamp* induit une distorsion supplémentaire de l'ADN qui, en cas de mésappariement C-C, pourrait perturber l'avancée de l'ADN polymérase et donc de la re-synthèse du brin (LeBlanc et al., 2018). Enfin, certains chercheurs suggèrent que les erreurs répliquatives sont mieux réparées sur le brin retardé par rapport au brin continu. En effet, les fragments d'Okazaki présentent de nombreuses extrémités 5' ce qui est favorable pour l'activité nucléase d'EXO1 et plus généralement qui stimulent le processus de réparation *via* le système MMR (Youri I. Pavlov, et al., 2003). Toutefois, les erreurs observées entre les deux brins en fin de répllication semblent similaires : l'efficacité de réparation du système MMR, qui semble supérieure sur le brin retardé, est compensée par la meilleure fidélité de la polymérase δ (brin continu) vis-à-vis de la polymérase ε (brin retardé) (Ganai and Johansson, 2016).

1.2. Le système MMR et le remodelage de la chromatine

Les nucléosomes sont des unités élémentaires de la chromatine nécessaires à la compaction de l'ADN. Ils sont chacun constitués d'un segment d'ADN de 146 paires de bases enroulés autour d'un octamère d'histones (un tétramère H3-H4 et deux dimères H2A-H2B), et ils sont reliés entre eux par une région internucléosomique (Luger et al., 1997). Les modifications post-traductionnelles des extrémités amino-terminales (*N-terminal tails*) des histones – acétylation, méthylation, phosphorylation, ubiquitination – sont impliquées dans de nombreuses voies de signalisation de plusieurs processus cellulaires (transcription, réparation de l'ADN, apoptose, cycle cellulaire) (Zhang et al., 2023).

Lorsque la chromatine est compactée, les mécanismes de réparation de l'ADN ne peuvent pas intervenir ; cela peut s'expliquer par l'inaccessibilité de l'ADN. Il a été démontré que les nucléosomes inhibent l'activité catalytique de MutS (ATPase) empêchant son changement conformationnel en *sliding clamp* nécessaire au processus de réparation du MMR (Li et al., 2009). Pendant la phase S, les nucléosomes sont désassemblés au passage de la fourche de réplication afin de permettre la réplication de l'ADN, puis réassemblés (Ransom et al., 2010). La triméthylation de la lysine 36 de l'histone H3 (H3K36me3), une marque épigénétique de la chromatine active, recrute le complexe MutS du système MMR *via* le domaine de fixation PWWP (*Proline-Tryptophane-Tryptophane-Proline*) de la protéine MSH6 (Figure 2A) (Li et al., 2013). Après l'identification d'un mésappariement, le complexe MutS interagit avec le complexe protéique CAF1 (*Chromatin Assembly Factor-1*) et inhibe son activité de réassemblage des nucléosomes. Une autre protéine de remodelage de la chromatine, l'hélicase SMARCAD1 (*SWI/SNF-related, matrix-associated actin-dependent regulator of chromatin, subfamily a, containing DEAD/H box 1*), est également recrutée par MSH2 pour empêcher la reformation des nucléosomes et assurer l'accessibilité de l'ADN au système de réparation pour assurer son activité (Figure 2B) (Huang and Li, 2020; Terui et al., 2018).

Des travaux récents ont montré que la présence de la marque H3K36me3 est plus enrichie dans les régions transcrites que dans les régions introniques ou non-transcrites. L'absence de marque H3K36me3 ou une perturbation de l'interaction H3K36me3 avec le complexe MutS induit une augmentation importante des mutations dans les régions transcrites sans évolution significative dans les régions non-transcrites. Il a été suggéré que

l'action de réparation du système MMR soit couplée au processus de transcription, *via* la marque épigénétique H3K36me3 (Frigola et al., 2017; Huang and Li, 2018).

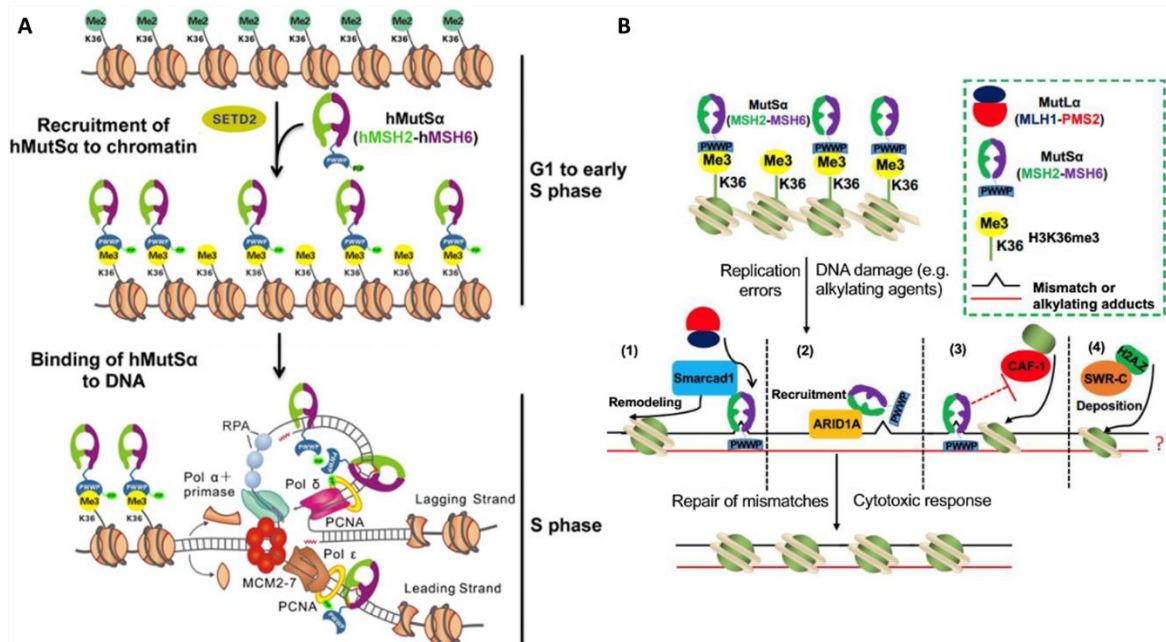


Figure 2 | Le remodelage de la chromatine régule le système de réparation MMR.

(A) Le MMR est recruté *via* l'interaction du complexe MutSα et de la marque épigénétique H3K36me3, avant la réplication. **(B)** Les protéines de remodelage de la chromatine régulent l'accès des protéines du système MMR à la chromatine, en particulier MutSα, en réponse aux erreurs de réplication ou aux dommages causés par l'alkylation (Huang et al., 2020).

Les protéines de remodelage de la chromatine (CAF1, SMARCD1) et les protéines qui modifient l'histone H3 – déméthylase : famille des KDM (*Histone Lysine Demethylase*), méthyltransférase : SETD2/3 (*SET Domain Containing 2/3*), ASH1L (*Absent Small And Homeotic Disks Protein 1 Homolog*) – sont nécessaires dans le processus de réparation des mésappariements par le système MMR. Des mutations dans les gènes qui codent pour ces protéines peuvent entraîner une diminution de l'activité voire une déficience de l'activité du système MMR et donc une augmentation des mutations en empêchant le recrutement des complexes sur l'ADN (Dalgliesh et al., 2010; Gerlinger M, Rowan AJ, 2012; Huang and Li, 2020; Li et al., 2013).

1.3. Le rôle du système MMR dans le processus apoptotique

Le système MMR intervient également dans la signalisation de réponse aux dommages à l'ADN suite à l'exposition à des agents génotoxiques tels que les agents alkylants, les espèces réactives de l'oxygène ou les radiations (UV, γ) (Gupta and Heinen, 2019; Ijsselsteijn et al., 2020). Les modifications des bases peuvent provoquer des lésions persistantes dans l'ADN si elles ne sont pas prises en charge par des mécanismes adaptés. C'est le cas des méthylations du résidu O6 de la guanine (O6-MeG, *O6-methyl-guanine*) induites par des agents alkylants (par exemple, le Témazolomide ou le Cisplatine, des agents utilisés dans divers traitements anti-cancéreux), normalement dé-méthylé par la MGMT (*O6-MethylGuanine-MethylTransferase*) (Xu-Welliver and Pegg, 2002). En absence de cette enzyme, les lésions O6-MeG entraînent une signalisation apoptotique MMR-dépendante de réponse aux dommages à l'ADN, impliquant des protéines du système MMR et de la voie ATM et/ou ATR (*Ataxia Telangiectasia Mutated* et *Ataxia Telangiectasia and Rad3-related*) (Fu et al., 2012).

Les lésions O6-MeG dans un contexte MGMT-déficient sont particulièrement intéressantes dans notre étude des événements initiateurs de la tumorigenèse MSI puisqu'elles peuvent apporter un avantage sélectif (survie des cellules) et donc favoriser la perte du MMR. Ce phénomène est brièvement décrit dans les deux encadrés suivants (« La signalisation apoptotique MMR-dépendante » et « MMR & dommages alkylants ») (Fu et al., 2012; Li, 2008; Wang and Qin, 2003).

La signalisation apoptotique MMR-dépendante

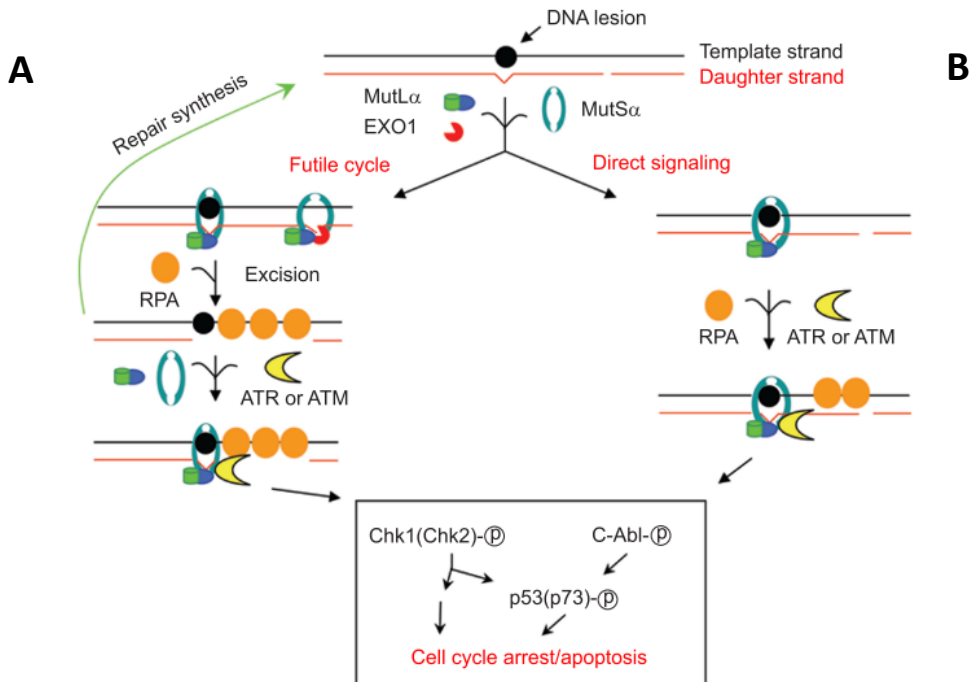


Figure 3 | La signalisation apoptotique MMR-dépendante de la réponse aux dommages à l'ADN.

Deux modèles sont proposés par les chercheurs : le cycle futile **(A)** et la signalisation directe **(B)**.

Dans les deux modèles, les complexes MutL et MutS interviennent en premier pour initier la réparation d'une lésion de l'ADN sur le brin matrice qui provoque un mésappariement sur le brin néosynthétisé (Schaetzlein et al., 2007). L'activité d'EXO1 ne peut pas être compensée par PMS2 (Klapacz et al., 2009). Le recrutement d'ATR/ATM induit la phosphorylation des kinases du checkpoint Chk1 et Chk2, ainsi que p53, ce qui provoque l'arrêt du cycle cellulaire et l'apoptose des cellules (Brown et al., 2002).

(A) Dans le modèle de cycle futile, le mésappariement persiste malgré la réparation du MMR, puisque la lésion initiale est située sur le brin matrice alors que le MMR est spécifique du néo-brin. Se succèdent alors des cycles de réparation infructueux qui provoquent des cassures simples puis double brins ; ces dommages à l'ADN induisent l'arrêt du cycle cellulaire en phase G2 et ensuite l'apoptose (Gutpa et al., 2019).

(B) Dans le modèle de signalisation directe, le MMR envoie un signal d'arrêt du cycle cellulaire et d'apoptose, *via* le recrutement d'ATR/ATM par MutS/MutL, sans initier de cycles de réparation (Yoshioka et al., 2006).

Dans les deux modèles, c'est la prise en charge par le MMR qui est cytotoxique pour les cellules, non la lésion initiale.

MMR & dommages alkylants

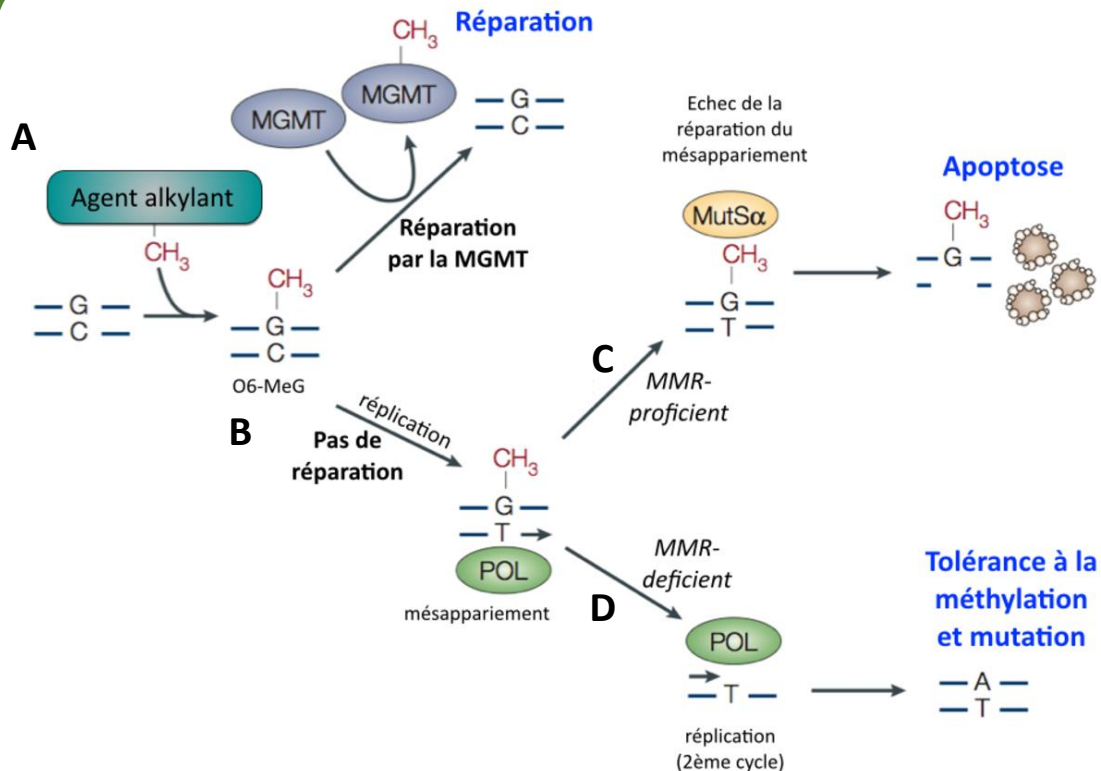


Figure 4 | La prise en charge des lésions O6-MeG induites par des agents alkylants, dans des contextes de présence ou absence des systèmes de réparation MGMT et MMR.

(Gutpa et al., 2019 ; Thèse Bodo)

(A) L'enzyme MGMT est une déméthylase spécifique de la lésion O6-MeG, elle transfère le groupe méthyl de la guanine modifiée sur un de ses résidus cystéine (Xu-Welliver et al., 2002).

(B) Lorsque la MGMT est déficiente, suite à une mutation perte de fonction ou hyperméthylation du promoteur du gène qui code pour la protéine, la modification persiste. Au cycle de réplication suivant, la base modifiée O6-MeG n'est pas correctement identifiée par l'ADN polymérase qui introduit un mésappariement GT (Kaina et al., 1990).

(C) Dans les cellules MMR-proficientes, le mésappariement induit des cycles futiles de réparation et donc la signalisation apoptotique MMR-dépendante.

(D) Dans les cellules MMR-déficientes, le mésappariement subsiste et la cellule survie. Le deuxième cycle de réplication conduit à une mutation : transition G-C vers A-T ; les cellules MMR-déficientes sont dites tolérantes à la méthylation.

Ainsi la perte du système MMR peut apporter un avantage sélectif aux cellules MGMT-déficientes, en contrepartie de l'accumulation de mutations.

2. Les conséquences de l'inactivation du MMR

Une mutation perte de fonction bi-allélique d'un des quatre gènes principaux du MMR (*MSH2*, *MLH1*, *MSH6* ou *PMS2*) entraîne l'inactivation du mécanisme de réparation. La déficience du MMR n'est pas en soit l'évènement transformateur des cellules saines en cellules cancéreuses, mais elle est responsable d'un processus général d'instabilité génomique qui favorise le développement tumoral (Duval and Hamelin, 2003). Dans ce contexte MMR-déficient, les mutations s'accumulent dans tout le génome, que ce soit dans les régions intergéniques (98% du génome) comme dans les régions intragéniques (Cortes-Ciriano et al., 2017). Cependant, même si certaines séquences intergéniques (promoteurs et *enhancers* transcriptionnels) sont impliquées dans la régulation de l'expression des gènes, le rôle fonctionnel physiologique de la plupart de ces séquences n'est que partiellement élucidé ; de ce fait l'implication des mutations intergéniques dans la tumorigenèse MSI est peu connue (Mathieu et al., 2014; Palazzo and Gregory, 2014). La fonction des séquences intragéniques est quant à elle mieux comprise, et les altérations génétiques associées – lorsqu'il s'agit de mutations dans des régions codantes – peuvent être corrélées au phénotype des tumeurs MSI (Bettington et al., 2013; Boland and Shike, 2010). Dans un contexte MMR-déficient, les séquences répétées sont des cibles mutationnelles privilégiées : elles représentent 59,8% des mutations observées dans les cancers MSI contre 40,2% pour les séquences non-répétées (SNV, *Single Nucleotide Variant*) (Jonchere et al., 2018).

2.1 Les mutations dans les séquences répétées : les microsatellites

L'origine des microsatellites est très débattue par les spécialistes, cependant ils s'accordent pour dire que les néo-microsatellites proviennent de mutations successives (substitutions ou *indels*) ou de microsatellites importés par des éléments transposables (SINE ou LINE, *Short/Long Interspersed Elements*) (Walter Messier, Shou-Hsien Li, 1996; Zhu et al., 2000; Wilder and Hollocher, 2001). À la suite d'évènements de sélection au cours de l'évolution, les microsatellites codants ont été contraints à de petites et moyennes tailles de répétition (5 à 11 pb) afin de diminuer la probabilité de mutations. En effet, en raison de la nature répétée de ces séquences, des erreurs peuvent être introduites à cause du glissement de la polymérase sur l'ADN (*slippage*) lors de la réplication, ce qui provoque des délétions ou

des insertions de base dans respectivement 80% et 20% des cas (Levinson and Gutman, 1987; Tautz, 1989). Plus la taille de répétition d'un microsatellite est grande, plus sa propension à être mutée est grande (Jonchere et al., 2018; Woerner et al., 2010).

Les microsatellites non-codants – qui n'ont pas un rôle fonctionnel direct – peuvent être de petites, moyennes ou grandes tailles (5 à 100 pb) sans, toutefois, que leur mutation ait une conséquence néfaste avérée (Ellegren, 2004).

2.1.1. Les microsatellites codants

Les microsatellites codants (cMS, *coding MicroSatellite*) sont présents dans environ 20% des gènes humains, soit plus de 4 000 gènes (El-Bchiri et al., 2008). L'instabilité des cMS induit fréquemment un décalage du cadre de lecture (mutation *frameshift*) qui peut entraîner l'apparition d'un codon stop prématuré (PTC, *Premature Termination Codon*) – après transcription – dans l'ARN. Ces ARN aberrants sont généralement pris en charge et dégradés par le système de surveillance des ARN (NMD, *Nonsense-Mediated Decay*) (voir chapitre 1, partie 4), ce qui a pour conséquence une diminution de l'expression de la protéine associée (Woerner et al., 2005). Parfois, le PTC n'est pas identifié par le NMD ce qui permet la synthèse d'une protéine tronquée (fonctionnelle ou non, selon la position du PTC dans le gène) ainsi qu'un peptide dont la séquence en acides aminés est aberrante suite au décalage du cadre de lecture (FSP, *FrameShift Peptide*) et potentiellement immunogène pour la cellule (Figure 5) (Kloor and Von Knebel Doeberitz, 2016; Schwitalle et al., 2008).

2.1.2. Les microsatellites non-codants

Les microsatellites non-codants (ncMS, *non-coding MicroSatellite*) représentent le plus grand nombre de microsatellites dans le génome, que ce soit dans les régions intragéniques – *i.e.* introniques et exoniques non-codantes (UTR, *UnTranslated Region*) – et intergéniques (promoteurs et *enhancers* transcriptionnels) (Cortes-Ciriano et al., 2017; Mathieu et al., 2014; Palazzo and Gregory, 2014). Ils sont très mutés dans un contexte MMR-déficient, étant donné leur grande taille de répétition ; cependant, les conséquences de ces mutations sur l'expression ou l'activité de la protéine restent à élucider (Figure 5).

Également, les régions non-traduites (UTR, *UnTranslated Region*), situées en amont (5'UTR) ou en aval (3'UTR) de la séquence codante des ARN messagers (ARNm), permettent

l'initiation (séquence consensus Kozak) et la terminaison de la traduction (Barrett et al., 2012; Kozak, 1989). La coiffe (5'UTR) et la queue poly A (3'UTR) protègent les ARNm de la dégradation par les nucléases. Des mutations dans les microsatellites des UTR pourraient influencer le niveau d'expression protéique, par inhibition ou ralentissement de la traduction, ou dégradation des ARNm *via* de petits ARN régulateurs (miRNA, *microRNA*) (Morris and Geballe, 2000; Rana, 2007; Sawaya et al., 2013). Des travaux ont démontré une corrélation entre l'instabilité des microsatellites des régions UTR et des variations d'expression, grâce à des analyses génomiques et transcriptomiques couplées (Kwiyeom Yoon et al., 2013; Suraweera et al., 2001).

Enfin, les régions intergéniques, telles que les promoteurs et *enhancers*, présentent de nombreux microsatellites fréquemment mutés dans un contexte MMR-déficient. C'est le cas du promoteur du gène *MMP3* (*Matrix MetalloProteinase-3*) qui code pour une protéine impliquée dans l'organisation de la matrice extracellulaire ; l'instabilité du microsatellite situé dans la séquence promotrice du gène *MMP3* est corrélée avec la perte d'expression de la protéine dans les tumeurs MSI coliques (Morán et al., 2002).

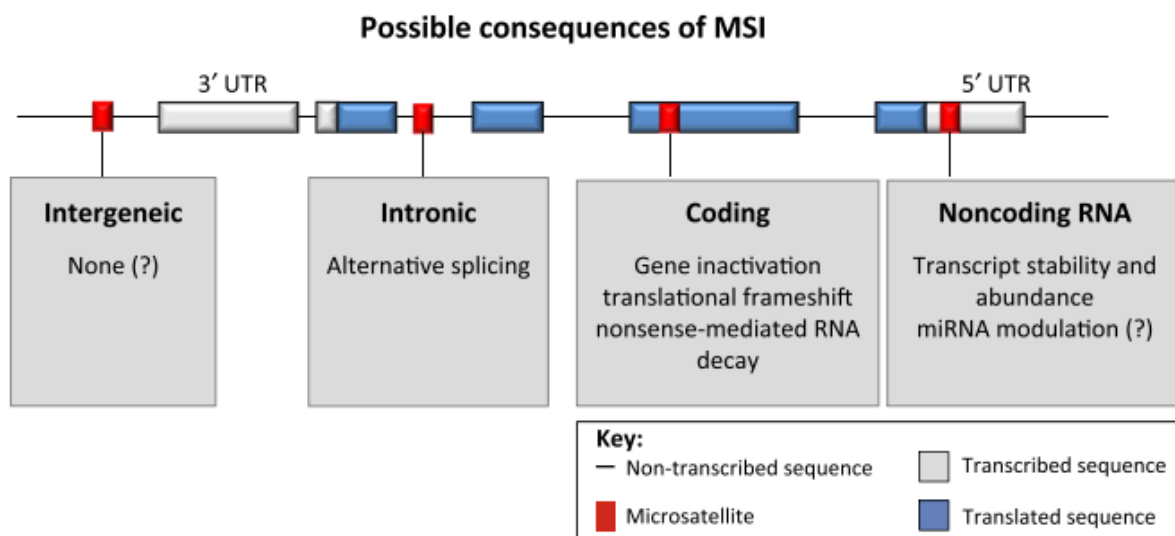


Figure 5 | La distribution des microsatellites dans le génome humain.

Les microsatellites des régions non-codantes intergéniques (promoteurs et *enhancers*) et intragéniques (introns et UTR), ainsi que des régions codantes du génome. Les conséquences fonctionnelles de l'instabilité microsatellitaire varient selon la localisation de ces séquences répétées dans les régions citées précédemment (Kloor and Von Knebel Doeberitz, 2016).

2.2. Les mutations dans les séquences non-répétées

Les SNV sont dues à des mutations ponctuelles de type substitution, insertion ou délétion d'un seul nucléotide dans des séquences non-répétées du génome. Elles peuvent induire une modification non-synonyme ou synonyme, c'est-à-dire provoquer respectivement un changement ou non d'un acide aminé dans la séquence protéique (Chu and Wei, 2019). L'inactivation du système MMR induit l'accumulation de centaines de mutations SNV dans le génome, parfois dans des régions codantes, qui peuvent également induire un décalage du cadre de lecture avec l'apparition d'un PTC. En revanche, la capacité des SNV à générer des peptides néo-antigéniques est bien moindre (Turajlic et al., 2017).

Enfin, les variations de copie de gènes (CNV, *Copy Number Variation*) sont des variations structurelles (SV, *Structural Variation*) dues à des duplications ou des délétions de grandes régions chromosomiques, somatiques ou germinales. Ces mutations sont très fréquentes dans les cancers MSS contrairement aux cancers MSI, où peu ou pas d'évènements de ce type sont identifiés dans les cellules tumorales (Le Borgne et al., 2013; Muleris et al., 2008).

3. L'épissage des ARN pré-messagers

Chez les eucaryotes, les gènes contiennent en moyenne neuf exons (séquences codantes) et huit introns (séquences non-codantes), structurés en alternance dans les ARN transcrits (Deutsch and Long, 1999). Parmi les 20 000 gènes dans le génome humain, 90% sont épissés de manière alternative ce qui participe à la diversité de la synthèse protéique (plus de 100 000 protéines différentes) (Sultan et al., 2008). L'épissage alternatif des gènes contribue à la régulation de plusieurs processus cellulaires comme la prolifération, la différenciation des cellules, la transition épithélio-mésenchymateuse et l'activation des lymphocytes T (Baralle and Giudice, 2017; Blencowe, 2006). L'altération de l'épissage est associée au développement de nombreuses maladies (cancers, maladies génétiques héréditaires) (Abramowicz and Gos, 2019; Gehring and Roignant, 2021). Le processus d'épissage permet l'excision des introns, une étape essentielle pour l'expression des ARN-messagers et hautement régulée par des facteurs protéiques *trans*-régulateurs qui interagissent avec des séquences *cis*-régulatrices dans le transcrit (Bradley and Anczuków, 2023). Également, l'épissage et la transcription des ARN sont

couplés dans le temps et l'espace ce qui entraîne une co-régulation réciproque de ces deux mécanismes (Wang et al., 2008).

De nombreux ncMS présents dans les séquences introniques régulatrices de l'épissage des ARN (*polypyrimidine tract*) sont fréquemment mutés dans les cancers MSI. L'instabilité de ces microsatellites peut induire un épissage aberrant du gène associé, comme cela a été rapporté pour les gènes *MRE11* et *HSP110* dans les cancers colorectaux MSI (Dorard et al., 2011; Giannini et al., 2004). Mes travaux de thèse ont permis de décrire ses altérations de l'épissage des ARN associées à l'instabilité des ncMS et leurs conséquences fonctionnelles, au cours de la tumorigenèse MSI colique (Article 1). Dans la partie suivante, je décris succinctement les mécanismes d'épissage et ses éléments régulateurs qui peuvent être altérés dans les cancers.

3.1. Le mécanisme d'épissage des ARN pré-messagers

Il existe deux complexes d'épissage, appelé splicéosome : le U2-splicéosome est spécifique des sites majoritaires d'épissage (GT-AG) et compte plus de 300 protéines et petites ribonucléoprotéines snRNP (*Small Nuclear Ribonucleoprotein*) ; le U12-splicéosome reconnaît les sites AT-AC (et GT-AG) enrichis dans les gènes de la réplication, la réparation de l'ADN, la transcription, l'épissage et la traduction (Akinyi and Frilander, 2021). Le splicéosome fait intervenir séquentiellement différents complexes pour : 1) identifier et se fixer aux sites d'épissage (complexe A) ; 2) la première *trans*-estérification clive en 5' la jonction exon-intron et forme un lasso (complexe B) et 3) la deuxième *trans*-estérification permet l'élimination du lasso et la jonction des deux exons (complexe C) (Figure 6A) (Hang et al., 2015; Urbanski et al., 2018). Plusieurs séquences nucléotides sont essentielles dans le déroulé de l'épissage. Les sites consensus donneur et accepteur d'épissage sont respectivement situés aux jonctions 5' exon-intron et 3' intron-exon ; les deux sites sont composés d'une séquence dinucléotidique GT (5' donneur ou 5'SS, *5'Splice Site*) et AG (3' accepteur ou 3'SS, *3'Splice Site*). Le point de branchement (BPS, *Branch Point Site*) et de l'ensemble polypyrimidique (PY, *Polypyrimidine Tract*) sont des séquences nucléotidiques dégénérées, c'est-à-dire très diverses (Lee and Rio, 2015). La ribonucléoprotéine snRNP-U1 et le complexe protéique U2AF1/2 reconnaissent respectivement le 5'SS et le PY-3'SS (Figure 6B) (Bradley and Anczuków, 2023).

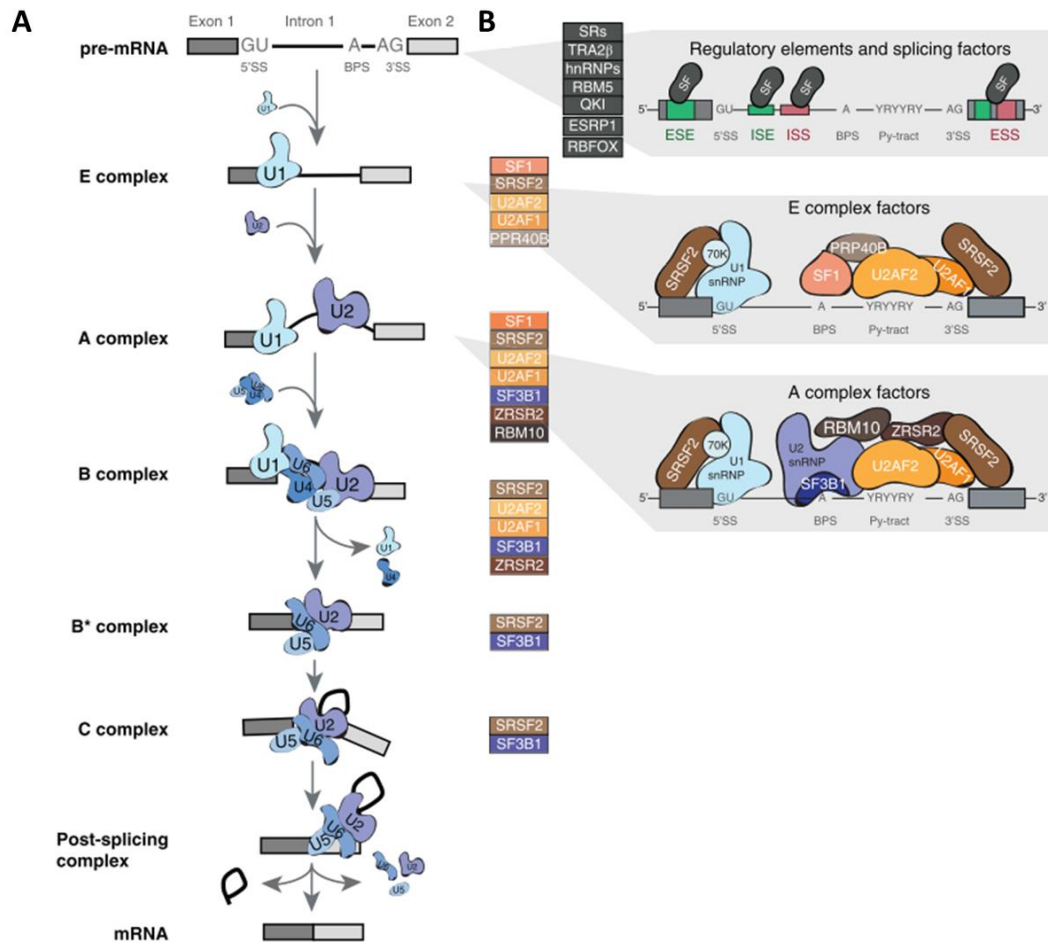


Figure 6 | Le mécanisme d'épissage des ARN pré-messagers et ses composants régulateurs.

(A) Représentation graphique de l'assemblage de la machinerie d'épissage (spliceosome) sur une molécule de pré-ARNm et de la réaction d'épissage pour générer un ARNm mature épissé. **(Complexe E)** L'épissage est initié par la formation du complexe E *via* la liaison de snRNP U1 au site d'épissage 5' (5'SS) de l'intron. SF1 et U2AF2 (U2) se lient respectivement au site du point de branchement (BPS) et au PY. **(Complexe A)** L'interaction de snRNP U2 avec le BPS stabilisée par U2 (SF1, U2AF2, U2AF1, SF3A/B) conduit à la formation du complexe A et entraîne le retrait de SF1 du BPS. **(Complexe B)** Le recrutement de la tri-snRNP préassemblée U4/U6/U5 marque la formation du complexe B catalytiquement inactif. **(Complexe B*)** Des changements conformationnels majeurs (libération de U1 et U4), conduisent à l'activation du spliceosome et à la formation du complexe B*. **(Complexe C)** La première étape catalytique de l'épissage génère le complexe C et aboutit à la formation du lasso (intron cyclisé au niveau du BPS). Le complexe C réalise la deuxième étape catalytique de l'épissage, qui aboutit à la jonction des deux exons. **(Complexe post-épissage)** Après l'épissage, le spliceosome se désassemble de manière ordonnée, libérant l'ARNm ainsi que le lasso lié par U2/U5/U6. Les snRNP sont ensuite dissociées et recyclées. **(B)** Les séquences et protéines régulatrices de l'épissage mutées dans les cancers. Les facteurs d'épissage régulateurs (SF) qui peuvent se lier à des séquences *enhancers* exoniques ou introniques de l'épissage (ESE ou ISE) ou *silencers* (ESS ou ISS) sont également altérés dans les tumeurs humaines (Urbanski et al., 2018).

3.2. La régulation de l'épissage des ARN

L'épissage alternatif implique une *trans*-régulation par deux familles de protéines qui reconnaissent et se lient à des éléments *cis*-régulateurs. D'une part, les protéines riches en sérines et arginines (SR, *Serine-Arginine Rich*) se fixent sur des séquences activatrices, exoniques ou introniques (ESE ou ISE, *Exonic/Intronic Splicing Enhancer*) (Howard and Sanford, 2015). D'autre part, les hnRNP (*heterogenous nuclear ribonucleoproteins*) répriment l'épissage alternatif *via* l'interaction avec des séquences exoniques ou introniques (ESS ou ISS, *Exonic/Intronic Splicing Silencer*) (Geuens et al., 2016). Les différents *trans*-facteurs régulent de manière concentration-dépendante l'épissage alternatif (Bradley and Anczuków, 2023; Cáceres and Kornblihtt, 2002). On observe majoritairement des exclusions d'exon (ES, *Exon Skipping*) mais d'autres types d'épissage alternatif existent comme les exons mutuellement exclusifs (2 exons sont exprimés différemment l'un par rapport à l'autre), la rétention d'intron ou de site d'épissage 3' et/ou 5' (Corcos and Solier, 2005). Enfin, certains introns de grandes tailles (>50 kb) sont épissés par un mécanisme non-canonique dit récursif : des 3'SS cryptiques présents dans la séquence intronique en amont du 3'SS canonique permettent d'épisser progressivement ces grands introns (Sibley et al., 2015).

Les processus de transcription et d'épissage des ARN sont couplés dans le temps et l'espace : les ARN nouvellement transcrits sont immédiatement pris en charge par le splicéosome (Orphanides and Reinberg, 2002). De cette manière, la machinerie de transcription peut directement influencer la maturation des ARN messagers en modulant la processivité de l'ARN polymérase II *via* les éléments *cis*-régulateurs de la transcription – *i.e.* promoteurs et *enhancers* – et les marqueurs épigénétiques – *i.e.* H3K36me3 peut induire l'exclusion d'exons dans les cellules souches humaine – (Kadener et al., 2002; Luco et al., 2010; Tilgner et al., 2012). Réciproquement, certains éléments du splicéosome, comme la snRNP U1, peuvent stimuler ou inhiber la transcription (Théoleyre and Baklouti, 2004). Également, des études ont montré que le splicéosome est recruté et se fixe sur le domaine C-terminal de la grande sous-unité l'ARN polymérase II (Nojima et al., 2018). Ces données suggèrent que le système MMR pourrait interagir directement ou indirectement avec le splicéosome, *via* la marque épigénétique H3K36me3 (Frigola et al., 2017; Huang and Li, 2018).

Dans de nombreuses pathologies, l'épissage des ARN est dérégulé soit par des mutations dans les séquences régulatrices (*cis*-altérations) ou dans les gènes qui codent les protéines du splicéosome (*trans*-altération). C'est le cas de mutations germinales en *cis* dans le gène *CFTR* (*Cystic Fibrosis Transmembrane Conductance Regulator*) mis en cause dans la mucoviscidose, ou bien de mutations somatiques en *trans* affectant le gène *SF3B1* (*Splicing Factor 3B Subunit 1*) impliqués dans plusieurs cancers comme les mélanomes uvéaux, le cancer du sein ou du pancréas (Biankin et al., 2012; Dujardin et al., 2016; Harbour et al., 2013; Koboldt et al., 2012). Les gènes du splicéosome *SF3B1*, *SRSF2* (*Serine/Arginine-Rich Splicing Factor 2*) et *U2AF1* (*U2 Small Nuclear RNA Auxiliary Factor 1*) sont les plus fréquemment mutés dans les cancers (Bradley et al., 2023). Également, de nombreuses mutations observées en *cis* (dans le *polypyrimidine tract* ou dans les séquences régulatrices) sont associées à une altération de l'épissage alternatif des gènes – exclusion d'exon ou rétention d'intron anormale – dans les cancers (Figure 7) (Jung et al., 2015; Sveen et al., 2016). Ces altérations sont fréquentes dans les gènes clés des voies de signalisation impliquées dans le développement tumoral, *Hallmarks of cancer* (Hanahan, 2022).

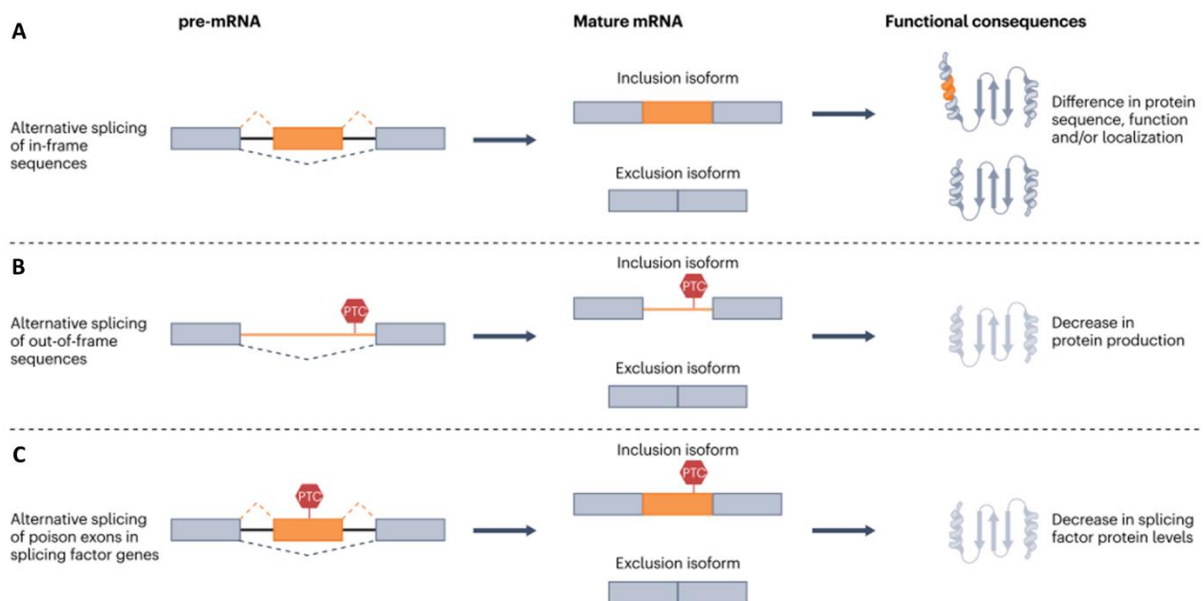


Figure 7 | Les évènements d'épissage alternatif les plus fréquemment observés dans les cancers.

(A) La rétention d'intron ou le saut d'exon, sans décalage du cadre lecture, peut conduire à l'ajout ou à la suppression de nucléotides codants pour des acides aminés, ce qui a un impact sur la structure, la fonction et/ou la localisation de la protéine. **(B)** La rétention d'intron ou le saut d'exon, avec décalage du cadre lecture, peut introduire un codon de terminaison prématurée, et donc un ARNm aberrant généralement dégradé par le NMD. **(C)** Les exons dits « empoisonnés » sont un type d'évènement d'épissage alternatif particulier correspondant à des exons qui, lorsqu'ils sont inclus, introduisent un PTC dans le transcrit. Ils sont très fréquents dans les gènes codant pour des facteurs d'épissage, ce qui permet de réguler leurs niveaux d'expression protéiques (régulation endogène). Dans les cancers, ces « exons poisons » sont associées à la surexpression ou à la sous-expression de facteurs d'épissage, respectivement, oncogènes ou suppresseurs de tumeurs. Également, ces évènements d'épissage alternatif coïncident à des isoformes initialement exprimées au cours du développement embryonnaire, puis modifiées lors de la différenciation des cellules adultes. Ce retour aux modèles embryonnaires est supposé permettre la prolifération des cellules cancéreuses et la plasticité phénotypique (Bradley and Anczuków, 2023).

4. La surveillance et la dégradation des ARNm aberrants : le système NMD

Le système NMD (*Non-Mediated Decay*) est responsable de la surveillance des transcrits messagers : il reconnaît et dégrade les ARNm qui présentent un PTC consécutif à une erreur de transcription, une mutation non-sens ou à un épissage aberrant. Ce processus est essentiel pour éliminer les protéines tronquées et les peptides néo-antigéniques, aux effets qui pourraient être délétères lorsqu'ils s'accumulent dans la cellule (Gardner, 2010). Dans la majorité des situations, le système NMD identifie les PTC *via* les complexes protéiques présents aux jonctions exon-exon (EJC, *Exon Junction Complexe*) (Figure 8) selon la règle qu'un

codon stop normal est situé à moins de 55 nucléotides en amont de la dernière jonction exon-exon (souvent localisé dans le 3'UTR) ; cependant, il existe certaines exceptions (Figure 9) (Kurosaki et al., 2019; Lindeboom et al., 2016; Tan et al., 2022).

Les gènes qui codent les protéines clés du NMD (*UPF1/2/3B* et *SMG1/5/6/7*) sont fréquemment mutés dans les cancers, ce qui induit une augmentation – comme dans les cancers colorectaux MSI – ou une diminution de l'activité du NMD (Bokhari et al., 2018; El-Bchiri et al., 2008; Tan et al., 2022). Également l'expression de gènes régulateurs du NMD peut être dérégulée : c'est le cas du répresseur du NMD, *UPF3A*, moins exprimé dans les carcinomes à cellules rénales et les cancers du côlon MSI (Gotoh et al., 2014; Michalak et al., 2020).

Des travaux suggèrent que le NMD favorise la progression tumorale en induisant la dégradation des ARN des gènes suppresseurs de tumeurs mutés avec PTC, ce qui a pour conséquence une diminution de leur expression. Également, il a été suggéré que les mutations engendrant un PTC mais qui ne déclenchent pas leur dégradation par le NMD, en majorité des oncogènes pro-tumoraux, ont tendance à être sélectionnées (Kishor et al., 2019; Lindeboom et al., 2016). Les mutations qui induisent des PTC peuvent aussi entraîner la synthèse de peptides néo-antigéniques présentés à la surface des cellules tumorales par le complexe d'histocompatibilité (MHC, *Major Histocompatibility Complex*). Les lymphocytes T CD4/8 reconnaissent ces antigènes comme des peptides du non-soi et initient une réponse immunitaire anti-tumorale qui cible spécifiquement les cellules qui expriment ces antigènes (Marrak and Kappler, 1987). Dans les mélanomes, les mutations avec PTC et queues néo-antigéniques qui échappent à la surveillance du NMD sont souvent sélectionnées négativement au sein de la tumeur, ce qui induit une faible présentation de peptides tumoraux par le MHC (Litchfield et al., 2020).

Le système NMD

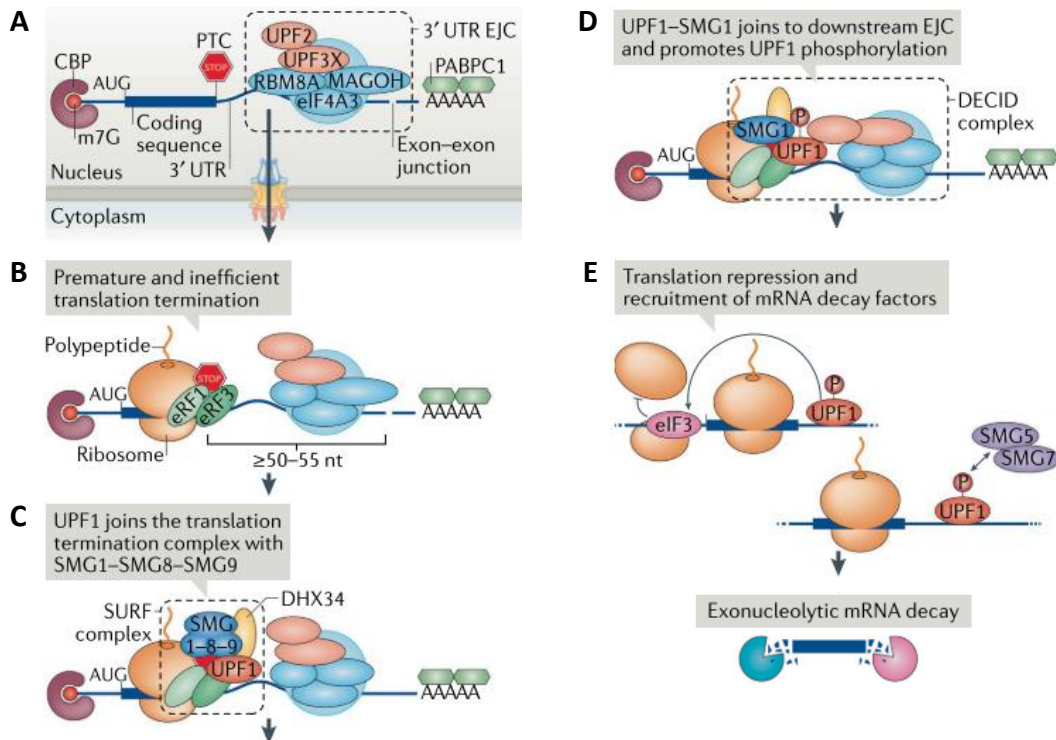


Figure 8 | La voie de signalisation EJC du système NMD.

(A) La coiffe (5'UTR) et la queue poly-A (3'UTR) sont essentielles pour le transport, la stabilité et la traduction des ARNm. Les complexes multi-protéiques principalement composés de eIF4A3, MAGOH, RBM8A (aussi appelé Y14) permettent d'identifier les jonctions exon-exon. PABPC1 (*PolyA-Binding Protein Nuclear 1 & Cytoplasmic 1*) est fixé à la queue poly-A.

(B) Les codons stop situés en amont de la dernière jonction exon-exon, à plus de 55 nucléotides sont reconnus comme aberrants par le ribosome, lors du premier cycle de traduction. Le complexe eRF1/3 (*Eukaryotic Release Factor 1/3*) interagit avec le ribosome ce qui active le système NMD.

(C) UPF1 (*UP-Frameshift 1*), une hélicase ATP-dépendante clé dans le mécanisme et SMG1/8/9 (*Suppressor with Morphogenic effect on Genitalia 1/8/9*) sont recrutés au niveau d'eRF1/3 pour former le complexe SURF.

(D) L'interaction du complexe EJC et de SURF (complexe DECID) *via* UPF2/3 entraîne un changement conformationnel d'UPF1, ce qui favorise sa phosphorylation activatrice par SMG1/8/9.

(E) L'activation d'UPF1 induit la répression de la traduction et le recrutement des endonucléases SMG5 et SMG7. Parallèlement, l'ARNm est décoiffé ce qui favorise la dégradation du transcrit.

Les signaux activateurs du système NMD

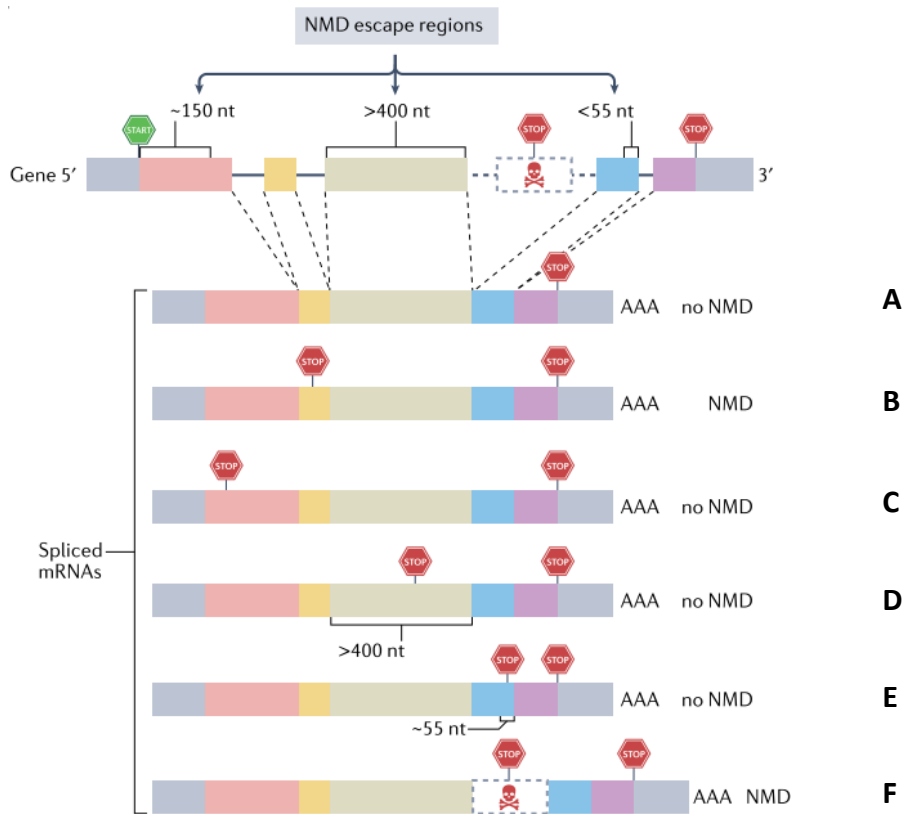


Figure 9 | Les différentes situations où un PTC déclenche (ou non) la dégradation des ARNm par le système NMD.

(A) Le codon stop est situé dans le dernier exon, c'est la situation normale, le système NMD n'est pas déclenché.

(B) Chez les mammifères, un codon stop est identifié comme prématuré lorsqu'il se situe en amont de la dernière jonction exon-exon, ce qui déclenche la dégradation de l'ARN par le système NMD.

Cependant, il existe des exceptions où cette règle ne s'applique pas. Les PTC localisés dans les ~150 nucléotides après le codon initiateur **(C)**, dans un grand exon (>400 nucléotides) au milieu de l'ARNm **(D)**, dans les ~55 nucléotides en amont de la dernière jonction exon-exon **(E)**, échappent à la surveillance du NMD ; les ARNm mutés ne sont donc pas éliminés.

(F) Enfin, lorsqu'un épissage alternatif (*frameshift*) induit un PTC, il est reconnu par le système NMD et l'ARNm est dégradé.

La majorité des PTC (51%) induisent une réponse forte, 27% une réponse faible et 22% n'induit pas de réponse du NMD. L'outil bio-informatique « *NMDetective* » prédit la réponse du NMD selon la mutation génératrice d'un PTC.

5. La surveillance immunitaire et ses mécanismes d'échappement

Comme je l'ai décrit dans la partie 2 de ce chapitre, la majorité des mutations associées à l'instabilité microsatellitaire génère des peptides néo-antigéniques dont certains peuvent être présentés par les cellules cancéreuses et ainsi déclencher une réponse immunitaire anti-tumorale (Lothe et al., 1993). L'accumulation de ces mutations immunogènes au cours de la tumorigenèse MSI explique la forte infiltration lymphocytaire observée dans les tumeurs de type MSI (Fridman et al., 2012; Galon et al., 2006; Mlecnik et al., 2016; Shia et al., 2013). La surveillance immunitaire vise à éliminer les cellules tumorales afin de protéger l'organisme. L'immunité anti-tumorale exerce ainsi une pression de sélection sur ces cellules cancéreuses qui peut favoriser l'émergence de mécanismes d'échappement immunitaire au sein de la tumeur (Hanahan, 2022).

Dans cette partie, je décris brièvement le mécanisme de surveillance immunitaire anti-tumorale. La compréhension de ce processus a été essentielle pour développer l'immunothérapie inhibitrice des points de contrôle de la réponse immunitaire et des thérapies vaccinales (voir chapitre 2, partie 5.1.3) ainsi que de le modèle de sélection des mutations au cours de la tumorigenèse (voir chapitre 3, partie 2.1.2).

5.1. La surveillance immunitaire anti-tumorale

La surveillance immunitaire anti-tumorale est assurée par de nombreuses cellules immunitaires qui agissent ensemble pour identifier les cellules cancéreuses et les éliminer efficacement. Parmi ces cellules, les lymphocytes T et B ainsi que NK (*Natural Killer*) sont les principaux effecteurs de la réponse immunitaire anti-tumorale ; elles reconnaissent les cellules tumorales *via* l'interaction de leur récepteur (T/BCR, *T/B Cell Receptor* et KIR, *Killer cell Ig-Receptor*) avec les peptides néo-antigéniques tumoraux présentés par le MHC. La lyse des cellules cancéreuses par les cellules NK et lymphocytes T entraîne le relargage d'antigènes tumoraux qui stimulent à nouveau les cellules immunitaires et ainsi intensifie la réponse immunitaire anti-tumorale. L'environnement immunitaire tumorale est également composé de cellules présentatrices d'antigènes (APC, *Antigen Presenting Cell*) – *i.e.* les cellules dendritiques et les macrophages - et de structures lymphoïdes tertiaires (TLS, *Tertiary Lymphoid Structure*) – ganglions lymphatiques riches en lymphocytes B et an APC – qui

permettent d'amplifier la réponse immunitaire anti-tumorale par les cellules effectrices (Figure 10) (Dieu-Nosjean et al., 2008; Fridman et al., 2012; Granier et al., 2017; Llosa et al., 2015; Pansy et al., 2021).

Les molécules de surface MHC – ou HLA (*Human Leukocyte Antigen*), chez l'homme – servent à la présentation des marqueurs du soi (antigènes normaux des cellules saines) ou du non-soi (antigènes tumoraux des cellules cancéreuses). Le HLA de classe I est composé d'une chaîne lourde α (à trois sous-domaines, HLA-A/B/C) et d'une chaîne légère β_2m pour β -2-microglobuline (codé par le gène *B2M*). Il permet d'activer les cellules effectrices NK (directement, *via* la cellule tumorale) ou des lymphocytes T (indirectement, *via* les APC). Le HLA de classe II est quant à lui constitué d'une chaîne lourde α (à deux sous-domaines) et d'une chaîne légère β (à deux sous-domaines) (Ahadova et al., 2018; Pansy et al., 2021; Yang et al., 2023). Le processus de présentation des antigènes débute dans le réticulum endoplasmique des cellules, où les peptides antigéniques sont transportés et chargés sur le HLA par les protéines TAP1/2 (*Transporter associated with Antigen-Processing 1/2*); le complexe HLA-antigène est ensuite exocyté à la surface des cellules (Axelrod et al., 2019; Kambayashi and Laufer, 2014).

La surveillance immunitaire exerce une pression de sélection (immunosélection ou *immunoediting*) lors de la tumorigenèse qui peut entraîner l'émergence de clones tumoraux qui échappent au contrôle immunitaire; ce processus, dynamique au cours du développement de la tumeur, a été théorisé en trois étapes critiques: (i) l'élimination, où les cellules immunitaires détectent et neutralisent les cellules tumorales; (ii) l'équilibre, le système immunitaire ne parvient plus à détruire toutes les cellules tumorales mais celles-ci restent sous son contrôle; (iii) l'échappement, les cellules tumorales ont développé des mécanismes de résistance à la surveillance immunitaire qui est alors dépassée. C'est l'altération de la surveillance immunitaire par différents mécanismes dans les cellules ou dans l'environnement tumoral, qui provoque le déséquilibre en faveur des cellules cancéreuses résistantes et ainsi l'échappement (Granier et al., 2017; Schreiber et al., 2011).

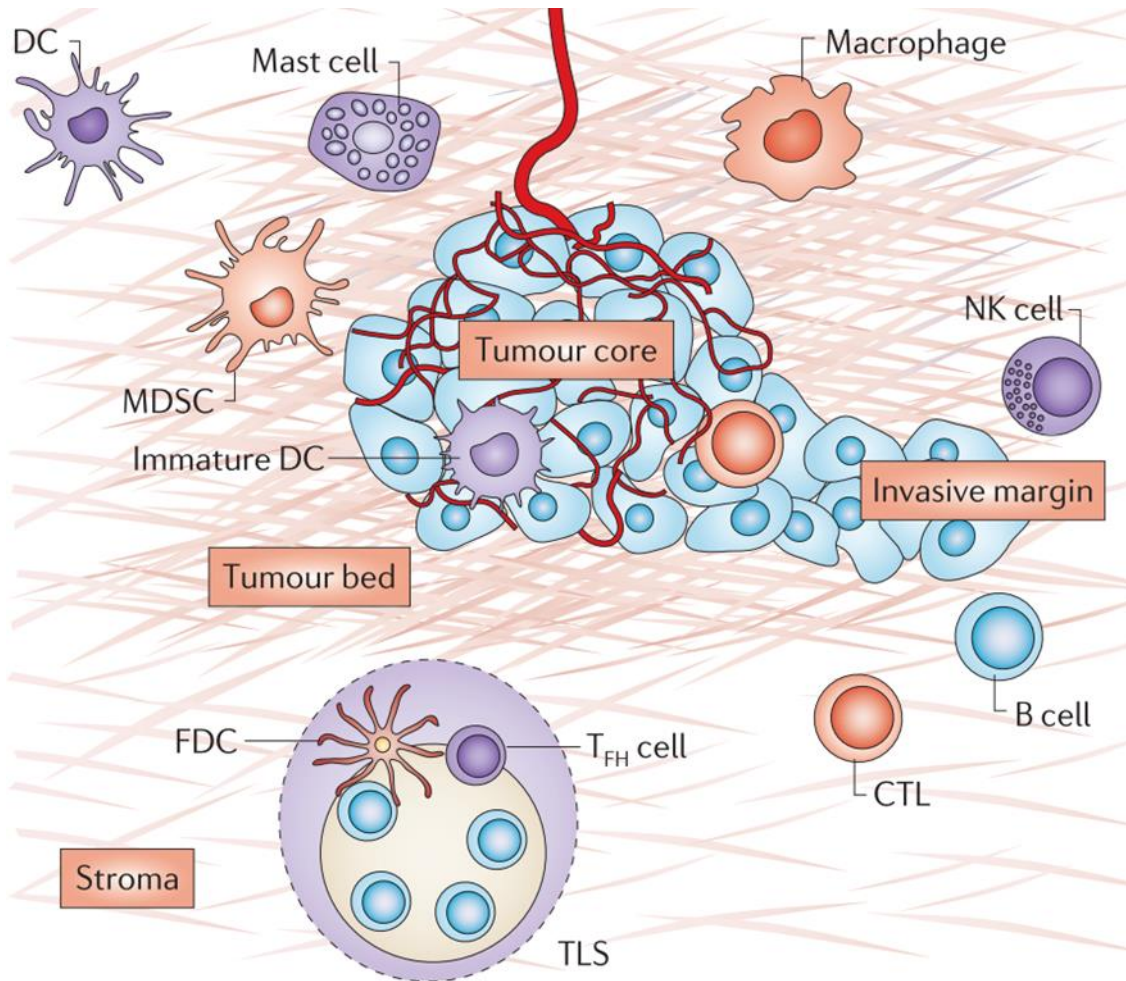


Figure 10 | L'environnement tumoral immunitaire.

Représentation schématique des cellules immunitaires infiltrantes de la tumeur impliquées dans la réponse immunitaire anti-tumorale. DC : cellule dendritique, MDSC : cellule suppressive d'origine myéloïde, CTL : lymphocyte T cytotoxique, FDC : cellules dendritiques folliculaire (Fridman et al., 2012).

5.2. Les mécanismes d'échappement de la surveillance immunitaire

5.2.1. L'altération de la présentation antigénique par le HLA

La reconnaissance des cellules tumorales par les cellules immunitaires effectrices *via* la présentation antigénique du HLA est essentielle pour déclencher la réponse immunitaire anti-tumorale (Figure 11A). L'altération de ce processus permettrait aux cellules cancéreuses de contourner la surveillance immunitaire et ainsi de maintenir la progression tumorale (Hanahan, 2022). Plusieurs mécanismes d'échappement des cellules tumorales de la surveillance immunitaire ont été décrits par les chercheurs (Pansy et al., 2021) ; dans cette partie, je présente les plus pertinents dans le contexte des cancers MSI.

L'altération de la présentation néo-antigénique par le HLA peut suffire à empêcher la reconnaissance des cellules tumorales par les cellules immunitaires (Figure 11B). La mutation d'un microsatellite du gène *B2M* observée dans 20-40% des cancers MSI et associée à la perte de fonction de la protéine, suffit à bloquer la formation du complexe HLA et la présentation des antigènes tumoraux (Kloor and Von Knebel Doeberitz, 2016; Yeon Yeon et al., 2019). Également, les mutations perte de fonction dans les autres gènes qui codent pour le HLA de classe I (*HLA-A/B/C*) ou de classe II (*HLA-DP/DQ/DR*) ont les mêmes conséquences (Bohaumilitzky et al., 2022; Kloor and Von Knebel Doeberitz, 2016). La répression transcriptionnelle de ces gènes (hyperméthylation des promoteurs) ou de la voie de signalisation NF- κ B qui régule l'expression du HLA de classe I, entraîne l'absence de présentation d'antigènes (Pansy et al., 2021). De manière générale, il a été démontré que les mécanismes de remodelage de la chromatine (méthylation, modifications des histones), essentiels pour l'interaction des cellules tumorales avec les lymphocytes, sont souvent détournés pour permettre l'échappement de la surveillance immunitaire des cellules cancéreuses (Cao et al., 2020; Kishor et al., 2019). Enfin, dans un contexte physiologique sain, les cellules qui n'expriment pas de HLA sont identifiées comme des cellules anormales par les cellules NK et éliminées. Cependant, dans les cellules tumorales, l'expression des gènes *HLA-E/F/G* qui codent pour les molécules HLA non-canoniques est observée. Ces molécules interagissent avec les cellules immunitaires effectrices et inhibent leur action cytotoxique (Pansy et al., 2021).

5.2.2. L'expression des inhibiteurs des points de contrôles immunitaires

Le déclenchement de la réponse immunitaire passe par la reconnaissance des antigènes présentés *via* les récepteurs des lymphocytes (T et B, NK). Cependant, ce signal n'est pas suffisant pour activer les cellules immunitaires et induire une réponse anti-tumorale. Un second signal doit être transmis par les co-récepteurs (activateurs ou inhibiteurs) : il s'agit des points de contrôles immunitaires (ICK, *Immune Checkpoint*) (Granier et al., 2017). Les co-récepteurs activateurs (tels que CD28, CD80 ou CD40L) stimulent la production de cytokines (IL-2, TNF- α) essentielle aux capacités prolifératives et cytotoxiques (activation) des lymphocytes. À l'inverse, les co-récepteurs PD-1 (*Programmed cell Death protein 1*) et CTLA-4 (*Cytotoxic T-Lymphocyte-Associated protein 4*) inhibent l'activation des lymphocytes effecteurs (Ishida et al., 1992; Leach et al., 1996).

La stimulation prolongée des co-récepteurs inhibiteurs par les ligands (CD80/86 et PD-L1 respectivement pour CTLA-4 et PD-1) induit une perte d'activité des cellules immunitaires ce qui leur confère un état d'épuisement (Day et al., 2006; Fourcade et al., 2010; Wherry et al., 2007). De cette manière, la surexpression de ces ligands inhibiteurs dans les cellules tumorales permet un échappement de la surveillance immunitaire, comme c'est le cas dans les cancers MSI (Figure 11C) (Dong et al., 2002; Marisa et al., 2018; Thompson et al., 2004). L'expression des ligands inhibiteurs par les cellules tumorales peut être antérieure (résistance primaire) ou postérieure (résistance secondaire) à l'activation des lymphocytes (Granier et al., 2017). L'expression innée de PD-L1 peut être associée à des événements oncogéniques comme la suractivation de la voie de signalisation PI3K-Akt suite à la mutation de PTEN ou de PI3KCA (voir chapitre 3, partie 2.1) ; alors que, le mécanisme secondaire adaptatif est modulé par la libération de cytokines inflammatoires – comme les interférons γ (INF- γ) ou l'interleukine-2 (IL-2) – par les cellules immunitaires qui induisent la transcription du ligand PD-L1 (Green et al., 2010; Kinter et al., 2008; Sznol and Chen, 2013).

5.2.3. Le rôle immunosuppresseur du microenvironnement

L'action des cellules immunitaires effectrices est régulée par la présence d'autres cellules immunitaires dans le microenvironnement tumoral comme les lymphocytes T régulateurs (Tregs) et les cellules suppressives d'origine myéloïde (MDSC, *Myeloid Dendritic Suppressive Cell*) (Figure 11D) (Eisenbarth et al., 2002; Wakkach et al., 2003). Ces cellules, recrutées par la tumeur (sécrétion de prostaglandine, PGE-2), inhibent l'activation des lymphocytes *via* la synthèse de cytokines immunosuppressive (comme IL-10 ou TGF- β) afin de bloquer la réponse immunitaire anti-tumorale et ainsi favoriser l'échappement immunitaire (Batlle and Massagué, 2019; Mannino et al., 2015; Oft, 2014). Également, les fibroblastes associés au cancer (CAF, *Cancer Associated Fibroblast*) présentes dans le microenvironnement tumoral, peuvent moduler la réponse immunitaire anti-tumorale en sécrétant des cytokines anti-inflammatoires (*i.e.* PGE-2) ou en remodelant la matrice extracellulaire (ECM, *Extra-Cellular Matrix*), ce qui perturbe l'activation des lymphocytes (Mhaidly and Mechta-Grigoriou, 2020; Monteran and Erez, 2019; Peltanova et al., 2019). Enfin, les protéines transmembranaires de la famille des Sémaphorines (comme, SEMA3A/B, SEMA4C/D), impliquées dans la migration des cellules immunitaires, sont souvent mutées dans les cancers MSI ce qui induit une activité immunosuppressive (Chen and Cuang, 2019; Hung et al., 2022).

ANTI-TUMOR IMMUNE RESPONSE

IMMUNE EVASION BY TUMORS

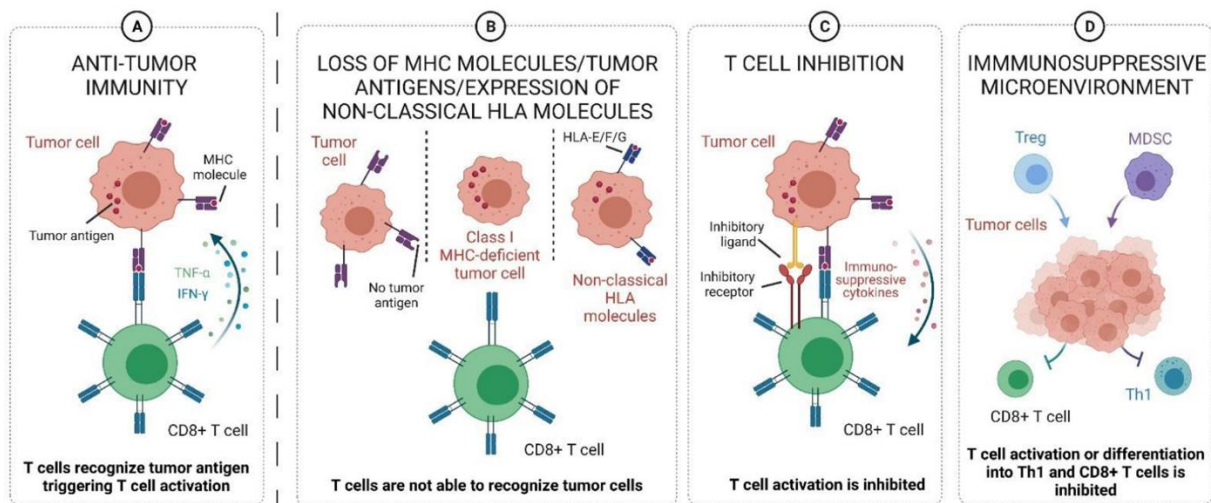


Figure 11 | Les mécanismes d'évasion immunitaire des tumeurs.

(A) La réponse immunitaire anti-tumorale classique médiée par les lymphocytes T. **(B)** L'altération de la présentation des antigènes tumoraux par le HLA inhibe l'activation des cellules immunitaires effectrices, incapables de reconnaître les cellules tumorales. **(C)** L'activation des inhibiteurs de checkpoints immunitaires *via* l'interaction des récepteurs à la surface des cellules effectrices et l'expression des ligands à la surface des cellules tumorales. Les cytokines immuno-suppressives entraînent l'inhibition de la réponse immunitaire anti-tumorale. **(D)** Les lymphocytes T régulateurs (Treg) et les cellules suppressives d'origine myéloïde (MDSC) s'accumulent et suppriment la réponse anti-tumorale des lymphocytes T cytotoxiques (Pansy et al., 2021).

Chapitre 2 – Les caractéristiques cliniques des cancers MSI

Les cancers MSI représentent 12-15% des cancers du côlon et du rectum, de l'endomètre et de l'estomac, sporadiques ou héréditaires (Aaltonen et al., 1993; Ionov et al., 1993; Thibodeau et al., 1993). Ils sont également présents dans plus de 50 localisations tumorales à plus faibles fréquences (Bonneville et al., 2021; Hause et al., 2016; Schwark and Stadler, 2018). Comme je l'ai décrit dans le chapitre précédent, l'instabilité génomique consécutive à l'inactivation du système MMR, est propice au développement de tumeurs caractérisées par un phénotype hypermutateur. Dans 90% des cancers colorectaux, l'inactivation du MMR est d'origine sporadique suite à une modification épigénétique (hyperméthylation du promoteur du gène MLH1). Dans 10% des cas, les cancers colorectaux MSI sont d'origine héréditaire et associés à une mutation hétérozygote germinale d'un des quatre gènes du MMR : c'est le syndrome de Lynch (Aaltonen et al., 1993; Cunningham et al., 1998; Peltomäki et al., 1992; Veigl et al., 1998). Enfin, des patients sont atteints d'une pathologie pédiatrique très rare (moins de 1% des cas) associée aux tumeurs MSI – *i.e.* le syndrome de déficience constitutionnelle du MMR (CMMRD, *Constitutional MMR-Deficiency*) ; les patients naissent avec un système MMR-déficient suite à une mutation bi-allélique germinale d'un des gènes du MMR. Actuellement, les spécialistes estiment qu'environ 200 personnes sont atteintes par le CMMRD dans le monde (Ricciardone and Tayfun, 1999; Wang et al., 1999).

Ce deuxième chapitre portera sur la caractérisation épidémiologique, moléculaire et clinique des cancers MSI sporadiques et associés au syndrome de Lynch, dans la localisation colorectale.

1. La classification moléculaire et les signatures mutationnelles des CRC MSI

Le type d'instabilité génomique permet de distinguer deux catégories de cancers coliques : (i) les tumeurs avec instabilité microsatellitaire (MSI) et (ii) les tumeurs avec instabilité chromosomique (CIN+/MSS). Les cancers colorectaux représentent une maladie très hétérogène d'un point de vue moléculaire, morphologique et clinique. Mon laboratoire d'accueil en collaboration avec le CIT (Carte d'identité des tumeurs, programme de recherche

dédié à la génomique, initié par la Ligue Nationale Contre le Cancer) a publié une première classification des cancers colorectaux en six sous-types grâce aux données transcriptomiques de 750 patients (Marisa et al., 2013). En 2015, à partir de plus de 4 000 patients, le travail d'un consortium international, a permis de classifier les cancers colorectaux en quatre sous-types. Il s'agit de la classification CMS (CMS, *Consensus Molecular Subtype*) : le groupe CMS1 (14% des tumeurs) enrichi en tumeurs MSI, est caractérisé par un phénotype hypermutateur, une infiltration immunitaire importante et par des altérations épigénétiques élevées dans les tumeurs, objectivées par un phénotype de méthylation des îlots CpG (CIMP). Les CMS2-4 regroupent les cancers CIN selon une surexpression des voies de signalisation WNT et MYC (CMS2 canonique, 37%), une dérégulation métabolique (CMS3 métabolique, 13%) et l'activation de TGF- β et de l'angiogenèse (CMS4 mésenchymateux, 23%). Il est à noter que 21% des CRC n'appartiennent à aucun de ces quatre groupes (Guinney et al., 2015).

Alexandrov et ses collègues ont analysé 84 729 690 mutations somatiques provenant de séquençage de l'exome et du génome entier (WES/WGS, *Whole Exome/Genome Sequencing*) de 23 829 tumeurs, toutes localisations et tous types confondus. Les résultats de ces travaux, disponibles sur la base de données COSMIC (*Catalogue Of Somatic Mutations In Cancer*), ont permis de déterminer 81 signatures mutationnelles distinctes. Ces signatures sont pour la plupart associées à des processus biologiques tels que la (l'in)fidélité de l'ADN polymérase, l'exposition à des agents mutagènes (endo/exogènes), à des modifications de bases ou encore à la déficience des mécanismes de réparation de l'ADN. Certaines signatures sont communes à plusieurs cancers (localisations et/ou sous-types) alors que d'autres sont spécifiques. L'inactivation du système MMR est associée aux signatures mutationnelles SBS6/15/21/26/44, DSB7/10 et ID7 qui décrivent des transitions C>T et des *indels* caractéristiques des cancers MSI (Alexandrov et al., 2020, 2013).

La classification moléculaire CMS et les signatures mutationnelles distinguent les voies de développement des cancers colorectaux (CIN *versus* MSI). Cependant, à ce jour, aucune de ces caractéristiques ne permet de différencier la tumorigenèse MSI sporadique par rapport aux formes héréditaires.

2. Les cancers colorectaux MSI sporadiques

En France, environ 47 000 nouveaux cas de CRC sont recensés chaque année et 17 000 en décèdent ; il s'agit respectivement du 3^{ème} et 2^{ème} cancer le plus fréquent, chez l'homme et la femme (INCa 2023). Les cancers MSI sporadiques représentent environ 12% des CRC (Colle and Cohen, 2019; Kavun et al., 2023). Les cancers colorectaux MSI sporadiques se développent majoritairement dans le côlon droit (81%), et plus fréquemment chez les femmes de plus de 70 ans (Cohen et al., 2017; Malkhosyan et al., 2000; Tsai et al., 2018). Des études ont montré que ce risque plus élevé chez la femme pourrait être associé à la diminution de la production d'œstrogène après la ménopause et il a été suggéré que ce risque pourrait être compensé par un traitement hormonal préventif (Kim et al., 2015; Rossouw et al., 2002; Slattery et al., 2001). Chez les hommes, la consommation d'alcool et de tabac accroît le risque de développer un cancer colorectal (Longnecker, 1990; Richman, 2015). Également, des chercheurs supposent que la localisation proximale (côlon droit) des cancers MSI soit associée au microbiote et à un environnement immunitaire favorable pouvant être corrélé à la présence de cellules de Paneth – l'existence de ces cellules dans le côlon distal est rare – (Fodde et al., 2023; Jin et al., 2021; Ong et al., 2014).

2.1. Caractéristiques génétiques et moléculaires

Dans la majorité des cas (70-95%), la répression transcriptionnelle du gène MLH1 – hyperméthylation bi-allélique du promoteur – induit l'inactivation du système MMR et le développement de tumeurs MSI sporadiques (Cunningham et al., 1998; Kane et al., 1997; Veigl et al., 1998). Des analyses épigénétiques ont montré que l'hyperméthylation du promoteur de MLH1 augmente en vieillissant, ce qui peut expliquer l'âge tardif de développement des tumeurs MSI sporadiques (Kakar et al., 2003). Le gène BRAF (oncogène de la voie RAS/RAF/MAPK) est fréquemment muté dans les cancers colorectaux MSI sporadiques (60%) (Kambara et al., 2004; Parsons et al., 2012; Weisenberger et al., 2006). Plus rarement (5-30%), le déclenchement de la tumorigenèse MSI sporadique est consécutif à des simples ou doubles mutations somatiques des gènes du MMR, *i.e.* MLH1, MSH2, MSH6 ou PMS2 (Baretti and Le, 2018; Nicolaidis et al., 1994; Pearlman et al., 2019; Sourrouille et al., 2013). Des mutations dans les gènes codants pour la sous-unité catalytique des ADN polymérase ϵ et δ (POLE et

POLD1) – responsable de l’activité de relecture du complexe répliatif – ont été observées dans les cancers colorectaux MSI. La coexistence des mutations POLD1/POLE/MMR semble participer au phénotype hypermutateur, sans pour autant être associée à l’instabilité dans les séquences microsatellitaires (Koboldt et al., 2012; Palles et al., 2013).

Des altérations des mécanismes de *trans*-régulation de l’expression ou de l’activité du système MMR ont été mises en évidence dans les cancers colorectaux MSI sporadiques. C’est le cas de la surexpression des micro-ARN (miRNA) miR-155 et miR-21 qui régulent négativement l’expression des protéines du MMR – les miRNA se lient par complémentarité de séquences aux ARNm cibles et répriment leur traduction ou induisent leur hydrolyse *via* le complexe RISC – (Fire et al., 1998; Svrcek et al., 2013; Valeri et al., 2010). Des mutations perte de fonction du gène SETD2, responsable de la triméthylation de la lysine 36 de l’histone 3 (H3K36me3) sont observées dans certains cancers colorectaux MSI sporadiques. Cette mutation pourrait être à l’origine d’une inactivation du système MMR (voir chapitre 1, partie 1.2) (Huang and Li, 2020; Li et al., 2013).

2.2. Valeur pronostic des cancers colorectaux MSI sporadiques

Une méta-analyse (7 642 patients atteints de CRC, tous stades confondus, dont 1 277 tumeurs MSI) a montré qu’à stade de progression tumorale égale, les patients atteints de cancers colorectaux MSI ont un taux de survie plus élevé que ceux atteints de cancers MSS (Popat et al., 2005; Samowitz et al., 2007; Sinicrope et al., 2006). En intra-MSI, les tumeurs de stade II (localisées, non-métastatiques) semblent être de meilleurs pronostics que celles de stade III (localisées, avec des métastases ganglionnaires) (Cohen et al., 2017; Roth et al., 2012). En revanche, les tumeurs MSI de stade IV (présence de métastase à distance), peu fréquentes (3-5%), sont de moins bons pronostics que les tumeurs métastatiques de type MSS (Koopman et al., 2009; Malesci et al., 2007). Quatre essais cliniques randomisés (CAIRO, CAIRO2, FOCUS et COIN ; 3 063 patients au total) montrent ce pronostic défavorable des CRC métastatiques (mCRC) de phénotype MSI : la survie globale (OS, *Overall Survival*) et la survie sans aggravation de la maladie (PFS, *Progression Free Survival*) diminuent significativement chez les patients de la cohorte MSI par rapport à celle des patients atteints d’un CRC MSS (Venderbosch et al., 2014). Il semblerait que la mutation BRAF V600E, fréquente dans les tumeurs comme dans les

métastases MSI, soit associée à ce mauvais pronostic (Goldstein et al., 2014; Venderbosch et al., 2014).

3. Les cancers MSI héréditaires : le syndrome de Lynch

Le syndrome de Lynch (LS) est décrit pour la première fois par Aldred Scott Warthin en 1895 dans une famille atteinte par de nombreux cancers du côlon, de l'estomac, et de l'endomètre à un âge précoce (Warthin, 1985). Initialement nommé HNPCC (*Hereditary Non-Polyposis Colorectal Cancer*) par opposition à la FAP (*Familial Adenomatous Polyposis*) – un syndrome de prédisposition aux cancers digestifs de type CIN, caractérisé par la présence d'un grand nombre de polypes intestinaux, associé à une mutation du gène APC – les cancers associés au syndrome de Lynch présentent peu de lésions précoces adénomateuses (tumeurs bénignes) (Peltomäki et al., 1992). Les patients atteints par le syndrome de Lynch sont porteurs d'une mutation mono-allélique germinale dans un gène du système MMR, au mode de transmission autosomique dominant, comme l'a démontré Henry T. Lynch qui donnera son nom à la pathologie (Lynch et al., 1966).

3.1. Caractéristiques épidémiologiques et prédispositions génétiques

Le syndrome de Lynch est responsable de 2-5% des cancers colorectaux, soit 37% des cancers colorectaux MSI et 26% des cancers de l'endomètre MSI (16,3% des tumeurs MSI, toutes localisations confondues) (Colle and Cohen, 2019). Dans la population générale, 1 personne sur 249 est porteuse d'une mutation germinale d'un des gènes du MMR caractéristique du LS, soit 20 à 25 millions d'individus atteints dans le monde (Peltomäki, 2005; Win et al., 2017). Chez les patients atteints par le LS, le risque de développer un cancer avant 70 ans est bien plus élevé que dans la population générale (la médiane de survenue du premier cancer est de 46 ans) ; ce risque est influencé par le sexe de l'individu, ainsi que le gène touché par la mutation germinale (Kavun et al., 2023; Sehgal et al., 2014). Les femmes atteintes par le LS ont une probabilité moins élevée de développer un CRC MSI par rapport aux hommes (10-52% *versus* 20-69%), mais elles développent un cancer de l'endomètre dans 40 à 60% des cas (Burton-Chase et al., 2017).

Les mutations germinales du gène MSH2 et MLH1 sont les plus fréquentes dans les cancers colorectaux (40-45%) alors que les mutations des gènes MSH6 et PMS2 ne représentent que 5-10% (Evrard and Alexandre, 2021; Poulogiannis et al., 2010; Rossi et al., 2017). Il a été décrit, dans 1% des cas, qu'une large délétion de la région 3' du gène EpCAM (*Epithelial Cell Adhesion Molecule*) – en amont du gène MSH2 – peut induire l'inactivation transcriptionnelle du gène MSH2 (Ligtenberg et al., 2009). Dans de rares cas (3%), la déficience du système MMR est consécutive à l'épimutation constitutionnelle des promoteurs des gènes MLH1 et/ou MSH2, qui entraîne une répression transcriptionnelle ; cette modification épigénétique est stable dans le temps et héréditaire (Hitchins and Ward, 2009; Pinto et al., 2018). Dans de rare cas (<10%), l'inactivation du MMR peut être consécutive à l'hyperméthylation du promoteur MLH1 – associée ou non à une mutation BRAF V600E – dans les CRC MSI associé au syndrome de Lynch (Kientz et al., 2019; Peltomäki, 2014; Yokoyama et al., 2018). Selon le modèle Knudson, une altération somatique du second allèle – le plus souvent, de type perte d'hétérozygotie (LOH, *Loss Of Heterozygotie*), répression transcriptionnelle épigénétique ou mutation somatique – suffit à inactiver le système MMR et à déclencher la tumorigenèse MSI (Knudson, 1971). La pénétrance de la pathologie est incomplète – *i.e.* les patients sont atteints par la prédisposition génétique mais ne développent pas de cancers – selon les gènes du système MMR mutés : les patients porteurs d'une mutation germinale dans les gènes PMS2 ou MSH6 présentent une plus faible pénétrance que ceux avec une mutation MLH1 ou MSH2 (Truninger et al., 2005). Ceci a pour conséquence une sous-estimation des mutations MSH6 et PMS2 associées au syndrome de Lynch, puisque ces patients ne sont pas diagnostiqués (Espenschied et al., 2017).

Comme les tumeurs MSI coliques sporadiques, les formes héréditaires sont localisées principalement dans le côlon proximal (70%) et elles sont très infiltrées en cellules immunitaires. Cependant, les tumeurs coliques LS ont un risque plus important de progresser en métastase (35% des cas) (Lynch et al., 2009; Schneider et al., 2012; Vasen et al., 1999). Également, les tumeurs coliques du LS se développent trois fois plus vite que celles d'origine sporadique (3 ans *versus* 10 ans), ce qui peut s'expliquer par des voies de la tumorigenèse différentes (Carethers and Stoffel, 2015; Edelstein et al., 2011) ; ces aspects seront détaillés par la suite (voir chapitre 3, partie 1).

3.2. Critères d'identification du syndrome de Lynch

En 1991 lors d'une conférence à Amsterdam, des experts internationaux se sont réunis pour proposer des critères cliniques d'identification des patients atteints par le syndrome de Lynch (Vasen et al., 1991). Ces critères, trop stricts, ont été révisés et complétés pour inclure des critères anatomopathologiques ainsi que les cancers associés aux localisations extra-coliques (Table 1). Malgré cela, entre 28-50% des patients atteints par ce syndrome ne sont pas pris en compte dans les critères (Ryan et al., 2017). De ce fait, depuis 2009 les experts proposent que tous les patients atteints d'un cancer colorectal avant 70 ans soient dépistés pour un phénotype MSI (PCR et IHC voir partie 4.1) (Balmaña et al., 2013; Boland and Shike, 2010). En France, en dehors des contextes familiaux décrits dans les critères d'Amsterdam II/Bethesda (Table 1), l'Institut Nationale du Cancer (INCa) préconise un dépistage systématique des patients atteints d'un cancer colorectal avant 60 ans, à la recherche de mutations somatiques des gènes du MMR (INCa 2016).

Critères d'Amsterdam II
Au moins trois apparentés du cas index sont atteints d'un cancer du spectre LS*
<ol style="list-style-type: none">1. L'un des trois individus doit être apparenté au premier degré avec les autres2. Au moins deux générations successives sont concernées3. Au moins un des cas de cancer du spectre LS doit être diagnostiqué avant 50 ans4. La polypose adénomateuse familiale (FAP) doit être exclue dans les cas de CCR5. Le diagnostic histologique des tumeurs doit être vérifié
Critères de Bethesda révisés
Le statut MSI des tumeurs doit être analysé dans les cas suivants :
<ol style="list-style-type: none">1. CCR diagnostiqué avant 50 ans2. Présence de tumeurs du spectre LS synchrones et métachrones quel que soit l'âge3. CCR avec une histologie évocatrice** du phénotype MSI, diagnostiqué avant 60 ans4. Cancer du spectre LS diagnostiqué chez au moins un parent au premier degré avant 50 ans5. Cancer du spectre LS diagnostiqué chez au moins deux parents de premier ou deuxième degré quel que soit l'âge

Table 1 | Critères d'Amsterdam II et de Bethesda révisés.

*Le spectre de Lynch comprend les cancers colorectaux, endomètre, estomac, ovaire, intestin grêle. ** L'histologie évocatrice fait référence à la présence de lymphocytes intra-tumoraux, réaction inflammatoire de type Crohn (Lynch et al., 2015; Umar et al., 2004; Vasen et al., 1999).

3.3. Surveillance clinique

À stade égal, la survie des patients LS est meilleure que pour les patients atteints de cancers sporadiques, et ce malgré le développement précoce de tumeurs coliques (Stigliano et al., 2008) ; cela peut notamment s'expliquer par le suivi précoce et continu des patients diagnostiqués LS. La surveillance des CRC consiste à effectuer une coloscopie tous les deux ans, dès l'âge de 20 ans, puis tous les ans à partir de 40 ans ; une échographie endovaginale tous les deux ans complète la surveillance chez la femme à partir de 30 ans (INCa 2017). L'endoscopie colique est une mesure qui permet de diminuer la fréquence de développement des CRC à hauteur de 40% chez les patients LS, grâce à l'élimination des adénomes précoces (Järvinen et al., 2000). L'efficacité de la surveillance précoce et en continu n'a pas été spécifiquement démontrée pour les autres localisations.

Des thérapies préventives comme la chirurgie prophylactique ou la chimioprévention ont été suggérées (Lecomte et al., 2021). Chez les femmes, une hystérectomie (ablation partielle ou totale de l'utérus) est envisageable pour prévenir les cancers de l'endomètre (Schmeler et al., 2006). Cependant, l'ablation prophylactique du côlon n'est pas recommandée, considérant l'efficacité de la surveillance colique et les conséquences d'une colectomie partielle ou totale pour le confort du patient (diarrhée chronique, incontinence) (Boland et al., 2018).

4. Le diagnostic des cancers MSI

La détermination du phénotype d'instabilité microsatellitaire a une valeur pronostic et thérapeutique dans de nombreuses localisations tumorales (Sinicrope et al., 2006; Zhu et al., 2015). De ce fait, les sociétés internationales d'oncologie recommandent un dépistage systématique du statut MSI dans les cancers colorectaux (Luchini et al., 2019; Pai and Pai, 2016; Van Cutsem et al., 2016). Les approches standards utilisées en routine dans les services d'anatomopathologie sont l'évaluation de l'expression des protéines MMR par marquage immunohistochimique et la détermination du statut d'instabilité microsatellitaire par PCR (Svrcek et al., 2019). De nouvelles approches développées à partir du séquençage à haut-débit (NGS) ou de l'analyse d'image par des intelligences artificielles (AI, *Artificial Intelligence*) émergent (Gilson et al., 2021; Kather et al., 2019; Ratovomanana et al., 2021). Ces

technologies permettent d'assister, de faciliter et de standardiser le travail des cliniciens dans la recherche du phénotype MSI des tumeurs.

4.1. Approches standards reconnues

Le marquage immunochimique (IHC, *ImmunoHistoChemistry*) et la réaction en chaîne par polymérase (PCR, *Polymerase Chain Reaction*) sont des techniques simples à mettre en place dans les services hospitaliers adaptés. Elles permettent d'obtenir un résultat rapide et fiable. En complément, des analyses de l'hyperméthylation du promoteur MLH1 et de la mutation BRAF V600E par pyroséquençage (séquençage par synthèse nucléotidique marqué au pyrophosphate) permettent de préciser l'origine moléculaire du cancer (Balmaña et al., 2013; Xiao et al., 2020). Ces deux analyses complémentaires sont réalisées en routine à l'hôpital Saint-Antoine et au CHU de Lille, avec lesquels j'ai collaboré durant ma thèse.

4.1.1. L'immunohistochimie : évaluation de l'expression des protéines du MMR

Le marquage des protéines du système MMR (MLH1, MSH2, MSH6 et PMS2) par IHC permet de définir leur expression dans un tissu d'intérêt (tumoral) ; l'absence de marquage d'une de ces quatre protéines indique une déficience du système de réparation et donc un phénotype MSI de la tumeur (Figure 12A). Afin de contrôler la bonne qualité du marquage, il est nécessaire de valider la présence de cellules positives dans le tissu normal adjacent ainsi que les lymphocytes T infiltrant la tumeur (Richman, 2015). L'absence de marquage d'une des protéines du MMR est réciproque avec l'autre protéine du complexe : par exemple, la perte de MLH1 ou MSH2 est corrélée à une perte d'expression respectivement de PMS2 ou MSH6. Cela s'explique par une déstabilisation des hétérodimères, suite à la perte d'expression d'une des deux protéines du complexe, qui entraîne la dégradation par le protéasome du binôme protéique. Toutefois, une rétention d'expression de MLH1 ou MSH2 est parfois observée en cas de déficience respective de PMS2 et MSH6 (perte hétérogène : cela s'explique par la formation d'hétérodimères MLH1-PMS1 ou MSH2-MSH3 qui stabilise les protéines (Colas et al., 2012; Shia, 2008). Dans de très rares cas, la rétention des protéines MSH6 ou PMS2 est également possible lorsque MLH1 ou MSH2 sont respectivement déficients ; ce phénomène encore mal compris peut s'expliquer par l'existence de mutations stabilisatrices directes ou indirectes de MSH6 ou PMS2 (Gill et al., 2005; McCarthy et al., 2019).

Le marquage IHC est une technique de diagnostic très sensible (> 90%) et spécifique (98-100%), peu onéreuse et rapide mais compliquée à mettre en œuvre pour une analyse à haut débit (Jaffrelot et al., 2022; Lindor et al., 2002; Yuan et al., 2015). Dans de très rares cas (5-11%), une mutation faux-sens provoque la perte de fonction de l'un des gènes du MMR sans perturbation son expression : l'IHC ne permet pas l'identification du statut MSI dans ce cas-là (Baretti and Le, 2018; Vasen and Peltomäki, 2004). L'utilisation d'une deuxième méthode complémentaire (PCR) permet de pallier cette limite.

4.1.2. La PCR : détermination du statut d'instabilité microsatellitaire

Comme nous l'avons déjà vu, l'inactivation du système MMR déclenche une instabilité génomique notamment dans les séquences microsatellitaires qui subissent des variations de leur taille de répétition (voir chapitre 1, partie 2.1). La PCR Pentaplex est une technique de détection du statut MSI qui repose sur l'analyse moléculaire de la taille de répétition de cinq marqueurs microsatellitaires non-codants, en comparant les tissus normaux de référence et tumoraux d'un même patient. Dans un premier temps, un panel de deux microsatellites mononucléotidiques (BAT25 et BAT26) et trois dinucléotidiques (D5S346, D2S123 et D17S250) a été proposé : il s'agit de la version NCI-Pentaplex (Boland et al., 1998). En 2006, mon équipe a proposé l'utilisation d'un nouveau panel de cinq marqueurs microsatellitaires mononucléotidiques (BAT25, BAT26, NR21, NR24 et NR27) (Figure 12B). Les tumeurs sont diagnostiquées MSI lorsqu'au moins 2 des 5 marqueurs sont instables (Buhard et al., 2006). Ce panel permet de s'affranchir du tissu sain de référence nécessaire dans la version NCI-Pentaplex, étant donné le caractère quasi-monomorphe de ses marqueurs ainsi que d'assurer la détection de l'instabilité microsatellitaire consécutive aux mutations perte de fonction du gène *MSH6*. En effet, lors de la réparation MMR, la protéine MSH6 reconnaît seulement les petits mésappariements comme ceux observés dans les microsatellites mononucléotidiques, alors que la protéine MSH3 répare les plus grands mésappariements tels que ceux dans les microsatellites dinucléotidiques (voir chapitre 1, partie 1.1) qui constituent le panel NCI-pentaplex (Umar et al., 2004; Wong et al., 2006; You et al., 2010). La pentaplex est dorénavant le panel de référence mondiale à utiliser en routine pour le diagnostic du phénotype MSI par PCR (Svrcek et al., 2019).

De nouveaux marqueurs microsatellitaires, indépendamment ou en complément, ont été suggérés afin d'améliorer le diagnostic du phénotype MSI. Par exemple, les microsatellites mononucléotidiques (MT1XT20, EWS16T), localisés en 3' UTR respectivement des gènes *MT1X* (*Metallothionein 1X*) et *EWSR1* (*EWing Sarcoma breakpoint Region 1*) (Kishore et al., 2014; Kondelin et al., 2015; Morandi et al., 2012). Également, mon équipe a proposé l'utilisation du microsatellite HT17 (très monomorphe) – situé dans la séquence polypyrimidique en amont de l'exon 9 du gène HSP110 – afin de diagnostiquer le statut MSI dans les cancers colorectaux (Buhard et al., 2016). La grande sensibilité de ce marqueur a été confirmée dans une étude indépendante ; cependant, il doit être validé dans d'autres localisations tumorales pour généraliser son utilisation (Berardinelli et al., 2018).

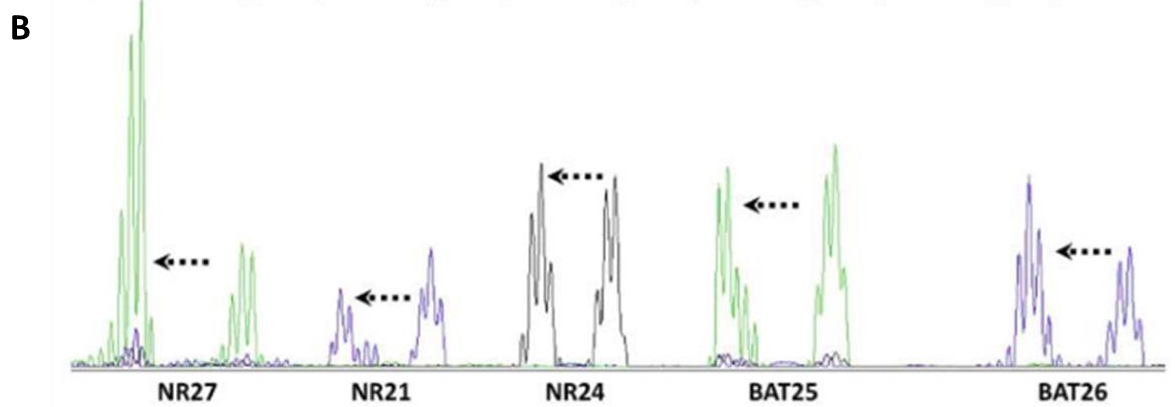
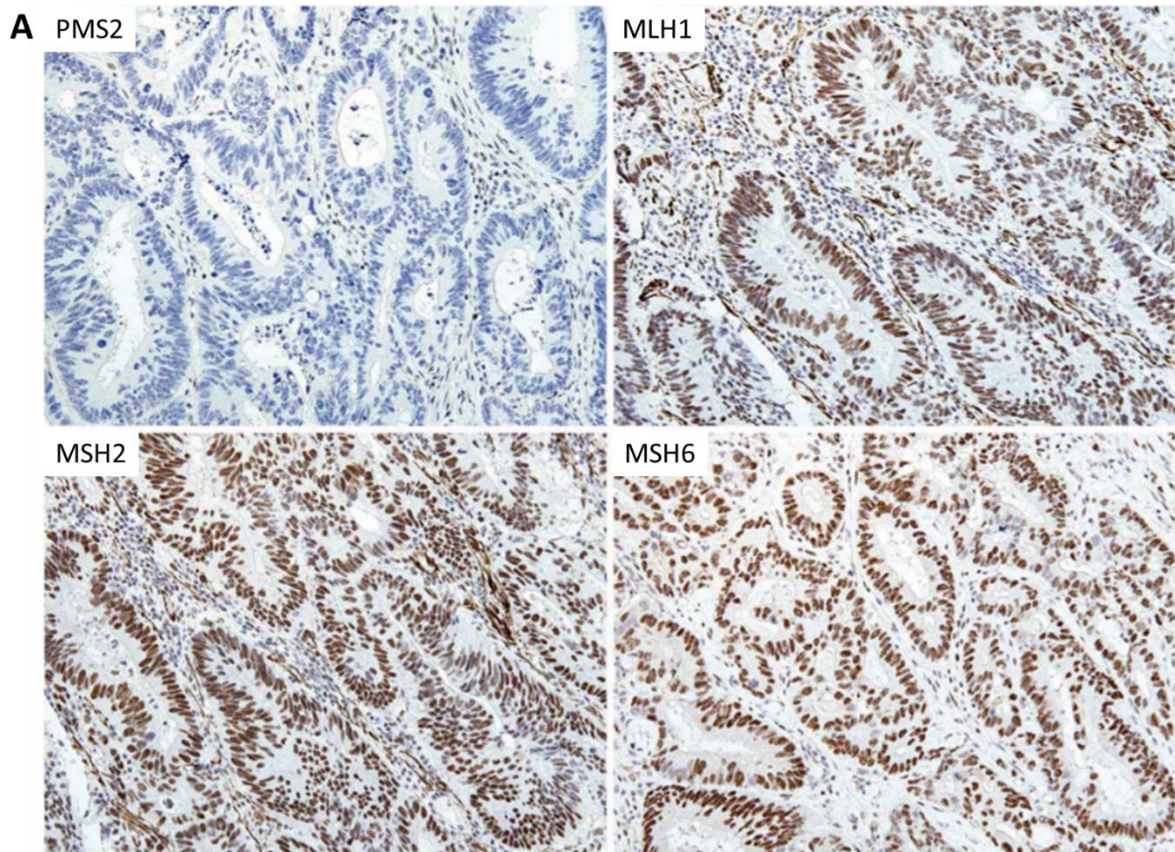


Figure 12 | Les outils conventionnels du diagnostic des cancers colorectaux MSI.

(A) Le marquage immunohistochimique des 4 protéines MMR chez ce patient permet de déterminer la perte d'expression isolée de PMS2 dans une tumeur colique, avec conservation des protéines MLH1, MSH2 et MSH6. **(B)** Amplification par PCR des 5 microsatellites du panel Pentaplex. L'instabilité des microsatellites se traduit par un raccourcissement des séquences répétées dans l'ADN tumoral par rapport à l'ADN normal (symbolisé par la flèche) (Adapté de Boland and Shike, 2010; Richman, 2015).

Des cas de discordance entre les résultats d'analyse de la PCR pentaplex et de l'IHC peuvent se présenter. Dans la majorité des cas, la discordance est due à l'utilisation de panel de deux anticorps uniquement (anti-MLH1, anti-MSH2) moins fiable que l'utilisation de quatre anticorps (concordance PCR-IHC de 99,6% sur 3228 patients testés) (Guyot D'Asnières De Salins et al., 2021; Jaffrelot et al., 2022). Les cas de discordances peuvent également être dus à une mauvaise interprétation des résultats, lorsque le tissu normal de référence est contaminé en tissu tumoral (PCR NCI-Pentaplex) et/ou lorsque le marquage IHC des protéines du MMR des lames histologiques est trop hétérogène. En cas de discordance persistante, les techniques de PCR-IHC doivent être réitérées afin de statuer sur des résultats de confiance ; cela permet d'éviter des erreurs diagnostiques du statut MSI et d'entraîner une mauvaise prise en charge des patients (Cohen et al., 2019a).

4.1.3. Diagnostic pathologique : sporadique *versus* syndrome de Lynch

Les approches standards présentées précédemment permettent de diagnostiquer le phénotype MSI d'une tumeur. Outre les critères de Bethesda – qui ne suffisent pas toujours à identifier le caractère héréditaire d'un cancer MSI – des analyses moléculaires complémentaires basées sur le pyroséquençage du promoteur de MLH1 et du gène BRAF, peuvent orienter les cliniciens dans le cas d'une suspicion de LS. Les cancers sporadiques sont caractérisés par une hyperméthylation du promoteur de MLH1, corrélées dans 90% des cas avec la mutation BRAF V600E, alors que les cancers du LS présentent très rarement ces attributs (Kientz et al., 2019; Parsons et al., 2012; Yokoyama et al., 2018). Afin d'identifier l'origine des cancers MSI des patients, l'Institut National du Cancer a publié des recommandations cliniques (Figure 13).

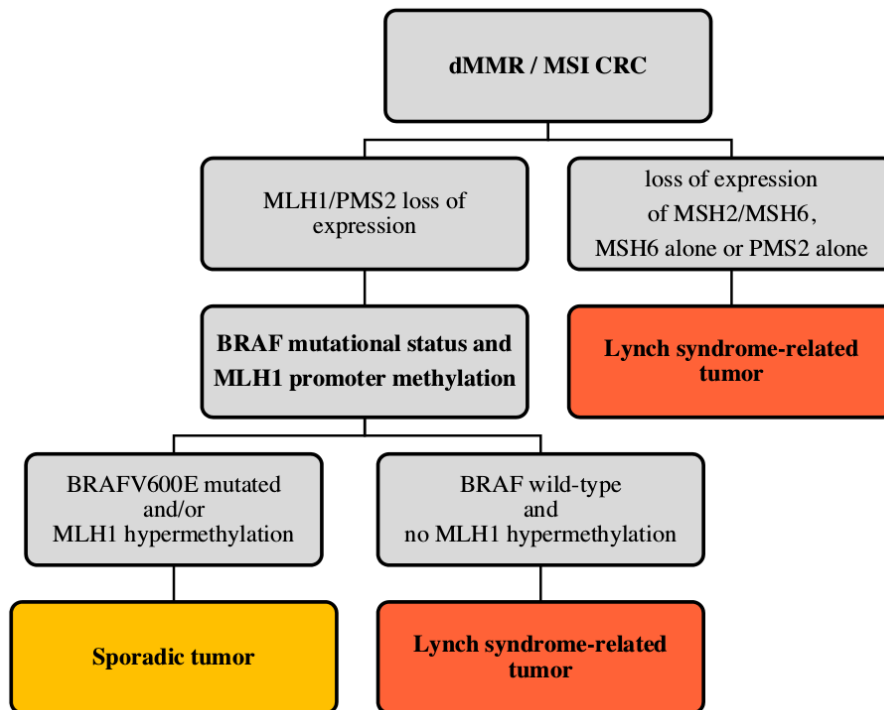


Figure 13 | L’algorithme décisionnel pour la distinction entre les cancers colorectaux sporadiques et ceux associés au spectre du syndrome de Lynch.

Lorsqu’un cancer colorectal est diagnostiqué MSI avec une absence de mutation somatique BRAF V600E et une absence d’hyperméthylation du promoteur du gène MLH1, une consultation oncogénétique est proposée aux patients afin de faire une analyse constitutionnelle des gènes MMR (Colle et al., 2017).

4.2. Les approches émergentes pour le diagnostic MSI

Les approches conventionnelles ont permis de standardiser le diagnostic du phénotype MSI et des formes pathologiques associées. Cependant, ces approches montrent des insuffisances techniques liées à l’interprétation des résultats qui peuvent mener à des discordances, ou à des difficultés logistiques (réalisation à bas-débit). L’émergence des technologies de séquençage haut-débit donne accès à de nouvelles capacités d’analyses à grande échelle, soutenues par le développement d’outils bio-informatiques *ad hoc*, afin d’aider les cliniciens. Grâce aux performances du NGS (profondeur de séquençage, fiabilité, haut-débit), le diagnostic des tumeurs pourrait être réalisé à partir de biopsie liquide (ADN tumoral circulant), peu invasive pour le patient. Plus récemment, la recherche dans le domaine des intelligences artificielles offre de nouvelles opportunités, notamment dans le diagnostic à partir de l’imagerie. À terme (après validation clinique de leur efficacité), ces nouvelles

technologies pourraient remplacer les approches conventionnelles de diagnostic des cancers MSI.

4.2.1. Les outils bio-informatiques

À l'instar de la PCR pentaplex qui définit le statut MSI de tumeurs grâce à des biomarqueurs microsatellitaires, des outils bio-informatiques ont été conçus pour diagnostiquer et évaluer le niveau d'instabilité des tumeurs à partir de données NGS. Parmi eux, *MSISensor* est le seul à avoir été approuvé par l'agence américaine du médicament (FDA, *Food & Drugs Administration*) (Niu et al., 2014). L'algorithme fonctionne à partir de panels de gènes cibles (TGS, *Target Genes Sequencing*) ou d'exome (WES, *Whole Exome Sequencing*) d'échantillons normaux-tumoraux pairés. *MSISensor* calcule un score d'instabilité (ratio du nombre de microsatellites mutés par rapport au nombre de microsatellites analysés) : si le score est supérieur ou égal à 10%, alors la tumeur est diagnostiquée MSI (Middha et al., 2017). Récemment, mon équipe a développé une nouvelle méthodologie nommée *MSICare* pour l'analyse des microsatellites mononucléotidiques de grande taille (≥ 12 répétitions) plus susceptible d'être instable dans un contexte MMR-déficient. De cette manière, il a été démontré que *MSICare* est plus sensible (96-100%) et plus spécifique (100%) que *MSISensor* (Ratovomanana et al., 2021). Des recherches en cours de validation permettront à *MSICare* de s'affranchir de l'utilisation de tissu normaux-tumoraux pairés, comme c'est déjà le cas pour *MSISensor* (Jia et al., 2020). Ces améliorations donnent l'opportunité de diagnostiquer des tumeurs en cas d'absence du tissu normal pairé.

Enfin, la FDA a autorisé l'utilisation des panels de gènes *FoundationOne CDx* et *MSK-IMPACT* pour diagnostiquer les cancers MSI (Cheng et al., 2015; Marabelle et al., 2020). Ces panels constitués de quelques centaines de gènes (contrairement au WES qui en contient 20 000) permettent de minimiser les coûts du diagnostic et d'avoir des rendements à très haut-débit (plusieurs centaines de patients séquencés simultanément). Les panels de gènes peuvent également contenir des marqueurs pronostic de la réponse aux traitements comme BRAF/KRAF.

4.2.2. ADN tumoral circulant

Le diagnostic à partir de biopsie solide est très efficace (lorsque la tumeur est détectable macroscopiquement par coloscopie) mais il est aussi très invasif pour les patients. L'utilisation de l'ADN tumoral circulant (ctDNA, *Circulating tumoral DNA*) présent dans le sang est un enjeu majeur pour améliorer la prise en charge des patients atteints de cancers MSI comme MSS (Gilson et al., 2021; Perakis et al., 2017; Thierry et al., 2017). Cependant, l'exploitation de ctDNA est bien plus complexe que les biopsies solides, en raison de la très faible quantité d'ADN tumoral dans le sang ainsi qu'à sa mauvaise qualité (très fragmentée par les nucléases) (Diehl et al., 2008; Rapisuwon et al., 2016). Les progrès technologiques du NGS et notamment l'usage d'identifiants uniques de molécules d'ADN (UMI, *Unique Molecular Identifier*) permet de séquencer des échantillons dont le matériel est disponible en très faible quantité et facilite l'identification de mutations rares *via* un séquençage « ultra-profond » (5 000 X - 10 000 X) (Nikanjam et al., 2022).

Deux études sur des petits effectifs de patients ont montré une bonne concordance dans le diagnostic du statut MSI entre les approches conventionnelles et le diagnostic *via* une approche à partir de ctDNA. La validation de ces résultats sur une plus grande cohorte de patients pour lesquels une biopsie liquide est disponible, pourrait valider l'utilisation des ctDNA en routine clinique pour le diagnostic des cancers MSI (Silveira et al., 2020; Willis et al., 2019).

Une meilleure compréhension du processus tumoral MSI, pourrait permettre d'identifier des mutations *drivers* de la tumorigenèse et donc de nouveaux biomarqueurs génomiques pour améliorer le diagnostic ainsi que de suggérer des facteurs prédictifs de la progression tumorale (Chen and Zhao, 2019; Herberts et al., 2022).

4.2.3. L'imagerie et l'intelligence artificielle pour diagnostiquer les tumeurs MSI

Une étude publiée dans le journal *Nature Medecine* propose une nouvelle approche diagnostique pour distinguer les cancers gastrointestinaux de type MSI et MSS basée sur l'analyse par intelligence artificielle (AI, *Artificial Intelligence*) de lames histologiques marquées en Hématoxyline et Éosine (HE) (Kather et al., 2019). Cette nouvelle méthode utilise des algorithmes d'apprentissage en profondeur (*deep learning*) avec un modèle de réseau de neurones convolutifs (CNN, *Convolution Neuronal Network*). Le principe est d'entraîner le

réseau avec une grande quantité d'images et donc d'informations afin que l'AI définisse des caractéristiques propres aux deux phénotypes moléculaires MSI et MSS pour les différencier (Kavitha et al., 2022). A ce stade les critères de distinction proposés sont la présence de régions de cellules dédifférenciées et d'infiltration lymphocytaire, propre au MSI, alors que les tumeurs MSS sont caractérisées par des régions glandulaires avec de la nécrose et des plages de mucus extracellulaire (Figure 14). D'autres travaux ont confirmé cette approche en utilisant les CNN dans des localisations de cancers colorectaux et gastriques (Echle et al., 2022, 2020; Yamashita et al., 2021).

Ces algorithmes montrent une sensibilité (91-95%) et une spécificité (67-77%) moindre par rapport aux approches standards ; en cause, un manque de puissance d'apprentissage dû à une quantité insuffisante d'images/informations disponibles, qui devra être solutionné afin d'utiliser l'AI en routine clinique (Cao et al., 2020; Echle et al., 2020). À terme, ce type d'approche permettra de présélectionner les patients probablement MSI afin de confirmer le diagnostic par les méthodes conventionnelles de PCR-IHC (Echle et al., 2021). Des recherches complémentaires sont en cours pour entraîner l'AI à différencier les cancers MSI sporadiques des formes héréditaires du LS (Alam et al., 2022).

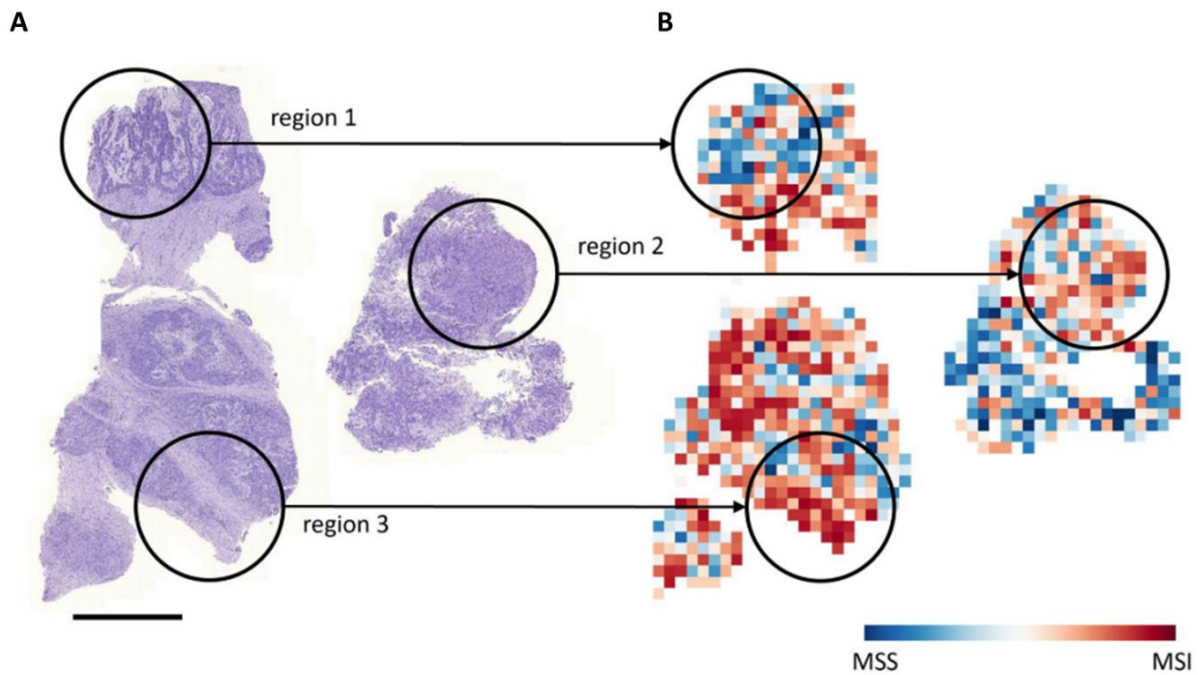


Figure 14 | La détermination du statut MSI grâce à l'imagerie et l'intelligence artificielle. (A) Image histologique d'un cancer colique dont le statut MSI doit être déterminé. (B) L'identification de trois régions suffit à l'intelligence artificielle pour diagnostiquer le statut MSI des tumeurs. La région 1 est une région glandulaire nécrosée présentant du mucus extracellulaire, caractéristique des cancers MSS. Les régions 2 et 3, respectivement une région solide dédifférenciée et une région fortement infiltrée en cellules immunitaires, prédisent le phénotype MSI. Ensemble, ces exemples représentatifs montrent que des morphologies différentes suscitent des prédictions différentes et que ces prédictions peuvent être ramenées à des modèles compréhensibles pour l'homme (Kather et al., 2019).

5. Les traitements des cancers colorectaux MSI

Les différents protocoles thérapeutiques des patients atteints de CRC, sont utilisés seuls ou en combinaison (mono/combo thérapie), selon les stades de progression du cancer défini par la classification TNM (T : tumeur primitive, N : ganglions lymphatiques, M : métastatique) (Flecchia et al., 2022; Kavun et al., 2023; Lecomte et al., 2021). Cette classification se base sur des critères anatomopathologiques comprenant l'envahissement de la paroi intestinale (de la sous-muqueuse à la séreuse, T), l'envahissement des ganglions régionaux (métastase ganglionnaire, N) et la dissémination de cellules tumorales primitives à un ou plusieurs organes (tumeur métastatique à distance, M) (Figure 15).

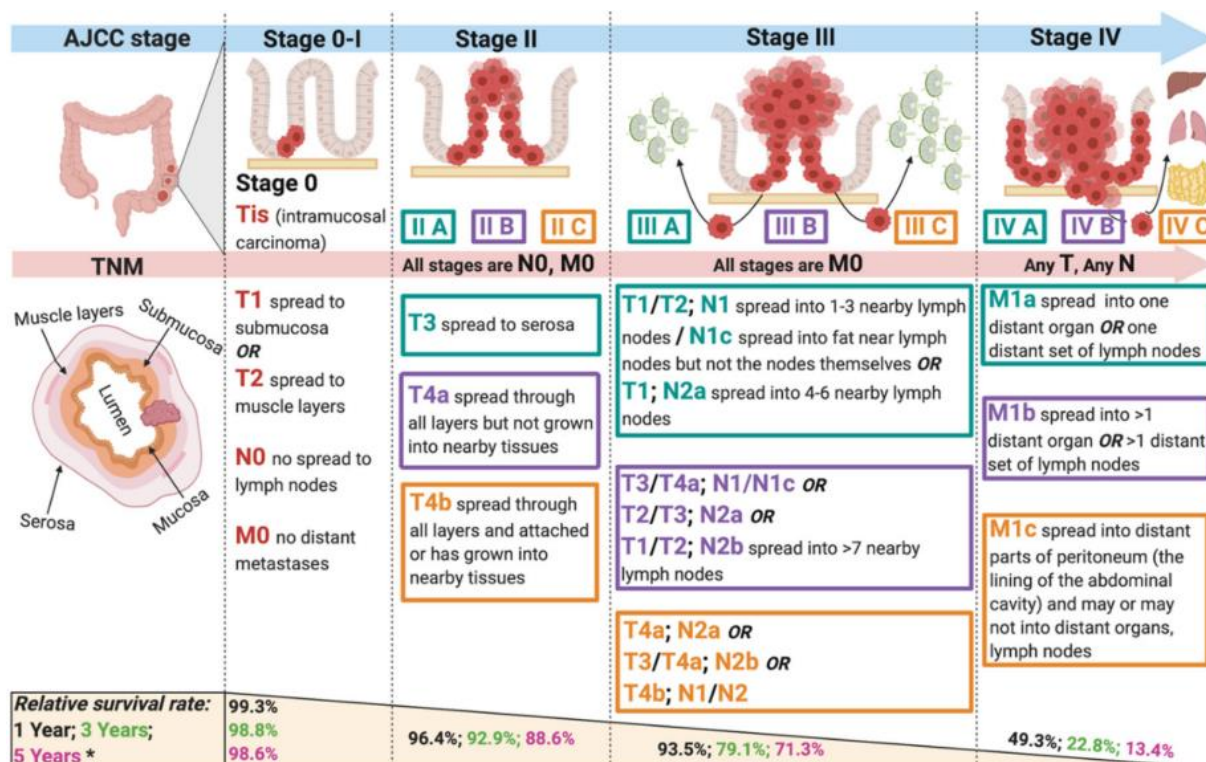


Figure 15 | La classification TNM des cancers colorectaux.

La taille de la tumeur, la croissance dans les ganglions lymphatiques, l'invasion localisée du tissu et les métastases à distance dans les organes et/ou les tissus définissent les stades d'avancées du cancer. Le taux de survie des patients diminue considérablement lorsque le cancer colorectal passe du stade 0 au stade 4 (Shek et al., 2021).

5.1. Les protocoles thérapeutiques conventionnels

De nos jours, le traitement par chirurgie de la tumeur est l'approche la plus efficace et la plus couramment pratiquée dans les cancers colorectaux MSI. Elle peut être complétée par des traitements médicamenteux néo-adjuvant ou adjuvant (avant ou après la chirurgie) de chimiothérapie ou d'immunothérapie (André et al., 2015; Le et al., 2015).

5.1.1. La thérapie chirurgicale

La résection chirurgicale consiste à retirer la tumeur de l'organe (exérèse du cancer). Elle impose une marge proximale et distale d'au moins 5 cm par rapport à la tumeur (limite d'exérèse), ainsi que le curage des ganglions régionaux tout en maintenant la vascularisation des segments coliques préservés (anastomosés). De cette manière, les chirurgiens peuvent pratiquer une hémicolectomie (droite, gauche, transverse ou sigmoïde) selon la localisation de la tumeur, ou une colectomie totale lorsqu'il y a plusieurs tumeurs ou chez les patients atteints par le syndrome de Lynch (Lecomte et al., 2021).

Pour les stades I et II de bas grade (tumeurs locales sans métastase ganglionnaire), la chirurgie seule est à privilégier, alors que pour les stades II de haut grade et les stades III, elle est combinée avec une chimiothérapie adjuvante (Flecchia et al., 2022; Lecomte et al., 2021). Cette association thérapeutique permet d'éviter une rechute cancéreuse *via* l'élimination des cellules tumorales persistante par la chimiothérapie (secondaire à la chirurgie). Pour les stades IV (métastatiques), l'exérèse de la tumeur locale est recommandée combinée avec une chimiothérapie adjuvante, ou néo-adjuvante afin de réduire la taille de la tumeur et ainsi faciliter sa résection (Lecomte et al., 2021; Morton et al., 2023).

5.1.2. La chimiothérapie

La prise en charge thérapeutique par chimiothérapie est identique pour les CRC de phénotype MSI comme MSS. Le protocole FOLFOX, une combinaison de 5-fluorouracile (5-FU), d'oxaliplatine (OXA) et de leucovorine (LV), est la chimiothérapie de référence utilisée en adjuvant de la chirurgie pour les stades II et III (André et al., 2015). Ce traitement induit la mort des cellules tumorales par différents mécanismes. Le 5-FU, une molécule de la classe des médicaments antimétabolites, inhibe la synthèse de la thymidine et donc la biosynthèse de l'ADN et de l'ARN, ce qui provoque l'arrêt de la fourche de réplication et de la transcription *via* la formation d'adduits (Hammond et al., 2016). La leucovorine permet la synthèse de la coenzyme CH₂THF (5-10 méthylène tétrahydrofolate) afin de stabiliser et d'améliorer les effets de cytotoxicité du 5-FU (Longley et al., 2003). L'oxaliplatine, un médicament de la classe des agents alkylants, induit la formation de ponts inter- et intra-brins entre deux guanines adjacentes, ce qui entraîne également une inhibition de la réplication et de la transcription et la formation de cassures doubles brins (Hector et al., 2001).

Le protocole néo-adjuvant FOLFIRI, une combinaison de 5-FU, de leucovorine et d'irinotécan (IRI), est privilégié pour les stades III et IV (Ychou et al., 2009). L'irinotécan, un analogue de la camptothécine, inhibe l'activité de la topoisomérase de type I *via* la formation d'un lien covalent entre cet enzyme et l'ADN ; cela provoque une cassure simple brin puis double brin après le passage de la fourche de réplication, bloquée au niveau de ces adduits, et ainsi la mort des cellules par apoptose (Hammond et al., 2016). Des traitements ciblés peuvent être inclus au protocole thérapeutique, seule ou en combinaison avec la chimiothérapie, pour traiter les mCRC. Le cetuximab et le bevacizumab sont des anticorps monoclonaux qui ciblent

respectivement les facteurs de croissance EGFR (*Epidermal Growth Factor Receptor*) et VEGF (*Vascular Endothelial Growth Factor*) (Cunningham et al., 2004; Kabbinavar et al., 2003).

Des études ont montré que les tumeurs MSI coliques semblent être résistantes au 5-FU avec un effet délétère pour les patients atteints dont la survie globale diminue par rapport au groupe de patients atteints par un CRC MSS (Ribic et al., 2003; Des Guetz et al., 2010; Popat et al., 2005). La chimiosensibilité des cellules tumorales au 5-FU est dépendante de l'activité du système MMR, ce qui pourrait expliquer la chimiorésistance des tumeurs MMR-déficientes (Arnold et al., 2003; Carethers et al., 1999; Meyers et al., 2001; Tajima et al., 2004). De manière intéressante, une analyse comparative intra-MSI a montré que cette résistance au 5-FU se limite aux cancers MSI sporadiques puisque qu'on observe une meilleure survie sans récurrence (DFS, *Disease Free Survival*) pour les CRC MSI associés au LS traités par 5-FU (Sinicrope et al., 2011). L'ajout de l'OXA au 5-FU (FOLFOX) pour traiter les CRC MSI sporadiques a apporté un effet thérapeutique bénéfique, étant donné que le mécanisme d'action de l'OXA est indépendant du MMR (André et al., 2015; Fink et al., 1996).

Également, les patients opérés d'une tumeur colique de stade II ou III avec une mutation du gène PI3KCA (*phosphatidylinositol-4,5-bisphosphonate-3-kinase*) peuvent recevoir un traitement de chimio-prévention à base d'aspirine, afin de prévenir une rechute et ainsi augmenter la PFS et l'OS des patients (Lecomte et al., 2021). L'aspirine inhibe la voie PI3K-Akt dans laquelle la protéine PI3KCA (et sa mutation tumorale activatrice) est impliquée (Kothari et al., 2015).

Enfin, les études CAPP2 et 3 ont montré que la prise régulière d'aspirine – chez des patients atteints par le LS – sur le long terme (100 à 600 mg par jour, pendant au moins 2 ans) semble induire une réduction du risque de développer un CRC de 63%. L'aspirine agirait en bloquant la production de l'onco-métabolite 2-hydroxyglutarate. Bien que des efforts considérables aient été consacrés à l'étude de l'anabolisme et du catabolisme de la 2-hydroxyglutarate, la machinerie moléculaire qui détecte et régule son métabolisme n'a pas été identifiée jusqu'à présent (Colvin et al., 2016). Cependant, la prise régulière d'aspirine n'est pas, à ce jour, recommandée car elle peut entraîner de graves effets secondaires comme des hémorragies digestives qui peuvent être fatales pour les patients (Burn et al., 2013, 2011).

5.1.3. L'immunothérapie anti-ICK (ou ICI)

Comme je l'ai décrit dans le chapitre 1 partie 5, les tumeurs MSI sont caractérisées par un infiltrat lymphocytaire dense associé à leur phénotype hypermutateur très immunogène (Fridman et al., 2012; Kloor and Von Knebel Doeberitz, 2016). Afin d'échapper à la surveillance, les cellules tumorales surexpriment des points de contrôles inhibiteurs de la réponse immunitaire anti-tumorale tels que CTLA-4 ou PD-1 (*Programmed Death-1 receptor*). L'immunothérapie anti-ICK consiste à bloquer ces récepteurs et/ou ligands *via* des anticorps monoclonaux (ICI, *Immune Checkpoint Inhibitor*).

En 2015, une étude clinique a montré un taux de réponse positive élevé (survie sans progression et survie globale) dans les CRC MSI de stade IV traités par le pembrolizumab (monothérapie), un anticorps monoclonal ciblant le récepteur PD-1, contrairement au CRC MSS qui ne sont pas répondeur (Figure 16) (Le et al., 2015). Cela s'explique notamment par la capacité des ICI à réactiver les CTL déjà infiltrés dans les tumeurs MSI, ce qui n'est pas le cas des tumeurs MSS (Colle et al., 2017; Fridman et al., 2012). Depuis, de nombreux nouveaux essais cliniques ont montré l'efficacité de l'immunothérapie mono- ou bi-thérapie (anti-CTLA-4 ou anti-PD-1 seul, ou en combinaison) dans les CRC MSI au stade métastatique, en 2^{ème} – après chimiothérapie – puis en 1^{ère} ligne de traitement (André et al., 2022, 2020; Diaz et al., 2022).

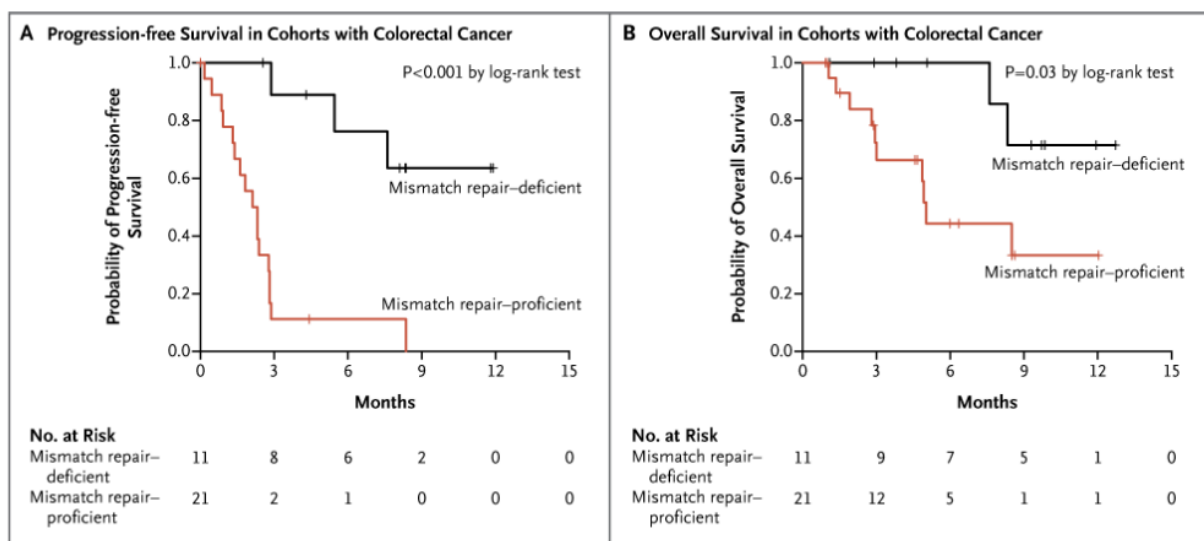


Figure 16 | La réponse à l'immunothérapie ICI (pembrolizumab) en fonction du statut MMR. Les courbes de Kaplan-Meier sont présentées dans des cohortes de CCR pour la survie sans progression (A) et pour la survie globale (B) et en fonction du statut MMR, proficient (MSS, rouge) ou déficient (MSI, noir) (Le et al., 2015).

Des travaux récents ont démontré l'efficacité des protocoles ICI néo-adjuvants dans les CRC MSI localisés (stade I-III) en 1^{ère} ou 2^{ème} ligne de traitement. Dans le cadre de l'essai clinique MSKCC, aucun traitement additionnel (chimiothérapie ou chirurgie) n'a été proposé aux patients traités et la réponse clinique a été complète, permettant ainsi d'envisager des stratégies de désescalade chirurgicale. Des modalités de surveillance (endoscopie, biopsies systématiques, imagerie) ont été mises en place afin de prévenir une éventuelle rechute pathologique (Cercek et al., 2022). Ce point est primordial pour les patients atteints du LS où l'on manque de données quant au risque de rechute après immunothérapie (résistance secondaire) ou de développement d'un cancer a posteriori (les patients atteints du LS peuvent avoir plusieurs cancers dans leur vie, indépendamment d'une rémission), potentiellement résistant au traitement ICI (Lecomte et al., 2021).

Au sein du groupe de patients MSI (intra-MSI), certaines tumeurs coliques de stades IV ne répondent pas aux traitements ICI. Cette résistance s'explique dans certains cas par l'inclusion à tort de patients faux positifs pour le phénotype MSI suite à des erreurs diagnostics (Cohen et al., 2019a). Pour les autres patients dont la tumeur est résistante au traitement anti-ICK, des biomarqueurs prédictifs tels que la charge mutationnelle et antigénique (Mandal et al., 2019), les mutations inactivatrices du gène $\beta 2M$ (Middha et al., 2019; Sade-Feldman et al., 2017) ou le rôle du microenvironnement stromal et des fibroblastes associés aux tumeurs (Cohen et al., 2019b; Endo et al., 2020) ont été proposés comme outils pronostics. Cependant, ils permettent d'expliquer seulement 10% des phénomènes de résistance à l'immunothérapie ICI (Foote et al., 2021; Ratovomanana et al., 2023). Afin de mieux identifier les patients potentiellement bons répondeurs à l'immunothérapie ICI, d'autres outils pronostics sont en cours de développement. C'est dans ce contexte que mon équipe a très récemment publié une étude qui propose de nouveaux biomarqueurs prédictifs (mutations et variations transcriptionnelles) de la réponse à l'immunothérapie ICI dans les CRC MSI de stade IV. Ces marqueurs sont associés à des signatures épithéliales (ADN) et stromales/fibroblastique (RNA), qui sembleraient avoir un rôle fonctionnel de séquestration des cellules immunitaires qui les empêcheraient d'éliminer les cellules tumorales (Ratovomanana et al., 2023).

5.2. Les approches thérapeutiques innovantes

Les stratégies thérapeutiques de prévention ont pour objectif de réduire le risque de développer un cancer MSI et de permettre une désescalade chirurgicale, essentiellement dans les formes héréditaires associées au syndrome de Lynch, pour le confort des patients (Sei et al., 2023; Syngal et al., 2015). À l’instar des cancers MSI sporadiques, les tumeurs associées au LS montrent une infiltration lymphocytaire importante, et ce déjà à des stades précoces qui suivent l’inactivation du système MMR (Bohaumilitzky et al., 2022). Pour rappel, l’immunogénicité des cancers MSI est associée à leur phénotype hypermutateur et plus exactement à l’accumulation de FSP – conséquence des mutations indels dans les microsatellites – exprimés à la surface des cellules tumorales (voir chapitre 1, partie 2 et 5). L’identification des FSP pourrait permettre de développer une stratégie vaccinale préventive dans les LS soit avec des FSP propres à chaque individu ou bien des FSP fréquemment présentés par le MHC I des tumeurs (Hernandez-Sanchez et al., 2022; Sei et al., 2023). Le groupe de recherche du Professeur Kloor a identifié et combiné trois FSP fréquemment mutés dans les tumeurs de patients atteints par le LS (AIM2 del-1, TAF1B del-1, HT001 del-1). Un essai clinique de phase IIa (16 patients atteints par un CRC MSI, traités pendant 6 mois) a démontré l’efficacité de la réponse immune cellulaire et humorale induite par ces FSP sans effet secondaire sévère notoire (Kloor et al., 2020). D’autres vaccins ont été testés dans des essais cliniques randomisés de phase I et II et ont montré des résultats similaires (Overman et al., 2021). Ces thérapies vaccinales s’apparentent à un traitement curatif pour amplifier une réponse immunitaire spécifique anticancéreuse en combinaison avec une immunothérapie ICI, dans un contexte de LS comme pour les formes sporadiques. Également, des résultats préliminaires obtenus chez la souris soutiennent une stratégie clinique de vaccination préventive dans le cadre du LS, en amont du déclenchement de la tumorigenèse (Gebert et al., 2021). L’amélioration des méthodes de diffusion avec des technologies comme les vaccins ARN, devraient permettre d’utiliser une plus grande gamme de peptides néo-antigéniques afin d’augmenter la réponse immune (Hernandez-Sanchez et al., 2022).

Par ailleurs, la compréhension des mécanismes de surveillance immunitaire dans les cancers MSI associés au LS a permis de développer des protocoles d’immunothérapie préventive (Gambini et al., 2022). Des essais cliniques sont en cours pour évaluer l’efficacité de ce protocole thérapeutique et décrire les potentiels effets secondaires à long terme d’un

traitement par immunothérapie, chez des patients diagnostiqués LS et présentant des lésions adénomateuses à haut risque. En effet, certains chercheurs supposent que la stimulation à très long terme par des immunothérapies dans le LS pourrait induire en conséquence des mécanismes d'échappement immunitaire et donc le développement de tumeurs résistantes à ces traitements (essai clinique numéro NCT04711434 et NCT03631641).

Chapitre 3 – La tumorigenèse MSI colique : de la crypte MMR-déficente à l'adénocarcinome

Afin de préserver son renouvellement continu, l'épithélium colique maintient un équilibre physiologique entre la prolifération, la différenciation et la mort cellulaire. L'altération de la régulation de ces mécanismes au profit de la prolifération cellulaire entraîne la transformation néoplasique des tissus (Carethers and Stoffel, 2015). Ainsi, et comme de nombreux cancers, les cancers MSI coliques sont caractérisés par un processus multi-étapes de transformation tumorale, conséquent à l'instabilité génétique associée à la déficience du système MMR. L'accumulation de mutations dans les cryptes MMR-déficientes (lésion pré-néoplasique) entraîne la transformation en adénome (tumeur bénigne) puis adénocarcinome (tumeur maligne) (Ahadova et al., 2018). Récemment, des travaux ont mis en lumière les phénomènes de sélections mutationnelles qui surviennent au cours de la tumorigenèse MSI, directement associés à l'instabilité microsatellitaire et à la réponse immunitaire anti-tumorale (Ballhausen et al., 2020; Jonchere et al., 2018).

Dans ce chapitre, je décrirai la tumorigenèse MSI colique associée au syndrome de Lynch et sporadique d'un point de vue anatomopathologique et génomique, notamment des cryptes MMR-déficientes l'objet principal d'étude au centre de mon projet de thèse.

1. Description anatomopathologique de la tumorigenèse MSI colique

L'anatomopathologie (également nommée « Anatomie et Cytologie Pathologique ») est une spécialité médicale qui consiste à prélever les tissus des pièces opératoires (ou de biopsies) pour les examiner et les caractériser. Cette discipline est aujourd'hui essentielle pour définir le parcours de soin des patients et d'identifier le phénotype MSI de la tumeur (voir chapitre 2). L'étude histopathologique de la muqueuse colique a été essentielle pour mes travaux de thèse, notamment dans la caractérisation des cryptes MMR-déficientes et l'élaboration du protocole technique développé spécifiquement pour l'étude de ces entités (Article 2).

1.1. L'initiation tumorale : les cryptes MMR-déficientes (syndrome de Lynch)

1.1.1. Généralités du rôle physiologique des cryptes saines dans l'homéostasie colique

Les cryptes de la muqueuse colique sont composées de cellules différenciées nécessaires à l'absorption d'eau et d'électrolytes (entérocytes, cellules de gobelet, cellules entéro-endocrines, cellules en houppes) et de cellules souches essentielles au renouvellement hebdomadaire de l'ensemble des cellules épithéliales (Barker, 2014). Les cellules souches, à la base de la crypte, génèrent les cellules TA (*Transit-Amplifying*) qui sont caractérisées par une prolifération rapide dans la moitié inférieure de la crypte. Elles se différencient ensuite en lignées matures lors de leur migration vers le sommet des cryptes où elles meurent par anoïkis (Figure 17). De ce fait, les cellules des cryptes sont organisées selon leur état de différenciation et on considère que l'ensemble de ces cellules proviennent d'une même population clonale (Barker et al., 2009).

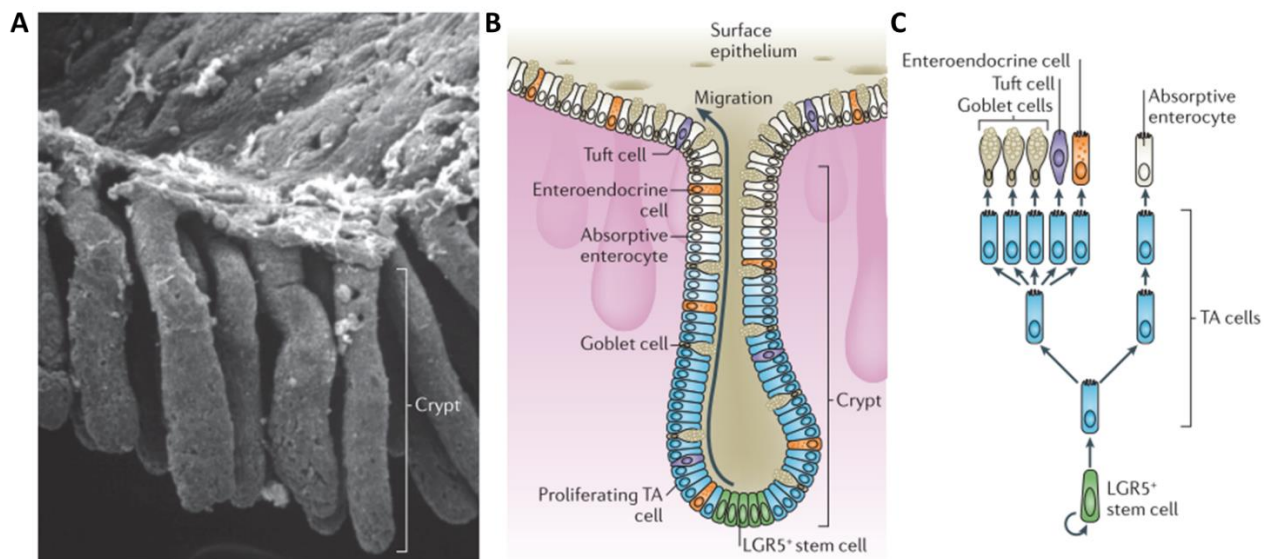


Figure 17 | L'organisation des cryptes de la muqueuse colique.

(A) Visualisation de cryptes de la muqueuse colique par microscopie électronique. (B) Représentation schématique de l'organisation des cellules dans la crypte. (C) Lignage des cellules souches vers les cellules matures (cellules de gobelet, entérocytes, cellules entéro-endocrines et cellules en houppes) dans les cryptes du côlon (Barker, 2014).

1.1.2. Les cryptes MMR-déficientes : mise en évidence et caractéristiques histologiques

Les cryptes MMR-déficientes sont des entités pré-cancéreuses qui ne présentent pas d'anomalie histologique mais une perte d'expression d'une des quatre principales protéines du MMR (Fearon, 2011). Elles ont été décrites pour la première fois dans la muqueuse saine de patients atteints par un cancer colorectal associé au syndrome de Lynch (Kloor et al., 2012). Initialement identifiées dans seulement 25% des patients, les cryptes MMR-déficientes seraient en réalité présentes chez 70-80% des patients (Brand et al., 2020; Thèse du Docteur Sarah Breton, 2021). Malgré des recherches approfondies, aucune crypte MMR-déficiente n'a été observée dans les cancers MSI sporadiques, ce qui suggère que ces entités pré-néoplasiques sont un marqueur histologique spécifique du syndrome de Lynch (Kloor et al., 2012; Pai et al., 2018; Staffa et al., 2015). Également, mon équipe d'accueil a montré que les patients Lynch atteints de tumeurs multiples présentent une densité de cryptes MMR-déficientes plus élevée par rapport aux patients qui ont développé une tumeur unique. Cette observation suggère que l'accumulation de ces lésions pré-néoplasiques soit associée avec un risque plus grand de développer une tumeur (Thèses du Docteur Isabelle Sourouille, 2019 ; Thèse du Docteur Sarah Breton, 2021).

Trois catégories ont été définies pour classifier les cryptes MMR-déficientes selon un critère histologique de regroupement en foyer des cryptes à une localisation donnée. Les foyers monocryptiques comportent une unique crypte MMR-déficiente, alors que les foyers oligocryptiques et polycryptiques sont respectivement composés de 2 à 4 ou de 4 et plus cryptes MMR-déficientes (Figure 18) (Kloor et al., 2012). Malgré ces différences histologiques, aucune corrélation entre la présence de ces entités et le statut clinique des patients ou un autre paramètre n'a pu être constatée (Brand et al., 2020; Pai et al., 2018; Staffa et al., 2015; Thèse du Docteur Sarah Breton, 2021). La présence des cryptes MMR-déficientes est observée dans la muqueuse saine adjacente à la tumeur comme dans la muqueuse saine à distance (à au moins 10 cm de la tumeur) ; il a été observé que la densité de foyers oligo-/poly-cryptiques est plus élevée dans la muqueuse saine à distance (Thèse du Docteur Sarah Breton, 2021). Ce résultat supporte l'hypothèse selon laquelle les cryptes MMR-déficientes dans la muqueuse saine à distance représentent des lésions pré-cancéreuses qui n'ont pas encore évolué vers un stade histologique néoplasique (Ahadova et al., 2018; Thèse du Docteur Sarah Breton, 2021).

Enfin, la visualisation des coupes en coloration HE confirme les descriptions quant au caractère histologiquement sain des cryptes MMR-déficientes, quelle que soit leur catégorie, dans un environnement MMR-proficient sain (Figure 18).

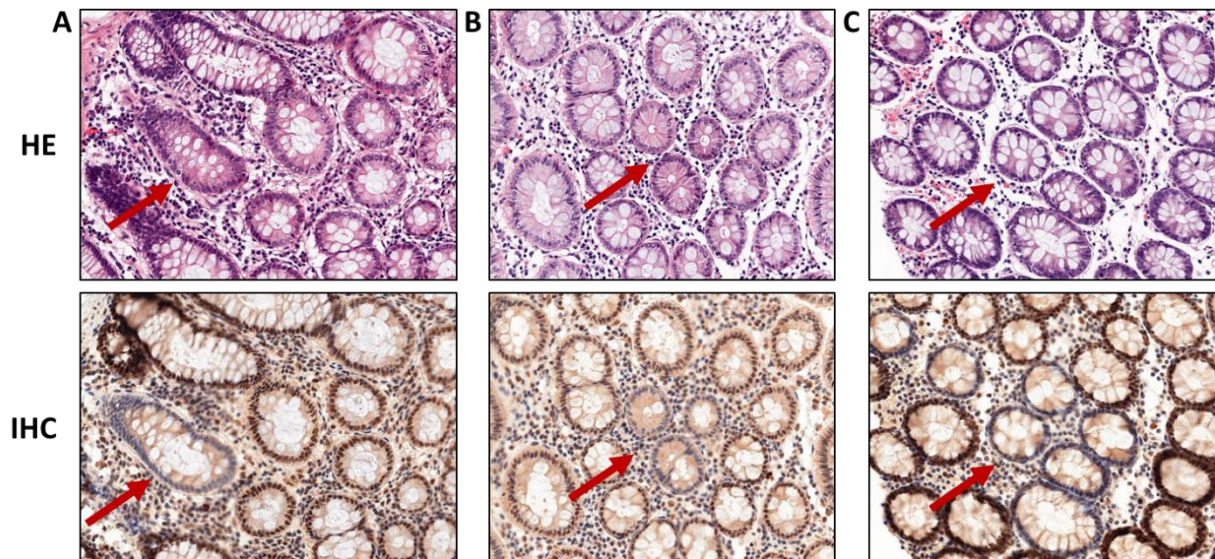


Figure 18 | Les cryptes MMR-déficientes de la muqueuse colique saine.

Les colorations HE (Hématoxyline et Éosine) (**haut**) et les marquages IHC (**bas**) pour la détection de la protéine du système MMR, MSH2, sont réalisées sur des coupes séquentielles. La flèche rouge indique la position d'un foyer monocryptique (**A**), oligocryptique (**B**) et polycryptique (**C**).

Pour maintenir l'homéostasie intestinale, les cellules souches doivent être capables de se renouveler et de générer différents types de cellules épithéliales différenciées (Barker et al., 2007). Les cellules souches peuvent se diviser de manière asymétrique pour générer une cellule souche fille et une cellule destinée à se différencier (division Q), ou de manière symétrique pour générer deux cellules souches filles (division P) ou deux cellules filles destinées à se différencier (division R) (Graham et al., 2011; Snippert et al., 2010; Xie et al., 2000). C'est ce processus qui permet de maintenir une seule population clonale au sein des cryptes saines (Barker et al., 2009). Ainsi, une cellule souche qui a subi une mutation (comme dans le cas d'une mutation perte de fonction d'une des protéines du MMR) transmettra cette mutation aux cellules filles souches comme différenciées. La cellule souche porteuse de la mutation, peut également remplacer les autres cellules souches de la crypte si ces dernières entrent en apoptose ou se répliquent par division symétrique R (Figure 19A et A') : on parle d'expansion clonale ou de conversion monoclonale (Humphries and Wright, 2008; Kim and Shibata, 2002; Nicolas et al., 2007). Ce mécanisme physiologique peut permettre de propager

une mutation par dérive génétique si celle-ci n'apporte pas d'avantage sélectif ou par sélection naturelle dans le cas où la mutation apporterait un avantage sélectif en survenant par exemple dans un oncogène ou gène suppresseur de tumeur ; dans ce dernier cas, la conversion monoclonale de la cellule souche mutée à l'ensemble des cellules souches de la crypte est plus rapide (Humphries and Wright, 2008; Maley et al., 2004).

La division des cryptes (ou *crypt fission*) est un mécanisme essentiel pour l'expansion de la muqueuse colique lors de l'embryogenèse mais aussi plus rarement – une à deux fois dans la vie – à l'âge adulte (Greaves et al., 2006; Langlands et al., 2016; Loeffler et al., 1997). Dans la muqueuse colique saine pré-cancéreuse, ce phénomène a été observé à des fréquences anormalement élevées ce qui semblerait être associé à une altération de la signalisation Wnt qui régule la division des cryptes (Figure 19B-C et B'-C') (Boman et al., 2004; Wasan et al., 1998). De ce fait, des chercheurs ont proposé que la fission des cryptes soit un mécanisme amplificateur de la tumorigenèse colique (van den Brink and Offerhaus, 2007; Wong W-M, Mandir N, Goodlad RA, Wong BC, Garcia SB, Lam SK, 2002).

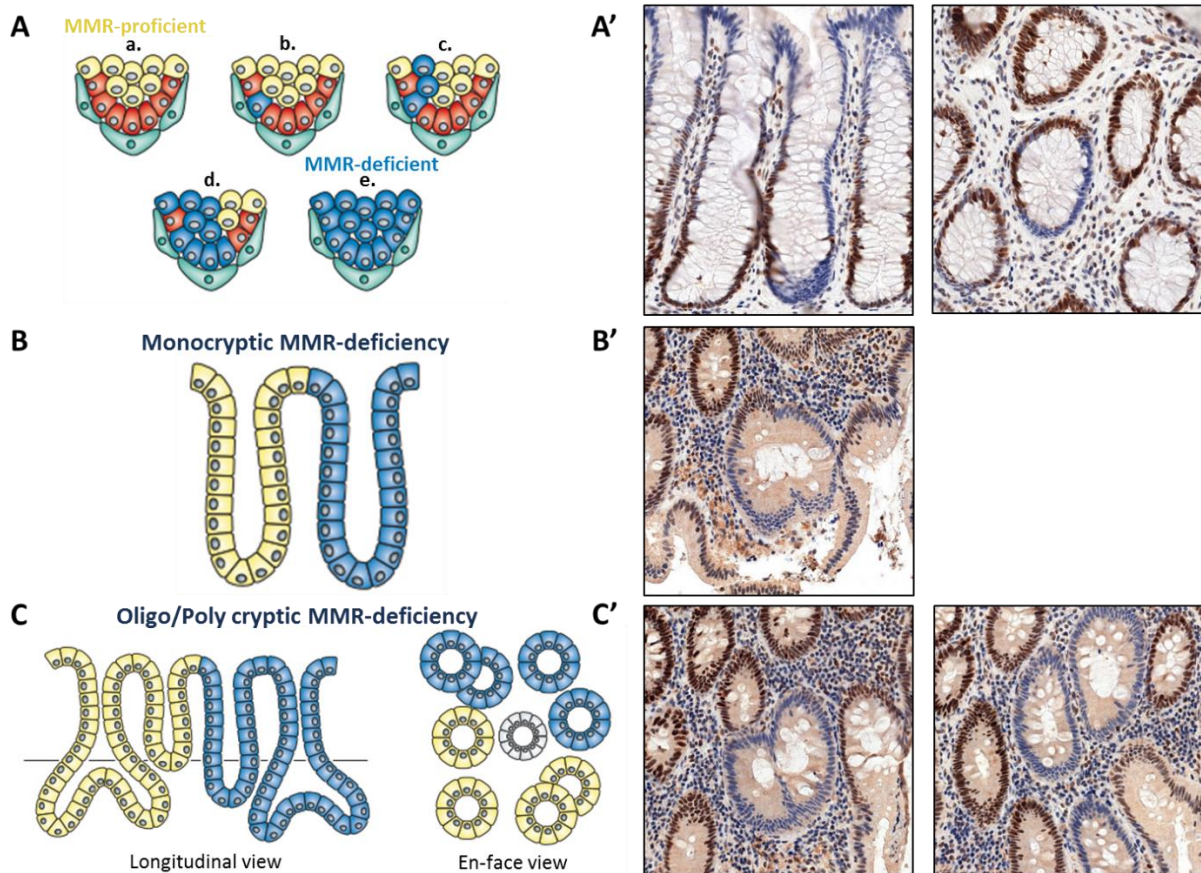


Figure 19 | L'expansion clonale des cellules souches de la niche et la division des cryptes amplifient la tumorigénèse dans un contexte MMR-déficient.

(A) Représentation schématique de l'organisation des cellules souches au fond des cryptes intestinales et description de la dynamique de conversion clonale des cellules souches : **(a.)** niche des cellules souches au fond de la crypte, **(b.)** apparition d'une mutation dans une seule des cellules souches, **(c.)** la mutation se dissémine par division asymétrique, **(d-e.)** *in fine*, la cellule souche clonale mutante devient majoritaire en remplaçant les autres cellules par division symétrique P. **(A')** La crypte observée est partiellement MMR-déficiente (perte focalisée) puisque les cellules souches au fond des cryptes ne sont pas toute déficientes (vue longitudinale) ; après conversion clonale, l'ensemble des cellules de la crypte pourront être MMR-déficientes. **(B-C)** Représentation schématique de la division des cryptes et **(B'-C')** illustration dans la muqueuse colique saine associée : à partir d'une unique crypte MMR-déficiente on observe la formation d'un foyer oligocryptique *via* la fission en deux cryptes indépendantes (d'après Humphries and Wright, 2008).

1.2. La progression tumorale MSI : de l'adénome à l'adénocarcinome

La découverte des cryptes MMR-déficientes a introduit une nouvelle étape précoce à la tumorigenèse colique classique, la séquence adénome-carcinome (Kloor et al., 2012). Comme nous l'avons vu, les cryptes MMR-déficientes sont décrites spécifiquement dans les cancers colorectaux MSI associés au syndrome de Lynch. Un schéma de carcinogenèse à trois options a été proposé pour décrire les différentes étapes : (i) le plus fréquemment le cancer se développe à partir de la transformation de cryptes MMR-déficientes en adénome puis en adénocarcinome, (ii) dans de rares cas la tumorigenèse MSI s'initie à l'étape adénome et progresse vers l'adénocarcinome (séquence adénome-carcinome classique), (iii) dans de très rares cas, les cryptes MMR-déficientes se transforment directement en adénocarcinome sans évolution intermédiaire en adénome (Ahadova et al., 2018).

La séquence adénome-carcinome décrit un schéma classique de progression de la tumorigenèse MSI colique sporadique et associée au syndrome de Lynch (Figure 20). Elle s'initie suite à une hyperprolifération des cellules de la muqueuse colique (polype hyperplasique) qui évolue en adénome – caractérisée par la présence de dysplasie (anomalies histologiques) – puis en adénocarcinome (invasif). Les pathologistes estiment que cette séquence dure en moyenne 10 ans (Carethers and Jung, 2015). Ensuite, les adénocarcinomes peuvent envahir la sous-muqueuse, la musculuse, la sous-séreuse puis la séreuse (différentes couches tissulaires) et s'étendre localement ou se disséminer à distance pour former des métastases (Phelip et al., 2014).

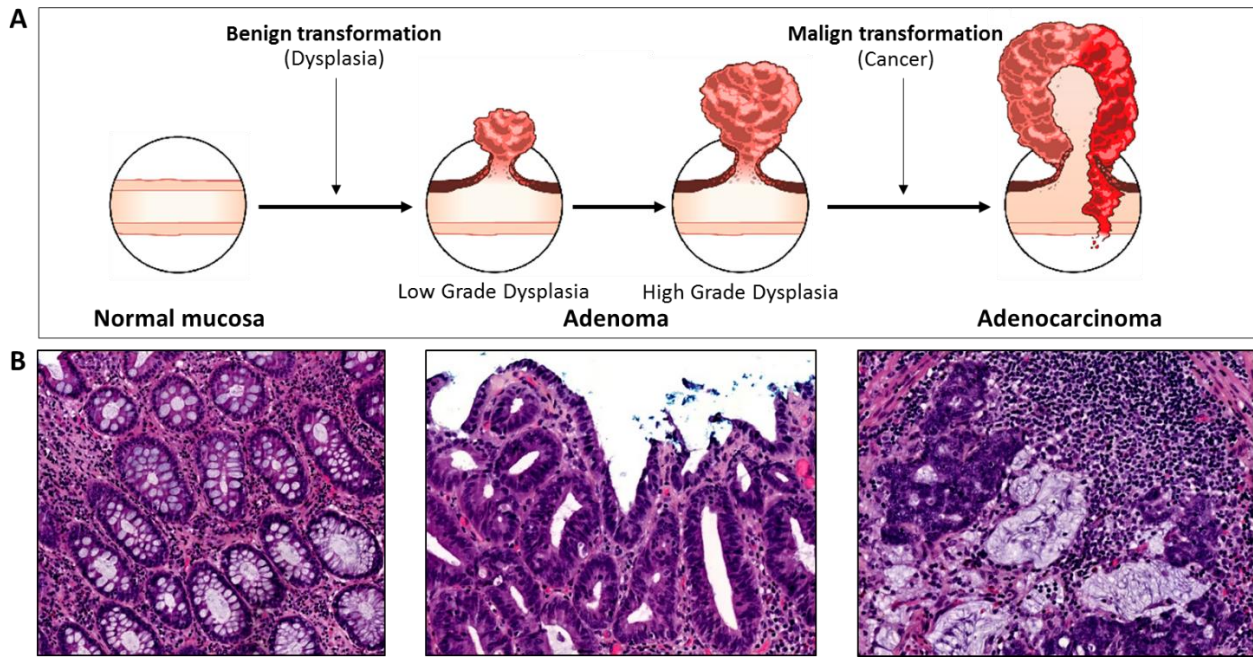


Figure 20 | La séquence adénome-carcinome.

(A) Représentation schématique et (B) visualisation histologique par coloration HE. La séquence s'initie par une transformation des cellules saines en adénome (tumeur bénigne), avec des lésions dysplasiques de bas puis et haut grade. L'accumulation de mutations entraîne la progression en adénocarcinome invasif (tumeur maligne) (Thèse du Docteur Florence Renaud, 2015).

1.2.1. Les adénomes

La terminologie en anatomopathologie est très précise pour décrire les différentes lésions pré-cancéreuses observées. Par exemple, les polypes sont des hyperplasies tissulaires sans dysplasie (déstructuration cellulaire, cytologique et nucléaire, ainsi que des altérations architecturales des cryptes mais sans infiltration de la *lamina propria*) et faisant saillie dans la lumière intestinale (observable par coloscopie), alors que les adénomes sont des polypes qui présentent de la dysplasie (Riddell et al., 1967; Bosman et al., 2010). Il existe deux niveaux de dysplasie : de bas grade ou de haut grade (Figure 21). La dysplasie de bas grade est caractérisée par des anomalies modérées tels que des noyaux basophiles, allongés et pseudostratifiés (Figure 21A et B). La dysplasie de haut grade se différencie par des anomalies architecturales plus importantes, avec des fusions glandulaires (cryptes) et des atypies nucléaires ainsi que des cellules en mitose (Figure 21C et D) (Snover, 2019; Thèse du Docteur Florence Renaud, 2015). L'évolution d'un adénome vers des stades plus avancés de la pathologie, est corrélée au niveau de dysplasie : plus la dysplasie est élevée, plus le risque de progresser en

adénocarcinome est important ; environ 5% des adénomes de haut grade progressent vers un cancer (Stryker et al., 1987; Bosman et al., 2010).

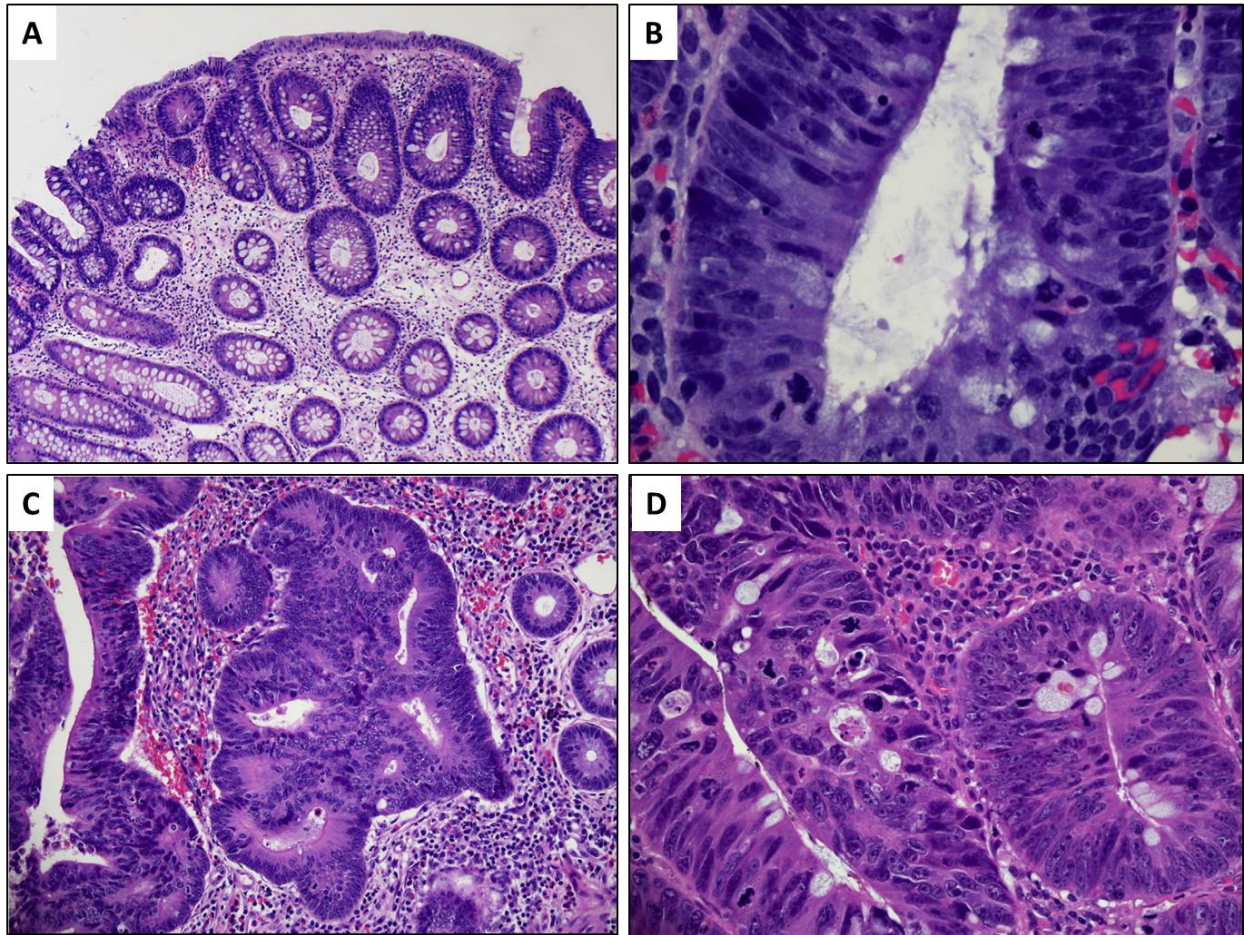


Figure 21 | La classification histologique des adénomes selon leur niveau de dysplasie. (A-B) Adénome de dysplasie de bas grade. (C-D) Adénome de dysplasie de haut-grade (d'après (Snover, 2019).

Dans la majorité des cas (90%), les adénomes coliques sont dits conventionnels (ConA, *Conventional Adenoma*) ; c'est le cas des adénomes qui surviennent dans la séquence adénome-carcinome des CRC CIN (MSS) mais aussi des CRC MSI associés au syndrome de Lynch (Snover, 2019). Des analyses ont montré que le développement des adénomes conventionnels serait corrélé avec la mutation du gène APC (voie de signalisation Wnt) notamment dans les cancers MSS (Powell et al., 1992). Trois sous-types d'adénome conventionnel sont décrits par les pathologistes (Ishii et al., 2011; Bosman et al., 2010): (i) les adénomes tubuleux qui, les plus fréquents (90%) sont caractérisés par des cryptes allongées en forme de tubes (Figure 22A) ; (ii) les adénomes vilieux représentant 10-25% des cas, se distinguent par des projections

villosités digitiformes plus susceptible de se transformer en carcinome (Figure 22B) ; (iii) les adénomes tubulovilleux, les plus rares (5-10%), possèdent des traits qui définissent les deux sous-types précédents (Figure 22C).

D'autres lésions adénomateuses ont été décrites par les pathologistes, comme les adénomes festonnés sessiles (SSA, *Sessile Serrated Adenoma*) associés plutôt à la tumorigenèse MSI sporadique (Figure 22D) (Snover, 2019; Torlakovic et al., 2003). Ces adénomes présentent une ondulation de la lumière intestinale au pourtour des cryptes, d'un aspect « dents de scie », qui serait lié à des anomalies de l'apoptose régulant l'exfoliation cellulaire à la surface des cryptes (Bettington et al., 2013). Les SSA (20-35% de l'ensemble des polypes colorectaux) sont plutôt localisés dans le côlon droit et associés au développement d'un cancer BRAF muté (*i.e.* BRAF V600E, 80% des cas), comme la plupart des CRC MSI sporadiques (Snover, 2019).

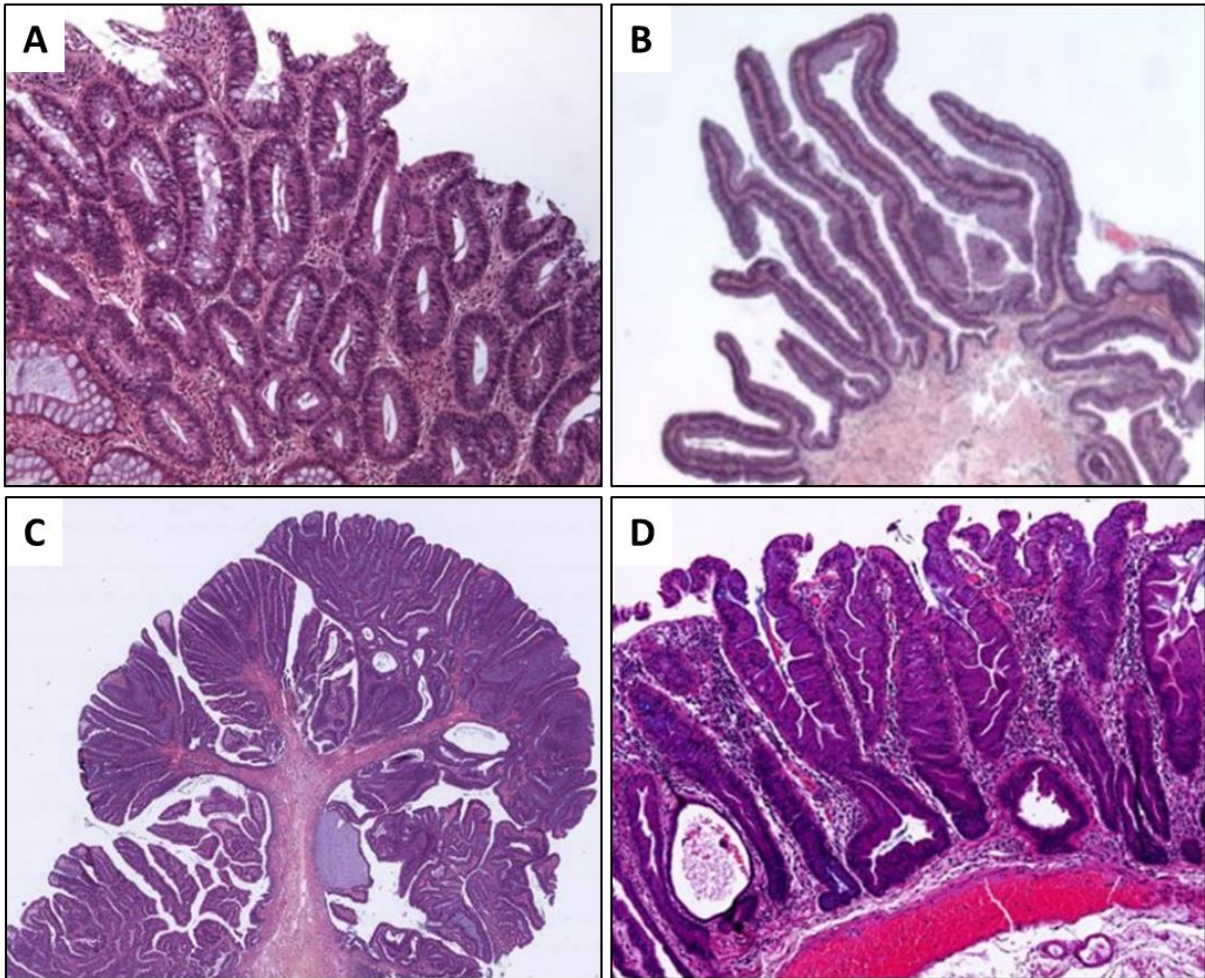


Figure 22 | Lésions adénomateuses conventionnelles et sessiles festonnées.

Les adénomes conventionnels sont classifiés en trois sous-types : **(A)** tubuleux, **(B)** villosus ou **(C)** tubulovillosus. **(D)** Observation d'un adénome festonné sessile, associé à la tumorigenèse MSI sporadique (Thèse du Docteur Florence Renaud, 2015).

1.2.2. Les adénocarcinomes

Les adénocarcinomes se différencient des adénomes selon un critère de franchissement de la musculaire muqueuse (fine couche tissulaire sur laquelle repose l'épithélium intestinale) et une infiltration de la sous-muqueuse. Quatre stades décrivent l'état de progression des adénocarcinomes selon leur niveau d'infiltration en profondeur des tissus et l'envahissement localisé ou à distance (voir chapitre 2) (Wittekind et al., 2009). Les métastases des CRC sont majoritairement localisées dans le foie (75%) et moins fréquemment dans le poumon et le cerveau (Phelip et al., 2014). Également, les adénocarcinomes sont classifiés selon d'autres critères tels que l'état de différenciation des cellules des cryptes coliques (peu, moyen ou bien différenciées) ainsi que le type histologique (Lieberkühnien, mucineux, médullaire ou à cellules indépendantes) (Figure 23). Les adénocarcinomes

mucineux sont caractérisés par une mucisécrétion importante (étendue de mucus, Figure 23A), l'adénocarcinome à cellules indépendantes comporte au moins 50% de cellules indépendantes (Figure 23B) et l'adénocarcinome médullaire est décrit par un état peu différencié et riche en lymphocytes (Figure 23C). Enfin, la majorité des CRC (90%) sont des adénocarcinomes Lieberkühnien bien différenciés (Figure 23D) (Jass, 2007; Bosman et al., 2010)(Jass et al., 1992).

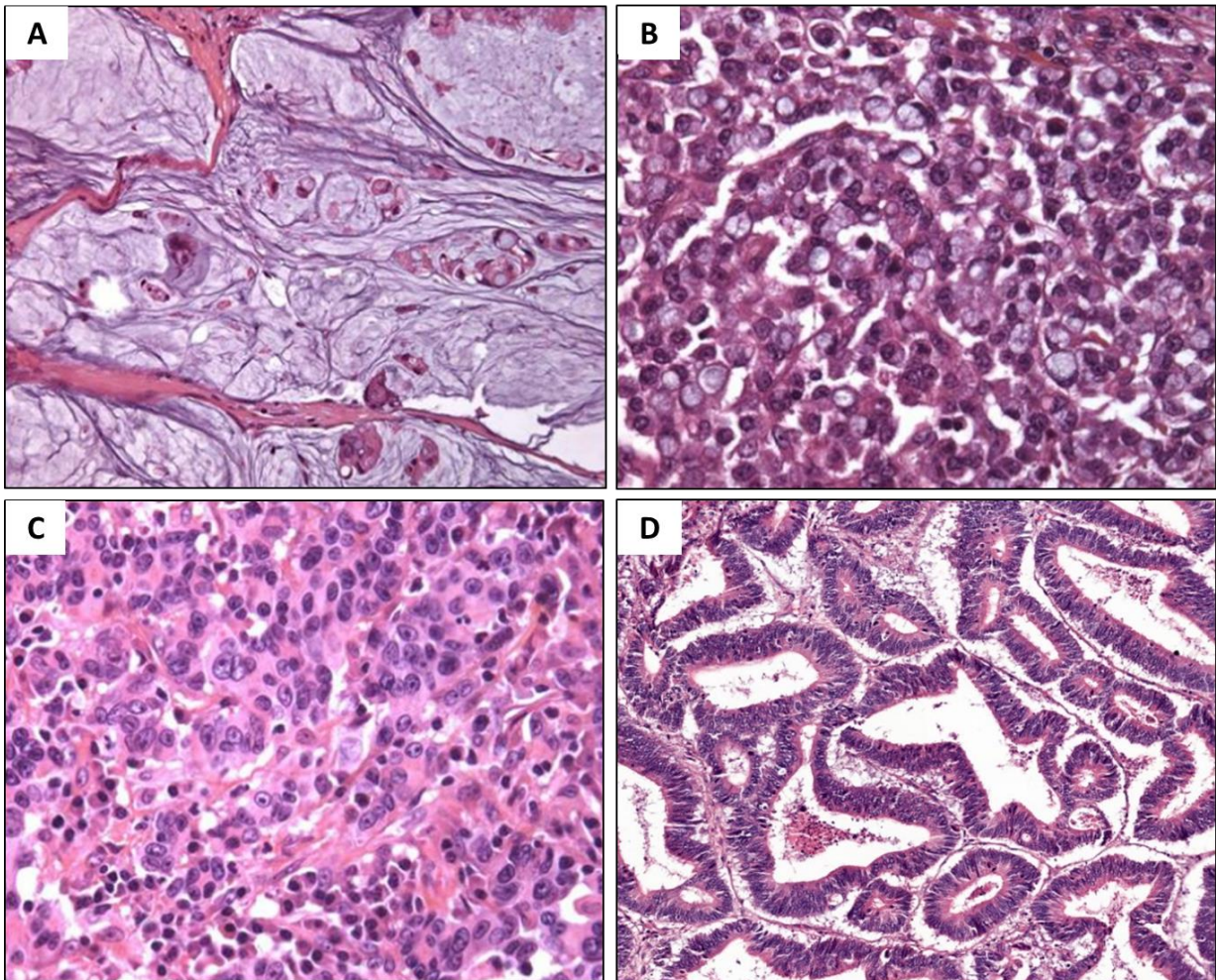


Figure 23 | La classification histologique des adénocarcinomes.

Adénocarcinome (A) mucineux, (B) à cellules indépendantes, (C) médullaire peu différencié et (D) Lieberkühnien bien différencié (Thèse du Docteur Florence Renaud, 2015).

2. Le paysage mutationnel des cancers colorectaux MSI

2.1. Les voies de signalisation de l'homéostasie colique et les principaux gènes mutés associés

Le maintien de l'homéostasie repose sur la communication intercellulaire des cellules de la crypte entre elles ainsi qu'avec les cellules stromales, *via* différentes voies de signalisation majeures telles que BMP/TGF- β , Wnt, EGFR/MAPK, Notch ou encore PI3K-Akt (Figure 24A). Ces voies qui régulent la prolifération, migration, différenciation, adhésion et mort cellulaire, sont fréquemment mutées dans la tumorigenèse colique, d'origine MSI comme MSS ; cependant, les perturbations des voies de signalisation sont différentes (les mutations observées sont spécifiques selon l'origine MSI ou MSS) ce qui distingue les deux types de cancers (Gehart and Clevers, 2019). Dans cette partie, je décris les mutations fréquentes des voies de signalisation de l'homéostasie intestinale, afin de décrire leur dérégulation en association avec les mutations fréquentes dans la tumorigenèse MSI colique.

2.1.1. La voie de signalisation Wnt/ β -caténine

La voie de signalisation Wnt/ β -caténine (*Wingless Int-1*) permet le maintien des propriétés des cellules souches (*stemness*) en favorisant la prolifération au détriment de la différenciation cellulaire (Figure 24B-1) (Pinto and Clevers, 2005).

La voie Wnt/ β -caténine est anormalement activée dans les CRC où le gène *APC* présente une mutation (séquence non-répétée) perte de fonction dans la plupart des cancers MSS (75%) contre seulement 20% des cancers MSI (Laurent-Puig et al., 1999; Markowitz and Bertagnolli, 2009). Pour autant, la voie est très dérégulée dans les cancers MSI par des mutations (microsatellites) des gènes *AXIN2*, *TCF-4*, *DVL1* ou encore *RNF43* (environ 40% des CRC MSI) (Duval and Hamelin, 2002a; Imai and Yamamoto, 2008; Jonchere et al., 2018; Koo et al., 2012; Yamamoto and Imai, 2015). Également, dans les CRC MSI associés au syndrome Lynch, on observe des mutations fréquentes du gène *CTNNB1* qui code pour la β -caténine, ce qui pourrait expliquer le développement d'adénome conventionnel (Jass, 2007; Mirabelli-Primdahl et al., 1999).

2.1.2. La voie de signalisation Notch

La voie de signalisation Notch est impliquée dans la transduction du signal des jonctions cellulaires, notamment dans l'inhibition latérale (régulation de l'expansion du compartiment de prolifération des cryptes) et régule la maintenance des cellules souches (*stemness*, bloque leur différenciation) (Figure 24B-2).

De nombreux microsatellites, localisés dans les gènes qui codent les protéines de la famille ADAM (*A Disintegrin And Metalloprotease*), sont souvent instables. Également, la voie de signalisation *Focal Adhesion*, ainsi que d'autres voies de jonction cellulaire associées à la régulation de la voie Notch sont massivement mutées dans les cancers MSI (Jensen et al., 2000; Lander et al., 2001; Vooijs et al., 2011).

2.1.3. La voie de signalisation BMP/TGF- β

La voie de signalisation BMP (*Bone Morphogenic Protein*) couplée à la voie TGF- β (*Transforming Growth Factor β*) réduit la prolifération (blocage en phase G1) et induit la différenciation cellulaire (Figure 24B-3) (Heldin et al., 1997; Jung et al., 2017).

Cette voie est inactivée dans la moitié des CRC MSS par une mutation (non-répétée) perte de fonction du gène *SMAD4* et/ou *SMAD2/3* qui empêche la transduction du signal (Markowitz and Bertagnolli, 2009). Dans les CRC MSI, la voie BMP/TGF- β est principalement inactive suite à des mutations (microsatellitaires) des gènes *TGFBR2* et/ou *ACVR2A* dans environ 90% des cas ; le gène *BMPR2* est également muté dans 50% des CRC MSI (Jonchere et al., 2018).

2.1.4. La voie de signalisation EGFR/MAPK

La voie de signalisation EGFR/MAPK, associée à la voie PI3K-Akt, est impliquée dans la prolifération, la survie (résistance à l'apoptose), la migration cellulaire ainsi que l'angiogenèse. Les voies MAPK et PI3K-Akt sont des cascades de signalisation associées à la signalisation EGFR, qui participent à la maintenance des propriétés de *stemness* des cellules souches des cryptes (Figure 24B-4) (Glaviano et al., 2023; Samuels and Ericson, 2006).

La voie EGFR/MAPK est aussi altérée dans les CRC MSI avec des mutations qui entraînent une activation constitutive de la voie et donc une hyperprolifération. Les mutations

(non-répétées) des gènes *KRAS* et *NRAS* sont présentes dans 10-20% des CRC MSI, alors que celles de *BRAF*, associées au CRC MSI sporadique, sont retrouvées chez 50-60% des patients (Markowitz and Bertagnolli, 2009; Samowitz et al., 2007; Seppälä et al., 2015). Dans cette même voie, le gène *PIK3CA* qui code pour la sous-unité catalytique de PI3K est muté dans les CRC (10-20%). Les mutations des gènes *KRAS* et *NRAS* sont également retrouvées dans 30-40% des CRC MSS contre seulement 5% pour *BRAF* (Samuels and Ericson, 2006; Velho et al., 2010; Whitehall et al., 2012). Le gène *PTEN* est également muté (microsatellites) dans certains CRC MSI (Jonchere et al., 2018; Richmond et al., 2015).

2.1.5. Les communications entre les voies de signalisation

Les cryptes sont organisées en compartiments (cellules souches, progénitrices et différenciées) dont la prolifération et la différenciation sont principalement contrôlées par un gradient d'expression de BMP et de Wnt ; ces deux voies dites antagonistes (Figure 24A). Les voies Wnt et Notch, pro-prolifération et anti-différenciation, sont essentielles pour maintenir la *stemness*. La surexpression des protéines activatrices de Wnt ou Notch a pour conséquence une hyperprolifération des cellules de la crypte alors que leur inhibition entraîne perte des compartiments (Fre et al., 2009; van Es et al., 2005). Au contraire, la voie BMP/TGF- β , pro-différenciation et anti-prolifération, permet de contrôler le devenir des cellules progénitrices et de maintenir l'état différencié des cellules du compartiment supérieur des cryptes. Les mutations inhibitrices de cette voie ont pour conséquence une augmentation de la prolifération des cellules et une perte de leur état différencié – ou dé-différenciation – (Itatani et al., 2019; Jung et al., 2017). Enfin, la voie EGF/MAPK/PI3K-Akt impliquée dans la prolifération et survie cellulaire est notamment active dans le compartiment des cellules souches. La dérégulation de cette voie peut induire une activation constitutive et ainsi l'hyperprolifération des cellules (Samuels and Ericson, 2006; Whitehall et al., 2012). L'ensemble de ces travaux permettent de comprendre le rôle clé des cellules souches dans l'initiation de la carcinogenèse colique (Barker et al., 2009; Li and Martin, 2016).

L'ensemble de ces mutations altère les voies de régulation du maintien des caractéristiques propres aux cellules souches et de l'organisation histologique des cryptes coliques. En effet, la plupart de ces mutations ont pour conséquence une hyperprolifération des cellules souches ainsi qu'une dé-différenciation des cellules des autres compartiments

(Barker et al., 2009). Également, la plupart des mutations surviennent dans des gènes qui codent les protéines responsables de la réception et/ou de la transduction du signal ; de cette manière une ou deux mutations suffisent à déréguler la voie. Ces altérations sont responsables d'une déstructuration cellulaire, cytotogique et nucléaire, ainsi que des altérations architecturales des cryptes (*i.e.* dysplasie), observées dans les adénomes de la séquence adénome-carcinome (Riddell et al., 1967; Bosman et al., 2010).

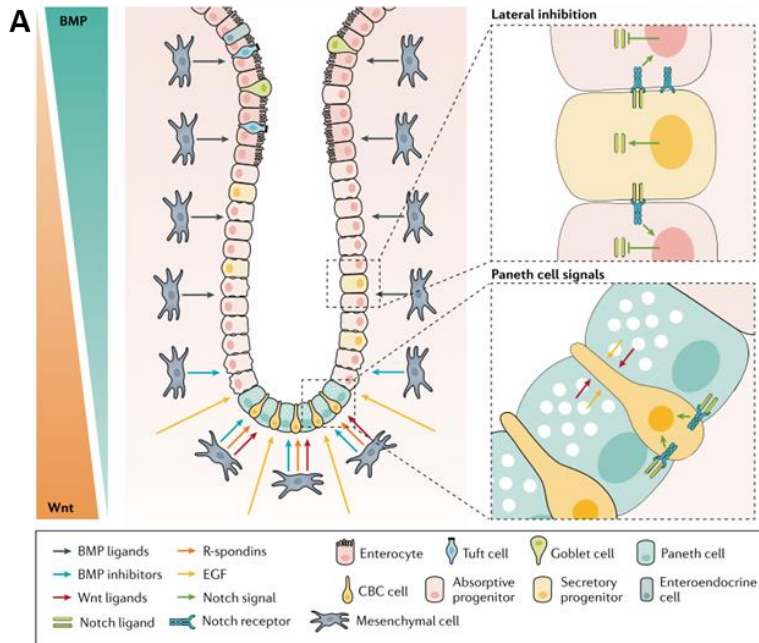
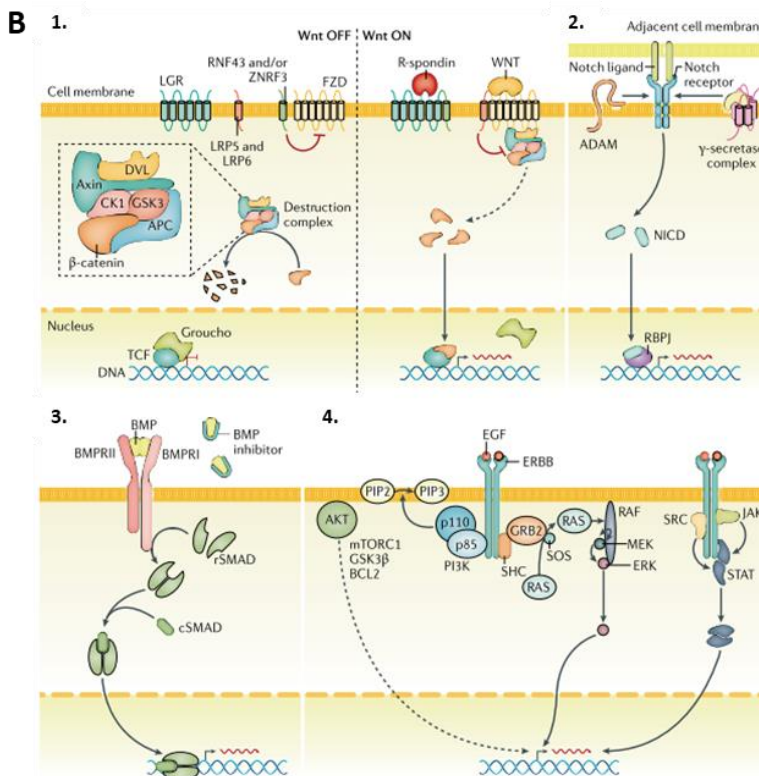


Figure 24 | Les voies de signalisation de l'homéostasie intestinale fréquemment mutées dans la carcinogénèse colorectale.

(A) Les cryptes intestinales sont organisées en trois compartiments où chaque voie de signalisation a une fonction spécifique. La niche est constituée des cellules souches qui peuvent se différencier en cellules progénitrices elles-mêmes précursseurs des cellules différenciées qui ont une fonction précise (entérocytes, cellules en houppes, gobelet ou entéro-endocrines).



(B) Les voies de signalisation de l'homéostasie intestinale sont impliquées dans la tumorigénèse colique.

- (1.)** Wnt/ β -caténine,
 - (2.)** Notch,
 - (3.)** BMP/TGF- β ,
 - (4.)** EGFR/MAPK.
- (Gehart and Clevers, 2019)

Enfin, d'autres mutations sur des séquences répétées sont retrouvées dans des gènes impliqués dans différentes voies de signalisation régulant l'apoptose (tels que BAX, CASP5, BCL10, FAS), la réplication et la réparation de l'ADN (tels que MSH3, MSH6, MBD4, MLH3, CHK1, ATR, RAD50, POLD3, BLM, DNA2) ou encore la transcription (tels que TAF1B, INO80E, FAM111B, SETD1B, CREBBP) (Duval and Hamelin, 2002b; Imai and Yamamoto, 2008; Yamamoto and Imai, 2015). Le gène suppresseur de tumeur *TP53* qui code pour la protéine p53 régulatrice du cycle cellulaire, de l'apoptose, de la réponse aux stress génotoxiques, est muté dans environ 30% des CRC MSI (Leslie et al., 2002; Rampino et al., 1997).

Malgré les nombreux travaux qui décrivent ces mutations de gènes cibles dans un contexte MMR déficient, l'identification des mutations *drivers* (mutations initiatrices et/ou nécessaire pour la progression tumorale), reste encore à élucider dans ce type de cancer caractérisé par une charge mutationnelle 100 fois plus importantes que dans les CRC de type MSS.

2.2 Les modèles de sélection mutationnelle dans les cancers MSI

Les CRC MSI, sporadiques comme associés au syndrome de Lynch, sont caractérisés par une hétérogénéité mutationnelle importante qui s'explique par le grand nombre de mutations consécutives à l'instabilité génétique induite par l'inactivation du système MMR (phénotype hypermutateur). Les cancers MSI sont caractérisés par une forte hétérogénéité intra-tumorale : au sein d'une tumeur, plusieurs populations sous-clonales – aux génotypes différents (certaines mutations sont partagées et d'autres sont spécifiques) – co-existent. La fréquence des allèles variants (VAF, *Variant Allele Frequency*) est un outil bio-informatique très utile pour estimer l'hétérogénéité intra-tumorale des mutations. La VAF permet d'identifier les populations de cellules cancéreuses dominantes au sein de la tumeur et ainsi de suivre la dynamique sous clonale d'une mutation dans un échantillon. Cet outil est également efficace pour évaluer l'état homozygote ou hétérozygote d'une mutation (Boscolo Bielo et al., 2023). Cependant, la VAF est peu utilisée lors des analyses bio-informatiques de l'instabilité microsatellites, étant donné le manque de standardisation dans sa mise œuvre.

Également, les cancers MSI présentent une hétérogénéité inter-patient importante, mesurée par la fréquence mutationnelle (MF, *Mutational Frequency*) : il s'agit du ratio du

nombre de patients qui présentent une mutation donnée par rapport au nombre total de patients analysés. La MF permet d'identifier facilement les mutations susceptibles d'être partagées par un grand nombre de patients ainsi que des mutations spécifiques. Pour cette raison, la MF est couramment utilisée dans les modèles de sélection mutationnelle développés ces dernières années, pour identifier les mutations possiblement pro/anti-tumorales dans un contexte MSI (Jonchere et al., 2018; Woerner et al., 2010).

Dans cette partie, je présente deux modèles de sélection qui ont été développés spécifiquement pour la compréhension de la tumorigenèse MSI, afin d'identifier les mutations microsatellites *drivers* par la fréquence mutationnelle selon 1) la taille de répétition et 2) les propriétés immunogènes (Ballhausen et al., 2020; Jonchere et al., 2018).

2.2.1. Sélection et classification des évènements mutationnels selon la taille de répétition des microsatellites

Pour rappel, les mutations *frameshift* observées dans les microsatellites peuvent entraîner une inactivation fonctionnelle qui peut être pro- ou anti-tumorale selon la classification du gène (*i.e.* oncogène ou suppresseur de tumeur). Également, la probabilité qu'un microsatellite soit muté dépend de sa taille de répétition : plus il est de grande taille, plus il est susceptible d'être muté (voir chapitre 1, partie 2.1). En 2010, un modèle de prédiction de la fréquence mutationnelle basé sur la taille de la répétition, a permis d'identifier des mutations sélectionnées ou contre-sélectionnées ou encore neutre en fonction de la fréquence mutationnelle respectivement plus élevées ou plus faibles ou identiques que celles attendues dans le modèle stochastique pour chaque taille de répétition. L'ensemble de ces mutations microsatellites (intergéniques, introniques et exoniques) identifiées dans les adénocarcinomes MSI, sont regroupées dans la base de données *Seltarbase* (Woerner et al., 2010).

En 2018, notre équipe a publié une analyse génomique de l'instabilité microsatellitaire de 47 tumeurs primaires MSI qui améliore le modèle initial de sélection selon la fréquence mutationnelle observée et la taille de répétition (Figure 25A). Ce modèle démontre l'incidence de la taille de répétition sur la fréquence mutationnelle des microsatellites exoniques uniquement, ce qui permet d'affiner le modèle stochastique. En effet, la densité est définie par l'ensemble des fréquences mutationnelles des microsatellites selon leur taille de répétition ;

les premiers et derniers percentiles définissent les limites (respectivement) basses et hautes de la courbe de densité (zone « bruit de fond », où la densité est la plus élevée). De cette manière, les valeurs supérieures et inférieures aux limites de densité (*outliers*) définissent les mutations sélectionnées positivement et négativement (Figure 25B). Parmi les mutations microsatellitaires qui subissent une sélection positive, nous retrouvons des gènes connus tels que TGFBR2, ACVR2A, BMPR2, BAX ou encore MSH3. Plus généralement, la nature de la sélection suggère le caractère pro- ou anti-tumorale des mutations au cours de la tumorigenèse (Figure 25A) (Jonchere et al., 2018).

Toutefois, ce modèle présente des limites puisque la densité observée atteint ses valeurs maximales pour un microsatellite d'une taille supérieure ou égale à 11 nucléotides et minimales pour un microsatellite d'une taille inférieur ou égale à 8 nucléotides (Figure 25B). *De facto*, l'identification des microsatellites sélectionnés positivement ou négativement est optimal pour les microsatellites de taille de répétition de 9 à 20 nucléotides (sélection négative), et pour ceux de taille 5 à 10 nucléotides (sélection positive).

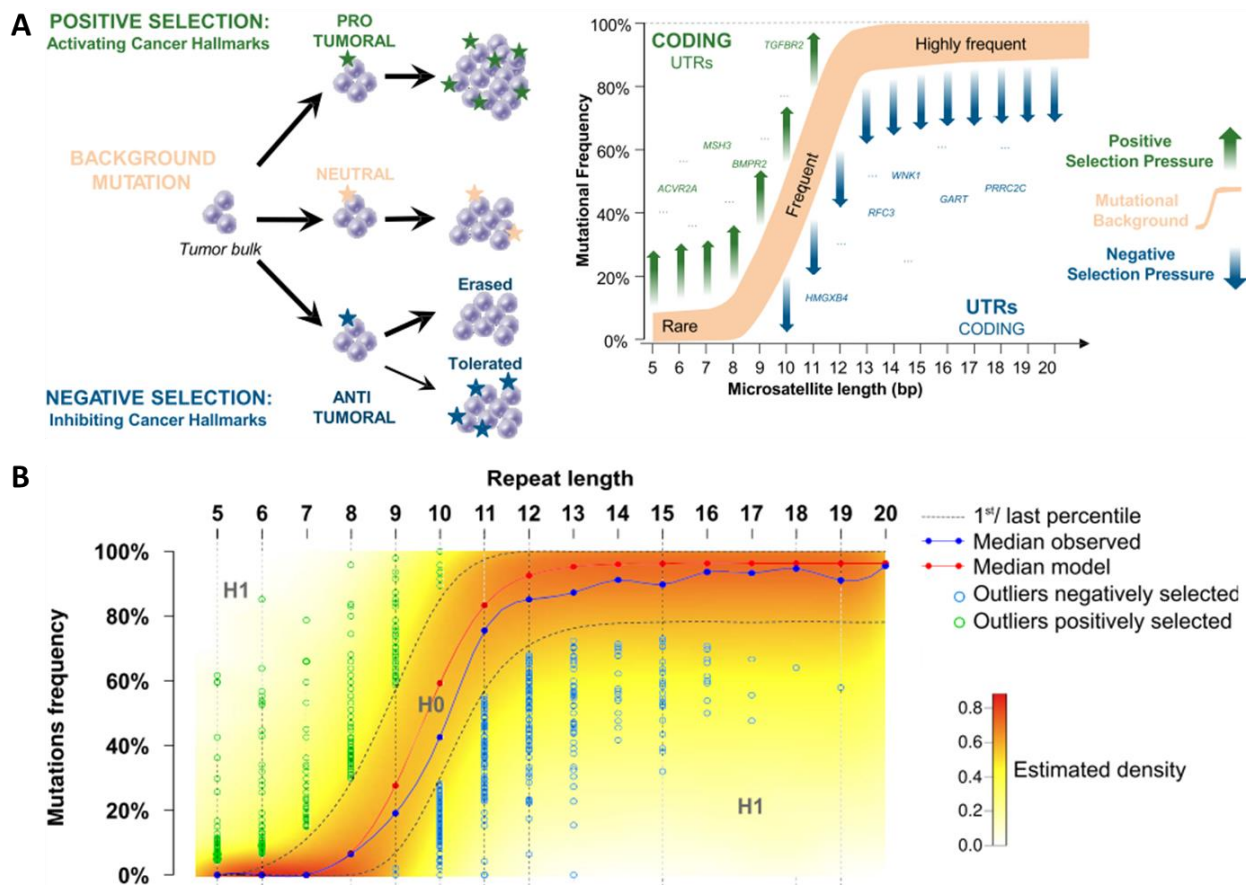


Figure 25 | Le modèle de sélection des mutations au cours de la tumorigenèse MSI selon la taille de répétition des microsatellites.

(A) Représentation schématique. **(B)** Distribution des mutations microsatellitaires exoniques (codant + UTR) selon leur fréquence mutationnelle et leur taille de répétition. Sélection des mutations : les évènements marqués en vert sont considérés comme positivement sélectionnés ; ceux en bleu sont considérés comme négativement sélectionnés (Jonchere et al., 2018).

2.2.2. Sélection et classification des évènements mutationnels selon les propriétés immunogènes

En 2020, Ballhausen et ses collègues ont suggéré l'implication du système immunitaire dans la sélection des mutations *drivers* au cours de la tumorigenèse qui pourrait intervenir dès le stade de cryptes MMR-déficientes (Figure 26A). Les chercheurs ont développé une stratégie pour identifier les mutations les plus immunogènes. Pour cela, ils ont sélectionné *in silico* les mutations microsatellitaires codantes les plus fréquentes qui génèrent des FSP néo-antigéniques ; la probabilité que les FSP soient présentés par le HLA de classe I et qu'ils déclenchent une réponse immunitaire anti-tumorale, a été évaluée grâce à un l'outil statistique de prédiction GELS, qu'ils ont développés (Ballhausen et al., 2020). Cette stratégie

a permis d'identifier une cinquantaine de mutations, dont certaines déjà connues comme ACVR2A, TGFBR2, CASP5, AIM2, TAF1B ou encore BAX. Malgré leur pouvoir immunogène important (anti-tumorale), ces mutations ne sont pas contre-sélectionnées au cours de la tumorigenèse MSI, au contraire elles sont sélectionnées positivement (Jonchere et al., 2018).

À partir de ces résultats, les chercheurs ont suggéré un modèle de sélection des mutations où la pression de sélection dépendrait de l'environnement immunitaire anti-tumoral (Figure 26B). De cette manière, les mutations *drivers* immunogéniques seraient tolérées et sélectionnées positivement étant donné leurs propriétés pro-tumorales, tout comme les mutations *drivers* non-immunogéniques ; les mutations *passengers* immunogéniques seraient quant à elles contre-sélectionnées en raison de leur propriété anti-tumorale. En revanche, en cas de défaillance de la présentation néo-antigénique des cellules tumorales – suite à une mutation perte de fonction dans le gène B2M observée dans 20-40% des CRC MSI (voir chapitre 1, partie 5) –, la pression de sélection immunitaire ne serait plus exercée et l'ensemble des mutations, qu'elles soient immunogènes ou non, seraient tolérées (Ballhausen et al., 2020).

Ce modèle permettrait ainsi d'identifier les mutations *drivers* pro-tumorales (immunogéniques) au cours de la tumorigenèse MSI colique. Cependant, ce modèle se limite à décrire des événements de sélection positive uniquement, pour des mutations dont les capacités immunogènes (prédiction de présentation néo-antigénique et stimulation d'une réponse immunitaire anti-tumorale) ont déjà été décrites.

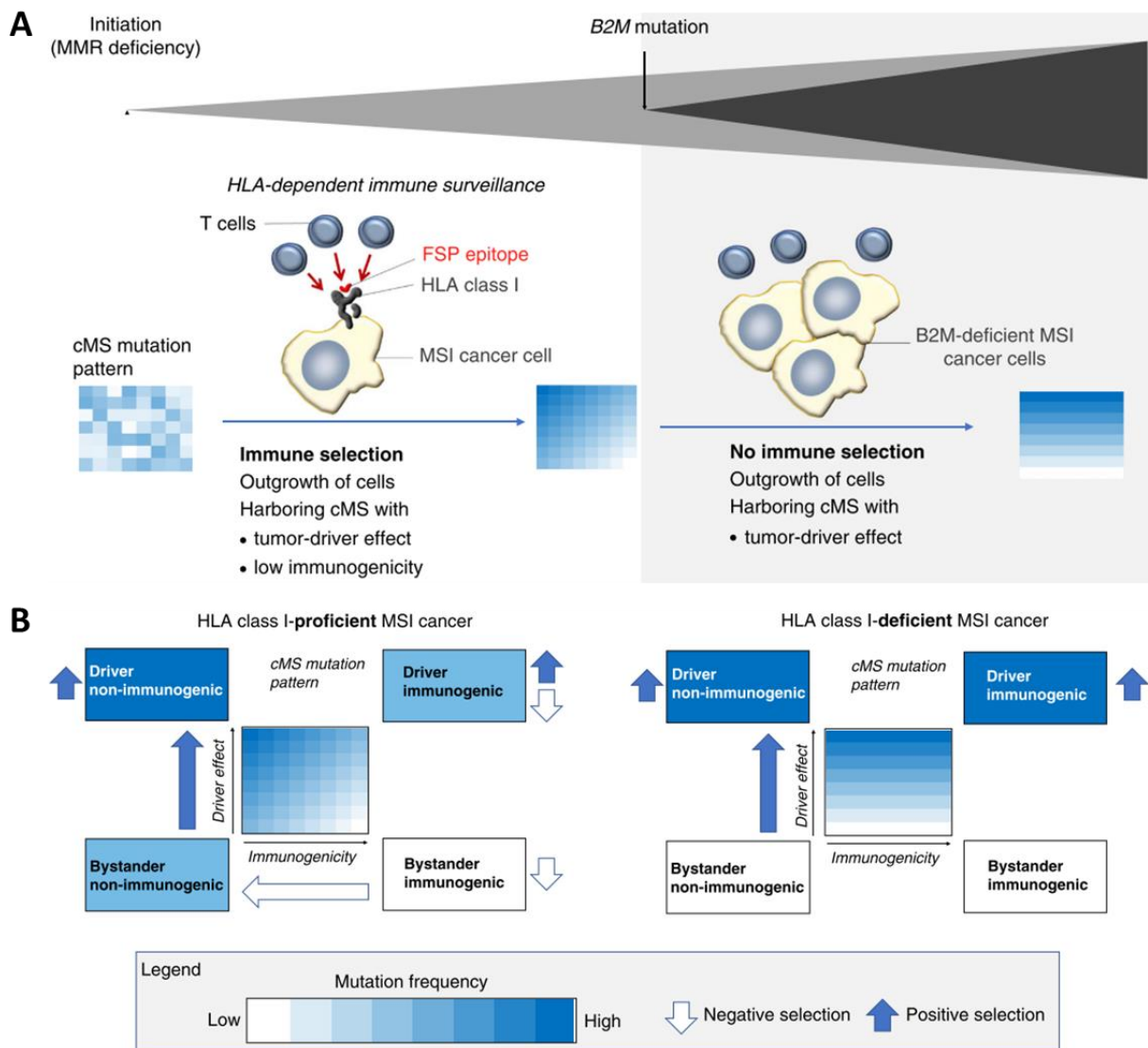


Figure 26 | Le modèle de sélection des mutations au cours de la tumorigenèse MSI selon des propriétés immunogènes (et pro-tumorales) des mutations.

(A) Représentation schématique de la tumorigenèse MSI colique selon le modèle de sélection d'immunogénicité des mutations. La mutation du gène B2M entraîne une défaillance de la présentation néo-antigénique par le HLA de classe I, ce qui représente un mécanisme d'échappement immunitaire acquis durant la progression tumorale. **(B)** Illustration du fonctionnement de la sélection des mutations *drivers* pro-tumorales par leur propriété immunogène. À gauche, le modèle montre une sélection des mutations selon leur capacité immunogène et pro-tumorale (*drivers*) alors qu'à droite, les mutations subissent uniquement une pression de sélection concernant leur propriété pro-tumorale, dans un contexte où la présentation néo-antigénique *via* le HLA de classe I est défaillante (*i.e.* suite à une mutation du gène B2M) (Ballhausen et al., 2020).

TRAVAUX DE RECHERCHE

ARTICLE 1

“Massive Cis-Perturbation of Alternative Splicing Due to Microsatellite Instability at U2AF-Binding Polypyrimidic Sites Early Impairs Cell Differentiation And Drives Colorectal Cancer Initiation”

Jonchère V. & Montémont H. et al., en révision dans le journal *Genome Biology*

Résumé de l'article n°1

En 2011, mon laboratoire d'accueil a démontré que l'instabilité d'un microsatellite intronique localisé dans la séquence polypyrimidique en amont de l'exon 9 du gène *HSP110* entraînait un épissage aberrant de cet exon (Dorard et al., 2011). De la même manière, des chercheurs avaient déjà montré une corrélation entre l'instabilité microsatellitaire intronique et l'épissage aberrant du gène *MRE11* (Giannini et al., 2004). À partir de ces observations nous avons voulu investiguer à grand échelle les perturbations du processus d'épissage et leur rôle dans l'initiation et la progression tumorale dans le contexte MSI de cancers coliques.

Pour cela, nous avons étudié la tumorigenèse MSI et rechercher les mutations dans les séquences régulatrices de l'épissage, à travers trois types de lésions : (i) des cryptes MMR-déficientes (une lésion pré-cancéreuse observée uniquement dans la muqueuse colique saine des patients atteints par le syndrome de Lynch) ; (ii) des lésions adénomateuses MMR-déficientes (tumeurs bénignes) ; (iii) des adénocarcinomes (tumeurs malignes) MSI. Nos analyses ont été réalisées à partir du séquençage des ARN des adénocarcinomes pour l'identification des épissages aberrants et du séquençage de l'exome des trois lésions, pour l'évaluation de l'instabilité microsatellitaire intronique.

Cette étude a montré l'implication de l'instabilité microsatellitaires dans les régions introniques situées dans la séquence polypyrimidique, sur l'épissage des ARN pré-messagers. Nous avons identifié une signature transcriptomique de 985 exons dont l'épissage est spécifiquement altéré dans les cancers MSI. Parmi ces 985 sauts d'exons, 96 sont directement corrélés à une instabilité microsatellitaire dans l'ADN correspondant et ceci à haute fréquence mutationnelle. Parmi ces 96 exons, environ 60% (61/96) sont retrouvés instable dès les phases précoces du développement tumorale MSI, dans les cryptes MMR-déficientes. Ces résultats suggèrent que l'altération de l'épissage des ARN spécifiques des cancers MSI soit impliquée dans l'initiation de la carcinogenèse MSI. Également, cette signature mutationnelle des 985 exons épissés alternativement se caractérise par un grand nombre de mutations *frameshift* dont certaines entraînent la synthèse de peptides néo-antigéniques présentés par le MHC des cellules tumorales aux cellules immunitaires. Aussi, la signature mutationnelle MSI contient 134 exons dont l'épissage alternatif, observé dans les tumeurs MSI, est responsable de la différenciation des cellules. Enfin, nous avons montré que cette signature mutationnelle est

corrélée à la signature oncogénique U2AF1-S34F déjà connue, puisque présente dans de nombreux cancers (tels que les cancers du poumon, du col de l'utérus, les mélanomes, ou encore des leucémies myéloïdes aigües), et également associée au mécanisme d'épissage des ARN pré-messagers.

Dans l'ensemble, ces travaux améliorent notre connaissance de la tumorigenèse MSI colique, de l'initiation à la progression, et ouvre un champ de recherche quant à l'instabilité microsatellitaire non-codante mais fonctionnelle, notamment dans la régulation de l'épissage des ARN pré-messagers.

Ce travail fait l'objet d'un article que je co-signe en premier auteur et qui a été soumis à publication en octobre 2023 et est en cours de révision dans le journal *Genome Biology*.

Paris, 24th October 2023

Dear Sir,

On behalf of my colleagues, please find attached our manuscript entitled “**Massive Cis-Perturbation of Alternative Splicing Due To Microsatellite Instability at U2AF-Binding Polypyrimidic Sites Early Impairs Cell Differentiation And Drives Colorectal Cancer Initiation**” that we are submitting for publication in *Genome Biology*.

Genomic instability creates the diversity that facilitates the development of cancer hallmarks in tumor cells [1]. High levels of genome instability are observed in a common type of cancer with a so-called mutator phenotype which stems from impairment of the DNA mismatch repair (MMR) system in tumor cells and briefly consists in the accumulation of numerous nucleotide mutations, particularly at repetitive DNA sequences such as microsatellites (MS). Cancers featuring this type of genomic instability are referred to as MSI (for Microsatellite Instability) cancers [2-4].

From a pathophysiological point of view, it has been first reported that MSI was likely to trigger important consequences through the occurrence of oncogenic mutations affecting coding microsatellites contained in cancer-related genes, e.g., in *TGFBR2* or *BAX* [5, 6]. In the noncoding DNA, MS located near the 3' splice acceptor site often lie within the cis-splicing intronic polypyrimidine tract. Though the instability of a few noncoding MS has been reported by our team and others to generate splicing anomalies in MSI CRC [7], this phenomenon has never been investigated in deep at the exome scale. In the present study, we show that MSI leads to hundreds of noncoding mutations endowed with cis-activity in splicing. These events occur prior to cell transformation and MSI coding mutations in the yet untransformed dMMR colonic crypt. By altering U2AF binding to pre-mRNA, they primarily affect the expression of numerous alternative exons encoding protein isoforms primarily governing cell fate, blocking cell differentiation and driving tumor initiation in MSI CRC as a cis-mimic of the U2AFS34F mutation in several other non-MSI cancers. We have carried out these analyses on normal and tumor material from patients or at the single-cell level.

We believe our findings are highly original in the field of cancer genetics and represent a major breakthrough in our understanding of a frequent mode of carcinogenesis associated with a large clinical spectrum of tumors. By connecting several fields of knowledge, namely tumor genomics, gene splicing, and tumor pathophysiology, these findings will be of broad interest to the scientific and medical research community. Finally, our discovery potentially paves the way for the identification of new targeted therapies against MSI cancers, at a time when precision medicine for MSI cancers using immune checkpoint inhibitors is growing fast, with the need however to find new targets in combination to overcome resistance to these new drugs, as recently illustrated by our team [8].

Sincerely Yours,

Alex Duval, MD, PhD, Corresponding Author

REFERENCES

1. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation.** *Cell* 2011, **144**:646-674.
2. Aaltonen LA, Peltomaki P, Leach FS, Sistonen P, Pylkkanen L, Mecklin JP, Jarvinen H, Powell SM, Jen J, Hamilton SR, et al.: **Clues to the pathogenesis of familial colorectal cancer.** *Science* 1993, **260**:812-816.
3. Branch P, Aquilina G, Bignami M, Karran P: **Defective mismatch binding and a mutator phenotype in cells tolerant to DNA damage.** *Nature* 1993, **362**:652-654.
4. Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J, Kane M, Kolodner R: **The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer.** *Cell* 1993, **75**:1027-1038.
5. Markowitz S, Wang J, Myeroff L, Parsons R, Sun L, Lutterbaugh J, Fan RS, Zborowska E, Kinzler KW, Vogelstein B, et al.: **Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability.** *Science* 1995, **268**:1336-1338.
6. Rampino N, Yamamoto H, Ionov Y, Li Y, Sawai H, Reed JC, Perucho M: **Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype.** *Science* 1997, **275**:967-969.
7. Dorard C, de Thonel A, Collura A, Marisa L, Svrcek M, Lagrange A, Jegou G, Wanherdrick K, Joly AL, Buhard O, et al: **Expression of a mutant HSP110 sensitizes colorectal cancer cells to chemotherapy and improves disease prognosis.** *Nat Med* 2011, **17**:1283-1289.
8. Ratovomanana T, Nicolle R, Cohen R, Diehl A, Siret A, Letourneur Q, Buhard O, Perrier A, Guillerme E, Coulet F, et al: **Prediction of Response to Immune Checkpoint Blockade in Patients with Metastatic Colorectal Cancer with Microsatellite Instability.** *Ann Oncol* 2023.

Massive Cis-Perturbation of Alternative Splicing Due To Microsatellite Instability at U2AF-Binding Polypyrimidic Sites Early Impairs Cell Differentiation And Drives Colorectal Cancer Initiation

5 Vincent Jonchère ^{1€}, Hugo Montémont ^{1€}, Aurélie Siret ¹, Emmanuel Tubacher ², Christophe Battail ³, Assane Fall ¹, Enora Le Scanf ⁴, Karim Labreche ¹, Victor Renault ², Toky Ratovomanana ¹, Quentin Letourneur ¹, Olivier Buhard ¹, Ariane Jolly ⁵, Philippe Le Rouzic ¹, Habib Zouali ², Florence Renaud ¹, Rémy Nicolle ⁶, Pascale Cervera ^{1,7}, Magali Svrcek ^{1,7}, Pierre Bourgoïn ^{1,7}, Hélène Blanché ², Anne Boland ³, Jérémy Lefèvre ^{1,8}, Yann Parc ^{1,8}, Mehdi Touat ^{1,9}, Franck Bielle ¹⁰, Romain Cohen ^{1,11}, Thierry André ^{1,11}, Florence Coulet ^{1,12}, Pierre de la Grange ⁵, Aurélien de Reyniès ⁶, Jean-François Fléjou ^{1,7}, Agusti Alentorn ¹, Laurent Corcos ^{4*}, Jean-François Deleuze ^{2*}, Ada Collura ^{1*}, Alex Duval ^{1,12*§}

€ Co-first Authors

* Co-senior Authors

15 § Corresponding author

1- Sorbonne Université, INSERM, Unité Mixte de Recherche Scientifique 938 and SIRIC CURAMUS, Centre de Recherche Saint-Antoine, Equipe Instabilité des Microsatellites et Cancer, Equipe labellisée par la Ligue Nationale contre le Cancer, F-75012 Paris, France.

20 2- Laboratory for Genomics, Foundation Jean Dausset–CEPH (Centre d’Etude du Polymorphisme Humain), Paris, France.

3- Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine (CNRGH), 91057, Evry, France.

4- INSERM, UMR 1078, Université de Brest, Génétique Génomique Fonctionnelle et Biotechnologies, Etablissement Français du Sang, Brest, France.

25 5- Genosplice, Paris, France.

6- Programme “Cartes d’Identité des Tumeurs”, Ligue Nationale Contre le Cancer, Paris, France.

7- Sorbonne Université, department of Pathology, AP-HP.sorbonne université, hôpital Saint-Antoine, 47-83 boulevard de l’hôpital, F-75012 Paris, France.

30 8- Sorbonne Université, department of Digestive Surgery, AP-HP, Hôpital Saint-Antoine, Paris, France.

9- Sorbonne Université, Inserm, CNRS, UMR S 1127, Institut du Cerveau et de la Moelle épinière, ICM, AP-HP, Hôpitaux Universitaires La Pitié Salpêtrière - Charles Foix, Service de Neurologie 2 Mazarin, Paris, France.

35 10- Sorbonne Université, Inserm, CNRS, UMR S 1127, Institut du Cerveau et de la Moelle épinière, ICM, AP-HP, Hôpitaux Universitaires La Pitié Salpêtrière - Charles Foix, Service de Neuropathologie Laboratoire Escourolle, Paris, France.

11- Sorbonne Université, department of Medical Oncology, AP-HP, Hôpital Saint-Antoine, Paris, France.

12- Genetics Department, AP-HP.sorbonne université, Paris, France.

40

DECLARATIONS

Ethics approval and consent to participate

This study was approved by our institutional review board/ethic committee (Sorbonne University) and informed consent was recorded. Research has been performed in accordance with the Declaration of Helsinki.

45

Availability of data and materials

Exome sequencing and RNA sequencing data have been deposited in the European genome–phenome archive. The data deposit is in progress and all data will be available on ebi/SRA before publication.

50

Competing interests

The authors declare no financial and non-financial competing interests.

Funding

This work was supported by grants from SIRIC CURAMUS and the “Ligue Nationale Contre le Cancer”. A. Duval’s team is labeled by the French National League against Cancer. The sponsor was Assistance Publique – Hôpitaux de Paris (Département de la Recherche Clinique et du Développement)”. The study was funded by a grant from Programme Hospitalier de Recherche Translationnelle en Cancérologie – PHRT-K 14056 (Ministère de la Santé)”. This work was also supported by the France Génomique National infrastructure, funded as part of the « Investissements d’Avenir » program managed by the Agence Nationale pour la Recherche (contract ANR-10-INBS-09)”.
60

60

Authors' contributions

All authors have made substantial contributions to the conception **OR** design of the work **OR** the acquisition, analysis, **OR** interpretation of data **OR** have drafted the work or substantively revised it. All authors have approved the submitted version. They have agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

65

Acknowledgements

We thank the CEPH-Biobank and CNRGH teams for their technical expertise. We thank the Tumorothèque CRB Cancer, Hôpital Tenon, APHP.sorbonne université, Paris, France.

70

Corresponding Author: Alex Duval, email: alex.duval@inserm.fr

75

ABSTRACT

Background Microsatellite instability (MSI) due to mismatch repair deficiency (dMMR) is frequent in colorectal cancer (CRC). To date, the role of MSI in CRC has been associated with somatic coding events, whereas the noncoding pathophysiological impact of this mode of genomic instability is yet poorly understood. Here we combined an extensive analysis of coding and noncoding MSI events at the different steps of colorectal tumorigenesis using whole exome and the search for associated splicing events using RNA sequencing at tumor bulk and single-cell level.

Results Our results demonstrate that MSI leads to hundreds of noncoding mutations in pre-mRNA at polypyrimidine U2AF binding sites endowed with cis-activity in splicing. We show that these events occur very early prior to cell transformation in the yet untransformed dMMR colonic crypt, acting as a cis-mimic of the U2AFS34F mutation in several other non-MSI cancer types. Our data also illustrate how such a burst of somatic MSI mutations impairs colonic cell differentiation by primarily affecting the expression of alternative exons encoding protein isoforms governing cell fate while also targeting constitutive exons, making dMMR cells early immunogenic, well before the onset of coding mutations.

Conclusions Overall, these data revisit the role of MSI in colorectal tumorigenesis. They show evidence for a new very early cis-acting noncoding MSI splicing pathway mimicking the mutation in trans of U2AF1 in several other cancers which promotes MSI cancer initiation, far before coding MSI which accumulate later in MSI CRC during tumor progression.

KEY WORDS

Colorectal Cancer, Mismatch Repair Deficiency, Microsatellite Instability, Whole Exome and RNA sequencing, Alternative Splicing, Polypyrimidine U2AF binding site, Colonic Crypt, Cancer initiation, Cell differentiation

BACKGROUND

Genomic instability creates the diversity that facilitates development of the hallmark properties of tumor cells [1]. Following impairment of the DNA mismatch repair (MMR) system, high levels of genomic instability are observed in cancers with the mutator phenotype. This results in the accumulation of numerous mutations, particularly in repetitive DNA sequences such as microsatellites (MS) [2, 3]. Tumors featuring this type of genomic instability are known as MSI (for Microsatellite Instability) cancers [4-6]. They can arise in patients with the hereditary cancer condition known as Lynch syndrome (LS) or sporadically, as in approximately 15-20 % of all colorectal, gastric and endometrial cancers [7]. In dMMR tumor cells, it has been clearly demonstrated that the level of instability of a MS is positively and strongly correlated with its size [8, 9]. So far, somatic MS alterations leading to changes in gene homeostasis have been primarily observed in poorly unstable small coding MS-containing genes, large microsatellites (i.e. > 10 bp) not being present in the coding portion of the human genome due to the counter-selection of such more highly unstable sequences during evolution of species (for review see [10]). Although some of these coding MS mutations have been proposed to play a role in the tumorigenic process, with no real functional evidence behind the claim in many cases, their occurrence has been mostly related to the synthesis of numerous aberrant immunogenic neoantigens [11] and the infiltration of MSI tumors with activated cytotoxic T-cell lymphocytes (CTLs) and Th1 cells, thus creating a hostile, anti-tumor microenvironment. Consequently, MSI colorectal cancer (CRC) has a more favorable prognosis than MS stable (MSS) CRC, at least for localized tumors [12], while they can be treated effectively with immune checkpoint inhibitors [13] although resistance to these new drugs can still develop [14-16]. Besides, it is noteworthy that the DNA alterations due to MSI in cancer are much more numerous in the noncoding part of the genome [8, 9]. However, the functional consequences of MSI noncoding mutations remain largely unknown.

Human genes are generally comprised of multiple (coding) exons interspersed with (noncoding) introns that undergo splicing to generate mature messenger RNAs (mRNA) and subsequently proteins (reviewed in [17]). Pre-mRNA splicing is a complex process that involves constitutive intron removal and ligation of most exons in the gene, as well as the inclusion or exclusion of certain exons from the final mRNA [17]. Alternative RNA splicing is observed at physiologically regulated exons in more than 95% of the 19,804 transcribed human genes [17]. This process enables the production of thousands of alternative mRNAs encoding numerous self-protein isoforms generated specifically within a tissue during cell differentiation [17, 18]. Alternative mRNA splicing dominates changes in the transcriptome as stem cells differentiate into tissue progenitor cells, notably in the colonic crypt [19]. Accumulating evidence indicated that both alternative and constitutive exons are aberrantly spliced during cancer development, in part due to somatic mutations resulting from genome instability [20]. An excess of alternative splicing has recently been observed using long-read nanopore sequencing in MSI CRCs as compared to MSS CRCs, although the authors did not examine the reason for such an observation, nor the pathophysiological role that these alternative transcripts may have in these tumors [21]. Interestingly, numerous MS that can be highly unstable in the genome of MSI cancers due to their large size are intronic and located near the polypyrimidine tract (PY) of the intron/exon boundary [22, 23]. These noncoding MS are important cis-acting sequence elements for the binding of trans-acting core spliceosomal factors like U2AF2, upstream the

intronic AG splice acceptor site which binds U2AF1, for guiding intron removal during pre-mRNA splicing [24].

150 In this work, we combine an extensive analysis of coding and non-coding MSI events at the different steps of colorectal tumorigenesis, from the dMMR crypt to adenocarcinoma, using high-throughput whole exome sequencing, with the search for associated splicing events at both alternative and constitutive exons in the RNA of these tumors using high-throughput RNA sequencing. We have carried out these analyses on normal and tumor material from patients, by considering either the tumor bulk or the single cancer cell. Overall, our results completely revisit the role of MSI in colorectal tumorigenesis. They show evidence for a new very early cis-acting noncoding MSI splicing pathway promoting dMMR cell transformation and yet MSI cancer initiation, far before coding MSI which accumulate later in established MSI CRC. They also establish how, from a mechanistic point of view, this pathway acts as a cis-mimic of the mutation in trans of U2AF1 in several other cancers.

RESULTS

160 Working Hypotheses and Study Design

MS distribution within human genes is not homogenous, with intronic regions displaying more frequent and larger MS than coding exons (**Fig. 1A left panel**). A highly MS-enriched intronic region is particularly observed in the PY (**Fig. 1A, right panel**), which is a major cis-acting determinant of pre-mRNA splicing [25] (**Fig. 1B**). In the present work, we hypothesized (**WH1**) that RNA changes specific of MSI CRCs would occur because of the high mutational burden of these tumors whose mutations could destabilize the physiological splicing process which is dependent on the binding of trans-factor complexes from the spliceosome to the PY cis-sequence sites of pre-mRNAs (**Fig. 1B and 1C**). We also hypothesized (**WH2**) that this noncoding MSI mutational burden disrupting splicing would have an early impact during tumor evolution, as it targets large microsatellites, which have a very high mutation rate in dMMR cells and whose alteration might thus facilitate MSI tumor initiation in particular (**Fig. 1B and 1C**). Finally, we assumed (**WH3**) that the splicing changes due to MSI should mimic those observed in other tumors with disruption of U2AFs complexes since these spliceosomal factors normally bind DNA at the PY for correct splicing (**Fig. 1B and 1C**).

175 To investigate this, whole exome sequencing was performed to study MSI at all the human intron/exon boundaries and up to 50 bp into the intronic sequence. We analyzed a series of dMMR precancerous lesions, i.e., a pool of 10 different dMMR crypts foci from 8 MSI CRC patient and also 14 dMMR adenoma, with the aim to examine the status of MS endowed with a role in splicing at the very early steps of the MSI-driven tumorigenic process (**Fig. 1D**). In addition, 46 MSI CRC samples with matched normal mucosa samples were also analyzed to investigate the mutational burden at the same noncoding MS in established colorectal tumors. RNA sequencing was performed on 101 established MSI CRC including the series of 46 MSI CRC also analyzed by Whole Exome, 32 MSS CRC, and matched normal mucosa samples ($n = 133$). Due to the lack of sufficient material in quantity and quality we were unable to sequence the RNA on the microdissected dMMR crypts and adenoma. Finally, we also examine the issue of alternative splicing in CRC in relation to their MSI or MSS phenotype at the single cell (SC) level using a public dataset from patients with MSI (34 patients - 33,333 cells) or MSS (28 patients – 33,334 cells) CRC [26] (**Fig. 1D**).

Cis-splicing intronic polypyrimidine tract regions are enriched in DNA microsatellites frequently mutated very early in dMMR colonic crypts and colonic dysplasia prior to their transformation into MSI CRCs

In **Fig.2A**, we outline our approach for longitudinal analysis of MSI at crucial steps along the path of colorectal tumorigenesis, i.e., deficient MMR crypt, adenoma and established adenocarcinoma. The results show, through quantitative analysis of MSI in the genome using MSICare, that MSI is already observed in dMMR crypts, albeit at moderate levels, and rapidly reaches a plateau early in adenoma and yet in adenocarcinoma (**Fig. 2B** and also **Supp. Fig. S1A**). Due to differences in microsatellite size, we observed a much higher number of unstable noncoding MS in the vicinity of the PY than unstable coding MS (**Fig. 2C right panel**). In a dynamic way, our data establish that the level of noncoding MSI whose sizes are frequently above 10 nucleotides, reaching 15-20 base pairs and beyond, is already very high in the dMMR crypts whereas the instability of coding MS whose lengths exceptionally exceed 10 nucleotides in the human genome is quantitatively much lower (**Fig. 2C left panel** and **Supp. Fig. S1B**). Among the dMMR cryptic lesions we were able to examine, it appeared that the level of MSI gradually increased from monocryptic dMMR foci to slightly more advanced oligo/polycryptic dMMR foci (**Fig. 2D**). It is clear that in the dMMR crypts, it is the instability of MS very close to the AG site that already prevails within the 3' portion of PY known to play a key role in splicing, with a further increase in the number of such MSI events during the progression towards cancer. (**Fig. 2E**).

Massive perturbation of splicing In MSI CRC

We then turned to the question of pre-mRNA splicing and its disruption in MSI CRC by performing RNA sequencing within the tumor bulk. We deliberately limited our analyses to the issue of exon insertion/exclusion within mRNAs as it represents the most likely parameter of RNA metabolism to be directly influenced by cis-acting MSI at intron boundaries, notably at the PY. Inclusion or exclusion of all human 3' exon-flanking a microsatellite (3' Exon-FMS) was measured by the percent spliced-in index (PSI) as described [27]. This is a standard method that represents the ratio between reads that include or exclude specific exons (**Supp. Fig. S2**). Overall, we identified 985 exons that showed significantly altered expression in MSI CRC (**Fig. 3A**). These included 839 deregulated skipping events at 3' Exon-FMS in MSI CRC compared to normal colonic tissue, and a further 276 compared to MSS CRC. A higher frequency of deregulated exon splicing events was observed in MSI CRC compared to MSS CRC (two-sided p -value $< 2 \times 10^{-20}$) (**Supp. Fig. S3 and Supp. Table S1**). Splicing changes mostly involved exon exclusion (see also the data on aberrant exon inclusion in MSI CRC on **Supp. Fig. S3**). They were observed at high levels in all MSI CRC samples and were a distinctive feature of the transcriptome profile of both hereditary (Lynch syndrome) and sporadic MSI CRC, independently of TNM stage (stage 2 or 3) or the presence of *BRAF/KRAS* driver mutations. Of note, the average number of skipping events per tumor was likely similar to the number of coding MS mutations in this MSI CRC cohort (**Fig. 3B**) [8].

The 985 deregulated exons were then characterized with respect to their alternative or constitutive nature. Their expression pattern was first analyzed in the 133 normal colorectal mucosa samples (57 030 analyzed exons). As expected, the majority of these exons (88.6 %) were constitutively expressed. The remaining exons (11.4 %) were alternatively expressed, are physiologically regulated and also have an expected role in cell differentiation in the normal colonic mucosa (for review see [17, 18]) (**Fig. 3C left panel**). In MSI CRC samples, a minority of

235 the deregulated exons were constitutive ($n = 293/985$, 29.7 %) whereas the majority ($n =$
692/985, 70.3 %) belonged to the 11.4 % of exons that displayed an alternative expression
pattern in normal colonic mucosa (significant enrichment by 6.2-fold; two-sided p -value $< 2 \times 10^{-16}$) (Fig. 3C right panel, Supp. Fig. S4). These alternative exons were deregulated at low,
240 moderate or high frequencies in MSI CRC (Fig. 3D left panel), as opposed to the constitutive
exons that were mostly skipped at lower frequencies (Fig. 3D right panel) (two-sided p -value
 $< 2 \times 10^{-16}$). Finally, we were able to confirm that exon skipping also recurrently occurred at a
number of the above 839 candidate exons in independent cohorts of MSI tumors from the
TCGA. These results were obtained considering not only MSI CRC, but also MSI gastric and
endometrial cancer, highlighting that the splicing signature is a shared pathophysiological
245 feature of these frequent MSI tumors deriving from 3 distinct primary sites, and probably in
pan-cancer (Fig. 3E).

Confirmation of the increased prevalence of alternative splicing events in MSI versus MSS CRC at the single-cell level

In line with the above results, we wanted to further investigate the issue of alternative
splicing in CRC in relation to their MSI or MSS phenotype at the single cell (SC) level.
250 Independent public scRNA sequencing dataset from both MSI and MSS CRC were used from
the Pelka et al study [26]. In both types of CRC, i.e., MSI and MSS tumors, the predominance of
tumor clones with the same TA (Transit Amplifying cells) state was observed using Census
(<https://doi.org/10.1101/2022.10.19.512926>) and CellTypist [28] (Fig. 3F) and the scRNA
sequencing analyses therefore focused on these predominant clones within the tumor bulk.
255 With the MARVEL package [29] which enables systematic and integrated splicing and gene
expression analysis of single cells to characterize the splicing landscape and reveal biological
insights, we were able to measure the degree of alternative exon inclusion from scRNA
sequencing data in both MSI and MSS contexts. The overall number of alternative splicing
events between TA MSS and TA MSI cell types are represented in Fig. 3G. Results clearly show
260 an overall highly significant increase of such events in MSI compared to MSS TA malignant
clones (p -value $< 2 \times 10^{-74}$). Noteworthy, these analyses cumulatively detect all types of
alternative splicing events and not only exon skipping. They support at the single-cell level a
strong link between MSI and alternative splicing during transcription in CRC.

265 Splicing anomalies found in MSI CRCs are frequently caused by MSI through mutations that occur very early during cancer development

We then sought to address whether a substantial proportion of splicing anomalies in MSI
CRCs were directly induced by MSI through a cis-acting mechanism at the PY intronic site. This
was done by assessing whether somatic deletions in noncoding 3' exon-flanking MS would
promote exon skipping in MSI tumors. We examined the status of DNA MS by whole exome
270 sequencing (up to 50 bp into the intronic sequence) in the initial MSI CRC cohort ($n = 46$) and
compared this to paired normal mucosa ($n = 46$). Of the 985 deregulated exons, 277 MS/exon
couples were selected for this analysis as they showed sufficiently high coverage and mutation
frequency at 3' noncoding MS to provide adequate statistical power for the tumor cohort
(Supp. Table S2). In Figure 4A, we show 6 examples of genes in which the expression level of a
275 single exon is closely related to the status of the DNA microsatellite in the PY. As the size of
somatic deletions in the intronic MS of MSI tumors became larger, the exclusion of these 96
exons in tumor RNAs increased. Overall, a significant correlation between somatic deletions at
flanking 3' noncoding MS and exon skipping (PSI Index) was observed for 96 of these 277

280 candidates (34.7 %) (Fig. 4B). Alternative exons were more readily skipped by such cumulative somatic deletions due to MSI compared to constitutive exons in MMR-deficient tumors (Fig. 4B, Supp. Fig. S5, S6 and Supp. Table S3). For a subset of the 96 noncoding MS (also called ES96), a mini-gene reporter system [30] was used to transfect 2 cell lines (SNU-1, MSI and AGS, MSS). The deletions of 3 MS (i.e., a T14, T11 and T10) was indeed a trigger of exon skipping compared to negative controls (i.e., noncoding MS whose instability did not cause exon skipping in MMR-deficient CRC) (Fig. 4C and Supp. Fig. S7).

285
290
295
Next, we sought to clarify the kinetics of occurrence of the above MSI-driven splicing mutations during the multi-step process that paves the way to cancer in the colorectal mucosa, with the aim to identify the most precocious ones since they could be considered as driving and initiating splicing mutations. Of the 96 mutations whose occurrence was associated with exon skipping in tumors, a majority of them, i.e., 61/96 (64.9%), were already present in dMMR colonic crypts, and at high frequencies for a number of them, e.g., in *ORC2L* (88%), *PDS5A* and *NUP153* (75%), and *RAB3GAP2* (71%) (Fig. 4D, 4E, Supp. Fig. S8 and Supp. Table S4). Taken together, these genes provide a signature that testifies to the earliness of many aberrant MSI-driven splicing MSI mutations in the dMMR crypt, prior to its transformation, in a much more prevalent way than what can be observed at the level of coding microsatellites at this pre-tumoral stage. These early events are associated in particular with cell cycle, focal adhesion and autophagy or PI3K-AKT pathways for instance, suggesting an early perturbation of these processes highly connected with tumor initiation in the yet untransformed dMMR colonic mucosa (Fig. 4E and Supp. Fig. S9, Supp. Table S5).

300 **The MSI splicing signature primarily deregulates alternative mRNA isoforms that normally drive cell differentiation in the colonic crypt**

305
310
We next sought to identify functional relevance of the splicing signature identified here in CRC ($n = 985$ exons). When we compared our MSI splicing signature ($n = 985$ exons) to that of Habowski et al. which groups alternative mRNA changes that drive cell differentiation in the normal murine colonic crypt [19], an enrichment in genes whose alternative splicing is modified as stem cells differentiate first into progenitors and then into mature cell types in the normal colonic crypt was observed [19] (Fig. 5A and Supp. Table S6). This enrichment comprised a total of 134 unique genes. Compared to coding MS mutations in CRC, MSI-driven exon skipping events mainly deregulate the fine tuning of self-protein isoforms that are normally tightly regulated in the normal colonic mucosa under physiological conditions (Fig. 5B). Deregulation of these self-protein isoforms, while being expected to be well tolerated from an immune point of view, should impair the differentiation of MMR-deficient cancer cells, in line with the poorly differentiated histological pattern that is a characteristic of MSI CRC as compared to MSS CRC, being observed in the great majority of MSI tumors [7, 31].

315
In line with this, alternative exon deregulation in MSI CRC was significantly associated with several cancer-related pathways known to play an important role in tissue differentiation in the colon along the colonic crypt and also during colorectal cancer initiation, e.g., Hippo signaling pathways such as BMP and Wnt/Wingless signaling (two-sided p -value < 0.01) [32] (Fig. 5C).

320 **The MSI splicing signature also generates early immunogenic aberrant frameshifts at constitutive exons**

In addition, we also investigated whether splicing defects affecting constitutive exons in mRNAs might have an early immunogenic impact, since this kind of exon skipping is not physiologically observed in the normal colonic tissue. Taking into account the HLA-MHC status of each patient (CRC only) and the presentation of aberrant peptides generated in this setting, a large number of immunogenic neopeptides were predicted to be presented by the patient's tumor cells (Fig. 5D). These data strongly suggest an immunogenic nature of the MSI splicing signature, and we further examine its implementation at the early steps of MSI CRC tumorigenesis. A significant amount of these constitutive immunogenic exons are present in dMMR crypts (20/27, 74.1%) including 10 of them with high mutational frequency (MF \geq 50%) (Fig. 5E). These results suppose immunogenic property already in early stage which can stimulate immune infiltration in normal colonic mucosa as it was recently published [33].

Aberrant splicing events observed in MMR-deficient CRC are shared with those induced by oncogenic dysfunction of the spliceosomal factor U2AF1 that fosters blockade of cell differentiation in several human cancer types

We next investigated whether exon skipping events in MSI CRC could be shared with oncogenic signatures observed following mutations of core spliceosomal proteins or associated trans-acting RNA splicing factors in other cancer types (Fig. 5F left panel). The signature ($n = 985$ exons) was compared to results from 20 publications that described the exon signature of 9 RNA splicing factors (ESRP1, ESRP2, FUBP1, RBM10, RBM4, SF3B1, SRRM4, SRSF2, U2AF1) in different cancer types. Strikingly, the MSI splicing signature was significantly enriched in exons that were skipped following U2AF1 dysfunction in several cancer types [34-36] (Fig. 5F right panel, Supp. Fig. S10 and Supp. Table S7). This enrichment corresponded to a total of 125 unique exons and the results concur with the known role of mutations in U2AF1 in cancer, particularly U2AF1 S34F in lung carcinoma, melanoma and acute myeloid leukemia [37, 38]. The present results suggest that, mechanistically, the cis-acting, MSI-driven splicing pathway promotes tumor development by mimicking the oncogenic impact of U2AF1 dysfunction on several gene targets (Fig. 5F left panel). To further illustrate this functional scenario, we then characterize the category of non-coding MS in the PY whose instability prone to induce RNA instability at 3' Exon-FMS. Compared to controls, non-coding MS ($n = 96$) that were shown to directly influence RNA splicing in MSI CRC were primarily located very close to the intron/exon junction and to the U2AF1 binding region within the functional polypyrimidine tract [39] (Fig. 5G). This underlines further that aberrant splicing due to MSI can promote tumor development through cis-somatic MS deletions that may mimic U2AF1 perturbation of function and result in skipping of numerous exons, thereby defining a U2AF1-like MSI-driven oncogenic signature (see also Supp. Fig. S11 for the influence on exon skipping of length and deletion size of these intronic MS). In contrast, the impact of additional mutations that occurred outside MS at the donor (GT) or acceptor (AG) sites on exon exclusion levels was shown to be very low (Supp. Fig. S12).

DISCUSSION

MSI is a common type of genetic instability in human solid tumors with around 700,000 new cases of MSI cancers diagnosed worldwide every year. To date, this mode of instability and its pathophysiological consequences with regard to tumorigenesis have been mostly reported at the level of coding DNA [3, 10, 16, 40, 41]. Here we report for the first time the deep and early

365 impact of MSI on gene splicing in both the untransformed and transformed dMMR colonic
mucosa through a newly identified MSI-driven noncoding genomic pathway. In brief, coding
MS mutations are frameshift and are usually observed as 1 or 2 bp deletions in short DNA
370 repeats and at low/moderate frequencies [8]. They invariably lead to truncation of the
corresponding protein in tumor cells, resulting in aberrant proteins which may have lost their
function. Although a role in tumorigenesis has been established for several coding MS
mutations, they mostly result in heavy tumor infiltration with cytotoxic T cells and hence an
improved prognosis due to their neo-antigenic properties [11, 42]. Our data suggest here that
375 these coding MSI variants are late and hardly observed during the early steps of MSI
tumorigenesis. In contrast, the MS often found in intronic sequences around the PY are shown
here to undergo frequent deletions of one or several bp in MSI cancer cells but also in yet
untransformed dMMR colonic crypts, particularly the long MS contained in the cis-regulatory
intronic polypyrimidine tract. Consequently, these noncoding changes early deregulates at high
380 frequencies the expression of dozens of exons, the number of MSI-driven exon skipping events
in established MSI colon tumor being approximatively similar to the number of MSI-driven
somatic mutations in coding MS. Importantly, alternative exons are shown here to constitute
the main early target of this process in MSI CRC, probably due to their greater susceptibility to
undergo skipping at the intron/exon junction. We confirm this close link between the MSI
phenotype of tumor clones and incrementation of splicing events in an independent, single-cell
level dataset. Thus, the MSI-driven noncoding pathway we highlight in this study mainly
385 deregulates the fine tuning of many self-protein isoforms that are tightly regulated under
physiological conditions and govern the process of cell fate determination in the normal colonic
mucosa [17-19]. Therefore, we suspect that this MSI-driven process would drastically impair
the differentiation of dMMR colonic cells. Besides, a number of constitutive exons are however
also targeted by this MSI pathway in MSI cancer cells which may also initiate immunization
390 mechanisms prior to cell transformation. These data are especially meaningful in the context
of recent studies where elevated mucosal T-cell infiltration was demonstrated in Lynch patients
even in the absence of cancer [33].

Of pathophysiological interest is the significant overlap we document between this
noncoding MSI splicing signature and the oncogenic U2AF-S34F signature in other cancers [34-
395 36]. This constitutes an important argument for its oncogenic and transforming role in the
dMMR colonic crypt, well before the burst of coding mutations that arises much later in
transformed MSI cancer cells. The S34F mutation in U2AF1 is commonly found in
myelodysplastic syndromes and secondary acute myeloid leukemia [43]. In line with our
findings, it is responsible for blocking erythroid differentiation and causing aberrant alternative
400 splicing in hematopoietic progenitors associated with different disease phenotypes [43, 44]. A
number of targets common to both U2AF1 and MSI splicing signatures (*e.g.*, ADAM10 [45],
CENP-E [46]) may underpin this functional impact. MSI tumor progression is associated with
dedifferentiated histopathology and with cancer cell survival in a hostile and highly
immunogenic microenvironment [47]. From all our results, we speculate that MSI-driven cell
405 dedifferentiation fuels tumor heterogeneity, thereby favoring the development of malignant
clones to this local unfavorable environment and yet drug resistance. From our findings, it
appears, that pro-differentiation or splice-correcting agents might be considered as drugs of
choice for the early treatment of MSI lesions and more generally for MMR-deficient cancers.
When combined with immune checkpoint inhibitors, these drugs could also be proposed to
410 improve the personalized therapy of MSI cancers, at a time when precision medicine for MSI
cancers using immune checkpoint inhibitors is growing fast, with the need however to find new

targets in combination to overcome resistance to these new drugs, as recently illustrated by our team and others [16].

415 There are limitations to this study. Mechanistically, our results suggests that MSI-related somatic mutations in splice acceptor sites prevents them from attracting the trans-acting spliceosome machinery, in particular U2AF factors. Though cis-acting, MSI-driven splicing aberrations are likely to promote the development of MSI colon in a similar way to what is observed when U2AF1 is mutated by the S34F mutation, we have not precisely unraveled in this work the exact consequences of MSI on the binding of spliceosome factors to RNA. Some
420 might object that the number of splicing anomalies we report here in MSI CRC is not exceptionally high. This is partly due to the fact that we have willingly chosen to detect only the events with a frequency exceeding 5% in the tumor bulk. Note worthily, this process is here validated in several tumor series including tumor samples from colorectal and other epithelial tissue locations, suggesting that it probably plays a universal major role in MSI tumorigenesis.
425 The next step will be to more accurately describe MSI splicing anomalies in pan-cancer, using in particular long-read sequencing technology to establish their combinations precisely in tumors, and earlier as we initiated here in pre-tumor lesions.

METHODS

430 Samples and Patients

This study included an experimental cohorts composed of 101 MSI and 32 MSS CRC for which both tumor tissue and matched colonic mucosa were available. They were obtained from a retrospective series of patients who underwent surgical resection of mainly stage II-III CRC from 2004 to 2015 at Saint Antoine hospital, Paris, France (Cohort Microsplicother, NI14027HLJ). Also, 22 precancerous tissue were collected: n=14 adenoma - n=5 Low Grade
435 Dysplasia (LGD) and n=9 High Grade Dysplasia (HGD) - and n=8 dMMR healthy crypts (pool of 10 crypts foci, for each patient) associated to their normal mucosa (n = 22). They were obtained from a retrospective series of patients who underwent surgical resection of mainly stage II-III CRC from 2013 to 2020 at Lille University Hospital, France. We have included patients who
440 presented with a tumor showing loss of MLH1 or MSH2 expression at pathological examination. MSI/dMMR status was identified prospectively at diagnosis using the pentaplex PCR method and immunohistochemistry as previously described [7, 31]. Patients who received preoperative chemotherapy and/or radiation therapy were excluded. Extensive clinical data was available for all MSI CRC patients. This study was approved by our institutional review board/ethic
445 committee (Sorbonne University) and informed consent was recorded. Research has been performed in accordance with the Declaration of Helsinki .

RNA Sequencing

RNA extractions and quality controls were performed at the CEPH-Biobank. Before RNA
450 extraction, frozen tissue sections were lysed in QIAzol Lysis Reagent using a TissueLyzer (Qiagen). After chloroform separation, RNA extractions were performed using the miRNeasy Mini Kit (Qiagen) with a QIAcube instrument (Qiagen). Manufacturer's instructions were followed and DNase treatment was included in the protocol. RNA concentrations were evaluated with a Nanodrop 2000C (ThermoFischer Scientific) spectrophotometer. RNA integrity
455 was assessed on a Bioanalyzer 2100 (Agilent) using the RNA 6000 Nano Kit. The average RIN

(RNA Integrity Number) calculated for tumor and normal adjacent tissue was equal 8.1 and 7.2, respectively. Downstream RNA sequencing experiments were performed on selected pairs of samples, each with RIN \geq 7 and DNA yield \geq 2.5 μ g.

460 Stranded mRNAseq sequencing was performed by the « Centre National de Recherche en
Génomique Humaine, Institut de Biologie François Jacob, CEA ». After complete RNA quality
control of each sample (quantification in duplicate and RNA6000 Nano LabChip analysis on
Bioanalyzer from Agilent), libraries were prepared using the "TruSeq Stranded mRNA Library
465 Prep Kit" from Illumina according to the manufacturer's instructions and with an input of 1 μ g
(selection of poly(A) RNAs). Library quality was checked by Bioanalyzer analysis and sample
libraries were then pooled before sequencing to reach the expected sequencing depth.
Sequencing was performed on an Illumina HiSeq 2000 sequencer as paired-end 100 bp reads
using Illumina sequencing reagents. Libraries were generally pooled by 4 samples per lane.
FASTQC v.0.11.7 was used for sequencing quality control and reads were aligned against
470 reference genome hg37 (GRCh37) using TopHat2.

470 Identification of exon skipping events

MSIsensor [48] (scan method) 0.5 was used to identify genomic MS and annotation was by
Annovar [49]. We identified all exons flanked by a MS near an intron/exon junction (up to 50bp
475 into the intronic sequence) and designed a custom junction FASTA reference (50 bases on each
side of the junction). Identification of reads spanning an exon-exon junction was done by BlastN
alignment against our custom reference for each 3'-Exon flanked by a MS (3' Exon-FMS). A
spanning read was considered if its sequence matched at least 8 bases on one side of the
junction and the remainder of the sequence strictly matched the other side of the junction.
Percent spliced in (PSI) [27] scores ranging from zero to one were calculated by dividing
480 Inclusion Reads by Inclusion + Exclusion reads and considered when the event was covered by
at least 5 reads (Inclusion or Exclusion reads).

For each 3' Exon-FMS, the PSI of MSI or MSS tumors was compared to the level in normal
tissues. Adjusted *P*-values were obtained using Benjamini–Hochberg correction of two-sided *t*-
test (*p*-adjust <0.05). In parallel, we also compared the PSI of MSI tumors to the PSI of MSS
485 tumors (*p*-adjust <0.05). For 3'Exon-FMS identified by the two-sided *t*-test adjusted as skipped,
we compared the PSI of each tumor or PSI of each normal matched tissue. Exons considered as
skipped had a fold change of PSI between tumor and normal of greater than 1.5. Finally, only
exons with a mutation frequency > 5% in this cohort were considered. IGV (Integrative
Genomics Viewer) was used to generate sashimi plot. From the 133 available healthy tissues,
490 we considered an exon as alternative if at least 5% of the tissues presented 3 or more reads of
exon exclusion. We did not replicate the same analysis on publicly available cohorts because
we provided recently clear evidence that the performance of MSISensor for the detection of
MSI vs. MSS CRC has to be improved. This was shown in large cohorts of mCRC and nmCRC
495 samples that were previously confirmed as MSI/dMMR or MSS/pMMR by IHC and MSI-PCR
methods performed in large, specialized test centers [50].

Whole exome sequencing

DNA was purified with AllPrep DNA/RNA/miRNA Universal Kit for frozen tissue and AllPrep
DNA/RNA FFPE Kit for FFPE cores, according to the manufacturer's instructions (Qiagen). Their
500 concentrations were determined by the NanoDrop ND-1000 spectrophotometer at 260/280
nm (NanoDrop Technologies, Inc.). Quality and integrity of the nucleic acids were analyzed with

the TapeStation 2200 (Agilent). The whole exome sequencing procedure was performed following manufacturer recommendations (SureSelect Human Exon Kit v5, 75 MB; Agilent). The generated reads were mapped to reference genome hg19 (GRCh37). DNA from adenoma associated colonic mucosa (punches) were purified with Quick-DNA/RNA TM FFPE Kit, according to the manufacturer's instructions (Zymo Research). dMMR crypts were microdissected with LMD7000 Laser (Leica) assisted by IHC (hMSH2 1/125 FE11 clone, Calbiochem), as matched colonic mucosa. Their concentrations were determined by the Qubit Fluorometer (Qubit® dsDNA HS Assay Kits). Quality and integrity of the nucleic acids were analyzed with the TapeStation 2200 (Agilent). The whole exome sequencing procedure was performed following manufacturer recommendations (KAPA HyperExome 45Mb, Roche, Illumina NovaSeq 6000 S1 Reagent Kit - 2x100 cycles). For microdissected tissue we used Unique Molecular Identifier (UMI), following manufacturer recommendations. The generated reads were trimmed using fastp with a minimum read length of 15 and mapped to reference genome hg17 (GRCh37) using bwa v0.7.17.

Variant calling

Reading of Single Nucleotide Variations (SNV) and small insertions/deletions was performed using MuTect tool 2.0 [51]. The somatic mutations were filtered as follows: Total Reads > 10, Somatic score > 3, Mutated Allele Frequency in Tumor tissue \geq 5%, Mutated Allele Count in Tumor tissue \geq 3, Mutated Allele Frequency in Constitutional tissue < 4%. A somatic mutation observed at the donor (GU) or acceptor (AG) site is considered to affect splicing if it induces a delta of PSI >0.1.

525 Microsatellite mutations

Mononucleotide MS with length greater or equal to 5 bp were considered if they were covered with at least 10 mapping reads in both normal and tumor paired samples. MSIsensor 0.5 was initially used to generate the read count distribution for each MS. The total number of reads covering each MS was normalized (arbitrary value of 100) in tumor and healthy matched tissue. For each MS, the normalized value of reads in healthy tissue was subtracted from the tumor tissue. If the ratio difference between normal and tumor was > 10% for deletion the MS was identified as mutated. The deletion size considered for each MS corresponds to the largest deletion value with a ratio of difference between normal and tumor > 10%. Prediction of branch site was performed using SVM-BPfinder using default parameters [52].

535

Minigene construction and cell transfection

Genomic fragments encompassing the putative cassette exons were synthesized (Genscript, NJ 08854, USA), sequence verified and inserted at the unique EcoRI site in the pSPLICE minigene plasmid [30] in the 5'>3' orientation and sequence-verified. The mutant sequence, as identified from our exome sequencing data, was then produced by site-directed mutagenesis. The plasmids were amplified and purified with the endotoxin-free plasmid prep kit (Macherey Nagel, IRDT, France) and used to transfect three gastric cancer cell lines (AGS (MSS), HGT-1 (MSS) and SNU-I (MSI)) using the Lipofectamine 2000 reagent according to the manufacturer's instructions (Invitrogen, Illkirch, France). Cells were harvested after 24 hours and total RNA was extracted with the NucleoSpin RNA prep kit (Macherey Nagel, IRDT, France).

545

Functional analysis

550 Genes were analyzed using Enrichr in order to analyze for the enrichment of genes belonging to signaling pathways (BioPlanet 2019) [53] (two sided P -value $< .05$). The prognostic significance of candidate gene expression and the availability of drugs was evaluated with Human protein atlas [37, 38].

RT-PCR analysis

555 Total RNA (1-2 μ g) from transfected cells was used for first-strand cDNA synthesis using MuLV reverse transcriptase and random primers (NEB Biolabs, Boston, USA). PCR was performed with 10 ng cDNA using the Forward 5'-CAACTTCAAGCTCCTAAGCCACTGC-3' and Reverse 5'-TAGGATCCGGTCACCAGGAAGTTGGTTAAATCA-3' PolR2G primers to analyze cassette transcripts produced from the minigene constructs. RT-PCR products were then gel analyzed.

560

Survival analysis

Survival analysis were performed using the R package 'survival' and overall survival (OS) censored at 5-years as the survival end point. Univariate models were independently computed using Cox proportional hazards regression, with a two sided P -value cutoff of .05 to consider the statistical significance of exon skipping. Genes common between the U2AF-exon signature and our MSI-exon signature and for which gene expression is prognostic (significantly enriched in liver, pancreas and renal) were evaluated in the survival analysis.

565

Single cell analysis

570 Single cell RNA-seq (scRNA-seq) data was obtained from the Pelka et al study [26] accessed via dbGaP (phs002407.v1.p1). Raw FASTQ files were aligned to the GRCh38 reference genome using Cell Ranger v2.1.1, generating BAM files. These BAM files were used as input for STARsolo (bioRxiv, 2021.2005.2005.442755), part of STAR v2.7.8a, to generate gene expression count matrices. SingCellaR package [54] was then used to filter for high quality cells based on UMI counts and number of detected genes. The preprocessed data was further analyzed using Scanpy [55] for dimensionality reduction with uniform manifold approximation and projection (UMAP) [56] and plotting.

575

580 Cell types were identified using Census (<https://doi.org/10.1101/2022.10.19.512926>), which implements a collection of hierarchically organized gradient-boosted decision tree models for cell type classification. Specifically, we used Census to identify cancer cells. Additionally, we used CellTypist [28] and its pre-trained '*Cells_Intestinal_Tract.pkl*' model for automated cell type prediction based on gene expression profiles.

585

Splice junction counts were generated from the FASTQ files using STARsolo. The junction counts and reference GTF file were input to MARVEL package [29] to create an R object for downstream analyses. MARVEL was used to annotate junctions, filter unannotated and multi-mapped junctions, and estimate percent spliced-in (PSI) values directly from junction reads.

590

Finally, we performed splicing analysis comparing PSI values between pseudo-bulk samples using MARVEL's permutation-based approach [57] to compare transit amplifying (TA) cells based on MSI/MSS phenotype.

590

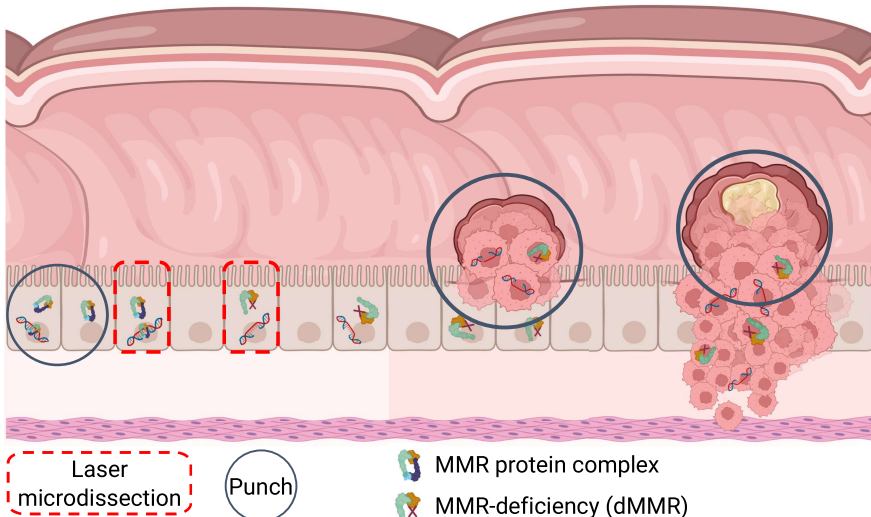
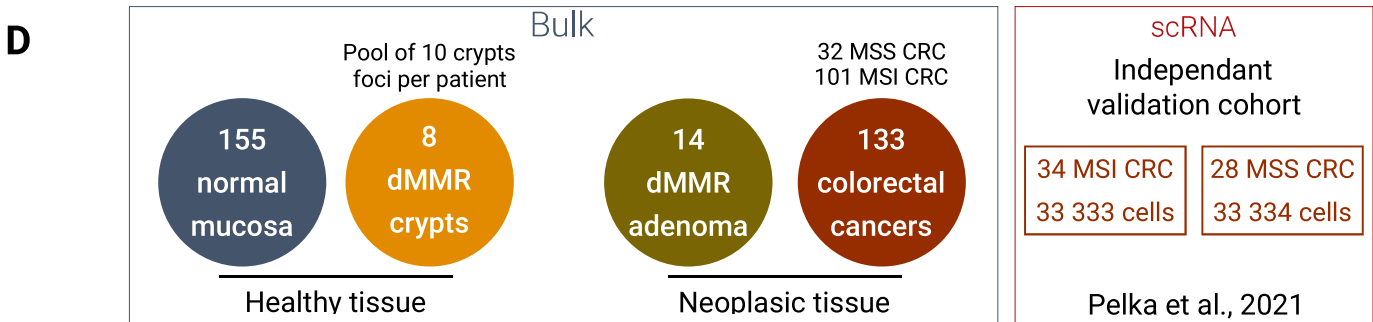
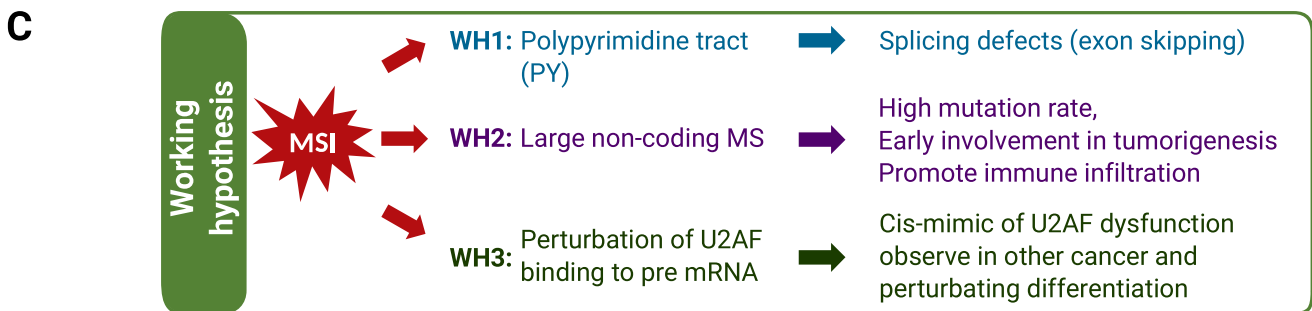
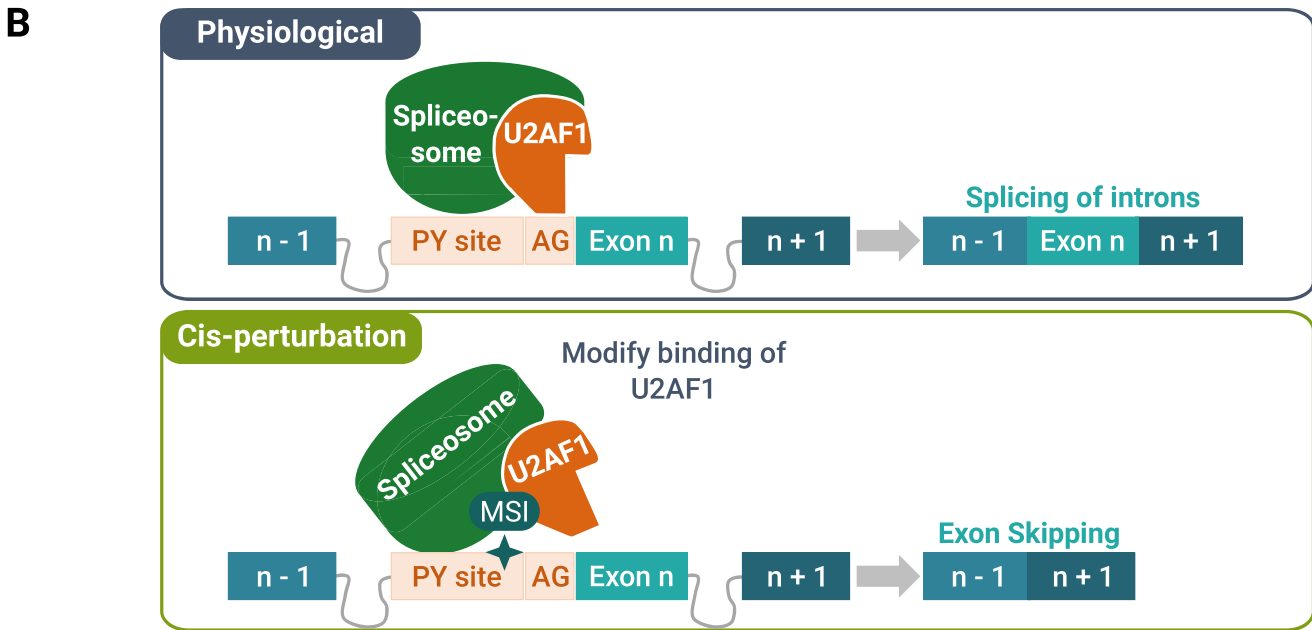
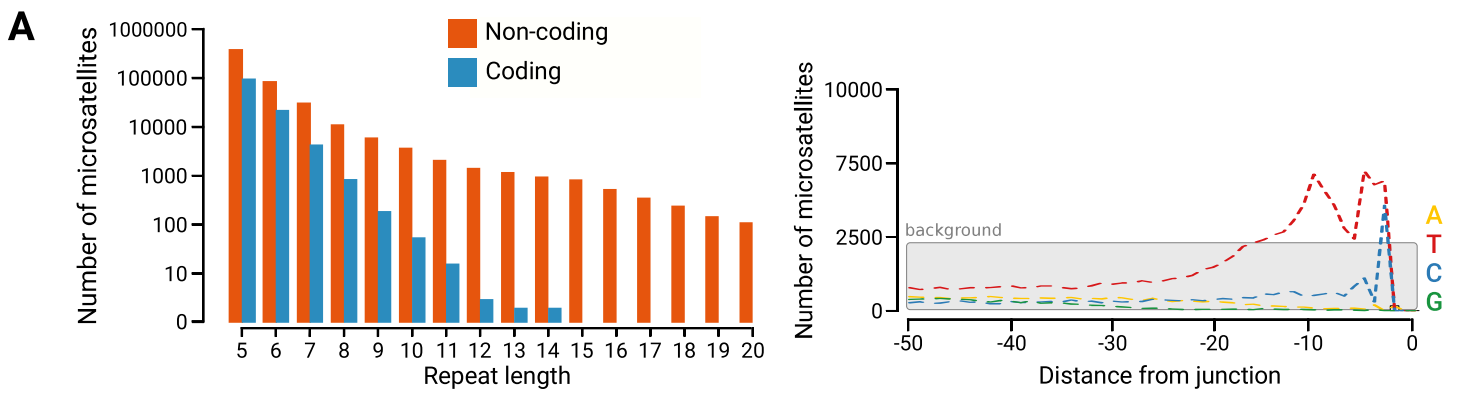
REFERENCES

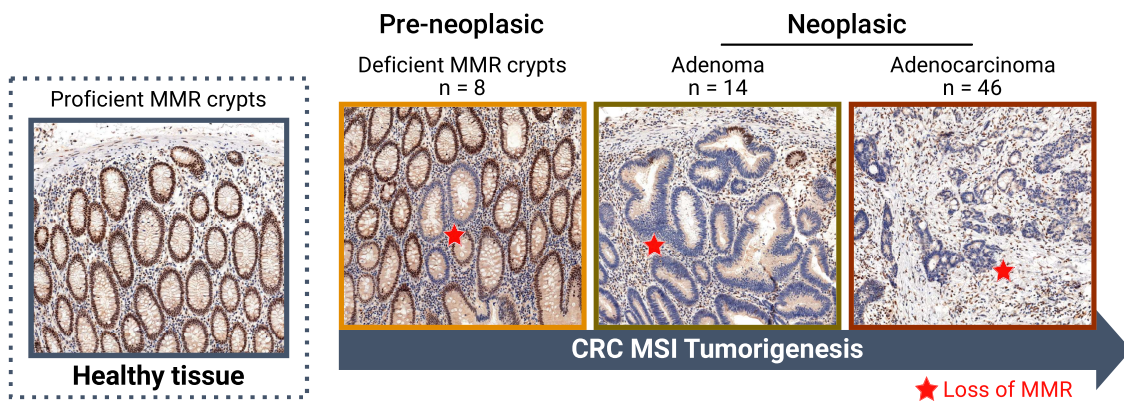
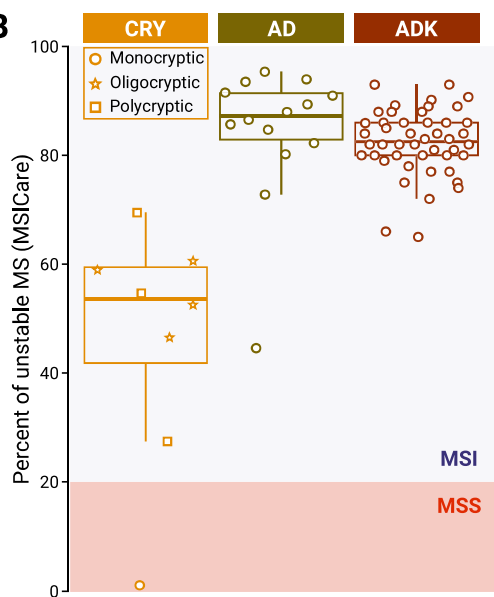
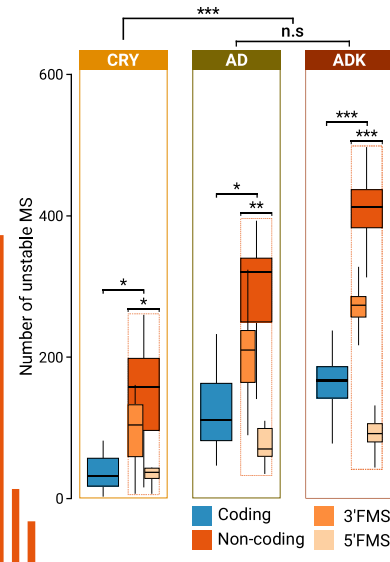
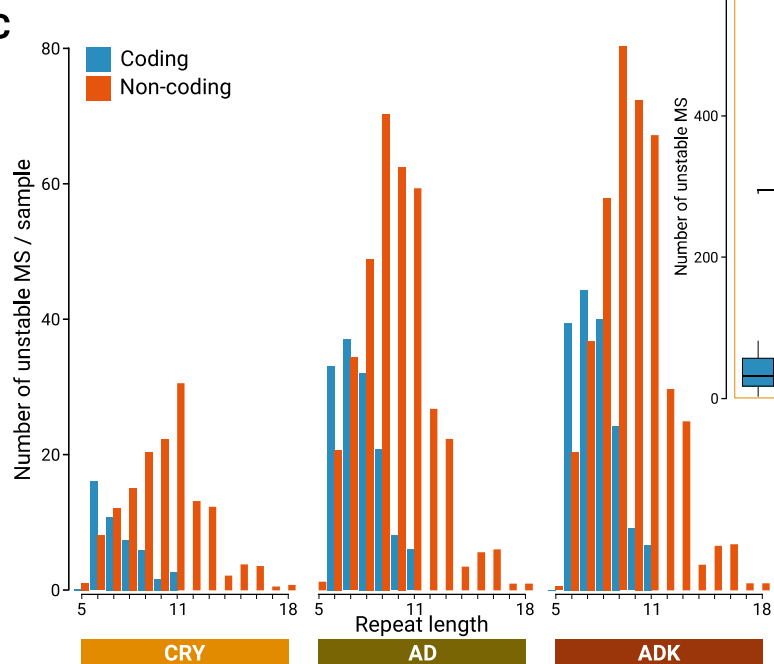
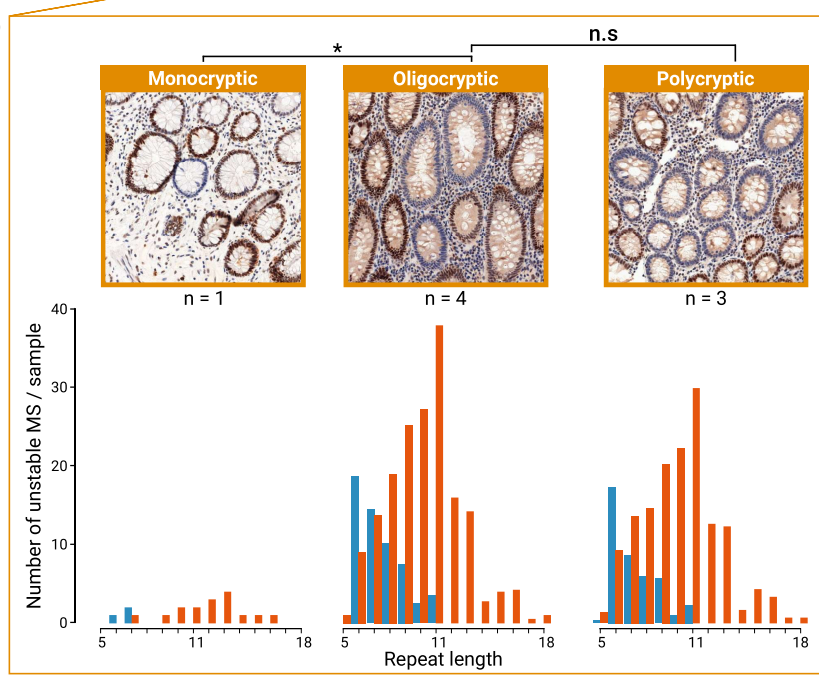
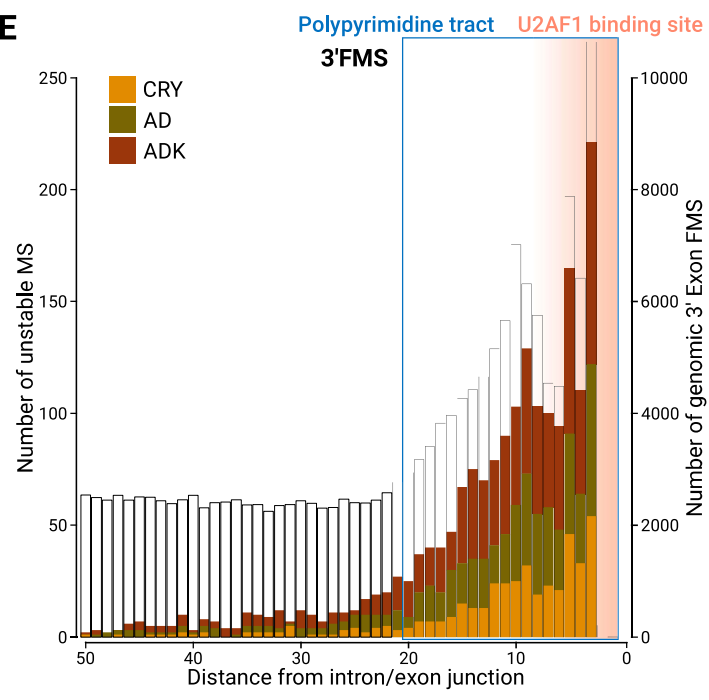
1. Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation.** *Cell* 2011, **144**:646-674.
- 595 2. Markowitz S, Wang J, Myeroff L, Parsons R, Sun L, Lutterbaugh J, Fan RS, Zborowska E, Kinzler KW, Vogelstein B, et al.: **Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability.** *Science* 1995, **268**:1336-1338.
3. Rampino N, Yamamoto H, Ionov Y, Li Y, Sawai H, Reed JC, Perucho M: **Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype.** *Science* 1997, **275**:967-969.
- 600 4. Aaltonen LA, Peltomaki P, Leach FS, Sistonen P, Pylkkanen L, Mecklin JP, Jarvinen H, Powell SM, Jen J, Hamilton SR, et al.: **Clues to the pathogenesis of familial colorectal cancer.** *Science* 1993, **260**:812-816.
5. Branch P, Aquilina G, Bignami M, Karran P: **Defective mismatch binding and a mutator phenotype in cells tolerant to DNA damage.** *Nature* 1993, **362**:652-654.
- 605 6. Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J, Kane M, Kolodner R: **The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer.** *Cell* 1993, **75**:1027-1038.
7. Svrcek M, Lascols O, Cohen R, Collura A, Jonchere V, Flejou JF, Buhard O, Duval A: **MSI/MMR-deficient tumor diagnosis: Which standard for screening and for diagnosis? Diagnostic modalities for the colon and other sites: Differences between tumors.** *Bull Cancer* 2019, **106**:119-128.
- 610 8. Jonchere V, Marisa L, Greene M, Virouleau A, Buhard O, Bertrand R, Svrcek M, Cervera P, Goloudina A, Guillerme E, et al: **Identification of Positively and Negatively Selected Driver Gene Mutations Associated With Colorectal Cancer With Microsatellite Instability.** *Cell Mol Gastroenterol Hepatol* 2018, **6**:277-300.
- 615 9. Kim TM, Laird PW, Park PJ: **The landscape of microsatellite instability in colorectal and endometrial cancer genomes.** *Cell* 2013, **155**:858-868.
10. Duval A, Hamelin R: **Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability.** *Cancer Res* 2002, **62**:2447-2454.
- 620 11. Lothe RA, Peltomaki P, Meling GI, Aaltonen LA, Nystrom-Lahti M, Pylkkanen L, Heimdal K, Andersen TI, Moller P, Rognum TO, et al.: **Genomic instability in colorectal cancer: relationship to clinicopathological variables and family history.** *Cancer Res* 1993, **53**:5849-5852.
- 625 12. Marisa L, Svrcek M, Collura A, Becht E, Cervera P, Wanherdrick K, Buhard O, Goloudina A, Jonchere V, Selves J, et al: **The Balance Between Cytotoxic T-cell Lymphocytes and Immune Checkpoint Expression in the Prognosis of Colon Tumors.** *J Natl Cancer Inst* 2018, **110**.
- 630 13. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, Skora AD, Luber BS, Azad NS, Laheru D, et al: **PD-1 Blockade in Tumors with Mismatch-Repair Deficiency.** *N Engl J Med* 2015, **372**:2509-2520.
14. Cohen R, Bennouna J, Meurisse A, Tournigand C, De La Fouchardiere C, Tougeron D, Borg C, Mazard T, Chibaudel B, Garcia-Larnicol ML, et al: **RECIST and iRECIST criteria for the evaluation of nivolumab plus ipilimumab in patients with microsatellite instability-high/mismatch repair-deficient metastatic colorectal cancer: the GERCOR NIPICOL phase II study.** *J Immunother Cancer* 2020, **8**.
- 635

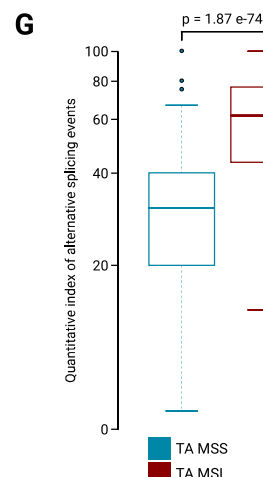
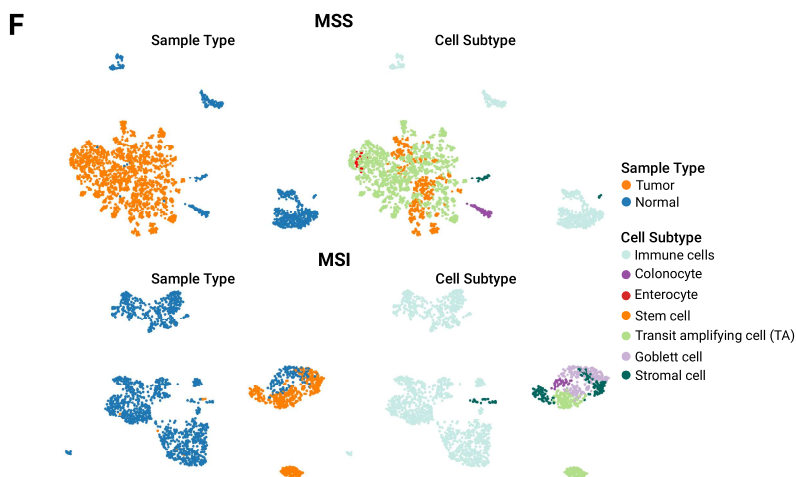
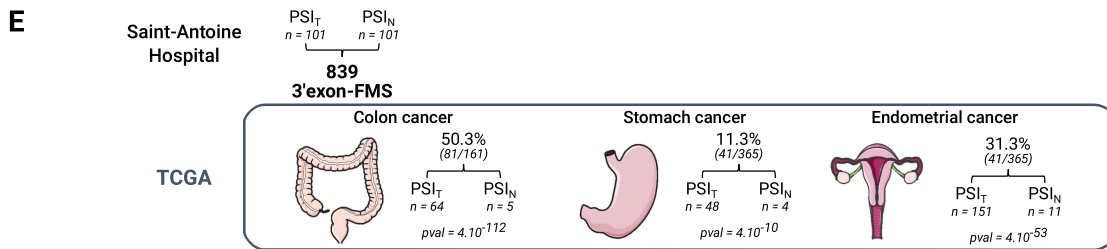
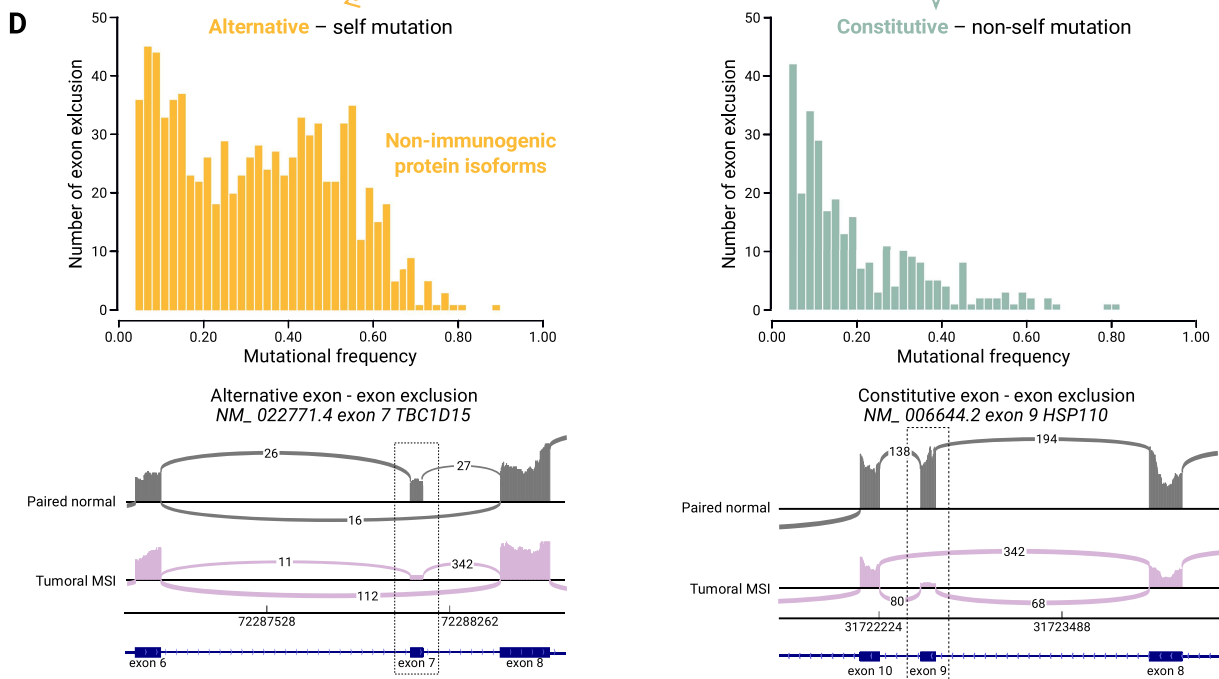
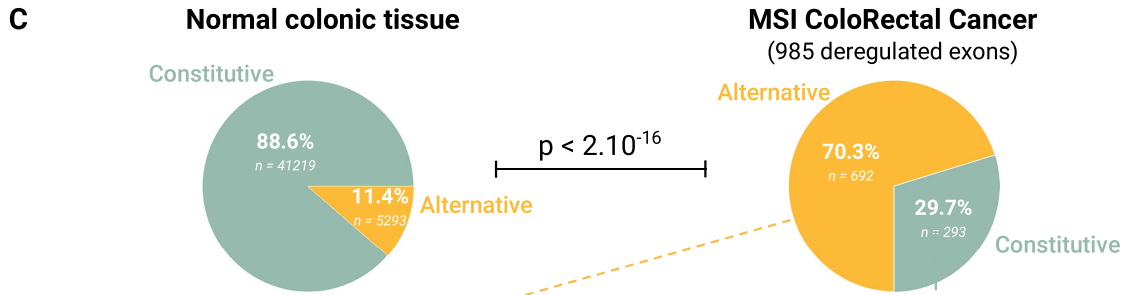
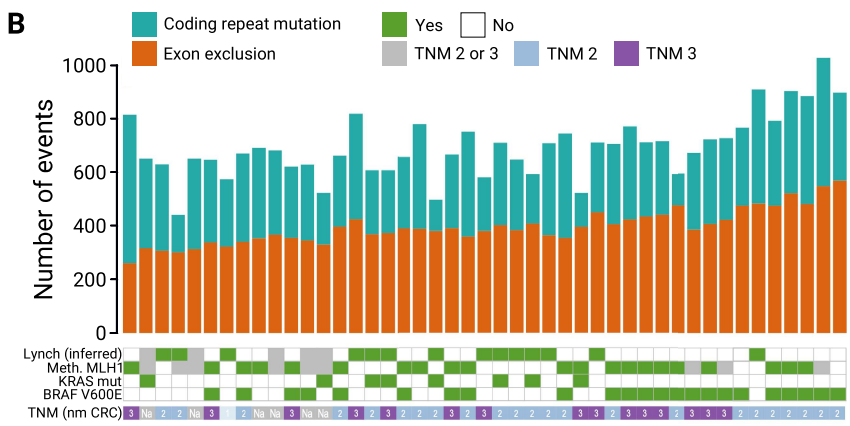
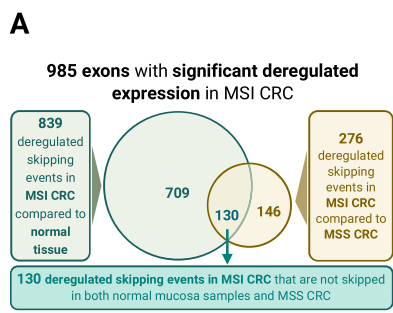
15. Cohen R, Svrcek M, Dreyer C, Cervera P, Duval A, Pocard M, Flejou JF, de Gramont A, Andre T: **New Therapeutic Opportunities Based on DNA Mismatch Repair and BRAF Status in Metastatic Colorectal Cancer.** *Curr Oncol Rep* 2016, **18**:18.
- 640 16. Ratovomanana T, Nicolle R, Cohen R, Diehl A, Siret A, Letourneur Q, Buhard O, Perrier A, Guillerme E, Coulet F, et al: **Prediction of Response to Immune Checkpoint Blockade in Patients with Metastatic Colorectal Cancer with Microsatellite Instability.** *Ann Oncol* 2023.
- 645 17. Baralle FE, Giudice J: **Alternative splicing as a regulator of development and tissue identity.** *Nat Rev Mol Cell Biol* 2017, **18**:437-451.
18. Fiszbein A, Kornblihtt AR: **Alternative splicing switches: Important players in cell differentiation.** *Bioessays* 2017, **39**.
- 650 19. Habowski AN, Flesher JL, Bates JM, Tsai CF, Martin K, Zhao R, Ganesan AK, Edwards RA, Shi T, Wiley HS, et al: **Transcriptomic and proteomic signatures of stemness and differentiation in the colon crypt.** *Commun Biol* 2020, **3**:453.
20. Buhard O, Lagrange A, Guilloux A, Colas C, Chouchene M, Wanherdrick K, Coulet F, Guillerme E, Dorard C, Marisa L, et al: **HSP110 T17 simplifies and improves the microsatellite instability testing in patients with colorectal cancer.** *J Med Genet* 2016, **53**:377-384.
- 655 21. Qu H, Wang Z, Zhang Y, Zhao B, Jing S, Zhang J, Ye C, Xue Y, Yang L: **Long-Read Nanopore Sequencing Identifies Mismatch Repair-Deficient Related Genes with Alternative Splicing in Colorectal Cancer.** *Dis Markers* 2022, **2022**:4433270.
22. Dorard C, de Thonel A, Collura A, Marisa L, Svrcek M, Lagrange A, Jego G, Wanherdrick K, Joly AL, Buhard O, et al: **Expression of a mutant HSP110 sensitizes colorectal cancer cells to chemotherapy and improves disease prognosis.** *Nat Med* 2011, **17**:1283-1289.
- 660 23. Giannini G, Rinaldi C, Ristori E, Ambrosini MI, Cerignoli F, Viel A, Bidoli E, Berni S, D'Amati G, Scambia G, et al: **Mutations of an intronic repeat induce impaired MRE11 expression in primary human cancer with microsatellite instability.** *Oncogene* 2004, **23**:2640-2647.
- 665 24. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al: **Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.** *Nucleic Acids Res* 2016, **44**:W90-97.
- 670 25. Romanelli MG, Diani E, Lievens PM: **New insights into functional roles of the polypyrimidine tract-binding protein.** *Int J Mol Sci* 2013, **14**:22906-22932.
26. Pelka K, Hofree M, Chen JH, Sarkizova S, Pirl JD, Jorgji V, Bejnood A, Dionne D, Ge WH, Xu KH, et al: **Spatially organized multicellular immune hubs in human colorectal cancer.** *Cell* 2021, **184**:4734-4752 e4720.
- 675 27. Schafer S, Miao K, Benson CC, Heinig M, Cook SA, Hubner N: **Alternative Splicing Signatures in RNA-seq Data: Percent Spliced in (PSI).** *Curr Protoc Hum Genet* 2015, **87**:11 16 11-11 16 14.
28. Dominguez Conde C, Xu C, Jarvis LB, Rainbow DB, Wells SB, Gomes T, Howlett SK, Suchanek O, Polanski K, King HW, et al: **Cross-tissue immune cell analysis reveals tissue-specific features in humans.** *Science* 2022, **376**:eabl5197.
- 680 29. Wen WX, Mead AJ, Thongjuea S: **MARVEL: an integrated alternative splicing analysis platform for single-cell RNA sequencing data.** *Nucleic Acids Res* 2023, **51**:e29.

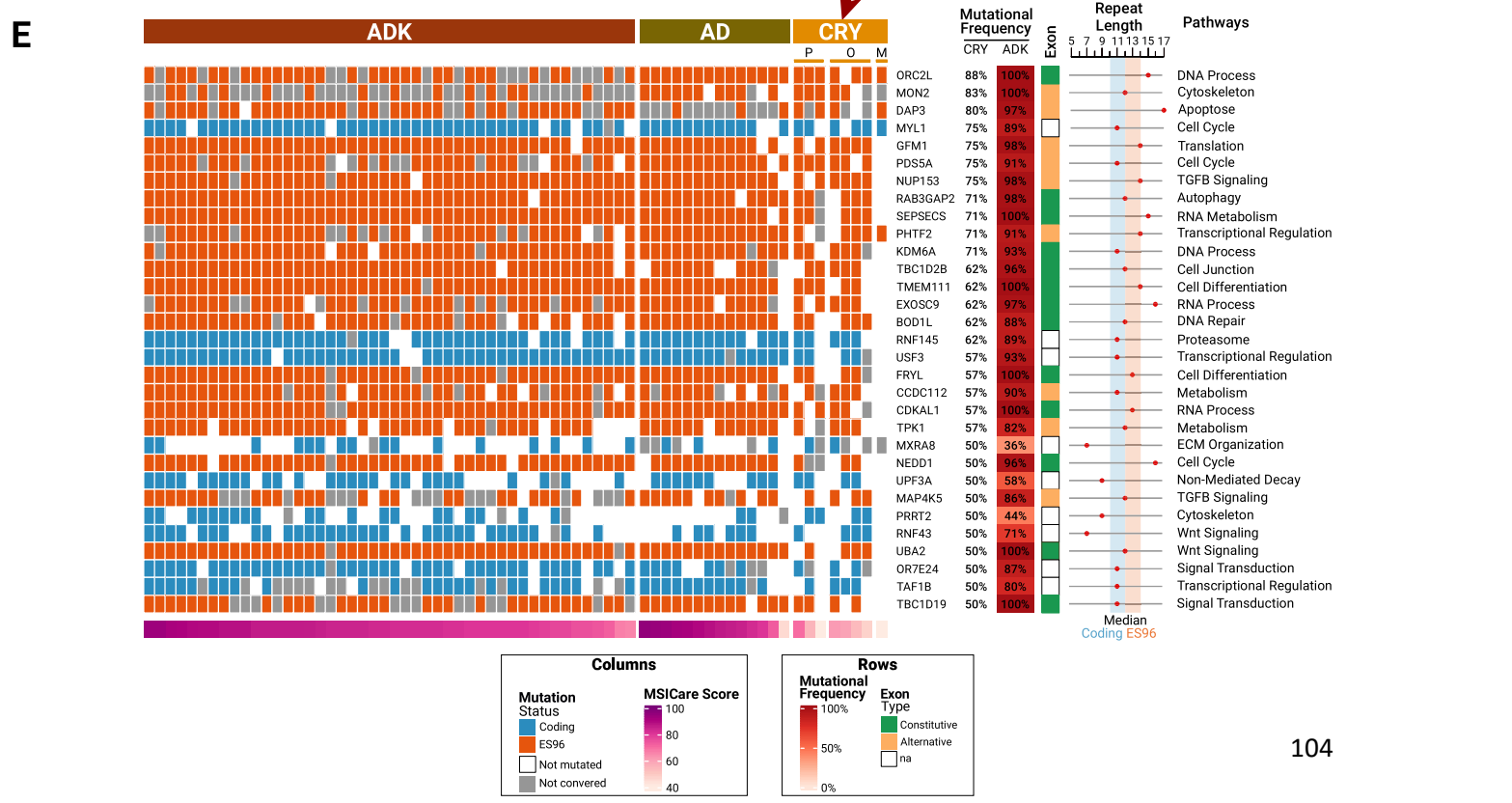
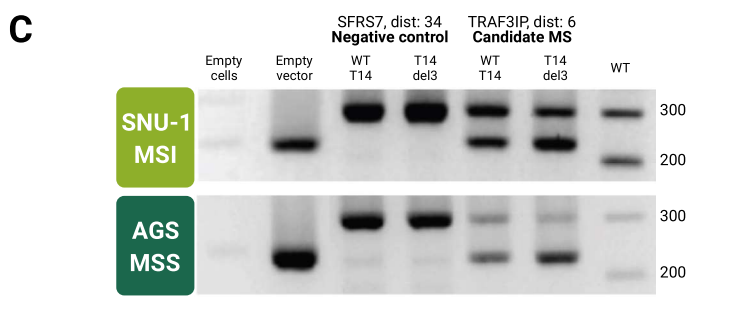
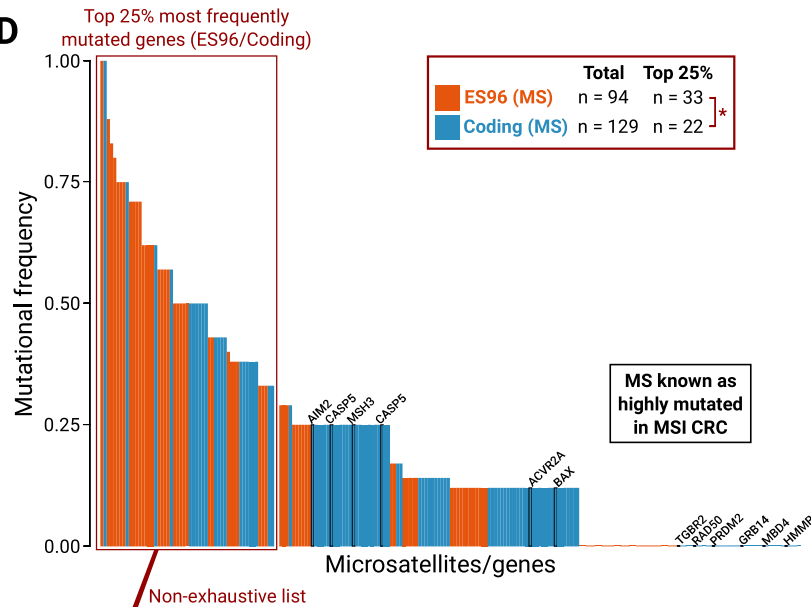
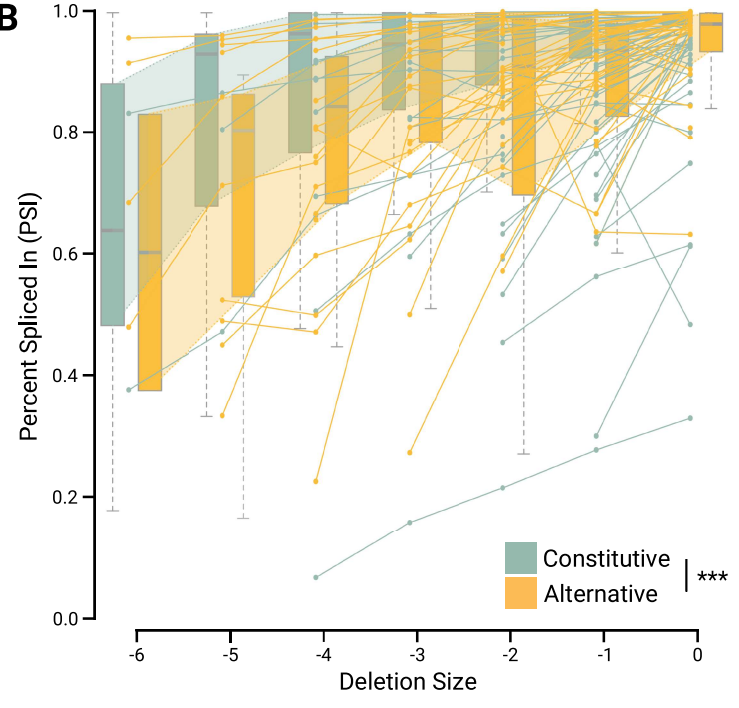
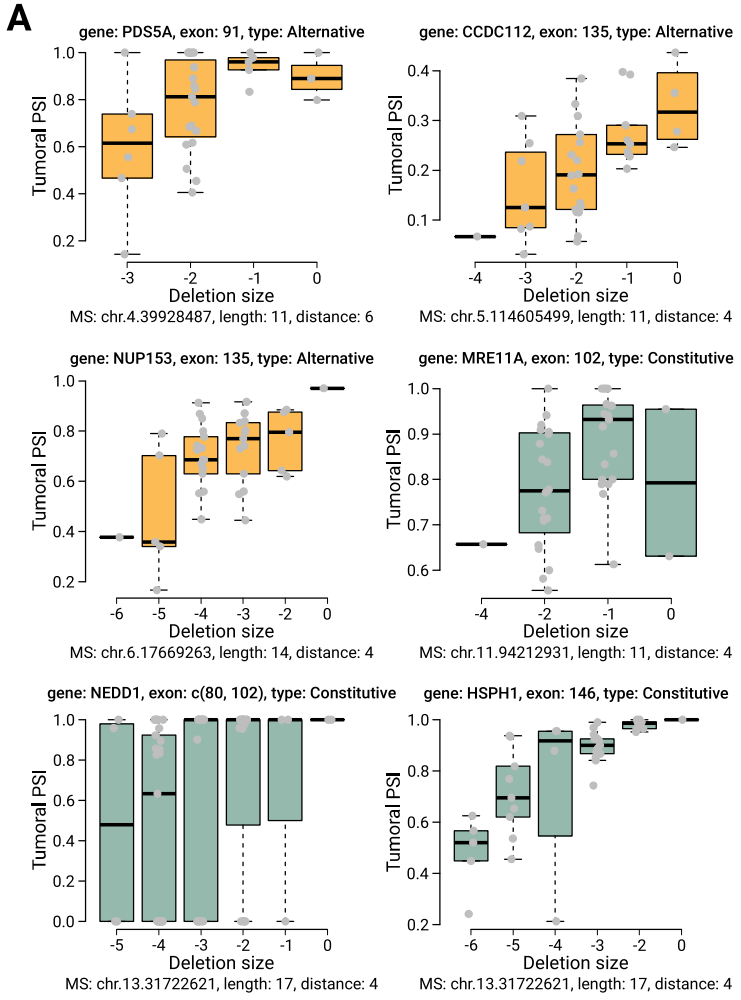
30. Fichou Y, Gehannin P, Corre M, Le Guern A, Le Marechal C, Le Gac G, Ferec C: **Extensive functional analyses of RHD splice site variants: Insights into the potential role of splicing in the physiology of Rh.** *Transfusion* 2015, **55**:1432-1443.
- 685 31. Cohen R, Hain E, Buhard O, Guilloux A, Bardier A, Kaci R, Bertheau P, Renaud F, Bibeau F, Flejou JF, et al: **Association of Primary Resistance to Immune Checkpoint Inhibitors in Metastatic Colorectal Cancer With Misdiagnosis of Microsatellite Instability or Mismatch Repair Deficiency Status.** *JAMA Oncol* 2019, **5**:551-555.
- 690 32. Beumer J, Clevers H: **Cell fate specification and differentiation in the adult mammalian intestine.** *Nat Rev Mol Cell Biol* 2021, **22**:39-53.
33. Bohaumilitzky L, Kluck K, Huneburg R, Gallon R, Nattermann J, Kirchner M, Kristiansen G, Hommerding O, Pfuderer PL, Wagner L, et al: **The Different Immune Profiles of Normal Colonic Mucosa in Cancer-Free Lynch Syndrome Carriers and Lynch Syndrome Colorectal Cancer Patients.** *Gastroenterology* 2022, **162**:907-919 e910.
- 695 34. Shao C, Yang B, Wu T, Huang J, Tang P, Zhou Y, Zhou J, Qiu J, Jiang L, Li H, et al: **Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human genome.** *Nat Struct Mol Biol* 2014, **21**:997-1005.
- 700 35. Brooks AN, Choi PS, de Waal L, Sharifnia T, Imielinski M, Saksena G, Pedamallu CS, Sivachenko A, Rosenberg M, Chmielecki J, et al: **A pan-cancer analysis of transcriptome changes associated with somatic mutations in U2AF1 reveals commonly altered splicing events.** *PLoS One* 2014, **9**:e87361.
36. Ilagan JO, Ramakrishnan A, Hayes B, Murphy ME, Zebari AS, Bradley P, Bradley RK: **U2AF1 mutations alter splice site recognition in hematological malignancies.** *Genome Res* 2015, **25**:14-26.
- 705 37. Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhori G, Benfeitas R, Arif M, Liu Z, Edfors F, et al: **A pathology atlas of the human cancer transcriptome.** *Science* 2017, **357**.
38. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, et al: **Proteomics. Tissue-based map of the human proteome.** *Science* 2015, **347**:1260419.
- 710 39. Yoshida H, Park SY, Sakashita G, Nariai Y, Kuwasako K, Muto Y, Urano T, Obayashi E: **Elucidation of the aberrant 3' splice site selection by cancer-associated mutations on the U2AF1.** *Nat Commun* 2020, **11**:4744.
- 715 40. Hamelin R, Chalastanis A, Colas C, El Bchiri J, Mercier D, Schreurs AS, Simon V, Svrcek M, Zaanani A, Borie C, et al: **[Clinical and molecular consequences of microsatellite instability in human cancers].** *Bull Cancer* 2008, **95**:121-132.
41. Collura A, Lefevre JH, Svrcek M, Tougeron D, Zaanani A, Duval A: **[Microsatellite instability and cancer: from genomic instability to personalized medicine].** *Med Sci (Paris)* 2019, **35**:535-543.
- 720 42. Llosa NJ, Cruise M, Tam A, Wicks EC, Hechenbleikner EM, Taube JM, Blosser RL, Fan H, Wang H, Luber BS, et al: **The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints.** *Cancer Discov* 2015, **5**:43-51.
- 725 43. Graubert TA, Shen D, Ding L, Okeyo-Owuor T, Lunn CL, Shao J, Krysiak K, Harris CC, Koboldt DC, Larson DE, et al: **Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes.** *Nat Genet* 2011, **44**:53-57.

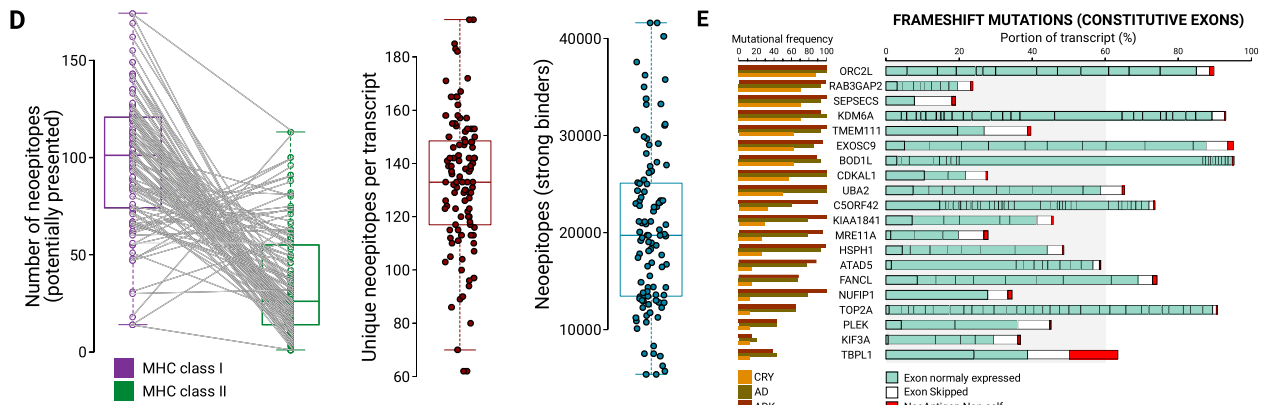
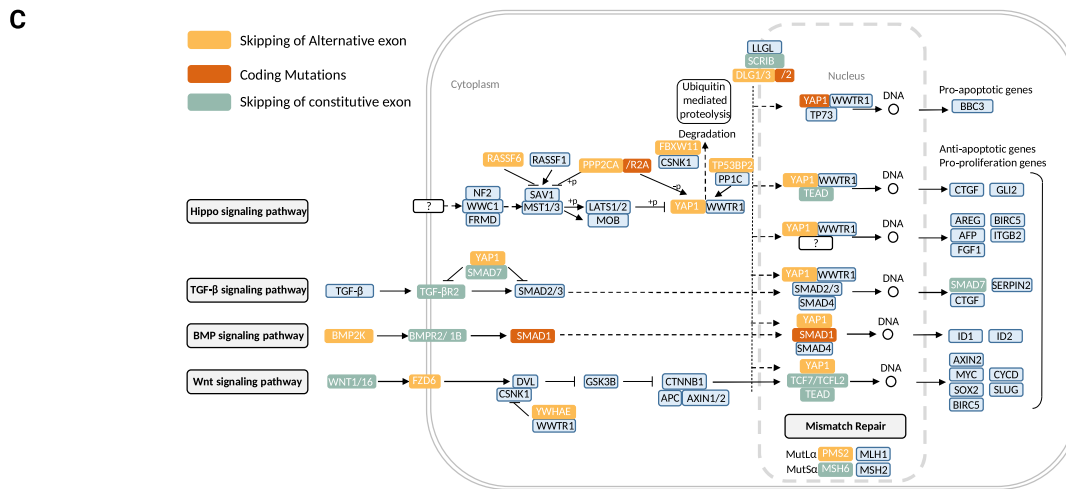
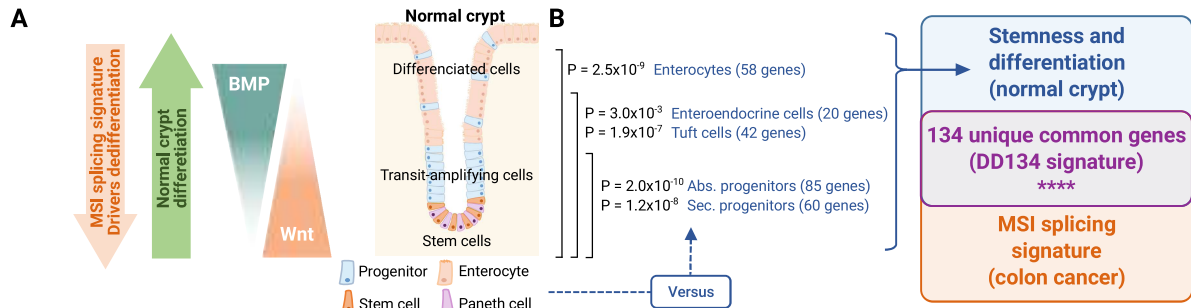
44. Yip BH, Steeples V, Repapi E, Armstrong RN, Llorian M, Roy S, Shaw J, Dolatshad H, Taylor S, Verma A, et al: **The U2AF1S34F mutation induces lineage-specific splicing alterations in myelodysplastic syndromes.** *J Clin Invest* 2017, **127**:2206-2221.
730
45. Tsai YH, VanDussen KL, Sawey ET, Wade AW, Kasper C, Rakshit S, Bhatt RG, Stoeck A, Maillard I, Crawford HC, et al: **ADAM10 regulates Notch function in intestinal stem cells of mice.** *Gastroenterology* 2014, **147**:822-834 e813.
46. Garcia Del Arco A, Edgar BA, Erhardt S: **In Vivo Analysis of Centromeric Proteins Reveals a Stem Cell-Specific Asymmetry and an Essential Role in Differentiated, Non-proliferating Cells.** *Cell Rep* 2018, **22**:1982-1993.
735
47. Lee LJ, Papadopoli D, Jewer M, Del Rincon S, Topisirovic I, Lawrence MG, Postovit LM: **Cancer Plasticity: The Role of mRNA Translation.** *Trends Cancer* 2021, **7**:134-145.
48. Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD, Wendl MC, Ding L: **MSIsensor: microsatellite instability detection using paired tumor-normal sequence data.** *Bioinformatics* 2014, **30**:1015-1016.
740
49. Wang K, Li M, Hakonarson H: **ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.** *Nucleic Acids Res* 2010, **38**:e164.
50. Ratovomanana T, Cohen R, Svrcek M, Renaud F, Cervera P, Siret A, Letourneur Q, Buhard O, Bourgoin P, Guilleme E, et al: **Performance of Next-Generation Sequencing for the Detection of Microsatellite Instability in Colorectal Cancer With Deficient DNA Mismatch Repair.** *Gastroenterology* 2021, **161**:814-826 e817.
745
51. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G: **Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples.** *Nat Biotechnol* 2013, **31**:213-219.
750
52. Corvelo A, Hallegger M, Smith CW, Eyraes E: **Genome-wide association between branch point properties and alternative splicing.** *PLoS Comput Biol* 2010, **6**:e1001016.
53. Bauer K, Nelius N, Reuschenbach M, Koch M, Weitz J, Steinert G, Kopitz J, Beckhove P, Tariverdian M, von Knebel Doeberitz M, Kloor M: **T cell responses against microsatellite instability-induced frameshift peptides and influence of regulatory T cells in colorectal cancer.** *Cancer Immunol Immunother* 2013, **62**:27-37.
755
54. Wang G, Wen WX, Mead AJ, Roy A, Psaila B, Thongjuea S: **Processing single-cell RNA-seq datasets using SingCellaR.** *STAR Protoc* 2022, **3**:101266.
- 760 55. Wolf FA, Angerer P, Theis FJ: **SCANPY: large-scale single-cell gene expression data analysis.** *Genome Biol* 2018, **19**:15.
56. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, Ginhoux F, Newell EW: **Dimensionality reduction for visualizing single-cell data using UMAP.** *Nat Biotechnol* 2018.
- 765 57. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R: **CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes.** *Nat Protoc* 2020, **15**:1484-1506.



A**B****C****D****E**







F 125 shared genes between MSI splicing signature and U2AF-S34F oncogenic signature of sever human tumors

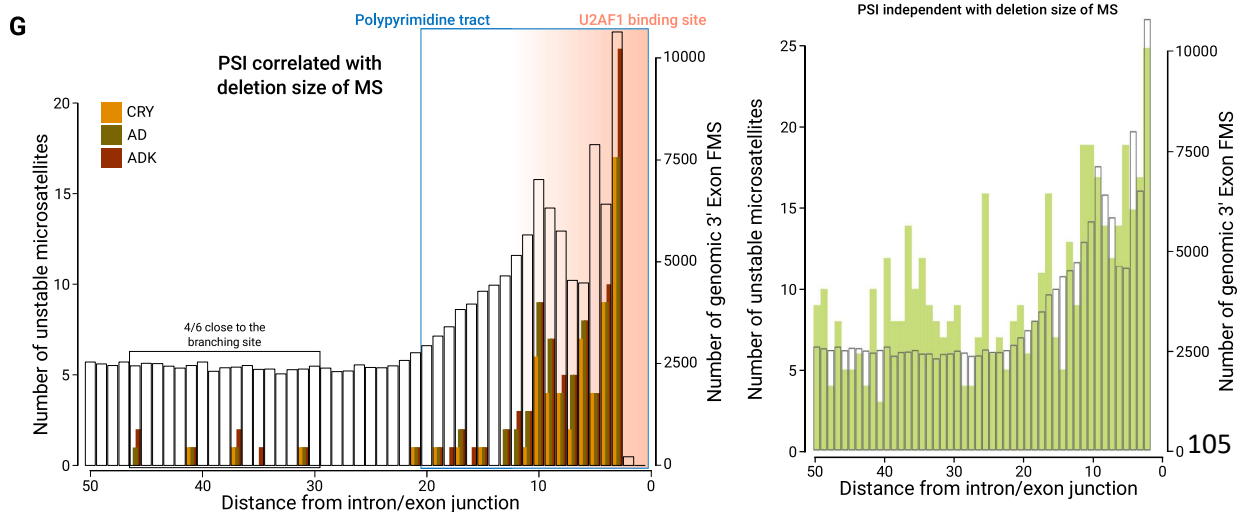
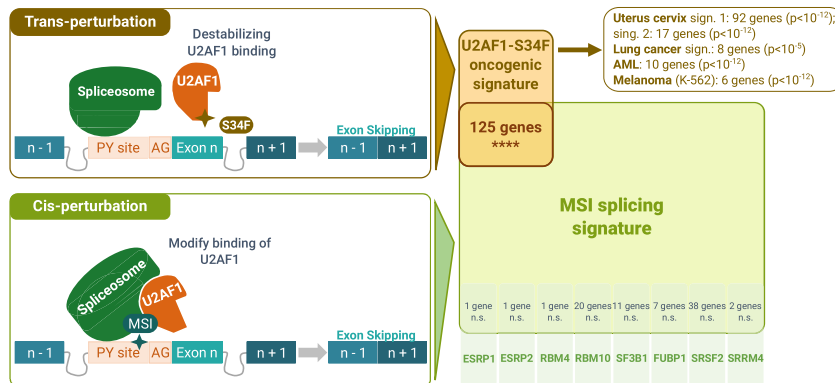


FIGURE LEGENDS

Figure 1. Frequently mutated intronic microsatellites may cause aberrant exon skipping in MSI CRC. (A) **Left panel**, genomic distribution of microsatellite (MS) mutations (log 10 scale) of two gene regions (exonic and intronic) according to repeat length. **Right panel**, distribution of all flanking MS (3' splice acceptor site, 3'FMS) according to their distance from the intron-exon boundary. The distribution of these MS is indicated according to their nucleotide composition (yellow: Adenosine, red: Thymine, blue: Cytosine, green: Guanine). Background corresponds to the distribution of all MS across the human genome. (B) **Upper panel**, schematic representation of physiological splicing. PY: polypyrimidine tract, AG: acceptor site. **Lower panel**, schematic representation of MSI cis-perturbation of splicing. (C) Working hypothesis (WH). (D) Experimental design to investigate splicing cis-perturbation in MSI colorectal cancers (CRC).

Figure 2. MSI-driven mutational events in non-coding regions lead mRNA changes which appears in early stages of MSI tumorigenesis. (A) Histological description of MSI CRC tumorigenesis, from pre-neoplastic lesion (dMMR crypts) to neoplastic steps (adenoma & adenocarcinoma). Staining was performed by immunohistochemistry of MSH2 protein. (B) MSI status for each sample is define by MSICare {Ratovomanana, 2021 #1585}. (C) **Left panel**, number of unstable MS by sample according to their length and considering their genomic position (coding or non-coding). (C) **Right panel**, Boxplot represent total number of unstable MS according to their region and condition. Statistic: Student test, * P -value < 0.05, ** P -value < 1.10^{-3} , *** P -value < 1.10^{-4} , **** P -value < 1.10^{-5} . (D) Loss of MMR in colonic crypt is categorized in three groups, depending on the number of dMMR crypts in a foci: monocryptic (1), oligocryptic (2-4) and polycryptic (> 4). In all the groups, crypts are identical and morphologically healthy (E) Distribution of 3'FMS according to the distance of the 3' splice site (AG); the genomic background distribution of all flanking MS ($n = 335,564$) is indicated (phantom white bars).

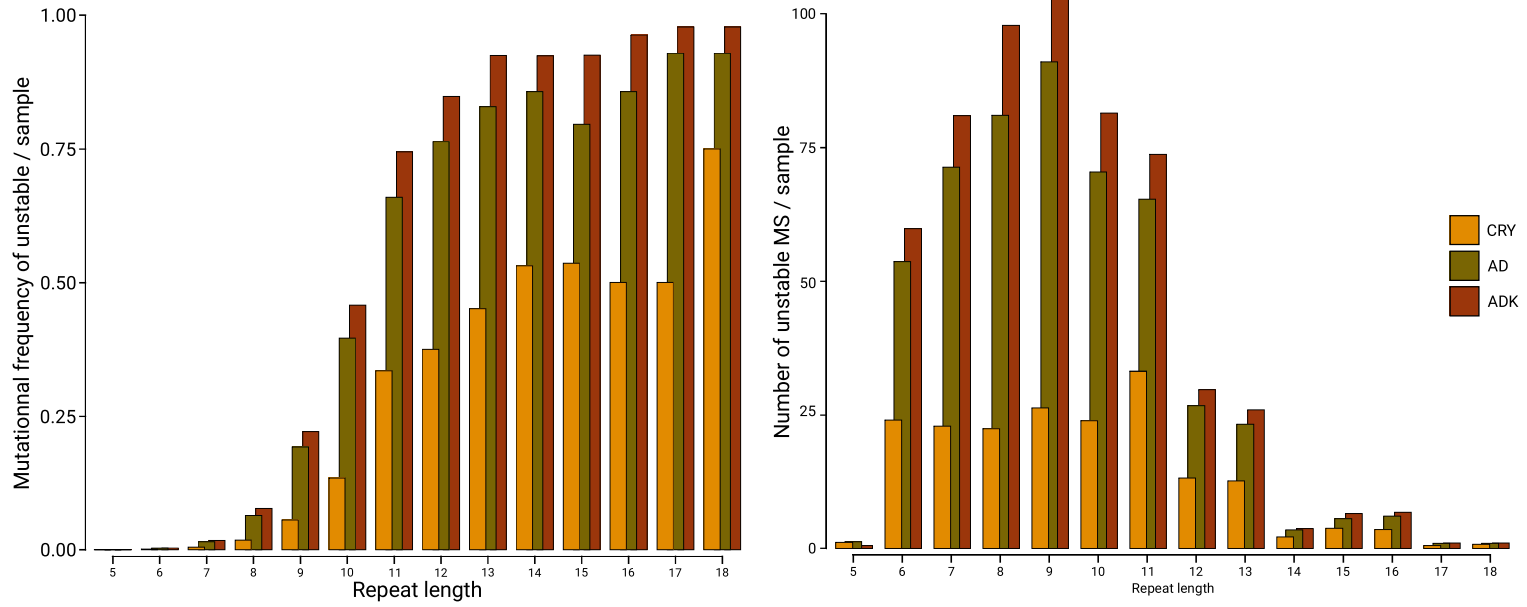
Figure 3. MSI-driven mutational events in non-coding regions lead to skipping of mainly alternative regulated exons in MSI CRC. (A) **Left panel**, Venn diagram showing the overlap between exon skipping events in MSI CRC compared to normal colonic tissue and to MSS CRC. (B) Barplot displaying the number of mutations in coding MS (expected frameshift mutations) and the number of skipped exons across the whole exome for 46 non-metastatic MSI tumors (TNM stage 2 or 3). (C) **Left panel**, pie chart represents the portion of alternative and constitutive exons in normal colorectal tissue ($n = 57030$ exons, $n = 133$ normal tissue samples). **Right panel**, pie chart represents the portion of the same alternative and constitutive exons whose expression was deregulated in MSI tumors. Results of the Chi-square test between the two pie charts are indicated. (D) Distribution of exon skipping events according to their frequency in MSI tumors. **Left and right panels** correspond respectively to exons classified as constitutive ($n = 293$) or alternative ($n = 692$) in normal colonic tissue. **Bottom panel**, example of IGV-Sashimi plots showing the read coverage for 2 examples of significantly deregulated exons between tumors and normal tissue (**right**, constitutive and **left**, alternative). (E) Percentage overlap of MSI exon skipping between our study and three independent TCGA cohorts. The number of overlapping events and total number of events analyzed are indicated in brackets. P -value of enrichments are indicated. (F) Cells distribution in MSS (**Top panel**) or MSI samples (**Bottom panel**) by tumor or normal status (**Left**) and cell subsets (**Right**). (G)

Quantitative index of alternative splicing events in transit amplifying cells (TA) in MSS or MSI tumors.

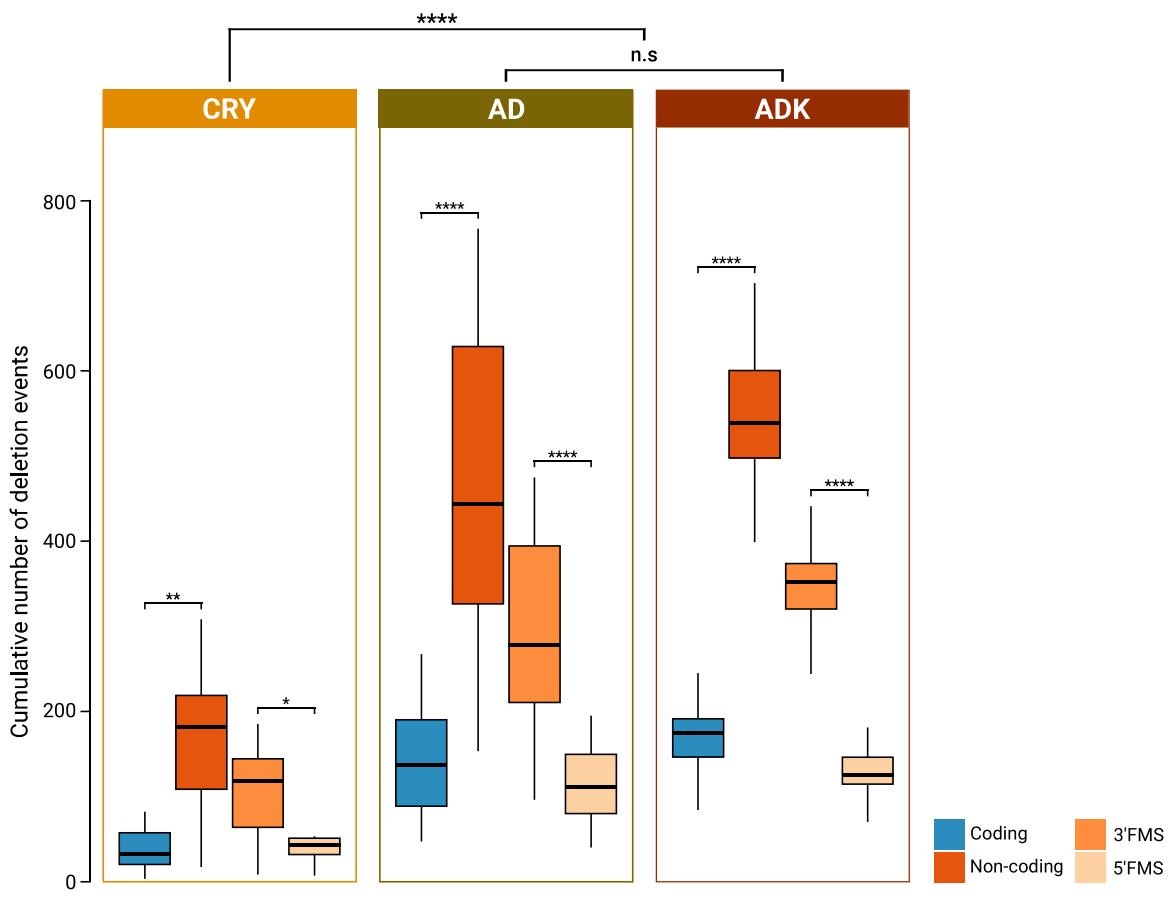
Figure 4: Splicing anomalies found in MSI CRCs are frequently caused by MSI through mutations that occur very early in cancer development, as early as the untransformed dMMR crypt state. (A) Six examples of genes in which exon skipping is closely related to the status of the DNA microsatellite in the PY at the intronic boundary. (B) Boxplots represent the Percent Spliced Included (PSI) according to deletion size for the 96 MS (ES96). Result of two-sided anova test between exons classified as alternative and those classified as constitutive was as follows: *** P -value < .001. In the background, each MS displaying a significant correlation is indicated by a line and dots. (C) Plasmids containing the mini-genes of interest were transfected into AGS and SNU-1 cancer cells and mini-gene derived transcripts were analyzed by RT-PCR. The figures illustrate results from 3 independent experiments with similar results. A faint band corresponds to the endogenous PolR2G mRNA. A more abundant band of identical size is observed in lanes corresponding to the empty vector that carries the same inserted genomic sequence. (D) Barplots of ES96 and C129 (selection of coding MS covered in at least 50% of the samples and mutated in overall 30% of the lesions, $n = 129$) microsatellites according to their mutational frequency in early stage i.e. dMMR crypts. Top 25% most frequently mutated MS (ES96/Coding) corresponds to the first quartile of mutational frequency. Coding MS known as highly mutated in MSI Colorectal Cancer are indicated in black. ES96 signature is significantly enriched in Top Genes; * P -value < 1.10^{-2} . (E) Non-exhaustive list of most frequently mutated MS (ES96/Coding) in early stage of MSI CRC tumorigenesis (more details in **Supp. Table S6**). MS are order by mutational frequency in dMMR crypts.

Figure 5: mRNA changes impair cell differentiation and mimic the oncogenic U2AF1 inactivation in MSI CRC. (A) Schematic representation of the normal colonic crypt. (B) Overlapping signature of 134 genes shared between our MSI splicing signature and that of Habowski et al. that regroups alternative mRNA changes driving cell differentiation in the normal colonic crypt {Habowski, 2020 #1596}. SC: Stem cells; Abs.: Absorptive; Sec.: Secretary/deep crypt secretory cells/goblet; EEC: Enteroendocrine; Ent: Enterocytes, TuftC: Tuft Cells. (C) Interactions between genes related to coding repeat mutations and/or genes with alterations of splicing on Hippo signaling pathways with a role in CRC and in cell differentiation in the colonic crypt (adapted from KEGG: hsa04390). (D) **Left panel**, number of neoepitopes potentially presented by HLA-MHC Class I or II for each patients ($n = 101$ CRC MSI). **Middle panel**, boxplot displaying the number of unique neoepitopes in average per transcript and per patients. **Right panel**, boxplot displaying the number of neoepitopes showing a high affinity with HLA-MHC Class I or II for each patients. (E) Constitutive exon skipping with frameshift (ES96) exposing neoantigen tail; genes are ordering by mutational frequency in early stage. (F) Overlapping signature of 125 genes containing common exons between the U2AF-S34F signature and the MSI splicing signature. (G) Distribution of MS in intronic boundary (-50 to 0 of intron/exon junction). **Left panel**, the 96 MS (ES96) show a very close localization next to the AG site, especially in U2AF1 binding region in the PY. **Right panel**, the 181 MS for which the PSI value is independent of deletion size are not specifically concentrated in the PY next to AG site. For both panels, the background distribution of all flanking MS ($n = 335,564$) is indicated (grey).

A



B



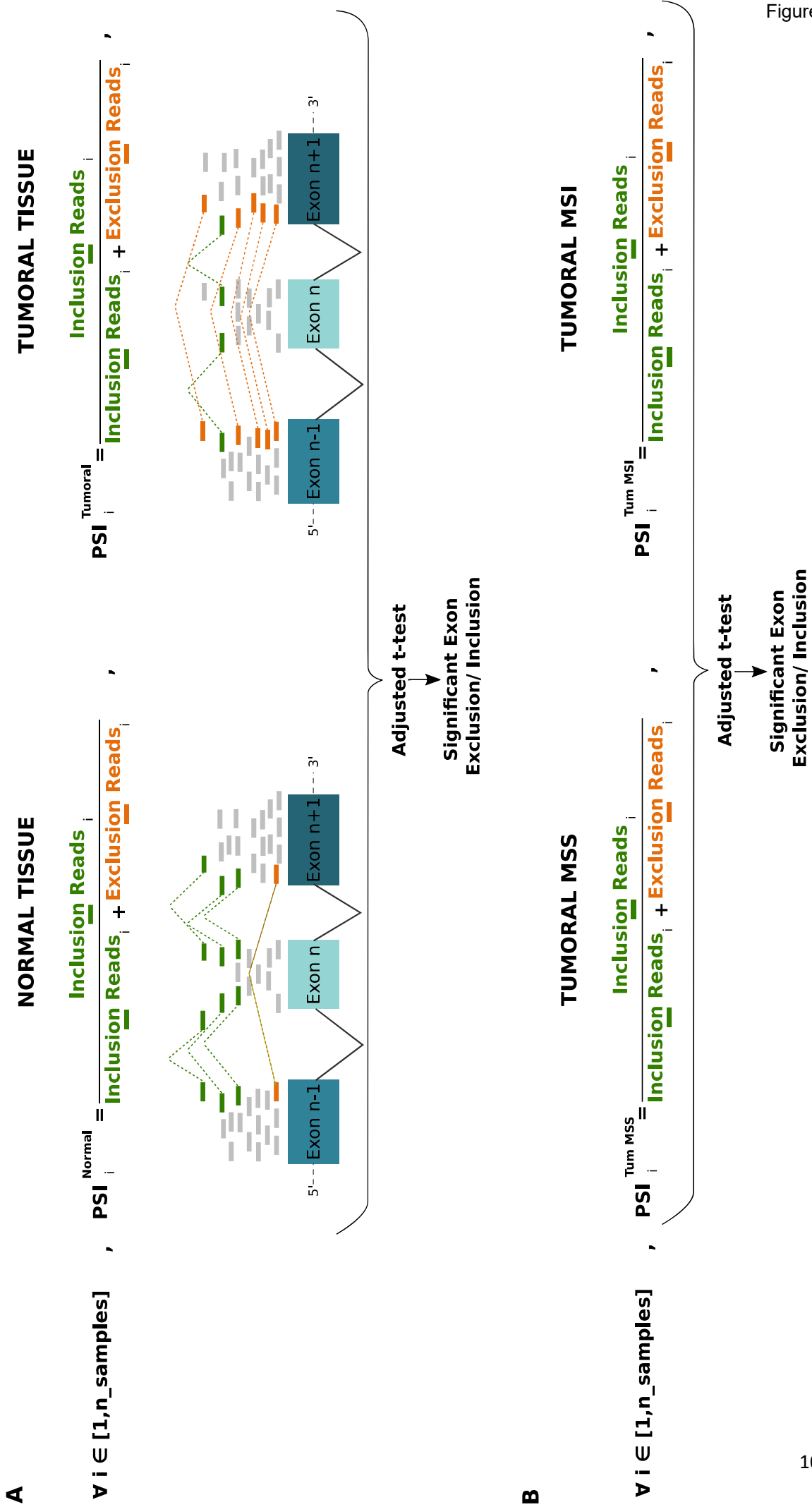
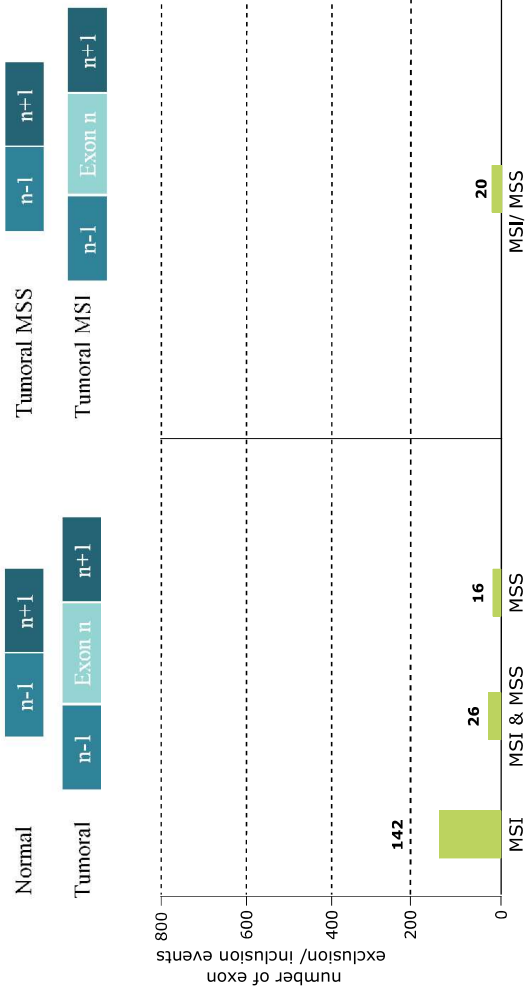
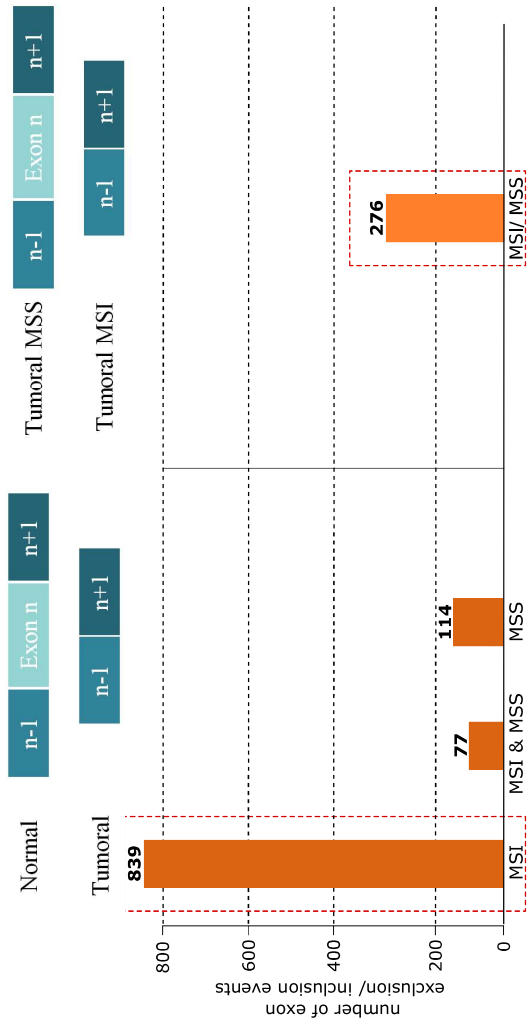


Figure Supp 2

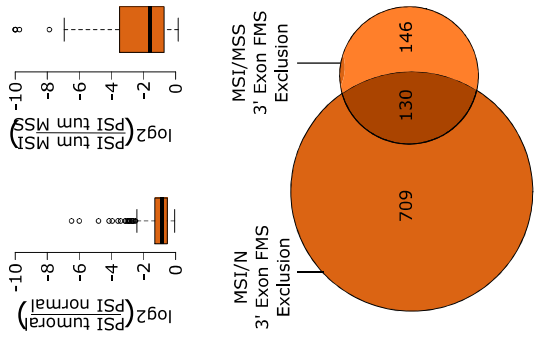
C



A



B



D

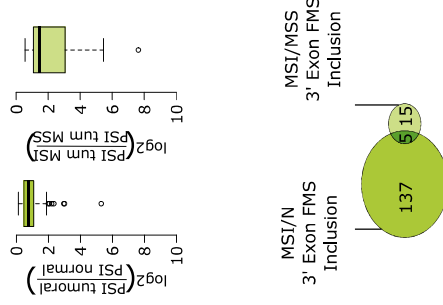
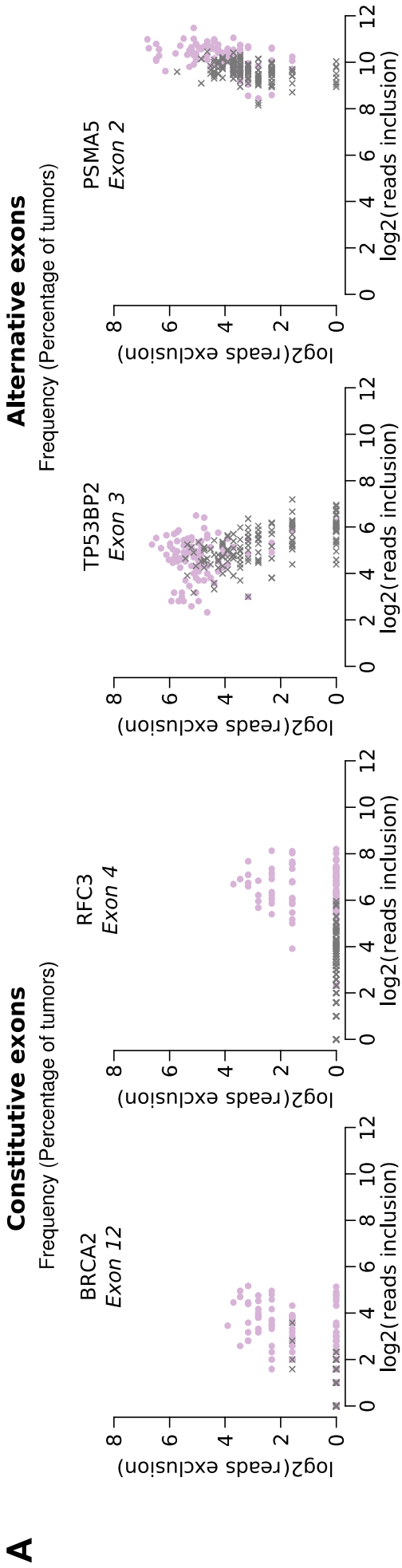
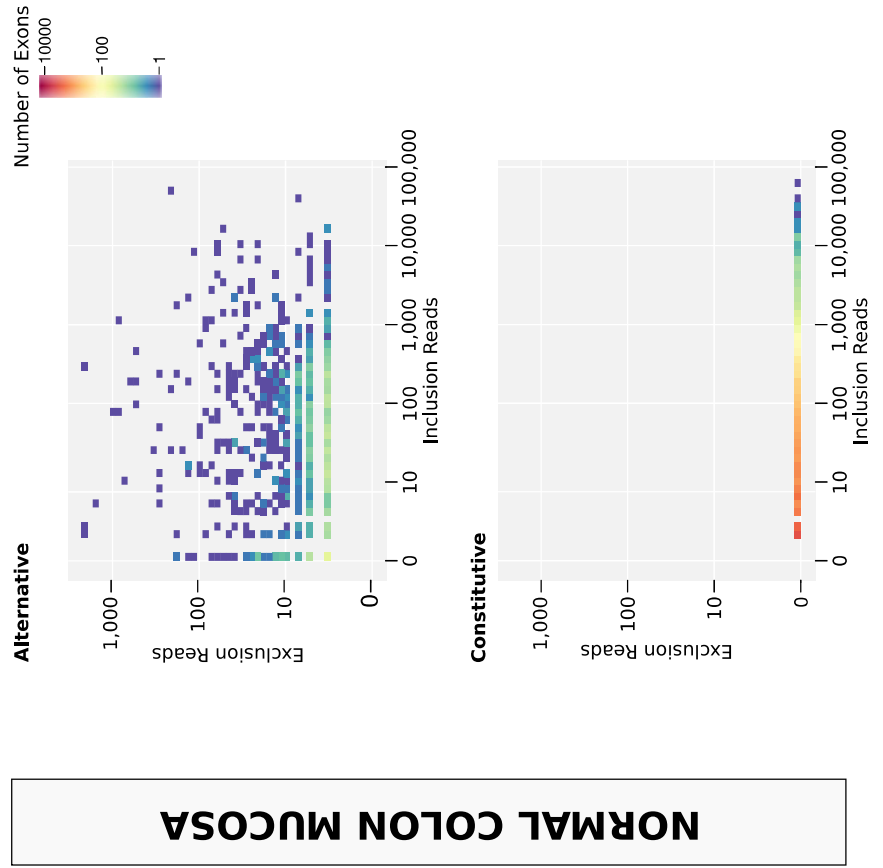
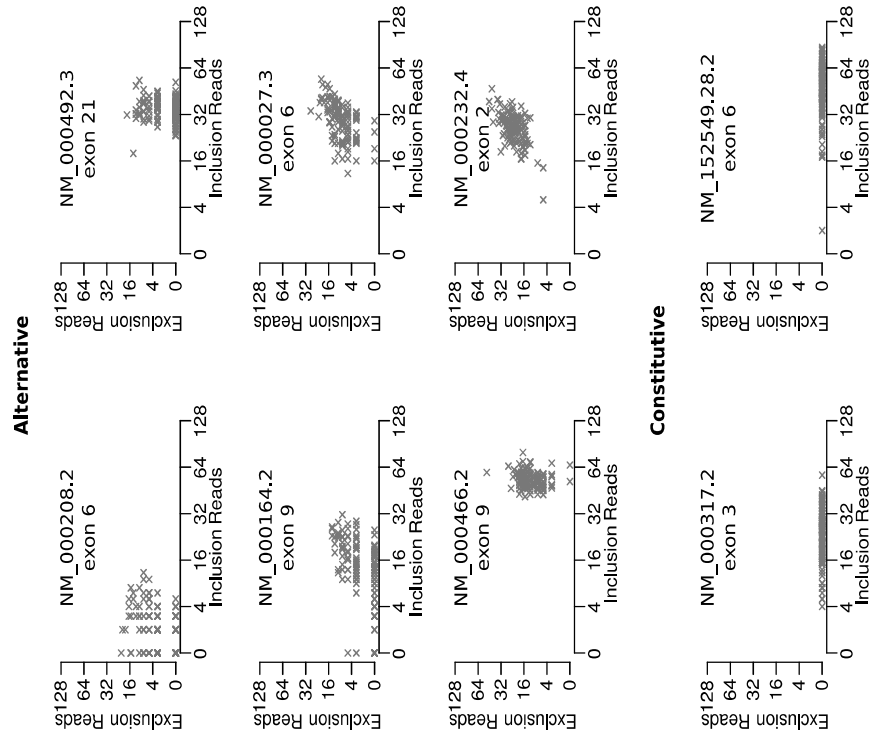


Figure Supp 3

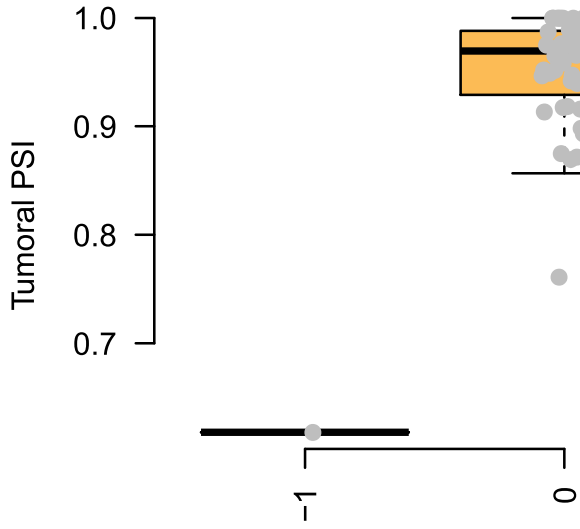


● Paired normal X Tumoral MSI

C

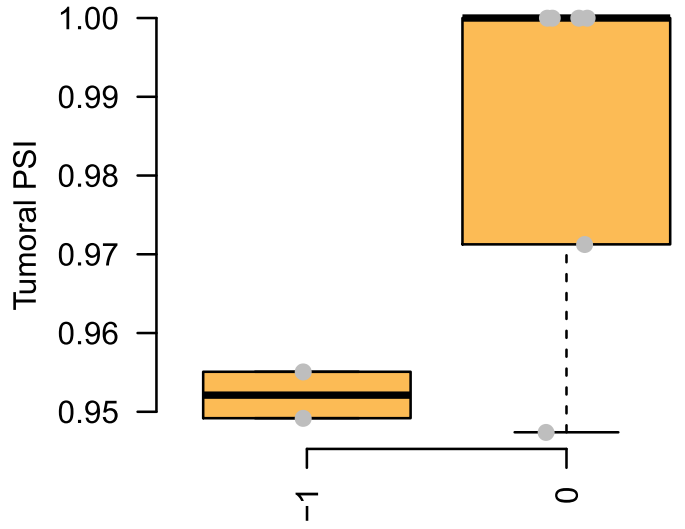


gene: IFRD1 ,exon: 146 ,type: Alternative



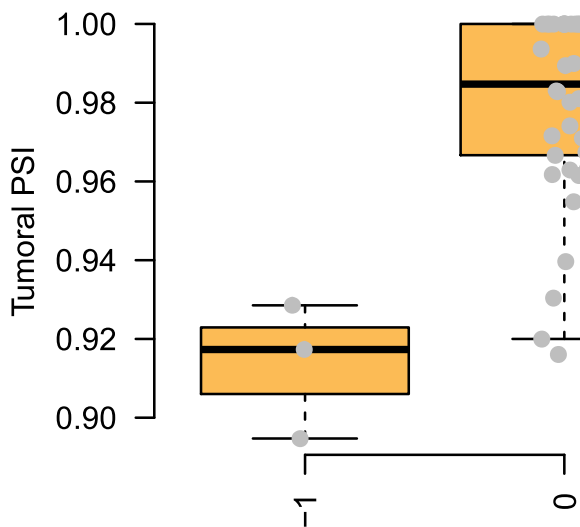
MS: chr.7.112102317 ,length: 6 ,distance: 3

gene: UBXN2B ,exon: 113 ,type: Alternative



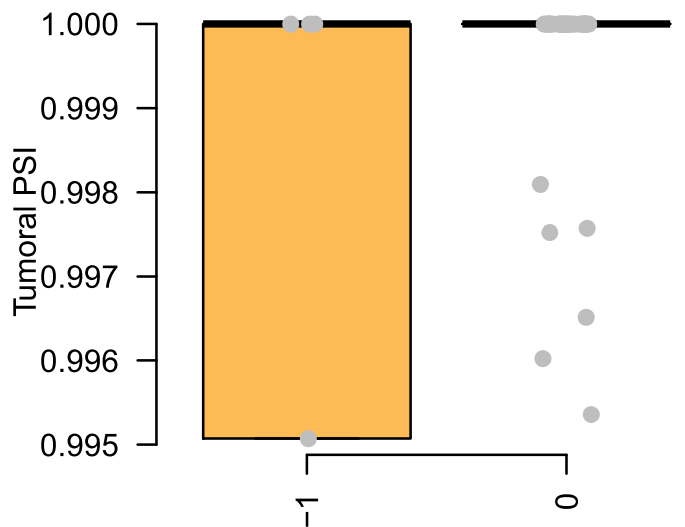
MS: chr.8.59352170 ,length: 6 ,distance: 17

gene: C13orf23 ,exon: 12 ,type: Alternative



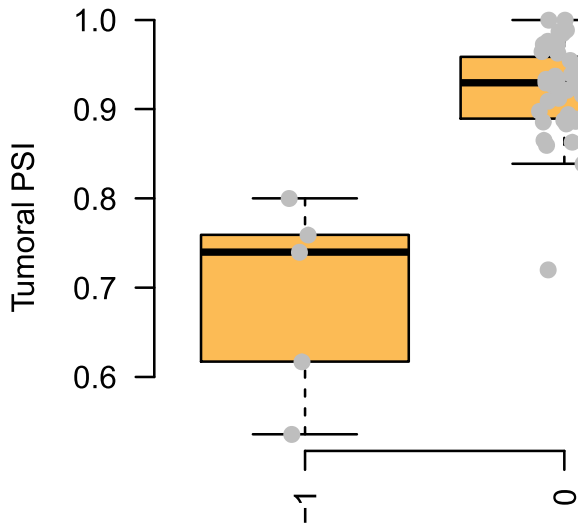
MS: chr.13.39588629 ,length: 6 ,distance: 16

gene: ACAT1 ,exon: 1 ,type: Alternative



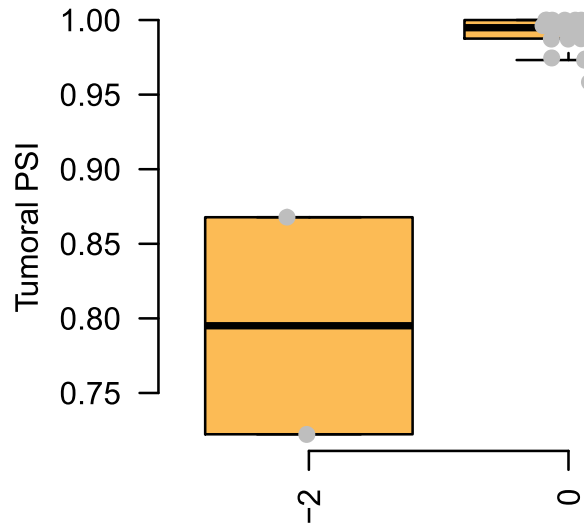
MS: chr.11.108014697 ,length: 7 ,distance: 7

gene: CDC45L ,exon: 91 ,type: Alternative



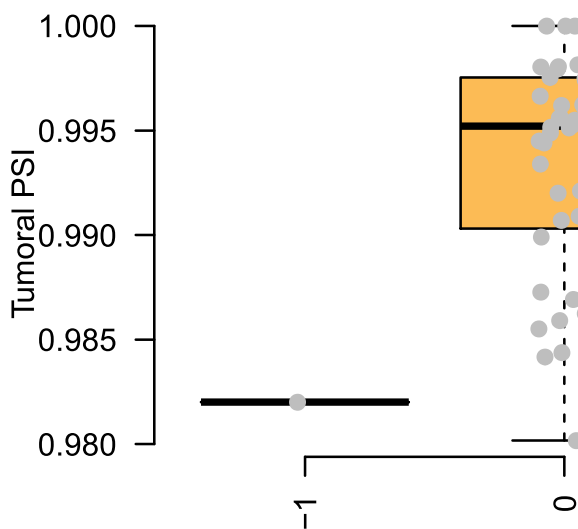
MS: chr.22.19470203 ,length: 7 ,distance: 4

gene: DDX3X ,exon: 80 ,type: Alternative



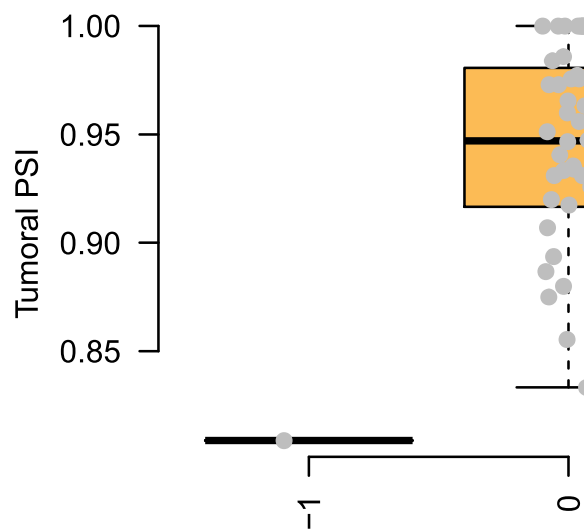
MS: chr.X.41198280 ,length: 7 ,distance: 3

gene: PSMD4 ,exon: 102 ,type: Alternative



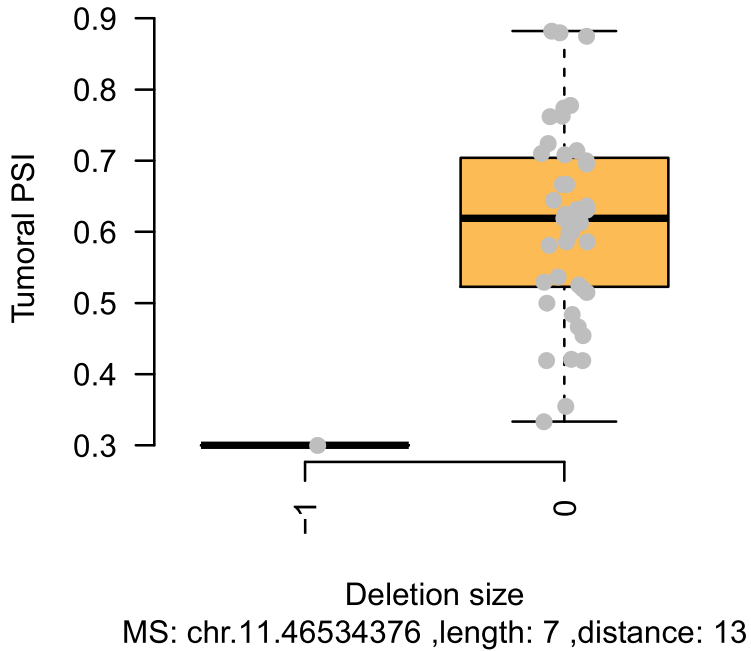
MS: chr.1.151237590 ,length: 7 ,distance: 46

gene: ZNF3 ,exon: 80 ,type: Alternative

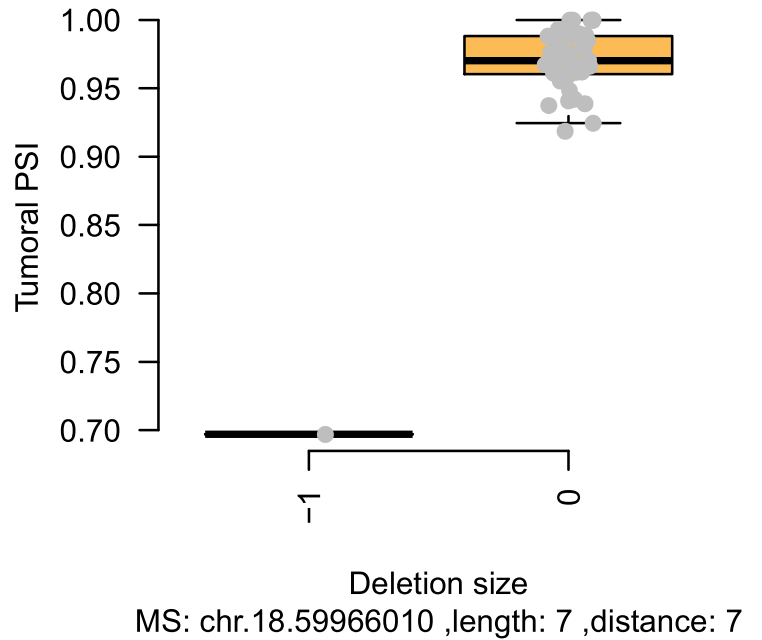


MS: chr.7.99675093 ,length: 7 ,distance: 37

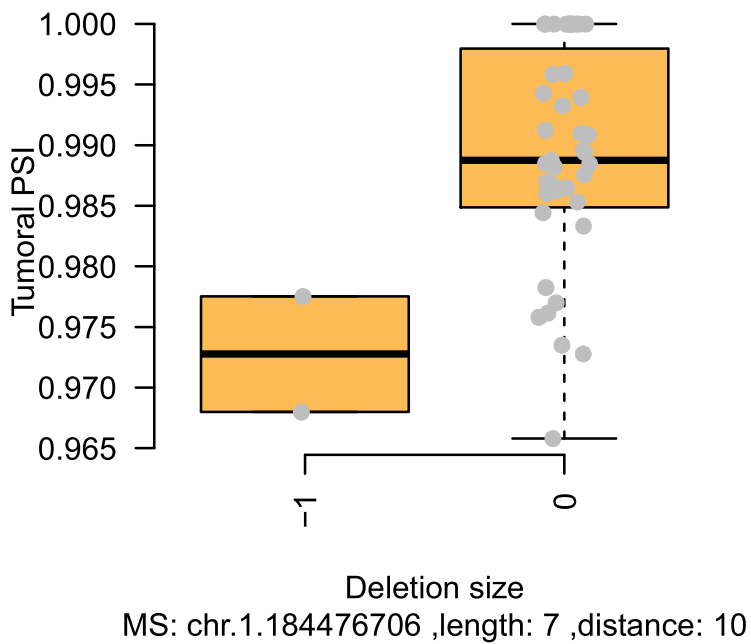
gene: **AMBRA1** ,exon: 146 ,type: Alternative



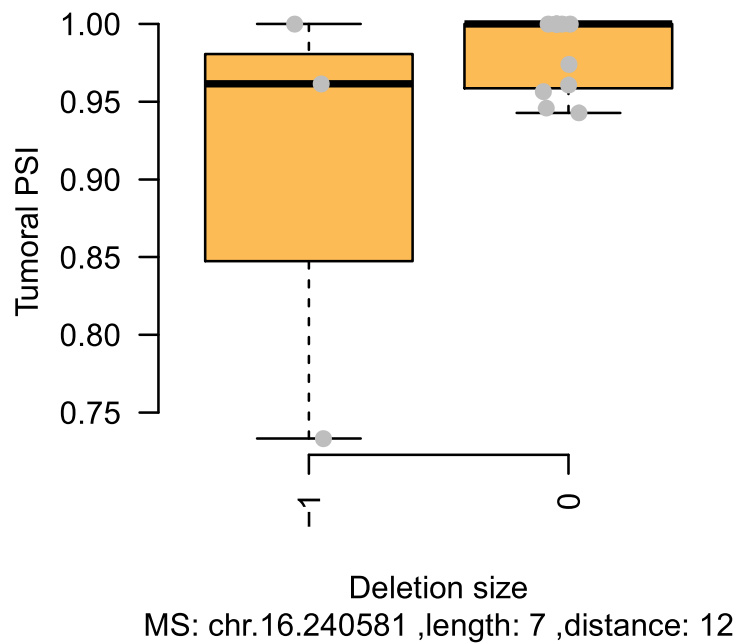
gene: **KIAA1468** ,exon: 78 ,type: Alternative



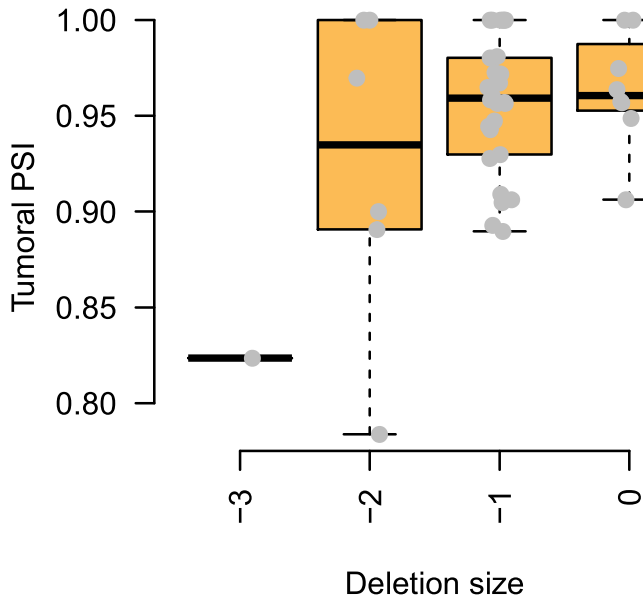
gene: **C1orf21** ,exon: 80 ,type: Alternative



gene: **LUC7L** ,exon: 135 ,type: Alternative

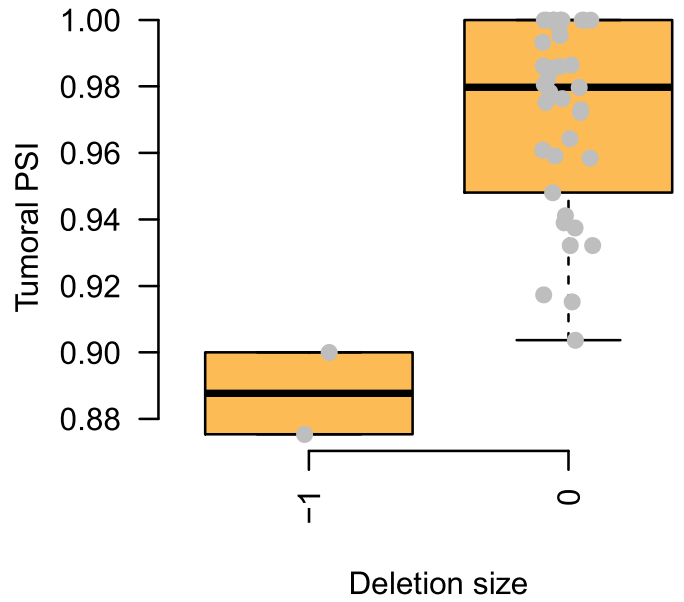


gene: RFWD2 ,exon: 80 ,type: Alternative



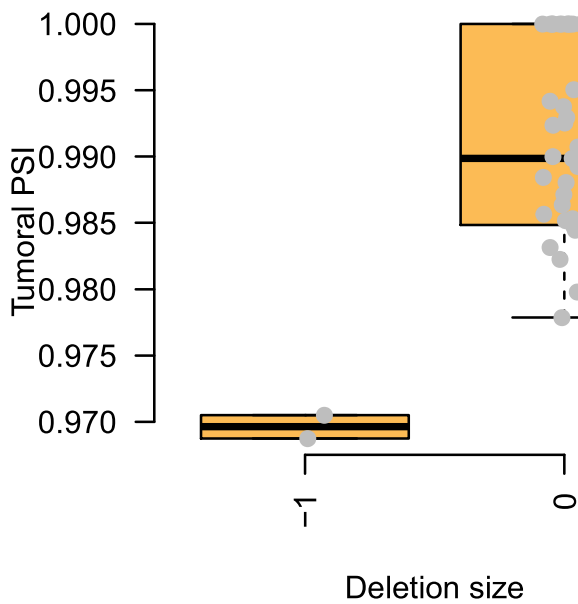
MS: chr.1.176145160 ,length: 8 ,distance: 17

gene: MX1 ,exon: 48 ,type: Alternative



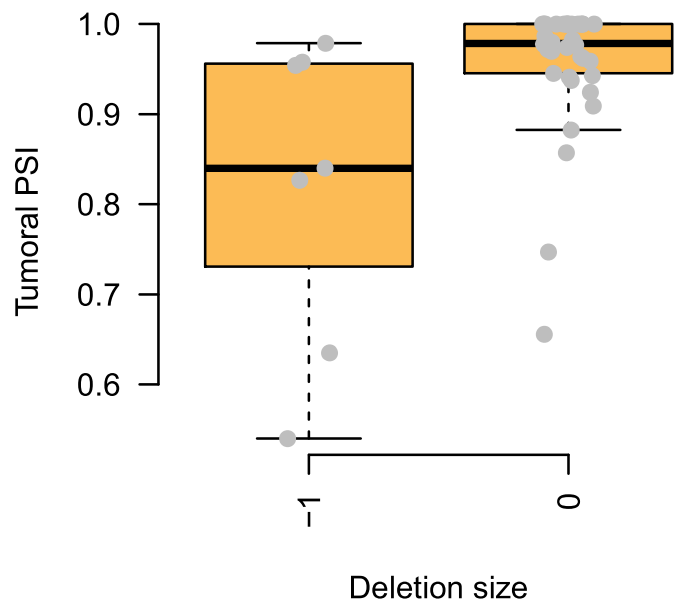
MS: chr.21.42821046 ,length: 8 ,distance: 11

gene: PPP2R5C ,exon: 113 ,type: Alternative



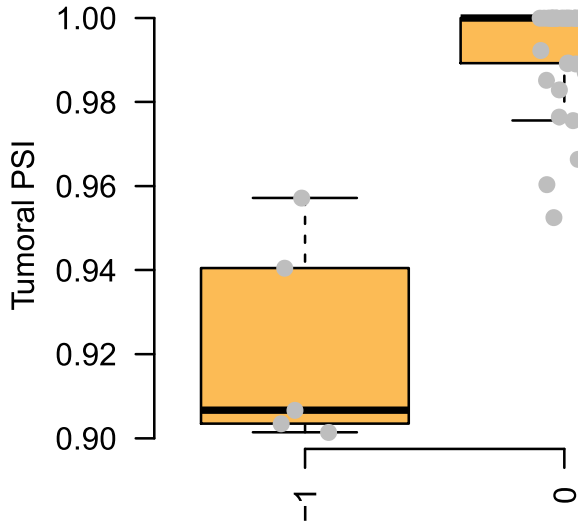
MS: chr.14.102349578 ,length: 8 ,distance: 9

gene: GOLGA4 ,exon: 146 ,type: Alternative



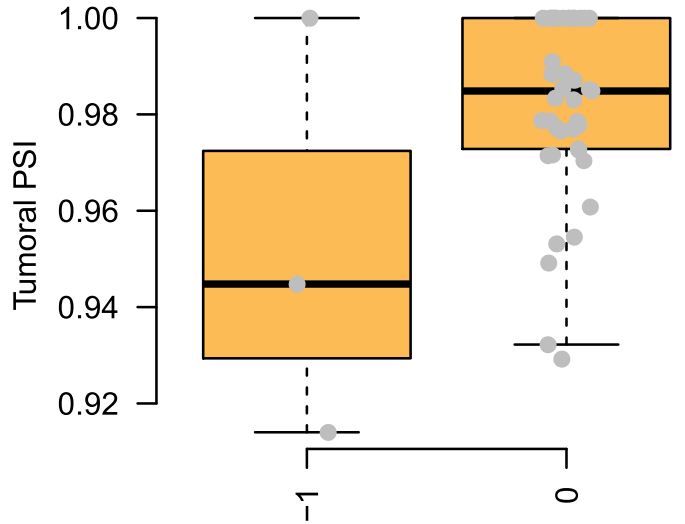
MS: chr.3.37340769 ,length: 8 ,distance: 3

gene: POLB ,exon: 30 ,type: Alternative



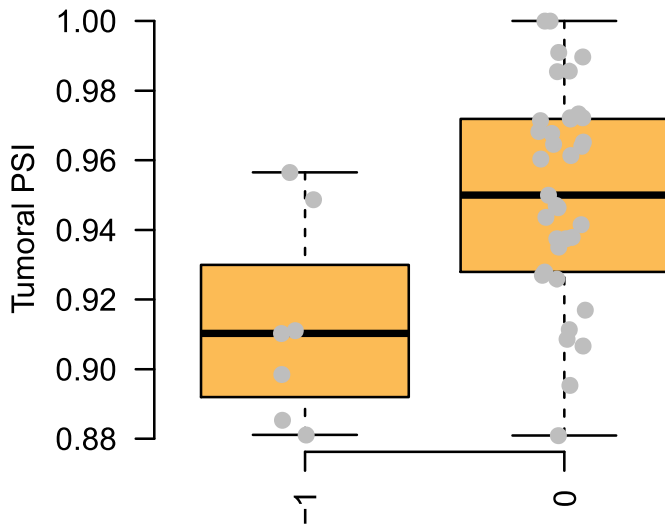
MS: chr.8.42227357 ,length: 8 ,distance: 3

gene: RPS6KB1 ,exon: 91 ,type: Alternative



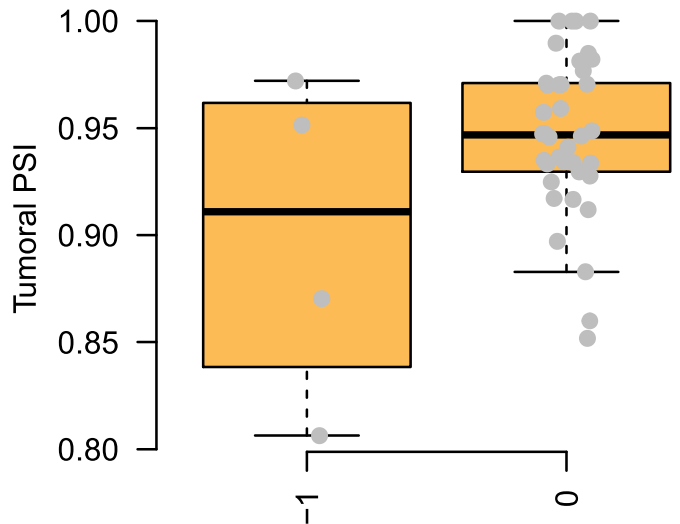
MS: chr.17.57991981 ,length: 8 ,distance: 8

gene: VPS36 ,exon: 12 ,type: Alternative



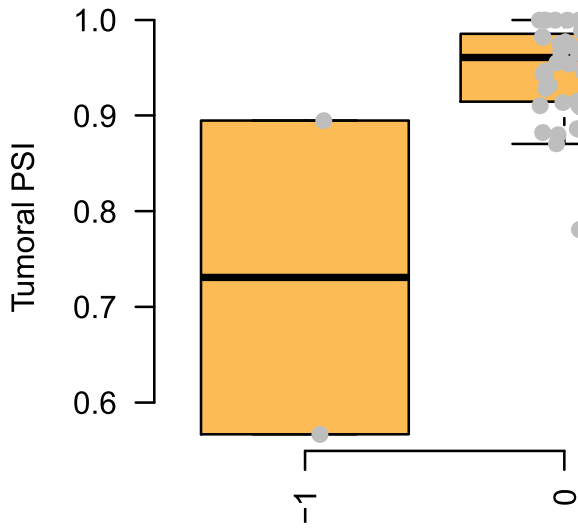
MS: chr.13.52992201 ,length: 8 ,distance: 10

gene: C21orf66 ,exon: 43 ,type: Alternative



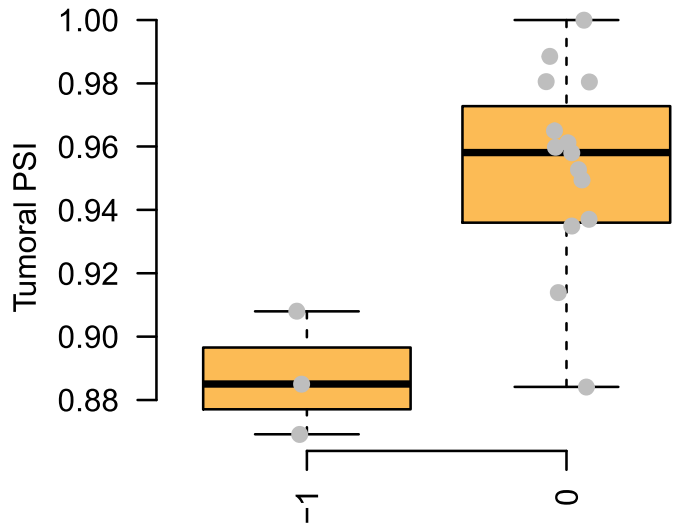
MS: chr.21.34115632 ,length: 8 ,distance: 7

gene: **ASH1L** ,exon: 113 ,type: Alternative



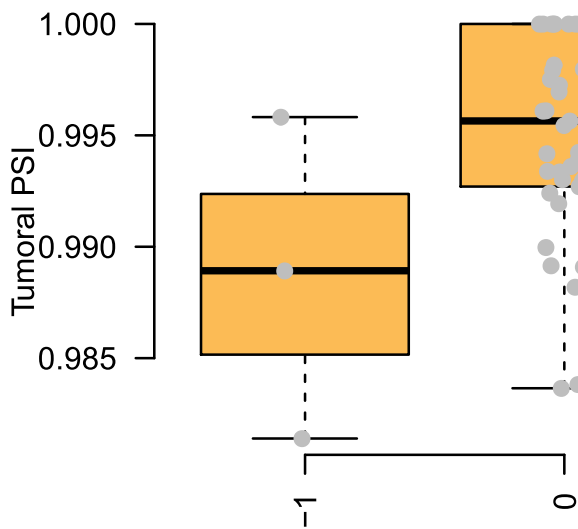
MS: chr.1.155385749 ,length: 8 ,distance: 35

gene: **PIGP** ,exon: 69 ,type: Alternative



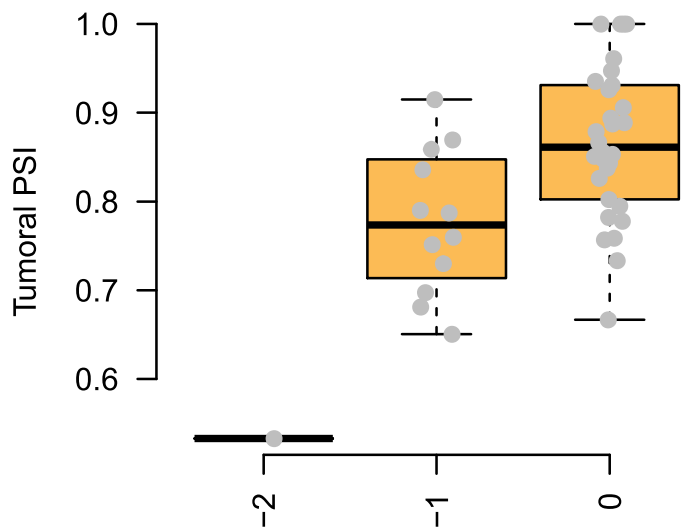
MS: chr.21.38441942 ,length: 8 ,distance: 18

gene: **ATP2A2** ,exon: 113 ,type: Alternative



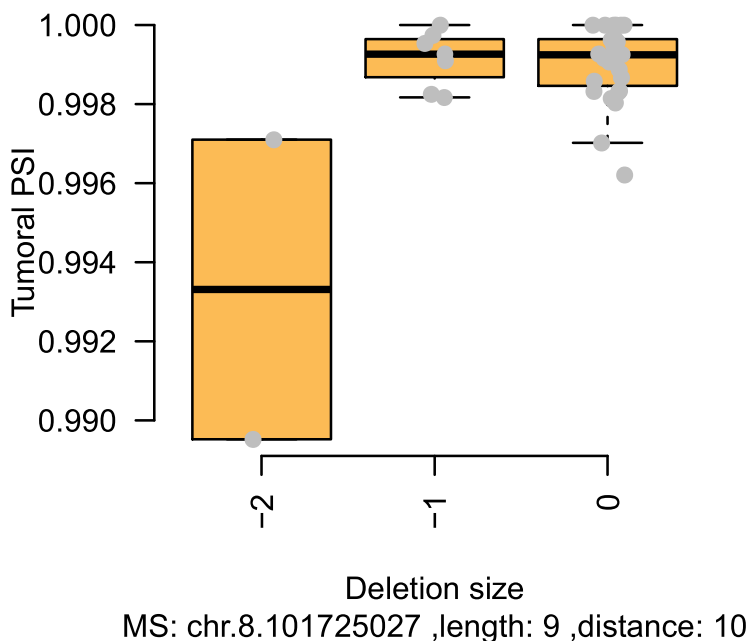
MS: chr.12.110760778 ,length: 8 ,distance: 12

gene: **MBIP** ,exon: 124 ,type: Alternative

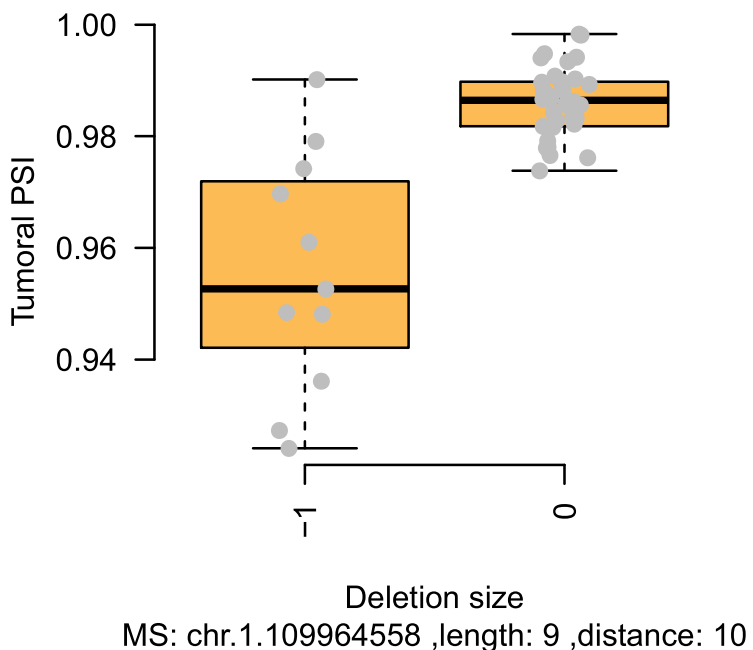


MS: chr.14.36777407 ,length: 9 ,distance: 12

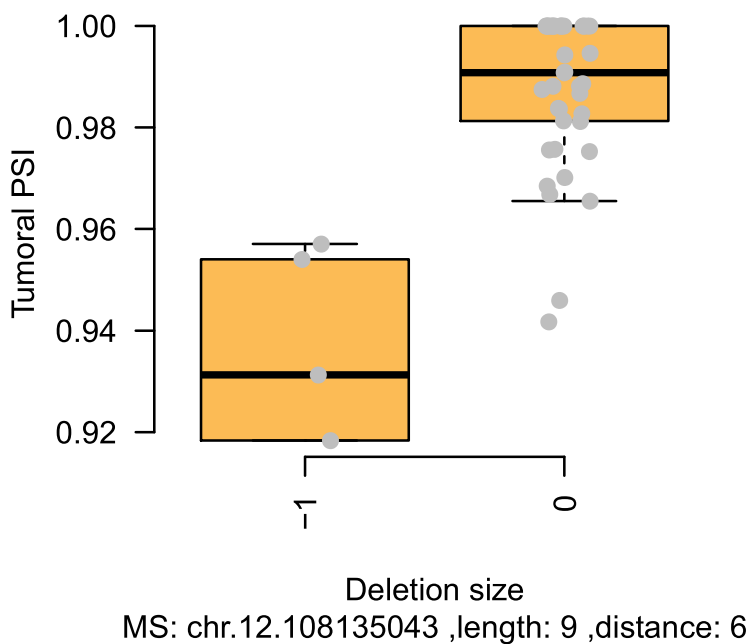
gene: PABPC1 ,exon: 113 ,type: Alternative



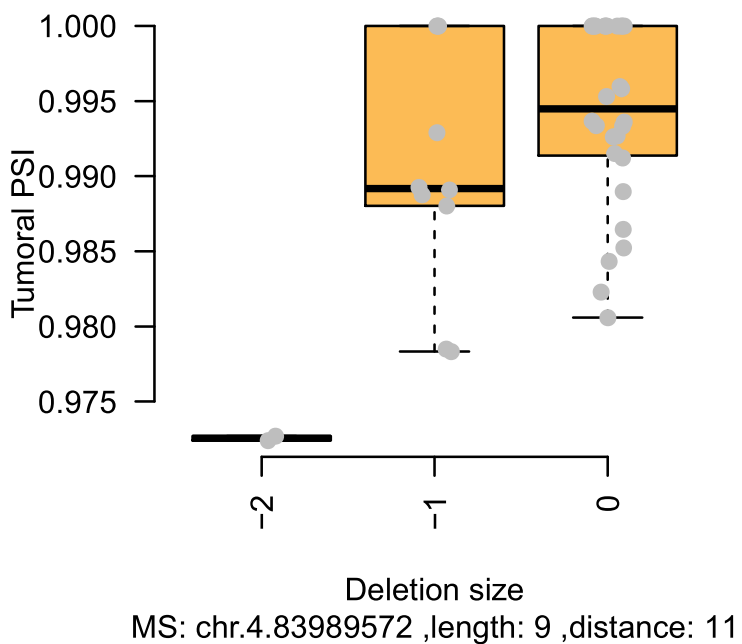
gene: PSMA5 ,exon: 69 ,type: Alternative



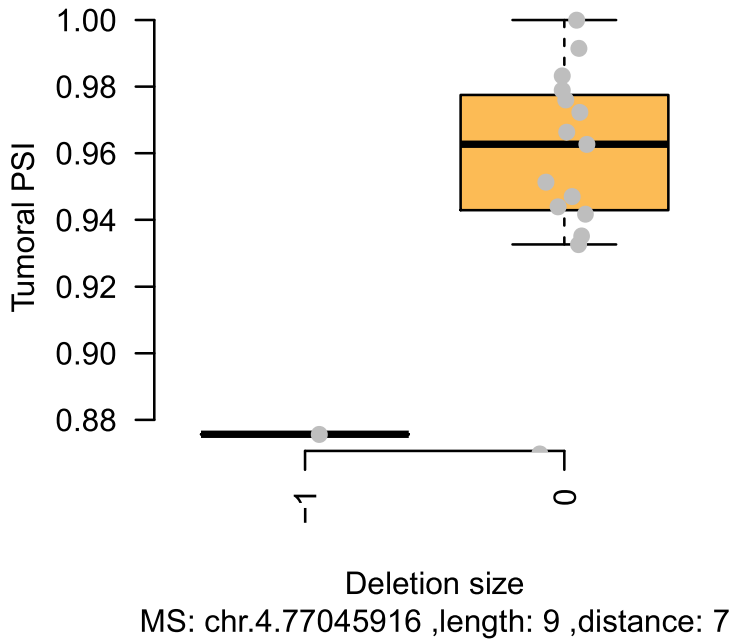
gene: PRDM4 ,exon: 1 ,type: Alternative



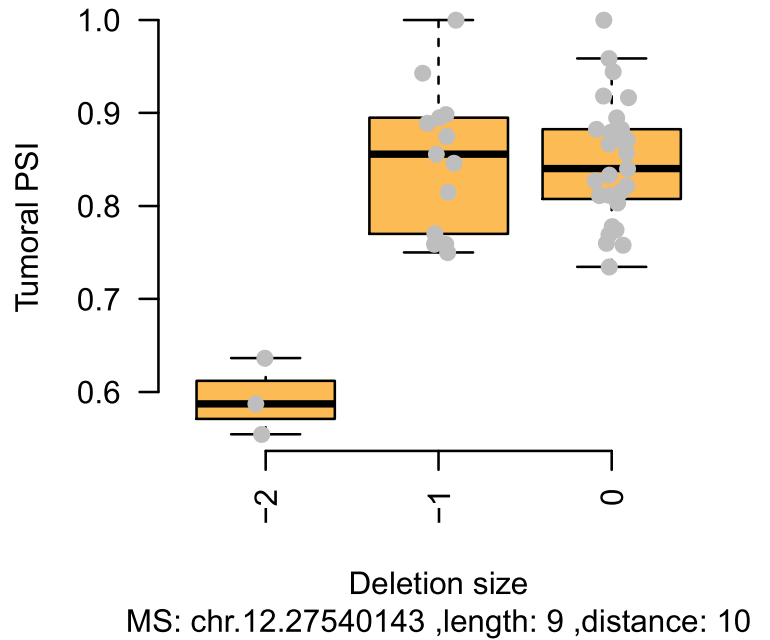
gene: COPS4 ,exon: 146 ,type: Alternative



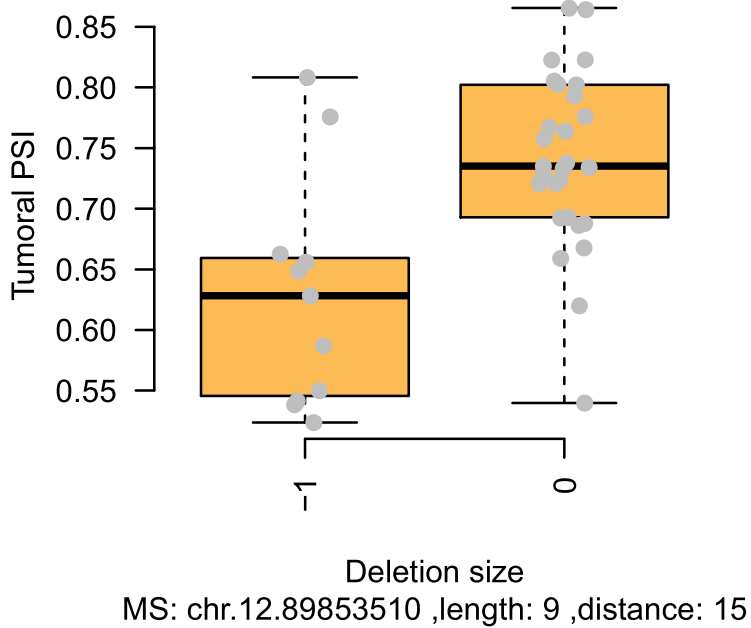
gene: NUP54 ,exon: 146 ,type: Alternative



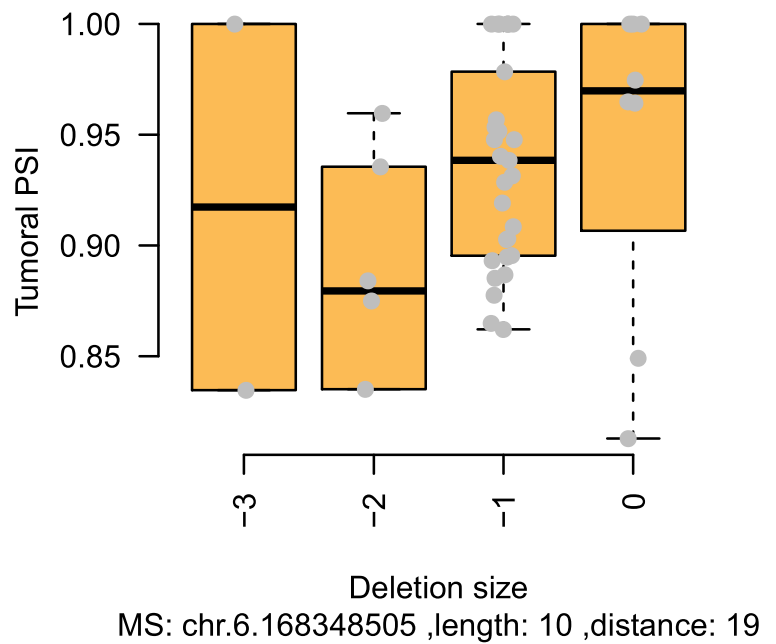
gene: ARNTL2 ,exon: 124 ,type: Alternative



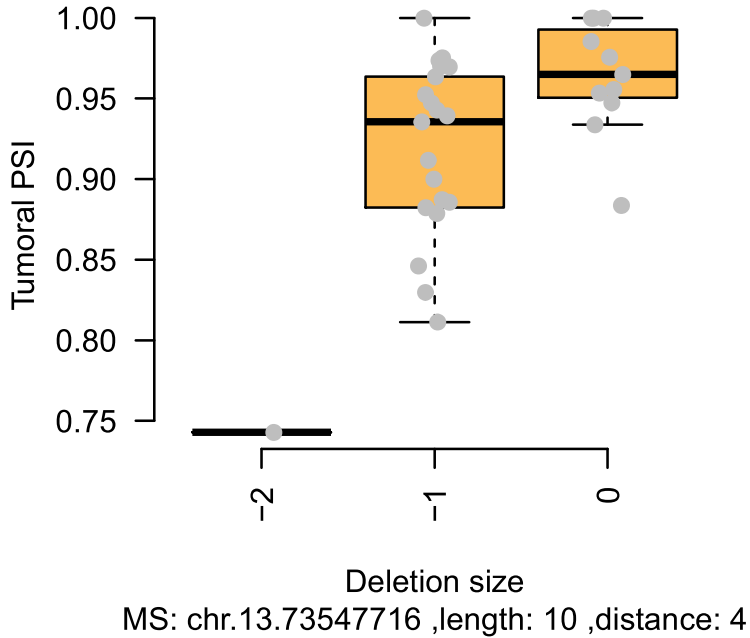
gene: WDR51B ,exon: 1 ,type: Alternative



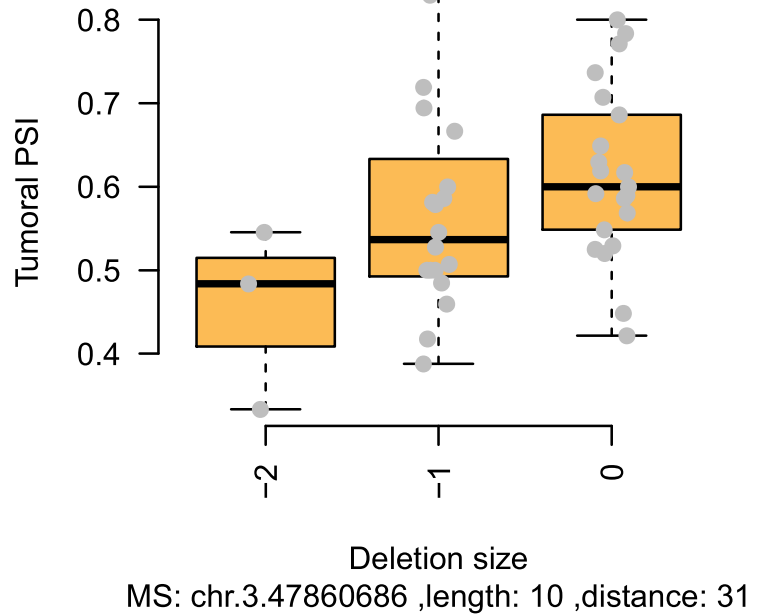
gene: MLLT4 ,exon: 77 ,type: Alternative



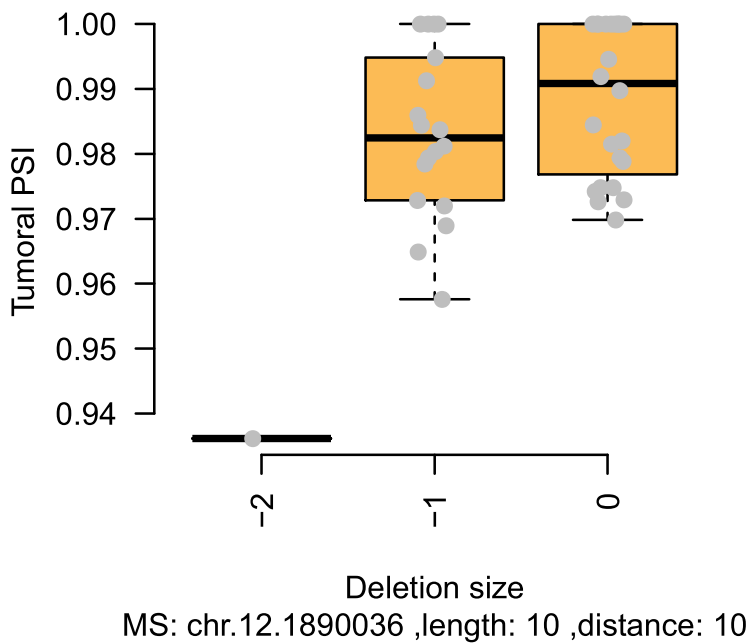
gene: PIBF1 ,exon: 48 ,type: Alternative



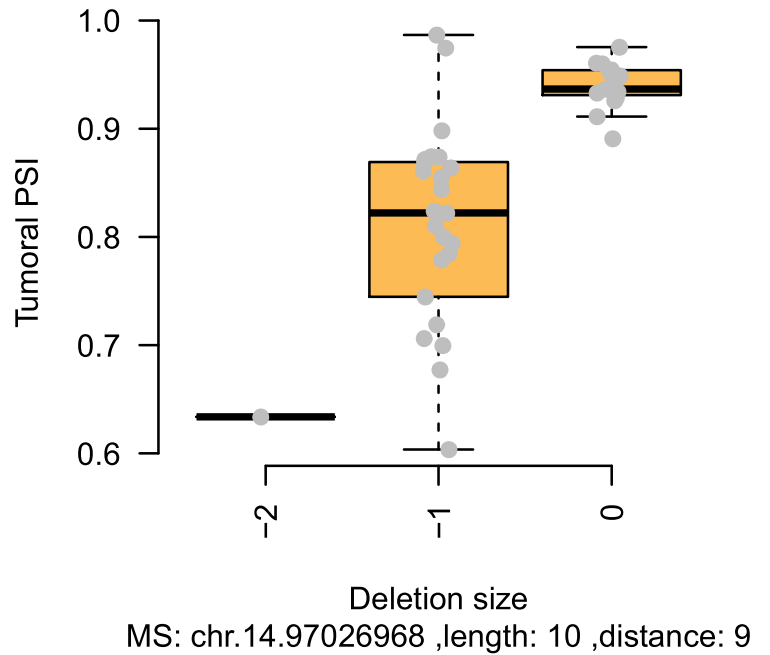
gene: DHX30 ,exon: 102 ,type: Alternative



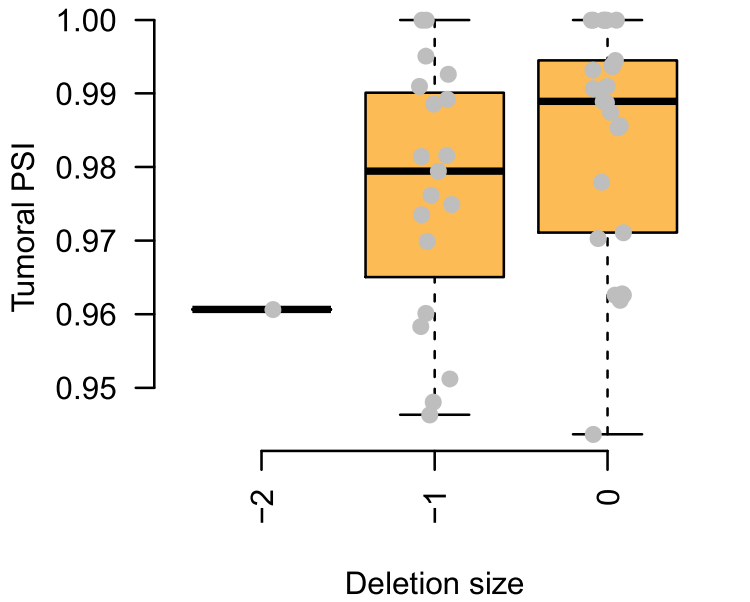
gene: ADIPOR2 ,exon: 113 ,type: Alternative



gene: PAPOLA ,exon: 70 ,type: Alternative

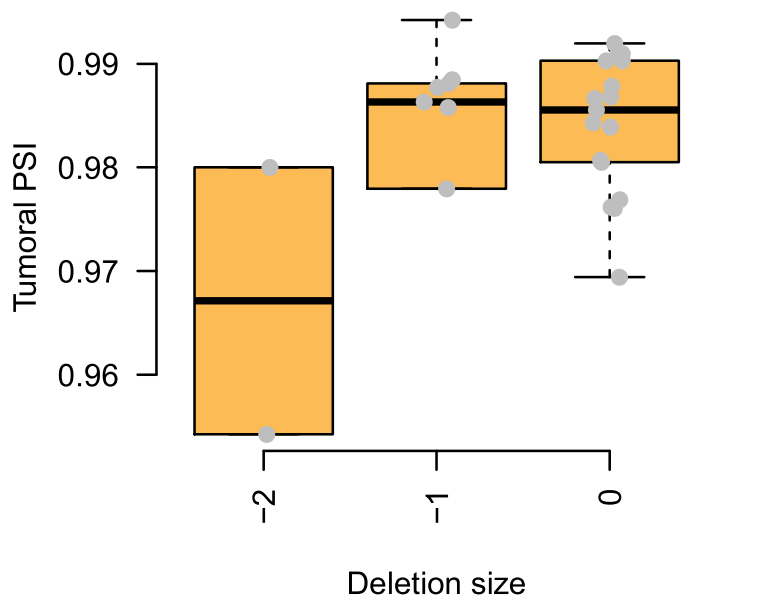


gene: C8orf76 ,exon: 91 ,type: Alternative



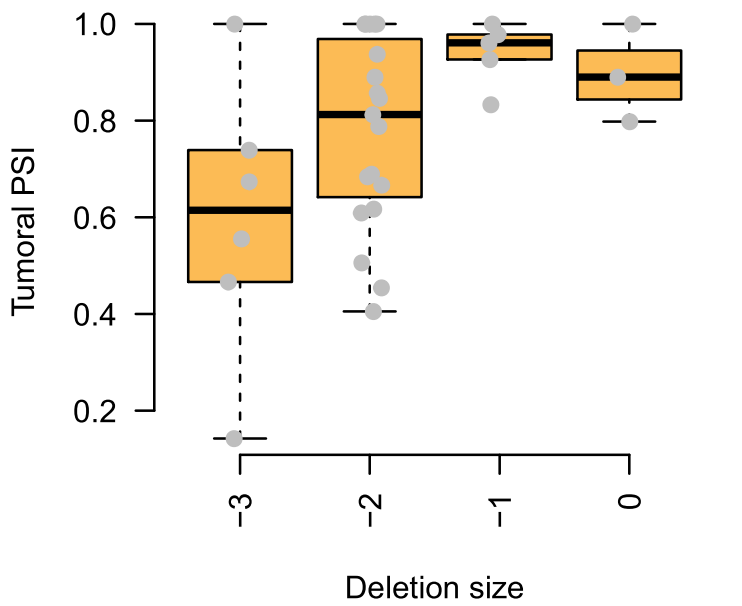
MS: chr.8.124244006 ,length: 10 ,distance: 9

gene: NAP1L1 ,exon: 113 ,type: Alternative



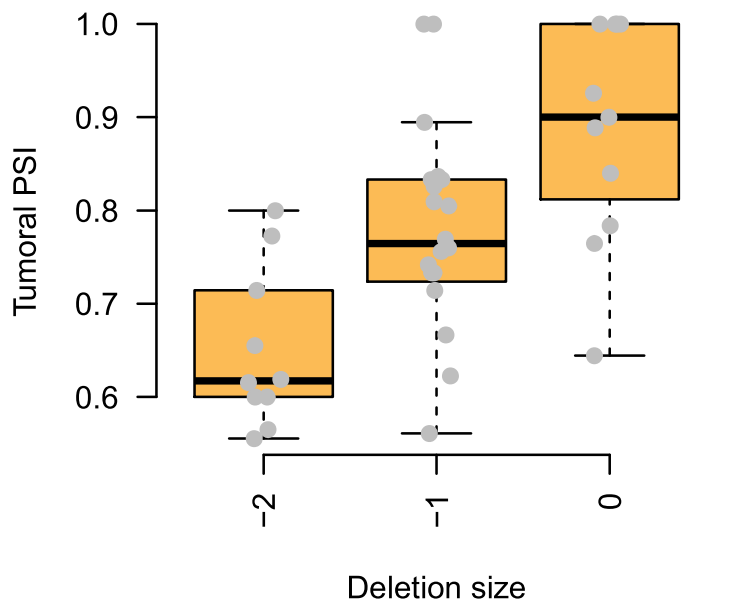
MS: chr.12.76453671 ,length: 10 ,distance: 13

gene: PDS5A ,exon: 91 ,type: Alternative



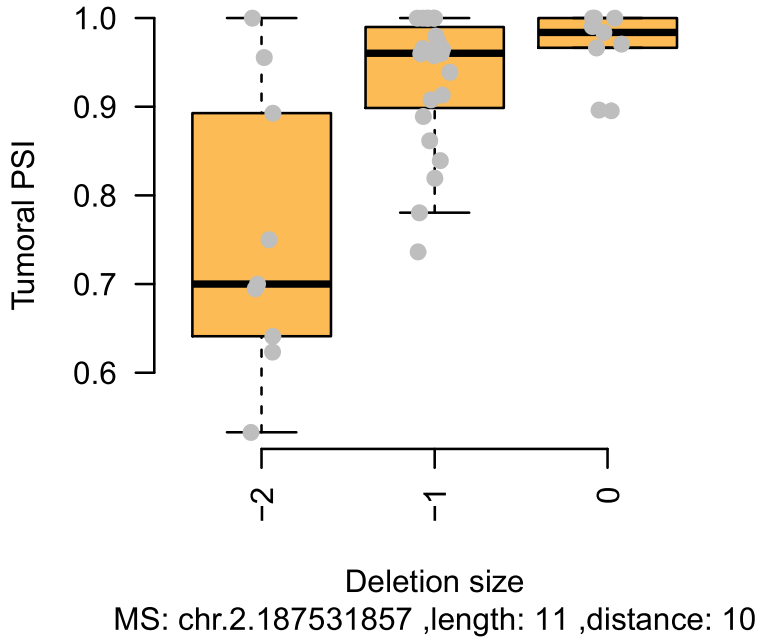
MS: chr.4.39928487 ,length: 11 ,distance: 6

gene: KIAA1432 ,exon: 35 ,type: Alternative

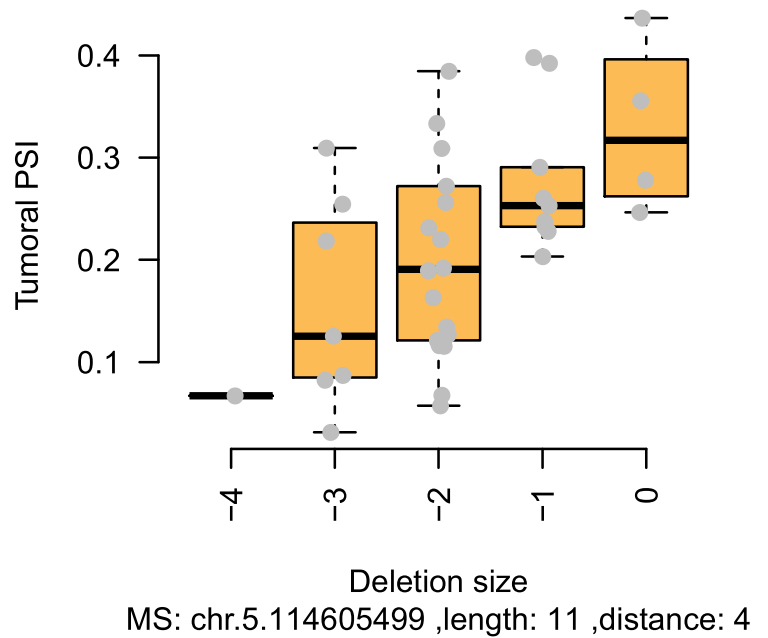


MS: chr.9.5753519 ,length: 11 ,distance: 7

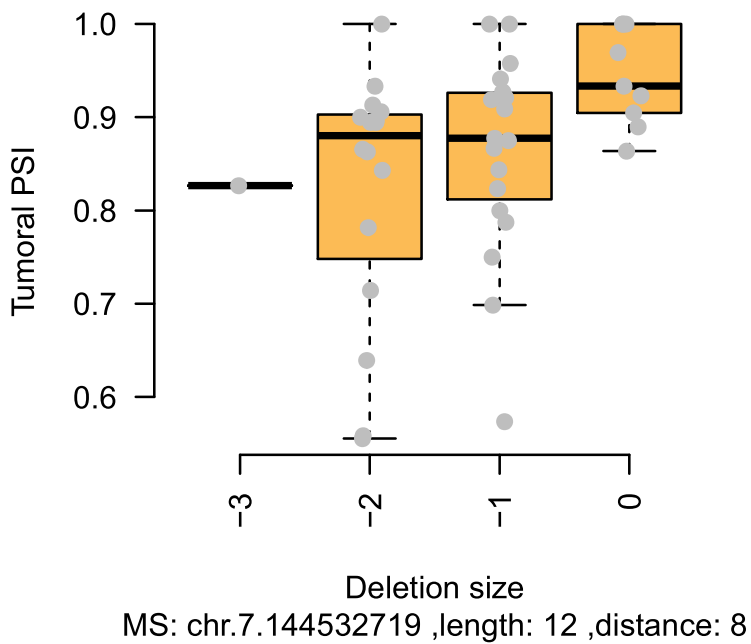
gene: ITGAV ,exon: 71 ,type: Alternative



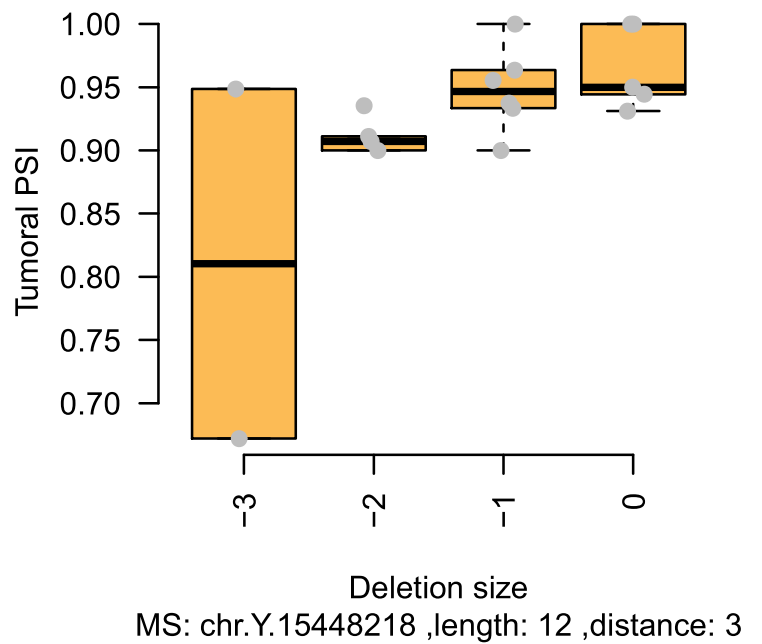
gene: CCDC112 ,exon: 135 ,type: Alternative



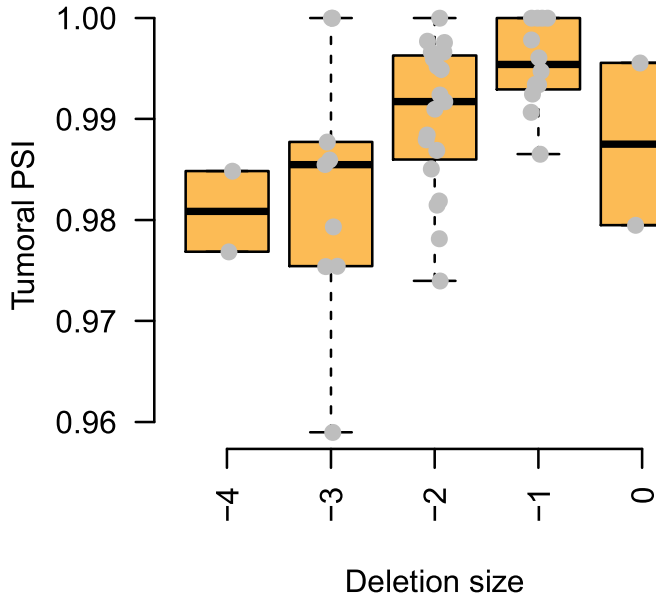
gene: TPK1 ,exon: 69 ,type: Alternative



gene: UTY ,exon: 48 ,type: Alternative

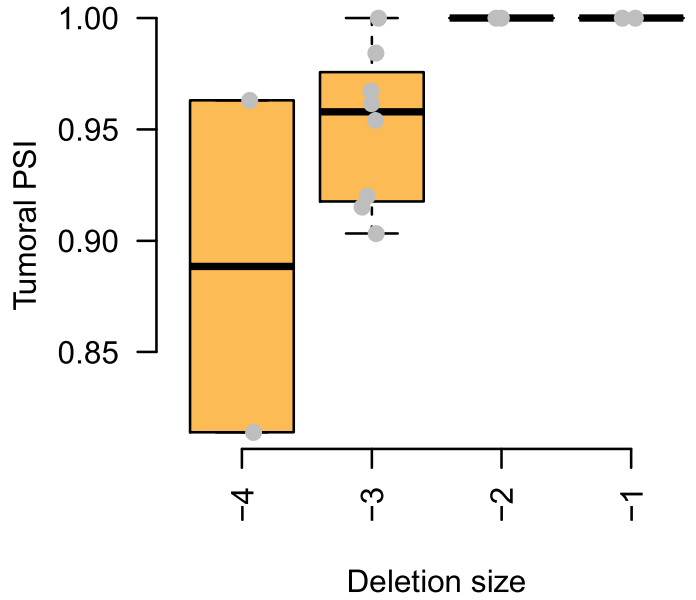


gene: MED4 ,exon: 69 ,type: Alternative



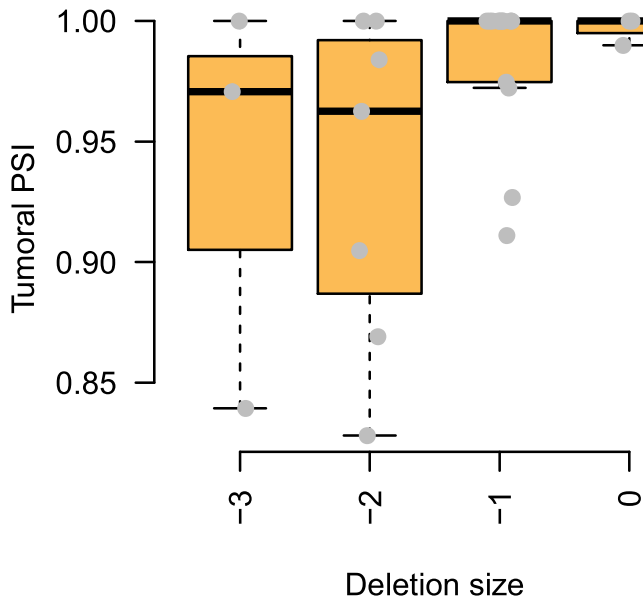
MS: chr.13.48664557 ,length: 12 ,distance: 3

gene: MON2 ,exon: 82 ,type: Alternative



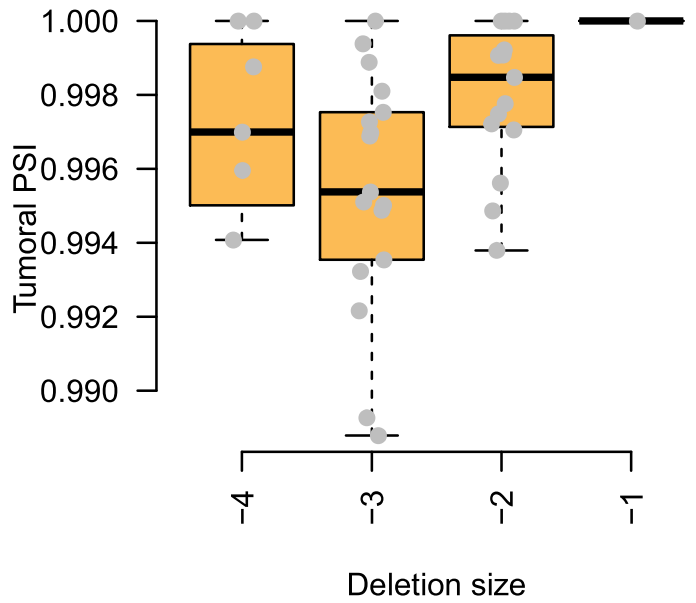
MS: chr.12.62972211 ,length: 12 ,distance: 3

gene: MAP4K5 ,exon: 81 ,type: Alternative



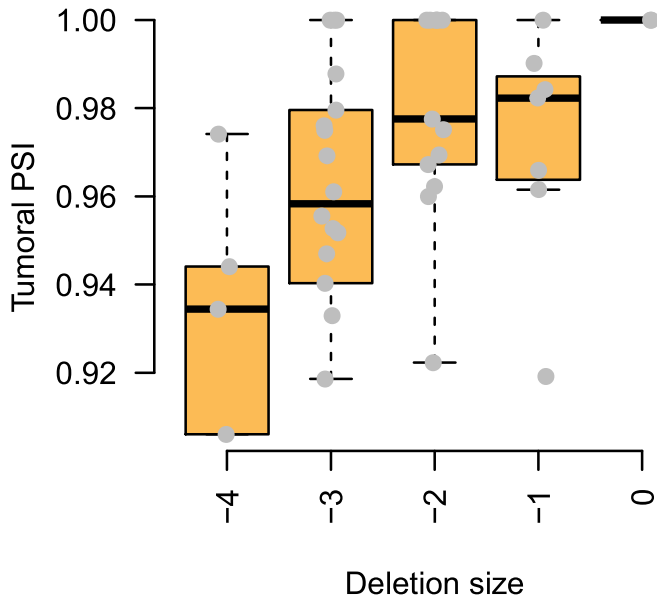
MS: chr.14.50895476 ,length: 12 ,distance: 4

gene: YWHAE ,exon: 80 ,type: Alternative



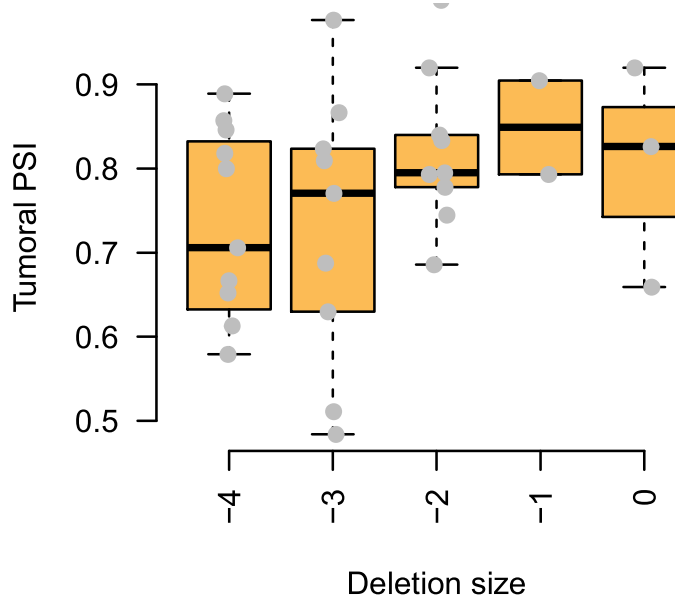
MS: chr.17.1265305 ,length: 13 ,distance: 3

gene: CCDC52 ,exon: 43 ,type: Alternative



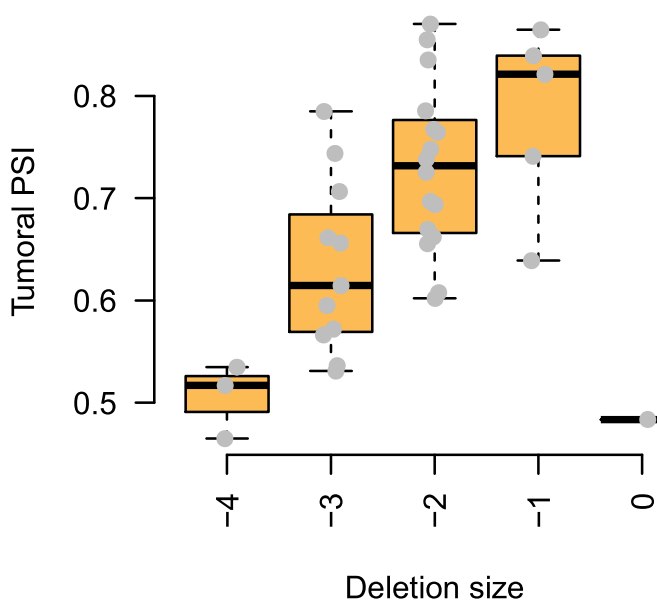
MS: chr.3.113169373 ,length: 13 ,distance: 10

gene: PHTF2 ,exon: 12 ,type: Alternative



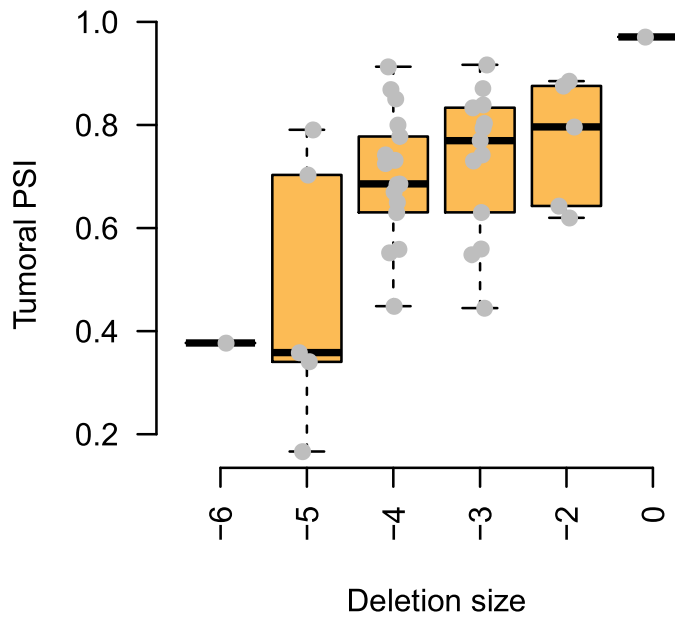
MS: chr.7.77567009 ,length: 14 ,distance: 5

gene: PTP4A2 ,exon: 69 ,type: Alternative



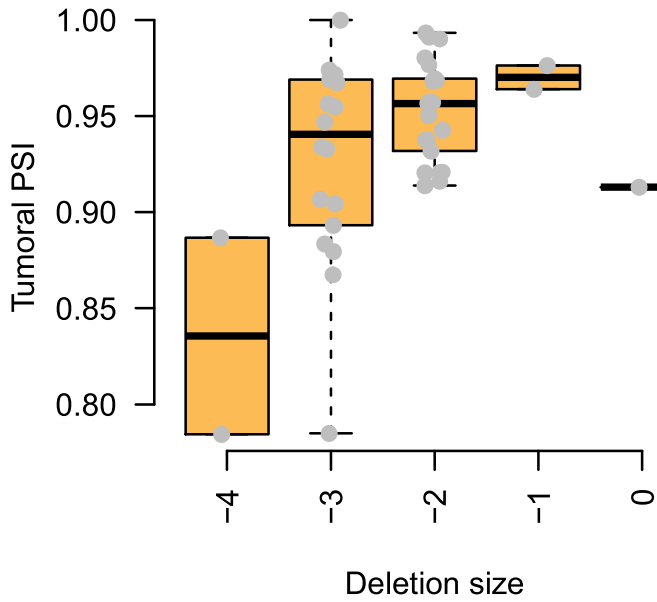
MS: chr.1.32381593 ,length: 14 ,distance: 5

gene: NUP153 ,exon: 135 ,type: Alternative



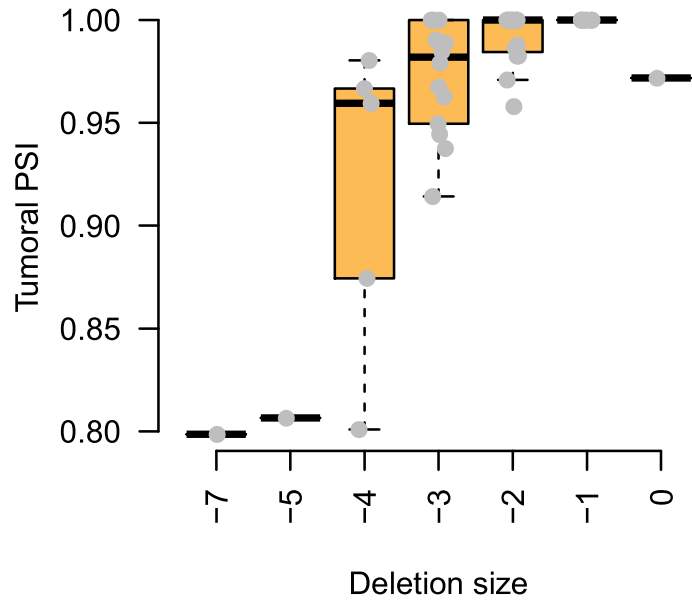
MS: chr.6.17669263 ,length: 14 ,distance: 4

gene: GFM1 ,exon: 48 ,type: Alternative



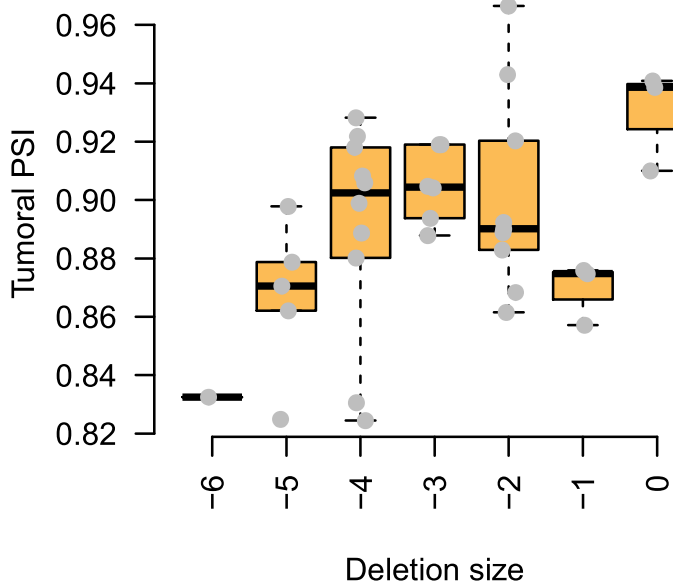
MS: chr.3.158407931 ,length: 14 ,distance: 8

gene: RB1CC1 ,exon: 71 ,type: Alternative



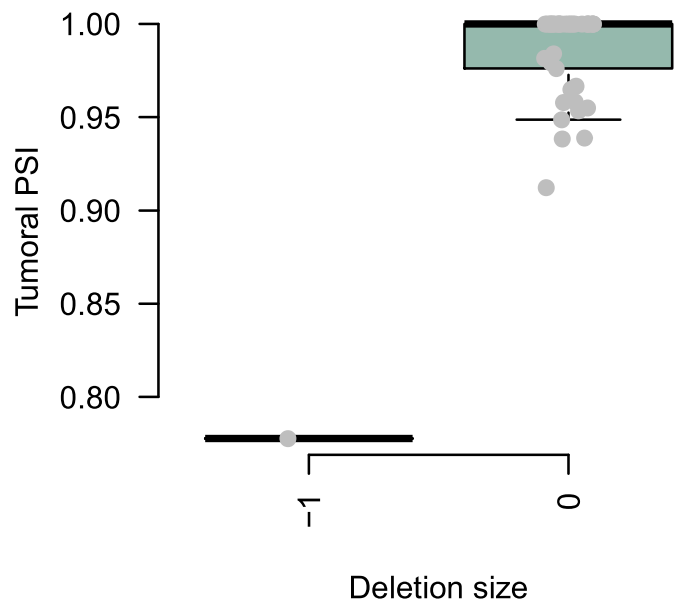
MS: chr.8.53543091 ,length: 15 ,distance: 3

gene: DAP3 ,exon: 80 ,type: Alternative



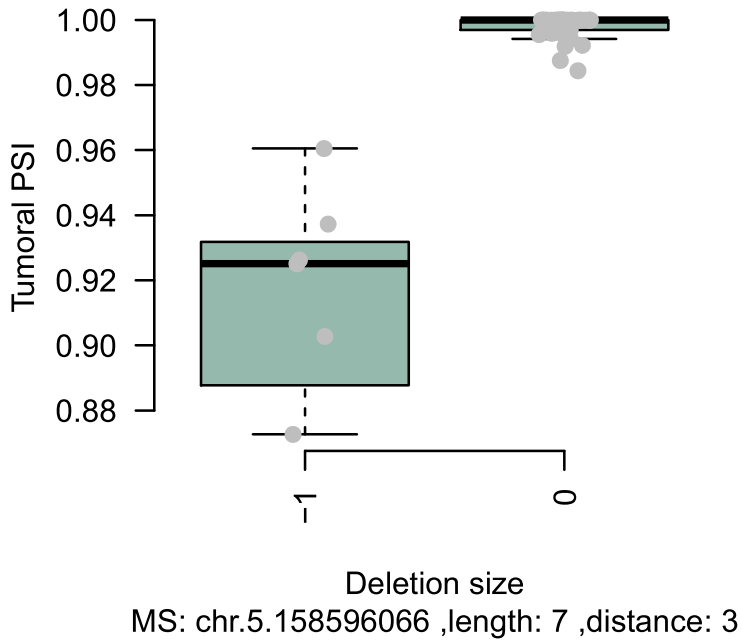
MS: chr.1.155686740 ,length: 17 ,distance: 41

gene: FCHO2 ,exon: 71 ,type: Constitutive

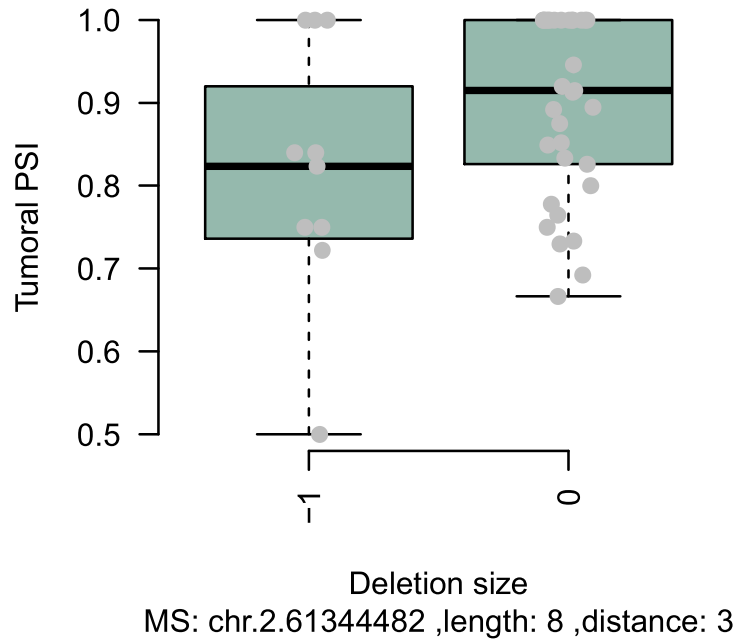


MS: chr.5.72373991 ,length: 6 ,distance: 11

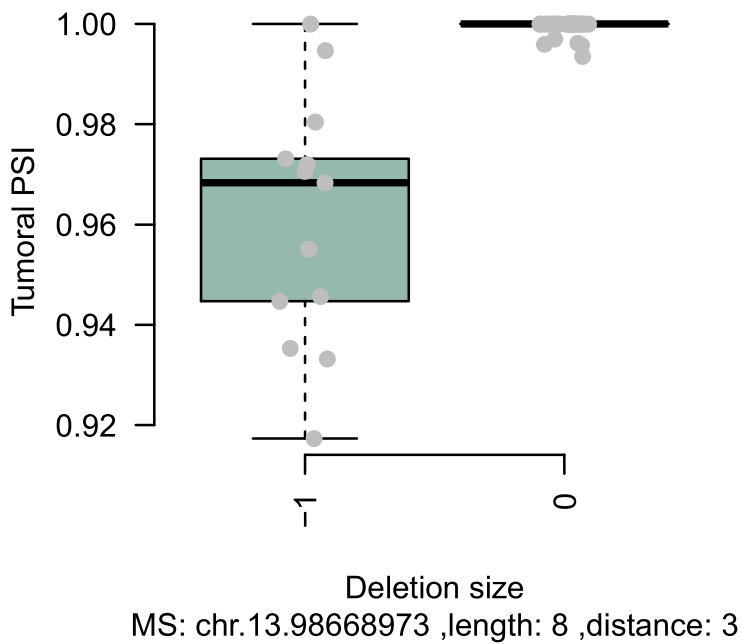
gene: RNF145 ,exon: 135 ,type: Constitutive



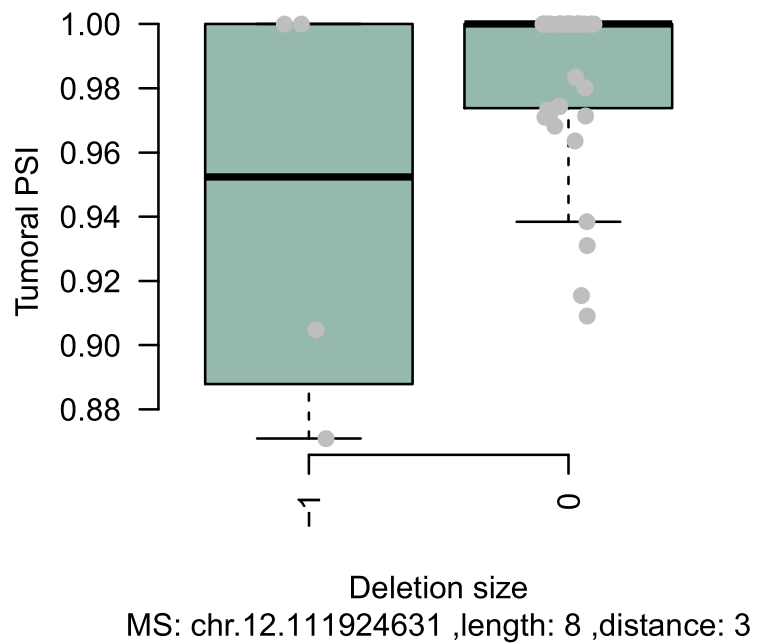
gene: KIAA1841 ,exon: 60 ,type: Constitutive



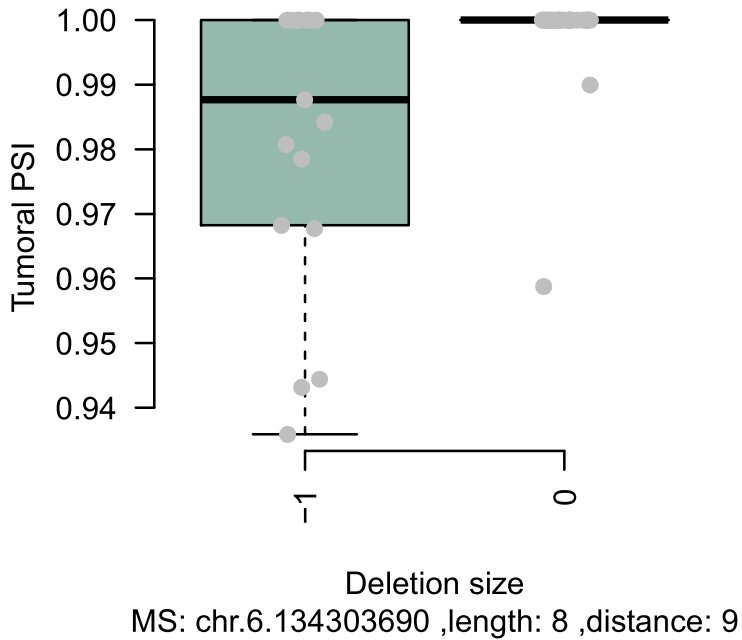
gene: IPO5 ,exon: 75 ,type: Constitutive



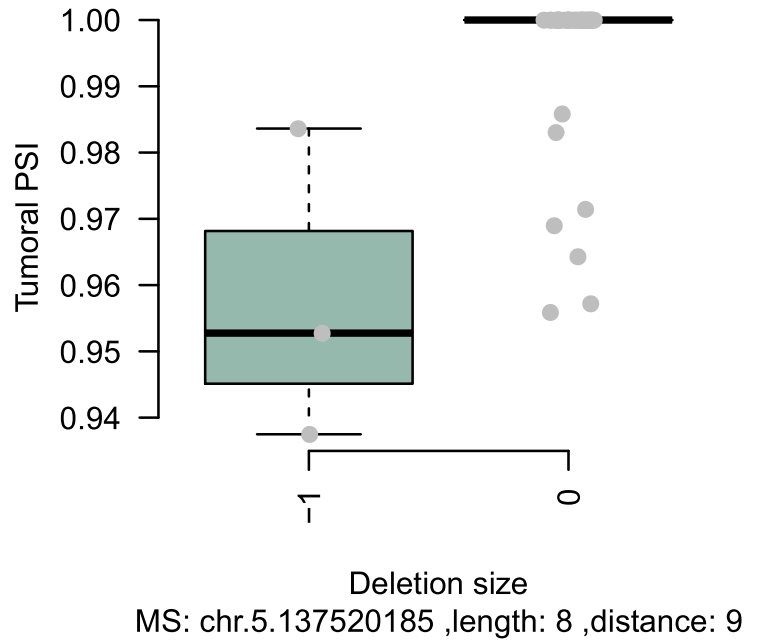
gene: ATXN2 ,exon: 48 ,type: Constitutive



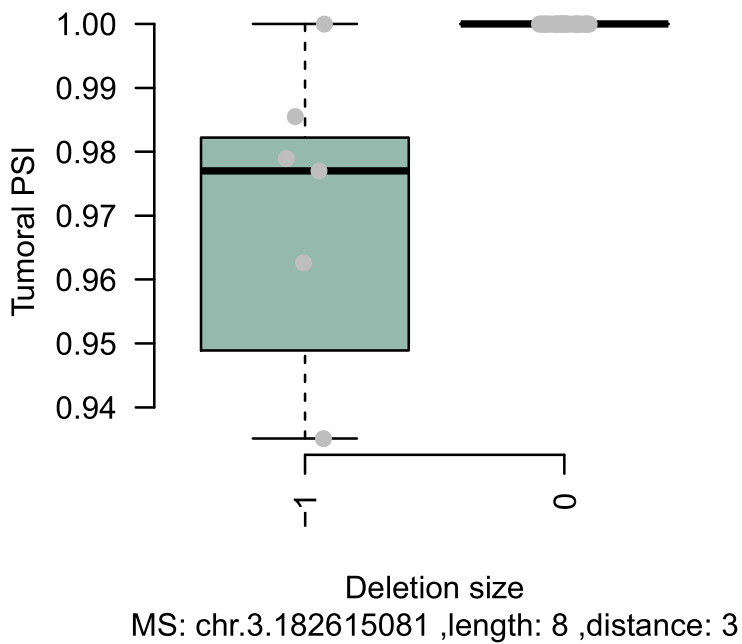
gene: TBPL1 ,exon: 80 ,type: Constitutive



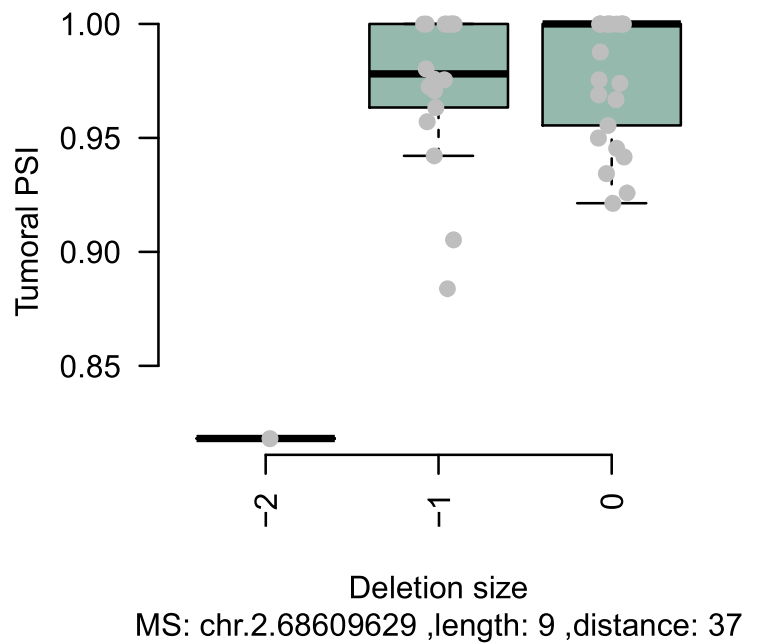
gene: KIF20A ,exon: 30 ,type: Constitutive



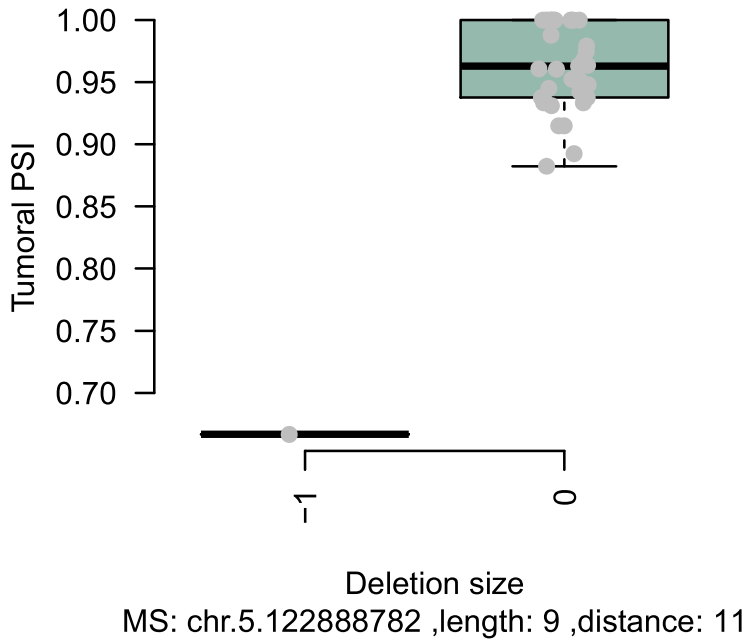
gene: ATP11B ,exon: 77 ,type: Constitutive



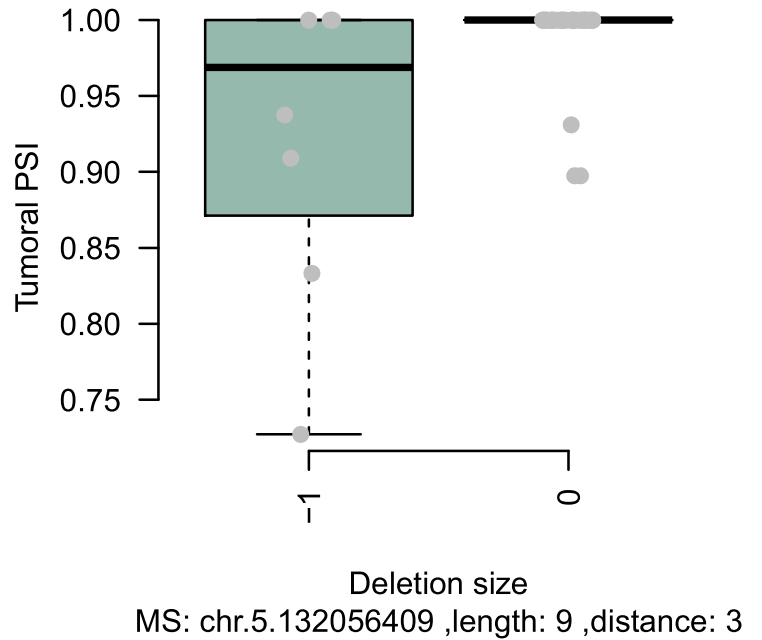
gene: PLEK ,exon: 91 ,type: Constitutive



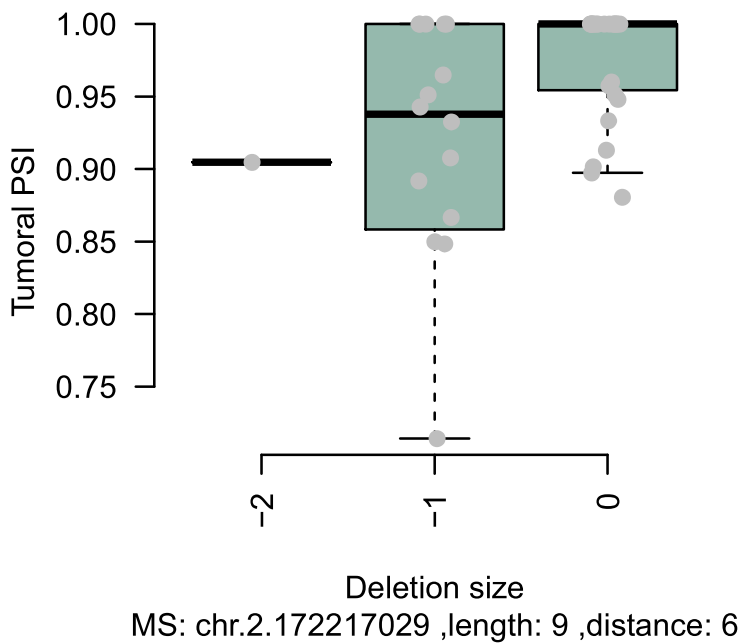
gene: CSNK1G3 ,exon: 80 ,type: Constitutive



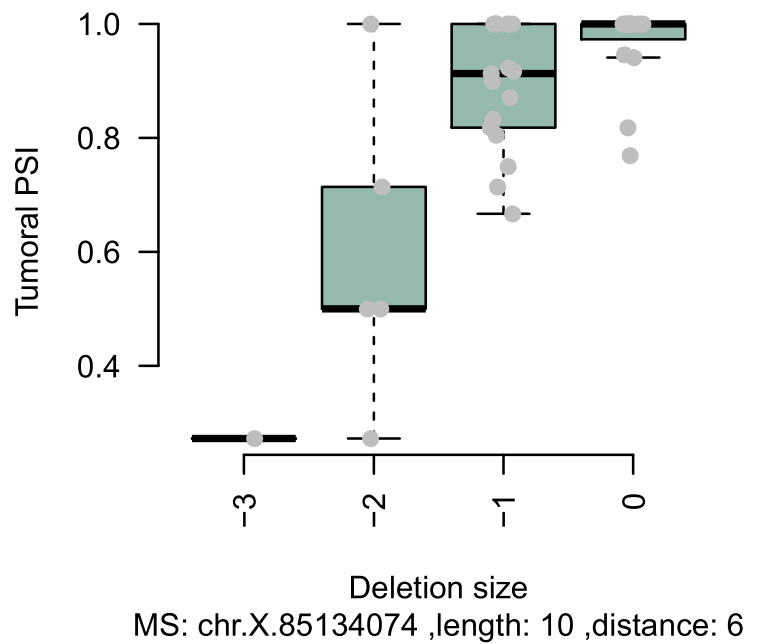
gene: KIF3A ,exon: 102 ,type: Constitutive



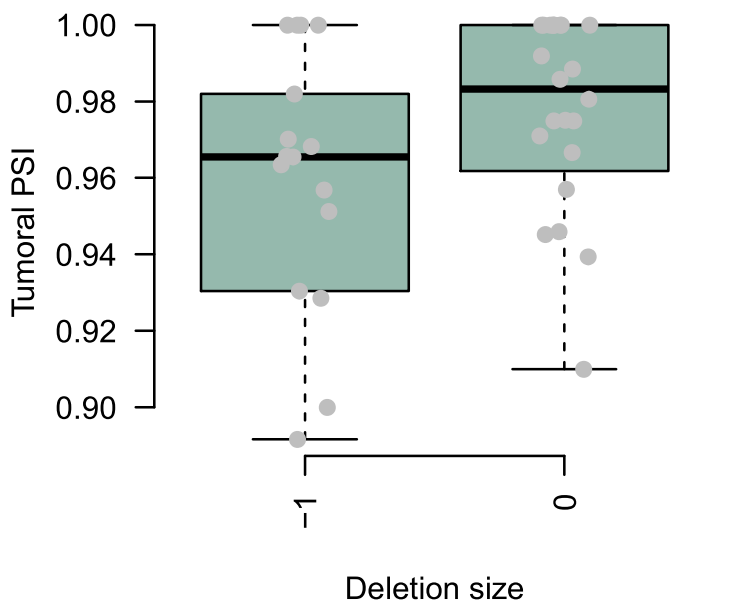
gene: METTL8 ,exon: 80 ,type: Constitutive



gene: CHM ,exon: 30 ,type: Constitutive

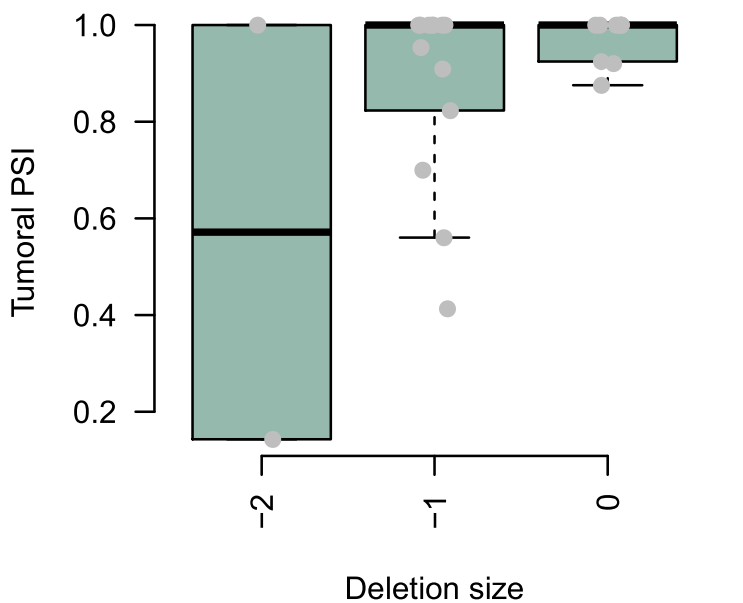


gene: ARHGAP11A ,exon: 12 ,type: Constitutive



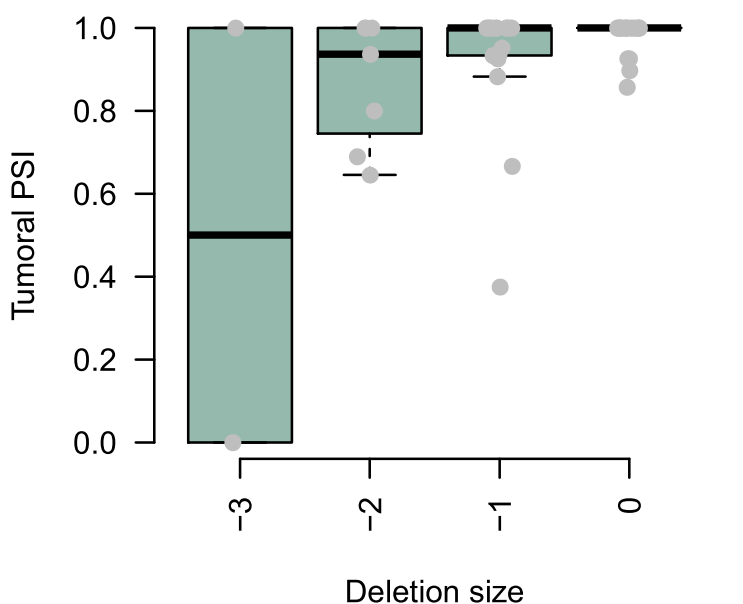
MS: chr.15.32927960 ,length: 10 ,distance: 9

gene: TAF9B ,exon: 91 ,type: Constitutive



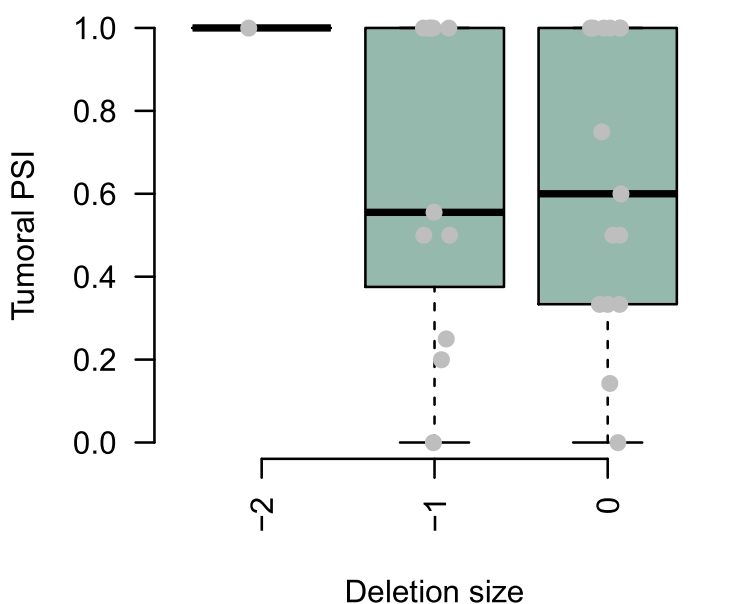
MS: chr.X.77393384 ,length: 10 ,distance: 4

gene: C5orf42 ,exon: 96 ,type: Constitutive



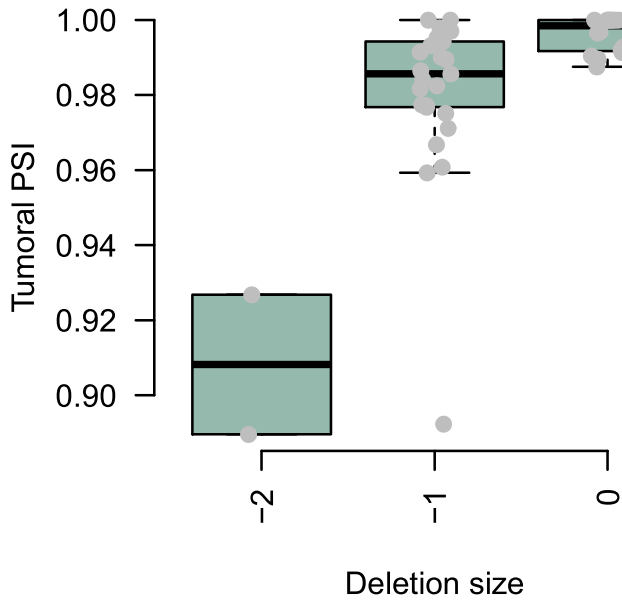
MS: chr.5.37139475 ,length: 10 ,distance: 3

gene: NEK5 ,exon: 60 ,type: Constitutive



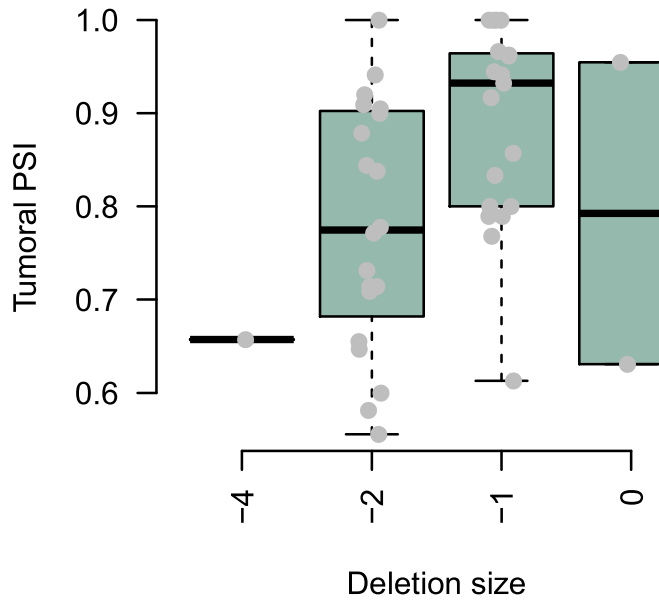
MS: chr.13.52656451 ,length: 10 ,distance: 3

gene: TOP2A ,exon: 83 ,type: Constitutive



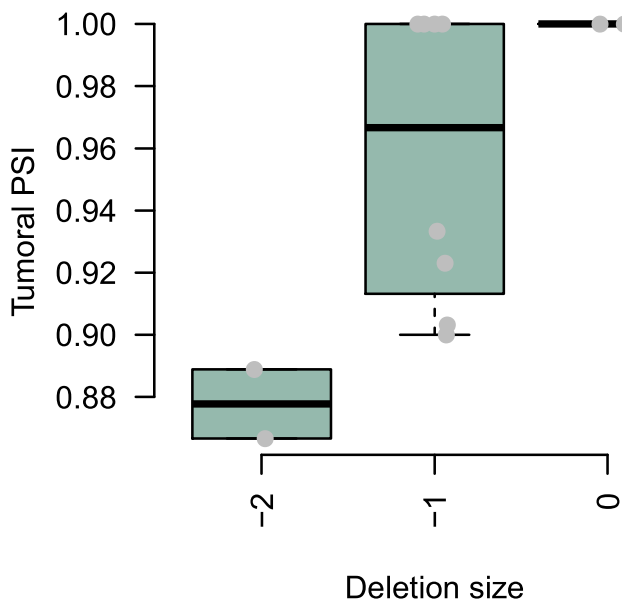
MS: chr.17.38548372 ,length: 11 ,distance: 9

gene: MRE11A ,exon: 102 ,type: Constitutive



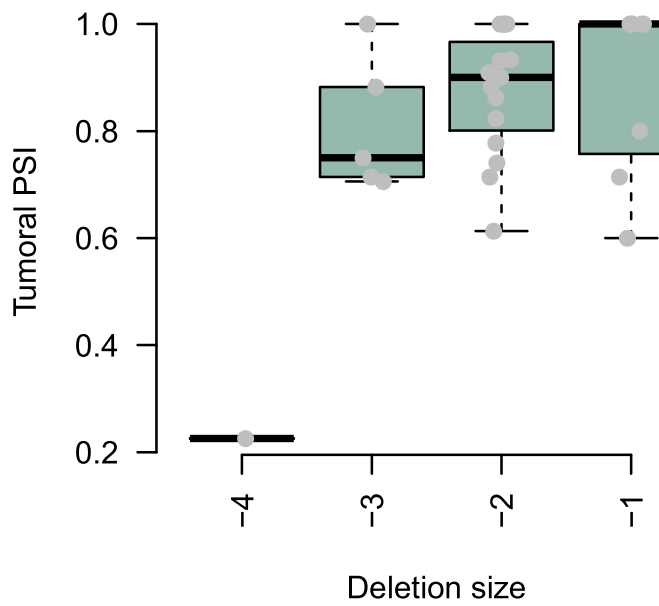
MS: chr.11.94212931 ,length: 11 ,distance: 4

gene: NUFIP1 ,exon: 69 ,type: Constitutive



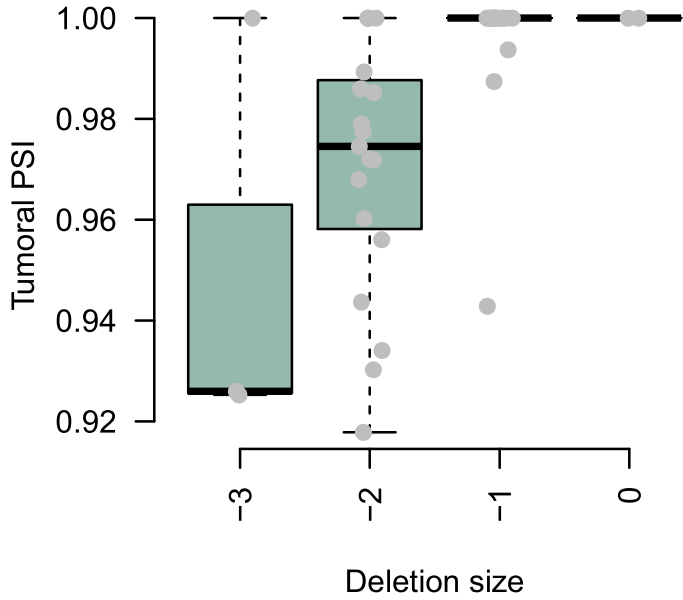
MS: chr.13.45556292 ,length: 11 ,distance: 3

gene: TBC1D19 ,exon: 48 ,type: Constitutive



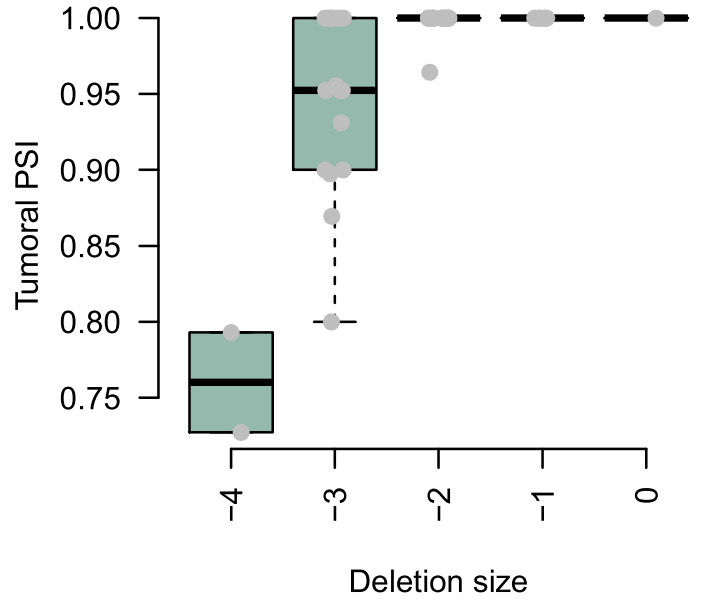
MS: chr.4.26737064 ,length: 11 ,distance: 3

gene: KDM6A ,exon: 76 ,type: Constitutive



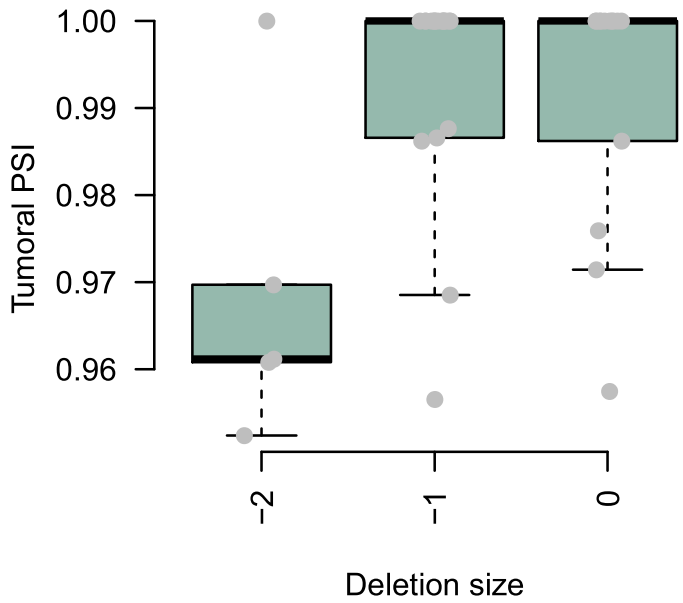
MS: chr.X.44949952 ,length: 11 ,distance: 6

gene: RNF214 ,exon: 1 ,type: Constitutive



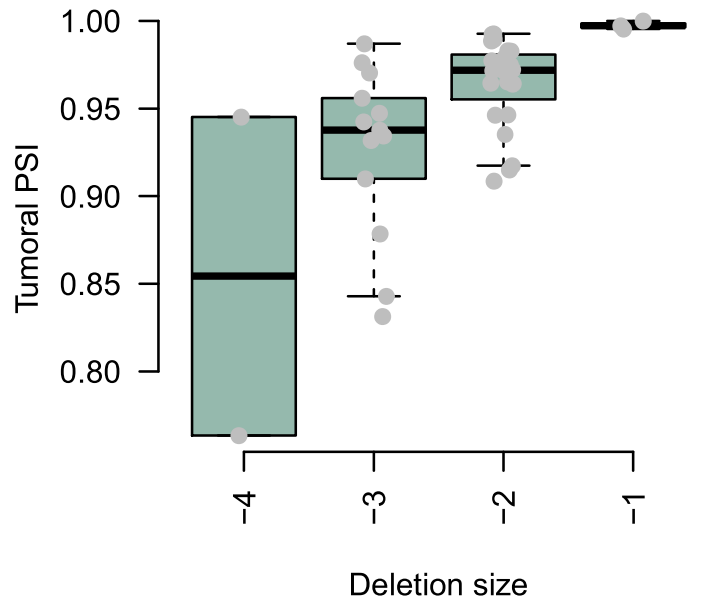
MS: chr.11.117152303 ,length: 11 ,distance: 3

gene: FANCL ,exon: 1 ,type: Constitutive



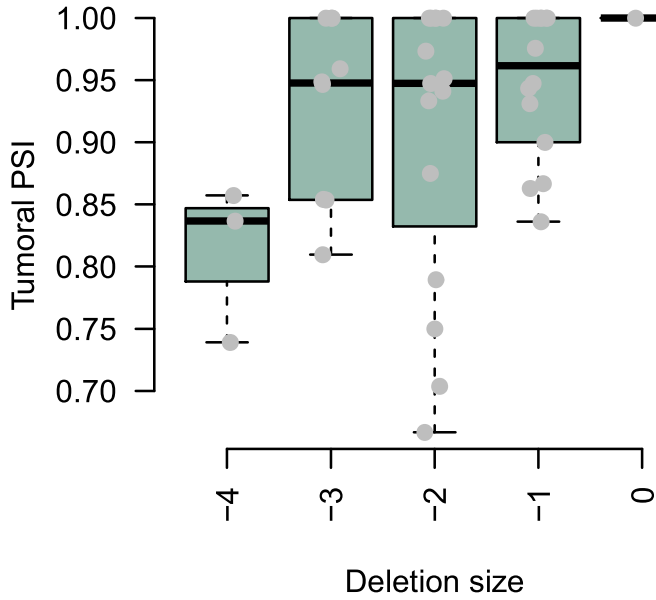
MS: chr.2.58390219 ,length: 12 ,distance: 10

gene: UBA2 ,exon: 23 ,type: Constitutive



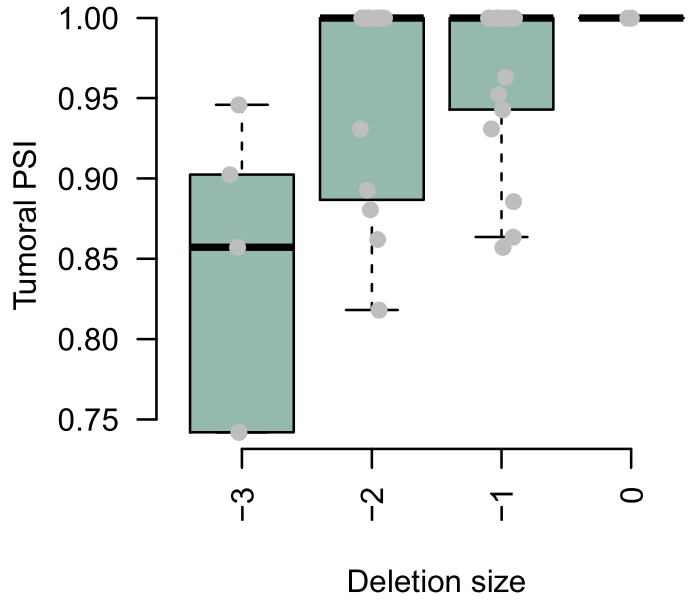
MS: chr.19.34945332 ,length: 12 ,distance: 6

gene: RAB3GAP2 ,exon: 1 ,type: Constitutive



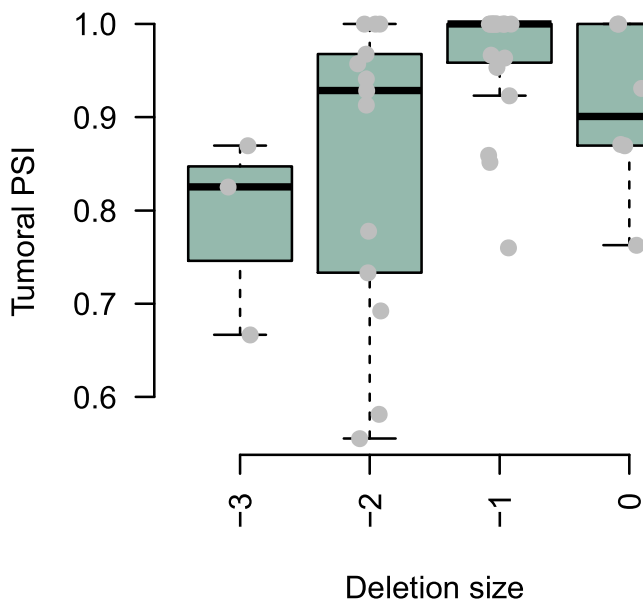
MS: chr.1.220369746 ,length: 12 ,distance: 6

gene: TBC1D2B ,exon: 124 ,type: Constitutive



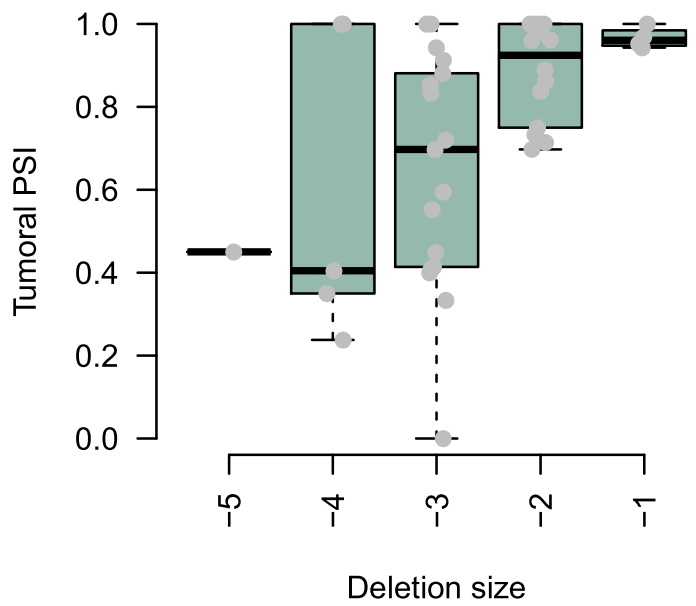
MS: chr.15.78310307 ,length: 12 ,distance: 8

gene: BOD1L ,exon: 71 ,type: Constitutive



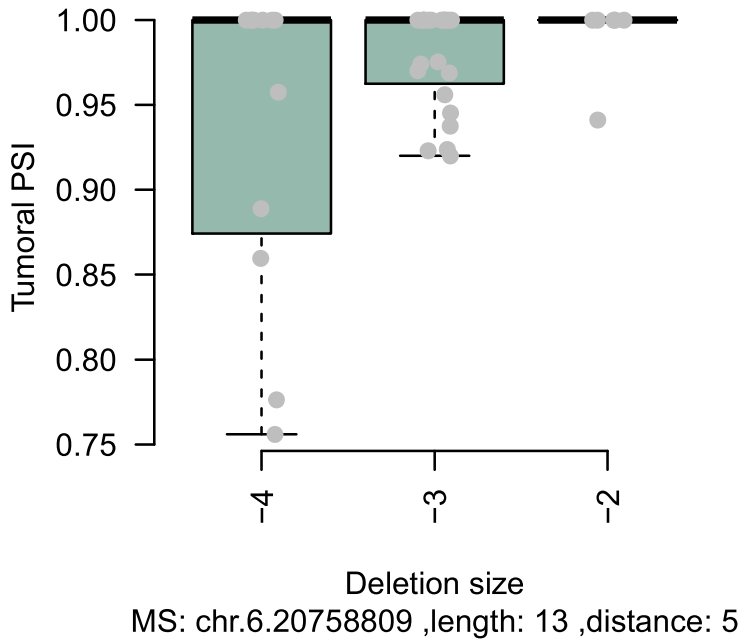
MS: chr.4.13582681 ,length: 12 ,distance: 3

gene: FRYL ,exon: 48 ,type: Constitutive

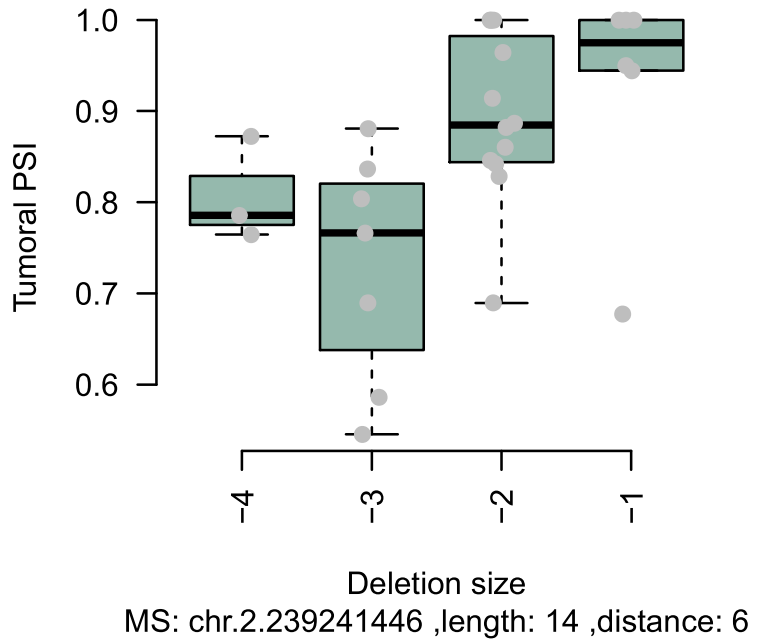


MS: chr.4.48596037 ,length: 13 ,distance: 4

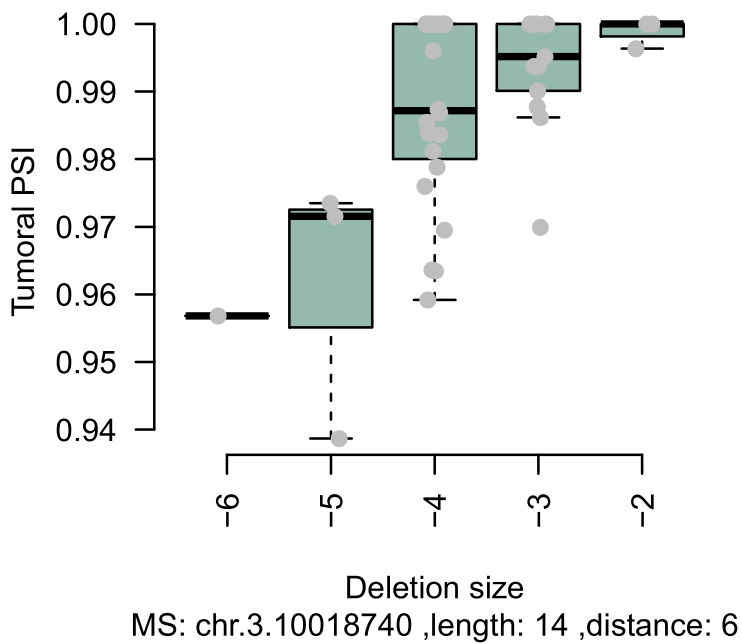
gene: CDKAL1 ,exon: 124 ,type: Constitutive



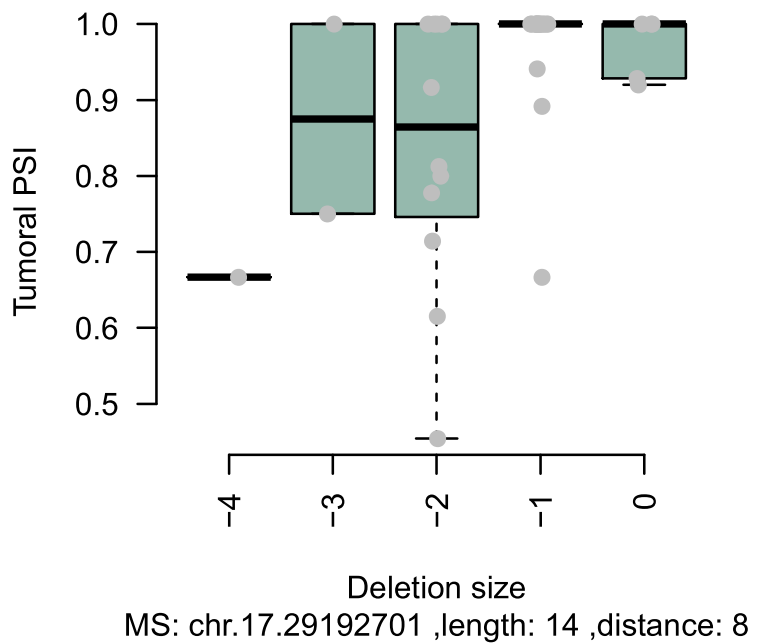
gene: TRAF3IP1 ,exon: 113 ,type: Constitutive



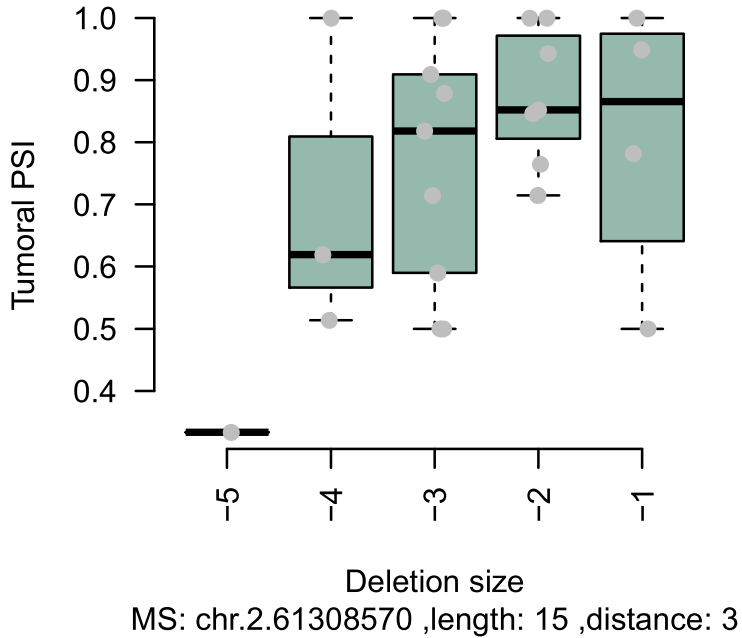
gene: TMEM111 ,exon: 80 ,type: Constitutive



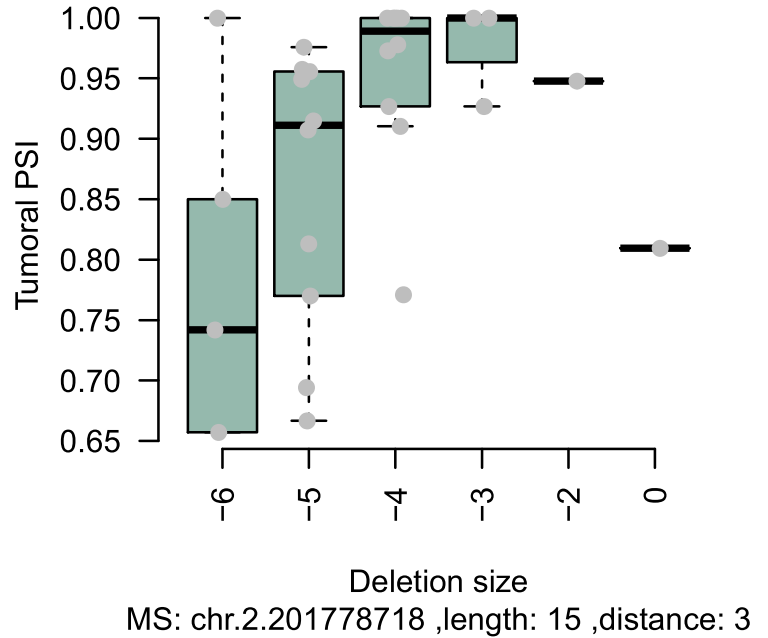
gene: ATAD5 ,exon: 12 ,type: Constitutive



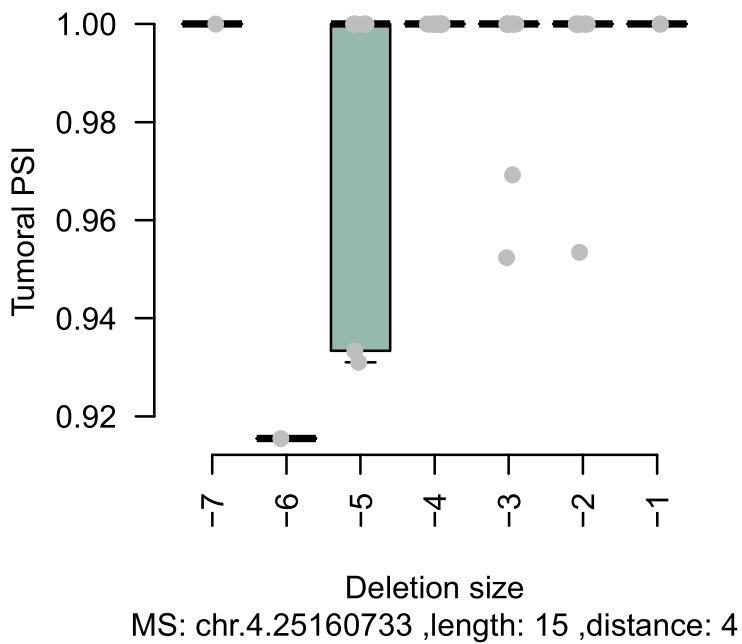
gene: KIAA1841 ,exon: 124 ,type: Constitutive



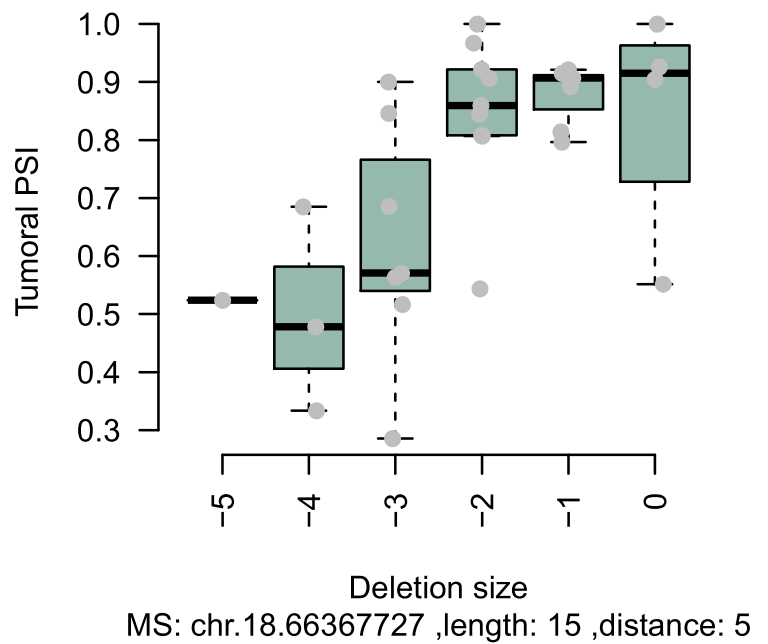
gene: ORC2L ,exon: 48 ,type: Constitutive



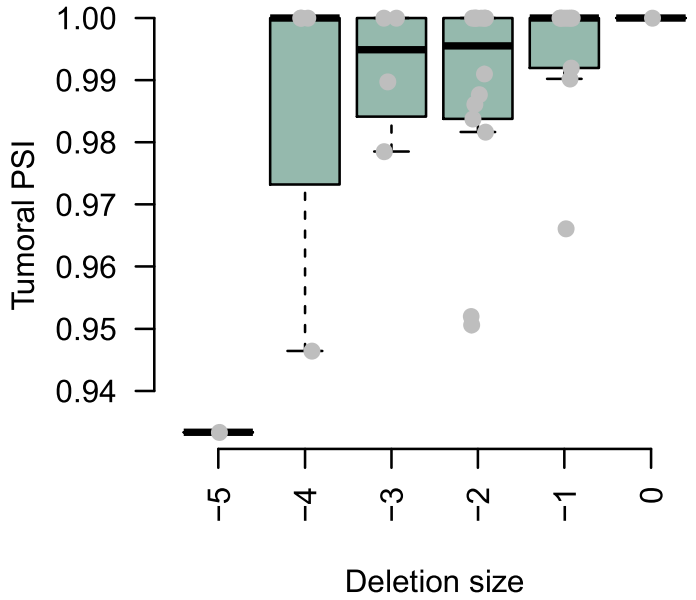
gene: SEPSECS ,exon: 69 ,type: Constitutive



gene: TMX3 ,exon: 113 ,type: Constitutive

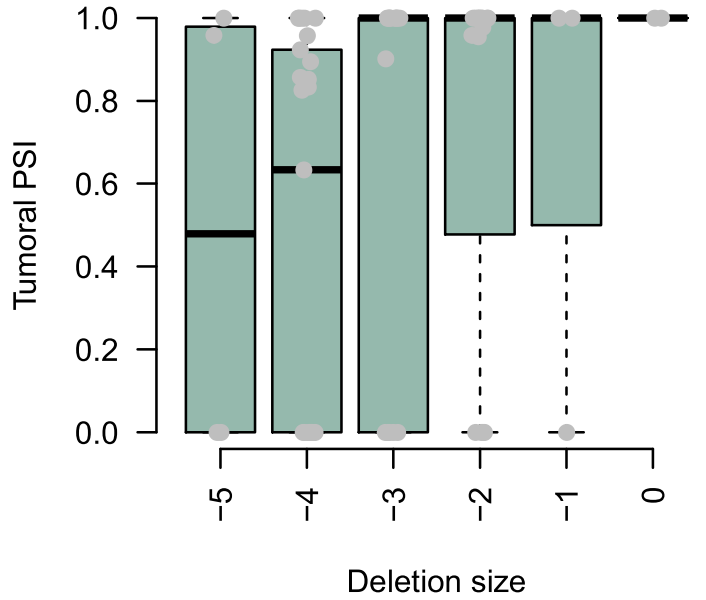


gene: EXOSC9 ,exon: 23 ,type: Constitutive



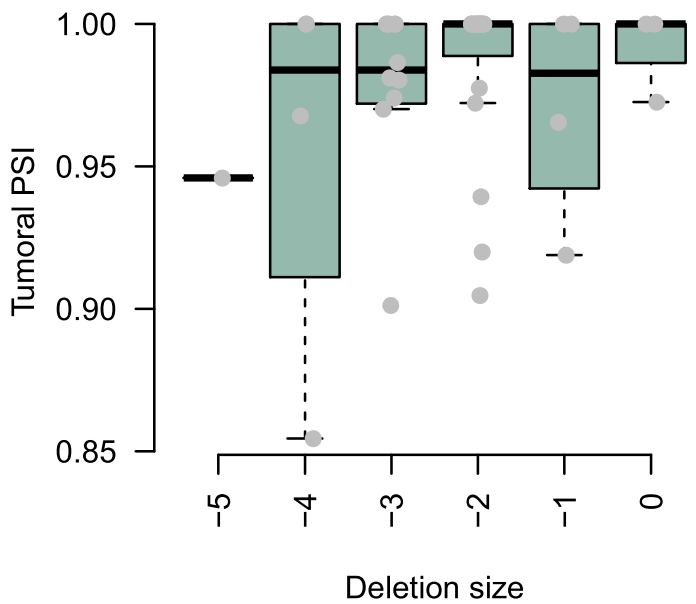
MS: chr.4.122737488 ,length: 16 ,distance: 21

gene: NEDD1 ,exon: c(80, 102) ,type: Constitutive



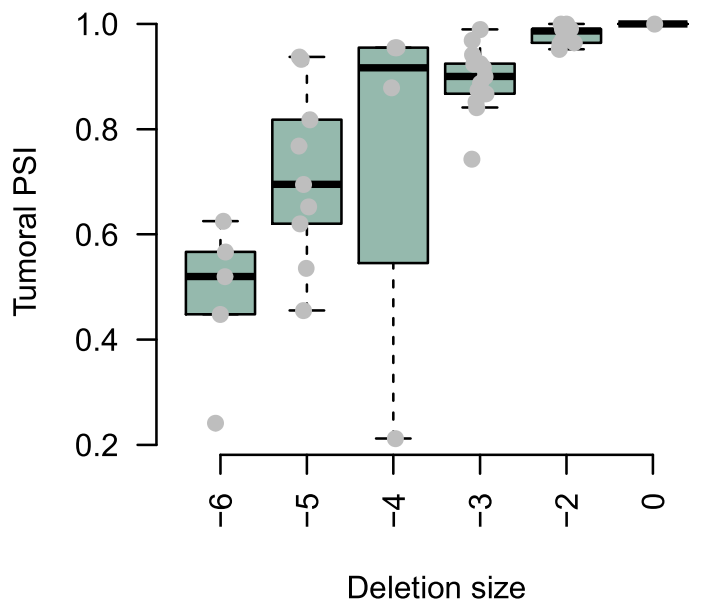
MS: chr.12.97311377 ,length: 16 ,distance: 7

gene: KIF13B ,exon: 80 ,type: Constitutive



MS: chr.8.29053762 ,length: 16 ,distance: 46

gene: HSPH1 ,exon: 146 ,type: Constitutive

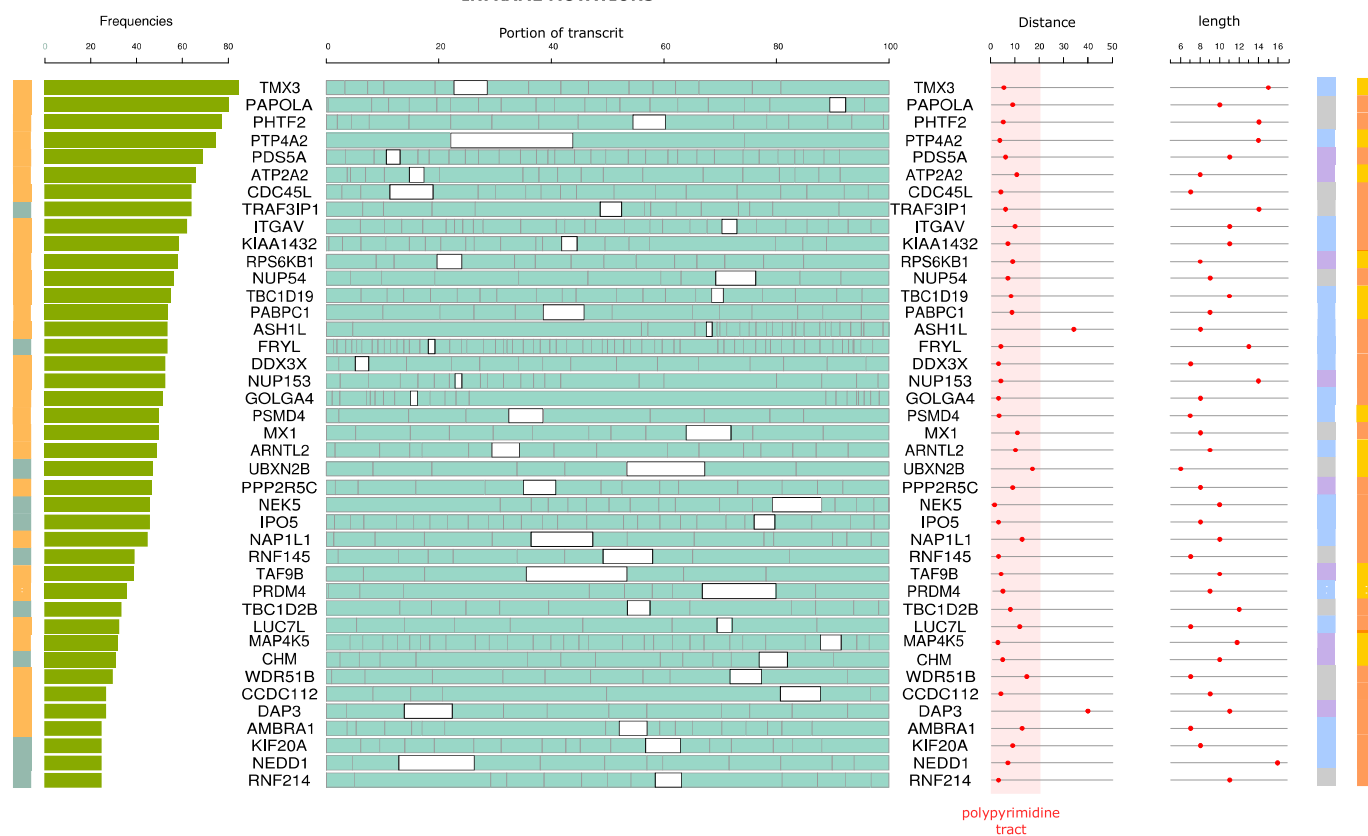


MS: chr.13.31722621 ,length: 17 ,distance: 4

FRAMESHIT MUTATIONS



INFRAME MUTATIONS

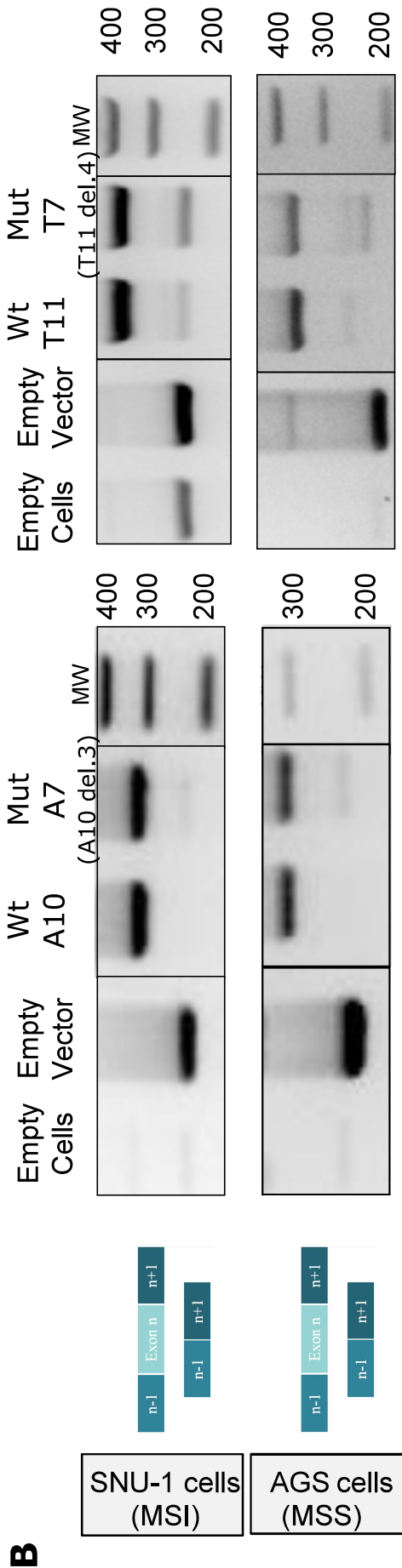


Exon Skipping Constitutive
Exon Skipping Alternative

Exon Included
Exon Skipped
NeoAntigen non-self
Self protein

Expected
Anti-tumoral
NA
Pro-tumoral

Expected
Loss of Function
NA
Functional Protein



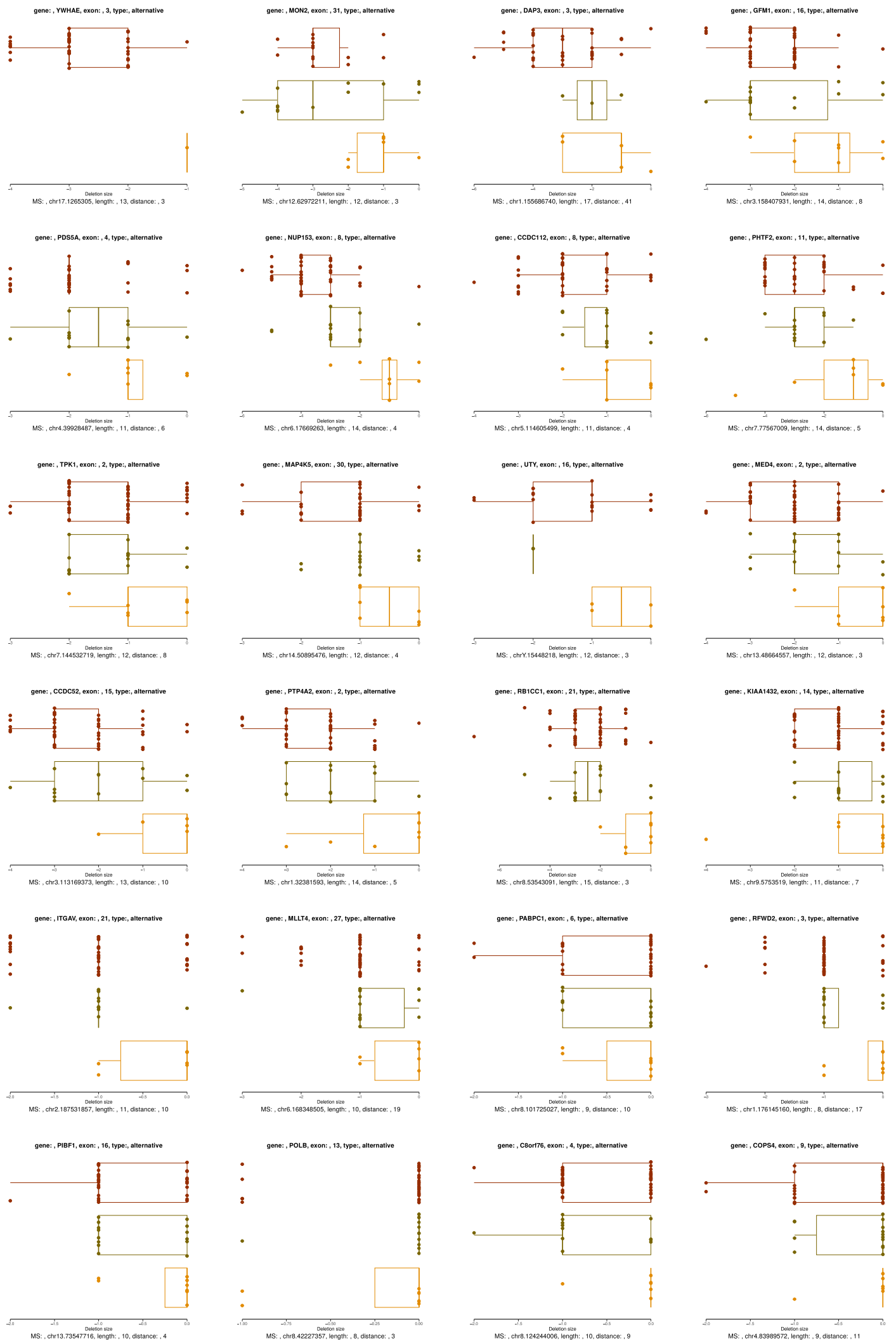
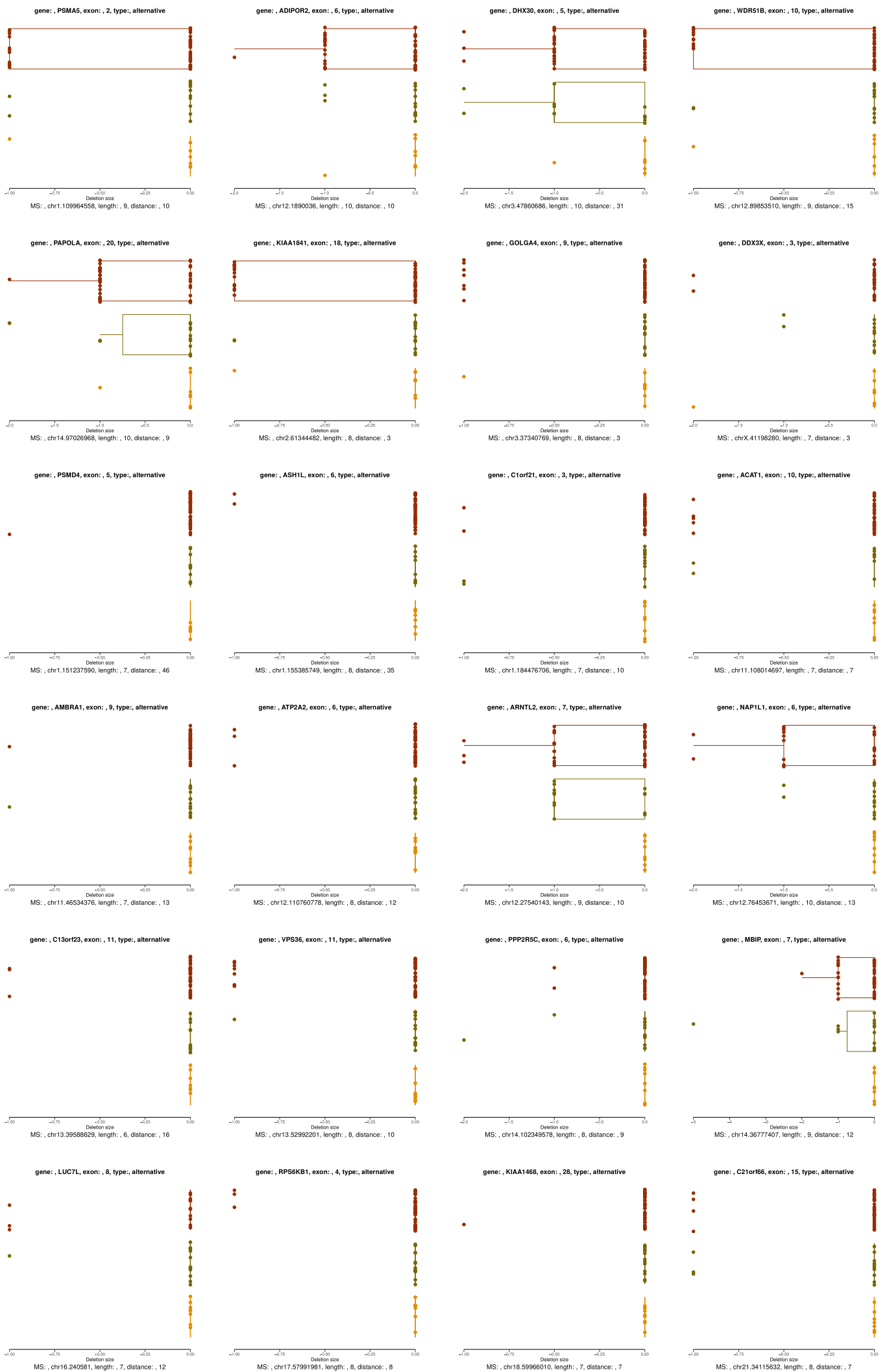
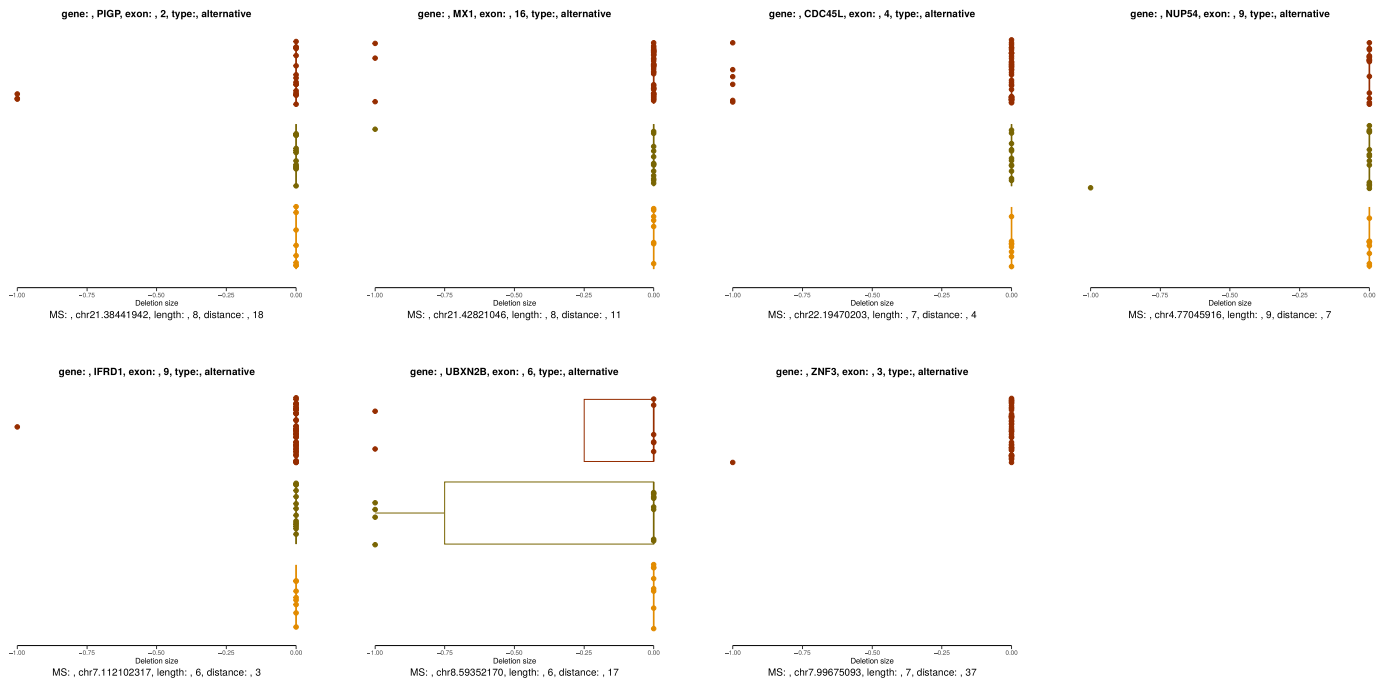
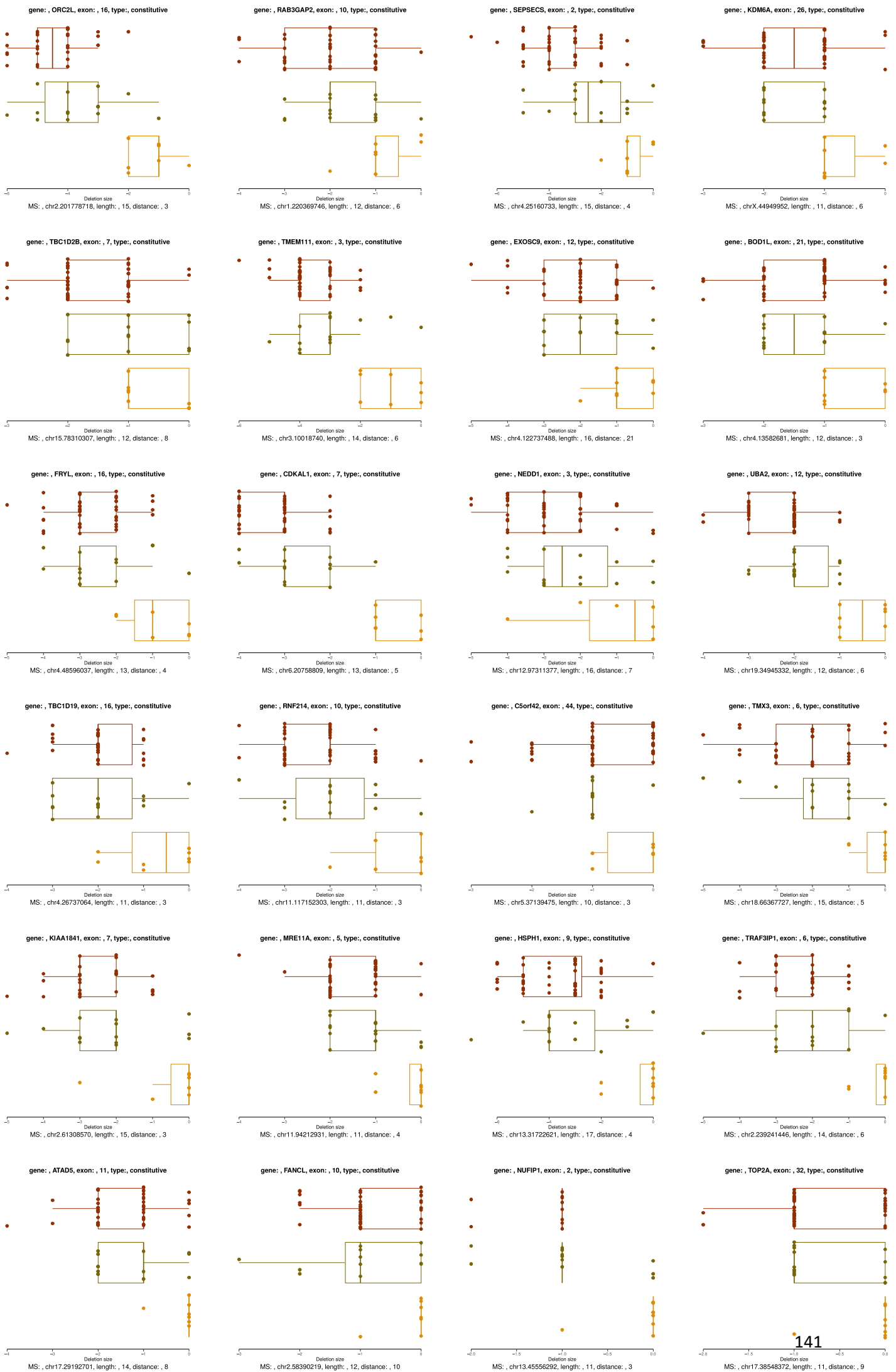
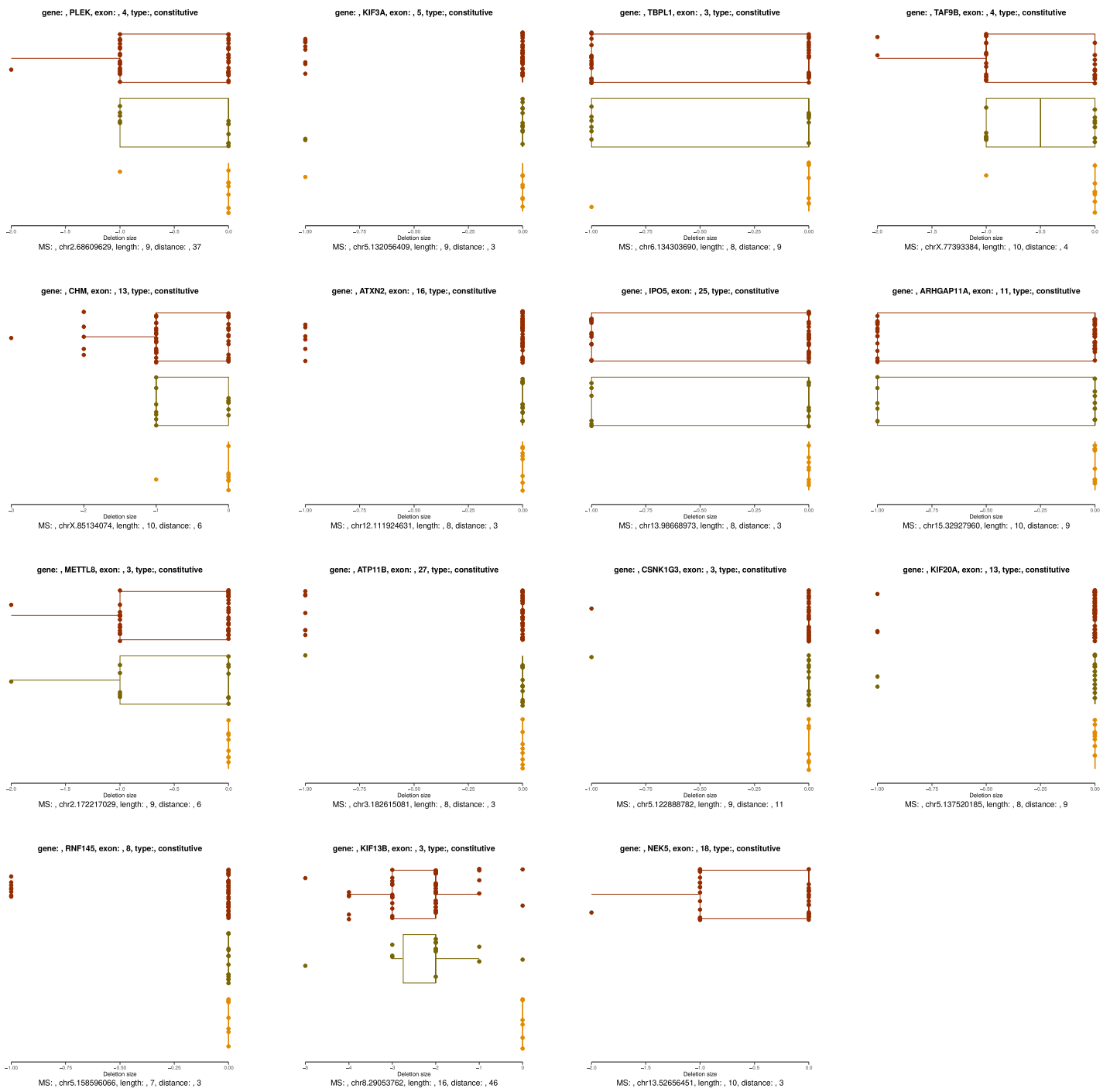


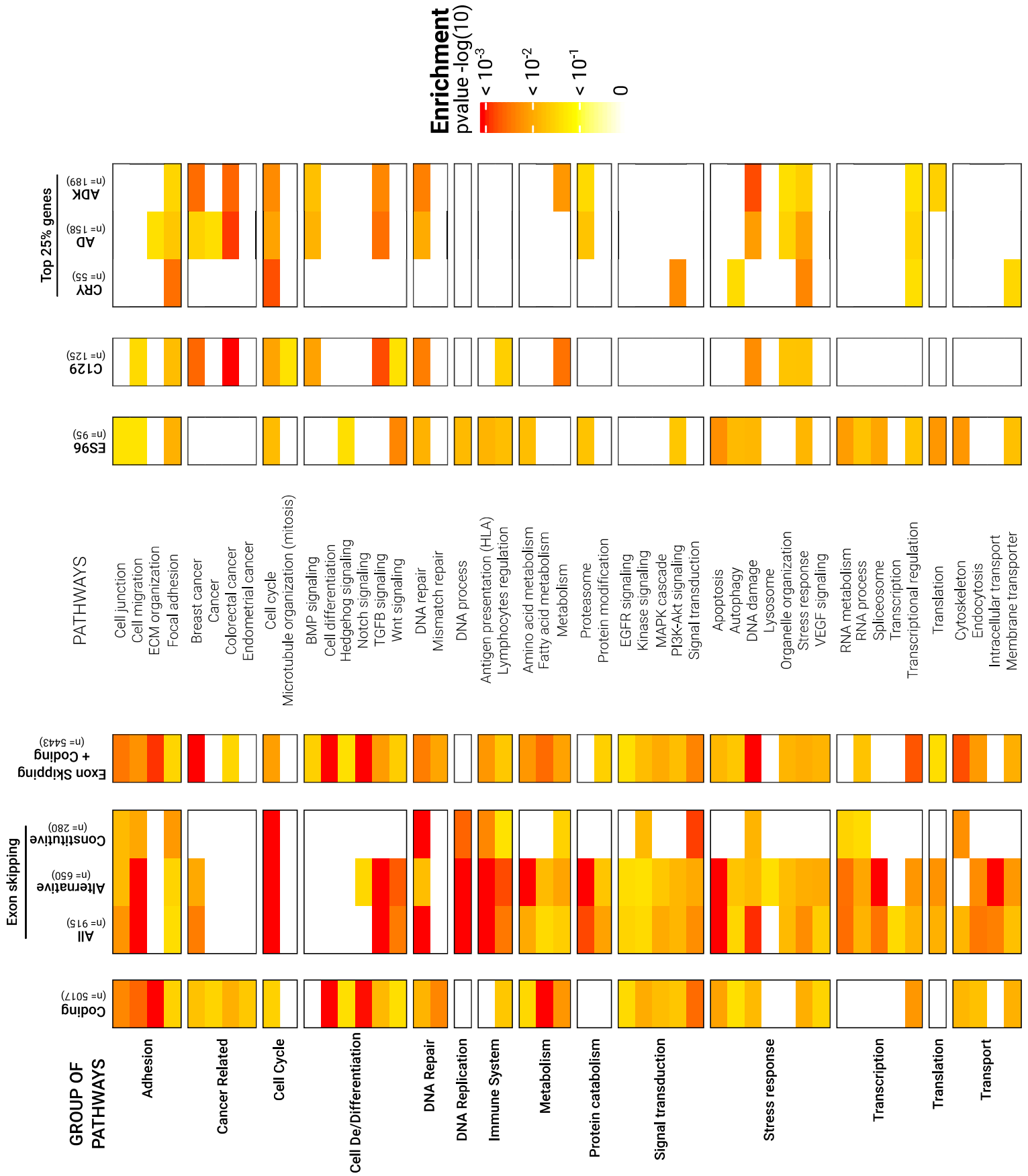
Figure Supp 8138

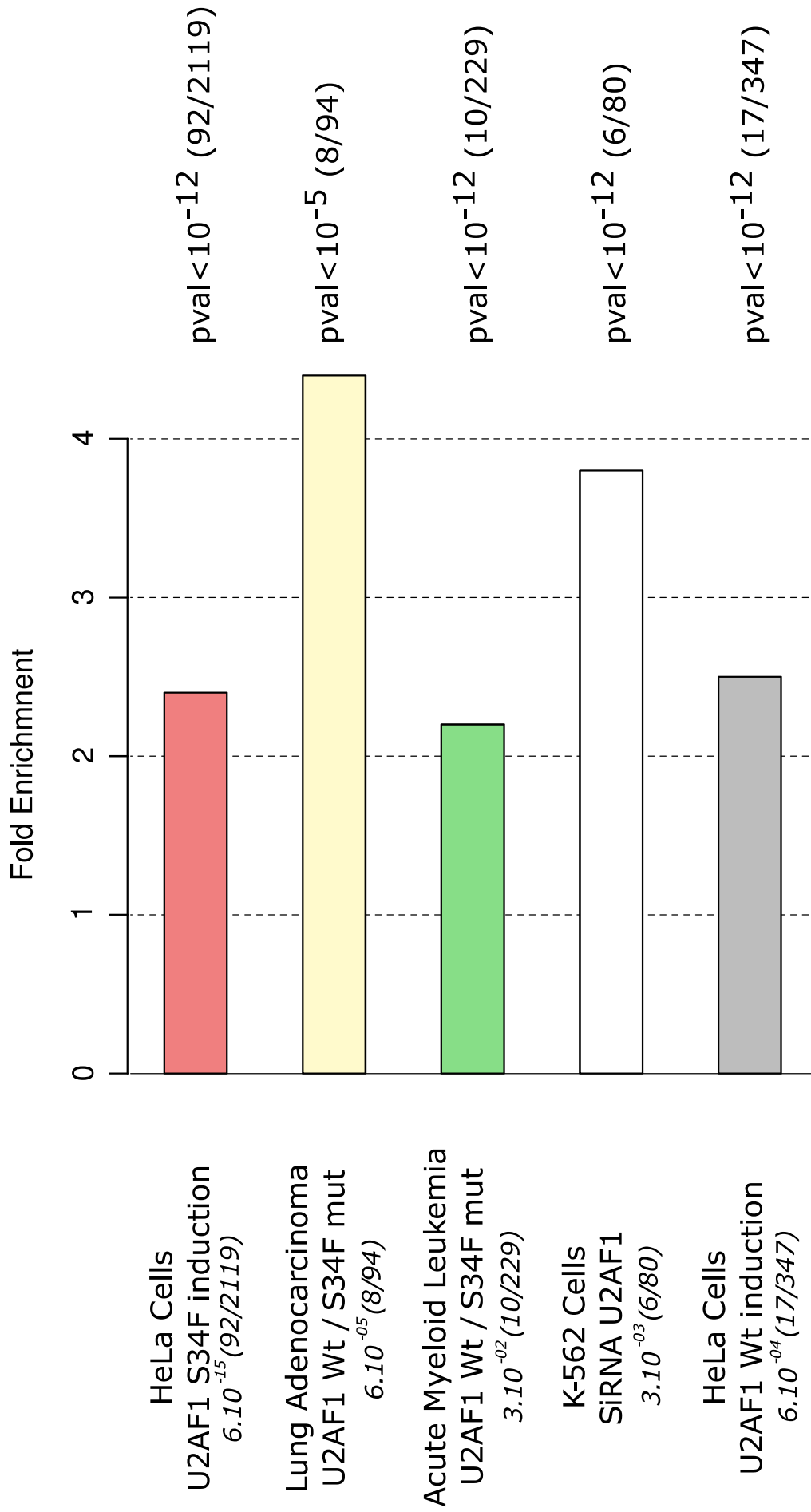




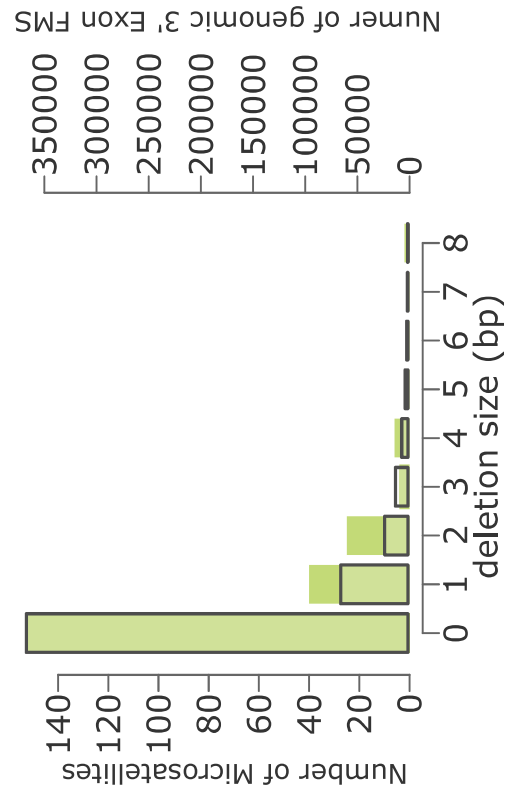
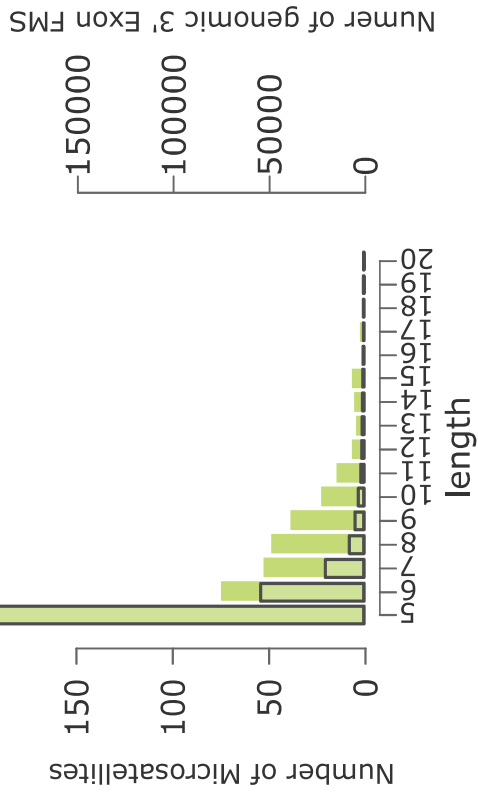




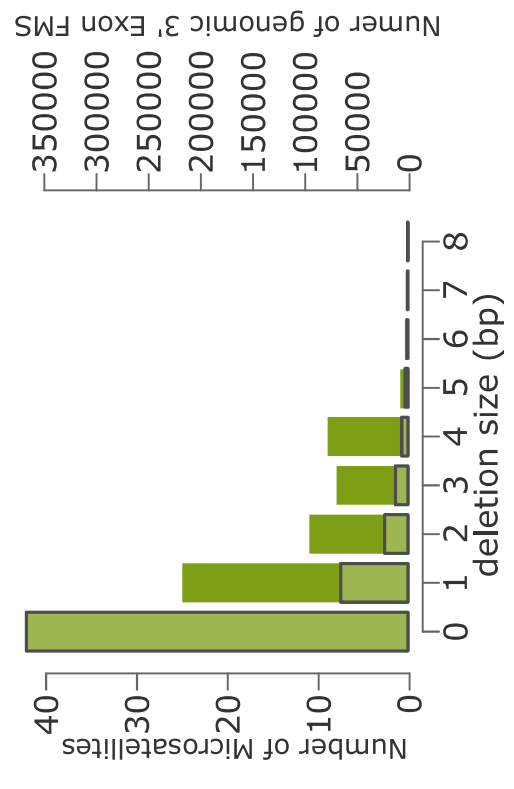
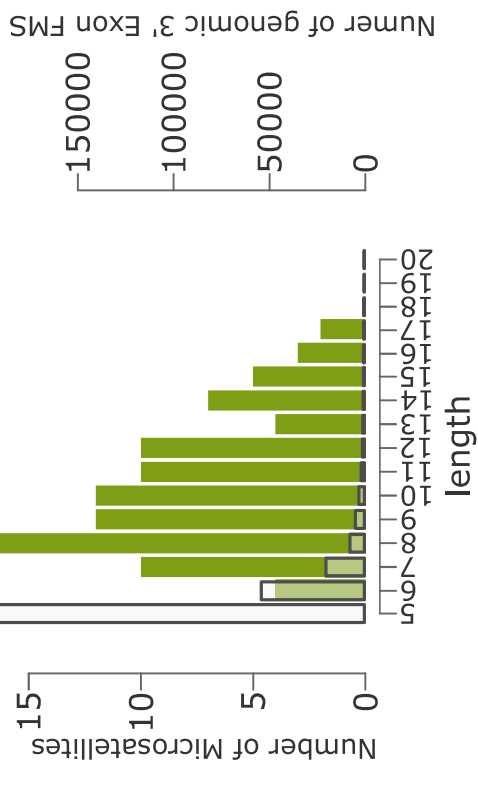


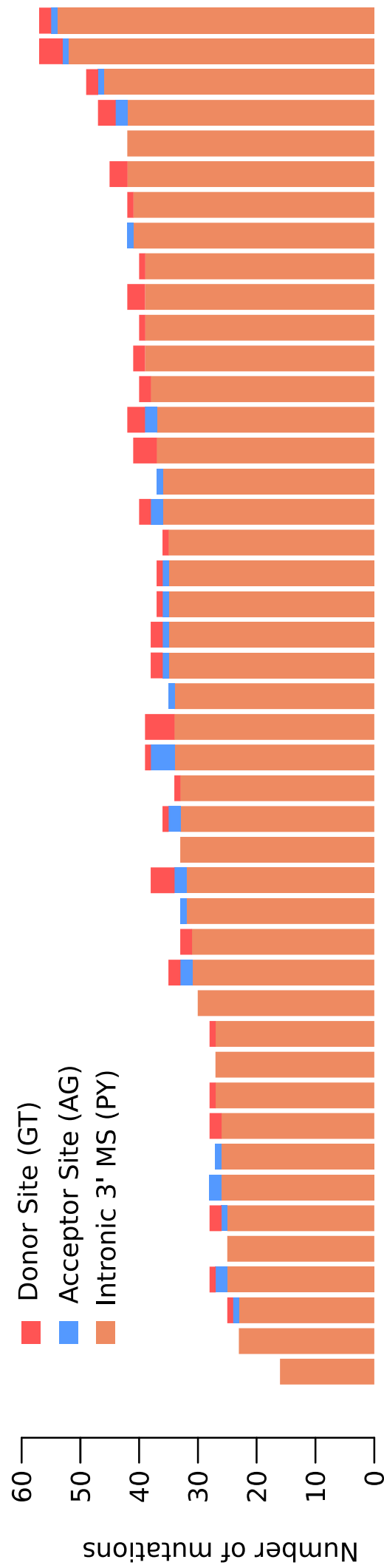


PSI independent of deletion size of MS



PSI correlated with deletion size of MS





NM_201412.1	16	240540	240569	8	10 frameshift	alternative	LUC7L	NM_201412.1:ENST00000095647.9207990.1.742574257.81.65346534.0.970297029.0.973105134.0.988116890.0.014370156.Up.msi	51,5625	0.4375	74.25	1,1875	0.991515151 0.984006734 0.782321186 n.s	0.0348959581 Up.msi	UP.MSI.ac
NM_201997.1	11	6454092	6454097	3	13 Inframe	alternative	SF1	NM_201997.1:ENST00000037228.8118811.16.7272722.99.5247547.6.514851485.0.882994374.0.934540389.1.206296404.Up.msi	215,90625	16	93.75	8.625	0.925893761 0.908	0.016334337 Up.msi	UP.MSI.ac
NM_206808.1	13	100515245	100515346	4	9 frameshift	alternative	CLVL	NM_206808.1:ENST0000003756.43564356.6.811881188.99.9594950.11.98019801.0.879298245.0.879590008.0.961148757 n.s	86,59375	5.5	107.21875	15.625	0.936485023 0.854269892 0.006156597 Down.msi	0.018306476 Up.msi	UP.MSI.ac
NM_207343.1	10	117152316	117152414	10	15 frameshift	constitutive	RNF214	NM_207343.1:ENST0000003033.29702970.L287128712.16.64356435	28.5	0.0625	15.53125	0	0.997807017	0.012842925 Up.msi	UP.MSI.ac

Line	Account	Debit	Credit	Balance
1	1000000000			
2	1000000000			
3	1000000000			
4	1000000000			
5	1000000000			
6	1000000000			
7	1000000000			
8	1000000000			
9	1000000000			
10	1000000000			
11	1000000000			
12	1000000000			
13	1000000000			
14	1000000000			
15	1000000000			
16	1000000000			
17	1000000000			
18	1000000000			
19	1000000000			
20	1000000000			
21	1000000000			
22	1000000000			
23	1000000000			
24	1000000000			
25	1000000000			
26	1000000000			
27	1000000000			
28	1000000000			
29	1000000000			
30	1000000000			
31	1000000000			
32	1000000000			
33	1000000000			
34	1000000000			
35	1000000000			
36	1000000000			
37	1000000000			
38	1000000000			
39	1000000000			
40	1000000000			
41	1000000000			
42	1000000000			
43	1000000000			
44	1000000000			
45	1000000000			
46	1000000000			
47	1000000000			
48	1000000000			
49	1000000000			
50	1000000000			
51	1000000000			
52	1000000000			
53	1000000000			
54	1000000000			
55	1000000000			
56	1000000000			
57	1000000000			
58	1000000000			
59	1000000000			
60	1000000000			
61	1000000000			
62	1000000000			
63	1000000000			
64	1000000000			
65	1000000000			
66	1000000000			
67	1000000000			
68	1000000000			
69	1000000000			
70	1000000000			
71	1000000000			
72	1000000000			
73	1000000000			
74	1000000000			
75	1000000000			
76	1000000000			
77	1000000000			
78	1000000000			
79	1000000000			
80	1000000000			
81	1000000000			
82	1000000000			
83	1000000000			
84	1000000000			
85	1000000000			
86	1000000000			
87	1000000000			
88	1000000000			
89	1000000000			
90	1000000000			
91	1000000000			
92	1000000000			
93	1000000000			
94	1000000000			
95	1000000000			
96	1000000000			
97	1000000000			
98	1000000000			
99	1000000000			
100	1000000000			

Line	Account	Debit	Credit	Balance
1	1111111111			
2	1111111111			
3	1111111111			
4	1111111111			
5	1111111111			
6	1111111111			
7	1111111111			
8	1111111111			
9	1111111111			
10	1111111111			
11	1111111111			
12	1111111111			
13	1111111111			
14	1111111111			
15	1111111111			
16	1111111111			
17	1111111111			
18	1111111111			
19	1111111111			
20	1111111111			
21	1111111111			
22	1111111111			
23	1111111111			
24	1111111111			
25	1111111111			
26	1111111111			
27	1111111111			
28	1111111111			
29	1111111111			
30	1111111111			
31	1111111111			
32	1111111111			
33	1111111111			
34	1111111111			
35	1111111111			
36	1111111111			
37	1111111111			
38	1111111111			
39	1111111111			
40	1111111111			
41	1111111111			
42	1111111111			
43	1111111111			
44	1111111111			
45	1111111111			
46	1111111111			
47	1111111111			
48	1111111111			
49	1111111111			
50	1111111111			
51	1111111111			
52	1111111111			
53	1111111111			
54	1111111111			
55	1111111111			
56	1111111111			
57	1111111111			
58	1111111111			
59	1111111111			
60	1111111111			
61	1111111111			
62	1111111111			
63	1111111111			
64	1111111111			
65	1111111111			
66	1111111111			
67	1111111111			
68	1111111111			
69	1111111111			
70	1111111111			
71	1111111111			
72	1111111111			
73	1111111111			
74	1111111111			
75	1111111111			
76	1111111111			
77	1111111111			
78	1111111111			
79	1111111111			
80	1111111111			
81	1111111111			
82	1111111111			
83	1111111111			
84	1111111111			
85	1111111111			
86	1111111111			
87	1111111111			
88	1111111111			
89	1111111111			
90	1111111111			
91	1111111111			
92	1111111111			
93	1111111111			
94	1111111111			
95	1111111111			
96	1111111111			
97	1111111111			
98	1111111111			
99	1111111111			
100	1111111111			

Supplementary table.S4

Supplementary Table S4

MS.id	gene	category	repeat.length	MF.all	MF.ADK	MF.AD	MF.CRY	MD.ADK	MD.AD	MD.CRY
chr17.1265305	YWHAE	ES96	13	1	1	NA	1	-3	NA	-1
chr6.30545853	ABCF1	coding	10	0.5	0.49	0.5	1	-1	-1	-1
chr2.201778718	ORC2L	ES96	15	0.98	1	1	0.88	-4	-4	-1
chr12.62972211	MON2	ES96	12	0.88	1	0.77	0.83	-3	-4	-1
chr1.155686740	DAP3	ES96	17	0.95	0.97	1	0.8	-3	-2	-2
chr3.158407931	GFM1	ES96	14	0.93	0.98	0.86	0.75	-2	-3	-2
chr4.39928487	PDS5A	ES96	11	0.89	0.91	0.93	0.75	-2	-2	-1
chr6.17669263	NUP153	ES96	14	0.94	0.98	0.93	0.75	-4	-3	-1
chr2.211179765	MYL1	coding	11	0.87	0.89	0.86	0.75	-1	-1	-1
chr1.220369746	RAB3GAP2	ES96	12	0.94	0.98	0.93	0.71	-2	-2	-1
chr4.25160733	SEPSECS	ES96	15	0.96	1	0.93	0.71	-4	-3	-1
chr7.77567009	PHTF2	ES96	14	0.91	0.91	1	0.71	-3	-3	-1
chrX.44949952	KDM6A	ES96	11	0.92	0.93	1	0.71	-2	-2	-1
chr15.78310307	TBC1D2B	ES96	12	0.87	0.96	0.69	0.62	-2	-1	-1
chr3.10018740	TMEM111	ES96	14	0.94	1	0.93	0.62	-4	-3	-2
chr4.122737488	EXOSC9	ES96	16	0.89	0.97	0.85	0.62	-2	-2	-1
chr4.13582681	BOD1L	ES96	12	0.86	0.88	0.93	0.62	-1	-2	-1
chr5.158630629	RNF145	coding	11	0.87	0.89	0.93	0.62	-1	-1	-1
chr4.48596037	FRYL	ES96	13	0.94	1	0.93	0.57	-3	-3	-2
chr5.114605499	CCDC112	ES96	11	0.84	0.9	0.82	0.57	-2	-1	-1
chr6.20758809	CDKAL1	ES96	13	0.95	1	1	0.57	-3	-3	-1
chr7.144532719	TPK1	ES96	12	0.8	0.82	0.86	0.57	-1	-2	-1
chr3.113377481	USF3	coding	11	0.91	0.93	1	0.57	-2	-2	-2
chr12.97311377	NEDD1	ES96	16	0.89	0.96	0.86	0.5	-3	-3	-2
chr14.50895476	MAP4K5	ES96	12	0.78	0.86	0.77	0.5	-1	-1	-1
chr19.34945332	UBA2	ES96	12	0.94	1	1	0.5	-2	-2	-1
chr4.26737064	TBC1D19	ES96	11	0.9	1	0.93	0.5	-2	-2	-2
chrY.15448218	UTY	ES96	12	0.75	0.78	1	0.5	-2	-2	-1
chr1.1290109	MXRA8	coding	7	0.36	0.36	0.3	0.5	-1	-1	-1
chr13.115057210	UPF3A	coding	9	0.58	0.58	0.64	0.5	-1	-1	-1
chr16.29825015	PRRT2	coding	9	0.39	0.44	0.15	0.5	-1	-1	-1
chr17.56435160	RNF43	coding	7	0.63	0.71	0.43	0.5	-1	-1	-1
chr19.9361740	OR7E24	coding	11	0.79	0.87	0.64	0.5	-1	-1	-1
chr2.9989570	TAF1B	coding	11	0.76	0.8	0.85	0.5	-1	-1	-1
chr11.117152303	RNF214	ES96	11	0.91	0.98	0.93	0.43	-2	-2	-1
chr13.48664557	MED4	ES96	12	0.89	0.98	0.86	0.43	-2	-2	-1
chr1.152195728	HRNR	coding	11	0.6	0.67	0.4	0.43	-1	-2	-1
chr11.63149670	SLC22A9	coding	11	0.88	0.96	0.86	0.43	-1	-1	-1
chr2.165551295	COBLL1	coding	9	0.65	0.71	0.54	0.43	-1	-1	-1
chr2.197531518	CCDC150	coding	11	0.8	0.86	0.77	0.43	-1	-1	-1
chr3.113169373	CCDC52	ES96	13	0.89	0.96	0.85	0.4	-3	-3	-2
chr1.32381593	PTP4A2	ES96	14	0.88	0.97	0.92	0.38	-2	-2	-2
chr8.53543091	RB1CC1	ES96	15	0.88	0.98	0.86	0.38	-2	-3	-1
chr9.5753519	KIAA1432	ES96	11	0.71	0.77	0.71	0.38	-1	-1	-1
chr1.89449508	RBMXL1	coding	11	0.79	0.91	0.62	0.38	-1	-1	-1
chr11.117222647	CEP164	coding	11	0.74	0.85	0.57	0.38	-1	-1	-1
chr11.26587426	MUC15	coding	9	0.42	0.42	0.42	0.38	-1	-1	-1
chr11.77784146	NDUFC2;NDUFC2-KCTD14	coding	9	0.64	0.7	0.57	0.38	-1	-1	-1
chr12.122242657	SETD1B	coding	8	0.79	0.92	0.69	0.38	-1	-1	-1
chr21.30339205	LTN1	coding	11	0.81	0.89	0.79	0.38	-2	-2	-1
chr2.187531857	ITGAV	ES96	11	0.77	0.79	0.92	0.33	-1	-1	-1
chr5.37139475	C5orf42	ES96	10	0.62	0.6	0.9	0.33	-1	-1	-1
chr6.168348505	MLLT4	ES96	10	0.76	0.83	0.71	0.33	-1	-1	-1
chr12.123351891	VPS37B	coding	9	0.55	0.54	0.67	0.33	-1	-1	-1
chr20.58587783	CDH26	coding	10	0.77	0.8	0.85	0.33	-1	-1	-1
chr18.66367727	TMX3	ES96	15	0.82	0.91	0.92	0.29	-2	-2	-1
chr2.61308570	KIAA1841	ES96	15	0.83	1	0.79	0.29	-3	-3	-2
chr8.101725027	PABPC1	ES96	9	0.33	0.31	0.42	0.29	-1	-1	-1
chr1.23240245	EPHB2	coding	9	0.48	0.48	0.57	0.29	-1	-1	-1
chr1.176145160	RFWD2	ES96	8	0.74	0.83	0.75	0.25	-1	-1	-1
chr11.94212931	MRE11A	ES96	11	0.83	0.95	0.79	0.25	-2	-2	-1
chr13.31722621	HSPH1	ES96	17	0.88	0.98	0.93	0.25	-3	-4	-2
chr13.73547716	PIBF1	ES96	10	0.58	0.67	0.57	0.25	-1	-1	-1
chr2.239241446	TRAF3IP1	ES96	14	0.86	1	0.92	0.25	-2	-2	-1
chr8.42227357	POLB	ES96	8	0.12	0.11	0.07	0.25	-1	-1	-1
chr1.159032486	AIM2	MFMC	10	0.74	0.84	0.71	0.25	-1	-1	-1
chr10.70182520	DNA2	coding	9	0.45	0.52	0.36	0.25	-1	-1	-1
chr11.104878040	CASP5	MFMC	10	0.67	0.8	0.5	0.25	-1	-1	-1
chr11.104879686	CASP5	MFMC	10	0.63	0.73	0.54	0.25	-1	-1	-1
chr11.58892376	FAM111B	coding	10	0.7	0.78	0.71	0.25	-1	-1	-1
chr13.103381995	CCDC168	coding	9	0.55	0.65	0.38	0.25	-1	-1	-1
chr14.24040435	JPH4	coding	9	0.36	0.42	0.23	0.25	-1	-1	-1
chr15.60745781	ICE2	coding	9	0.37	0.42	0.29	0.25	-1	-1	-1

Supplementary table.S4

chr15.79750585	KIAA1024	coding	8	0.34	0.39	0.21	0.25	-1	-1	-1
chr16.30016652	INO80E	coding	9	0.66	0.7	0.7	0.25	-1	-1	-2
chr17.59465991	BCAS3	coding	10	0.41	0.4	0.54	0.25	-1	-1	-1
chr18.57013193	LMAN1	coding	9	0.48	0.56	0.36	0.25	-1	-1	-1
chr19.49850472	TEAD2	coding	8	0.31	0.38	0.14	0.25	-1	-1	-1
chr2.190670539	PMS1	coding	11	0.59	0.61	0.71	0.25	-1	-1	-1
chr2.203922057	NBEAL1	coding	9	0.5	0.55	0.5	0.25	-1	-1	-1
chr3.100039735	TBC1D23	coding	9	0.44	0.51	0.31	0.25	-1	-1	-1
chr3.130733046	ASTE1	coding	11	0.84	0.93	0.86	0.25	-2	-1	-1
chr3.160156367	TRIM59	coding	10	0.45	0.5	0.38	0.25	-1	-1	-1
chr3.51417603	DOCK3	coding	7	0.49	0.51	0.57	0.25	-1	-1	-1
chr4.15938177	FGFBP1	coding	9	0.38	0.39	0.43	0.25	-1	-1	-1
chr5.145647319	RBM27	coding	9	0.33	0.35	0.33	0.25	-1	-1	-1
chr5.158630641	RNF145	coding	11	0.63	0.7	0.64	0.25	-1	-1	-1
chr5.79970914	MSH3	MFMC	8	0.52	0.59	0.46	0.25	-1	-1	-1
chr6.80751896	TTK	coding	9	0.45	0.45	0.57	0.25	-1	-1	-1
chr7.108205525	THAP5	coding	9	0.34	0.39	0.21	0.25	-1	-1	-1
chr4.83989572	COPS4	ES96	9	0.29	0.31	0.29	0.17	-1	-1	-1
chr8.124244006	C8orf76	ES96	10	0.46	0.44	0.67	0.17	-1	-1	-1
chr11.5799651	OR52N5	coding	10	0.44	0.44	0.62	0.17	-1	-1	-2
chr19.11577604	ELAVL3	coding	9	0.33	0.31	0.5	0.17	-1	-1	-1
chr1.109964558	PSMA5	ES96	9	0.23	0.27	0.15	0.14	-1	-1	-1
chr12.1890036	ADIPOR2	ES96	10	0.37	0.45	0.21	0.14	-1	-1	-1
chr17.29192701	ATAD5	ES96	14	0.77	0.88	0.77	0.14	-1	-2	-1
chr2.58390219	FANCL	ES96	12	0.62	0.68	0.67	0.14	-1	-1	-1
chr3.47860686	DHX30	ES96	10	0.48	0.52	0.54	0.14	-1	-1	-1
chr1.1116222	TTLL10	coding	8	0.32	0.36	0.25	0.14	-1	-1	-1
chr1.202407189	PPP1R12B	coding	10	0.65	0.78	0.5	0.14	-1	-1	-1
chr1.35370730	DLGAP3	coding	8	0.3	0.39	0.09	0.14	-1	-1	-1
chr10.114925316	TCF7L2	coding	9	0.35	0.37	0.42	0.14	-1	-1	-1
chr16.15802686	MYH11	coding	8	0.46	0.49	0.54	0.14	-1	-1	-1
chr19.6477250	DENND1C	coding	9	0.44	0.46	0.67	0.14	-1	-1	-1
chr20.31022441	ASXL1	coding	8	0.31	0.36	0.22	0.14	-1	-1	-1
chr4.113578482	LARP7	coding	10	0.41	0.43	0.5	0.14	-1	-1	-1
chr6.158508008	SYNJ2	coding	8	0.35	0.41	0.25	0.14	-1	-1	-1
chr7.31697913	CCDC129	coding	10	0.47	0.53	0.5	0.14	-1	-1	-1
chr12.89853510	WDR51B	ES96	9	0.23	0.29	0.14	0.12	-1	-1	-1
chr13.45556292	NUFIP1	ES96	11	0.69	1	0.79	0.12	-1	-1	-1
chr14.97026968	PAPOLA	ES96	10	0.51	0.65	0.29	0.12	-1	-2	-1
chr17.38548372	TOP2A	ES96	11	0.58	0.64	0.64	0.12	-1	-1	-1
chr2.61344482	KIAA1841	ES96	8	0.24	0.28	0.14	0.12	-1	-1	-1
chr2.68609629	PLEK	ES96	9	0.39	0.43	0.42	0.12	-1	-1	-1
chr3.37340769	GOLGA4	ES96	8	0.12	0.16	0	0.12	-1	NA	-1
chr5.132056409	KIF3A	ES96	9	0.18	0.2	0.14	0.12	-1	-1	-1
chr6.134303690	TBPL1	ES96	8	0.36	0.39	0.43	0.12	-1	-1	-1
chrX.41198280	DDX3X	ES96	7	0.1	0.07	0.17	0.12	-2	-1	-2
chrX.77393384	TAF9B	ES96	10	0.49	0.59	0.5	0.12	-1	-1	-1
chrX.85134074	CHM	ES96	10	0.54	0.61	0.54	0.12	-1	-1	-1
chr1.110906426	SLC16A4	coding	9	0.36	0.39	0.42	0.12	-1	-1	-1
chr1.115537600	SYCP1	coding	10	0.38	0.36	0.57	0.12	-1	-1	-1
chr1.27621107	WDTC1	coding	8	0.33	0.3	0.64	0.12	-1	-1	-1
chr1.45407181	EIF2B3	coding	9	0.37	0.41	0.36	0.12	-1	-1	-1
chr10.60148569	TFAM	coding	10	0.57	0.6	0.77	0.12	-1	-1	-1
chr11.126137086	SRPRA	coding	8	0.34	0.4	0.29	0.12	-1	-1	-1
chr12.66531936	TMBIM4	coding	10	0.36	0.36	0.5	0.12	-1	-1	-1
chr14.45716018	MIS18BP1	coding	11	0.64	0.77	0.54	0.12	-1	-1	-1
chr15.56736722	MNS1	coding	10	0.49	0.59	0.38	0.12	-1	-1	-1
chr16.10867202	TVP23A	coding	9	0.46	0.53	0.43	0.12	-1	-1	-1
chr17.48433966	XYLT2	coding	7	0.32	0.36	0.29	0.12	-1	-1	-1
chr17.59465980	BCAS3	coding	9	0.49	0.62	0.29	0.12	-1	-1	-1
chr19.48197890	BICRA	coding	8	0.37	0.34	0.57	0.12	-1	-1	-1
chr19.49458970	BAX	MFMC	8	0.37	0.42	0.36	0.12	-1	-1	-1
chr2.114500276	SLC35F5	coding	10	0.57	0.6	0.75	0.12	-1	-1	-1
chr2.148683685	ACVR2A	MFMC	8	0.88	0.98	1	0.12	-1	-1	-1
chr2.172549338	DYNC112	coding	9	0.36	0.4	0.36	0.12	-1	-1	-1
chr20.47858503	DDX27	coding	8	0.49	0.59	0.36	0.12	-1	-1	-1
chr20.48467300	SLC9A8	coding	10	0.48	0.6	0.29	0.12	-1	-1	-1
chr3.57509312	DNAH12	coding	9	0.4	0.39	0.57	0.12	-1	-1	-1
chr4.83785564	SEC31A	coding	9	0.54	0.6	0.57	0.12	-1	-1	-1
chr5.140049101	WDR55	coding	8	0.36	0.4	0.36	0.12	-1	-1	-1
chr6.102503431	GRIK2	coding	8	0.35	0.35	0.5	0.12	-1	-1	-1
chr6.108214754	SEC63	coding	10	0.6	0.72	0.5	0.12	-1	-1	-1
chr6.114181209	MARCKS	coding	11	0.75	0.87	0.71	0.12	-1	-1	-1
chr7.120971878	WNT16	coding	7	0.32	0.36	0.31	0.12	-1	-1	-1
chr8.103289348	UBR5	coding	8	0.49	0.5	0.69	0.12	-1	-1	-1
chr8.48866909	PRKDC	coding	10	0.37	0.29	0.64	0.12	-1	-1	-1

Supplementary table.S4

chr8.95686610	ESRP1	coding	8	0.34	0.36	0.43	0.12	-1	-1	-1
chr1.151237590	PSMD4	ES96	7	0.02	0.02	0	0	-1	NA	NA
chr1.155385749	ASH1L	ES96	8	0.04	0.05	0	0	-1	NA	NA
chr1.184476706	C1orf21	ES96	7	0.06	0.05	0.14	0	-1	-1	NA
chr11.108014697	ACAT1	ES96	7	0.12	0.12	0.22	0	-1	-1	NA
chr11.46534376	AMBRA1	ES96	7	0.03	0.02	0.07	0	-1	-1	NA
chr12.110760778	ATP2A2	ES96	8	0.04	0.07	0	0	-1	NA	NA
chr12.111924631	ATXN2	ES96	8	0.09	0.13	0	0	-1	NA	NA
chr12.27540143	ARNTL2	ES96	9	0.36	0.36	0.62	0	-1	-1	NA
chr12.76453671	NAP1L1	ES96	10	0.29	0.46	0.15	0	-1	-1	NA
chr13.39588629	C13orf23	ES96	6	0.04	0.07	0	0	-1	NA	NA
chr13.52992201	VPS36	ES96	8	0.12	0.16	0.07	0	-1	-1	NA
chr13.98668973	IPO5	ES96	8	0.3	0.31	0.43	0	-1	-1	NA
chr14.102349578	PPP2R5C	ES96	8	0.06	0.05	0.14	0	-1	-2	NA
chr14.36777407	MBIP	ES96	9	0.25	0.29	0.29	0	-1	-1	NA
chr15.32927960	ARHGAP11A	ES96	10	0.36	0.4	0.46	0	-1	-1	NA
chr16.240581	LUC7L	ES96	7	0.11	0.2	0.07	0	-1	-1	NA
chr17.57991981	RPS6KB1	ES96	8	0.04	0.07	0	0	-1	NA	NA
chr18.59966010	KIAA1468	ES96	7	0.01	0.02	0	0	-1	NA	NA
chr2.172217029	METTL8	ES96	9	0.34	0.38	0.43	0	-1	-1	NA
chr21.34115632	C21orf66	ES96	8	0.11	0.09	0.21	0	-1	-1	NA
chr21.38441942	PIGP	ES96	8	0.07	0.16	0	0	-1	NA	NA
chr21.42821046	MX1	ES96	8	0.06	0.07	0.07	0	-1	-1	NA
chr22.19470203	CDC45L	ES96	7	0.09	0.13	0	0	-1	NA	NA
chr3.182615081	ATP11B	ES96	8	0.12	0.16	0.07	0	-1	-1	NA
chr4.77045916	NUP54	ES96	9	0.03	0	0.07	0	NA	-1	NA
chr5.122888782	CSNK1G3	ES96	9	0.03	0.03	0.07	0	-1	-1	NA
chr5.137520185	KIF20A	ES96	8	0.07	0.07	0.14	0	-1	-1	NA
chr5.158596066	RNF145	ES96	7	0.12	0.17	0	0	-1	NA	NA
chr7.112102317	IFRD1	ES96	6	0.01	0.02	0	0	-1	NA	NA
chr8.29053762	KIF13B	ES96	16	0.86	0.96	0.93	0	-2	-2	NA
chr8.59352170	UBXN2B	ES96	6	0.2	0.25	0.29	0	-1	-1	NA
chr1.14108748	PRDM2	MFMC	9	0.31	0.33	0.43	0	-1	-1	NA
chr1.28785729	PHACTR4	coding	10	0.44	0.52	0.43	0	-1	-1	NA
chr1.6257784	RPL22	coding	8	0.83	0.84	NA	0	-1	NA	NA
chr1.91967356	CDC7	coding	9	0.31	0.43	0.07	0	-1	-1	NA
chr1.93667515	CCDC18	coding	9	0.3	0.33	0.38	0	-1	-1	NA
chr10.111893349	ADD3	coding	8	0.45	0.54	0.36	0	-1	-1	NA
chr10.118618627	ENO4	coding	8	0.48	0.6	0.38	0	-1	-1	NA
chr10.64159512	ZNF365	coding	8	0.37	0.42	0.43	0	-1	-1	NA
chr10.97918855	ZNF518A	coding	9	0.36	0.44	0.29	0	-1	-1	NA
chr14.45693721	MIS18BP1	coding	9	0.31	0.35	0.36	0	-1	-1	NA
chr14.74205772	ELMSAN1	coding	7	0.31	0.43	0.14	0	-1	-1	NA
chr16.19725705	KNOP1	coding	10	0.31	0.35	0.36	0	-1	-1	NA
chr17.26653806	TMEM97	coding	10	0.33	0.35	0.46	0	-1	-1	NA
chr18.30913142	CCDC178	coding	10	0.66	0.76	0.64	0	-1	-1	NA
chr2.165365287	GRB14	MFMC	9	0.3	0.36	0.25	0	-1	-1	NA
chr2.203420129	BMPR2	coding	7	0.37	0.41	0.43	0	-1	-1	NA
chr2.27248516	MAPRE3	coding	8	0.32	0.36	0.38	0	-1	-1	NA
chr20.4850568	SLC23A2	coding	9	0.47	0.49	0.69	0	-1	-1	NA
chr22.32334104	C22orf24	coding	9	0.49	0.48	0.79	0	-1	-1	NA
chr3.112253058	ATG3	coding	10	0.35	0.4	0.38	0	-1	-1	NA
chr3.129155547	MBD4	MFMC	10	0.33	0.4	0.31	0	-1	-1	NA
chr3.136573485	SLC35G2	coding	9	0.38	0.49	0.25	0	-1	-1	NA
chr3.142274739	ATR	coding	10	0.43	0.56	0.29	0	-1	-1	NA
chr3.183665256	ABCC5	coding	9	0.31	0.38	0.29	0	-1	-1	NA
chr3.30691871	TGFBR2	MFMC	10	0.76	0.95	0.57	0	-1	-1	NA
chr4.56336953	CLOCK	coding	9	0.47	0.57	0.43	0	-1	-1	NA
chr4.57220268	AASDH	coding	10	0.38	0.46	0.36	0	-1	-2	NA
chr5.131931451	RAD50	MFMC	9	0.39	0.51	0.23	0	-1	-1	NA
chr5.140795208	PCDHGA10	coding	11	0.51	0.62	0.5	0	-1	-2	NA
chr5.147499874	SPINK5	coding	10	0.35	0.43	0.29	0	-1	-1	NA
chr5.147861098	HTR4	coding	10	0.48	0.56	0.54	0	-1	-1	NA
chr5.162917425	HMMR	MFMC	9	0.31	0.33	0.43	0	-1	-1	NA
chr5.175796243	ARL10	coding	10	0.32	0.38	0.33	0	-1	-1	NA
chr6.111587360	MFSD4B	coding	9	0.51	0.6	0.5	0	-1	-1	NA
chr7.151874147	KMT2C	coding	9	0.36	0.43	0.31	0	-1	-1	NA
chr7.77423459	TMEM60	coding	9	0.39	0.44	0.43	0	-1	-1	NA
chr8.145689658	CYHR1	coding	9	0.4	0.44	0.54	0	-1	-1	NA
chr8.37791833	GOT1L1	coding	9	0.31	0.38	0.29	0	-1	-1	NA
chr9.96422611	PHF2	coding	8	0.39	0.47	0.36	0	-1	-1	NA
chr13.52656451	NEK5	ES96	10	0.45	0.45	NA	NA	-1	NA	NA
chr7.99675093	ZNF3	ES96	7	0.02	0.02	NA	NA	-1	NA	NA
chr10.81841421	TMEM254	coding	10	0.4	0.4	NA	NA	-1	NA	NA

Supp Table 5

gene	Stem Vs tuft cells	Stem Vs Enterocytes	Stem Vs enteroendocrine	stem Vs Secretory progenitor/deep crypt secretory cells/goblet (overlap)	Stem Vs Absorptive Progenitor
AGAP1	overlap	overlap	no	no	overlap
ALPK1	no	no	no	no	overlap
AMBRA1	overlap	overlap	no	no	overlap
ARAP1	no	overlap	no	overlap	no
ARAP2	overlap	overlap	no	overlap	no
ARHGAP11A	no	no	no	no	overlap
ARHGEF7	no	overlap	overlap	no	no
ARID4B	no	no	no	no	overlap
ARNT	overlap	no	no	overlap	overlap
ASH1L	no	no	no	no	overlap
ATF2	overlap	overlap	overlap	overlap	overlap
ATXN2	no	overlap	no	no	overlap
ATXN7L1	overlap	overlap	no	no	no
BPNT1	no	no	no	overlap	no
COBL	no	no	no	overlap	no
CPEB2	no	overlap	no	overlap	no
CRYZL1	no	no	no	no	overlap
CSDE1	no	overlap	no	no	no
CSNK1G3	no	no	overlap	no	no
CSPP1	no	overlap	no	overlap	overlap
CTNND1	overlap	overlap	overlap	overlap	overlap
CUL1	overlap	overlap	no	overlap	overlap
DBT	no	no	no	no	overlap
DHX30	overlap	overlap	no	overlap	overlap
DIP2A	no	overlap	no	no	no
DLG1	no	no	no	no	overlap
DRAM2	no	overlap	no	overlap	overlap
DYNC112	no	no	overlap	no	no
E2F6	no	no	no	no	overlap
EIF3M	no	no	no	overlap	no
EIF4A2	overlap	no	overlap	overlap	no
EIF4G3	overlap	overlap	no	overlap	overlap
ENTPD5	no	overlap	no	no	overlap
EPS15	no	no	no	no	overlap
ERC1	overlap	overlap	no	no	overlap
ESCO1	overlap	no	no	overlap	no
FAM135A	no	overlap	no	no	overlap
FAM13B	no	overlap	overlap	no	no
FAM162A	no	no	no	no	overlap
FAM76A	no	no	no	overlap	overlap
FBXW11	no	no	no	no	overlap
FDPS	overlap	no	no	no	no
FGGY	overlap	no	overlap	no	no
FRYL	no	no	no	no	overlap
GOLGA2	no	no	no	overlap	no
GOLGA4	no	overlap	no	overlap	overlap
GPBP1	no	no	no	overlap	no
GTF2A2	overlap	overlap	no	no	overlap
HDHD2	no	overlap	no	no	no
HMGCS1	no	overlap	no	no	no
HNRNPD	no	no	overlap	overlap	overlap
HNRNPH1	no	no	no	overlap	overlap
HNRNPM	no	no	no	overlap	no
IMMT	overlap	no	overlap	overlap	overlap
KCTD3	no	no	no	no	overlap
LPIN2	no	no	no	no	overlap
LSM14A	no	no	overlap	no	no
MAP4K3	no	overlap	no	overlap	no
METAP1	no	no	no	overlap	overlap
MPND	overlap	overlap	no	overlap	overlap
MPZL1	no	no	overlap	no	no
MRPS18C	no	no	no	overlap	no
MYO6	no	no	no	no	overlap
NACA	no	overlap	no	overlap	no
NAMPT	no	no	no	no	overlap
NCOA6	no	no	no	no	overlap
NDUFAF2	no	overlap	no	overlap	overlap

NDUFS4	overlap	no	overlap	overlap	no
NEDD4L	no	overlap	no	no	overlap
NFIA	overlap	overlap	no	overlap	overlap
NIPA2	overlap	no	no	overlap	overlap
NOL10	no	no	no	no	overlap
NOX1	overlap	no	no	no	no
NQO2	no	overlap	no	no	no
NUP54	no	no	no	overlap	no
OPA1	no	no	no	overlap	no
PAM	no	no	overlap	no	overlap
PAPOLA	no	no	no	no	overlap
PARK7	no	no	no	no	overlap
PCCA	no	no	no	no	overlap
PDCD6IP	no	no	no	overlap	no
PDIA5	no	no	no	no	overlap
PDXDC1	overlap	no	no	overlap	overlap
PEX2	no	overlap	no	overlap	no
PHF3	overlap	overlap	overlap	overlap	overlap
PKN2	overlap	overlap	no	overlap	overlap
PLEKHA5	no	overlap	no	no	overlap
PLSCR1	no	no	no	no	overlap
PPA2	no	overlap	no	overlap	overlap
PPHLN1	overlap	overlap	no	no	overlap
PPP1R12A	no	overlap	no	no	overlap
PPP2R2A	no	no	no	no	overlap
PPP4R2	no	overlap	no	overlap	no
PTGES3	no	overlap	no	no	overlap
PTK2	overlap	no	no	no	no
PTPN12	no	no	no	no	overlap
PTPN13	no	overlap	no	no	overlap
RBL2	no	no	no	no	overlap
RBM5	no	no	no	no	overlap
RC3H1	overlap	overlap	overlap	overlap	overlap
RNF138	no	no	overlap	no	no
RNF145	no	no	no	no	overlap
RNF214	overlap	overlap	no	no	overlap
SAE1	no	no	no	no	overlap
SEC31A	overlap	overlap	no	overlap	overlap
SECISBP2L	overlap	overlap	no	no	no
SEL1L3	no	overlap	no	no	no
SFT2D2	no	overlap	no	no	overlap
SIPA1L1	no	no	no	overlap	no
SLC10A7	no	no	no	overlap	no
SLC12A2	no	no	no	no	overlap
SLC35A1	overlap	overlap	no	overlap	overlap
SLC44A1	overlap	overlap	no	overlap	no
SMC6	no	no	no	no	overlap
ST14	overlap	overlap	no	overlap	no
STAG1	overlap	no	no	overlap	no
STK39	no	no	no	no	overlap
STX8	no	no	no	no	overlap
SUMO1	no	overlap	no	overlap	overlap
SYNE2	overlap	overlap	no	no	overlap
TANC1	overlap	no	overlap	overlap	overlap
TLK2	overlap	overlap	no	no	overlap
TM9SF2	no	no	no	overlap	no
TMEM14C	no	no	no	overlap	no
TMEM161B	no	no	no	overlap	no
TOP2B	overlap	overlap	no	overlap	overlap
TTC3	overlap	overlap	overlap	overlap	overlap
UCHL3	no	no	no	no	overlap
UFM1	no	no	no	no	overlap
UHMK1	no	no	no	overlap	no
USO1	no	no	no	no	overlap
USP34	overlap	overlap	no	overlap	overlap
UTY	no	no	no	overlap	no
YTHDF3	overlap	overlap	overlap	no	overlap

Legends to Supplemental Figures and tables

Supplemental Figure S1:

(A) Distribution of mutational frequency (left) and number of unstable MS (right) by samples according to the MS repeat length; samples are sorted by condition. (B) Boxplot of cumulative deletion events in MS according to their region and condition. Statistic: Student test, * P-value < 0.05, ** P-value < 1.10^{-3} , *** P-value < 1.10^{-4} , **** P-value < 1.10^{-5} .

Supplemental Figure S2:

(A) The Percent Spliced In (PSI) which corresponds to the fraction of isoforms that include the exon (ranging from 0 to 1) was calculated for all samples and was used to identify the significant deregulated exons skipping events (two-sided adjusted *t*-test) between normal tissue and tumours. (B) The same approach has been used in order to identify the significant deregulated exon skipping events between MSI tumours and MSS tumours.

Supplemental Figure S3:

Barplots represent the number of exon exclusion (A) or inclusion (B) events in MSI and MSS tumors (C, D) compared to the normal tissue (left) or in MSI tumors compared to MSS tumors (right) (see for details Supplementary Table S1). All the candidate exons are flanked by at least one microsatellite in 3' part of the previous intron (up to 50 bp in intronic sequence from the AG acceptor splicing site, 3' exon-FMS). The majority of deregulated exons corresponds to exon skipping events in MSI tumors (surrounded by a red rectangle).

Supplemental Figure S4:

(A) Distribution of exclusion reads according to inclusion reads (log₂ scale) for 2 examples of constitutive exons (left) and for 2 examples of alternative exons (right) deregulated between tumors and normal tissue. (B) Two-dimensional histograms show the number of exclusion reads according to inclusion reads for all exons. The density of exons is displaying by a logarithmic color code. The upper panel represent exons classified as alternative and the lower panel classified as constitutive. This classification is based on the interpretation of data from all normal tissues (n=133). (C) Distribution of exclusion reads according to inclusion reads (log₂ scale) in all normal tissues (n=133) for 8 examples of exons classified as alternative (n=6) or constitutive (n=2).

Supplemental Figure S5:

Boxplots represent the Percent Spliced Excluded (PSI) according to the size of deletion for each 96 microsatellites displaying a significant correlation (named ES96) (see for detail supplementary table S2).

Supplemental Figure S6:

Plasmids containing the mini-genes of interest were transfected into AGS and SNU-1 cells, and mini-gene derived transcripts were analyzed by RT-PCR. The figures illustrate the results from 3 independent experiments with similar results. The "Empty Cells" lanes represent results from un-transfected cells. A faint band corresponds to the endogenous PolR2G mRNA. A more abundant band of identical size is observed in the lanes corresponding to the empty vector (EV) that carries the inserted same genomic sequence.

Supplemental Figure S7:

Figure represents the biological consequence of exon skipping leading to frame shift (upper panel) or inframe (lower). Left part represents the type of exon (determined as constitutive or alternative in normal tissue) and their mutational frequency. The center showed the protein representation, the proportion of protein lost and size of neoantigenic tail. Right, the distance of microsatellite from 3' splice site (AG) position, the length of the intronic 3' MS, the expected function impact are indicated (n=96, Details of all proteins are available in Supplementary Table S2).

Supplemental Figure S8:

Boxplots represent the size of deletion for each 96 microsatellites displaying a significant correlation (ES96), according to the conditions; brown: Adenocarcinoma, green: adenoma, yellow: dMMR crypts. The microsatellites are classified by exon type (alternative or constitutive) and ordered by mutational frequency in the dMMR crypts (see for detail supplementary table S4).

Supplemental Figure S9:

Heatmap displaying pathway analysis in global coding, exon skipping signature, exon skipping + coding, ES96, (C129), top 25% most mutated genes (ES96 + C129) through tumorigenesis; colors represent enrich pathway pvalue, $-\log(10)$. Enrichment was realized with EnrichR - BioPlanet 2019. Term were sort by pvalue (< 0.05) and regroup first in pathway (rows, middle) then in group of pathways (rows, left). Columns are sorted by enriched condition parameters.

Supplemental Figure S10:

Barplot represents the fold enrichment of U2AF exons-signature identified by 3 studies (Brooks et al 2014, Ilagan et al 2015, Kim et al 2015) in our MSI exon-signature (see for detail supplementary table S3).

Supplemental Figure S11:

Upper panel represent the distribution of microsatellites according to the repeat length and the lower panel represent the distribution of microsatellites according to the deletion size. Microsatellites whose somatic deletions in MSI CRC influence splicing are not framed. Left, distribution of the 96 MS displaying a significant correlation between deletion size of the intronic 3' MS and PSI. Right, distribution of the 181 MS for which the PSI value is independent of deletion size. For both panels, the background distribution of all flanking MS ($n = 335,564$) is indicated (grey).

Supplemental Figure S12:

Barplots display the number of mutations in donor site (GT), in acceptor site (AG) and number of exons skipping for each 46 MSI tumors.

Supplementary Table S1:

List of 3'Exon-FMS deregulated between normal tissue and MSI (Table S1A) or MSS (Table S1B) tumors or between MSI and MSS tumors (Table S1C).

Supplementary Table S2:

List of the MS associated with the 985 exons skipping; PSI index is indicated.

Supplementary Table S3:

List of the 96 genes which somatic deletions at flanking 3' non-coding MS are correlated with exon skipping (ES96).

Supplementary Table S4:

List of mutated genes in dMMR crypts from ES96 and coding, as well as known coding MS/genes of MSI CRC (independently of their MSI status, named MFMC); mutational frequency (MF) and MS deletion size (MD) are indicated through tumorigenesis. MF : Mutational Frequency; MD : Median of deletion

Supplementary Table S5:

List of common genes between the specific genes/exons associated with intestinal stem cells and our MSI-exon signature.

Supplementary Table S6:

List of publications which catalogue the exons-signature of 9 RNA splicing factors and the potential overlapping with our MSI-exons signatures (n=985).

RÉSULTATS PRÉLIMINAIRES

Résultats Préliminaires : Analyse de l'instabilité microsatellitaire codante pour l'identification des mutations impliquées dans l'initiation et la progression tumorale des cancers colorectaux MSI.

Dans cette deuxième partie, je présente les résultats préliminaires que j'ai obtenus sur l'analyse de l'instabilité microsatellitaire codante afin de décrire la cinétique mutationnelle de la tumorigenèse MSI colique, de la crypte MMR-déficiente à l'adénocarcinome. Le but de cette analyse est d'identifier des gènes cibles et des mutations potentiellement *drivers* du processus de transformation précoce et/ou tardif afin de mieux comprendre les différentes étapes de la physiopathologie tumorale MSI.

L'ensemble des données présentées ci-dessous doivent être considérées avec précaution étant donné le petit échantillonnage de patients disponible. Ces résultats devront être approfondis à l'aide d'une plus large cohorte de patients afin d'être validés.

Remarque : le matériel et méthode entrepris pour réaliser cette analyse de l'instabilité microsatellitaire codante est le même que celui décrit dans l'article 1. Certains points de méthode bio-informatique seront explicités dans des encadrés.

1. Analyse quantitative de l'instabilité microsatellitaire globale selon l'origine héréditaire *versus* sporadique des lésions

Grâce à une cohorte de 22 patients atteints d'un CRC MSI, nous avons obtenu les données génomiques de n= 23 adénocarcinomes dont n= 16 associés au syndrome de Lynch (68%) et n= 7 associés à une forme sporadique (32%). En parallèle à ces lésions tumorales nous avons analysé les résultats de séquençage de 11 adénomes et 8 cryptes MMR-déficientes (étape précoce) liés au syndrome de Lynch, et 3 adénomes MMR-déficients d'origine sporadique (Figure 1).

Il existe dans la littérature deux modèles pour décrire, selon des critères moléculaires et histologiques, le développement des cancers MSI qu'ils soient héréditaires ou sporadiques (voir chapitre 3). De la sorte, j'ai analysé les données de cette cohorte en tenant compte de leur origine. Étant donné le faible nombre d'adénomes MMR-déficient de formes sporadiques (n= 3), nous n'avons pas réalisé de comparaison entre les stades adénomes LS et SP (Figure 1).

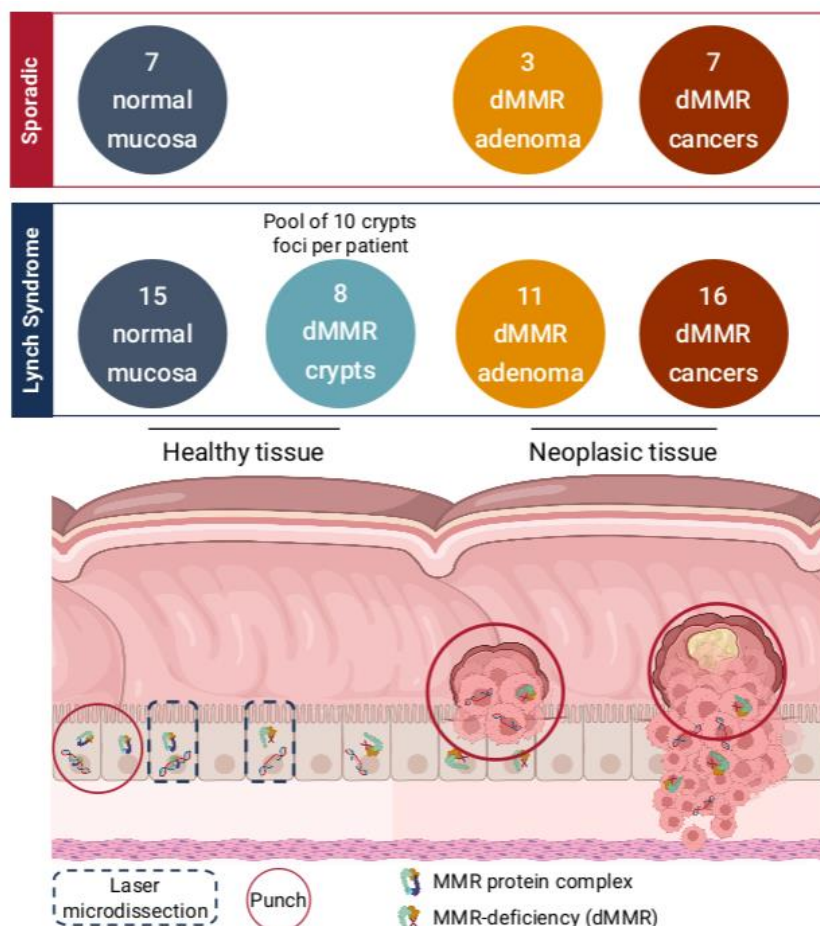


Figure 1 | La cohorte d'étude des évènements de la cinétique mutationnelles de la tumorigenèse MSI.

Les patients sont séparés en deux groupes selon l'origine de la carcinogenèse, sporadique ou associé au syndrome de Lynch. Par la suite, les analyses ont été menées de manière indépendante entre ces deux groupes.

Afin de pouvoir comparer les échantillons les uns par rapport aux autres, nous avons sélectionné les microsatellites pour lesquels la couverture de séquençage est présente dans tous les échantillons (avec un minimum de 10 reads par microsatellites).

Dans un premier temps, nous avons mesuré le niveau d'instabilité microsatellitaire dans les différentes lésions LS et SP, grâce à l'outil MSICare (voir chapitre 2). De la même manière que dans l'article 1, nous observons que l'instabilité microsatellitaire évolue nettement en phase précoce de la tumorigenèse (cryptes MMR-déficiente) avant d'atteindre un plateau (adénome et adénocarcinome) que ce soit pour les grands microsatellites ciblés par MSICare (Figure 2A) ainsi qu'à l'échelle de tous les microsatellites de 5 à 20 répétitions (Figure 2B).

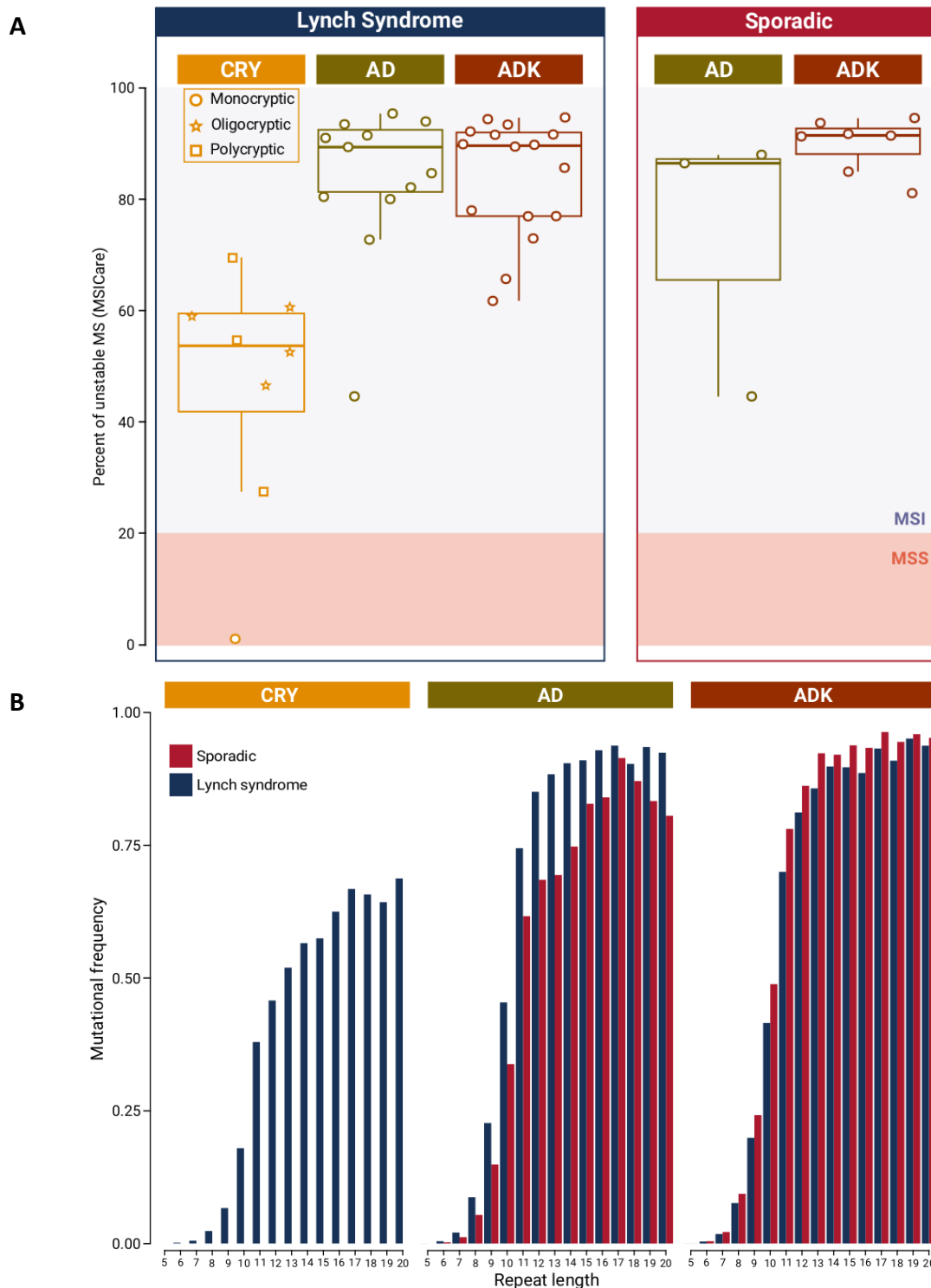


Figure 2 | L'instabilité microsatellitaire globale dans les différentes lésions des cancers MSI sporadiques (en rouge) versus celle associée au syndrome de Lynch (bleu).

(A) L'outil MSICare permet de définir un score qui représente le niveau d'instabilité microsatellitaire à partir du ratio des microsatellites de taille de répétition 12 et plus mutés par rapport à l'ensemble de ces microsatellites observés ($\text{MSICare score (\%)} = \frac{\text{MS mutés}}{\text{MS observés}} \times 100$) (Ratovomanana et al., 2020) ; les échantillons dont le score est supérieur à 20% sont définis comme MSI – en-dessous de cette limite ils sont considérés MSS. **(B)** Représentation de la MF selon la taille des microsatellites (de 5 à 20 répétitions) et le type de lésion. CRY : cryptes MMR-déficientes, AD : adénomes, ADK : adénocarcinome.

En différenciant les microsatellites codants et non-codants, nous observons que la charge mutationnelle microsatellitaire est plus importante dans les régions non-codantes que codantes du génome et qu'elles augmentent au cours de la tumorigenèse (entre la crypte MMR-déficiente et l'adénome) avant d'atteindre un plateau (adénome et adénocarcinome) (Figure 3). Ce résultat est attendu puisque les microsatellites non-codants sont de plus grandes tailles que les codants et donc plus susceptibles d'être mutés (voir chapitre 1).

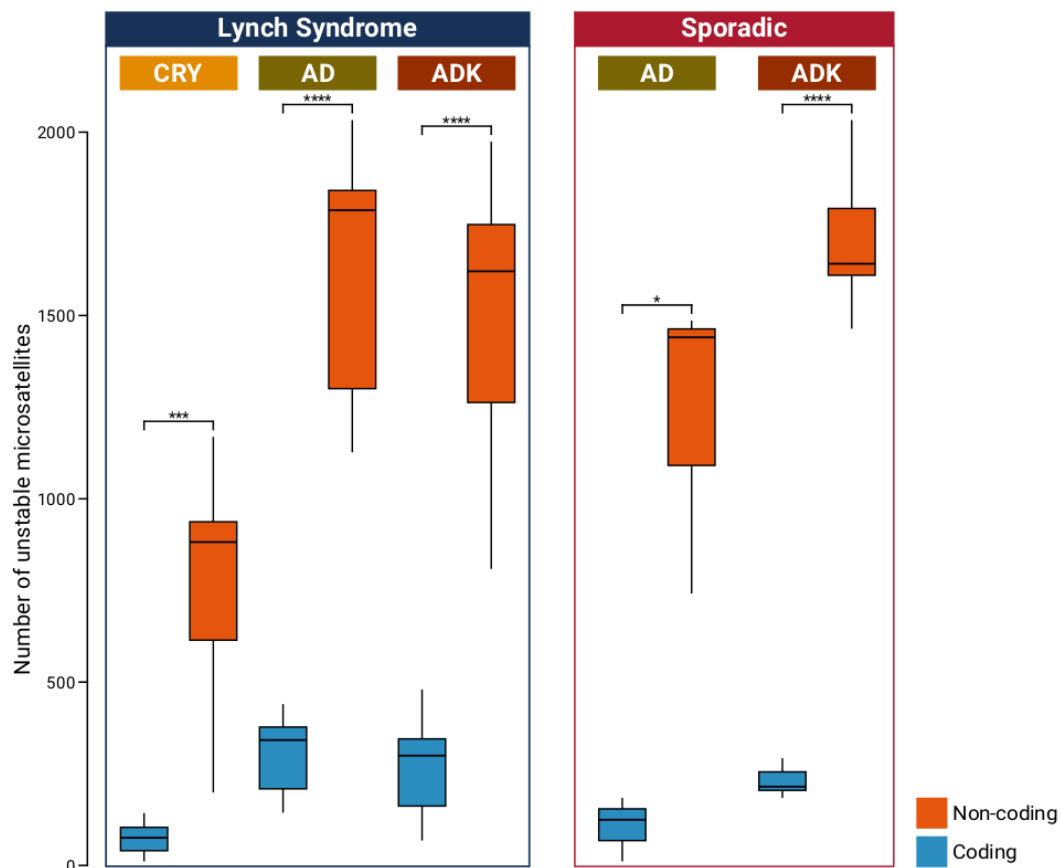


Figure 3 | Distribution du nombre de microsatellites instables codants (bleu) et non-codants (orange) par patient selon le type de lésion.

CRY : cryptes MMR-déficientes, AD : adénomes, ADK : adénocarcinome. * P-value < 0.05, *** P-value < 1.10-4, **** P-value < 1.10-5.

2. Identification des mutations potentiellement *drivers* de la tumorigenèse MSI (syndrome de Lynch)

Pour la suite des analyses, tous les microsatellites codants ont été pris en compte, avec toutefois au moins 50% des échantillons couverts, par lésion.

La fréquence mutationnelle (MF, Mutational Frequency) est définie par le pourcentage d'échantillons qui comportent la mutation par rapport au nombre total d'échantillons.
($MF = \frac{n_{\text{échantillons mutés}}}{N_{\text{échantillons total}}}$)

Dans un premier temps, nous avons comparé les mutations entre les différentes lésions, soit un total de $n = 508$ mutations dont $n = 144$ mutations dans les cryptes MMR-déficientes, $n = 408$ mutations dans les adénomes et $n = 307$ mutations dans les adénocarcinomes (Figure 4A). Parmi ces mutations, $n = 67$ sont partagées entre toutes les lésions (12,2% du total) alors que $n = 52$ (9,5%), $n = 174$ (31,8%) et $n = 78$ (14,2%) sont respectivement spécifiques des cryptes MMR-déficientes, des adénomes et des adénocarcinomes (Figure 4A). Ces résultats laissent supposer que la moitié des mutations qui apparaissent dans les cryptes MMR-déficientes se maintiennent au cours de la tumorigenèse : ces mutations pourraient avoir un rôle pro-tumoral initiateur. Également, $n = 326$ mutations apparaissent au stade adénome dont 45% sont maintenus au stade adénocarcinome : ces mutations pourraient avoir un rôle pro-tumoral dans la progression.

L'analyse d'enrichissement des voies de signalisation montre que les mutations partagées sont associées à la transcription, la réparation et la réplication de l'ADN communément altérées dans les cancers, ainsi qu'à l'organisation des cryptes coliques (voies Wnt/ β -caténine, BMP/TGF- β , Notch) et à l'adhésion (jonction cellule-cellule, organisation de la matrice extracellulaire et migration cellulaire), ce qui est caractéristique des cancers épithéliaux comme le CRC (Figure 4B) (Bienz and Clevers, 2000; Lietha et al., 2007; Sulzmaier et al., 2014).

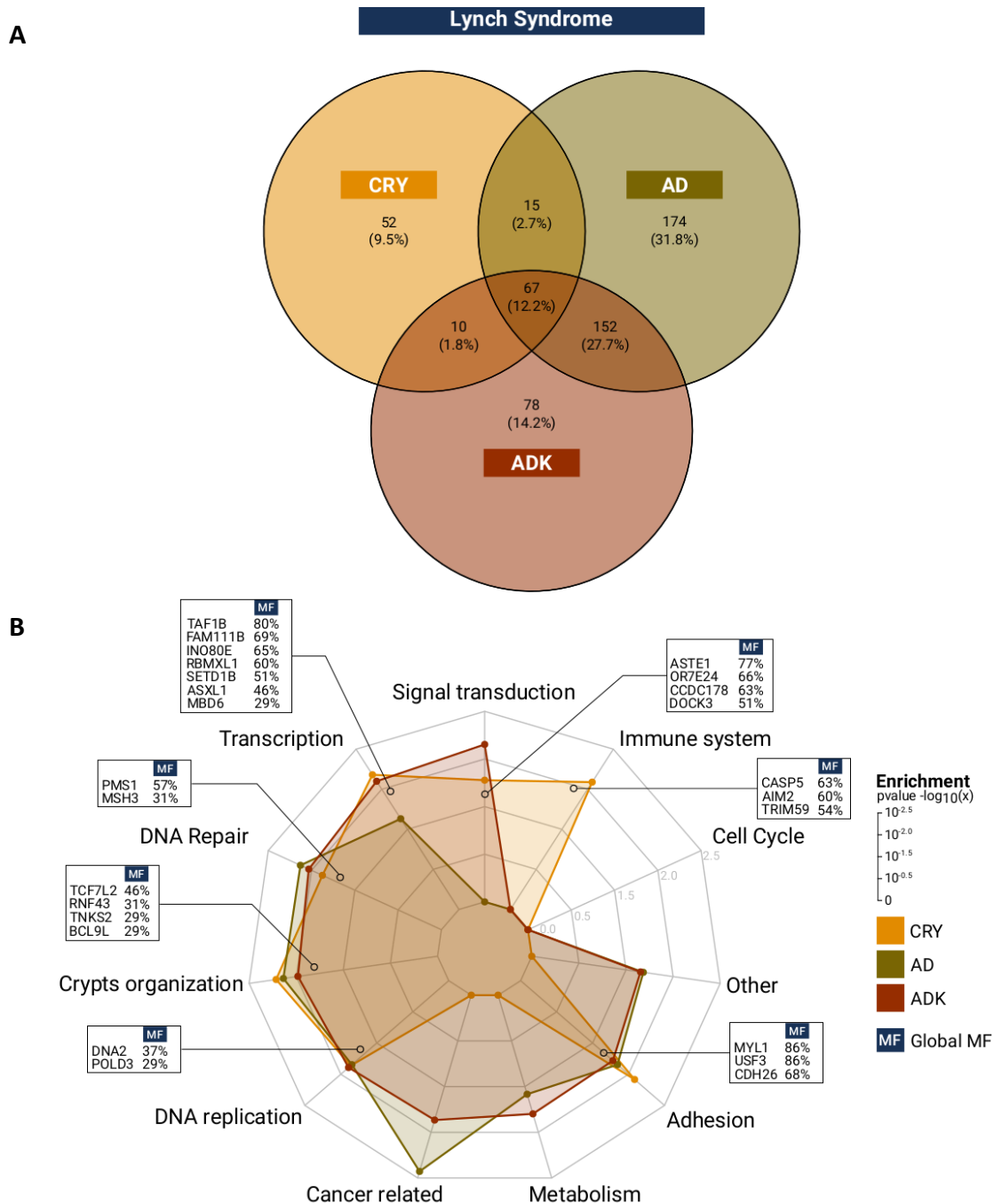


Figure 4 | Représentation des mutations partagées à travers la tumorigenèse MSI du syndrome de Lynch, et des voies de signalisation associées.

(A) Le diagramme de Venn permet de visualiser les mutations partagées ou spécifiques de chaque lésion. Afin de réduire l'hétérogénéité intra-lésionnelle, seules les mutations avec une fréquence mutationnelle supérieure ou égale à 30% ont été prises en compte (limite définit arbitrairement). **(B)** Le diagramme « en toile d'araignée » représente l'enrichissement des voies de signalisation associées aux mutations observées dans chaque lésion, partagées ou spécifiques. L'analyse d'enrichissement a été réalisée avec l'outil EnrichR, filtrée ($pval < 0,05$) puis adaptée pour regrouper les voies de signalisation en groupe selon leur rôle cellulaire. Chaque niveau du diagramme correspond à la valeur statistique ($pvalue$) d'enrichissement, en $-\log_{10}(x)$. Des exemples de mutations sont indiqués avec leur fréquence mutationnelle globale correspondant à la moyenne des fréquences mutationnelles de chaque lésion. CRY : cryptes MMR-déficientes, AD : adénomes, ADK : adénocarcinome.

Parmi les n= 67 mutations partagées entre toutes les lésions, nous avons supposé que certaines soient impliquées dans l'initiation de la tumorigenèse. Pour ce faire, nous avons sélectionné les mutations avec la fréquence mutationnelle la plus élevée. Nous avons ainsi identifié n= 19 mutations classées en deux catégories (Table). Dans un premier groupe, n= 9 mutations précoces dont la MF est supérieure ou égale à 50% dans les cryptes MMR-déficientes ainsi augmentent dans les lésions tardives (MF \geq 75%) ; parmi ces mutations TAF1B, MYL1, USF3 et UPF3A. Dans un second groupe, les mutations tardives dont la MF, inférieure à 50% dans le stade précoce, franchit les 75% dans les stades tardifs (adénomes et adénocarcinomes) ; c'est le cas de certaines mutations déjà connues puisque fréquemment mutées dans les cancers : ACVR2A, ASTE1, TGFBR2, CASP5 (Figure 5A). L'étude des valeurs moyennes des VAF montre que la proportion des mutations dans la population cellulaire globale (poids de la mutation) augmente drastiquement au cours de la tumorigenèse (Figure 5B).

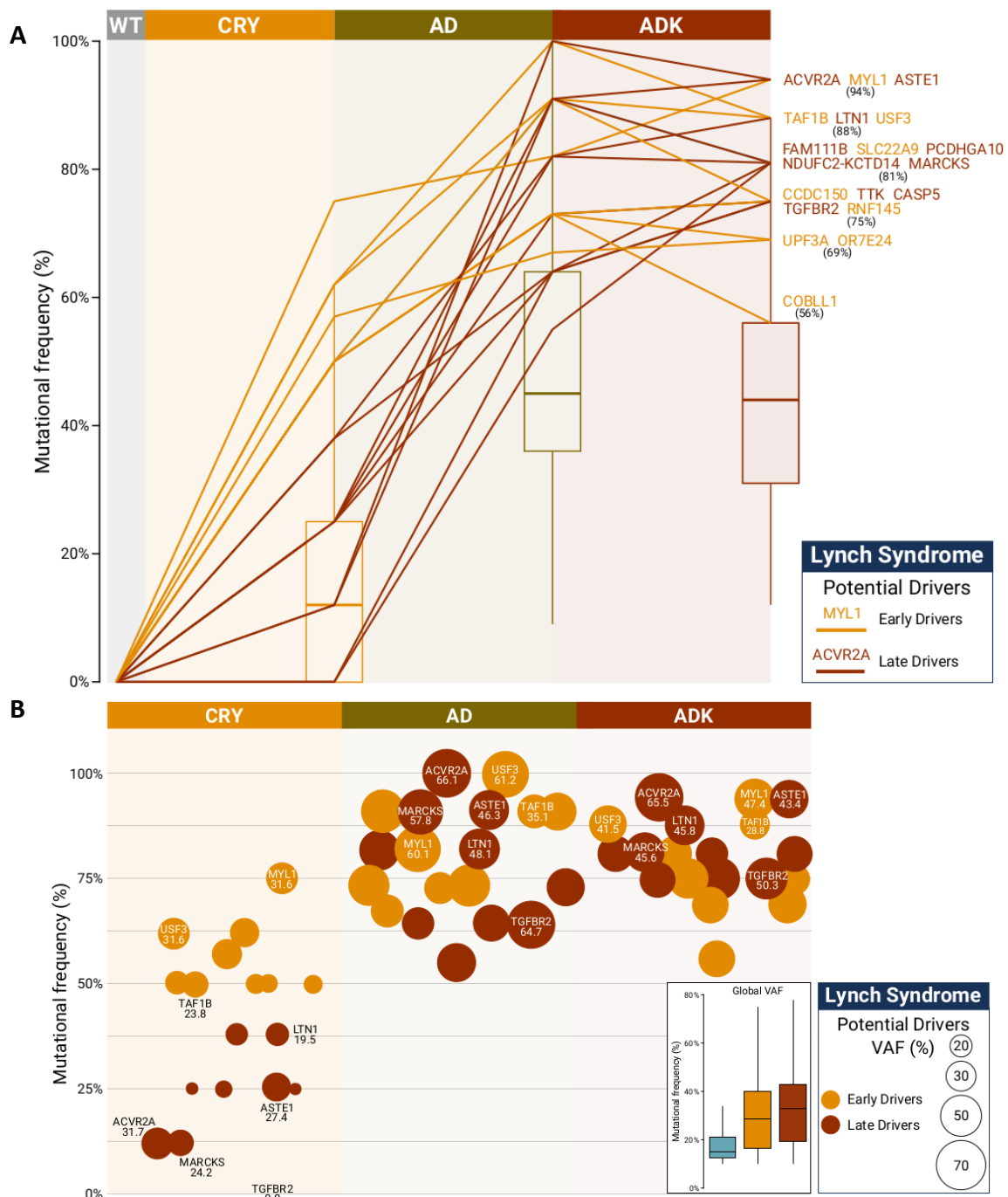


Figure 5 | Identification des mutations potentiellement *drivers* de la tumorigenèse MSI colique associée au syndrome de Lynch.

(A) Représentation de la MF des microsatellites codants les plus instables par type de lésion. Les diagrammes en boîte reflètent la distribution de la MF pour l'ensemble des lésions. Nous avons utilisé deux filtres stricts (arbitraires) pour sélectionner les mutations potentiellement *drivers* : MF \geq 50% pour les cryptes MMR-déficientes (> 75^{ème} percentile) et MF \geq 75% dans les adénomes et les adénocarcinomes (> 75^{ème} percentile). **(B)** Représentation de l'évolution de la MF et de la moyenne des VAF au cours de la tumorigenèse. Chaque cercle désigne une mutation, dont la circonférence définit la valeur de la VAF. La distribution des VAF au global par lésion est présentée dans l'encadré en bas à droite. CRY : cryptes MMR-déficientes, AD : adénomes, ADK : adénocarcinome.

General information				Mutational Frequency (MF)			Variant Allele Frequency (VAF)		
MS position	Gene	Repeat	Drivers	MF _{ADK}	MF _{AD}	MF _{CRY}	VA _{ADK}	VA _{FA} _D	VA _{CR} _Y
chr2.210315041	MYL1	11	early	0.94	0.82	0.75	31.6	39.5	33.7
chr3.113658634	USF3	11	early	0.88	1.00	0.62	31.6	48.3	17.9
chr5.159203621	RNF145	11	early	0.75	0.91	0.62	28.0	34.2	28.4
chr19.9251064	OR7E24	11	early	0.69	0.67	0.57	39.8	33.5	29.1
chr11.63382198	SLC22A9	11	early	0.81	0.91	0.50	16.1	40.2	36.5
chr13.114291735	UPF3A	9	early	0.69	0.73	0.50	43.1	50.9	16.3
chr2.164694785	COBLL1	9	early	0.56	0.73	0.50	40.1	48.6	17.2
chr2.196666794	CCDC150	11	early	0.75	0.73	0.50	20.9	32.2	21.6
chr2.9849441	TAF1B	11	early	0.88	0.91	0.50	23.8	34.8	19.2
chr11.78073100	NDUFC2-KCTD14	9	late	0.81	0.64	0.38	20.5	32.1	15.3
chr21.28966883	LTN1	11	late	0.88	0.82	0.38	19.5	41.8	35.4
chr11.105008959	CASP5	10	late	0.75	0.73	0.25	11.9	30.8	0.00
chr11.59124903	FAM111B	10	late	0.81	0.82	0.25	12.0	28.4	23.7
chr3.131014202	ASTE1	11	late	0.94	0.91	0.25	27.5	46.4	37.7
chr6.80042179	TTK	9	late	0.75	0.64	0.25	15.0	28.5	22.9
chr2.147926116	ACVR2A	8	late	0.94	1.00	0.12	31.7	55.7	38.7
chr6.113860033	MARCKS	11	late	0.81	0.91	0.12	24.2	31.5	10.3
chr3.30650379	TGFBR2	10	late	0.75	0.64	0.00	0.00	38.2	24.9
chr5.141415641	PCDHGA10	11	late	0.81	0.55	0.00	0.00	34.9	48.6

Table | Liste des gènes identifiés comme potentiellement *drivers* de la tumorigenèse MSI associé au syndrome de Lynch.

« MS position » : position génomique du microsatellite, « Gene » : annotation du nom de la séquence codante où se trouve le microsatellite, « Repeat » : taille de répétition du microsatellite, « Drivers » : mutation identifiée comme précoce ou tardive, « Mutational Frequency » : fréquence mutationnelle observée selon le type de lésion, « Variant Allele Frequency » : fréquence du variant observé selon le type de lésion. NDUFC2-KCTD14 : fusion des gènes NDUFC2 et KCTD14.

Le concept de sélection mutationnelle est essentiel pour mieux comprendre l'implication des mutations (pro-/anti-tumorale) dans la tumorigenèse MSI. Afin de définir l'état de sélection des mutations observées dans les différentes lésions, nous avons utilisé les deux modèles de sélection décrits dans les adénocarcinomes coliques MSI (voir chapitre 3, partie 2). Tout d'abord, nous avons reproduit le modèle de sélection par la taille de répétition à partir de nos données, au stade précoce (cryptes MMR-déficientes) et tardif (adénocarcinome seulement) indépendamment, qui s'apparente à celui publié par mon équipe en 2018 (Figure 6A et B). Nos résultats préliminaires montrent qu'au stade précoce, plusieurs mutations codantes sont sélectionnées positivement par leur taille de répétition ; certaines mutations comme celle du gène SETD1B (remodelage de la chromatine), RNF43 et TNK2 (signalisation Wnt/ β -caténine et EGFR/PI3K-Akt) ont déjà été décrites et associées à la tumorigenèse colique MSI. Parmi les mutations potentiellement *drivers* initiatrices de la tumorigenèse, les mutations microsatellitaires des gènes COBLL1 (endocytose) et UPF3A (régulation du système NMD) subissent une pression de sélection positive selon le modèle de la taille de répétition. En revanche, malgré la très haute fréquence mutationnelle des autres mutations (MYL1, USF3, RNF145, OR7E24, TAF1B, SLC22A9, CCDC150), nous ne pouvons pas statuer sur leur état de sélection étant donné les limites que présente ce modèle (voir chapitre 3, partie 2.1) (Figure 6A).

Pour les mêmes raisons, le type de sélection des mutations précoces comme tardives au stade adénocarcinome ne peut être défini, excepté pour n= 4 mutations (ACVR2A, NDUFC2-KCTD14, TTK, UPF3A) (Figure 6B). Également, plusieurs de ces mutations potentiellement *drivers* ont été mises en évidence par le modèle de pression de sélection dépendante des propriétés immunogènes (et pro-tumorales) des mutations qu'elles soient précoces (TAF1B, SLC22A9, CCDC150, RNF145) ou tardives (ACVR2A, TGBR2, ASTE1, LTN1, TTK, FAM11B, CAPS5, NDUFC2, MARCKS) (Ballhausen et al., 2020).

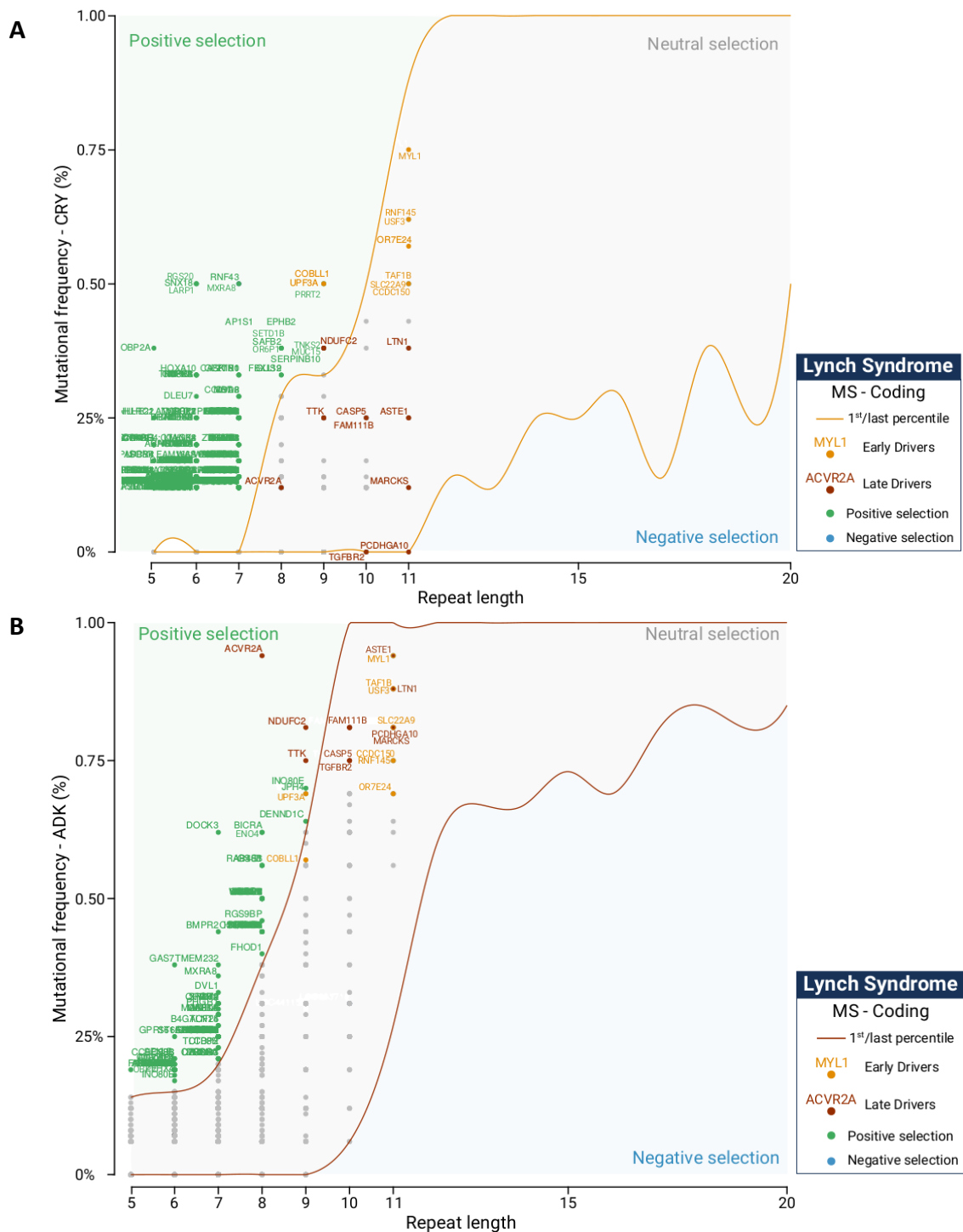


Figure 6 | Évaluation de la pression de sélection subie par les mutations potentiellement *drivers* de la tumorigenèse MSI associée au syndrome de Lynch. Modèle de sélection par la taille de répétition des microsatellites dans les cryptes MMR-déficientes (A) et dans les adénocarcinomes (B).

L'évaluation du type de sélection (positive, neutre, négative) est définie par la distribution des 1^{ers} et derniers percentiles des fréquences mutationnelles observées pour chaque taille de répétition. La zone de sélection neutre définit le bruit de fond mutationnelle. Les mutations potentiellement *drivers* de la tumorigenèse MSI associées au syndrome de Lynch sont indiquées en jaune (précoces) et marron (tardives) CRY : cryptes MMR-déficientes, ADK : adénocarcinome.

L'étude des mutations potentiellement *drivers* de la tumorigenèse MSI associée au syndrome de Lynch, montre plusieurs interactions entre les différents gènes ciblés et un regroupement notable autour des signalisations de l'organisation des cryptes et de la réplication de l'ADN (Figure 7). Le gène ASTE1 (signalisation EGFR/PI3K-Akt), au centre du réseau, montre le plus grand nombre d'interactions avec les autres gènes qu'elles soient tardives ou précoces ; dans le groupe des mutations précoces, USF3 (transition épithélio-mésenchymateuse) et TAF1B (initiation de la transcription) sont les gènes qui affichent le plus d'interactions avec les autres gènes (Figure 7). L'analyse *in silico* des conséquences des mutations sur les protéines met en avant des pertes de fonction évidentes ainsi que l'existence de queues néo-antigéniques (mutations *frameshift*, voir chapitre 1) pouvant être présentée par le CMH de classe I et ainsi stimuler une réponse immunitaire (Figure 8). Les néo-antigènes induits par les mutations microsatellites dans les gènes ASTE1 et TAF1B, ont été utilisés dans le développement d'un vaccin anti-cancéreux qui a montré son efficacité dans un essai clinique de phase I/IIa (Kloor et al., 2020).

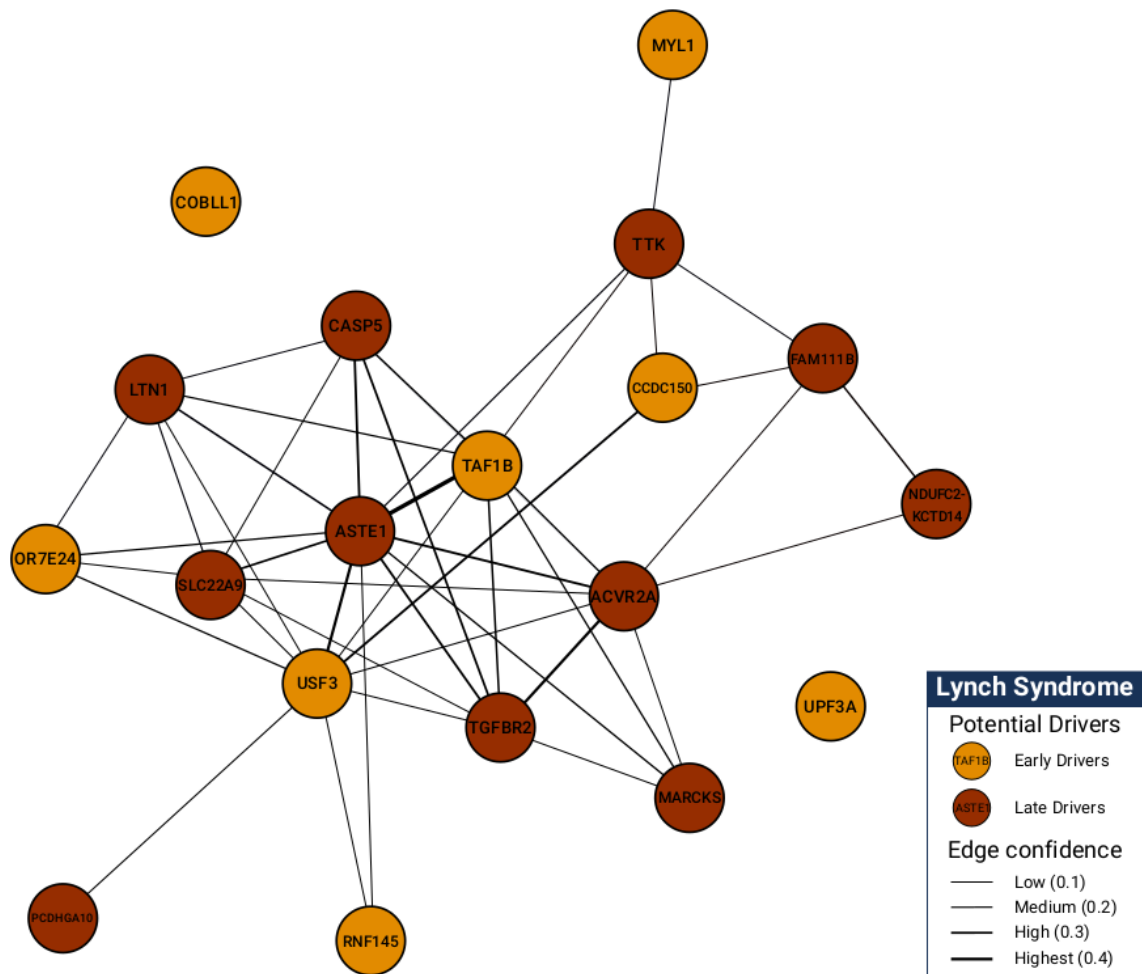


Figure 7 | Représentation en réseau des gènes mutés et potentiellement *drivers* de la tumorigénèse MSI associée au syndrome de Lynch.

Le réseau de gènes est effectué à partir de l'outil « String – Interaction Network ». L'analyse est faite à partir des n= 19 mutations (gènes) potentiellement *drivers* et les interactions montrées sont fonctionnelles et physiques. Les paramètres suivants ont été utilisés pour définir le réseau, liens entre les nœuds : « confiance », score d'interaction : de 0.1 à 0.4, autre : un seul ensemble de gènes, sans clusterisation. Les gènes sont identifiés selon leur mutation, précoce ou tardive.

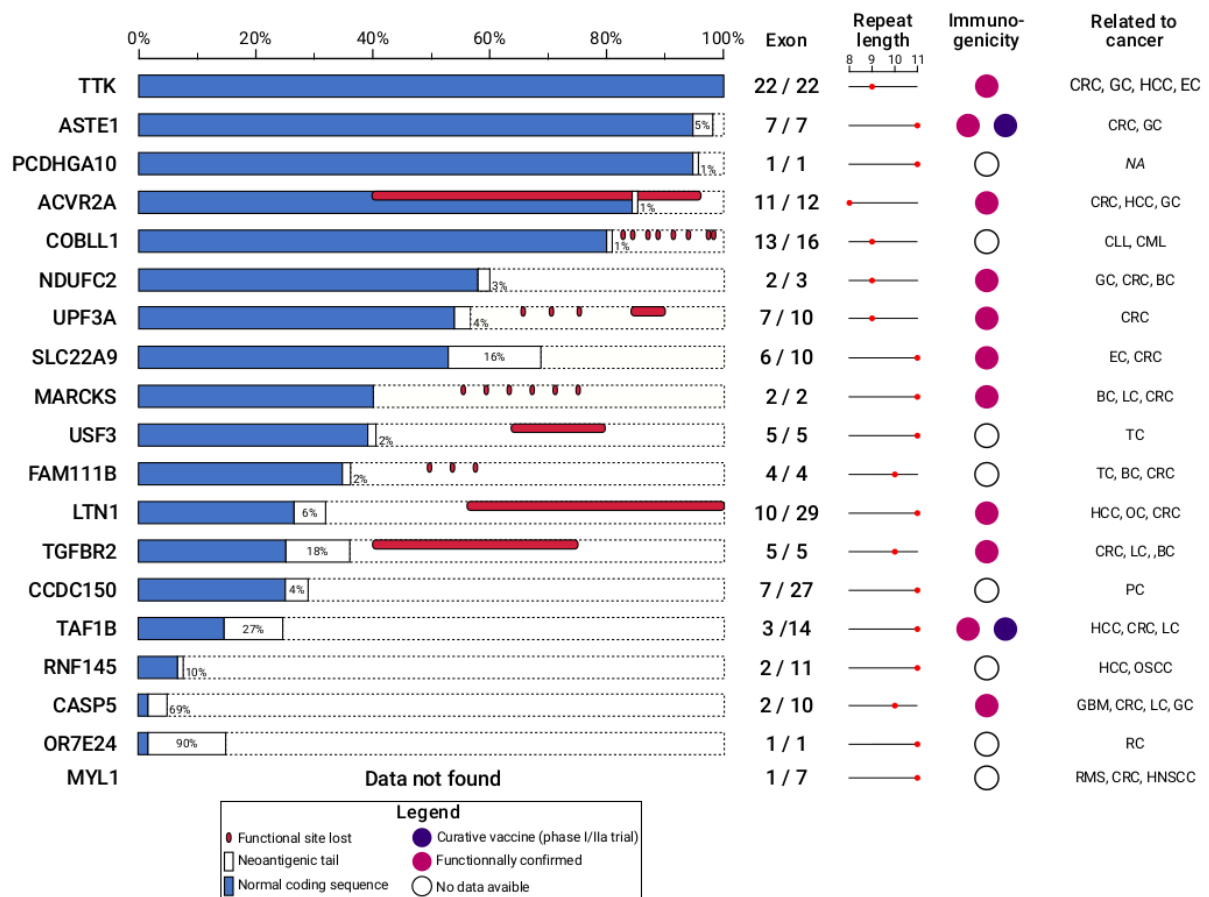


Figure 8 | Détermination *in silico* de la conséquence des mutations sur la fonction des gènes.

Les séquences d'ADN complémentaires ont été obtenues *via* la base de données CCDS afin de générer *in silico* la mutation dans le microsatellite spécifié. Ces séquences mutées ont ensuite été traduites en séquences protéiques (acides aminés) afin de déterminer le codon stop prématuré (PTC) et la séquence néo-peptidique (outils *expasy* et *Pairwise Sequence Alignment*). CRC : cancer colorectal, HCC : carcinome hépatocellulaire, GC : cancer de l'estomac, EC : cancer de l'endomètre, CLL : leucémie lymphoïde chronique, CML : leucémie myéloïde chronique, BC : cancer du sein, TC : cancer de la thyroïde, OC : cancer de la cavité buccale, PC : cancer de la prostate, LC : cancer du poumon, OSCC : carcinome épidermoïde buccal, RC : cancer du rectum, GBM : glioblastome, RMS : rhabdomyosarcome, HNSCC : carcinome épidermoïde de la tête et du cou.

3. Biomarqueurs distinctifs (potentiels) des formes héréditaires et sporadiques, au stade adénocarcinome

Dans cette dernière partie, nous avons comparé les mutations observées dans les adénocarcinomes associées au syndrome de Lynch *versus* sporadique, afin d'identifier les mutations spécifiques et communes entre ces deux formes. L'analyse comparative par diagramme de Venn met en évidence $n = 189$ mutations partagées entre les deux formes alors que $n = 118$ et $n = 269$ sont respectivement spécifiques aux adénocarcinomes du syndrome de Lynch et sporadiques, soit en proportion 40% de toutes les mutations ADK LS *versus* 60% des mutations ADK SP. Cet écart important suggère l'existence de mutations *drivers* différentes impliquées dans la tumorigenèse des deux formes, en accord avec les observations histologiques divergentes décrites dans la littérature par des données moléculaires (Figure 9A). Ces résultats sont supportés par l'analyse d'enrichissement des voies de signalisation qui montre des altérations spécifiques aux cancers sporadiques (organisation et régulation du système immunitaire et du cycle cellulaire) alors que d'autres sont communes aux deux formes (transcription, réplication et réparation de l'ADN, différenciation cellulaire, adhésion) (Figure 9B).

Enfin, par une analyse statistique, nous avons montré que les mutations de gènes potentiellement *drivers* (initiateurs comme progresseurs) de la tumorigenèse MSI associées au syndrome de Lynch, sont aussi très fréquemment mutées dans les cancers sporadiques ($MF_{ADK\ SP} > 90\%$) et pourrait ainsi avoir le même rôle pro-tumoral que ce soit dans un contexte héréditaire ou sporadique (Figure 10). Cette analyse suggère également l'existence de mutations spécifiques de l'une ou l'autre forme MSI au stade adénocarcinomes : $n = 5$ mutations ADK LS contre $n = 33$ mutation ADK SP ; cet écart peut s'expliquer du fait qu'il y ait plus de mutations spécifiques (en nombre) dans les cancers sporadiques (Figure 9A).

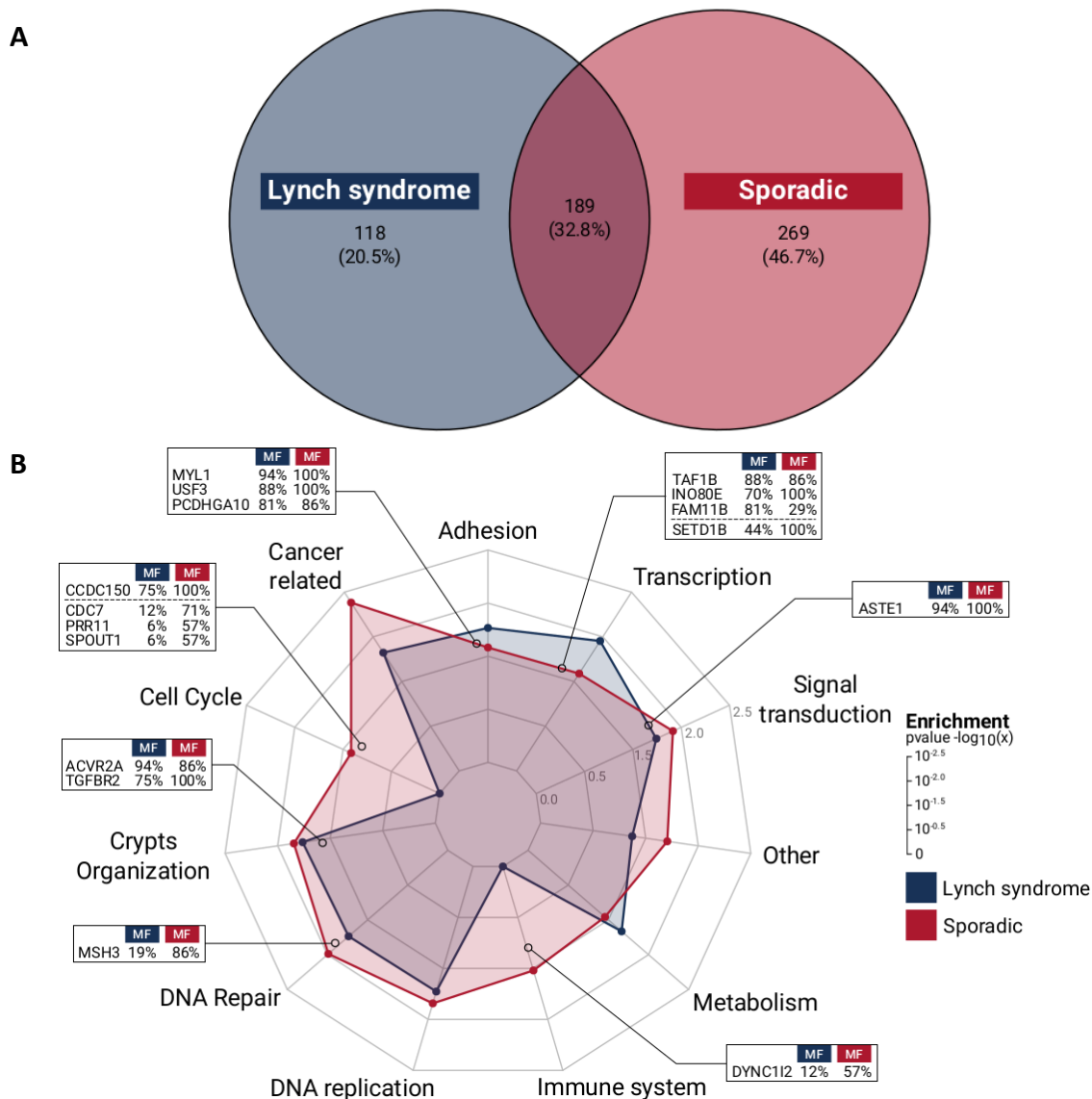


Figure 9 | Représentation des mutations partagées et spécifiques des cancers sporadiques versus du syndrome de Lynch, et des voies de signalisation associées, au stade adénocarcinome.

(A) Le diagramme de Venn permet de visualiser les mutations partagées ou spécifiques de chaque lésion. Afin de réduire l'hétérogénéité intra-tumorale, seules les mutations avec une fréquence mutationnelle supérieure ou égale à 30% ont été conservées (limite définie arbitrairement). (B) Le diagramme « en toile d'araignée » représente l'enrichissement des voies de signalisation associées aux mutations observées selon l'origine héréditaire ou sporadique. L'analyse d'enrichissement a été réalisée avec l'outil EnrichR, filtrée ($pval < 0,05$) puis adaptée pour regrouper les voies de signalisation en groupe selon leur rôle cellulaire. Chaque niveau du diagramme correspond à la valeur statistique ($pvalue$) d'enrichissement, en $-\log_{10}(x)$. La fréquence mutationnelle présentée correspond à celle observée dans les adénocarcinomes associés au syndrome de Lynch (bleu) ou à la forme sporadique (rouge).

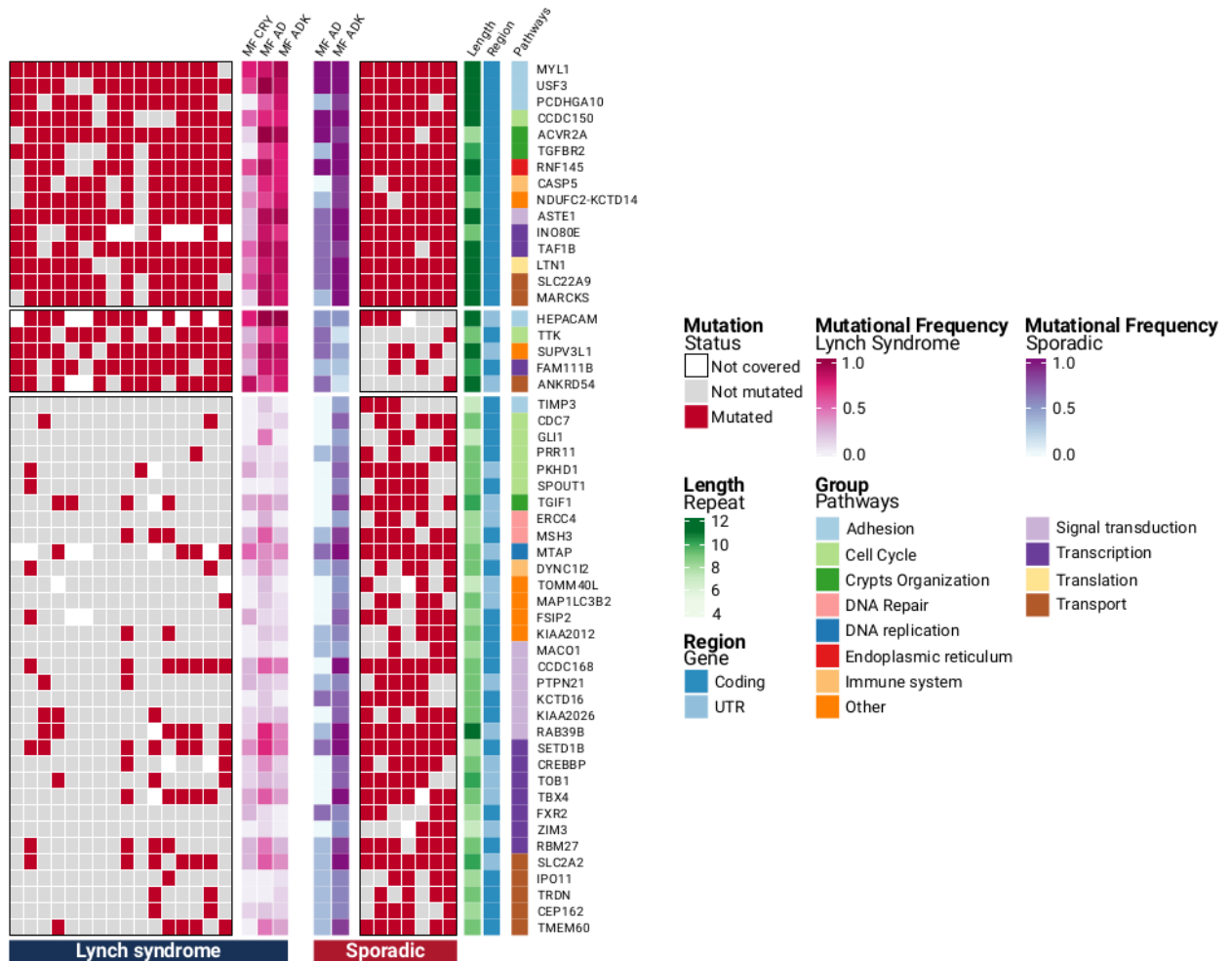


Figure 10 | Détermination des mutations partagées et spécifiques dans les adénocarcinomes sporadiques et associées au syndrome de Lynch.

La mise en évidence des mutations spécifiques aux deux formes a été exécutée par une étude statistique (χ^2) à partir du nombre d'adénocarcinomes LS mutés et non-mutés, par rapport aux adénocarcinomes SP mutés et non-mutés (matrice = $n_{\text{échantillons LS mutés}}$, $n_{\text{échantillons SP mutés}}$, $n_{\text{échantillons LS non-mutés}}$, $n_{\text{échantillons SP non-mutés}}$, 2, 2). Les fréquences mutationnelles à travers la tumorigenèse des deux formes sont indiquées, ainsi que la région du génome où se trouve le microsatellite (codant ou UTR), la taille de répétition et la voie de signalisation dans laquelle le gène muté intervient.

CONCLUSION, DISCUSSION & PERSPECTIVES

Conclusion, discussion et perspectives

Les travaux de recherche de mon équipe d'accueil participent à la description de la tumorigenèse MSI colique, nécessaire pour développer de nouveaux outils diagnostic et thérapeutiques. Mon projet de thèse s'intègre aux objectifs du laboratoire puisqu'il vise à comprendre les mécanismes moléculaires et physiopathologiques impliqués dans l'initiation et dans la progression tumorale MSI colique des formes sporadiques et du syndrome de Lynch.

Au cours de ma thèse, je me suis appliqué à étudier la cinétique mutationnelle microsatellitaire de la tumorigenèse MSI sporadique et du syndrome de Lynch. Dans un contexte tumoral hypermuté, l'identification des mutations *drivers* de la tumorigenèse s'avère particulièrement compliquée. De ce fait, j'ai exploré le paysage mutationnel MSI dans les lésions précoces observées spécifiquement dans le syndrome de Lynch, les cryptes MMR-déficientes, où mes résultats montrent que la charge mutationnelle est moins importante que dans les adénocarcinomes. Mes résultats décrivent ainsi une cinétique mutationnelle et suggèrent un rôle fonctionnel de certaines mutations microsatelitaires introniques et exoniques dans l'initiation et progression tumorale. Ces données devront être validées par la suite sur une plus large cohorte afin d'améliorer nos connaissances sur l'oncogenèse MSI, et ainsi à terme ouvrir de nouvelles pistes pour développer des approches diagnostiques et thérapeutiques.

Mes travaux de thèse s'inscrivent dans une recherche exploratoire des phases précoces de la carcinogenèse colorectale MSI. L'étude de la cinétique mutationnelle MSI que j'ai réalisé pendant mon doctorat, a permis de proposer des mutations potentiellement *drivers* de l'initiation et de la progression tumorale. Mes résultats de thèse peuvent être considérés comme précurseurs dans cette thématique de recherche et dans leur continuité pourront permettre d'améliorer nos connaissances sur la physiopathologie tumorale MSI. Le but ultime que je vois à long terme de mes recherches est de pouvoir participer au développement des outils de médecine personnalisée préventive pour les malades atteints de cancer colique MSI.

Le syndrome de Lynch, un modèle d'étude de la tumorigenèse MSI colique précoce ?

L'exploration du paysage mutationnel de la tumorigenèse MSI précoce grâce à la méthodologie ILAM-NGS

Afin de pouvoir réaliser les analyses génomiques des cryptes MMR-déficientes, nous avons dû développer un protocole de grande précision et spécificité (ILAM-NGS), basé sur la microdissection laser assisté par un marquage IHC (ILAM, *Immunohistochemistry Laser-Assisted Microdissection*) et le séquençage haut-débit (NGS : WES et 3'RNA-seq). À notre connaissance, il s'agit d'une nouvelle méthodologie permettant à la fois de microdisséquer des entités identifiées à l'aide d'un marquage IHC et de séquencer l'ADN et l'ARN extraits de ces dernières.

Initialement conçu pour récupérer spécifiquement les cryptes MMR-déficientes avec une précision au niveau cellulaire, le protocole ILAM-NGS a montré une très bonne efficacité sur les adénomes et adénocarcinomes MSI. Ces résultats suggèrent que ce protocole pourrait être appliqué à d'autres travaux en cancérologie comme l'étude de l'hétérogénéité tumorale mais également pour des recherches physiopathologiques, où la pratique de l'histologie et du séquençage haut-débit est devenue essentielle. Cette approche technique décrite en annexe, fera l'objet d'un article méthodologique prochainement.

La progression « crypte MMR-déficiente – adénome – adénocarcinome » des cancers MSI associés au syndrome de Lynch

Les cryptes MMR-déficientes sont les lésions histologiques les plus précoces jusqu'alors identifiées dans le développement des cancers MSI puisqu'elles sont présentes avant la transformation tumorale (pré-néoplasique). Leur rôle dans l'initiation de la tumorigenèse MSI a été proposé sans toutefois que l'altération de certains mécanismes cellulaires ou les mutations soient décrits. Présentes en foyers dans la muqueuse colique saine des patients LS, les cryptes MMR-déficientes semblent évoluer histologiquement au gré de la dynamique d'apparition des mutations, suite à l'inactivation du système MMR. Les résultats de ma thèse indiquent pour la première fois, que les foyers monocryptiques MMR-déficients sont très peu

instables par rapport aux oligo/polycryptiques et définies de ce fait comme MSS. Cela suggère que ces foyers monocryptiques représenteraient une étape particulièrement précoce avant probablement de se diviser (par fission cryptique) pour devenir des foyers oligo/polycryptiques. En amont de la fission des cryptes, l'expansion clonale des cellules souches propagent la mutation responsable de la déficience du système MMR, par dérive génétique ou sélection naturelle. Toutefois, les mécanismes à l'origine de l'inactivation du système MMR restent inconnus et devront être élucidés à l'avenir.

L'accumulation de mutations survenant dans des gènes impliqués dans les voies de signalisation de l'homéostasie intestinale entraînent la déstructuration cellulaire, cytotogique et nucléaire, ainsi que des altérations architecturales des cryptes. Ces changements s'apparentent à une transformation des cellules des cryptes MMR-déficientes histologiquement saines en adénome dysplasique. L'évolution d'un adénome en adénocarcinome se caractérise par une transformation maligne qui permet aux cellules de se propager de la muqueuse à la sous-muqueuse (jusqu'à la séreuse, en cas d'envahissement localisé du tissu). Moléculairement, cela ne semble pas résulter d'une augmentation de la charge mutationnelle, puisque les adénomes et les adénocarcinomes présentent environ la même quantité de mutations, à des fréquences mutationnelles similaires. L'apparition de mutations qui altèrent la polarité apico-basale (caractéristique des cellules épithéliales), notamment des jonctions inter-cellulaires, des molécules d'adhésion à la matrice extracellulaire, pourraient déclencher la transition épithélio-mésenchymateuse (EMT, *Epithelial–Mesenchymal Transition*) des cellules et ainsi leur propagation dans le tissu. En effet, ces voies de signalisation sont massivement mutées dans les adénocarcinomes, ce qui pourrait s'expliquer par une sélection des clones cellulaires porteurs de ces mutations. Il serait intéressant d'approfondir nos analyses dans ce sens, notamment par l'exploitation des données de VAF, afin d'identifier les mutations sous-clonales sélectionnées qui permettent la transformation maligne.

Comme décrit dans la littérature, la tumorigenèse MSI sporadique et celle associée au syndrome de Lynch sont différentes. Tout d'abord parce que les cryptes MMR-déficientes sont spécifiquement observées chez les patients atteints par le LS et ce malgré des recherches approfondies dans la muqueuse colique des patients atteints de cancers MSI sporadiques. Les

cancers MSI sporadiques se développeraient, dès l'inactivation du système MMR, à un stade adénome ou bien il pourrait s'agir d'une transformation plus lente à partir d'un adénome MMR-proficient où une hyperméthylation du promoteur MLH1 entraînerait un phénotype MSI. Ce point est particulièrement intéressant et devra être approfondi pour permettre une meilleure compréhension de l'initiation de la tumorigenèse sporadique. Une autre différence réside dans l'histologie des tumeurs MSI d'origine distincte, en effet les adénomes LS sont majoritairement de type conventionnel alors qu'ils sont de type festonné sessile dans les formes sporadiques. Une hypothèse est que cette différence est due aux mutations du gène *APC* et *CNNTB1* (qui code pour la β -caténine) dans les adénomes LS, ce qui s'apparente à la transformation bénigne observée dans les cancers MSS. Pour autant, le paysage mutationnel des adénocarcinomes LS ressemble plutôt (sans être identique) à celui des adénocarcinomes sporadiques, ce qui suggère que l'instabilité microsatellitaire impacte drastiquement la transformation maligne du tissu.

Le manque de connaissance quant à l'initiation de la tumorigenèse des cancers MSI sporadiques nous empêche de pouvoir comparer les mutations et mécanismes précoces des deux formes de cancers MSI. Toutefois, le développement tumoral LS et sporadique présente les mêmes caractéristiques phénotypiques, que ce soit de l'infiltration lymphocytaire ou de la dédifférenciation/indifférenciation des cellules cancéreuses. Des analyses comparatives des deux formes de cancers MSI, au stade adénome et adénocarcinome, pourront permettre de mieux décrire la cinétique mutationnelle de la tumorigenèse MSI au global et peut être d'identifier des signatures mutationnelles spécifiques.

La cinétique mutationnelle issue de l'instabilité microsatellitaire codante et intronique au cours de la tumorigenèse et ses implications fonctionnelles

De nombreux microsatellites introniques se situent dans les séquences régulatrices de l'épissage des ARN. Des travaux précédemment publiés ont montré une corrélation entre l'instabilité microsatellitaire de ces séquences et l'épissage alternatif des exons en aval. C'est le cas notamment de l'exon 9 du gène *HSP110*, comme l'a démontré mon équipe, ou bien du gène *MRE11* (Dorard et al., 2011; Giannini et al., 2004). Dès lors, mon laboratoire a souhaité

investiguer l'implication de l'instabilité microsatellitaire dans la dérégulation de l'épissage des ARN, durant la tumorigenèse MSI.

Les résultats que j'ai obtenus pendant ma thèse montrent l'existence d'une signature mutationnelle de l'altération de l'épissage des ARN spécifique des cancers MSI caractérisé par des hautes fréquences mutationnelles dès les phases précoces de la tumorigenèse. Aussi, la signature mutationnelle MSI-dépendante se caractérise par un nombre important de sauts d'exons lors de l'épissage des ARN, qui suggère une dérégulation importante de ce mécanisme. De manière intéressante, cette dérégulation et l'épissage aberrant qui suit, touche des exons connus comme alternatifs, notamment impliqués dans la différenciation ainsi que la présentation de peptides néo-antigéniques immunogènes. Cette signature est aussi associée à la signature oncogénique U2AF-S34F connue dans de nombreux cancers.

Ces résultats très prometteurs pourront être approfondis notamment sur les aspects précoces. D'une part, nous avons récemment mis au point et réalisé un nouveau protocole de séquençage des ARN matures entiers (*RNA-seq full length*) à partir des cryptes MMR-déficientes et des adénomes, dont la qualité des ARN que nous avons obtenue pendant ma thèse, ne nous permettait pas d'obtenir ces données en premier lieu. À l'instar des adénocarcinomes, l'analyse de ces nouvelles données de séquençage des ARN matures entiers nous donnerait la possibilité d'identifier les événements d'épissage alternatif dans les cryptes MMR-déficientes et les adénomes. D'autre part, seulement 277 microsatellites (suffisamment bien couverts) ont été analysés pour réaliser la corrélation statistique entre l'instabilité microsatellitaire intronique et la dérégulation de l'épissage des exons associés (34,7% ; 96/277). Il serait ainsi pertinent de confirmer ces résultats sur un plus grand nombre de microsatellites couverts. Ensemble, ces résultats permettront d'appréhender la cinétique des altérations transcriptionnelles dans la tumorigenèse MSI colique du LS.

Les altérations de l'épissage des ARN observées spécifiquement dans les cancers MSI, ne semblent pas être entièrement dépendantes de l'instabilité des séquences microsatellitaires (régulation en *cis*). Des travaux ont récemment montré que la machinerie d'épissage des ARN et le système MMR pouvait interagir directement ou indirectement *via* la machinerie de transcription (Huang and Li, 2020). De ce fait, le système MMR pourrait avoir une incidence sur le fonctionnement physiologique de l'épissage, et ainsi en cas de déficience du MMR, cela pourrait entraîner des altérations lors de la transcription et/ou de l'épissage

(régulation en *trans*). Une autre hypothèse serait que le saut d'exon observé dans certains gènes soit conséquent à d'autres mutations en *cis*, notamment dans les séquences *enhancers* et *silencers* de l'épissage exoniques (ESE, ESI) comme introniques (SSE, SSI) qui régulent l'inclusion ou l'exclusion des exons. Enfin, de nombreux exons associés à des facteurs de la régulation endogène de l'épissage sont dérégulés et pourraient ainsi entraîner une amplification générale de la dérégulation de l'épissage, comme c'est le cas dans d'autres cancers (mélanomes uvéaux, cancers du sein ou du pancréas).

L'étude de l'instabilité microsatellitaire codante à partir des données de séquençage de l'exome des cryptes MMR-déficientes, des adénomes et des adénocarcinomes, que j'ai développé pendant ma thèse, a permis d'initier une analyse sur les mutations à haute fréquence mutationnelle dès l'initiation et au cours de la progression tumorale. Ces résultats restent préliminaires et devront être validés. Néanmoins, ces données ont permis d'évaluer certains paramètres que je discuterai ici. La majorité de ces mutations surviennent dans des gènes impliqués dans les processus cellulaires de la prolifération, la différenciation et de la survie (résistance à la mort cellulaire).

Ces résultats suggèrent que ces mutations jouent un rôle critique (*driver*) dans le développement tumoral. Nous proposons de répartir les mutations en deux classes : les mutations *drivers* précoces qui sont présentes dès l'initiation de la tumorigenèse, dans les cryptes MMR-déficientes (9/19) ; les mutations *drivers* tardives apparaissent durant la progression tumorale (10/19), au stade adénome et/ou adénocarcinome. Les mutations des gènes *TGFBR2*, *ACVR2A* ou *CASP5* – jusqu'alors décrites comme les mutations *drivers* principales du développement tumoral – n'apparaissent qu'à partir du stade adénome. Alors que les mutations des gènes *TAF1B* et *UPF3A* sont particulièrement intéressantes puisqu'elles apparaissent dans le stade précoce et se maintiennent (à haute fréquence mutationnelle) dans les stades tardifs. La présence de la mutation du gène *UPF3A* laisse supposer que le système NMD soit impliqué dans l'initiation tumorale afin de ralentir ou d'empêcher la sélection de mutation anti-tumorale.

Également, l'analyse d'enrichissement des mutations des microsatellites codants nous apporte des précisions sur les voies de signalisation impliquées dans la tumorigenèse. Dans les

cryptes MMR-déficientes, les voies associées au système immunitaire (c'est-à-dire l'inflammation et la présentation des peptides antigéniques) semblent très altérées ce qui pourrait s'apparenter aux prémices d'un mécanisme d'échappement immunitaire. En effet, des travaux récents ont montré une infiltration immunitaire précoce dans la muqueuse colique saine des patients atteints par le syndrome de Lynch (Bohaumilitzky et al., 2022). Nos données suggèrent que le déclenchement de la surveillance immunitaire fait suite à l'accumulation de nombreuses mutations *frameshift* dans les cryptes MMR-déficientes, dont certaines immunogènes, seraient présentées par le MHC de classe I des cellules. Les mutations dans le processus de présentation des peptides antigéniques pourraient ainsi permettre aux cryptes MMR-déficientes de contourner la surveillance immunitaire et donc de continuer à se développer.

L'ensemble de ces résultats devront être confirmés sur une plus large cohorte de patients afin d'avoir une puissance statistique suffisamment fiable. Également, il serait intéressant de valider l'étude de la cinétique mutationnelle avec une approche *single cell* et/ou en séquençage régionalisé afin de mieux décrire la dynamique sous-clonale au sein des adénomes et adénocarcinomes. En particulier, ces techniques pourraient nous permettre d'identifier les clones adénomateux (et les mutations) responsables de la progression en adénocarcinome.

Vers le développement de thérapie personnalisée préventive pour les patients atteints de cancers MSI du spectre héréditaire

En 2020, un vaccin curatif composé de trois FSP néo-antigéniques issus des mutations des gènes *TAF1B*, *AIM2* et *ASTE1*, a été validé dans un essai clinique de phase I/IIa. Nous avons suggéré que les deux mutations des gènes *TAF1B* et *ASTE1* seraient *drivers* de la tumorigenèse MSI colique du LS, dont *TAF1B* serait impliqué dans l'initiation tumorale. Également, mes résultats ont montré que parmi les 19 mutations potentiellement *drivers* de la tumorigenèse, 11 mutations possèdent des propriétés immunogènes (testé *in vitro*) dont 3 mutations (*TAF1B*, *SLC22A9*, *UPF3A*) sont déjà présentes à haute fréquence mutationnelle au stade précoce et 6 autres mutations immunogènes sont présentes dans les cryptes MMR-déficientes, avec des fréquences mutationnelles de 25-40% (*TTK*, *FAM111B*, *CASP5*, *ASTE1*, *NDUFC2*, *LTN1*). Ces 9

mutations pourraient avoir un intérêt thérapeutique car, après validation dans une cohorte plus importante, elles pourraient être utilisées pour élaborer des approches vaccinales préventives contre les cancers colorectaux MSI associés au syndrome de Lynch.

Pour conclure, l'instabilité microsatellitaire intronique et codante est impliquée dans l'initiation et dans la progression tumorale MSI colique, de manière complémentaire. La charge mutationnelle intronique plus importante, en nombre et en fréquence, altère l'épissage des ARN notamment des oncogènes (dont certains sont retrouvés dans la signature oncogénique U2AF-S34F décrite dans de nombreux cancers), ce qui a pour conséquence des gains, des pertes et des perversions de fonction. Cette signature mutationnelle pourrait aussi expliquer en partie le phénotype indifférencié des tumeurs MSI. Une hypothèse très intéressante sur la dynamique mutationnelle est que les mutations codantes sélectionnées au cours de la tumorigenèse surviennent plutôt dans des gènes suppresseurs de tumeur dont la fonction est perdue et qui entraîne (souvent) la synthèse d'un FSP néo-antigénique qui peut participer à stimuler une réponse immunitaire anti-tumorale. L'exposition des cellules tumorales à la surveillance immunitaire favorise la sélection de mécanismes d'échappement, tels que la mutation du gène *B2M* ou la surexpression des inhibiteurs de point (PD-L1/PD-1, CTLA-4). Ces biomarqueurs sont des cibles d'intérêts pour la recherche clinique et le développement de nouvelles approches diagnostics, pronostics et thérapeutiques.

ANNEXE

MÉTHODOLOGIE

“Immunoguided Laser-Assisted Microdissection for Whole Exome and RNA-Sequencing: A Precise Technology to Explore Genomics and Transcriptomics Alteration in Early Stage of Cancer”

Montémont H. et al., en cours de rédaction

Résumé de la méthodologie ILAM-NGS, *Immunohistochemistry Laser-Assisted Microdissection Next-Generation Sequencing*

La découverte des lésions pré-cancéreuses (les cryptes MMR-déficientes) observées dans la muqueuse colique saine des patients atteints par le syndrome de Lynch offre de nouvelles opportunités pour améliorer nos connaissances sur la physiopathologie tumorale MSI, pour élucider les mécanismes d'initiation tumorale, qui à terme seront essentiels pour développer de nouveaux outils diagnostique et thérapeutiques.

Durant ma thèse, j'ai mis au point un protocole technique basé sur la microdissection laser assisté par un marquage immunohistochimique (ILAM, *Immunohistochemistry Laser-Assisted Microdissection*) afin de réaliser le séquençage à haut-débit des ADN et des ARN (NGS, *Next-Generation Sequencing*) des cryptes MMR-déficientes. Cette technologie consiste à (i) marquer immunohistochimiquement des lames histologiques d'intérêts (issu de conservation en FFPE) avec des anticorps au choix (anti-protéines du MMR, ici) ; (ii) microdisséquer les entités d'intérêts (dans notre cas les cryptes MMR-déficientes) sur les lames marquées, à l'aide d'un laser (précision au niveau cellulaire, x40) ; (iii) extraire les acides nucléiques (en parallèle) à partir des échantillons microdisséqués, dont la quantité en matériel est très faible (de 0,5 à 5 mm² de surface tissulaire) ; (iv) séquencer les ADN (exome) et les ARN (3'RNA ou ARNm *full length*) à partir d'acides nucléiques de très faible qualité (conservés FFPPE et ayant subi des traitements chimiques) et quantité (inférieure à 5 ng) ; (v) analyser les mutations des séquences répétées et non-répétées exoniques et introniques (+/- 50 pb flanquant les exons), l'expression différentielle des gènes, les variants d'épissage et la prédiction de présentation de peptides néo-antigéniques en bio-informatique.

Initialement conçu pour récupérer spécifiquement les cryptes MMR-déficientes avec une haute précision, spécificité et sensibilité, le protocole ILAM-NGS a montré une très bonne efficacité sur les adénomes et adénocarcinomes MSI. Ces résultats suggèrent que ce protocole pourrait être appliqué à d'autres travaux en cancérologie comme l'étude de l'hétérogénéité tumorale mais également pour des recherches physiopathologiques, où la pratique de l'histologie et du séquençage haut-débit est devenue essentielle.

Ce travail, en cours de rédaction, fera l'objet d'un article de méthodologie que je signerai en premier auteur.

BIBLIOGRAPHIE

Bibliographie

- Aaltonen, L.A., Peltomäki, P., Leach, F.S., Sistonen, P., Pylkkänen, L., Mecklin, J.P., Järvinen, H., Powell, S.M., Jen, J., Hamilton, S.R., Petersen, G.M., Kinzler, K.W., Vogelstein, B., De La Chapelle, A., 1993. Clues to the pathogenesis of familial colorectal cancer. *Science* (80-). 260, 812–816.
- Abramowicz, A., Gos, M., 2019. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *JAMA Oncol.* 5, 1631–1633.
- Ahadova, A., Gallon, R., Gebert, J., Ballhausen, A., Endris, V., Kirchner, M., Stenzinger, A., Burn, J., von Knebel Doeberitz, M., Bläker, H., Kloor, M., 2018. Three molecular pathways model colorectal carcinogenesis in Lynch syndrome. *Int. J. Cancer* 143, 139–150.
- Akinyi, M. V., Frilander, M.J., 2021. At the Intersection of Major and Minor Spliceosomes: Crosstalk Mechanisms and Their Impact on Gene Expression. *Front. Genet.* 12, 1–10.
- Alam, M.R., Abdul-Ghafar, J., Yim, K., Thakur, N., Lee, S.H., Jang, H.J., Jung, C.K., Chong, Y., 2022. Recent Applications of Artificial Intelligence from Histopathologic Image-Based Prediction of Microsatellite Instability in Solid Cancers: A Systematic Review. *Cancers (Basel)*. 14, 1–18.
- Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., Islam, S.M.A., Lopez-Bigas, N., Klimczak, L.J., McPherson, J.R., Morganella, S., Sabarinathan, R., Wheeler, D.A., Mustonen, V., Boutros, P., Chan, K., Fujimoto, A., Getz, G., Huang, M.N., Kazanov, M., Lawrence, M., Martincorena, I., Morganella, S., Nakagawa, H., Polak, P., Prokopec, S., Roberts, S.A., Rozen, S.G., Saini, N., Shibata, T., Shiraishi, Y., Stratton, M.R., Teh, B.T., Vázquez-García, I., Yousif, F., Yu, W., 2020. The repertoire of mutational signatures in human cancer. *Nature* 578, 94–101.
- Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A. V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., Boyault, S., Burkhardt, B., Butler, A.P., Caldas, C., Davies, H.R., Desmedt, C., Eils, R., Eyfjörd, J.E., Foekens, J.A., Greaves, M., Hosoda, F., Hutter, B., Illicic, T., Imbeaud, S., Imielinsk, M., Jäger, N., Jones, D.T.W., Jonas, D., Knappskog, S., Koo, M., Lakhani, S.R., López-Otín, C., Martin, S., Munshi, N.C., Nakamura, H., Northcott, P.A., Pajic, M., Papaemmanuil, E., Paradiso, A., Pearson, J. V., Puente, X.S., Raine, K., Ramakrishna, M., Richardson, A.L., Richter, J., Rosenstiel, P., Schlesner, M., Schumacher, T.N., Span, P.N., Teague, J.W., Totoki, Y., Tutt, A.N.J., Valdés-Mas, R., Van Buuren, M.M., Van 't Veer, L., Vincent-Salomon, A., Waddell, N., Yates, L.R., Zucman-Rossi, J., Andrew Futreal, P., McDermott, U., Lichter, P., Meyerson, M., Grimmond, S.M., Siebert, R., Campo, E., Shibata, T., Pfister, S.M., Campbell, P.J., Stratton, M.R., 2013. Signatures of mutational processes in human cancer. *Nature* 500, 415–421.
- Alfred G. Knudson, J., 1971. Mutation and Cancer: Statistical Study of Retinoblastoma. *Proc. Natl. Acad. Sci.* 68, 820–823.

- André, T., De Gramont, Armand, Vernerey, D., Chibaudel, B., Bonnetain, F., Tijeras-Raballand, A., Scriver, A., Hickish, T., Tabernero, J., Van Laethem, J.L., Banzi, M., Maartense, E., Shmueli, E., Carlsson, G.U., Scheithauer, W., Papamichael, D., Möehler, M., Landolfi, S., Demetter, P., Colote, S., Tournigand, C., Louvet, C., Duval, A., Fléjou, J.F., De Gramont, Aimery, 2015. Adjuvant fluorouracil, leucovorin, and oxaliplatin in stage II to III colon cancer: Updated 10-year survival and outcomes according to BRAF mutation and mismatch repair status of the MOSAIC study. *J. Clin. Oncol.* 33, 4176–4187.
- André, T., Lonardi, S., Wong, K.Y.M., Lenz, H.J., Gelsomino, F., Aglietta, M., Morse, M.A., Van Cutsem, E., McDermott, R., Hill, A., Sawyer, M.B., Hendlisz, A., Neyns, B., Abdullaev, S., Memaj, A., Lei, M., Dixon, M., Kopetz, S., Overman, M.J., 2022. Nivolumab plus low-dose ipilimumab in previously treated patients with microsatellite instability-high/mismatch repair-deficient metastatic colorectal cancer: 4-year follow-up from CheckMate 142. *Ann. Oncol.* 33, 1052–1060.
- André, T., Shiu, K.-K., Kim, T.W., Jensen, B.V., Jensen, L.H., Punt, C., Smith, D., Garcia-Carbonero, R., Benavides, M., Gibbs, P., de la Fouchardiere, C., Rivera, F., Elez, E., Bendell, J., Le, D.T., Yoshino, T., Van Cutsem, E., Yang, P., Farooqui, M.Z.H., Marinello, P., Diaz, L.A., 2020. Pembrolizumab in Microsatellite-Instability–High Advanced Colorectal Cancer. *N. Engl. J. Med.* 383, 2207–2218.
- Arnold, C.N., Goel, A., Boland, C.R., 2003. Role of hMLH1 promoter hypermethylation in drug resistance to 5-fluorouracil in colorectal cancer cell lines. *Int. J. Cancer* 106, 66–73.
- Axelrod, M.L., Cook, R.S., Johnson, D.B., Balko, J.M., 2019. Biological consequences of MHC-II expression by tumor cells in cancer. *Clin. Cancer Res.* 25, 2392–2402.
- Ballhausen, A., Przybilla, M.J., Jendrusch, M., Haupt, S., Pfaffendorf, E., Seidler, F., Witt, J., Hernandez Sanchez, A., Urban, K., Draxlbauer, M., Krausert, S., Ahadova, A., Kalteis, M.S., Pfuderer, P.L., Heid, D., Stichel, D., Gebert, J., Bonsack, M., Schott, S., Bläker, H., Seppälä, T., Mecklin, J.P., Ten Broeke, S., Nielsen, M., Heuveline, V., Krzykalla, J., Benner, A., Riemer, A.B., von Knebel Doeberitz, M., Kloor, M., 2020. The shared frameshift mutation landscape of microsatellite-unstable cancers suggests immunoediting during tumor evolution. *Nat. Commun.* 11, 1–13.
- Balmaña, J., Balaguer, F., Cervantes, A., Arnold, D., 2013. Familial risk-colorectal cancer: ESMO clinical practice guidelines. *Ann. Oncol.* 24, vi73–vi80.
- Baralle, F.E., Giudice, J., 2017. Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.* 18, 437–451.
- Baretti, M., Le, D.T., 2018. DNA mismatch repair in cancer. *Pharmacol. Ther.* 189, 45–62.
- Barker, N., 2014. Adult intestinal stem cells: Critical drivers of epithelial homeostasis and regeneration. *Nat. Rev. Mol. Cell Biol.* 15, 19–33.
- Barker, N., Ridgway, R.A., Van Es, J.H., Van De Wetering, M., Begthel, H., Van Den Born, M., Danenberg, E., Clarke, A.R., Sansom, O.J., Clevers, H., 2009. Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature.*
- Barker, N., Van Es, J.H., Kuipers, J., Kujala, P., Van Den Born, M., Cozijnsen, M., Haegebarth,

- A., Korving, J., Begthel, H., Peters, P.J., Clevers, H., 2007. Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* 449, 1003–1007.
- Barrett, L.W., Fletcher, S., Wilton, S.D., 2012. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell. Mol. Life Sci.* 69, 3613–3634.
- Batlle, E., Massagué, J., 2019. Transforming Growth Factor- β Signaling in Immunity and Cancer. *Immunity* 50, 924–940.
- Berardinelli, G.N., Scapulatempo-Neto, C., Durães, R., de Oliveira, M.A., Guimarães, D., Reis, R.M., 2018. Advantage of HSP110 (T17) marker inclusion for microsatellite instability (MSI) detection in colorectal cancer patients. *Oncotarget* 9, 28691–28701.
- Bettington, M., Walker, N., Clouston, A., Brown, I., Leggett, B., Whitehall, V., 2013. The serrated pathway to colorectal carcinoma: Current concepts and challenges. *Histopathology* 62, 367–386.
- Biankin, A. V., Waddell, N., Kassahn, K.S., Gingras, M.C., Muthuswamy, L.B., Johns, A.L., Miller, D.K., Wilson, P.J., Patch, A.M., Wu, J., Chang, D.K., Cowley, M.J., Gardiner, B.B., Song, S., Harliwong, I., Idrisoglu, S., Nourse, C., Nourbakhsh, E., Manning, S., Wani, S., Gongora, M., Pajic, M., Scarlett, C.J., Gill, A.J., Pinho, A. V., Rooman, I., Anderson, M., Holmes, O., Leonard, C., Taylor, D., Wood, S., Xu, Q., Nones, K., Fink, J.L., Christ, A., Bruxner, T., Cloonan, N., Kolle, G., Newell, F., Pinese, M., Mead, R.S., Humphris, J.L., Kaplan, W., Jones, M.D., Colvin, E.K., Nagrial, A.M., Humphrey, E.S., Chou, A., Chin, V.T., Chantrill, L.A., Mawson, A., Samra, J.S., Kench, J.G., Lovell, J.A., Daly, R.J., Merrett, N.D., Toon, C., Epari, K., Nguyen, N.Q., Barbour, A., Zeps, N., Kakkar, N., Zhao, F., Wu, Y.Q., Wang, M., Muzny, D.M., Fisher, W.E., Brunicardi, F.C., Hodges, S.E., Reid, J.G., Drummond, J., Chang, K., Han, Y., Lewis, L.R., Dinh, H., Buhay, C.J., Beck, T., Timms, L., Sam, M., Begley, K., Brown, A., Pai, D., Panchal, A., Buchner, N., De Borja, R., Denroche, R.E., Yung, C.K., Serra, S., Onetto, N., Mukhopadhyay, D., Tsao, M.S., Shaw, P.A., Petersen, G.M., Gallinger, S., Hruban, R.H., Maitra, A., Iacobuzio-Donahue, C.A., Schulick, R.D., Wolfgang, C.L., Morgan, R.A., Lawlor, R.T., Capelli, P., Corbo, V., Scardoni, M., Tortora, G., Tempero, M.A., Mann, K.M., Jenkins, N.A., Perez-Mancera, P.A., Adams, D.J., Largaespada, D.A., Wessels, L.F.A., Rust, A.G., Stein, L.D., Tuveson, D.A., Copeland, N.G., Musgrove, E.A., Scarpa, A., Eshleman, J.R., Hudson, T.J., Sutherland, R.L., Wheeler, D.A., Pearson, J. V., McPherson, J.D., Gibbs, R.A., Grimmond, S.M., 2012. Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* 491, 399–405.
- Bienz, M., Clevers, H., 2000. Linking colorectal cancer to Wnt signaling. *Cell* 103, 311–320.
- Blencowe, B.J., 2006. Alternative Splicing: New Insights from Global Analyses. *Cell* 126, 37–47.
- Bohaumilitzky, L., Kluck, K., Hüneburg, R., Gallon, R., Nattermann, J., Kirchner, M., Kristiansen, G., Hommerding, O., Pfuderer, P.L., Wagner, L., Echterdiek, F., Kösegi, S., Müller, N., Fischer, K., Nelius, N., Hartog, B., Borthwick, G., Busch, E., Haag, G.M., Bläker, H., Möslein, G., von Knebel Doeberitz, M., Seppälä, T.T., Ahtiainen, M., Mecklin, J.P., Bishop, D.T., Burn, J., Stenzinger, A., Budczies, J., Kloor, M., Ahadova, A., 2022. The Different Immune Profiles of Normal Colonic Mucosa in Cancer-Free Lynch Syndrome

- Carriers and Lynch Syndrome Colorectal Cancer Patients. *Gastroenterology* 162, 907-919.e10.
- Bokhari, A., Jonchere, V., Lagrange, A., Bertrand, R., Svrcek, M., Marisa, L., Buhard, O., Greene, M., Anastasia Demidova, J.J., Adriaenssens, E., Chassat, T., Biard, D.S., Flejou, J.-F., Duval, A., Collura, A., 2018. Targeting nonsense-mediated mRNA decay in colorectal cancers with microsatellite instability. *Oncogenesis* 7.
- Boland, C.R., Boland, C.R., Thibodeau, S.N., Thibodeau, S.N., Hamilton, S.R., Hamilton, S.R., Sidransky, D., Sidransky, D., Eshleman, J.R., Eshleman, J.R., Burt, R.W., Burt, R.W., Meltzer, S.J., Meltzer, S.J., Rodriguez-Bigas, M.A., Rodriguez-Bigas, M.A., Fodde, R., Fodde, R., Ranzani, G.N., Ranzani, G.N., Srivastava, S., Srivastava, S., 1998. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* 58, 5248–5257.
- Boland, C.R., Shike, M., 2010. Report From the Jerusalem Workshop on Lynch Syndrome-Hereditary Nonpolyposis Colorectal Cancer. *Gastroenterology* 138, 2197.e1-2197.e7.
- Boland, P.M., Yurgelun, M.B., Boland, C.R., 2018. Recent progress in Lynch syndrome and other familial colorectal cancer syndromes. *CA. Cancer J. Clin.* 68, 217–231.
- Boman, B.M., Walters, R., Fields, J.Z., Kovatich, A.J., Zhang, T., Isenberg, G.A., Goldstein, S.D., Palazzo, J.P., 2004. Colonic crypt changes during adenoma development in familial adenomatous polyposis: Immunohistochemical evidence for expansion of the crypt base cell population. *Am. J. Pathol.* 165, 1489–1498.
- Bonneville, R., Paruchuri, A., Wing, M.R., Krook, M.A., Reeser, J.W., Chen, H.Z., Dao, T., Samorodnitsky, E., Smith, A.M., Yu, L., Nowacki, N., Chen, W., Roychowdhury, S., 2021. Characterization of clonal evolution in microsatellite unstable metastatic cancers through multiregional tumor sequencing. *Mol. Cancer Res.* 19, 465–475.
- Boscolo Bielo, L., Trapani, D., Repetto, M., Crimini, E., Valenza, C., Belli, C., Criscitiello, C., Marra, A., Subbiah, V., Curigliano, G., 2023. Variant allele frequency: a decision-making tool in precision oncology? *Trends in Cancer* 1–11.
- Bradley, R.K., Anczuków, O., 2023. RNA splicing dysregulation and the hallmarks of cancer. *Nat. Rev. Cancer* 23, 135–155.
- Brand, R.E., Dudley, B., Karloski, E., Das, R., Fuhrer, K., Pai, Rish K., Pai, Reetesh K., 2020. Detection of DNA mismatch repair deficient crypts in random colonoscopic biopsies identifies Lynch syndrome patients. *Fam. Cancer* 19, 169–175.
- Buhard, O., Cattaneo, F., Yick, F.W., So, F.Y., Friedman, E., Flejou, J.F., Duval, A., Hamelin, R., 2006. Multipopulation analysis of polymorphisms in five mononucleotide repeats used to determine the microsatellite instability status of human tumors. *J. Clin. Oncol.* 24, 241–251.
- Buhard, O., Lagrange, A., Guilloux, A., Colas, C., Chouchène, M., Wanherdrick, K., Coulet, F., Guillerm, E., Dorard, C., Marisa, L., Bokhari, A., Greene, M., El-Murr, N., Bodo, S., Muleris, M., Sourouille, I., Svrcek, M., Cervera, P., Blanché, H., Lefevre, J.H., Parc, Y., Lepage, C., Chapusot, C., Bouvier, A.M., Gaub, M.P., Selves, J., Garrett, K., Iacopetta, B.,

- Soong, R., Hamelin, R., Garrido, C., Lascols, O., André, T., Fléjou, J.F., Collura, A., Duval, A., 2016. HSP110 T17 simplifies and improves the microsatellite instability testing in patients with colorectal cancer. *J. Med. Genet.* 53, 377–384.
- Burn, J., Gerdes, A.M., MacRae, F., Mecklin, J.P., Moeslein, G., Olschwang, S., Eccles, D., Evans, D.G., Maher, E.R., Bertario, L., Bisgaard, M.L., Dunlop, M.G., Ho, J.W.C., Hodgson, S. V., Lindblom, A., Lubinski, J., Morrison, P.J., Murday, V., Ramesar, R., Side, L., Scott, R.J., Thomas, H.J.W., Vasen, H.F., Barker, G., Crawford, G., Elliott, F., Movahedi, M., Pylvanainen, K., Wijnen, J.T., Fodde, R., Lynch, H.T., Mathers, J.C., Bishop, D.T., 2011. Long-term effect of aspirin on cancer risk in carriers of hereditary colorectal cancer: An analysis from the CAPP2 randomised controlled trial. *Lancet* 378, 2081–2087.
- Burn, J., Mathers, J.C., Bishop, D.T., 2013. Chemoprevention in Lynch syndrome. *Fam. Cancer* 12, 707–718.
- Burton-Chase, A.M., Parker, W.M., Polivka, K.M., Gritz, E.R., Amos, C.I., Lu, K.H., Lynch, P.M., Rodriguez-Bigas, M.A., Nancy You, Y., Peterson, S.K., 2017. A comparison between Lynch syndrome and sporadic colorectal cancer survivors' satisfaction with their healthcare providers. *Cancer Med.* 6, 698–707.
- Cáceres, J.F., Kornblihtt, A.R., 2002. Alternative splicing: Multiple control mechanisms and involvement in human disease. *Trends Genet.* 18, 186–193.
- Cao, R., Yang, F., Ma, S.C., Liu, L., Zhao, Y., Li, Y., Wu, D.H., Wang, T., Lu, W.J., Cai, W.J., Zhu, H.B., Guo, X.J., Lu, Y.W., Kuang, J.J., Huan, W.J., Tang, W.M., Huang, K., Huang, J., Yao, J., Dong, Z.Y., 2020. Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in Colorectal Cancer. *Theranostics* 10, 11080–11091.
- Carethers, J.M., Chauhan, D.P., Fink, D., Nebel, S., Bresalier, R.S., Howell, S.B., Boland, C.R., 1999. Mismatch repair proficiency and in vitro response to 5-fluorouracil. *Gastroenterology* 117, 123–131.
- Carethers, J.M., Jung, B.H., 2015. Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer. *Gastroenterology* 149, 1177-1190.e3.
- Carethers, J.M., Stoffel, E.M., 2015. Lynch syndrome and Lynch syndrome mimics: The growing complex landscape of hereditary colon cancer. *World J. Gastroenterol.* 21, 9253–9261.
- Cercek, A., Lumish, M., Sinopoli, J., Weiss, J., Shia, J., Lamendola-Essel, M., El Dika, I.H., Segal, N., Shcherba, M., Sugarman, R., Stadler, Z., Yaeger, R., Smith, J.J., Rousseau, B., Argiles, G., Patel, M., Desai, A., Saltz, L.B., Widmar, M., Iyer, K., Zhang, J., Gianino, N., Crane, C., Romesser, P.B., Pappou, E.P., Paty, P., Garcia-Aguilar, J., Gonen, M., Gollub, M., Weiser, M.R., Schalper, K.A., Diaz, L.A., 2022. PD-1 Blockade in Mismatch Repair–Deficient, Locally Advanced Rectal Cancer. *N. Engl. J. Med.* 386, 2363–2376.
- Chen, L.H., Cuang, E.Y., 2019. Importance of semaphorins in cancer immunity. *Transl. Lung Cancer Res.* 8, S468–S470.
- Chen, M., Zhao, H., 2019. Next-generation sequencing in liquid biopsy: cancer screening and early detection. *Hum. Genomics* 13, 34.

- Cheng, D.T., Mitchell, T.N., Zehir, A., Shah, R.H., Benayed, R., Syed, A., Chandramohan, R., Liu, Z.Y., Won, H.H., Scott, S.N., Rose Brannon, A., O'Reilly, C., Sadowska, J., Casanova, J., Yannes, A., Hechtman, J.F., Yao, J., Song, W., Ross, D.S., Oultache, A., Dogan, S., Borsu, L., Hameed, M., Nafa, K., Arcila, M.E., Ladanyi, M., Berger, M.F., 2015. Memorial sloan kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): A hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J. Mol. Diagnostics* 17, 251–264.
- Chu, D., Wei, L., 2019. Nonsynonymous, synonymous and nonsense mutations in human cancer-related genes undergo stronger purifying selections than expectation. *BMC Cancer* 19, 1–12.
- Cohen, R., Buhard, O., Cervera, P., Hain, E., Dumont, S., Bardier, A., Bachet, J.B., Gornet, J.M., Lopez-Trabada, D., Dumont, S., Kaci, R., Bertheau, P., Renaud, F., Bibeau, F., Parc, Y., Vernerey, D., Duval, A., Svrcek, M., André, T., 2017. Clinical and molecular characterisation of hereditary and sporadic metastatic colorectal cancers harbouring microsatellite instability/DNA mismatch repair deficiency. *Eur. J. Cancer* 86, 266–274.
- Cohen, R., Hain, E., Buhard, O., Guilloux, A., Bardier, A., Kaci, R., Bertheau, P., Renaud, F., Bibeau, F., Fléjou, J.F., André, T., Svrcek, M., Duval, A., 2019a. Association of Primary Resistance to Immune Checkpoint Inhibitors in Metastatic Colorectal Cancer with Misdiagnosis of Microsatellite Instability or Mismatch Repair Deficiency Status. *JAMA Oncol.* 5, 551–555.
- Cohen, R., Pellat, A., Boussion, H., Svrcek, M., Lopez-Trabada, D., Trouilloud, I., Afchain, P., André, T., 2019b. Immunotherapy and metastatic colorectal cancers with microsatellite instability or mismatch repair deficiency. *Bull. Cancer*.
- Colas, C., Coulet, F., Svrcek, M., Collura, A., Fléjou, J.F., Duval, A., Hamelin, R., 2012. Lynch or Not Lynch? Is that Always a Question?, 1st ed, *Advances in Cancer Research*. Elsevier Inc.
- Colle, R., Cohen, R., 2019. Epidemiology of microsatellite instability across solid neoplasms. *Bull. Cancer* 106, 114–118.
- Colle, R., Cohen, R., Cochereau, D., Duval, A., Lascols, O., Lopez-Trabada, D., Afchain, P., Trouilloud, I., Parc, Y., Lefevre, J.H., Fléjou, J.F., Svrcek, M., André, T., 2017. Immunothérapie et patients traités pour cancer avec instabilité des microsatellites. *Bull. Cancer* 104, 42–51.
- Colvin, H., Nishida, N., Konno, M., Haraguchi, N., Takahashi, H., Nishimura, J., Hata, T., Kawamoto, K., Asai, A., Tsunekuni, K., Koseki, J., Mizushima, T., Satoh, T., Doki, Y., Mori, M., Ishii, H., 2016. Oncometabolite D-2-Hydroxyglurate Directly Induces Epithelial-Mesenchymal Transition and is Associated with Distant Metastasis in Colorectal Cancer. *Sci. Rep.* 6, 1–11.
- Corcus, L., Solier, S., 2005. Épissage Alternatif, Pathologie Et Thérapeutique Moléculaire . *Médecine/Sciences* 21, 253–260.
- Cortes-Ciriano, I., Lee, S., Park, W.Y., Kim, T.M., Park, P.J., 2017. A molecular portrait of microsatellite instability across multiple cancers. *Nat. Commun.* 8, 1–12.

- Cunningham, D., Humblet, Y., Siena, S., Khayat, D., Bleiberg, H., Santoro, A., Bets, D., Mueser, M., Harstrick, A., Verslype, C., Chau, I., Van Cutsem, E., 2004. Cetuximab Monotherapy and Cetuximab plus Irinotecan in Irinotecan-Refractory Metastatic Colorectal Cancer. *N. Engl. J. Med.* 351, 337–345.
- Cunningham, J.M., Christensen, E.R., Tester, D.J., Kim, C.Y., Roche, P.C., Burgart, L.J., Thibodeau, S.N., 1998. Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. *Cancer Res.* 58, 3455–3460.
- Dalgliesh, G.L., Furge, K., Greenman, C., Chen, L., Bignell, G., Butler, A., Davies, H., Edkins, S., Hardy, C., Latimer, C., Teague, J., Andrews, J., Barthorpe, S., Beare, D., Buck, G., Campbell, P.J., Forbes, S., Jia, M., Jones, D., Knott, H., Kok, C.Y., Lau, K.W., Leroy, C., Lin, M.L., McBride, D.J., Maddison, M., Maguire, S., McLay, K., Menzies, A., Mironenko, T., Mulderrig, L., Mudie, L., Ogmeara, S., Pleasance, E., Rajasingham, A., Shepherd, R., Smith, R., Stebbings, L., Stephens, P., Tang, G., Tarpey, P.S., Turrell, K., Dykema, K.J., Khoo, S.K., Petillo, D., Wondergem, B., Anema, J., Kahnoski, R.J., Teh, B.T., Stratton, M.R., Futreal, P.A., 2010. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* 463, 360–363.
- Day, C.L., Kaufmann, D.E., Kiepiela, P., Brown, J.A., Moodley, E.S., Reddy, S., Mackey, E.W., Miller, J.D., Leslie, A.J., DePierres, C., Mncube, Z., Duraiswamy, J., Zhu, B., Eichbaum, Q., Altfeld, M., Wherry, E.J., Coovadia, H.M., Goulder, P.J.R., Klenerman, P., Ahmed, R., Freeman, G.J., Walker, B.D., 2006. PD-1 expression on HIV-specific T cells is associated with T-cell exhaustion and disease progression. *Nature* 443, 350–354.
- Des Guez, G., Lecaille, C., Mariani, P., Bennamoun, M., Uzzan, B., Nicolas, P., Boisseau, A., Sastre, X., Cucherousset, J., Lagorce, C., Schischmanoff, P.O., Morere, J.F., 2010. Prognostic impact of microsatellite instability in colorectal cancer patients treated with adjuvant FOLFOX. *Anticancer Res.* 30, 4297–4301.
- Deutsch, M., Long, M., 1999. Intron-exon structures of eukaryotic model organisms. *Nucleic Acids Res.* 27, 3219–3228.
- Diaz, L.A., Shiu, K.K., Kim, T.W., Jensen, B.V., Jensen, L.H., Punt, C., Smith, D., Garcia-Carbonero, R., Benavides, M., Gibbs, P., de la Fourchardiere, C., Rivera, F., Elez, E., Le, D.T., Yoshino, T., Zhong, W.Y., Fogelman, D., Marinello, P., Andre, T., 2022. Pembrolizumab versus chemotherapy for microsatellite instability-high or mismatch repair-deficient metastatic colorectal cancer (KEYNOTE-177): final analysis of a randomised, open-label, phase 3 study. *Lancet Oncol.* 23, 659–670.
- Diehl, F., Schmidt, K., Choti, M.A., Romans, K., Goodman, S., Li, M., Thornton, K., Agrawal, N., Sokoll, L., Szabo, S.A., Kinzler, K.W., Vogelstein, B., Diaz, L.A., 2008. Circulating mutant DNA to assess tumor dynamics. *Nat. Med.* 14, 985–990.
- Dieu-Nosjean, M.C., Antoine, M., Danel, C., Heudes, D., Wislez, M., Poulot, V., Rabbe, N., Laurans, L., Tartour, E., De Chaisemartin, L., Lebecque, S., Fridman, W.H., Cadranel, J., 2008. Long-term survival for patients with non-small-cell lung cancer with intratumoral lymphoid structures. *J. Clin. Oncol.* 26, 4410–4417.
- Dong, H., Strome, S.E., Salomao, D.R., Tamura, H., Hirano, F., Flies, D.B., Roche, P.C., Lu, J., Zhu, G., Tamada, K., Lennon, V.A., Cells, E., Chen, L., 2002. Tumor-associated B7-H1

- promotes T-cell apoptosis: A potential mechanism of immune evasion. *Nat. Med.* 8, 793–800.
- Dorard, C., Collura, A., Marisa, L., Garrido, C., Duval, A., 2011. Expression of a mutant HSP110 sensitizes colorectal cancer cells to chemotherapy and improves disease prognosis. *Nat. Med.* 17, 1283–1289.
- Dujardin, G., Dagueneat, É., Bernard, D.G., Flodrops, M., Durand, S., Chauveau, A., El Khoury, F., Le Jossic-Corcoss, C., Corcos, L., 2016. L'épissage des ARN pré-messagers: quand le spliceosome perd pied. *Medecine/Sciences* 32, 1103–1110.
- Duval, A., Hamelin, R., 2002a. Genetic instability in human mismatch repair deficient cancers. *Ann. Genet.* 45, 71–75.
- Duval, A., Hamelin, R., 2002b. Mutations at Coding Repeat Sequences in Mismatch Repair-deficient Human Cancers. *Cancer Res.* 62, 2447–2454.
- Duval, A., Hamelin, R., 2003. Réparation des erreurs de réplication, microsatellites et cancer. *Médecine/Sciences* 19, 55–62.
- Dzantiev, L., Constantin, N., Genschel, J., Iyer, R.R., Burgers, P.M., Modrich, P., 2004. A defined human system that supports bidirectional mismatch-provoked excision. *Mol. Cell* 15, 31–41.
- Echle, A., Ghaffari Laleh, N., Quirke, P., Grabsch, H.I., Muti, H.S., Saldanha, O.L., Brockmoeller, S.F., van den Brandt, P.A., Hutchins, G.G.A., Richman, S.D., Horisberger, K., Galata, C., Ebert, M.P., Eckardt, M., Boutros, M., Horst, D., Reissfelder, C., Alwers, E., Brinker, T.J., Langer, R., Jenniskens, J.C.A., Offermans, K., Mueller, W., Gray, R., Gruber, S.B., Greenson, J.K., Rennert, G., Bonner, J.D., Schmolze, D., Chang-Claude, J., Brenner, H., Trautwein, C., Boor, P., Jaeger, D., Gaisa, N.T., Hoffmeister, M., West, N.P., Kather, J.N., 2022. Artificial intelligence for detection of microsatellite instability in colorectal cancer—a multicentric analysis of a pre-screening tool for clinical application. *ESMO Open* 7, 100400.
- Echle, A., Grabsch, H.I., Quirke, P., van den Brandt, P.A., West, N.P., Hutchins, G.G.A., Heij, L.R., Tan, X., Richman, S.D., Krause, J., Alwers, E., Jenniskens, J., Offermans, K., Gray, R., Brenner, H., Chang-Claude, J., Trautwein, C., Pearson, A.T., Boor, P., Luedde, T., Gaisa, N.T., Hoffmeister, M., Kather, J.N., 2020. Clinical-Grade Detection of Microsatellite Instability in Colorectal Tumors by Deep Learning. *Gastroenterology* 159, 1406-1416.e11.
- Echle, A., Rindtorff, N.T., Brinker, T.J., Luedde, T., Pearson, A.T., Kather, J.N., 2021. Deep learning in cancer pathology: a new generation of clinical biomarkers. *Br. J. Cancer* 124, 686–696.
- Edelstein, D.L., Axilbund, J., Baxter, M., Hylind, L.M., Romans, K., Griffin, C.A., Cruz-Correa, M., Giardiello, F.M., 2011. Rapid Development of Colorectal Neoplasia in Patients With Lynch Syndrome. *Clin. Gastroenterol. Hepatol.* 9, 340–343.
- Eisenbarth, S.C., Piggott, D.A., Huleatt, J.W., Visintin, I., Herrick, C.A., Bottomly, K., 2002. Lipopolysaccharide-enhanced, toll-like receptor 4-dependent T helper cell type 2 responses to inhaled antigen. *J. Exp. Med.* 196, 1645–1651.

- El-Bchiri, J., Guilloux, A., Dartigues, P., Loire, E., Mercier, D., Buhard, O., Sobhani, I., de la Grange, P., Auboeuf, D., Praz, F., Fléjou, J.F., Duval, A., 2008. Nonsense-mediated mRNA decay impacts MSI-driven carcinogenesis and anti-tumor immunity in colorectal cancers. *PLoS One* 3, 1–10.
- Ellegren, H., 2004. Microsatellites: Simple sequences with complex evolution. *Nat. Rev. Genet.* 5, 435–445.
- Endo, E., Okayama, H., Saito, K., Nakajima, S., Yamada, L., Ujiie, D., Kase, K., Fujita, S., Endo, H., Sakamoto, W., Saito, M., Saze, Z., Momma, T., Ohki, S., Mimura, K., Kono, K., 2020. A TGFb-dependent stromal subset underlies immune checkpoint inhibitor efficacy in DNA mismatch repair-deficient/microsatellite instability-high colorectal cancer. *Mol. Cancer Res.* 18, 1402–1413.
- Espenschied, C.R., LaDuca, H., Li, S., McFarland, R., Gau, C.L., Hampel, H., 2017. Multigene panel testing provides a new perspective on Lynch syndrome. *J. Clin. Oncol.* 35, 2568–2575.
- Evrard, C., Alexandre, J., 2021. Predictive and prognostic value of microsatellite instability in gynecologic cancer (Endometrial and ovarian). *Cancers (Basel).* 13, 1–15.
- Fearon, E.R., 2011. Molecular genetics of colorectal cancer. *Annu. Rev. Pathol. Mech. Dis.* 6, 479–507.
- Fink, D., Nebel, S., Aebi, S., Zheng, H., Cenm, B., Nehmã, A., Christen, D., Howell, S.B., 1996. The Role of DNA Mismatch Repair in Platinum Drug Resistance. *Cancer Res.* 4881–4886.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C., 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811.
- Flechia, C., Zaanani, A., Lahlou, W., Basile, D., Broudin, C., Gallois, C., Pilla, L., Karoui, M., Manceau, G., Taieb, J., 2022. MSI colorectal cancer, all you need to know. *Clin. Res. Hepatol. Gastroenterol.* 46.
- Fodde, R., Verhagen, M., Joosten, R., Schmitt, M., Sacchetti, A., Choi, J., Välimäki, N., Aaltonen, L., Augenlicht, L., 2023. Paneth cells as the origin of intestinal cancer in the context of inflammation, *Research square*.
- Foot, M.B., Maron, S.B., Cercek, A., Argilés, G., Rousseau, B., Diaz, L.A., 2021. TMB cut-offs fail to predict benefit of PD-1 blockade in gastroesophageal adenocarcinoma in KEYNOTE-061. *Ann. Oncol.* 32, 1188–1189.
- Fourcade, J., Sun, Z., Benallaoua, M., Guillaume, P., Luescher, I.F., Sander, C., Kirkwood, J.M., Kuchroo, V., Zarour, H.M., 2010. Upregulation of Tim-3 and PD-1 expression is associated with tumor antigen-specific CD8+ T cell dysfunction in melanoma patients. *J. Exp. Med.* 207, 2175–2186.
- Fre, S., Pallavi, S.K., Huyghe, M., Laé, M., Janssen, K.P., Robine, S., Artavanis-Tsakonas, S., Louvard, D., 2009. Notch and Wnt signals cooperatively control cell proliferation and tumorigenesis in the intestine. *Proc. Natl. Acad. Sci. U. S. A.* 106, 6309–6314.
- Fridman, W.H., Pagès, F., Sautès-Fridman, C., Galon, J., 2012. The immune contexture in

- human tumours: Impact on clinical outcome. *Nat. Rev. Cancer* 12, 298–306.
- Frigola, J., Sabarinathan, R., Mularoni, L., Muinõs, F., Gonzalez-Perez, A., López-Bigas, N., 2017. Reduced mutation rate in exons due to differential mismatch repair. *Nat. Genet.* 49, 1684–1692.
- Fu, D., Calvo, J.A., Samson, L.D., 2012. Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat. Rev. Cancer* 12, 104–120.
- Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., Tosolini, M., Camus, M., Berger, A., Wind, P., Zinzindohoué, F., Bruneval, P., Cugnenc, P.H., Trajanoski, Z., Fridman, W.H., Pagès, F., 2006. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* (80-). 313, 1960–1964.
- Gambini, D., Ferrero, S., Kuhn, E., 2022. Lynch Syndrome: From Carcinogenesis to Prevention Interventions. *Cancers (Basel)*. 14, 1–21.
- Ganai, R.A., Johansson, E., 2016. DNA Replication-A Matter of Fidelity. *Mol. Cell* 62, 745–755.
- Gardner, L.B., 2010. Nonsense-mediated RNA decay regulation by cellular stress: Implications for tumorigenesis. *Mol. Cancer Res.* 8, 295–308.
- Gebert, J., Gelincik, O., Oezcan-Wahlbrink, M., Marshall, J.D., Hernandez-Sanchez, A., Urban, K., Long, M., Cortes, E., Tosti, E., Katzenmaier, E.M., Song, Y., Elsaadi, A., Deng, N., Vilar, E., Fuchs, V., Nelius, N., Yuan, Y.P., Ahadova, A., Sei, S., Shoemaker, R.H., Umar, A., Wei, L., Liu, S., Bork, P., Edelmann, W., von Knebel Doeberitz, M., Lipkin, S.M., Kloor, M., 2021. Recurrent Frameshift Neoantigen Vaccine Elicits Protective Immunity With Reduced Tumor Burden and Improved Overall Survival in a Lynch Syndrome Mouse Model. *Gastroenterology* 161, 1288-1302.e13.
- Gehart, H., Clevers, H., 2019. Tales from the crypt: new insights into intestinal stem cells. *Nat. Rev. Gastroenterol. Hepatol.* 16, 19–34.
- Gehring, N.H., Roignant, J.Y., 2021. Anything but Ordinary – Emerging Splicing Mechanisms in Eukaryotic Gene Regulation. *Trends Genet.* 37, 355–372.
- Genschel, J., Kadyrova, L.Y., Iyer, R.R., Dahal, B.K., Kadyrov, F.A., Modrich, P., 2017. Interaction of proliferating cell nuclear antigen with PMS2 is required for MutL α activation and function in mismatch repair. *Proc. Natl. Acad. Sci. U. S. A.* 114, 4930–4935.
- Gerlinger M, Rowan AJ, H.S., 2012. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* 22, 2012–2014.
- Geuens, T., Bouhy, D., Timmerman, V., 2016. The hnRNP family: insights into their role in health and disease. *Hum. Genet.* 135, 851–867.
- Giannini, G., Rinaldi, C., Ristori, E., Ambrosini, M.I., Cerignoli, F., Viel, A., Bidoli, E., Berni, S., D’Amati, G., Scambia, G., Frati, L., Screpanti, I., Gulino, A., 2004. Mutations of an intronic repeat induce impaired MRE11 expression in primary human cancer with microsatellite instability. *Oncogene* 23, 2640–2647.
- Gill, S., Lindor, N.M., Burgart, L.J., Smalley, R., Leontovich, O., French, A., Goldberg, R.M.,

- Sargent, D.J., Jass, J.R., Hopper, J.L., Jenkins, M.A., Young, J., Barker, M.A., Walsh, M.D., Ruzskiewicz, A.R., Thibodeau, S.N., 2005. Isolated loss of PMS2 expression in colorectal cancers: Frequency, patient age, and familial aggregation. *Clin. Cancer Res.* 11, 6466–6471.
- Gilson, P., Merlin, J.L., Harlé, A., 2021. Detection of microsatellite instability: State of the art and future applications in circulating tumour dna (ctdna). *Cancers (Basel)*. 13.
- Glaviano, A., Foo, A.S.C., Lam, H.Y., Yap, K.C.H., Jacot, W., Jones, R.H., Eng, H., Nair, M.G., Makvandi, P., Georger, B., Kulke, M.H., Baird, R.D., Prabhu, J.S., Carbone, D., Pecoraro, C., Teh, D.B.L., Sethi, G., Cavalieri, V., Lin, K.H., Javidi-Sharifi, N.R., Toska, E., Davids, M.S., Brown, J.R., Diana, P., Stebbing, J., Fruman, D.A., Kumar, A.P., 2023. PI3K/AKT/mTOR signaling transduction pathway and targeted therapies in cancer. *Mol. Cancer* 22, 1–37.
- Goldstein, J., Tran, B., Ensor, J., Gibbs, P., Wong, H.L., Wong, S.F., Vilar, E., Tie, J., Broaddus, R., Kopetz, S., Desai, J., Overman, M.J., 2014. Multicenter retrospective analysis of metastatic colorectal cancer (CRC) with high-level microsatellite instability (MSI-H). *Ann. Oncol.* 25, 1032–1038.
- Gotoh, M., Ichikawa, H., Arai, E., Chiku, S., Sakamoto, H., Fujimoto, H., Hiramoto, M., Nammo, T., Yasuda, K., Yoshida, T., Kanai, Y., 2014. Comprehensive exploration of novel chimeric transcripts in clear cell renal cell carcinomas using whole transcriptome analysis. *Genes Chromosom. Cancer* 396, 389–396.
- Graham, T.A., Humphries, A., Sanders, T., Rodriguezjusto, M., Tadrous, P.J., Preston, S.L., Novelli, M.R., Leedham, S.J., McDonald, S.A.C., Wright, N.A., 2011. Use of methylation patterns to determine expansion of stem cell clones in human colon tissue. *Gastroenterology* 140, 1241-1250.e9.
- Granier, C., Soumelis, V., Mandavit, M., Gibault, L., Belazzoug, R., de Guillebon, E., Badoual, C., Tartour, E., Roussel, H., 2017. The “immune checkpoints”, how does it work. *Ann. Pathol.* 37, 18–28.
- Greaves, L.C., Preston, S.L., Tadrous, P.J., Taylor, R.W., Barron, M.J., Oukrif, D., Leedham, S.J., Deheragoda, M., Sasieni, P., Novelli, M.R., Jankowski, J.A.Z., Turnbull, D.M., Wright, N.A., McDonald, S.A.C., 2006. Mitochondrial DNA mutations are established in human colonic stem cells, and mutated clones expand by crypt fission. *Proc. Natl. Acad. Sci. U. S. A.* 103, 714–719.
- Green, M.R., Monti, S., Rodig, S.J., Juszczynski, P., Currie, T., O’Donnell, E., Chapuy, B., Takeyama, K., Neuberg, D., Golub, T.R., Kutok, J.L., Shipp, M.A., 2010. Integrative analysis reveals selective 9p24.1 amplification, increased PD-1 ligand expression, and further induction via JAK2 in nodular sclerosing Hodgkin lymphoma and primary mediastinal large B-cell lymphoma. *Blood* 116, 3268–3277.
- Groothuizen, F.S., Sixma, T.K., 2016. The conserved molecular machinery in DNA mismatch repair enzyme structures. *DNA Repair (Amst)*. 38, 14–23.
- Guinney, J., Dienstmann, R., Wang, X., De Reyniès, A., Schlicker, A., Sonesson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., Bot, B.M., Morris, J.S., Simon, I.M., Gerster, S., Fessler, E., De Sousa .E Melo, F., Missiaglia, E., Ramay, H., Barras, D.,

- Homicsko, K., Maru, D., Manyam, G.C., Broom, B., Boige, V., Perez-Villamil, B., Laderas, T., Salazar, R., Gray, J.W., Hanahan, D., Tabernero, J., Bernardis, R., Friend, S.H., Laurent-Puig, P., Medema, J.P., Sadanandam, A., Wessels, L., Delorenzi, M., Kopetz, S., Vermeulen, L., Tejpar, S., 2015. The consensus molecular subtypes of colorectal cancer. *Nat. Med.* 21, 1350–1356.
- Guo, S., Zhang, Y., Yuan, F., Gao, Y., Gu, L., Wong, I., Li, G.M., 2006. Regulation of replication protein A functions in DNA mismatch repair by phosphorylation. *J. Biol. Chem.* 281, 21607–21616.
- Gupta, D., Heinen, C.D., 2019. The mismatch repair-dependent DNA damage response: Mechanisms and implications. *DNA Repair (Amst)*. 78, 60–69.
- Guyot D’Asnières De Salins, A., Tachon, G., Cohen, R., Karayan-Tapon, L., Junca, A., Frouin, E., Godet, J., Evrard, C., Randrian, V., Duval, A., Svrcek, M., Lascols, O., Vignot, S., Coulet, F., André, T., Fléjou, J.F., Cervera, P., Tougeron, D., 2021. Discordance between immunochemistry of mismatch repair proteins and molecular testing of microsatellite instability in colorectal cancer. *ESMO Open* 6, 100120.
- Hammond, W.A., Swaika, A., Mody, K., 2016. Pharmacologic resistance in colorectal cancer: A review. *Ther. Adv. Med. Oncol.* 8, 57–84.
- Hanahan, D., 2022. Hallmarks of Cancer: New Dimensions. *Cancer Discov.* 12, 31–46.
- Hang, J., Wan, R., Yan, C., 2015. Structural basis of pre-mRNA splicing. *Science (80-)*. 4–7.
- Harbour, J.W., Roberson, E.D.O., Anbunathan, H., Onken, M.D., Worley, L.A., Bowcock, A.M., 2013. Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. *Nat. Genet.* 45, 133–135.
- Hause, R.J., Pritchard, C.C., Shendure, J., Salipante, S.J., 2016. Classification and characterization of microsatellite instability across 18 cancer types. *Nat. Med.* 22, 1342–1350.
- Hector, S., Bolanowska-Higdon, W., Zdanowicz, J., Hitt, S., Pendyala, L., 2001. In vitro studies on the mechanisms of oxaliplatin resistance. *Cancer Chemother. Pharmacol.* 48, 398–406.
- Heldin, C.H., Miyazono, K., Ten Dijke, P., 1997. TGF- β signalling from cell membrane to nucleus through SMAD proteins. *Nature* 390, 465–471.
- Herberts, C., Annala, M., Sipola, J., Ng, S.W.S., Chen, X.E., Nurminen, A., Korhonen, O. V., Munzur, A.D., Beja, K., Schönlau, E., Bernales, C.Q., Ritch, E., Bacon, J.V.W., Lack, N.A., Nykter, M., Aggarwal, R., Small, E.J., Gleave, M.E., Quigley, D.A., Feng, F.Y., Chi, K.N., Wyatt, A.W., 2022. Deep whole-genome ctDNA chronology of treatment-resistant prostate cancer. *Nature* 608, 199–208.
- Hernandez-Sanchez, A., Grossman, M., Yeung, K., Sei, S.S., Lipkin, S., Kloor, M., 2022. Vaccines for immunoprevention of DNA mismatch repair deficient cancers. *J. Immunother. Cancer* 10.
- Hitchins, M.P., Ward, R.L., 2009. Constitutional (germline) MLH1 epimutation as an aetiological mechanism for hereditary non-polyposis colorectal cancer. *J. Med. Genet.*

46, 793–802.

- Howard, J.M., Sanford, J.R., 2015. The RNAissance family: SR proteins as multifaceted regulators of gene expression. *Wiley Interdiscip. Rev. RNA* 6, 93–110.
- Hsieh, P., Yamane, K., 2008. DNA mismatch repair: Molecular mechanism, cancer, and ageing. *Mech. Ageing Dev.* 129, 391–407.
- Huang, Y., Li, G.M., 2018. DNA mismatch repair preferentially safeguards actively transcribed genes. *DNA Repair (Amst)*. 71, 82–86.
- Huang, Y., Li, G.M., 2020. DNA mismatch repair in the chromatin context: Mechanisms and therapeutic potential. *DNA Repair (Amst)*. 93, 102918.
- Humphries, A., Wright, N.A., 2008. Colonic crypt organization and tumorigenesis. *Nat. Rev. Cancer* 8, 415–424.
- Hung, Y.-H., Lai, M.-D., Hung, W.-C., Chen, L.-T., 2022. Semaphorin 4C promotes motility and immunosuppressive activity of cancer cells via CRMP3 and PD-L1. *Am J Cancer Res* 12, 713–728.
- Ijsselsteijn, R., Jansen, J.G., de Wind, N., 2020. DNA mismatch repair-dependent DNA damage responses and cancer. *DNA Repair (Amst)*. 93, 102923.
- Imai, K., Yamamoto, H., 2008. Carcinogenesis and microsatellite instability: The interrelationship between genetics and epigenetics. *Carcinogenesis* 29, 673–680.
- Ionov, Y., Shibata, D., Perucho, M., Peinado, M.A., Malkhosyan, S., 1993. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature* 363, 558–561.
- Ishida, Y., Agata, Y., Shibahara, K., Honjo, T., 1992. Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death. *EMBO J.* 11, 3887–3895.
- Ishii, T., Notohara, K., Umopathy, A., Mallitt, K.A., Chikuba, H., Moritani, Y., Tanaka, N., Rosty, C., Matsubara, N., Jass, J., Leggett, B., Whitehall, V., 2011. Tubular adenomas with minor villous changes show molecular features characteristic of tubulovillous adenomas. *Am. J. Surg. Pathol.* 35, 212–220.
- Itatani, Y., Kawada, K., Sakai, Y., 2019. Transforming growth factor- β signaling pathway in colorectal cancer and its tumor microenvironment. *Int. J. Mol. Sci.* 20.
- Jaffrelot, M., Farés, N., Brunac, A.C., Laurenty, A.P., Danjoux, M., Grand, D., Icher, S., Meilleroux, J., Mery, E., Buscail, E., Maulat, C., Toulas, C., Vande Perre, P., Chipoulet, E., Bonnet, D., Staub, A., Guimbaud, R., Selves, J., 2022. An unusual phenotype occurs in 15% of mismatch repair-deficient tumors and is associated with non-colorectal cancers and genetic syndromes. *Mod. Pathol.* 35, 427–437.
- Järvinen, H.J., Aarnio, M., Mustonen, H., AktanCollan, K., Aaltonen, L.A., Peltomäki, P., Chapelle, A.D. La, Mecklin, J.P., 2000. Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer. *Gastroenterology* 118, 829–834.

- Jass, J.R., 2007. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology* 50, 113–130.
- Jensen, J., Pedersen, E.E., Galante, P., Hald, J., Scott Heller, R., Ishibashi, M., Kageyama, R., Guillemot, F., Serup, P., Madsen, O.D., 2000. Control of endodermal endocrine development by Hes-1 Development of endocrine cells in the endoderm involves Atonal and Achaete/Scute-related basic helix-loop-helix. *Nat. Genet.* 24, 36–44.
- Jia, P., Yang, X., Guo, L., Liu, B., Lin, J., Liang, H., Sun, J., Zhang, C., Ye, K., 2020. MSIsensor-pro: Fast, Accurate, and Matched-normal-sample-free Detection of Microsatellite Instability. *Genomics, Proteomics Bioinforma.* 18, 65–71.
- Jin, M., Shang, F., Wu, J., Fan, Q., Chen, C., Fan, J., Liu, L., Nie, X., Zhang, T., Cai, K., Ogino, S., Liu, H., 2021. Tumor-Associated Microbiota in Proximal and Distal Colorectal Cancer and Their Relationships With Clinical Outcomes. *Front. Microbiol.* 12.
- Jonchere, V., Marisa, L., Greene, M., Virouleau, A., Buhard, O., Bertrand, R., Svrcek, M., Cervera, P., Goloudina, A., Guillerme, E., Coulet, F., Landman, S., Ratovomanana, T., Job, S., Ayadi, M., Elarouci, N., Armenoult, L., Merabtene, F., Dumont, S., Parc, Y., Lefèvre, J.H., André, T., Fléjou, J.F., Guilloux, A., Collura, A., de Reyniès, A., Duval, A., 2018. Identification of Positively and Negatively Selected Driver Gene Mutations Associated With Colorectal Cancer With Microsatellite Instability. *Cmgh* 6, 277–300.
- Jung, B., Staudacher, J.J., Beauchamp, D., 2017. Transforming Growth Factor β Superfamily Signaling in Development of Colorectal Cancer. *Gastroenterology* 152, 36–52.
- Jung, H., Lee, D., Lee, J., Park, D., Kim, Y.J., Park, W.Y., Hong, D., Park, P.J., Lee, E., 2015. Intron retention is a widespread mechanism of tumor-suppressor inactivation. *Nat. Genet.* 47, 1242–1248.
- Kabbinavar, F., Hurwitz, H.I., Fehrenbacher, L., Meropol, N.J., Novotny, W.F., Lieberman, G., Griffing, S., Bergsland, E., 2003. Phase I, randomized trial comparing bevacizumab plus fluorouracil (FU)/Leucovorin (LV) with FU/LV alone in patients with metastatic colorectal cancer. *J. Clin. Oncol.* 21, 60–65.
- Kadener, S., Fededa, J.P., Rosbash, M., Kornblihtt, A.R., 2002. Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation. *Proc. Natl. Acad. Sci. U. S. A.* 99, 8185–8190.
- Kadyrov, F.A., Dzantiev, L., Constantin, N., Modrich, P., 2006. Endonucleolytic Function of MutL α in Human Mismatch Repair. *Cell* 126, 297–308.
- Kakar, S., Burgart, L.J., Thibodeau, S.N., Rabe, K.G., Petersen, G.M., Goldberg, R.M., Lindor, N.M., 2003. Frequency of loss of hMLH1 expression in colorectal carcinoma increases with advancing age. *Cancer* 97, 1421–1427.
- Kambara, T., Simms, L.A., Whitehall, V.L.J., Spring, K.J., Wynter, C.V.A., Walsh, M.D., Barker, M.A., Arnold, S., McGivern, A., Matsubara, N., Tanaka, N., Higuchi, T., Young, J., Jass, J.R., Leggett, B.A., 2004. BRAF mutation is associated with DNA methylation in serrated polyps and cancers of the colorectum. *Gut* 53, 1137–1144.
- Kambayashi, T., Laufer, T.M., 2014. Atypical MHC class II-expressing antigen-presenting cells: Can anything replace a dendritic cell? *Nat. Rev. Immunol.* 14, 719–730.

- Kane, M.F., Loda, M., Gaida, G.M., Lipman, J., Mishra, R., Goldman, H., Jessup, J.M., Kolodner, R., 1997. Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res.* 57, 808–811.
- Kather, J.N., Pearson, A.T., Halama, N., Jäger, D., Krause, J., Loosen, S.H., Marx, A., Boor, P., Tacke, F., Neumann, U.P., Grabsch, H.I., Yoshikawa, T., Brenner, H., Chang-Claude, J., Hoffmeister, M., Trautwein, C., Luedde, T., 2019. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat. Med.* 25, 1054–1056.
- Kavitha, M.S., Gangadaran, P., Jackson, A., Venmathi Maran, B.A., Kurita, T., Ahn, B.C., 2022. Deep Neural Network Models for Colon Cancer Screening. *Cancers (Basel)*. 14.
- Kavun, A., Veselovsky, E., Lebedeva, A., Belova, E., Kuznetsova, O., Yakushina, V., Grigoreva, T., Mileyko, V., Fedyanin, M., Ivanov, M., 2023. Microsatellite Instability: A Review of Molecular Epidemiology and Implications for Immune Checkpoint Inhibitor Therapy. *Cancers (Basel)*. 15.
- Kientz, C., Prieur, F., Clemenson, A., Odile, M., Stachowicz, M.L., Auclair, J., Attignon, V., Schiappa, R., Wang, Q., 2019. MLH1 promoter hypermethylation : are you absolutely sure about the absence of MLH1 germline mutation ? About a new case. *Fam. Cancer* 5–8.
- Kim, K.M., Shibata, D., 2002. Methylation reveals a niche: Stem cell succession in human colon crypts. *Oncogene* 21–35, 5441–5449.
- Kim, S.E., Paik, H.Y., Yoon, H., Lee, J.E., Kim, N., Sung, M.K., 2015. Sex- and gender-specific disparities in colorectal cancer risk. *World J. Gastroenterol.* 21, 5167–5175.
- Kinter, A.L., Godbout, E.J., McNally, J.P., Sereti, I., Roby, G.A., Shea, M.A.O., Fauci, A.S., 2008. The Common γ -Chain Cytokines IL-2, IL-7, IL-15, and IL-21 Induce the Expression of Programmed Death-1 and Its Ligands. *J. Immunol.*
- Kishor, A., Ge, Z., Hogg, J.R., 2019. hnRNP L-dependent protection of normal mRNAs from NMD subverts quality control in B cell lymphoma. *EMBO J.* 38, 1–19.
- Kishore, S., Piscuoglio, S., Kovac, M.B., Gylling, A., Wenzel, F., Trapani, F., Altermatt, H.J., Mele, V., Marra, G., Peltomäki, P., Terracciano, L., Zavolan, M., Heinimann, K., 2014. 3'-UTR Poly(T/U) tract deletions and altered expression of EWSR1 are a hallmark of mismatch repair-deficient cancers. *Cancer Res.* 74, 224–234.
- Kloor, M., Huth, C., Voigt, A.Y., Benner, A., Schirmacher, P., von Knebel Doeberitz, M., Bläker, H., 2012. Prevalence of mismatch repair-deficient crypt foci in Lynch syndrome: A pathological study. *Lancet Oncol.* 13, 598–606.
- Kloor, M., Reuschenbach, M., Pauligk, C., Karbach, J., Rafiyan, M.R., Al-Batran, S.E., Tariverdian, M., Jäger, E., von Knebel Doeberitz, M., 2020. A frameshift peptide neoantigen-based vaccine for mismatch repair-deficient cancers: A phase I/IIa clinical trial. *Clin. Cancer Res.* 26, 4503–4510.
- Kloor, M., Von Knebel Doeberitz, M., 2016. The immune biology of microsatellite-unstable cancer. *Trends in Cancer* 2, 121–133.

Koboldt, D.C., Fulton, R.S., McLellan, M.D., Schmidt, H., Kalicki-Veizer, J., McMichael, J.F., Fulton, L.L., Dooling, D.J., Ding, L., Mardis, E.R., Wilson, R.K., Ally, A., Balasundaram, M., Butterfield, Y.S.N., Carlsen, R., Carter, C., Chu, Andy, Chuah, E., Chun, H.J.E., Coope, R.J.N., Dhalla, N., Guin, R., Hirst, C., Hirst, M., Holt, R.A., Lee, D., Li, H.I., Mayo, M., Moore, R.A., Mungall, A.J., Pleasance, E., Robertson, A.G., Schein, J.E., Shafiei, A., Sipahimalani, P., Slobodan, J.R., Stoll, D., Tam, A., Thiessen, N., Varhol, R.J., Wye, N., Zeng, T., Zhao, Y., Birol, I., Jones, S.J.M., Marra, M.A., Cherniack, A.D., Saksena, G., Onofrio, R.C., Pho, N.H., Carter, S.L., Schumacher, S.E., Tabak, B., Hernandez, B., Gentry, J., Nguyen, H., Crenshaw, A., Ardlie, K., Beroukhir, R., Winckler, W., Getz, G., Gabriel, S.B., Meyerson, M., Chin, L., Kucherlapati, R., Hoadley, K.A., Auman, J.T., Fan, C., Turman, Y.J., Shi, Y., Li, L., Topal, M.D., He, X., Chao, H.H., Prat, A., Silva, G.O., Iglesia, M.D., Zhao, W., Usary, J., Berg, J.S., Adams, M., Booker, J., Wu, J., Gulabani, A., Bodenheimer, T., Hoyle, A.P., Simons, J. V., Soloway, M.G., Mose, L.E., Jefferys, S.R., Balu, S., Parker, J.S., Hayes, D.N., Perou, C.M., Malik, S., Mahurkar, S., Shen, H., Weisenberger, D.J., Triche, T., Lai, P.H., Bootwalla, M.S., Maglinte, D.T., Berman, B.P., Van Den Berg, D.J., Baylin, S.B., Laird, P.W., Creighton, C.J., Donehower, L.A., Noble, M., Voet, D., Gehlenborg, N., Di Cara, D., Zhang, Juinhua, Zhang, H., Wu, C.J., Yingchun Liu, S., Lawrence, M.S., Zou, L., Sivachenko, A., Lin, P., Stojanov, P., Jing, R., Cho, J., Sinha, Raktim, Park, R.W., Nazaire, M.D., Robinson, J., Thorvaldsdottir, H., Mesirov, J., Park, P.J., Reynolds, S., Kreisberg, R.B., Bernard, B., Bressler, R., Erkkila, T., Lin, J., Thorsson, V., Zhang, W., Shmulevich, I., Ciriello, G., Weinhold, N., Schultz, N., Gao, J., Cerami, E., Gross, B., Jacobsen, A., Sinha, Rileen, Aksoy, B.A., Antipin, Y., Reva, B., Shen, R., Taylor, B.S., Ladanyi, M., Sander, C., Anur, P., Spellman, P.T., Lu, Y., Liu, W., Verhaak, R.R.G., Mills, G.B., Akbani, R., Zhang, N., Broom, B.M., Casasent, T.D., Wakefield, C., Unruh, A.K., Baggerly, K., Coombes, K., Weinstein, J.N., Haussler, D., Benz, C.C., Stuart, J.M., Benz, S.C., Zhu, J., Szeto, C.C., Scott, G.K., Yau, C., Paull, E.O., Carlin, D., Wong, C., Sokolov, A., Thusberg, J., Mooney, S., Ng, S., Goldstein, T.C., Ellrott, K., Grifford, M., Wilks, C., Ma, S., Craft, B., Yan, C., Hu, Y., Meerzaman, D., Gastier-Foster, J.M., Bowen, J., Ramirez, N.C., Black, A.D., Pyatt, R.E., White, P., Zmuda, E.J., Frick, J., Lichtenberg, T.M., Brookens, R., George, M.M., Gerken, M.A., Harper, H.A., Leraas, K.M., Wise, L.J., Tabler, T.R., McAllister, C., Barr, T., Hart-Kothari, M., Tarvin, K., Saller, C., Sandusky, G., Mitchell, C., Iacocca, M. V., Brown, J., Rabeno, B., Czerwinski, C., Petrelli, N., Dolzhansky, O., Abramov, M., Voronina, O., Potapova, O., Marks, J.R., Suchorska, W.M., Murawa, D., Kycler, W., Ibbs, M., Korski, K., Spychała, A., Murawa, P., Brzeziński, J.J., Perz, H., łażniak, R., Teresiak, M., Tatka, H., Leporowska, E., Bogusz-Czerniewicz, M., Malicki, J., Mackiewicz, A., Wiznerowicz, M., Van Le, X., Kohl, B., Viet Tien, N., Thorp, R., Van Bang, N., Sussman, H., Phu, B.D., Hajek, R., Hung, N.P., Phuong, T.V.T., Thang, H.Q., Khan, K.Z., Penny, R., Mallery, D., Curley, E., Shelton, C., Yena, P., Ingle, J.N., Couch, F.J., Lingle, W.L., King, T.A., Gonzalez-Angulo, A.M., Dyer, M.D., Liu, S., Meng, X., Patangan, M., Waldman, F., Stöppler, H., Rathmell, W.K., Thorne, L., Huang, M., Boice, L., Hill, A., Morrison, C., Gaudioso, C., Bshara, W., Daily, K., Egea, S.C., Pegram, M.D., Gomez-Fernandez, C., Dhir, R., Bhargava, R., Brufsky, A., Shriver, C.D., Hooke, J.A., Campbell, J.L., Mural, R.J., Hu, H., Somiari, S., Larson, C., Deyarmin, B., Kvecher, L., Kovatich, A.J., Ellis, M.J., Stricker, T., White, K., Olopade, O., Luo, C., Chen, Y., Bose, R., Chang, L.W., Beck, A.H., Pihl, T., Jensen, M., Sfeir, R., Kahn, A., Chu, Anna, Kothiyal, P., Wang, Z., Snyder, E., Pontius, J., Ayala, B., Backus, M., Walton, J., Baboud, J., Berton, D., Nicholls, M., Srinivasan, D., Raman, R., Girshik, S., Kigonya, P., Alonso, S., Sanbhadti, R., Barletta,

- S., Pot, D., Sheth, M., Demchok, J.A., Shaw, K.R.M., Yang, L., Eley, G., Ferguson, M.L., Tarnuzzer, R.W., Zhang, Jiashan, Dillon, L.A.L., Buetow, K., Fielding, P., Ozenberger, B.A., Guyer, M.S., Sofia, H.J., Palchik, J.D., 2012. Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61–70.
- Kondelin, J., Tuupanen, S., Gylfe, A.E., Aavikko, M., Renkonen-Sinisalo, L., Järvinen, H., Böhm, J., Mecklin, J.P., Andersen, C.L., Vahteristo, P., Pitkänen, E., Aaltonen, L.A., 2015. 3'-UTR poly(T/U) repeat of EWSR1 is altered in microsatellite unstable colorectal cancer with nearly perfect sensitivity. *Fam. Cancer* 14, 449–453.
- Koo, B.K., Spit, M., Jordens, I., Low, T.Y., Stange, D.E., Van De Wetering, M., Van Es, J.H., Mohammed, S., Heck, A.J.R., Maurice, M.M., Clevers, H., 2012. Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors. *Nature* 488, 665–669.
- Koopman, M., Kortman, G.A.M., Mekenkamp, L., Ligtenberg, M.J.L., Hoogerbrugge, N., Antonini, N.F., Punt, C.J.A., Van Krieken, J.H.J.M., 2009. Deficient mismatch repair system in patients with sporadic advanced colorectal cancer. *Br. J. Cancer* 100, 266–273.
- Kothari, N., Kim, R., Jorissen, R.N., Desai, J., Tie, J., Wong, H.L., Farragher, I., Jones, I., Day, F.L., Li, S., Sakthinandeswaren, A., Palmieri, M., Lipton, L., Schell, M., Teer, J.K., Shibata, D., Yeatman, T., Sieber, O.M., Gibbs, P., Tran, B., 2015. Impact of regular aspirin use on overall and cancer-specific survival in patients with colorectal cancer harboring a PIK3CA mutation. *Acta Oncol. (Madr)*. 54, 487–492.
- Kozak, M., 1989. The scanning model for translation: An update. *J. Cell Biol.* 108, 229–241.
- Kurosaki, T., Popp, M.W., Maquat, L.E., 2019. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat. Rev. Mol. Cell Biol.* 20, 406–420.
- Kwiyeom Yoon, Lee, S., Han, T.-S., So Yeon Moon, Yun, S.M., Kong, S.-H., Sungwoong Jho, Jinny Choe, Jieun Yu⁴, Lee, H.-J., Ji Hyun Park, Hak-Min Kim, So Yeun Lee, Jongsun Park, Kim, W.-H., Bhak, J., Han-Kwang Yang, Kim, S.-J., 2013. Comprehensive genome- and transcriptome-wide analyses of mutations associated with microsatellite instability in Korean gastric cancers. *Genome Res.* 23, 1109–17.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, Christina, Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng,

- J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Hong, M.L., Dubois, J., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., De La Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G.R., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kasprzyk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F.A., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M. V., Kaul, R., Raymond, Christopher, Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J., 2001. Initial sequencing and analysis of the human genome. *Nature* 412, 565–566.
- Langlands, A.J., Almet, A.A., Appleton, P.L., Newton, I.P., Osborne, J.M., Näthke, I.S., 2016. Paneth Cell-Rich Regions Separated by a Cluster of Lgr5+ Cells Initiate Crypt Fission in the Intestinal Stem Cell Niche. *PLoS Biol.* 14, 1–31.
- Laurent-Puig, P., Blons, H., Cugenec, P.-H., 1999. Sequence of molecular genetic events in colorectal tumorigenesis. *Eur. J. Cancer Prev.*
- Le Borgne, M., Chartier, N., Billaud, M., 2013. L'oncogénomique: Chroniques d'un avènement (2): Instabilité chromosomique et cancer, enfin des CIN révèlateurs. *Medecine/Sciences* 29, 807–810.
- Le, D.T., Uram, J.N., Wang, H., Bartlett, B.R., Kemberling, H., Eyring, A.D., Skora, A.D., Azad, N.S., Laheru, D., Biedrzycki, B., Donehower, R.C., Zaheer, A., Fisher, G.A., Lee, J.J., Duffy, S.M., Goldberg, R.M., Koshiji, M., Bhajee, F., Hruban, R.H., Wood, L.D., Cuka, N., Pardoll, D.M., Papadopoulos, N., Kinzler, K.W., Cornish, T.C., Taube, J.M., Anders, R.A., Eshleman, J.R., Vogelstein, B., Diaz, L.A., 2015. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* 372, 2509–2520.
- Leach, D.R., Krummel, M.F., Allison, J.P., 1996. Enhancement of antitumor immunity by CTLA-4 blockade. *Science* (80-). 271, 1734–1736.
- LeBlanc, S.J., Gauer, J.W., Hao, P., Case, B.C., Hingorani, M.M., Weninger, K.R., Erie, D.A., 2018. Coordinated protein and DNA conformational changes govern mismatch repair initiation by MutS. *Nucleic Acids Res.* 46, 10782–10795.

- Lecomte, T., Tougeron, D., Chautard, R., Bressand, D., André, T., Bibeau, F., Blanc, B., Cohen, R., Jacques, J., Lagasse, J., Laurent-Puig, P., Lepage, C., Lucidarme, O., Martin-Babau, J., Panis, Y., Portales, F., Taïeb, J., Aparicio, T., Bouché, O., French Intergroup Clinical, 2021. Chapitre 3 Cancer du côlon non métastatique. *Thésaurus Natl. Cancérologie Dig.*
- Lee, Y., Rio, D.C., 2015. Mechanisms and regulation of alternative Pre-mRNA splicing. *Annu. Rev. Biochem.* 84, 291–323.
- Lengauer, C., Kinzler, K.W., Vogelstein, B., 1998. Genetic instabilities in human cancers. (cover story). *Nature* 396, 643. 7p. 3 Color Photographs.
- Leslie, A., Carey, F.A., Pratt, N.R., Steele, R.J.C., 2002. The colorectal adenoma carcinoma sequence. *Br. J. Surg.* 845–861.
- Levinson, G., Gutman, G.A., 1987. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* 4, 203–221.
- Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W., Li, G.M., 2013. The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutS α . *Cell* 153, 590–600.
- Li, F., Tian, L., Gu, L., Li, G.M., 2009. Evidence that nucleosomes inhibit mismatch repair in eukaryotic cells. *J. Biol. Chem.* 284, 33056–33061.
- Li, G.M., 2008. Mechanisms and functions of DNA mismatch repair. *Cell Res.* 18, 85–98.
- Li, S.K.H., Martin, A., 2016. Mismatch Repair and Colon Cancer: Mechanisms and Therapies Explored. *Trends Mol. Med.* 22, 274–289.
- Lietha, D., Cai, X., Ceccarelli, D.F.J., Li, Y., Schaller, M.D., Eck, M.J., 2007. Structural Basis for the Autoinhibition of Focal Adhesion Kinase. *Cell* 129, 1177–1187.
- Ligtenberg, M.J.L., Kuiper, R.P., Chan, T.L., Goossens, M., Hebeda, K.M., Voorendt, M., Lee, T.Y.H., Bodmer, D., Hoenselaar, E., Hendriks-Cornelissen, S.J.B., Tsui, W.Y., Kong, C.K., Brunner, H.G., Van Kessel, A.G., Yuen, S.T., Van Krieken, J.H.J.M., Leung, S.Y., Hoogerbrugge, N., 2009. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. *Nat. Genet.* 41, 112–117.
- Lindeboom, R.G.H., Supek, F., Lehner, B., 2016. The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat. Genet.* 48, 1112–1118.
- Lindor, B.N.M., Burgart, L.J., Leontovich, O., Goldberg, R.M., Cunningham, J.M., Sargent, D.J., Walsh-vockley, C., Petersen, G.M., Walsh, M.D., Leggett, B. a, Young, J.P., Barker, M. a, Jass, J.R., Hopper, J., Gallinger, S., Bapat, B., Redston, M., 2002. Testing in Phenotyping Colorectal Tumors. *Society* 20, 1043–1048.
- Litchfield, K., Reading, J.L., Lim, E.L., Xu, H., Liu, P., Al-Bakir, M., Wong, Y.N.S., Rowan, A., Funt, S.A., Merghoub, T., Perkins, D., Lauss, M., Svane, I.M., Jönsson, G., Herrero, J., Larkin, J., Quezada, S.A., Hellmann, M.D., Turajlic, S., Swanton, C., 2020. Escape from nonsense-mediated decay associates with anti-tumor immunogenicity. *Nat. Commun.* 11, 1–11.
- Llosa, N.J., Cruise, M., Tam, A., Wicks, E.C., Hechenbleikner, E.M., Taube, J.M., Blosser, R.L.,

- Fan, H., Wang, H., Lubber, B.S., Zhang, M., Papadopoulos, N., Kinzler, K.W., Vogelstein, B., Sears, C.L., Anders, R.A., Pardoll, D.M., Housseau, F., 2015. The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints. *Cancer Discov.* 5, 43–51.
- Loeffler, M., Bratke, T., Paulus, U., Li, Y.Q., Potten, C.S., 1997. Clonality and life cycles of intestinal crypts explained by a state dependent stochastic model of epithelial stem cell organization. *J. Theor. Biol.* 186, 41–54.
- Longley, D.B., Harkin, D.P., Johnston, P.G., 2003. 5-Fluorouracil: Mechanisms of action and clinical strategies. *Nat. Rev. Cancer* 3, 330–338.
- Longnecker, M.P., 1990. A case-control study of alcoholic beverage consumption in relation to risk of cancer of the right colon and rectum in men. *Cancer Causes Control* 1, 5–14.
- Luchini, C., Bibeau, F., Ligtenberg, M.J.L., Singh, N., Nottegar, A., Bosse, T., Miller, R., Riaz, N., Douillard, J.Y., Andre, F., Scarpa, A., 2019. ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: A systematic review-based approach. *Ann. Oncol.* 30, 1232–1243.
- Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., Misteli, T., 2010. Regulation of alternative splicing by histone modifications. *Science* (80-.). 327, 996–1000.
- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J., 1997. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260.
- Lynch, H., Shaw, M., Magnuson, C., Larsen, A., Krush, A., 1966. Hereditary Factors in Cancer. *Arch Intern Med* 68:105.
- Lynch, H.T., Lynch, P.M., Lanspa, S.J., Snyder, C.L., Lynch, J.F., Boland, C.R., 2009. Review of the Lynch syndrome: History, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin. Genet.* 76, 1–18.
- Lynch, H.T., Snyder, C.L., Shaw, T.G., Heinen, C.D., Hitchins, M.P., 2015. Milestones of Lynch syndrome: 1895-2015. *Nat. Rev. Cancer* 15, 181–194.
- Malesci, A., Laghi, L., Bianchi, P., Delconte, G., Randolph, A., Torri, V., Carnaghi, C., Doci, R., Rosati, R., Montorsi, M., Roncalli, M., Gennari, L., Santoro, A., 2007. Reduced likelihood of metastases in patients with microsatellite-unstable colorectal cancer. *Clin. Cancer Res.* 13, 3831–3839.
- Maley, C.C., Galipeau, P.C., Li, X., Sanchez, C.A., Paulson, T.G., Reid, B.J., 2004. Selectively advantageous mutations and hitchhikers in neoplasms: p16 lesions are selected in Barrett's esophagus. *Cancer Res.* 64, 3414–3427.
- Malkhosyan, S.R., Yamamoto, H., Piao, Z., Perucho, M., 2000. Late onset and high incidence of colon cancer of the mutator phenotype with hypermethylated hMLH1 gene in women. *Gastroenterology* 118, A1128.
- Mandal, R., Samstein, R.M., Lee, K.W., Havel, J.J., Wang, H., Krishna, C., Sabio, E.Y., Makarov,

- V., Kuo, F., Blecua, P., Ramaswamy, A.T., Durham, J.N., Bartlett, B., Ma, X., Srivastava, R., Middha, S., Zehir, A., Hechtman, J.F., Morris, L.G., Weinhold, N., Riaz, N., Le, D.T., Diaz, L.A., Chan, T.A., 2019. Genetic diversity of tumors with mismatch repair deficiency influences anti-PD-1 immunotherapy response. *Science* (80-.). 364, 485–491.
- Mannino, M.H., Zhu, Z., Xiao, H., Bai, Q., Wakefield, M.R., Fang, Y., 2015. The paradoxical role of IL-10 in immunity and cancer. *Cancer Lett.* 367, 103–107.
- Marabelle, A., Fakih, M., Lopez, J., Shah, M., Shapira-Frommer, R., Nakagawa, K., Chung, H.C., Kindler, H.L., Lopez-Martin, J.A., Miller, W.H., Italiano, A., Kao, S., Piha-Paul, S.A., Delord, J.P., McWilliams, R.R., Fabrizio, D.A., Aurora-Garg, D., Xu, L., Jin, F., Norwood, K., Bang, Y.J., 2020. Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study. *Lancet Oncol.* 21, 1353–1365.
- Marisa, L., de Reyniès, A., Duval, A., Selves, J., Gaub, M.P., Vescovo, L., Etienne-Grimaldi, M.C., Schiappa, R., Guenot, D., Ayadi, M., Kirzin, S., Chazal, M., Fléjou, J.F., Benchimol, D., Berger, A., Lagarde, A., Pencreach, E., Piard, F., Elias, D., Parc, Y., Olschwang, S., Milano, G., Laurent-Puig, P., Boige, V., 2013. Gene Expression Classification of Colon Cancer into Molecular Subtypes: Characterization, Validation, and Prognostic Value. *PLoS Med.* 10.
- Marisa, L., Svrcek, M., Collura, A., Becht, E., Cervera, P., Wanherdrick, K., Buhard, O., Goloudina, A., Jonchère, V., Selves, J., Milano, G., Guenot, D., Cohen, R., Colas, C., Laurent-Puig, P., Olschwang, S., Lefèvre, J.H., Parc, Y., Ghiringhelli, F., De Reynies, A., Duval, A., 2018. The balance between cytotoxic T-cell lymphocytes and immune checkpoint expression in the prognosis of colon tumors. *J. Natl. Cancer Inst.* 110, 68–77.
- Markowitz, S.D., Bertagnolli, M.M., 2009. Molecular Basis of Colorectal Cancer. *N. Engl. J. Med.* 361, 2449–2460.
- Marra, G., Iaccarino, I., Lettieri, T., Roscilli, G., Delmastro, P., Jiricny, J., 1998. Mismatch repair deficiency associated with overexpression of the MSH3 gene. *Proc. Natl. Acad. Sci. U. S. A.* 95, 8568–8573.
- Marrack, P., Kappler, J., 1987. The T Cell Receptor PHILIPPA. *Science* (80-.). 10, 1270.
- Mathieu, E.-L., Belhocine, M., Dao, L.T.M., Puthier, D., Spicuglia, S., 2014. [Functions of lncRNA in development and diseases]. *Médecine Sci. M/S* 30, 790–6.
- Mazurek, A., Berardini, M., Fishel, R., 2002. Activation of human MutS homologs by 8-oxo-guanine DNA damage. *J. Biol. Chem.* 277, 8260–8266.
- McCarthy, A.J., Capo-Chichi, J.M., Spence, T., Grenier, S., Stockley, T., Kamel-Reid, S., Serra, S., Sabatini, P., Chetty, R., 2019. Heterogenous loss of mismatch repair (MMR) protein expression: a challenge for immunohistochemical interpretation and microsatellite instability (MSI) evaluation. *J. Pathol. Clin. Res.* 5, 115–129.
- Messier, W., Li, S.-H., Stewart, C.-B., 1996. The birth of microsatellites. *Nature* 381, 1996.
- Meyers, M., Wagner, M.W., Hwang, H.S., Kinsella, T.J., Boothman, D.A., 2001. Role of the hMLH1 DNA mismatch repair protein in fluoropyrimidine-mediated cell death and cell

- cycle responses. *Cancer Res.* 61, 5193–5201.
- Mhaidly, R., Mehta-Grigoriou, F., 2020. Fibroblast heterogeneity in tumor micro-environment: Role in immunosuppression and new therapies. *Semin. Immunol.* 48, 101417.
- Michalak, M., Katzenmaier, E.M., Roeckel, N., Woerner, S.M., Fuchs, V., Warnken, U., Yuan, Y.P., Bork, P., Neu-Yilik, G., Kulozik, A., Doeberitz, M. von K., Kloor, M., Kopitz, J., Gebert, J., 2020. (Phospho)proteomic profiling of microsatellite unstable crc cells reveals alterations in nuclear signaling and cholesterol metabolism caused by frameshift mutation of nmd regulator upf3a. *Int. J. Mol. Sci.* 21, 1–22.
- Middha, S., Yaeger, R., Shia, J., Stadler, Z.K., King, S., Guercio, S., Paroder, V., Bates, D.D.B., Rana, S., Diaz, L.A., Saltz, L., Segal, N., Ladanyi, M., Zehir, A., Hechtman, J.F., 2019. Majority of B2M -Mutant and -Deficient Colorectal Carcinomas Achieve Clinical Benefit From Immune Checkpoint Inhibitor Therapy and Are Microsatellite Instability-High . *JCO Precis. Oncol.* 1–14.
- Middha, S., Zhang, L., Nafa, K., Jayakumaran, G., Wong, D., Kim, H.R., Sadowska, J., Berger, M.F., Delair, D.F., Shia, J., Stadler, Z., Klimstra, D.S., Ladanyi, M., Zehir, A., Hechtman, J.F., 2017. Reliable Pan-Cancer Microsatellite Instability Assessment by Using Targeted Next-Generation Sequencing Data. *JCO Precis. Oncol.* 1–17.
- Mirabelli-Primdahl, L., Gryfe, R., Kim, H., Millar, A., Luceri, C., Dale, D., Holowaty, E., Bapat, B., Gallinger, S., Redston, M., 1999. β -Catenin mutations are specific for colorectal carcinomas with microsatellite instability but occur in endometrial carcinomas irrespective of mutator pathway. *Cancer Res.* 59, 3346–3351.
- Mlecnik, B., Bindea, G., Angell, H.K., Maby, P., Angelova, M., Tougeron, D., Church, S.E., Lafontaine, L., Fischer, M., Fredriksen, T., Sasso, M., Bilocq, A.M., Kirilovsky, A., Obenauf, A.C., Hamieh, M., Berger, A., Bruneval, P., Tuech, J.J., Sabourin, J.C., Le Pessot, F., Mauillon, J., Rafii, A., Laurent-Puig, P., Speicher, M.R., Trajanoski, Z., Michel, P., Sesboüe, R., Frebourg, T., Pagès, F., Valge-Archer, V., Latouche, J.B., Galon, J., 2016. Integrative Analyses of Colorectal Cancer Show Immunoscore Is a Stronger Predictor of Patient Survival Than Microsatellite Instability. *Immunity* 44, 698–711.
- Monteran, L., Erez, N., 2019. The dark side of fibroblasts: Cancer-associated fibroblasts as mediators of immunosuppression in the tumor microenvironment. *Front. Immunol.* 10, 1–15.
- Morán, A., Iniesta, P., De Juan, C., González-Quevedo, R., Sánchez-Pernaute, A., Díaz-Rubio, E., Ramón y Cajal, S., Torres, A., Balibrea, J.L., Benito, M., 2002. Stromelysin-1 promoter mutations impair gelatinase B activation in high microsatellite instability sporadic colorectal tumors. *Cancer Res.* 62, 3855–3860.
- Morandi, L., De Biase, D., Visani, M., Monzoni, A., Tosi, A., Brulatti, M., Turchetti, D., Baccarini, P., Tallini, G., Pession, A., 2012. T[20] repeat in the 3'-untranslated region of the MT1X gene: A marker with high sensitivity and specificity to detect microsatellite instability in colorectal cancer. *Int. J. Colorectal Dis.* 27, 647–656.
- Morris, D.R., Geballe, A.P., 2000. Upstream Open Reading Frames as Regulators of mRNA Translation. *Mol. Cell. Biol.* 20, 8635–8642.

- Morton, D., Seymour, M., Magill, L., Handley, K., Glasbey, J., Glimelius, B., Palmer, A., Seligmann, J., Laurberg, S., Murakami, K., West, N., Quirke, P., Gray, R., 2023. Preoperative Chemotherapy for Operable Colon Cancer: Mature Results of an International Randomized Controlled Trial. *J. Clin. Oncol.* 41, 1541–1552.
- Muleris, M., Chalastanis, A., Meyer, N., Lee, M., Dutrillaux, B., Sastre-Garau, X., Hamelin, R., Fléjou, J.F., Duval, A., 2008. Chromosomal instability in near-diploid colorectal cancer: A link between numbers and structure. *PLoS One* 3.
- Nicolaides, N.C., Papadopoulos, N., Liu, B., Weit, Y.F., Carter, K.C., Ruben, S.M., Rosen, C.A., Haseltine, W.A., Fleischmann, R.D., Fraser, C.M., Adams, M.D., Venter, J.C., Dunlop, M.G., Hamilton, S.R., Petersen, G.M., De La Chapelle, A., Vogelstein, B., Kinzler, K.W., 1994. Mutations of two P/WS homologues in hereditary nonpolyposis colon cancer. *Nature* 371, 75–80.
- Nicolas, P., Kim, K.M., Shibata, D., Tavaré, S., 2007. The stem cell population of the human colon crypt: Analysis via methylation patterns. *PLoS Comput. Biol.* 3, 0364–0374.
- Nikanjam, M., Kato, S., Kurzrock, R., 2022. Liquid biopsy: current technology and clinical applications. *J. Hematol. Oncol.* 15, 1–14.
- Niu, B., Ye, K., Zhang, Q., Lu, C., Xie, M., McLellan, M.D., Wendl, M.C., Ding, L., 2014. MSIsensor: Microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* 30, 1015–1016.
- Nojima, T., Rebelo, K., Gomes, T., Grosso, A.R., Proudfoot, N.J., Carmo-Fonseca, M., 2018. RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Mol. Cell* 72, 369–379.e4.
- Oft, M., 2014. IL-10: master switch from tumor-promoting inflammation to antitumor immunity. *Cancer Immunol. Res.* 2, 194–199.
- Ong, B.A., Vega, K.J., Houchen, C.W., 2014. Intestinal stem cells and the colorectal cancer microenvironment. *World J. Gastroenterol.* 20, 1898–1909.
- Orphanides, G., Reinberg, D., 2002. A unified theory of gene expression. *Cell* 108, 439–451.
- Pai, Rish K., Dudley, B., Karloski, E., Brand, R.E., O’Callaghan, N., Rosty, C., Buchanan, D.D., Jenkins, M.A., Thibodeau, S.N., French, A.J., Lindor, N.M., Pai, Reetesh K., 2018. DNA mismatch repair protein deficient non-neoplastic colonic crypts: a novel indicator of Lynch syndrome. *Mod. Pathol.* 31, 1608–1618.
- Pai, Rish K., Pai, Reetesh K., 2016. A practical approach to the evaluation of gastrointestinal tract carcinomas for lynch syndrome. *Am. J. Surg. Pathol.* 40, e17–e34.
- Palazzo, A.F., Gregory, T.R., 2014. The Case for Junk DNA. *PLoS Genet.* 10.
- Palles, C., Cazier, J.B., Howarth, K.M., Domingo, E., Jones, A.M., Broderick, P., Kemp, Z., Spain, S.L., Almeida, E.G., Salguero, I., Sherborne, A., Chubb, D., Carvajal-Carmona, L.G., Ma, Y., Kaur, K., Dobbins, S., Barclay, E., Gorman, M., Martin, L., Kovac, M.B., Humphray, S., Lucassen, A., Holmes, C.H., Bentley, D., Donnelly, P., Taylor, J., Petridis, C., Roylance, R., Sawyer, E.J., Kerr, D.J., Clark, S., Grimes, J., Kearsley, S.E., Thomas, H.J.W., McVean, G., Houlston, R.S., Tomlinson, I., Thomas, H.J.W., Maher, E., Evans, G.,

- Cummings, C., Stevens, M., Walker, L., Halliday, D., Armstrong, R., Paterson, J., Hodgson, S., Homfray, T., Side, L., Izatt, L., Donaldson, A., Tomkins, S., Morrison, P., Goodman, S., Brewer, C., Henderson, A., Davidson, R., Murday, V., Cook, J., Haites, N., Bishop, T., Sheridan, E., Green, A., Marks, C., Carpenter, S., Broughton, M., Greenhalge, L., Suri, M., Bell, J., Ratcliffe, P., Wilkie, A., Broxholme, J., Buck, D., Cornall, R., Gregory, L., Knight, J., Lunter, G., Kingsbury, Z., Grocock, R., Hatton, E., Holmes, C.C., Hughes, L., Humburg, P., Kanapin, A., Murray, L., Rimmer, A., 2013. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat. Genet.* 45, 136–143.
- Pansy, K., Uhl, B., Krstic, J., Szmyra, M., Fechter, K., Santiso, A., Thüming, L., Greinix, H., Kargl, J., Prochazka, K., Feichtinger, J., Deutsch, A.J., 2021. Immune regulatory processes of the tumor microenvironment under malignant conditions. *Int. J. Mol. Sci.* 22.
- Parsons, M.T., Buchanan, D.D., Thompson, B., Young, J.P., Spurdle, A.B., 2012. Correlation of tumour BRAF mutations and MLH1 methylation with germline mismatch repair (MMR) gene mutation status: A literature review assessing utility of tumour features for MMR variant classification. *J. Med. Genet.* 49, 151–157.
- Pavlov, Y., Mian, I., Kunkel, T., 2003. EF-Tu Binding Peptides Identified, Dissected, and Affinity Optimized by Phage Display GDP conformation of EF-Tu. Because EF-Tu is abundant in the cell, other functions have been suggested, particularly during periods of cellular stress when protein syn-g. *Chem. Biol.* 10, 161–168.
- Pearlman, R., Haraldsdottir, S., De La Chapelle, A., Jonasson, J.G., Liyanarachchi, S., Frankel, W.L., Rafnar, T., Stefansson, K., Pritchard, C.C., Hampel, H., 2019. Clinical characteristics of patients with colorectal cancer with double somatic mismatch repair mutations compared with Lynch syndrome. *J. Med. Genet.* 56, 462–470.
- Peltanova, B., Raudenska, M., Masarik, M., 2019. Effect of tumor microenvironment on pathogenesis of the head and neck squamous cell carcinoma: A systematic review. *Mol. Cancer* 18, 1–24.
- Peltomäki, L., Sistonen, P., Mecklin, J.-P., Pylkkanen, L., Aaltonen, L., Nordling, S., Kere, J., Järvinen, H., Hamilton, S.R., Gloria Petersen, K.W.K., Vogelstein, B., Chapelle, A. de la, 1992. Evidence That the MCC-APC Gene Region in 5q21 Is Not the Site for Susceptibility to Hereditary Nonpolyposis Colorectal Carcinoma. *Cancer Res.* 52, 4530–4533.
- Peltomäki, P., 2005. Lynch syndrome genes. *Fam. Cancer* 4, 227–232.
- Peltomäki, P., 2014. Epigenetic mechanisms in the pathogenesis of Lynch syndrome. *Clin. Genet.* 85, 403–412.
- Perakis, S., Auer, M., Belic, J., Heitzer, E., 2017. *Advances in Circulating Tumor DNA Analysis*, 1st ed, *Advances in Clinical Chemistry*. Elsevier Inc.
- Pinto, D., Clevers, H., 2005. Wnt, stem cells and cancer in the intestine. *Biol. Cell* 97, 185–196.
- Pinto, D., Pinto, C., Guerra, J., Pinheiro, M., Santos, R., Vedeld, H.M., Yohannes, Z., Peixoto, A., Santos, C., Pinto, P., Lopes, P., Lothe, R., Lind, G.E., Henrique, R., Teixeira, M.R., 2018. Contribution of MLH1 constitutional methylation for Lynch syndrome diagnosis in

- patients with tumor MLH1 downregulation. *Cancer Med.* 7, 433–444.
- Pluciennik, A., Modrich, P., 2007. Protein roadblocks and helix discontinuities are barriers to the initiation of mismatch repair. *Proc. Natl. Acad. Sci. U. S. A.* 104, 12709–12713.
- Popat, S., Hubner, R., Houlston, R.S., 2005. Systematic review of microsatellite instability and colorectal cancer prognosis. *J. Clin. Oncol.* 23, 609–618.
- Poulogiannis, G., Frayling, I.M., Arends, M.J., 2010. DNA mismatch repair deficiency in sporadic colorectal cancer and Lynch syndrome. *Histopathology* 56, 167–179.
- Powell, S.M., Zilz, N., Beazer-Barclay, Y., Bryan, T.M., Hamilton, S.R., Thibodeau, S.N., Vogelstein, B., Kinzler, K.W., 1992. APC mutations occur early during colorectal tumorigenesis. *Nature* 359, 235–237.
- Rampino, N., Yamamoto, H., Li, Y., Sawai, H., Reed, J.C., Perucho, M., 1997. Somatic Frameshift Mutations in the BAX Gene in Colon Cancers of the Microsatellite Mutator Phenotype. *Science (80-)*. 275, 967–969.
- Rana, T.M., 2007. Illuminating the silence: Understanding the structure and function of small RNAs. *Nat. Rev. Mol. Cell Biol.* 8, 23–36.
- Ransom, M., Dennehey, B.K., Tyler, J.K., 2010. Chaperoning Histones during DNA Replication and Repair. *Cell* 140, 183–195.
- Rapisuwon, S., Vietsch, E.E., Wellstein, A., 2016. Circulating biomarkers to monitor cancer progression and treatment. *Comput. Struct. Biotechnol. J.* 14, 211–222.
- Ratovomanana, T., Cohen, R., Svrcek, M., Renaud, F., Cervera, P., Siret, A., Letourneur, Q., Buhard, O., Bourgoïn, P., Guillermin, E., Dorard, C., Nicolle, R., Ayadi, M., Touat, M., Bielle, F., Sanson, M., Le Rouzic, P., Buisine, M.P., Piessen, G., Collura, A., Fléjou, J.F., de Reyniès, A., Coulet, F., Ghiringhelli, F., André, T., Jonchère, V., Duval, A., 2021. Performance of Next-Generation Sequencing for the Detection of Microsatellite Instability in Colorectal Cancer With Deficient DNA Mismatch Repair. *Gastroenterology* 161, 814-826.e7.
- Ratovomanana, T., Nicolle, R., Cohen, R., Diehl, A., Siret, A., Letourneur, Q., Buhard, O., Perrier, A., Guillermin, E., Coulet, F., Cervera, P., Benusiglio, P., Labrèche, K., Colle, R., Collura, A., Despras, E., Rouzic, P. Le, Renaud, F., Cros, J., Alentorn, A., Touat, M., Ayadi, M., Bourgoïn, P., Prunier, C., Tournigand, C., Fouchardièrre, C. De, Tougeron, D., Jonchère, V., Bennouna, J., Reynies, A. De, Fléjou, J., Svrcek, M., André, T., Duval, A., 2023. Prediction of Response to Immune Checkpoint Blockade in Patients with Metastatic Colorectal Cancer with Microsatellite Instability. *Ann. Oncol.*
- Reyes, G.X., Schmidt, T.T., Kolodner, R.D., Hombauer, H., 2015. New insights into the mechanism of DNA mismatch repair. *Chromosoma* 124, 443–462.
- Ribic, C.M., Sargent, D.J., Moore, M.J., Stephen, T.N., French, A.J., Goldberg, R.M., Hamilton, S.R., Laurent-Puig, P., Gryfe, R., Sheph, L.E., 2003. Tumor Microsatellite-Instability Status as a Predictor of Benefit from Fluorouracil-Based Adjuvant Chemotherapy for Colon Cancer. *N. Engl. J. Med.* 7, 46–48.
- Ricciardone, M.D., Tayfun, O., 1999. Advances in Brief Human MLH1 Deficiency Predisposes

- to Hematological Malignancy and. *Cancer Res.* 59, 290–293.
- Richman, S., 2015. Deficient mismatch repair: Read all about it (Review). *Int. J. Oncol.* 47, 1189–1202.
- Richmond, C.A., Shah, M.S., Deary, L.T., Trotier, D.C., Thomas, H., Ambruzs, D.M., Jiang, L., Whiles, B.B., Rickner, H.D., Montgomery, R.K., Tovaglieri, A., Carlone, D.L., Breault, D.T., 2015. Dormant Intestinal Stem Cells Are Regulated by PTEN and Nutritional Status. *Cell Rep.* 13, 2403–2411.
- Riddell, R.H., Goldman, H., Ransohoff, D.F., Appelman, H.D., Fenoglio, C.M., Haggii-r, R.C., Correa, P., Hamilton, S.R., Morson, B.C., Sommers, S.C., Yardley, J.H., 1967. Perspectives in Pathology Dysplasia in Inflammatory Bowel Disease. *Perspect. Pathol.* 931.
- Rossi, L., Le Frere-Belda, M.A., Laurent-Puig, P., Buecher, B., De Pauw, A., Stoppa-Lyonnet, D., Canlorbe, G., Caron, O., Borghese, B., Colas, C., Delhomelle, H., Chabbert-Buffet, N., Grandjouan, S., Lecuru, F., Bats, A.S., 2017. Clinicopathologic characteristics of endometrial cancer in lynch syndrome A French multicenter study. *Int. J. Gynecol. Cancer* 27, 953–960.
- Rossouw, J., Anderson, G., Prentice, R., LaCroix, A., Kooperberg, C., Stefanick, M., Jackson, R., Beresford, A., Howard, B., Johnson, K., Morley Kotchen, J., Ockene, J., 2002. Risk and Benefit of Estrogen Postmenopauza. *JAMA* 288, 321–333.
- Roth, A.D., Delorenzi, M., Tejpar, S., Yan, P., Klingbiel, D., Fiocca, R., D’Ario, G., Cisar, L., Labianca, R., Cunningham, D., Nordlinger, B., Bosman, F., Van Cutsem, E., 2012. Integrated analysis of molecular and clinical prognostic factors in stage II/III colon cancer. *J. Natl. Cancer Inst.* 104, 1635–1646.
- Ryan, E., Sheahan, K., Creavin, B., Mohan, H.M., Winter, D.C., 2017. The current value of determining the mismatch repair status of colorectal cancer: A rationale for routine testing. *Crit. Rev. Oncol. Hematol.* 116, 38–57.
- Sade-Feldman, M., Jiao, Y.J., Chen, J.H., Rooney, M.S., Barzily-Rokni, M., Eliane, J.P., Bjorgaard, S.L., Hammond, M.R., Vitzthum, H., Blackmon, S.M., Frederick, D.T., Hazar-Rethinam, M., Nades, B.A., Van Seventer, E.E., Shukla, S.A., Yizhak, K., Ray, J.P., Rosebrock, D., Livitz, D., Adalsteinsson, V., Getz, G., Duncan, L.M., Li, B., Corcoran, R.B., Lawrence, D.P., Stemmer-Rachamimov, A., Boland, G.M., Landau, D.A., Flaherty, K.T., Sullivan, R.J., Hacohen, N., 2017. Resistance to checkpoint blockade therapy through inactivation of antigen presentation. *Nat. Commun.* 8.
- Samowitz, W.S., Slattery, M.L., Sweeney, C., Herrick, J., Wolff, R.K., Albertsen, H., 2007. APC mutations and other genetic and epigenetic changes in colon cancer. *Mol. Cancer Res.* 5, 165–170.
- Samuels, Y., Ericson, K., 2006. Oncogenic PI3K and its role in cancer. *Curr. Opin. Oncol.* 18, 77–82.
- Sawaya, S., Bagshaw, A., Buschiazzo, E., Kumar, P., Chowdhury, S., Black, M.A., Gemmell, N., 2013. Microsatellite Tandem Repeats Are Abundant in Human Promoters and Are Associated with Regulatory Elements. *PLoS One* 8.
- Schmeler, K.M., Lynch, H.T., Chen, L., Munsell, M.F., Soliman, P.T., Clark, M.B., Daniels, M.S.,

- White, K.G., Boyd-Rogers, S.G., Conrad, P.G., Yang, K.Y., Mary, R.M., Sun, C.C., Slomovitz, B.M., Gershenson, D.M., Lu, K.H., 2006. Prophylactic surgery to reduce the risk of gynecologic cancers in the lynch syndrome: Commentary. *N. Engl. J. Med.* 61, 319–321.
- Schneider, R., A.Fürst, Möslein, G., 2012. Lynch Syndrome – Epidemiology, Clinical Features, Molecular Genetics, Screening, Therapy. *Z Gastroenterol* 217–225.
- Schreiber, R.D., Old, L.J., Smyth, M.J., 2011. Cancer immunoediting: Integrating immunity's roles in cancer suppression and promotion. *Science* (80-). 331, 1565–1570.
- Schwark, A.L., Stadler, Z.K., 2018. Should We Lower Our Threshold for Germline Genetic Assessment in Pancreatic Adenocarcinoma? *JCO Precis. Oncol.* 1–4.
- Schwitalle, Y., Kloor, M., Eiermann, S., Linnebacher, M., Kienle, P., Knaebel, H.P., Tariverdian, M., Benner, A., von Knebel Doeberitz, M., 2008. Immune Response Against Frameshift-Induced Neopeptides in HNPCC Patients and Healthy HNPCC Mutation Carriers. *Gastroenterology* 134, 988–997.
- Sehgal, R., Sheahan, K., O'Connell, P.R., Hanly, A.M., Martin, S.T., Winter, D.C., 2014. Lynch Syndrome: An updated review. *Genes (Basel)*. 5, 497–507.
- Sei, S., Ahadova, A., Keskin, D.B., Bohaumilitzky, L., Gebert, J., von Knebel Doeberitz, M., Lipkin, S.M., Kloor, M., 2023. Lynch syndrome cancer vaccines: A roadmap for the development of precision immunoprevention strategies. *Front. Oncol.* 13, 1–17.
- Seppälä, T.T., Böhm, J.P., Friman, M., Lahtinen, L., Väyrynen, V.M.J., Liipo, T.K.E., Ristimäki, A.P., Kairaluoma, M.V.J., Kellokumpu, I.H., Kuopio, T.H.I., Mecklin, J.P., 2015. Combination of microsatellite instability and BRAF mutation status for subtyping colorectal cancer. *Br. J. Cancer* 112, 1966–1975.
- Shia, J., 2008. Immunohistochemistry versus microsatellite instability testing for screening colorectal cancer patients at risk for hereditary nonpolyposis colorectal cancer syndrome: Part I. The utility of immunohistochemistry. *J. Mol. Diagnostics* 10, 293–300.
- Shia, J., Holck, S., Depetris, G., Greenson, J.K., Klimstra, D.S., 2013. Lynch syndrome-associated neoplasms: A discussion on histopathology and immunohistochemistry. *Fam. Cancer* 12, 241–260.
- Sibley, C.R., Emmett, W., Blazquez, L., Faro, A., Haberman, N., Briese, M., Trabzuni, D., Ryten, M., Weale, M.E., Hardy, J., Modic, M., Curk, T., Wilson, S.W., Plagnol, V., Ule, J., 2015. Recursive splicing in long vertebrate genes. *Nature* 521, 371–375.
- Silveira, A.B., Bidard, F.C., Kasperek, A., Melaabi, S., Tanguy, M.L., Rodrigues, M., Bataillon, G., Cabel, L., Buecher, B., Pierga, J.Y., Proudhon, C., Stern, M.H., 2020. High-Accuracy Determination of Microsatellite Instability Compatible with Liquid Biopsies. *Clin. Chem.* 66, 606–613.
- Sinicrope, F.A., Foster, N.R., Thibodeau, S.N., Marsoni, S., Monges, G., Labianca, R., Yothers, G., Allegra, C., Moore, M.J., Gallinger, S., Sargent, D.J., 2011. DNA mismatch repair status and colon cancer recurrence and survival in clinical trials of 5-fluorouracil-based adjuvant therapy. *J. Natl. Cancer Inst.* 103, 863–875.

- Sinicrope, F.A., Rego, R.L., Foster, N., Sargent, D.J., Windschitl, H.E., Burgart, L.J., Witzig, T.E., Thibodeau, S.N., 2006. Microsatellite instability accounts for tumor site-related differences in clinicopathologic variables and prognosis in human colon cancers. *Am. J. Gastroenterol.* 101, 2818–2825.
- Slattery, M.L., Potter, J.D., Curtin, K., Edwards, S., Ma, K.N., Anderson, K., Schaffer, D., Samowitz, W.S., 2001. Estrogens reduce and withdrawal of estrogens increase risk of microsatellite instability-positive colon cancer. *Cancer Res.* 61, 126–130.
- Snippert, H.J., van der Flier, L.G., Sato, T., van Es, J.H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A.M., van Rheenen, J., Simons, B.D., Clevers, H., 2010. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* 143, 134–144.
- Snover, D.C., 2019. Diagnostic and reporting issues of preneoplastic polyps of the large intestine with early carcinoma. *Ann. Diagn. Pathol.* 39, 1–14.
- Sourrouille, I., Coulet, F., Lefevre, J.H., Colas, C., Eyries, M., Svrcek, M., Bardier-Dupas, A., Parc, Y., Soubrier, F., 2013. Somatic mosaicism and double somatic hits can lead to MSI colorectal tumors. *Fam. Cancer* 12, 27–33.
- Staffa, L., Echterdiek, F., Nelius, N., Benner, A., Werft, W., Lahrmann, B., Grabe, N., Schneider, M., Tariverdian, M., Von Knebel Doeberitz, M., Bläker, H., Kloor, M., 2015. Mismatch repair-deficient crypt foci in lynch syndrome-molecular alterations and association with clinical parameters. *PLoS One*.
- Stigliano, V., Assisi, D., Cosimelli, M., Palmirotta, R., Giannarelli, D., Mottolese, M., Mete, L.S., Mancini, R., Casale, V., 2008. Survival of hereditary non-polyposis colorectal cancer patients compared with sporadic colorectal cancer patients. *J. Exp. Clin. Cancer Res.* 27, 1–5.
- Stryker, S.J., Carney, J.A., Dozois, R.R., 1987. Multiple adenomatous polyps arising in a continent reservoir ileostomy. *Int J Color. Dis* 43–45.
- Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., Schmidt, D., O’Keeffe, S., Haas, S., Vingron, M., Lehrach, H., Yaspo, M.L., 2008. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* (80-). 321, 956–960.
- Sulzmaier, F.J., Jean, C., Schlaepfer, D.D., 2014. FAK in cancer: Mechanistic findings and clinical applications. *Nat. Rev. Cancer* 14, 598–610.
- Suraweera, N., Iacopetta, B., Duval, A., Compoin, A., Tubacher, E., Hamelin, R., 2001. Conservation of mononucleotide repeats within 3’ and 5’ untranslated regions and their instability in MSI-H colorectal cancer. *Oncogene* 20, 7472–7477.
- Sveen, A., Kilpinen, S., Ruusulehto, A., Lothe, R.A., Skotheim, R.I., 2016. Aberrant RNA splicing in cancer; Expression changes and driver mutations of splicing factor genes. *Oncogene* 35, 2413–2427.
- Svrcek, M., El-Murr, N., Wanherdrick, K., Dumont, S., Beaugerie, L., Cosnes, J., Colombel, J.F., Tiret, E., Fléjou, J.F., Lesuffleur, T., Duval, A., 2013. Overexpression of microRNAs-155 and 21 targeting mismatch repair proteins in inflammatory bowel diseases.

Carcinogenesis 34, 828–834.

- Svrcek, M., Lascols, O., Cohen, R., Collura, A., Jonchère, V., Fléjou, J.F., Buhard, O., Duval, A., 2019. MSI/MMR-deficient tumor diagnosis: Which standard for screening and for diagnosis? Diagnostic modalities for the colon and other sites: Differences between tumors. *Bull. Cancer* 106, 119–128.
- Syngal, S., Brand, R.E., Church, J.M., Giardiello, F.M., Hampel, H.L., Burt, R.W., 2015. ACG clinical guideline: Genetic testing and management of hereditary gastrointestinal cancer syndromes. *Am. J. Gastroenterol.* 110, 223–262.
- Sznol, M., Chen, L., 2013. Antagonist antibodies to PD-1 and B7-H1 (PD-L1) in the treatment of advanced human cancer. *Clin. Cancer Res.* 19, 1021–1034.
- Tajima, A., Hess, M.T., Cabrera, B.L., Kolodner, R.D., Carethers, J.M., 2004. The mismatch repair complex hMutS α recognizes 5-fluorouracil-modified DNA: Implications for chemosensitivity and resistance. *Gastroenterology* 127, 1678–1684.
- Tan, K., Stupack, D.G., Wilkinson, M.F., 2022. Nonsense-mediated RNA decay: an emerging modulator of malignancy. *Nat. Rev. Cancer* 22, 437–451.
- Tautz, 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* 17, 1–23.
- Terui, R., Nagao, K., Kawasoe, Y., Taki, K., Higashi, T.L., Tanaka, S., Nakagawa, T., Obuse, C., Masukata, H., Takahashi, T.S., 2018. Nucleosomes around a mismatched base pair are excluded via an Msh2-dependent reaction with the aid of SNF2 family ATPase Smarcd1. *Genes Dev.* 32, 806–821.
- Théoleyre, O., Baklouti, F., 2004. Des oncogènes à l'interface entre transcription et épissage. *Médecine/Sciences* 20, 1014–1019.
- Thibodeau, S.N., Bren, G., Schaid, D., 1993. Microsatellite instability in cancer of the proximal colon. *Science* (80-). 260, 816–819.
- Thierry, A.R., El Messaoudi, S., Mollevi, C., Raoul, J.L., Guimbaud, R., Pezet, D., Artru, P., Assenat, E., Borg, C., Mathonnet, M., De La Fouchardière, C., Bouché, O., Gavoille, C., Fiess, C., Auzemery, B., Meddeb, R., Lopez-Crapez, E., Sanchez, C., Pastor, B., Ychou, M., 2017. Clinical utility of circulating DNA analysis for rapid detection of actionable mutations to select metastatic colorectal patients for anti-EGFR treatment. *Ann. Oncol.* 28, 2149–2159.
- Thompson, R.H., Gillett, M.D., Cheville, J.C., Lohse, C.M., Dong, H., Webster, W.S., Krejci, K.G., Lobo, J.R., Sengupta, S., Chen, L., Zincke, H., Blute, M.L., Strome, S.E., Leibovich, B.C., Kwon, E.D., 2004. Costimulatory B7-H1 in renal cell carcinoma patients: Indicator of tumor aggressiveness and potential therapeutic target. *Proc. Natl. Acad. Sci. U. S. A.* 101, 17174–17179.
- Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakraborty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., Guigó, R., 2012. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* 22, 1616–1625.

- Torlakovic, E., Skovlund, E., Snover, D.C., Torlakovic, G., Nesland, J.M., 2003. Morphologic reappraisal of serrated colorectal polyps. *Am. J. Surg. Pathol.* 27, 65–81.
- Truninger, K., Menigatti, M., Luz, J., Russell, A., Haider, R., Gebbers, J.O., Bannwart, F., Yurtsever, H., Neuweiler, J., Riehle, H.M., Cattaruzza, M.S., Heinemann, K., Schär, P., Jiricny, J., Marra, G., 2005. Immunohistochemical analysis reveals high frequency of PMS2 defects in colorectal cancer. *Gastroenterology* 128, 1160–1171.
- Tsai, Y.J., Huang, S.C., Lin, H.H., Lin, C.C., Lan, Y.T., Wang, H.S., Yang, S.H., Jiang, J.K., Chen, W.S., Lin, T. chen, Lin, J.K., Chang, S.C., 2018. Differences in gene mutations according to gender among patients with colorectal cancer. *World J. Surg. Oncol.* 16, 1–5.
- Turajlic, S., Litchfield, K., Xu, H., Rosenthal, R., McGranahan, N., Reading, J.L., Wong, Y.N.S., Rowan, A., Kanu, N., Al Bakir, M., Chambers, T., Salgado, R., Savas, P., Loi, S., Birkbak, N.J., Sansregret, L., Gore, M., Larkin, J., Quezada, S.A., Swanton, C., 2017. Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol.* 18, 1009–1021.
- Umar, A., Boland, C.R., Terdiman, J.P., Syngal, S., de la Chapelle, A., Rüschoff, J., Fishel, R., Lindor, N.M., Burgart, L.J., Hamelin, R., Hamilton, S.R., Hiatt, R.A., Jass, J., Lindblom, A., Lynch, H.T., Peltomäki, P., Ramsey, S.D., Rodriguez-Bigas, M.A., Vasen, H.F.A., Hawk, E.T., Barrett, J.C., Freedman, A.N., Srivastava, S., 2004. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J. Natl. Cancer Inst.* 96, 261–268.
- Urbanski, L.M., Leclair, N., Anczuków, O., 2018. Alternative-splicing defects in cancer: Splicing regulators and their downstream targets, guiding the way to novel cancer therapeutics. *Wiley Interdiscip. Rev. RNA* 9, 1–36.
- Valeri, N., Gasparini, P., Braconi, C., Paone, A., Lovat, F., Fabbri, M., Sumani, K.M., Alder, H., Amadori, D., Patel, T., Nuovo, G.J., Fishel, R., Croce, C.M., 2010. MicroRNA-21 induces resistance to 5-fluorouracil by down-regulating human DNA MutS homolog 2 (hMSH2). *Proc. Natl. Acad. Sci. U. S. A.* 107, 21098–21103.
- Van Cutsem, E., Cervantes, A., Adam, R., Sobrero, A., Van Krieken, J.H., Aderka, D., Aranda Aguilar, E., Bardelli, A., Benson, A., Bodoky, G., Ciardiello, F., D’Hoore, A., Diaz-Rubio, E., Douillard, J.Y., Ducreux, M., Falcone, A., Grothey, A., Gruenberger, T., Haustermans, K., Heinemann, V., Hoff, P., Köhne, C.H., Labianca, R., Laurent-Puig, P., Ma, B., Maughan, T., Muro, K., Normanno, N., österlund, P., Oyen, W.J.G., Papamichael, D., Pentheroudakis, G., Pfeiffer, P., Price, T.J., Punt, C., Ricke, J., Roth, A., Salazar, R., Scheithauer, W., Schmoll, H.J., Tabernero, J., Taïeb, J., Tejpar, S., Wasan, H., Yoshino, T., Zaanan, A., Arnold, D., 2016. ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann. Oncol.* 27, 1386–1422.
- van den Brink, G.R., Offerhaus, G.J., 2007. The Morphogenetic Code and Colon Cancer Development. *Cancer Cell* 11, 109–117.
- van Es, J.H., Jay, P., Gregorieff, A., van Gijn, M.E., Jonkheer, S., Hatzis, P., Thiele, A., van den Born, M., Begthel, H., Brabletz, T., Taketo, M.M., Clevers, H., 2005. Wnt signalling induces maturation of Paneth cells in intestinal crypts. *Nat. Cell Biol.* 7, 381–386.
- Vasen, H., Peltomäki, P., 2004. Mutations associated with HNPCC predisposition -- Update of

- ICG-HNPCC/INSiGHT mutation database. *Dis. Markers* 20, 269–76.
- Vasen, H.F.A., Mecklin, J.P., Meera Khan, P., Lynch, H.T., 1991. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). *Dis. Colon Rectum* 34, 424–425.
- Vasen, H.F.A., Watson, P., Mecklin, J.P., Lynch, H.T., 1999. New Clinical Criteria for Hereditary Nonpolyposis Colorectal Definition of HNPCC. *Gastroenterology* 116, 1453–1456.
- Veigl, M.L., Kasturi, L., Olechnowicz, J., Ma, A., Lutterbaugh, J.D., Periyasamy, S., Li, G.M., Drummond, J., Modrich, P.L., Sedwick, W.D., Markowitz, S.D., 1998. Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. *Proc. Natl. Acad. Sci. U. S. A.* 95, 8698–8702.
- Velho, S., Oliveira, C., Paredes, J., Sousa, S., Leite, M., Matos, P., Milanezi, F., Ribeiro, A.S., Mendes, N., Licastro, D., Karhu, A., Oliveira, M.J., Ligtenberg, M., Hamelin, R., Carneiro, F., Lindblom, A., Peltomaki, P., Castedo, S., Schwartz, S., Jordan, P., Aaltonen, L.A., Hofstra, R.M.W., Suriano, G., Stupka, E., Fialho, A.M., Seruca, R., 2010. Mixed lineage kinase 3 gene mutations in mismatch repair deficient gastrointestinal tumours. *Hum. Mol. Genet.* 19, 697–706.
- Venderbosch, S., Nagtegaal, I.D., Maughan, T.S., Smith, C.G., Cheadle, J.P., Fisher, D., Kaplan, R., Quirke, P., Seymour, M.T., Richman, S.D., Meijer, G.A., Ylstra, B., Heideman, D.A.M., De Haan, A.F.J., Punt, C.J.A., Koopman, M., 2014. Mismatch repair status and BRAF mutation status in metastatic colorectal cancer patients: A pooled analysis of the CAIRO, CAIRO2, COIN, and FOCUS studies. *Clin. Cancer Res.* 20, 5322–5330.
- Vooijs, M., Liu, Z., Kopan, R., 2011. Notch: Architect, landscaper, and guardian of the intestine. *Gastroenterology* 141, 448–459.
- Wakkach, A., Fournier, N., Rie Brun, V., Breitmayer, J.-P., Oise Cottrez, F., Groux, H., 2003. Characterization of Dendritic Cells that Induce Tolerance and T Regulatory 1 Cell Differentiation In Vivo a direct effect on T cells but involved a modification of APC function (Wakkach et al., 2001). Dendritic cells (DCs) are professional APCs, special-i. *Immunity* 18, 605–617.
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., Burge, C.B., 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476.
- Wang, Q., Lasset, C., Desseigne, F., Frappaz, D., Bergeron, C., Navarro, C., Ruano, E., Puisieux, A., 1999. Neurofibromatosis and early onset of cancers in hMLH1-deficient children. *Cancer Res.* 59, 294–297.
- Wang, Y., Qin, J., 2003. MSH2 and ATR form a signaling module and regulate two branches of the damage response to DNA methylation. *Proc. Natl. Acad. Sci. U. S. A.* 100, 15387–15392.
- Warthin, 1985. Hereditywith Referenceto Carcinoma As ShownBythe Studyof the Cases Examinedin the PathologicalLaboratory Of the Universityof Michigan. *CA-A Cancer J. Clin.* 1–23.
- Wasan, H.S., Park, H.S., Liu, K.C., Mandir, N.K., Winnett, A., Sasieni, P., Bodmer, W.F.,

- Goodlad, R.A., Wright, N.A., 1998. APC in the regulation of intestinal crypt fission. *J. Pathol.* 185, 246–255.
- Weeden, C.E., Hill, W., Lim, E.L., Grönroos, E., Swanton, C., 2023. Impact of risk factors on early cancer evolution. *Cell* 186, 1541–1563.
- Weisenberger, D.J., Siegmund, K.D., Campan, M., Young, J., Long, T.I., Faasse, M.A., Kang, G.H., Widschwendter, M., Weener, D., Buchanan, D., Koh, H., Simms, L., Barker, M., Leggett, B., Levine, J., Kim, M., French, A.J., Thibodeau, S.N., Jass, J., Haile, R., Laird, P.W., 2006. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat. Genet.* 38, 787–793.
- Wherry, E.J., Ha, S.J., Kaech, S.M., Haining, W.N., Sarkar, S., Kalia, V., Subramaniam, S., Blattman, J.N., Barber, D.L., Ahmed, R., 2007. Molecular Signature of CD8+ T Cell Exhaustion during Chronic Viral Infection. *Immunity* 27, 670–684.
- Whitehall, V.L.J., Rickman, C., Bond, C.E., Ramsnes, I., Greco, S.A., Umapathy, A., McKeone, D., Faleiro, R.J., Buttenshaw, R.L., Worthley, D.L., Nayler, S., Zhao, Z.Z., Montgomery, G.W., Mallitt, K.A., Jass, J.R., Matsubara, N., Notohara, K., Ishii, T., Leggett, B.A., 2012. Oncogenic PIK3CA mutations in colorectal cancers and polyps. *Int. J. Cancer* 131, 813–820.
- Wilder, J., Hollocher, H., 2001. Mobile elements and the genesis of microsatellites in Dipterans. *Mol. Biol. Evol.* 18, 384–392.
- Willis, J., Lefterova, M.I., Artyomenko, A., Kasi, P.M., Nakamura, Y., Mody, K., Catenacci, D.V.T., Fakih, M., Barbacioru, C., Zhao, J., Sikora, M., Fairclough, S.R., Lee, H., Kim, K.M., Kim, S.T., Kim, J., Gavino, D., Benavides, M., Peled, N., Nguyen, T., Cusnir, M., Eskander, R.N., Azzi, G., Yoshino, T., Banks, K.C., Raymond, V.M., Lanman, R.B., Chudova, D.I., Talasaz, A.A., Kopetz, S., Lee, J., Odegaard, J.I., 2019. Validation of microsatellite instability detection using a comprehensive plasma-based genotyping panel. *Clin. Cancer Res.* 25, 7035–7045.
- Win, A.K., Jenkins, M.A., Dowty, J.G., Antoniou, A.C., Lee, A., Giles, G.G., Buchanan, D.D., Clendenning, M., Rosty, C., Ahnen, D.J., Thibodeau, S.N., Casey, G., Gallinger, S., Le Marchand, L., Haile, R.W., Potter, J.D., Zheng, Y., Lindor, N.M., Newcomb, P.A., Hopper, J.L., MacInnis, R.J., 2017. Prevalence and penetrance of major genes and polygenes for colorectal cancer. *Cancer Epidemiol. Biomarkers Prev.* 26, 404–412.
- Woerner, S.M., Kloor, M., Mueller, A., Rueschoff, J., Friedrichs, N., Buettner, R., Buzello, M., Kienle, P., Knaebel, H.P., Kunstmann, E., Pagenstecher, C., Schackert, H.K., Möslein, G., Vogelsang, H., Von Knebel Doeberitz, M., Gebert, J.F., 2005. Microsatellite instability of selective target genes in HNPCC-associated colon adenomas. *Oncogene* 24, 2525–2535.
- Woerner, S.M., Yuan, Y.P., Benner, A., Korff, S., Von Knebel Doeberitz, M., Bork, P., 2010. SelTarbase, a database of human mononucleotide-microsatellite mutations and their potential impact to tumorigenesis and immunology. *Nucleic Acids Res.* 38, 682–689.
- Wong W-M, Mandir N, Goodlad RA, Wong BC, Garcia SB, Lam SK, W.N., 2002. Crypt Fission. *Gut* 50, 212–218.

- Wong, Y.F., Cheung, T.H., Lo, K.W.K., Yim, S.F., Chan, L.K.Y., Buhard, O., Duval, A., Chung, T.K.H., Hamelin, R., 2006. Detection of microsatellite instability in endometrial cancer: Advantages of a panel of five mononucleotide repeats over the National Cancer Institute panel of markers. *Carcinogenesis* 27, 951–955.
- Xiao, B., Luo, J., Xie, E., Kong, L., Tang, J., Liu, D., Mao, L., Sui, Q., Li, W., Hong, Z., Pan, Z., Jiang, W., Ding, P.R., 2020. Comparisons of screening strategies for identifying Lynch syndrome among patients with MLH1-deficient colorectal cancer. *Eur. J. Hum. Genet.*
- Xie, T., Spradling, A.C., Xie, T., Spradling, A.C., 2000. A niche maintaining germ line stem cells in the *Drosophila* ovary. *Science* (80-.). 290, 328–330.
- Xu-Welliver, M., Pegg, A.E., 2002. Degradation of the alkylated form of the DNA repair protein, O6-alkylguanine-DNA alkyltransferase. *Carcinogenesis* 23, 823–830.
- Yamamoto, H., Imai, K., 2015. Microsatellite instability: an update. *Arch. Toxicol.* 89, 899–921.
- Yamashita, R., Long, J., Longacre, T., Peng, L., Berry, G., Martin, B., Higgins, J., Rubin, D.L., Shen, J., 2021. Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. *Lancet Oncol.* 22, 132–141.
- Yang, K., Halima, A., Chan, T.A., 2023. Antigen presentation in cancer — mechanisms and clinical implications for immunotherapy. *Nat. Rev. Clin. Oncol.* 20, 604–623.
- Ychou, M., Hohenberger, W., Thezenas, S., Navarro, M., Maurel, J., Bokemeyer, C., Shacham-Shmueli, E., Rivera, F., Kwok-Keung Choi, C., Santoro, A., 2009. A randomized phase III study comparing adjuvant 5-fluorouracil/folinic acid with FOLFIRI in patients following complete resection of liver metastases from colorectal cancer. *Ann. Oncol.* 20, 1964–1970.
- Yeon Yeon, S., Jung, S.H., Jo, Y.S., Choi, E.J., Kim, M.S., Chung, Y.J., Lee, S.H., 2019. Immune checkpoint blockade resistance-related B2M hotspot mutations in microsatellite-unstable colorectal carcinoma. *Pathol. Res. Pract.* 215, 209–214.
- Yokoyama, Takanori, Takehara, K., Sugimoto, N., Kaneko, K., Fujimoto, E., Okazawa-Sakai, M., Okame, S., Shiroyama, Y., Yokoyama, Takashi, Teramoto, N., Ohsumi, S., Saito, S., Imai, K., Sugano, K., 2018. Lynch syndrome-associated endometrial carcinoma with MLH1 germline mutation and MLH1 promoter hypermethylation: A case report and literature review. *BMC Cancer* 18.
- You, J.F., Buhard, O., Ligtenberg, M.J.L., Kets, C.M., Niessen, R.C., Hofstra, R.M.W., Wagner, A., Dinjens, W.N.M., Colas, C., Lascols, O., Collura, A., Flejou, J.F., Duval, A., Hamelin, R., 2010. Tumours with loss of MSH6 expression are MSI-H when screened with a pentaplex of five mononucleotide repeats. *Br. J. Cancer* 103, 1840–1845.
- Yuan, L., Chi, Y., Chen, W., Chen, X., Wei, P., Sheng, W., Zhou, X., Shi, D., 2015. Immunohistochemistry and microsatellite instability analysis in molecular subtyping of colorectal carcinoma based on mismatch repair competency. *Int. J. Clin. Exp. Med.* 8, 20988–21000.
- Zhang, Yaguang, Zhang, Q., Zhang, Yang, Han, J., 2023. The Role of Histone Modification in DNA Replication-Coupled Nucleosome Assembly and Cancer. *Int. J. Mol. Sci.* 24.

- Zhu, L., Li, Z., Wang, Y., Zhang, C., Liu, Y., Qu, X., 2015. Microsatellite instability and survival in gastric cancer: A systematic review and meta-analysis. *Mol. Clin. Oncol.* 3, 699–705.
- Zhu, Y., Estrassmann, J.E., Queller, D.C., 2000. Insertions, substitutions, and the origin of microsatellites. *Genet. Res.* 76, 227–236.

Valorisation de la Thèse

Communications scientifiques

Orales :

Massive Early Perturbation of Gene Splicing in Colorectal Cancer with Microsatellite Instability

Montémont H*, Jonchère J*, Tubacher E, Siret A, Bourgoïn P, Letourneur Q, Labreche K, Renaud F, Svrcek M, Deleuze J-F, Collura A, Duval A

17ème Séminaire annuel du CRSA, Lieusaint, France, 24 - 25 novembre 2022

Massive Early Perturbation of Gene Splicing in Colorectal Cancer with Microsatellite Instability

Montémont H*, Jonchère J*, Tubacher E, Siret A, Bourgoïn P, Letourneur Q, Labreche K, Renaud F, Svrcek M, Deleuze J-F, Collura A, Duval A

22nd Doctoral School Days, Paris, France, 20th - 21st June 2023

Identification of Molecular Drivers of Microsatellite Instability Colorectal Tumorigenesis

Montémont H*, Jonchère J*, Siret A, Bourgoïn P, Letourneur Q, Labreche K, Renaud F, Svrcek M, Duval A, Collura A

Mini-Symposium "Oncology and therapies" course, Paris, France, 16th October 2023

Massive Cis-Perturbation of Alternative Splicing Due to Microsatellite Instability at U2AF-Binding Polypyrimidic Sites Early Impairs Cell Differentiation and Drives Colorectal Cancer Initiation

Montémont H*, Jonchère J*, Tubacher E, Siret A, Bourgoïn P, Letourneur Q, Labreche K, Fall A, Renaud F, Svrcek M, Alentorn A, Deleuze J-F, Collura A, Duval A

1st National Congress on Tumors with Microsatellite Instability (SIRIC-CURAMUS), Paris, France, November 30th - December 1st 2023

Posters :

CrILES (Crypts Immunohistochemistry Laser Microdissection Extraction & Sequencing): A new technology to study early development of Lynch syndrome (inherited MSI CANCER), from pre-neoplastic healthy tissues.

Montémont H, Bourgoïn P, Siret A, Letourneur Q, Svrcek M, Duval A, Collura A

21st Doctoral School Days, Paris, France, 27th - 28th May 2021

Identification du statut mutationnel et transcriptionnel dans la carcinogénèse MSI : de l'inactivation du système MMR au stade adénocarcinome (sporadique et associé au syndrome de Lynch).

Montémont H, Bourgoïn P, Siret A, Letourneur Q, Renaud F, Svrcek M, Duval A, Collura A

16ème Séminaire annuel du CRSA, Lieusaint, France, 25 - 26 novembre 2021

Formations

Expérimentation animale, éthique et “bien-traitance” : formation des concepteurs de projets scientifiques utilisant des animaux Formation réglementaire

Sorbonne Université, Paris, France, 7 au 18 septembre 2020 (60 heures)

Introduction à l'histologie : exploration des tissus du corps humain

Université de Liège, en ligne (MOOC), 2020 (24 heures)

NGS & Cancer

Cancéropôle Ile-de-France, en ligne (MOOC), 2021 (25 heures)

BiG - Introduction to Bioinformatics for Medical Genetics

Université Paris Cité, en ligne (MOOC), 2022 (15 heures)

Animation scientifique

Co-Fondateur du Club des Jeunes Chercheurs du CRSA (CJ2C) et coordinateur, au Centre de Recherche Saint-Antoine, de novembre 2020 à septembre 2023. Organisation d'évènements scientifiques et valorisation des jeunes chercheurs.

À la découverte du Centre de Recherche Saint-Antoine : Plongez au cœur de la recherche en physiologie au service de la médecine ! (Activité destinée à une classe de Lycée)

Stand : Présentation des travaux des jeunes chercheurs du Centre de Recherche Saint-Antoine (CRSA). (Activité destinée au grand public)

Centre de Recherche Saint-Antoine et Sorbonne Université, Fête de la Science, initiateur et coordinateur des activités, Paris, octobre 2022

Représentation des personnels doctorants et post-doctorants lors de la visite du PDG de l'Inserm (avril 2023) et dans le cadre de l'évaluation HCERES (octobre 2023).