Ecole doctorale Biologie-santé – Université de Lille

Laboratoire PRISM INSERM U1192

THESE DE DOCTORAT

En vue de l'obtention du grade de Docteur en Science de la vie et de la santé,
Spécialité Biochimie et Biologie moléculaire de l'Université de Lille

Présenté par

Kamel BACHIRI

Identification des interactions oncogéniques du Polyomavirus à cellules de Merkel

Soutenu à Villeneuve-d'Ascq, le 18 décembre 2023

Jury :

Président du jury :	Pr HDR Jérôme SOLASSOL *Université de Montpellier*	Professeur
Rapportrice :	Dr HDR Muriel LE ROMANCER-CHERIFI *Université de Lyon*	Directrice de recherche
Rapportrice :	Dr HDR Elodie MONSELLIER *Université de Strasbourg*	Chargée de recherche
Examinatrice :	Dr. Amélie BONNEFOND *Université de Lille*	Directrice de recherche
Membre invité :	Pr HDR Antoine TOUZE *Université de Tours*	Professeur
Directeur de thèse :	Pr HDR Michel SALZET *Université de Lille*	Professeur
Co-encadrant :	Dr. Etienne COYAUD *Université de Lille*	Chargé de recherche

Résumé en français

Le carcinome à cellules de Merkel (CCM) est un cancer de la peau agressif et de très mauvais pronostic. Ce cancer est lié à la présence du Polyomavirus à cellules de Merkel dans 80% des cas. Ce Polyomavirus exprime deux oncoprotéines virales : sT et LT tronquée. L'expression des deux antigènes T est suffisante à l'émergence du cancer et responsable de la perturbation de checkpoints de signalisation, de la modification du profil épigénétique et de l'évasion immunitaire. L'interactomique et la protéomique nous ont permis d'identifier de nombreux facteurs épigénétiques en relation avec les antigènes T. Nous avons aussi mis en lumière des mécanismes candidats touchant la régulation de la stabilité protéique, l'expression génique et la stabilité génétique *via* la voie de réponse aux dommages de l'ADN. Les hypothèses issues de ces analyses ont été testées de manière ciblée. Un lien entre la peptidyl-prolyl cis/trans isomérase et la neddylation a pu être mis en évidence. Nous avons ainsi identifié un mécanisme candidat de régulation de l'activité des complexes d'ubiquitinylation à travers le recrutement de PIN1 et l'isomérisation par un primo de neddylation. L'étude d'EHMT2, régulateur épigénétique d'intérêt, a révélé l'importance de son rôle dans la réponse aux dommages de l'ADN au sein des CCM virus positifs (VP). Nous avons identifié un rôle protecteur de tLT et d'EHMT2 des dommages de l'ADN au sein des CCM-VP. Un rôle de EHMT2 dans la réparation suivant un stress réplicatif est aussi apparu. Nous avons découvert de nouveaux mécanismes importants dans l'oncogénèse des CCM-VP, rapportant pour la première fois un rôle central de tLT dans la gestion des dommages à l'ADN.

Résumé en anglais

Merkel cell carcinoma (MCC) is an extremely aggressive skin cancer with a high mortality rate. In 80% of cases, this cancer is linked to the presence of the Merkel cell polyomavirus which expresses two viral oncoproteins, small T and a truncated form of large T. The expression of these proteins is sufficient for carcinogenesis, inducing the disruption of cell cycle checkpoints, changes in the epigenetic profile, and immune evasion. The combination of interactomics and proteomics approaches allowed us to identify numerous epigenetic factors related to the T antigens. These studies have also highlighted potential mechanisms involving the regulation of protein stability, gene expression, and genetic stability via an altered DNA damage response pathway. The hypotheses formulated following these analyses were investigated in targeted assays. A link between the peptidyl-prolyl cis/trans isomerase and neddylation was demonstrated. We have thus identified a new potential mechanism for regulating the activity of ubiquitination complexes involving the recruitment of PIN1 to these complexes, thus their isomerization, by neddylation. The study of the EHMT2 functions, an epigenetic regulator of interest, revealed the importance of its role in the DNA damage response in virus positive MCC (VP-MCC). We were able to identify a protective role of tLT and EHMT2 in DNA damages in VP-MCC cells. More specifically, we also report a role of EHMT2 in solving single strand DNA breaks following replication stress. Our works have enabled the discovery of new important mechanisms in VP-MCC oncogenesis, reporting for the first time a central role of tLT in DNA damages management.

Liste des abréviations

ADN : Acide Désoxyribonucléique

ARN : Acide Ribonucléique

ARNm : ARN messager

BirA* : BirA (R118G)

CCM : Carcinome à Cellules de Merkel

CCM-VN : Carcinome à Cellules de Merkel Virus Négatif

CCM-VP : Carcinome à Cellules de Merkel Virus Positif

CRL : Cullin Ring Ligase

C-ter : Extrémité Carboxy-terminale

DDR : Réponse aux dommages de l'ADN

DSB : Cassure double brin

GST : Gluthatione S Transférase

H3K9Me1-2 : Mono- Di- méthylation de la lysine 9 de l'histone H3

HPV : Papillomavirus humain

HU : Hydroxyurée

IRES : Internal Ribosome Entry Site

LFQ : Quantifiquation sans marquage

LSD : Domaine de stabilisation de LT

LT : Antigène grand T

LT275 : Forme du LT de 275 acides aminés

LT330 : Forme du LT de 330 acides aminés

LT428 : Forme du LT de 428 acides aminés

MCPyV : Polyomavirus à Cellules de Merkel

MS : Spectrométrie de masse

N-ter : Extrémité amino-terminale

OTM : Olive Tail Moment

PCR : réaction de polymérisation en chaine

PLA : Proximity Ligation Assay

SCF : Skp Cullin F-box

shRNA : short hairpin ARN

shTA : shRNA dirigé contre les antigènes T

SSB : Cassure simple brin

sT : Antigène petit T

TA : Antigènes T

tLT : Antigène grand T tronqué

TMA : Tissue Micro Array

UV : Rayons Ultra-violets

WT : Forme sauvage

XL : Cross-link

Remerciements

Je tiens à remercier le Pr. Michel Salzet, directeur de PRISM, de m'avoir accueilli au sein du laboratoire. Merci également pour ces séances de préparation au concours de l'école doctorale qui n'ont pas été plaisantes sur le moment mais m'ont aidées à prendre mon sujet en main.

Je tiens à remercier le Dr. Etienne Coyaud, chargé de recherche INSERM au laboratoire, qui m'a encadré durant cette thèse. Merci pour le temps que tu m'as consacré, ton aide et tes conseils. Merci également pour la liberté que j'ai pu avoir tout au long de cette thèse. Je tiens également à souligner la bonne ambiance et la bonne entente que nous avons eu au labo comme en dehors, je n'ai pas été traumatisé par ma thèse et c'est déjà un bon point !

Je remercie le Dr. Muriel Le Romancer-Cherifi, le Dr. Elodie Monsellier, le Dr. Amélie Bonnefond et le Pr. Jérôme Solassol d'avoir accepté de de faire partie de mon jury de thèse et de prendre le temps de lire et évaluer mes travaux.

Je remercie les personnes avec qui j'ai pu collaborer sur ce projet, notamment le Pr. Antoine Touzé, le Dr. Thibault Kerverrec, Pauline Gaboriaud, le Dr. Roland Houben, le Dr. David Schrama, le Dr. Laëtitia Trapp-Fragnet et les membres de mon comité de suivi de thèse le Dr. Caroline Demeret et le Dr. Malika Hamdane.

Merci à Lucas Roussel presque docteur au moment où j'écris ces lignes. Tu n'as jamais été un collègue, tu étais là avant la thèse et tu le seras après. Ces trois dernières années n'auraient pas été les mêmes sans toi. Merci pour le soutien moral et pour tous ces moments où il fallait lâcher un peu la pression et où on avait 8 ans. Si je garde un bon souvenir de ma thèse c'est aussi grâce à toi, après cinq ans au même endroit on prend (enfin) des chemins différents et j'espère que ça ne s'arrêtera pas là.

Thanks to the doctor-to-be Diego Fernando Garcia Del Rio. I thought about writing it in Spanish but it would have been ridiculous. Meeting you was definitely one of the highlights of my PhD, I'm glad you didn't end up in Ghent. We're not colleagues anymore but I think we can say we are friends. Now you're off to Germany but this is not the end, I hope you'll keep sending memes, funky music and share your old man wisdom like you did these years "*Semper tempore, divsersis locis".*

Merci au Dr. Estelle Laurent avec qui j'ai passé beaucoup de temps au laboratoire. Merci pour tout ce que tu m'as appris, pour la patience dont tu as pu faire preuve à mon égard et pour tes conseils et suggestions. J'ai bien compris à quelle point ta rigueur était nécessaire quand tu as quitté le laboratoire et je suis persuadé que tout se passera bien pour toi dans la voie que tu as choisie.

Merci au Dr. Diala Kantar avec qui j'ai pu travailler sur mon projet. C'était un plaisir de te rencontrer même si tu n'avais pas une opinion très flatteuse de moi au départ. Tu m'as beaucoup aidé, surtout sur cette fin de thèse et je t'en suis reconnaissant. Ta bonne humeur constante a contribué à rendre les choses plus légères au laboratoire.

Merci à la presque Dr. Léa Ledoux avec qui j'ai partagé ces trois années de thèse (et une année de licence mais nous avons tous les deux oublié donc ça ne compte pas). C'était un plaisir de te (re)rencontrer.

Merci au Dr. Maheul Ploton qui est arrivé durant la dernière année mais qui s'est fait un malin plaisir de me rappeler TOUS les matins que j'avais une thèse à écrire.

Merci au futur Dr. Lydia Ziane Chaouche et au Dr. Alice Capuz pour ces discussions scientifiques ou non.

Merci à Soulaimane Aboulouard, Tristan Cardon, Sümeyye Cavdarli, Paul Chaillou, Blandine Chazarin, Marie Duhamel, Isabelle Fournier, Julien Franck, Jean-Pascal Gimeno, Laurine Lagache, Christophe Lefebvre, Nina Ogrinc, Antonella Raffo Romero, Franck Rodet, Christelle Van Camp, Louise Vansteenkiste, Nathalie Vasseur, Maxence Wisztorski et Yanis Zirem qui ont partagé mon quotidien et avec qui j'ai pu partager des moments de travail ou de pause.

Merci à tous les membres du laboratoire avec qui j'ai toujours pu discuter et vers qui je pouvais me tourner en cas de problème.

Merci au Dr. Mélanie Rose, ça fait déjà un moment que tu n'es plus au labo mais ça ne change rien. Merci pour ces moments partagés, pour tes conseils et tes mises en garde sur la thèse dont j'aurai dû tenir compte.

Merci à Antonio, Blandine, Manon et William de m'avoir fait changer d'air et sortir de ma thèse. Ça y est c'est la fin vous allez pouvoir arrêter de me charrier sur le fait que je sois

étudiant. Petite mention spéciale à William qui voulait que je précise que ce sont ses « super » notes prises en médicine qui ont contribuées à ce que je fasse de la Bio.

Merci à mes frères et sœur, Sarah, Nahel et Lounis pour le soutient et la motivation. Vous réussissez tous dans ce que vous faites, il fallait bien que je finisse ma thèse pour ne pas être à la traîne. Ça y est, il est enfin temps pour moi de trouver un « vrai » métier.

Merci à mes parents qui m'ont soutenu et encouragé tout au long de ma vie. J'ai toujours pu suivre la voie que je voulais sans jamais rencontrer d'obstacles de votre part, bien au contraire. Merci pour votre présence, votre aide et votre soutien.

Enfin, merci à toi Romane, tu as été là du début à la fin et tu as été d'un soutien sans faille dans les moments plus difficiles. Tu ne le réalises peut-être pas mais c'est vraiment toi qui m'as poussé à faire cette thèse (d'ailleurs, je t'en veux beaucoup pour ça actuellement mais ça ira mieux quand ce sera fini). Merci pour ta bonne humeur qui m'a accompagné, merci de m'avoir poussé. Et merci à ta grand-mère qui s'est inquiétée pour toutes ces cellules innocentes que j'ai maltraitées durant ma thèse. Merci d'avoir pris le temps de relire cette thèse qui a dû te paraître d'un ennui terrible en tant qu'historienne. Merci à toi d'être là et merci pour tout ce que tu as fait pour moi.

« *Pour ce qui est de l'avenir, il ne s'agit pas de le prévoir mais de le rendre possible* »

Antoine de Saint-Exupéry

« Les mythes que nous tissons, même s'ils renferment des erreurs, reflètent inévitablement un fragment de la vraie lumière »

J. R. R. Tolkien

Table des matières

1 Introduction générale 15
1.1 Contexte 15
1.2 Objectifs 18
1.2.1 Construction de l'interactome des antigènes T 18
1.2.2 Identification d'interacteurs/interactions à potentiel oncogénique 19
1.2.3 Analyses ciblées de mécanismes potentiels 19
2 Etat de l'art 20
2.1 Carcinome à cellule de Merkel 20
2.1.1 Virus et cancer 20
2.1.1.1 Onco-virologie 20
2.1.1.2 Oncovirus 20
2.1.2 Epidémiologie 22
2.1.3 Description clinique et traitement 23
2.1.3.1 Clinique 23
2.1.3.2 Traitement 25
2.1.4 Origine cellulaire 26
2.1.4.1 Cellules de Merkel 27
2.1.4.2 Cellules pre/pro B 27
2.1.4.3 Fibroblastes dermaux 28
2.1.4.4 Cellules épithéliales 28
2.2 Polyomavirus à cellules de Merkel 29
2.2.1 Découverte 29
2.2.2 Structure du génome 30
2.2.2.1 Région précoce 30
2.2.2.2 Région tardive 32
2.2.3 Cycle du virus 34
2.2.4 Implication dans les MCC 35
2.2.4.1 Statistiques / épidémiologie 35
2.2.4.2 Intégration 36
2.3 Oncogenèse des carcinomes à cellules de Merkel 41
2.3.1 Carcinomes à cellules de Merkel virus positifs 41
2.3.1.1 LT 42
2.3.1.1.1 RB1 42
2.3.1.1.2 p53 43
2.3.1.1.3 SOX2 et ATOH1 44
2.3.1.1.4 miR-375 45
2.3.1.1.5 Vam6p 47
2.3.1.2 sT 48
2.3.1.2.1 PP2A/PP4C 48

2.3.1.2.2 4E-BP1 51
2.3.1.2.3 MycL-EP400 51
2.3.1.2.4 FBXW7 53
2.3.1.3 Effets du MCPyV sur la régulation de la chromatine 54
2.3.1.3.1 Méthylation des histones 54
2.3.2 Carcinomes à cellules de Merkel virus négatif, une autre entité pathologique ? 56
2.4 Interactomique protéique 59
2.4.1 Interactomique supervisée 59
2.4.1.1 PLA in situ 59
2.4.1.2 Y2H 62
2.4.1.3 FRET/BRET 64
2.4.2 Interactomique semi-supervisée 65
2.4.2.1 Purification par affinité couplée à la spectrométrie de masse 65
2.4.2.2 Virotrap 67
2.4.2.3 BioID 68
2.4.2.4 Split-BioID 71
2.4.2.5 APEX 72
2.4.3 Interactomique non supervisée, le Cross-Link-MS 73
3 Matériel et méthodes 76
3.1 Culture cellulaire 76
3.2 Clonage 76
3.3 Préparation des échantillons 78
3.3.1 BioID et TurboID 78
3.3.2 GST Pull-down 79
3.3.3 Protéome total 79
3.3.4 Préparation des échantillons pour la spectrométrie de masse 80
3.4 Analyse en nLC-MS/MS 80
3.5 PLA 82
3.5.1 Préparation des cellules 82
3.5.2 Réalisation de la PLA 83
3.6 Tissue Micro Array 83
3.6.1 Echantillons 83
3.6.2 Immuno histochimie 84
3.7 Western-Blot 84
3.8 Immunofluorescence 85
3.9 Test de prolifération 86
3.9.1 MTS 86
3.9.2 Mixed culture assay 86
3.10 Comet assay 86
4 Résultats 87

4.1 Délimitation des interactions fonctionnelles des antigènes T 87
4.1.1 Identification des interacteurs proximaux des antigènes T par BioID 87
4.1.1.1 sT 89
4.1.1.2 LT et tLT 90
4.1.1.2.1 LT275 90
4.1.1.2.2 LT330/428, modifications épigénétiques et réparation de l'ADN 90
4.1.1.2.3 LT pleine taille 94
4.1.2 Identification des interacteurs des antigènes T par GST-Pull Down 94
4.1.3 Influence des antigènes T sur le protéome 97
4.2 PIN1 comme nouveau régulateur des CRLs 102
4.2.1 Implication de PIN1 dans les CCM 102
4.2.1.1 Interaction tLT-PIN1 102
4.2.1.2 Expression de PIN1 dans les CCM 104
4.2.2 PIN1 en interaction avec des composants du système d'ubiquitinylation 106
4.2.3 Relation PIN1 et neddylation 108
4.2.4 Confirmation de la localisation de PIN1 au niveau des CRLs 112
4.3 Régulation des dommages de l'ADN par tLT et EHMT2 114
4.3.1 Relation entre CCM, DDR et EHMT2 114
4.3.2 Remodelage médié par EHMT2 et LT 116
4.3.3 Influence de tLT sur les interactions formées par EHMT2 119
4.3.4 Une coopération entre EHMT2 et tLT dans la régulation de la DDR 121
5 Conclusions/Discussion générales et perspectives 127
6 Bibliographie 136
7 Formations et productions scientifiques 155
7.1 Formations 155
7.2 Productions scientifique 155
7.2.1 Présentations orales 155
7.2.2 Présentations Poster 156
7.2.3 Publications 156
8 Annexes 157
ANNEXE I : Table d'interactions proximales antigéniques 158
ANNEXE II : Protéomes totaux WaGa +/- TAs 167
ANNEXE III : Table d'interactions de EHMT2 +/- tLT 172
Article 1 : Tarbouriech et al.,2022, Viruses (auteur associé) 178
Article 2 : Leiendecker et al., 2023, Cancer Discovery (auteur associé) 197
Article 3 : Durand *et al.,* 2023, JID (auteur associé) 212

Table des illustrations

Figure 1 : Aspect clinique du carcinome à cellules de Merkel 24
Figure 2 : Structure du génome du MCPyV et domaines fonctionnels des antigènes T. 33
Figure 3 : Processus de mutation, de linéarisation et d'insertion du MCPyV par NHEJ et MMEJ. 40
Figure 4 : Mécanismes oncogéniques associés au LT du MCPyV 43
Figure 5 : Mécanismes oncogéniques connus impliquant le sT du MCPyV 50
Figure 6 : Mécanismes oncogéniques au sein des CCM-CN 58
Figure 7 : Aperçu des techniques d'interactomique supervisées et semi-supervisées 60
Figure 8 : Représentation schématique du principe de la PLA in situ. 61
Figure 9 : Etude d'interaction protéique proximale par BioID 70
Figure 10 : Interacteurs proximaux du LT tronqué réparti selon les processus biologiques les impliquants. 93
Figure 11 : Analyse de l'expression différentielle du protéome des cellules WaGa (CCM-VP) suivant l'inhibition de l'expression des antigènes T. 98
Figure 12 : Validation de l'interaction tLT-PIN1 par Proximity Ligation Assay 103
Figure 13 : Comparaison des niveaux de PIN1-FBXW7 et c-Jun au sein de cellules CCM -VP et -VN. 105
Figure 14 : Quantification de l'expression de PIN1 par Immuno Histochimie sur Tissue Micro Array. 106
Figure 15 : Environnement moléculaire au site d'interaction de PIN1 et NEDD8 111
Figure 16 : Visualisation d'interaction PIN1-NEDD8 et PIN1-RBX1 par PLA. 113
Figure 17 : Validation de l'interaction tLT-EHMT2. 116
Figure 18 : Effet de l'inhibition de EHMT2 sur la prolifération de cellules CCM-VP en présence et absence des antigènes T. 117
Figure 19 : Analyse protéomique de cellules CCM-VP sous inhibition chimique de EHMT2 118
Figure 20 : Effet de l'expression de tLT sur les interactions formées par EHMT2 121
Figure 21 : Quantification des foyers de γH2Ax en relation avec l'inhibition de EHMT2 et des antigènes T. 123
Figure 22 : Analyse des marqueurs de la DDR et des dommages réels de l'ADN. 124
Figure 23 : Analyse de plusieurs marqueurs de la DDR 126
Figure 24 : Schéma récapitulatif des points importants des conclusions 135

1 INTRODUCTION GENERALE

1.1 CONTEXTE

A l'échelle mondiale, près d'un décès sur 6 est attribué aux cancers (OMS). Ces pathologies touchent plus de 18 millions de personnes chaque année. De nombreux progrès ont été réalisés dans la compréhension, la détection et la prise en charge des cancers. Néanmoins, beaucoup de zones d'ombres demeurent dans la compréhension des mécanismes transformants. Une proportion non négligeable de ces cancers, 15 à 20%, est associée à la présence de pathogènes parmi lesquels les oncovirus occupent une place importante (Parkin, 2006). Bien que l'étiologie virale de certains cancers ait été découverte au début du vingtième siècle, les mécanismes oncogéniques associés restent souvent mal compris, notamment à cause de la complexité des interactions protéines hôtes – protéines virales et du manque de modèles d'études. Les cancers d'origine virale surviennent en règle générale de nombreuses années après l'infection, suite à une longue période de latence. En effet, des altérations de la cellule hôte (par exemple mutations, senescence), du génome viral, et/ou un affaiblissement de l'immunité (traitement antirejet de greffe, HIV, vieillissement) peuvent créer un environnement favorable à l'émergence d'un cancer. Ces cancers sont liés à l'expression des propriétés oncogéniques des protéines virales, telles que l'activation de voies oncogéniques et/ou la répression de facteurs suppresseurs de tumeur. L'étude systématique des interactions entre les protéines hôtes et virales est nécessaire pour comprendre les mécanismes mis en jeu lors de l'oncogenèse viro-induite. Plus largement, les oncovirus sont des entités transformantes bien délimitées génétiquement et peuvent ainsi servir de modèles d'étude simples pour découvrir de nouveaux mécanismes néoplasiques. L'étude des virus a par le passé permis d'identifier plusieurs oncogènes majeurs dont les orthologues chez l'homme sont très fréquemment mutés dans les cancers, comme v-myc, v-raf, erB1 ou AKT (Tsuchida, Murugan and Grieco, 2016). L'exemple le plus probant est celui de HRAS et KRAS, d'abord découverts dans le virus du sarcome murin de Harvey et Kirsten et dont les formes humaines sont mutées dans près de 25% des cancers solides (Kodaz *et al.*, 2017).

Les génomes viraux codent souvent des systèmes protéiques simples, et pourtant suffisants pour contourner efficacement les défenses de la cellule hôte et assurer leur

pérennité. C'est le cas du polyomavirus à cellules de Merkel (MCPyV), qui est classé comme un agent carcinogène de type 2 par l'IARC (International Agency for Research on Cancer).

Le Carcinome à Cellules de Merkel (CCM) est une forme rare (~2500 cas par an en Europe) et très agressive de cancer de la peau touchant principalement les personnes âgées et les sujets immunodéprimés (Becker *et al.*, 2018). Diagnostiqué à stade équivalent, le pronostic pour un CCM est plus péjoratif que pour un mélanome. Lors d'un diagnostic au stade 3 ou supérieur, les chances de survie à 5 ans sont inférieures à 40% (Schadendorf *et al.*, 2017). A ce jour, les traitements présentant les résultats les plus encourageants reposent sur l'utilisation d'immunothérapies ciblant PD-1 ou PD-1L. Cependant, une réponse complète au traitement n'a été observée que dans 23% des cas et la plupart des patients rechutent après une courte phase de rémission (Walker *et al.*, 2020). En 2008, le CCM a été associé au MCPyV, où il est retrouvé dans 80% des cas (Feng *et al.*, 2008). Sur le plan génomique, les CCM virus positif (CCM-VP) présentent un taux mutationnel bien inférieur à celui des CCM virus négatif (CCM-VN) (0,40-0,75 mutations/Mb contre ~40 mutations/Mb) (Goh *et al.*, 2016). L'observation clinique du peu de mutations additionnelles nécessaires à l'émergence d'un cancer lorsque MCPyV est présent appuie fortement un rôle oncogénique de ce virus. Le génome du MCPyV code donc, selon toute vraisemblance, les entités centrales de l'oncogenèse.

Le MCPyV est un virus à ADN double brin de ~5.4kb codant 6 protéines. Ce génome est séparé en deux régions : une région précoce et une région tardive. La région précoce code le LT (*Large T Antigen*), le sT (*small T Antigen*), la protéine 57kT et ALTO (*Alternative T Open reading frame*), une protéine alternative prédite dont l'existence est sujette à caution. La région tardive code quant à elle les protéines de la capside VP1 et VP2. Les protéines de la capside et ALTO ne sont pas détectées dans les MCC-VP. Alors qu'il reste épisomal chez les sujets sains, MCPyV est systématiquement intégré de façon clonale au génome des cellules dans les CCM. Cette intégration s'accompagne de mutations de LT menant à l'expression d'une protéine tronquée. Ces mutations ont récemment été mises en relation avec l'action de cytidines déaminases APOBEC3 (Soikkeli *et al.*, 2022). Ces enzymes qui sont normalement impliquées dans la réponse anti-virale auraient ici un rôle carcinogène. La forme tronquée de LT (tLT) et sT sont exprimées de manière constitutive dans les cellules de MCC.

Il a été démontré qu'une inhibition de LT ou de sT entraîne une réduction de la croissance des MCCs-VP (Houben *et al.*, 2012; Shuda, Chang and Moore, 2014). Les interactions

des antigènes T avec des facteurs hôtes présentent donc un grand intérêt pour identifier les mécanismes transformants. Il a notamment été démontré que LT séquestrait Rb, ce qui conduit à l'inhibition de l'interaction Rb-E2F1 (Chellappan *et al.*, 1991; Houben *et al.*, 2012). E2F1 reste ainsi libre, ce qui entraine la progression du cycle cellulaire. sT stabilise LT et supposément d'autres protéines importantes dans l'oncogenèse, telles que c-Myc ou la Cycline E en ciblant une l'ubiquitine ligase FBXW7 (Kwun *et al.*, 2013). Cependant, cette information est controversée dans la mesure où une autre étude avance que les antigènes T ne peuvent pas directement se fixer directement à FBXW7 (Dye *et al.*, 2019).

Outre ces mécanismes, les modifications épigénétiques semblent occuper une place importante dans l'oncogenèse liée aux antigènes T. En effet l'étude des profils de méthylation de l'ADN a montré une dérépression transcriptionnelle générale dans les CCM-VP suggérant une dédifférenciation de ces cellules (Chteinberg *et al.*, 2020). L'accessibilité de l'ADN, via la modulation de son niveau de compaction, est impactée par l'action du virus. Certains de ces mécanismes épigénétiques ont été identifié, comme le complexe formé par sT-MYCL et EP400. Il est désormais connu que sT participe à la formation de ce complexe, amenant ainsi EP400 au niveau de certains promoteurs ce qui a pour conséquence l'activation transcriptionnelle de ces gènes via l'activité Histone acétyltransférase de EP400 (Cheng *et al.*, 2017). Une régulation via des répresseurs épigénétique a également été observée récemment, comme illustré par EZH2 dans une étude à laquelle nous avons contribué. EZH2 est une sous-unité du Polycomb repressive complex 2 dont l'inhibition cause une perte de croissance des cellules CCM-VP (Gartin *et al.*, 2022). Bien qu'importants dans l'oncogenèse, ces mécanismes s'inscrivent dans une dérégulation complexe et sont très probablement associés à d'autres dérégulations épigénétiques essentielles qui restent à découvrir.

L'étude systématique des interactions de sT et tLT représente donc un projet bien délimité et pertinent pour découvrir de nouveaux mécanismes oncogéniques « co-essentiels » associés au développement des CCM.

La première description des interactions protéines-protéines (IPPs) entre cellule hôte et protéome complet d'un pathogène a été réalisée en 2011 sur le VIH en utilisant la technique de purification par affinité couplée à la spectrométrie de masse (IP-MS) (Jäger *et al.*, 2012). Depuis, l'émergence de nouveaux outils d'interactomique, tel BioID pour l'étude des interactions proximales, a permis de créer une carte plus complète des interactions protéiques

entre hôte et pathogène, comme illustré par l'étude de l'interactome proximal du virus Zika (Roux *et al.*, 2012; Coyaud *et al.*, 2018).

En effet, BioID permet d'identifier les interactions protéiques dans un contexte plus physiologique, par marquage covalent à la biotine en cellule vivante. Cette propriété est essentielle car elle permet l'accès aux IPP situées dans les compartiments peu solubles (ex : membranes, ADN, cytosquelette) et aux interactions transitoires et/ou de faible affinité qui peuvent néanmoins avoir des fonctions biologiques majeures. Cette technique consiste à exprimer un mutant de la biotine ligase d'*E.Coli* (BirA R118G, ou BirA*) fusionné à la protéine d'intérêt. Cette fusion exprimée dans des cellules vivantes mime les interactions de la protéine d'intérêt et permet la biotinylation des protéines proximales dans un rayon de 10 à 20 nm (Kim *et al.*, 2014). Ce marquage covalent permet de s'affranchir de la nécessité de préserver les IPP durant le processus de purification. Des conditions de lyse agressives (RIPA ou SDS) peuvent être ainsi utilisées, permettant la solubilisation des protéines de l'ensemble des compartiments cellulaires. Enfin, les interacteurs proximaux sont récupérés sur colonne streptavidine et identifiés par spectrométrie de masse (Gupta *et al.*, 2015). Dans mon projet, l'utilisation de ce type d'approche permettra donc de découvrir de nouvelles interactions potentiellement oncogéniques du MCPyV afin de bien comprendre les mécanismes moléculaires des CCM. Ces travaux représentent également un angle d'exploration de l'oncogenèse de manière générale. En effet, il existe de nombreuses voies d'oncogenèse canoniques, partagées entre de nombreuses néoplasies de types différents, dont certaines restent à identifier. L'étude du MCPyV a ainsi le potentiel de fournir une meilleure compréhension de mécanismes transformants impliqués dans de multiples cancers.

1.2 OBJECTIFS

1.2.1 Construction de l'interactome des antigènes T

Le premier objectif de cette thèse est d'obtenir les interactomes proximaux du small T antigen sT, du Large T antigen LT ainsi que de 3 formes tronquées du LT : LT275, LT330 et LT428. Il s'agit de trois points de troncation récurrents chez les patients. Ces interactomes ont été obtenus par BioID au sein de cellules 293 Flp-In T-REx dans un premier temps. Ces cellules d'origine embryonnaire permettent une expression inductible d'une seule copie intégrée au

génome du transgène codant chaque protéine de fusion BirA*-T Antigen (TA), ce qui limite le niveau de surexpression de la protéine de fusion. De plus, dû à leur caractère embryonnaire, elles expriment un large panel de groupes protéiques, ce qui permet d'observer un maximum d'interacteurs possibles. Elles représentent un « gold standard » en protéomique et sont utilisées ici comme bioréacteur afin de caractériser le potentiel biochimique d'interaction de chaque protéine en cellule vivante. Ces données visent à être combinées à d'autres études interactomiques notamment par GST-Pulldown. Les interactions présentant le potentiel oncogénique le plus important sont confirmées par méthode orthogonale.

1.2.2 Identification d'interacteurs/interactions à potentiel oncogénique

Une fois les données d'interactomique obtenues, le but est d'identifier les interactions ayant un potentiel oncogénique et de relier l'effet de ces interactions à des modifications du protéome, qui déterminent *in fine* les altérations phénotypiques. Pour y parvenir, les données obtenues sont comparées à une étude de protéomique globale permettant d'identifier les changements du protéome induits par l'expression des TAs, rendant possible l'inférence d'un rôle probable à certaines IPP en se basant sur la comparaison entre ces variations d'abondance de protéines et les fonctions connues des interacteurs identifiés. La mise en relation de ces informations et des données disponibles dans la littérature concernant les CCM et les interacteurs identifiés a permis la formulation de plusieurs hypothèses testées de manière ciblée.

1.2.3 Analyses ciblées de mécanismes potentiels

Une fois les données d'interactomique acquises, croisées à celles de génomique et de protéomique, les hypothèses formulées sur les mécanismes oncogéniques sont testées par des analyses ciblées. Les mécanismes identifiés dans cette partie concernent l'oncogenèse des CCM mais peuvent également intervenir dans un cadre plus large soulignant ainsi le potentiel du MCPyV comme modèle d'étude.

2 ETAT DE L'ART

2.1 CARCINOME A CELLULE DE MERKEL

2.1.1 Virus et cancer

2.1.1.1 Onco-virologie

La naissance de l'onco-virologie survient lors de l'isolement d'un « agent filtrable » au sein de sarcomes aviaire par Peyton Rous en 1911 (Rous, 1911). Cette découverte représente le premier isolement d'un virus dans un cancer solide et fonde les bases de l'onco-virologie. Cette discipline ne s'est réellement développée que plus tard, quand une plus grande partie de la communauté scientifique s'y est intéressée dans les années 60, suite à la mise en évidence du lien entre le virus d'Epstein-Barr et le lymphome de Burkitt, premier lien avéré entre cancer humain et virus (Epstein, Achong and Barr, 1964). La forte prévalence de ce virus dans la population adulte a cependant rendu compliqué l'affirmation d'un lien de causalité entre ce virus et le cancer. Ce n'est qu'en 1978 que ce lien a pu être établi par une étude épidémiologique (de-Thé *et al.*, 1978). Ces conclusions ont été confirmées et appuyées par une plus grande étude épidémiologique menée par la même équipe (Geser *et al.*, 1982). Depuis, l'étude des oncovirus s'est développée et a permis l'identification de nombreux oncogènes majeurs dont les orthologues ont été retrouvés chez l'homme par la suite, comme v-myc, v-raf, erB1 ou AKT (Tsuchida, Murugan and Grieco, 2016). L'exemple le plus marquant est celui de HRAS et KRAS, d'abord découvert dans le virus du sarcome murin de Harvey et Kirsten, les formes humaines ont ensuite été identifiées comme étant mutées dans près de 25% des cancers solides (Kodaz *et al.*, 2017).

2.1.1.2 Oncovirus

Depuis la découverte des oncovirus, l'intérêt croissant pour cette problématique et les progrès technologiques ont permis l'identification de plusieurs oncovirus humains (**Tableau 1**). Ces virus touchent un large éventail de tissus et présentent des différences quant à leur nature. Certains mécanismes sont néanmoins communs à plusieurs virus. L'inactivation des voies de RB1 est relativement repandue parmi les oncovirus humains. La voie RB1 est une voie

de signalisation cellulaire qui régule le cycle cellulaire, la différenciation cellulaire et l'apoptose. Cette protéine agit comme suppresseur de tumeur notamment en séquestrant des facteurs de transcriptions comme les E2F. L'inactivation de cette voie par des onco-protéines se fait de différentes manières selon le cancer. Certains virus produisent des protéines virales capables de se lier à RB1 et de la séquestrer, l'empêchant ainsi de remplir sa fonction de suppresseur de tumeur. C'est notamment le cas de la protéine E6 des papillomavirus humains et du Large T antigen du carcinome à cellules de Merkel. Cette voie est également impactée par une diminution de l'expression de RB1 comme c'est le cas dans le carcinome hépatocellulaire causé par les virus de l'hépatite B et C. Cette diminution d'expression est causée par l'hypermethylation du promoteur de RB1 ce qui réduit sa transcription (Edamoto *et al.*, 2003; Mileo *et al.*, 2015). Cette hyperméthylation est causée par une prise de contrôle de machineries de régulation épigénétiques par les protéines virales. La modification des processus épigénétiques et de régulation de l'ADN, est un moyen efficace pour les protéines virales de modifier l'expression des gènes d'une cellule. En agissant ainsi sur le mécanisme source de modification du protéome, peu de protéines virales sont nécessaires pour modifier radicalement le phénotype d'une cellule. Dans ce processus de transformation, les mutations somatiques de proto-oncogènes ou suppresseurs de tumeurs ne sont même plus forcément nécessaires à l'acquisition d'un phénotype néoplasique. Tous les virus considérés oncovirus humains agissent sur l'épigénome de la cellule afin de reprogrammer le profil d'expression génique (Soliman, Orlacchio and Verginelli, 2021).

Le dernier oncovirus humain identifié est le polyomavirus à cellules de Merkel, découvert en 2008 (Feng *et al.*, 2008).

Le Carcinome à cellules de Merkel est une forme rare de cancer de la peau présentant des caractéristiques neuroendocrines. La première description de cette pathologie remonte à 1972, elle est alors décrite comme « trabecular carcinoma of the skin » (Toker, 1972). La dénomination « carcinome à cellules de Merkel » n'est apparue que dans les années 80 après des études histologiques plus approfondies des tumeurs et l'attribution d'une nouvelle origine cellulaire à ce cancer, les cellules de Merkel (De Wolf-Peeters *et al.*, 1980).

VIRUS	ANNEE	CANCERS	MECANISMES	REFERENCES
Virus d'epstein-barr (EBV)	1964	Lymphome de Burkitt, carcinome du rhinopharynx, lymphome Hodgkinien	Inhibition apoptose, instabilité génomique	(Epstein, Achong and Barr, 1964)
Papillomavirus humain (HPV)	1974	Cancers du col de l'utérus, ano-génitaux, ORL	Inactivation p53 et RB1, instabilité chromosomique	(zur Hausen *et al.*, 1974)
Virus t-lymphotropique humain (HTLV-1)	1980	Leucémie/lymphome T de l'adulte	Activation transcriptionnelle, instabilité génomique	(Poiesz *et al.*, 1980)
Virus de l'hepatite b (HBV)	1965	Carcinome hépatocellulaire	Insertion ADN viral, inhibition p53	(Blumberg and Alter, 1965)
Virus de l'hepatite c (HCV)	1989	Carcinome hépatocellulaire	Inflammation chronique, cirrhose	(Choo *et al.*, 1989)
Virus herpes humain 8 (HHV-8)	1994	Sarcome de Kaposi	Homologues gènes prolifération/apoptose	(Chang *et al.*, 1994)
Polyomavirus à cellules de Merkel (MCPyV)	2008	Carcinome à cellules de Merkel	Inactivation RB1, inhibition apoptose	(Feng *et al.*, 2008)

Tableau 1 : Inventaire des oncovirus humain connus

Récapitulatif des sept oncovirus humain connus à ce jour. Chaque virus est associé à son année de découverte ainsi qu'au(x) cancer(s) dans lequel/lesquels ce virus est retrouvé. Quelques mécanismes oncogéniques associés à ces virus sont également présentés.

2.1.2 Epidémiologie

En raison de la faible incidence de cette pathologie, l'épidémiologie des CCMs est mal connue et les données varient selon les études. Les différentes sources rapportent cependant une rapide augmentation de l'incidence de cette maladie au sein des populations. Dans l'union Européenne, il est estimé qu'environ 2500 CCM sont diagnostiqués chaque année (Becker *et al.*, 2018). Très peu de données spécifiques à la France sont disponibles. A l'international, les données disponibles montrent une tendance à la hausse de l'incidence des CCMs au cours du temps. Entre 1997 et 2016, l'incidence chez les hommes aux Etats-Unis est passée de 0,55 à 1,03 cas pour 100 000 habitants. Une incidence supérieure est observée dans les pays de l'hémisphère sud. En effet, dans le Queensland an Australie, une incidence de 2,99 cas pour 100 000 habitants est rapportée en 2016 (Olsen, Pandeya and Whiteman, 2021).

L'exposition au soleil représente un facteur de risque important qui est reflété par les différences d'incidence entre les hémisphères nord et sud. Les hommes semblent être plus touchés que les femmes avec des incidences généralement supérieures : 2,99 contre 0,90 pour 100 000 habitants dans le Queensland en 2016 (Olsen, Pandeya and Whiteman, 2021).

2.1.3 Description clinique et traitement

2.1.3.1 Clinique

Les CCM se présentent cliniquement comme des tumeurs cutanées solides, généralement de couleur rouge à violacée (Yiengpruksawan *et al.*, 1991). Ils peuvent apparaître n'importe où sur le corps, mais sont plus fréquemment observés sur les zones exposées au soleil, telles que le visage, le cou et les membres supérieurs. Les lésions de CCM sont généralement asymptomatiques et peuvent être confondues avec d'autres lésions cutanées bénignes, ce qui peut retarder le diagnostic (**Figure 1**).

Les CCM se caractérisent par leur croissance rapide et leur fort potentiel d'invasion local. Les tumeurs primaires peuvent être fermes au toucher et mesurer de quelques millimètres à plusieurs centimètres de diamètre. Elles ont souvent un aspect lobulaire, avec des contours irréguliers et des vaisseaux sanguins dilatés visibles à la surface. Outre leur aspect clinique, les CCM se caractérisent par leur agressivité et leur tendance à se propager rapidement. La présence de métastases est détectée dans une part importante des patients au moment du diagnostic, de 30 à 50% selon les sources (Scholz, Rippmann and Spilker, 2004; Allen *et al.*, 2005). Les sites de métastases les plus courants comprennent les ganglions lymphatiques régionaux dans environ 40% des cas, la peau dans 25% des cas, le foie dans 23% des cas, des métastases osseuses dans 21% des cas, et les organes richement vascularisés comme les poumons et le cerveau dans 7 et 5% des cas, respectivement (Lewis *et al.*, 2020).

Bien que les CCM puissent survenir chez des patients de tout âge, ils sont plus fréquents chez les personnes âgées, avec un âge médian au diagnostic de 75 à 80 ans. Les CCM sont plus fréquents chez les individus à peau claire et sont plus fréquents chez les hommes que chez les femmes. Les patients immunodéprimés, tels que les transplantés d'organes ou les personnes atteintes du virus de l'immunodéficience humaine (VIH), présentent un risque accru de développer un CCM. Dans un but de sensibilisation des praticiens à la pathologie, un acronyme caractérisant le CCM a été mis en place. Ainsi, le CCM est une pathologie AEIOU, **A**symptomatique à **E**xpansion rapide touchant principalement les patients **I**mmunodéprimés et/ou de plus de 50 ans (**O**ver 50) sur des zones exposées aux **U**V (Heath *et al.*, 2008).

L’avancement de la maladie est défini par une répartition en 4 stades. Le stade I correspond à une lésion primaire localisée mesurant moins de 2 cm de diamètre, sans

envahissement ganglionnaire régional ni métastase distante. Le stade II se caractérise par des lésions primaires de plus de 2 cm restant localisées, sans envahissement des ganglions lymphatiques ni propagation métastatique à distance. Le stade III marque l'invasion des ganglions lymphatiques drainant la zone de la tumeur primitive. On distingue deux sous-groupes : le stade IIIA avec envahissement ganglionnaire microscopique et le stade IIIB avec atteinte ganglionnaire macroscopique. Enfin, le stade IV le plus avancé correspond à la dissémination métastatique à distance, le plus souvent pulmonaire, hépatique, osseuse ou cérébrale. Le taux de survie à 5 ans varie grandement selon le stade au diagnostic. Ce taux dépasse les 50% pour les stades I et II localisés, est d'environ 40% au stade III, mais chute à moins de 15% en cas de métastases disséminées de stade IV (Schadendorf *et al.*, 2017).

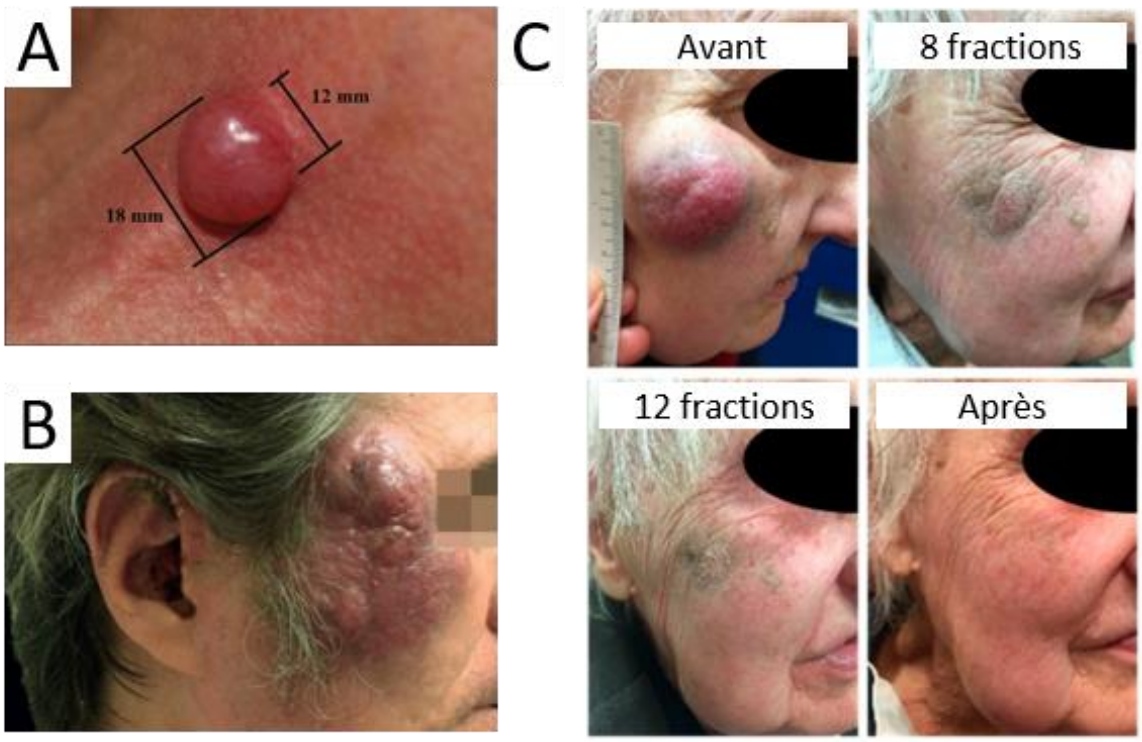


Figure 1 : Aspect clinique du carcinome à cellules de Merkel

***(A)** Lésion primaire observée chez un patient mesurant moins de deux centimètres sur la joue d'un patient (Errichetti et al., 2013) **(B)** Forme avancée de carcinome à cellules de Merkel (Góes et al., 2017). **(C)** Evolution d'un carcinome à cellules de Merkel suite à un traitement par radiothérapie (adapté de Zwijnenburg et al., 2021).*

2.1.3.2 Traitement

Comme pour tous les cancers, les options thérapeutiques envisageables dans les cas de CCM dépendent de plusieurs caractéristiques, dont l'avancée de la maladie. Dans le cas d'un CCM loco-régional, l'exérèse chirurgicale est la prise en charge classique. Les tumeurs de CCM se présentent fréquemment entourées de petites lésions satellites autour de la tumeur primaire. L'exérèse est donc généralement réalisée avec une marge importante (1 à 3 cm). La radiothérapie en traitement adjuvant de la chirurgie a permis de réduire le taux de récidive locale de 70% tout en permettant la diminution de la taille des marges réalisées lors de la chirurgie (Perez *et al.*, 2018). L'addition de radiothérapie mène à une augmentation du taux de survie à 5 ans de 10 à 15% par rapport à la chirurgie comme seule thérapie. Chez des patients présentant trop de comorbidités pour supporter une chirurgie suivie d'une radiothérapie dans un temps rapproché, la radiation en tant que monothérapie est également utilisée et permet la survie de patients présentant un avancement local ou des métastases loco-régionales (Harrington and Kwan, 2014). Les chimiothérapies sont en second rang dans l'arsenal thérapeutique contre les CCM. Ce type de traitement n'est jamais utilisé en première intention. Il a été observé que l'ajout de chimiothérapie de consolidation à la chirurgie et la radiothérapie n'améliorait pas la survie des patients des stades I à III. La chimiothérapie est cependant utilisée lorsque la maladie en est au stade métastatique. Les agents utilisés sont principalement le carboplatine, l'étoposide, la doxorubicine, la vincristine et des cyclophosphamides. La réponse au traitement par chimiothérapie est rapide mais généralement de courte durée (Zwijnenburg *et al.*, 2021).

Les CCM forment des tumeurs très immunogènes. Les développements de l'immunothérapie dans les cancers ont donc naturellement été appliqués au sein des CCM. La forme d'immunothérapie la plus utilisée dans ce cadre passe par le ciblage de la protéine de surface PD-L1 (« programmed death-ligand 1 », Ligand 1 de mort programmée). Le système PD-1 / PD-L1 est impliqué dans la régulation de la réponse immunitaire. Les lymphocytes T expriment à leur surface le récepteur PD-1 (« Programmed cell Death protein 1 ») (Han, Liu and Li, 2020). De nombreuses cellules peuvent exprimer à leur surface la protéine PD-L1. La liaison entre le récepteur PD-1 et son ligand PD-L1 entraine une inhibition des lymphocytes T. En condition physiologique, ce mécanisme permet de contrôler la réponse immunitaire via les lymphocytes T cytotoxiques et d'éviter des dommages tissulaires via des réactions auto-

immunes. Dans le cas des cancer, l'expression de PD-L1 par la cellule cancéreuse lui permet d'échapper au système immunitaire. Le blocage de ce ligand par immunothérapie permet la re-sensibilisation des cellules cancéreuse aux lymphocytes T cytotoxiques (Boyerinas *et al.*, 2015). Les CCM font partie de ces tumeurs où PD-L1 est surexprimé. Les effets de cette inhibition ont été testé au sein d'une cohorte CCM à l'aide de l'Avelumab, un anticorps monoclonal ciblant le récepteur PD-L1. Les 88 patients de cette cohorte présentaient un CCM de stade 4 ayant progressé après traitement par chimiothérapie. Cet essai clinique a pu révéler un taux de réponse objective de 33%, dont 11,4% de réponses complètes (D'Angelo *et al.*, 2020). Une autre étude utilisant le Pembrolizumab qui cible quant à lui le récepteur PD-1 a été réalisée sur des patients non traités en chimiothérapie. Le taux de réponses objectives lors de cette étude était de 56% (Nghiem *et al.*, 2016). Une étude comparant les différentes immunothérapies ciblant PD-1/PD-L1 a conclu que le Pembrolizumab, Nivolumab (ciblant PD-1) et l'Atezolizumab (ciblant PD-L1) avaient une efficacité similaire. L'Avelumab (ciblant PD-L1) avait quant à lui une efficacité légèrement inférieure mais semblait également mieux toléré (Al-Showbaki *et al.*, 2021). Le ciblage d'autres points de contrôles immunitaires est également étudié comme le récepteur CTLA-4. De manière similaire à la voie PD-1/PD-L1, le récepteur CTLA4 est exprimé à la surface des lymphocytes T. La fixation des protéines B7 (CD80 et CD86) au récepteur CTLA-4 entraine une inhibition des lymphocytes T. Ces protéines B7 sont exprimées à la surface des cellules présentatrices d'antigène et des cellules cancéreuses. Une étude récente a proposé une thérapie combinée utilisant le Nivolumab ciblant PD-1 et l'Ipilimumab ciblant CTLA-4. Les auteurs ont observé une amélioration du taux de réponse objective et une réponse plus durable lors de l'administration jointe des 2 immunothérapies (Kim *et al.*, 2022).

2.1.4 Origine cellulaire

L'origine cellulaire du CCM a fait l'objet de débats, avec plusieurs hypothèses proposées au fil des années. La découverte en 2008 de l'étiologie virale d'une large proportion des CCMs a permis une meilleure compréhension de cette pathologie y compris en ce qui concerne son origine cellulaire. En effet, la séparation en deux populations virus positives et virus négatives ainsi que l'étude des différences entre-elles a suggéré que ces deux sous-populations dérivent probablement de deux types cellulaires différents (Sunshine *et al.*, 2018).

2.1.4.1 Cellules de Merkel

L'hypothèse admise initialement soutenait que le CCM provenait des cellules de Merkel en raison de leurs similarités structurelles et immunohistochimiques (Sibley *et al.*, 1980). Les cellules de Merkel sont des mécanorécepteurs impliqués dans le toucher fin localisées dans la couche basale épidermique. L'un des principaux arguments en faveurs d'une origine dans les cellules de Merkel est l'expression de la cytokeratine 20 (CK20), une protéine servant de marqueur spécifique (Moll *et al.*, 1992). Outre CK20, les CCM expriment d'autres marqueurs évocateurs des cellules de Merkel. Les principaux marqueurs utilisés sont la cytokératine 20 mais également la synaptophysine et CD56 (Ortonne *et al.*, 1988; Kurokawa *et al.*, 2003). Cependant, cette hypothèse a depuis été considérée comme peu probable car les cellules de Merkel matures sont post-mitotiques (Moll, Zieger and Schmelz, 1996). Elles proviennent vraisemblablement de cellules à un stade de différenciation antérieur qui se sont multipliées puis différenciées en cellules « Merkel-like ». Il a été observé chez la souris, que l'induction de l'expression d'Atoh1 permettait l'apparition de nouvelles cellules de Merkel (Ostrowski *et al.*, 2015). Une autre étude a montré, toujours chez la souris, que l'induction de l'expression concomitante de sT et ATOH1 suffisait à l'émergence d'agrégats cellulaires présentant les caractéristiques des CCM (Verhaegen *et al.*, 2017).

2.1.4.2 Cellules pre/pro B

Certains arguments suggèrent que les CCM pourraient dériver de lymphocytes pré/pro-B. En effet, ces cellules tumorales expriment des facteurs des lignées B tels que PAX5, c-KIT, TdT, SCF et RAG1 (Sauer *et al.*, 2017). Outre la présence de ces marqueurs, une expression d'immunoglobulines a été observée au sein des CCM-VP (Murakami *et al.*, 2014). Des réarrangements clonaux des chaines lourdes et légères des immunoglobulines retrouvés dans les cellules pre/pro B ont également été détecté au sein de CCM-VP (zur Hausen *et al.*, 2013).

Cependant, d'autres observations vont à l'encontre d'une origine B des CCM. En effet, ces tumeurs contiennent des granules neuroendocrines absents des pré/pro B , et plutôt évocateurs d'une origine neuroendocrine (Munde *et al.*, 2013). De plus, l'infection par le virus MCPyV, retrouvé dans 80% des MCC, semble survenir dans des cellules progénitrices neurales ou cutanées plutôt que dans des précurseurs B (Verhaegen *et al.*, 2017).

2.1.4.3 Fibroblastes dermaux

L'hypothèse d'une origine fibroblastique pour les carcinomes à cellules de Merkel a été évoquée dans certains travaux, mais reste controversée. Plusieurs éléments appuient cette théorie, tout d'abord le taux de mutation somatique très faible et l'absence de signature UV (transition C->T) dans les formes CCM-VP qui suggère l'implication de cellules situées en profondeur dans la peau, moins exposées aux UV, comme le sont les fibroblastes (Sunshine *et al.*, 2018). Lors d'une étude *ex vivo* sur la permissivité des cellules à l'infection au MCPyV, Liu et collaborateurs ont montré que les fibroblastes étaient les seuls à pouvoir être efficacement infectés et exprimer sT et LT (Liu *et al.*, 2016). En plus de pouvoir être infectés et exprimer les protéines virales, les fibroblastes peuvent également être transformés par l'expression de sT *in vitro* (Angermeyer *et al.*, 2013; Shuda, Chang and Moore, 2014). La possibilité pour les fibroblastes de répliquer le génome de MCPyV et donc de produire de grandes quantités d'ADN viral augmente les risques d'une insertion accidentelle du génome de MCPyV, condition nécessaire à son oncogénicité (Feng *et al.*, 2008).

Certains éléments vont cependant à l'encontre de l'hypothèse des fibroblastes comme cellule d'origine pour les CCM-VP. L'absence d'ADN viral dans les fibroblastes avoisinant les tumeurs CCM-VP contredit en effet l'idée d'une infection des fibroblastes suivie d'une insertion accidentelle dans l'un des fibroblastes infectés (Sauer *et al.*, 2017).

2.1.4.4 Cellules épithéliales

Enfin, les cellules souches épithéliales ont également été envisagées comme cellules d'origine des CCM. L'étude des cellules de Merkel a révélé l'importance de plusieurs facteurs de transcription dans leur processus de différenciation. Il a notamment été observé que le facteur ATOH1 était impliqué dans la différenciation en cellules de Merkel à partir de cellules de la peau, excluant ainsi la possibilité d'une origine dans la crête neurale (Morrison *et al.*, 2009). Chez la souris, la délétion d'*Atoh1* dans les cellules souches épithéliales au niveau de l'épiderme entraine une absence de cellules de Merkel (Van Keymeulen *et al.*, 2009). L'analyse des profils de méthylation de l'ADN soutient également la thèse de l'origine au niveau de progéniteurs épithéliaux. Gravemeyer *et al.* ont observé que les CCM-VP et -VN présentaient un faible niveau de méthylation au niveau des promoteurs de marqueurs neuroendocrine

comme *SOX2* ou *INSM1*. Ils ont identifié à l'aide d'un modèle d'étude des profils de méthylation que les CCM présentaient une origine épithéliale (Gravemeyer *et al.*, 2022). La possibilité que les CCM puissent dériver de cellules épithéliales a également été montré par Kervarrec *et al.* Ils ont observé l'intégration du MCPyV au sein d'un trichoblastome qui est une tumeur bénigne du follicule pileux. Cette intégration montre la possibilité pour les CCM de trouver une origine dans les cellules épithéliales et suggère également les follicules pileux comme cellule d'origine (Kervarrec, Aljundi, *et al.*, 2020). En effet, les éléments tels que le faible nombre de mutations somatiques, l'absence de signatures UV et la localisation profonde des CCM sont cohérents avec une localisation d'origine située au sein des follicules pileux. Enfin, le même groupe a également montré que l'induction de l'expression des antigènes T combinée à l'expression de GLI1 (GLI family zinc finger 1) au sein de kératinocytes entrainait l'acquisition d'un phénotype similaire aux cellules de Merkel (Kervarrec, Samimi, *et al.*, 2020). L'apparition des CCM à partir de progéniteurs épithéliaux est aujourd'hui la thèse la plus probable. L'acquisition de caractères similaires aux cellules de Merkel serait alors le fruit notamment de la stabilisation de facteurs comme ATOH1 par la présence de l'antigène viral grand T (Large T antigen, LT) (Kervarrec, Samimi, *et al.*, 2020).

2.2 POLYOMAVIRUS A CELLULES DE MERKEL

2.2.1 Découverte

Le carcinome à cellule de Merkel est un cancer rare qui touche particulièrement les sujets présentant un déficit immunitaire. En effet, il a été observé que l'incidence des CCM était considérablement augmentée chez les patients immunodéprimés suite à une infection du VIH ou à la prise d'un traitement immunosuppresseur dans les cas de transplantations (Becker *et al.*, 2018). Cette particularité rappelle une autre pathologie, le sarcome de Kaposi dont l'étiologie virale a été établie par l'équipe de Patrick Moore en 1994 (Chang *et al.*, 1994). C'est cette similarité qui a poussé la même équipe à s'intéresser au CCM. En étudiant la possibilité d'une origine virale pour ce cancer, ils ont réalisé un séquençage haut débit du transcriptome de 4 tumeurs CCM. La méthode de soustraction digitale du transcriptome développée par la même équipe a été utilisée pour identifier des séquences de potentiels pathogènes. Il s'agit de séquençage du transcriptome suivie d'une soustraction de toutes les

séquences d'origine humaine par comparaison avec différentes banques de données humaines (Feng *et al.*, 2007). C'est par cette méthode que deux séquences d'origines inconnues ont été identifiées. L'une de ces séquences présentait une homologie de séquence avec les antigènes T de plusieurs polyomavirus. C'est ensuite par « genome walking » que la séquence complète d'un nouveau virus qu'ils nomment Polyomavirus à cellules de Merkel a été identifiée (Feng *et al.*, 2008). Le « genome walking » est une méthode qui permet de déterminer la séquence d'ADN adjacente et inconnue à une région d'ADN connue. Pour cela, des PCR successives sont réalisées avec des amorces dégénérées, c'est-à-dire avec des bases aléatoires, afin de progresser de proche en proche sur la région inconnue au départ (Leoni *et al.*, 2011). La présence du MCPyV a ensuite été testée par réaction de polymérisation en chaine (PCR) et Southern-blotting dans un échantillon de 10 tumeurs de CCM et 59 échantillons provenant de patients non-CCM. Ils ont observé une majorité de CCM présentant le MCPyV (80%) tandis que la totalité des échantillons non CCM étaient négatifs par PCR et seulement 5 présentaient un signal faible en Southern-blotting (Feng *et al.*, 2008).

2.2.2 Structure du génome

Le polyomavirus à cellules de Merkel est un virus à ADN double brin qui possède un génome circulaire généralement autour de 5,4 kb. Il respecte une organisation conservée parmi les polyomavirus. Le génome du MCPyV se compose d'une origine de réplication ainsi que de 2 régions distinctes, la région précoce et la région tardive. Ces deux régions contenant les gènes exprimés par le virus sont séparées par une région régulatrice non codante (« non-coding control region », NCCR). Cette NCCR représente un promoteur bidirectionnel qui permet un contrôle temporel de l'expression des gènes viraux (**Figure 2a**).

2.2.2.1 Région précoce

La région précoce du génome du MCPyV contient, comme son nom l'indique, les gènes qui sont transcrit dans les phases précoce de l'infection. Elle contient quatre transcrits superposés et différentiés par épissage alternatif. Les protéines codées par cette région sont l'antigène petit T (« small Tumor antigen », sT), l'antigène grand T (« Large Tumor antigen », LT), l'antigène 57kT (57kT) ainsi qu'une protéine appelée cadre de lecture alternatif de LT (« Alternative LT Open reading frame », ALTO). Excepté ALTO, ces protéines partagent leur région N-terminale, la partie C terminale de sT est épissée pour LT et 57kT qui ont la majeure

partie de leurs séquences en commun avec un exon de plus épissé pour 57kT. La protéine ALTO quant à elle débute plus tard dans la séquence est n'a pas de partie commune avec sT (**Figure 2a**). LT est une protéine de 817 acides aminés encodée par un transcrit composé de deux exons (Shuda *et al.*, 2008). Sa partie N-terminale est composée d'un domaine DnaJ qui contient plusieurs motifs connus (**Figure 2b**). On y retrouve notamment une séquence HPDKGG permettant l'interaction avec les protéines de choc thermique de la famille des Hsc70 (Heat shock cognate 70) ainsi que la région conservée 1 (CR1) LXXLL (Kwun *et al.*, 2009). Le deuxième exon de LT porte le motif LXCXE qui est un domaine d'interaction avec RB1. Cet exon code également le domaine de liaison à l'origine de réplication (OBD) et un signal de localisation nucléaire (NLS). La partie Carboxy-terminale de cette protéine virale contient le domaine hélicase qui est essentiel pour la réplication du virus (Shuda *et al.*, 2008). L'antigène grand T du polyomavirus à cellule de Merkel possède, également sur son deuxième exon, un domaine qui est spécifique à ce polyomavirus (Houben *et al.*, 2015). Il s'agit de la région unique du MCPyV (« MCPyV Unique Region », MUR), elle se situe en amont de l'OBD et représente environ 200 acides aminés. Cette région comprend notamment un motif de liaison à la protéine de tri vacuolaire Vamp6 et trois sérines (S147, S220, S239) dont la phosphorylation régule l'interaction avec les SCF/E3 ligase (Skp-Cullin-F-box), et donc la stabilité de LT. Le MUR contient également la séquence d'interaction avec RB1 (LXCXE) et le NLS (Liu *et al.*, 2011; Kwun, Chang and Moore, 2017).

La protéine 57kT a en commun avec LT sa partie N-terminal ainsi que le début de le deuxième exon comprenant le MUR. En raison d'un épissage alternatif, cet exon de LT n'est pas retrouvé intégralement chez 57kT, l'OBD notamment est absent et on ne retrouve pas intégralement le domaine hélicase mais uniquement les 100 acides aminés en position C-terminale sur un troisième exon (Spurgeon and Lambert, 2013).

La protéine sT partage également sa partie N-terminale avec LT et 57kT (Kwun *et al.*, 2013). On retrouve donc le domaine DnaJ contenant la région CR1 et la séquence HPDKGG (**Figure 2b**). Cette protéine est le produit de la transcription continue sans épissage au niveau du premier site d'épissage, qui marque la fin du premier exon de LT. Cette protéine se prolonge par une séquence de 106 acides aminés codée par la région intronique de LT située juste après ce premier site d'épissage. Cette région contient trois domaines qui ne sont pas retrouvés dans les séquences des autres antigènes T du MCPyV. Deux sites d'interactions avec

des phosphatases ont été identifiés, avec la protéine phosphatase 2A (PP2A) et la protéine phosphatase 4C (PP4C). Enfin, sT possède un domaine de stabilisation de LT (« LT Stabilization Domain », LSD) qui régulerait la stabilité du LT en ciblant les complexes SCF responsables de la dégradation de LT (Kwun *et al.*, 2013).

La quatrième protéine codée par la région précoce du génome des MCPyV est la protéines ALTO, celle-ci ne partage pas la région N-terminale commune aux trois autres antigènes T. La transcription du cadre de lecture alternatif codant ALTO débute sur un codon d'initiation présent dans la séquence du deuxième exon de LT. ALTO est essentiellement constituée du domaine MUR présent dans la séquence de LT et 57kT. Des études phylogénétiques suggèrent que malgré une faible identité de séquence, cette protéine présente des similarités avec l'antigène moyen T (« middle T », MT) du polyomavirus murin (Carter *et al.*, 2013).

En plus des protéines virales, la région précoce du génome du MCPyV permet également l'expression d'un micro ARN, le MCV-miR-M1. Il s'agit d'une courte séquence de 22 nucléotides qui est complémentaire inverse de la séquence de LT à proximité du motif LXCXE dans la région commune à LT, 57kT et ALTO. Cette complémentarité suggère une auto-régulation de l'expression des antigènes T par le virus. Ce MCV-miR-M1 est également capable de cibler le transcrit codant la protéine de l'immunité innée SP100 (Akhbari *et al.*, 2018).

2.2.2.2 Région tardive

A l'opposée de la région précoce par rapport au NCCR se trouve la région tardive dont les gènes s'expriment plus tard dans la cinétique de l'infection virale (**Figure 2a**). Cette région contient exclusivement les gènes codant des protéines structurales nécessaires à la formation de virions. Trois cadres de lectures ont été identifiés dans la région tardive du MCPyV. Le gène VP1 code la protéine la plus abondante de la capside virale, la protéine de capside majeure (Schowalter and Buck, 2013). Le gènes VP2 code la protéine de capside mineure qui est moins exprimée que VP1 et semble être impliquée dans l'assemblage de la capside. Enfin le troisième cadre de lecture identifié dans cette région tardive correspond au gène VP3. Il s'agit d'une autre protéine de capside mineure retrouvée dans d'autre polyomavirus. Contrairement à ce qui est retrouvé chez les autres polyomavirus, la séquence de VP3 ne présente pas à son extrémité N-terminale le motif conservé Methionine-Alanine-Leucine. L'expression de ce gène

est mise en cause par cette absence de l'extrémité N-terminale conservée mais également par l'impossibilité de détecter VP3 en cellules infectées et l'observation de mutations du codon d'initiation sur le génome de virus capable d'infecter des cellules et de se répliquer (Schowalter and Buck, 2013).

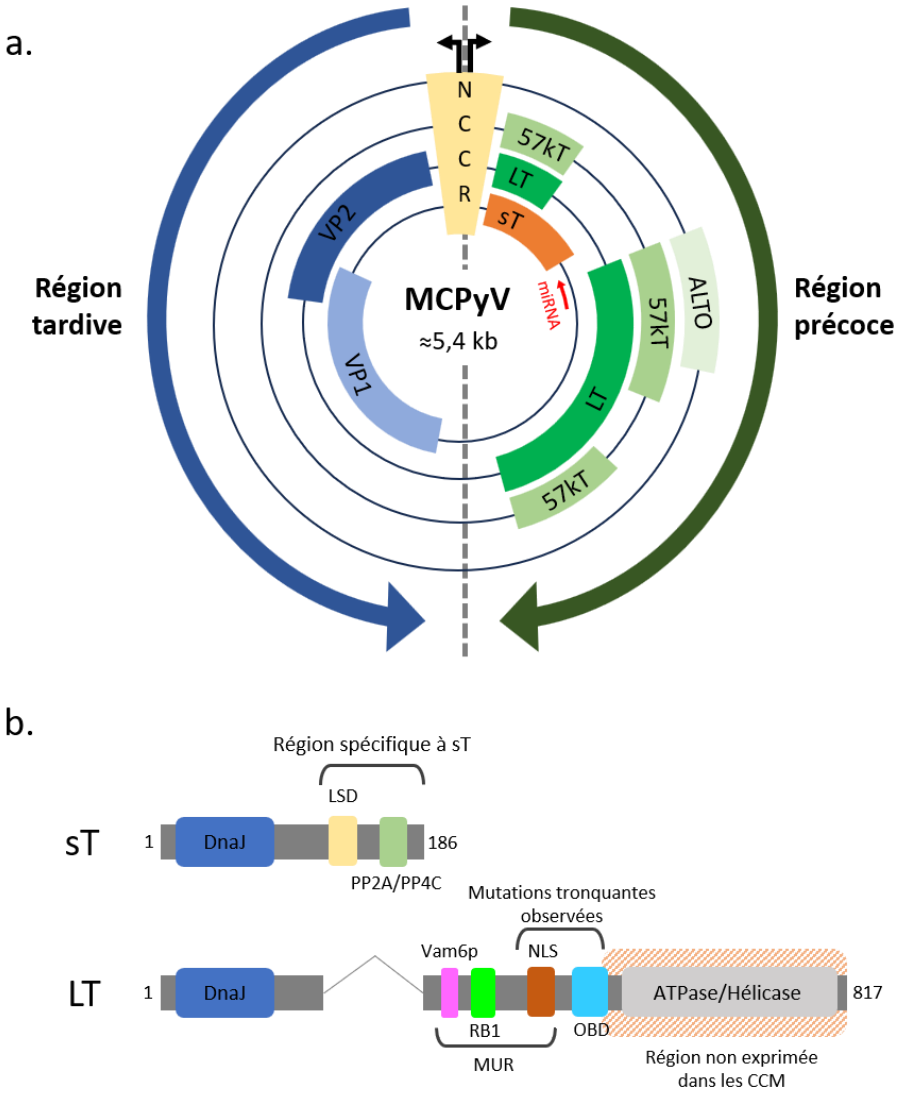


Figure 2 : Structure du génome du MCPyV et domaines fonctionnels des antigènes T.

***(A)** Le génome du MCPyV est répartit en deux régions : précoce et tardive. Ces régions sont séparées par le NCCR, promoteur bidirectionnel qui contrôle la régulation temporelle de l'expression des gènes du virus. Le NCCR contient également l'origine de réplication. La région précoce est composée des gènes codant sT, LT, 57kT et ALTO, dont l'existence est controversée. Cette région code également un micro ARN, le MCV-miR-M1 qui a une séquence complémentaire avec le début du deuxième exon de LT dans la partie commune avec 57kT et ALTO. **(B)** Les protéines sT et LT partagent une région N-terminale commune portant le domaine DnaJ. Ce domaine DnaJ contient la région conservée CR1 ainsi que la séquence HPDKGG nécessaire pour l'interaction avec les Hsc70. La région spécifique à sT (épissée dans la séquence de LT) porte le domaine LSD supposé stabiliser LT via son interaction avec les complexes SCF/E3. Cette région porte également les sites d'interaction avec les protéines phosphatases PP2A et PP4C. L'exon 2 de LT contient le domaine MUR (MCPyV Unic Region) qui porte les sites d'interaction avec Vam6p et RB1 ainsi que le signal de localisation nucléaire (NLS). Cet exon code également le domaine de liaison à l'origine de réplication (OBD) et la partie C-terminal qui porte l'activité ATPase/hélicase. Cette partie C-terminale est systématiquement tronquée dans les CCM. Les mutations tronquantes observées chez les patients CCM conservent toujours le domaine de liaison à RB1 et tronquent toujours le domaine ATPase/hélicase.*

2.2.3 Cycle du virus

Il est désormais connu que le MCPyV est l'agent étiologique principal du CCM. Il est cependant surprenant d'apprendre que ce virus est très prévalent dans la population et est considéré par certains comme faisant partie de la flore commensale de la peau (Kean *et al.*, 2009; Pastrana *et al.*, 2009; Tolstov *et al.*, 2009). Il existe donc des événements qui déterminent le potentiel oncogénique de ce virus à forte prévalence dans la population qui est à l'origine d'un cancer rare.

La première étape de l'infection au MCPyV est l'entrée du virus dans la cellule hôte. Les virions qui vont infecter les cellules sont constitués du génome encapsidé dans une capside virale de forme icosaédrique formée par les protéines structurales VP1 et VP2 codées par la région tardive du génome. Dans un premier temps, l'attachement des virions à la surface de la cellule se fait par l'interaction entre la protéine VP1 et des glycosaminoglycanes (GAG) sulfatés à la surface de la cellule hôte. Parmi les GAG, les héparanes sulfates sont essentiels à l'attachement des virions tandis que les chondroïtines sulfates ne semblent pas nécessaires. Dans un second temps, des glycanes sialylés servent de corécepteur post-attachement et représentent une étape essentielle pour l'entrée du virion dans la cellule (Schowalter, Pastrana and Buck, 2011).

L'internalisation des virions se fait par endocytose dépendante des cavéoles dans les radeaux lipidiques. Les virions entrent par le bourgeonnement des cavéoles et sont acheminés vers le réticulum endoplasmique au sein d'endosomes via les microtubules (Becker *et al.*, 2019). La décapsidation du génome viral est réalisée dans le réticulum endoplasmique et l'ADN viral est importé dans le noyau par les pores nucléaires (Giannecchini, 2020).

Les gènes de la région précoce sont les premiers à être exprimés. Les protéines LT et sT une fois traduites vont influencer la réplication virale. LT est capable de fixer l'origine de réplication se situant dans le NCCR grâce à son domaine OBD. Le domaine ATPase/hélicase présent à son extrémité C-terminale permet alors l'oligomérisation de LT essentielles pour l'initialisation de la réplication virale. Le rôle de sT dans la réplication virale semble principalement lié à la stabilisation de LT via son LSD. Il a été proposé que le ciblage de LT par les complexes SCF puisse servir de mécanisme de latence au MCPyV. La dégradation de LT empêcherait l'accumulation et le recrutement de la machinerie de réplication au niveau du

génome viral. En condition de stress cellulaire, la baisse d'activité des complexes SCF se traduirait alors pas une accumulation de LT et ainsi une activation accrue de la transcription et réplication du génome viral (Kwun, Chang and Moore, 2017). Il a également été observé que l'expression du MCV-miR-M1 avait un effet inhibiteur sur la réplication virale en diminuant l'expression des protéines virales. Cette diminution d'expression dépendante du MCV-miR-M1 améliore la persistance de l'infection virale sous forme épisomale en tant qu'infection latente (Theiss *et al.*, 2015).

Le mécanisme d'assemblage et de libération des virions de MCPyV est à ce jour assez mal connu. Une étude de Li *et al.* a pu identifier un autoassemblage de protéines VP1 formant de larges particules semblables aux virions MCPyV. Ils ont observé la formation de particules sphérique de 50 nm de diamètre renfermant des fragments d'ADN allant de 1,5 kb à 5kb provenant de l'ADN de la cellule hôte ainsi que de plus petites particules de 25 nm ne renfermant pas d'ADN. Une analyse par cryo-electro microscopie a montré que ces particules larges avaient un diamètre de 48nm et étaient formées de 72 capsomères pentamériques (Li *et al.*, 2015).

Le MCPyV forme généralement des infections longues durées où l'infection est silencieuse et le virus persiste dans la cellule sous forme latente et épisomale. La survenue de CCM impliquant le MCPyV nécessite la survenue d'évènements rare.

2.2.4 Implication dans les MCC

2.2.4.1 Statistiques / épidémiologie

Suite à la découverte du MCPyV en 2008, des études épidémiologiques ont été réalisées afin de connaitre sa prévalence. Il a ainsi pu être observé que l'infection au MCPyV était commune dans la population. Certains considèrent même ce virus comme constituant de la flore normale de la peau. En effet, le MCPyV est retrouvé à la fois sous forme de virion ou simplement en tant que génome viral sans sa capside. Le MCPyV comme d'autres polyomavirus présente donc une forte prévalence dans la population générale. Il a été observé que le sérum de 60 à 80% des adultes présentait une réactivité pour la protéine VP1 du MCPyV. Une prévalence importante est également retrouvée chez les enfants avec une séropositivité de 50% des sujets de moins de 15 ans (Tolstov *et al.*, 2009). Une autre étude a

identifié une séropositivité de 9% chez les enfants de 3 à 4 ans et de 35% chez les enfants de 4 à 13 ans. L'évolution de la séropositivité dans l'enfance combinée au fait que l'ADN viral n'est pas détecté dans des échantillons de sang néonatal suggère que l'infection au MCPyV est le fruit d'une exposition dans l'enfance (Chen *et al.*, 2011). Des anticorps dirigés contre les antigènes T sont également retrouvés chez les patients CCM-VP dans 40,5% des cas contre 0,9% chez les sujets sains. De plus, parmi les sujets CCM-VP, le taux d'anticorps est corrélé à l'agressivité de la maladie et est 8 fois inférieur dans les cas où une rechute n'a pas été observée. Il a donc été conclu que la présence d'anticorps dirigés contre les antigènes T ne freinait pas la progression de la pathologie. Il a également été observé que dans les cas de patients développant des métastases, l'augmentation du taux d'anticorps survient avant la détection clinique de la propagation du cancer. En conséquences, il a été proposé que la titration des anticorps dirigés contre les antigènes T pourrait servir d'indicateur de la progression de la maladie (Paulson *et al.*, 2010). Le mode de transmission du MCPyV reste à ce jour inconnu, l'ADN viral a été retrouvé dans différents tissus compliquant l'identification du lieu d'infection primaire. Il a été supposé que ce virus puisse être transmis par voies aériennes suite à l'identification de la présence de son ADN dans les voies respiratoires (Kantola *et al.*, 2009). Sa présence également retrouvée dans l'appareil digestif pointe quant à elle vers une transmission oro-fécale (Loyo *et al.*, 2010). L'une des hypothèses privilégiées reste néanmoins la transmission par contact avec la peau. De nombreuses études ont pu identifier la présence du virus sur la peau de sujets CCM ou sains. Ce virus pourrait donc se transmettre par contact peau à peau. Il a été observé que le contact entre la peau et une surface matérielle menait à la contamination de la surface par de l'ADN viral. Environ 5% de cet ADN viral était résistant aux DNases suggérant la présence de virions. Il est alors envisageable que l'infection au MCPyV puisse également survenir par contact avec une surface contaminée (Foulongne *et al.*, 2011).

2.2.4.2 Intégration

Le génome de MCPyV peut persister dans la cellule hôte durant de longue périodes de latence sous forme épisomale. Cette latence est médiée par le ciblage de LT par les ubiquitines ligases et par l'expression du MCV-miR-M1 (Theiss *et al.*, 2015; Kwun, Chang and Moore, 2017). Dans les cas de CCM, l'ADN viral n'est cependant pas présent sous forme épisomale mais systématiquement intégré au génome (Feng *et al.*, 2008). L'intégration du génome viral du MCPyV fait donc parti du processus oncogénique des CCM-VP. Le processus exact par lequel

le génome du MCPyV est intégré au génome est à ce jour inconnu, plusieurs hypothèses ont cependant été formulées. L'intégration et le potentiel oncogénique du MCPyV seraient le fruit de cassures doubles brins associées à des mutations du LT entrainant sa troncation et la perte du domaine C-terminal, supprimant ainsi son activité ATPase/Hélicase. Le site d'intégration du virus au sein du génome de la cellule hôte ne semble pas participer à ses effets oncogéniques. En effet, l'analyse du site d'intégration de l'ADN viral n'a révélé aucun site préférentiel et a montré une grande variabilité des sites d'intégration entre les patients (Duncavage *et al.*, 2011; Martel-Jantin *et al.*, 2012). La variabilité intra patient est quant à elle inexistante, une intégration clonale du MCPyV a été observée chez les patients CCM suggérant le développement de ces cancers à partir d'une seule cellule ayant subi les différents évènements rares nécessaire à l'oncogenèse CCM (Feng *et al.*, 2008).

Le premier évènement est la troncation de LT. Cet événement est nécessaire à l'émergence du CCM. L'expression de la forme complète de LT entraine une inhibition de la croissance cellulaire et l'arrêt du cycle cellulaire suite à l'activation de la voie de réponse aux dommages de l'ADN dépendante de p53 (DNA Damage Response, DDR) (Li *et al.*, 2013). Une troncation de cette partie C-terminale est systématiquement observée dans les CCM. Cette troncation est causée par des mutations qui entrainent l'apparition de codons stop prématurés. Une étude récente a proposé une origine de ces mutations comme conséquence d'un système antiviral de la cellule, les APOBECs (Soikkeli *et al.*, 2022). Les APOBECs (« Apolipoprotéins B mRNA editing enzyme, catalytic polypeptide-like », enzyme d'édition d'ARNm de l'apolipoprotéine B agissant comme un polypeptide catalitique) sont des protéines capables de catalyser la réaction de déamination qui converti l'acide nucléique Cytosine en Uracile. Les APOBEC3 sont impliquées dans la déamination de cytosines dans le cadre d'une réponse antivirale. Ces enzymes sont ainsi responsables de mutations dans le génome viral ayant pour but d'empêcher la réplication virale (Siriwardena, Chen and Bhagwat, 2016). Soikkeli *et al.* ont observé au sein d'une cohorte de patients CCM que les mutations présentes sur la séquence de LT correspondaient à la signature des déaminations induites par les APOBEC3. Il est donc plausible que la troncation donnant son caractère oncogénique à LT soit le fruit de la réponse antivirale de la cellule (Soikkeli *et al.*, 2022). Ces formes tronquées conservent toujours les domaines DnaJ et le domaine d'interaction avec la protéine RB1 (**Figure 2b**). L'observation des points de mutations au sein de CCM présentant une intégration en tandem du génome de

MCPyV a montré que l'évènement de mutation de LT précédait celui de l'intégration au génome (Schrama *et al.*, 2019). Il y aurait donc à la base de la naissance des CCM, après l'infection, la réactivation de la réplication virale, puis un évènement de mutation de la séquence de LT possiblement par intervention des APOBECs (Soikkeli *et al.*, 2022). Suite à ces mutations, des épisomes portant un LT tronqué et d'autres avec un LT sauvage (« Wild Type », WT) sont présents simultanément dans la cellule. Ainsi, la réplication des épisomes portant la mutation peut être assurée par le LT WT codée par d'autres épisomes. Une réplication circulaire continue a probablement lieu (**Figure 3a**), ce qui explique l'observation d'ADN viral intégré sous forme concaténée sans ADN humain entre les séquences virales (Schrama *et al.*, 2019).

Le deuxième événement nécessaire à l'émergence de CCM est donc l'intégration du génome viral au génome de la cellule hôte. L'hypothèse la plus probable est celle d'une erreur lors de la réplication menant à des cassures double brin (« Double Strand Break », DSB) associée à une réplication circulaire continue. Il s'agit d'une molécule d'ADN circulaire, ici le génome viral, qui est répliquée à partir de son origine de réplication de façon continue. Ainsi, après avoir réalisé un tour complet, la réplication ne s'arrête pas et un exemplaire de la séquence du génome viral est ajouté à la suite du premier (Bjursell, 1978). Cette version explique la présence d'intégration de génomes viraux en tandem (**Figure 3a**). L'intégration de ces molécules d'ADN linéaires est ensuite possible suite à des cassures double brin du génome de la cellule hôte. A ce jour, le scénario privilégié comprend donc cette réplication circulaire continue formant une longue molécule d'ADN linéaire associée à des cassures de l'ADN de la cellule hôte offrant une possibilité d'intégration pour le génome viral. Deux mécanismes ont été proposés pour le processus d'intégration. Le premier consiste en une jonction des extrémités non-homologues (« Non Homologous End-Joining », NHEJ). Il s'agit d'un mécanisme de réparation de l'ADN qui permet, en cas de DSB, de joindre deux extrémités d'ADN sans nécessiter une matrice homologue contrairement à la recombinaison homologue (**Figure 3b**). NHEJ est un mécanisme de réparation imparfait qui conduit fréquemment à des insertions ou délétions dans la séquence d'origine au niveau de la jonction. Il a été suggéré que la survenue de cassures double brins au niveau du génome de la cellule hôte et à proximité d'une molécule linéarisée d'ADN viral puisse intégrer le génome viral durant la réparation par NHEJ (Czech-Sioli *et al.*, 2020). L'étude menée par Czech-Sioli *et al.* a permis l'observation de

deux schémas d'intégration distincts. Ce premier schéma présentant une intégration linéaire d'une copie simple du génome virale ou d'une forme en tandem correspondant à l'intégration par NHEJ. Un second schéma plus complexe a également été observé et présente une duplication des régions adjacentes de l'ADN hôte. Ce schéma correspond à la formation de « Z-pattern » lors de la jonction d'extrémités dépendante de micro-homologies (« Microhomology-Mediated End Joining », MMEJ). Il s'agit d'un mécanisme de réparation des cassures double brin alternatif au NHEJ. Lors du MMEJ, les extrémités 5' de l'ADN viral et de l'ADN hôte sont réséquées. Un appariement peut alors s'effectuer par complémentarité entre les extrémités cohésives des ADN hôte et viral présentant une micro-homologie (**Figure 3c**). Une « invasion » est ensuite réalisée par l'autre extrémité 3' de l'ADN viral qui s'apparie par micro-homologie à une séquence de l'ADN hôte à proximité, formant ainsi une boucle. La synthèse d'ADN s'effectue ensuite en prenant le brin d'ADN hôte intact comme matrice. L'amplification de cette boucle se poursuit par une migration de branche et entraine au cours du processus l'intégration de la molécule d'ADN viral accompagnée d'une duplication d'une portion de l'ADN hôte formant un « Z-pattern ». Il s'agit de la réplication induite par les cassures dépendantes des micro-homologies (« Microhomology-Mediated Break-Induced Replication », MMBIR). Ce processus d'intégration du génome viral plus complexe que la voie NHEJ explique la présence de duplications de l'ADN hôte observée à certains sites d'intégration du MCPyV (Czech-Sioli *et al.*, 2020). Tous les sites d'intégrations ne présentant pas de duplication, il est probable que les deux mécanismes soient impliqués, et que l'un ou l'autre puisse se produire et aboutir à l'émergence de CCM.

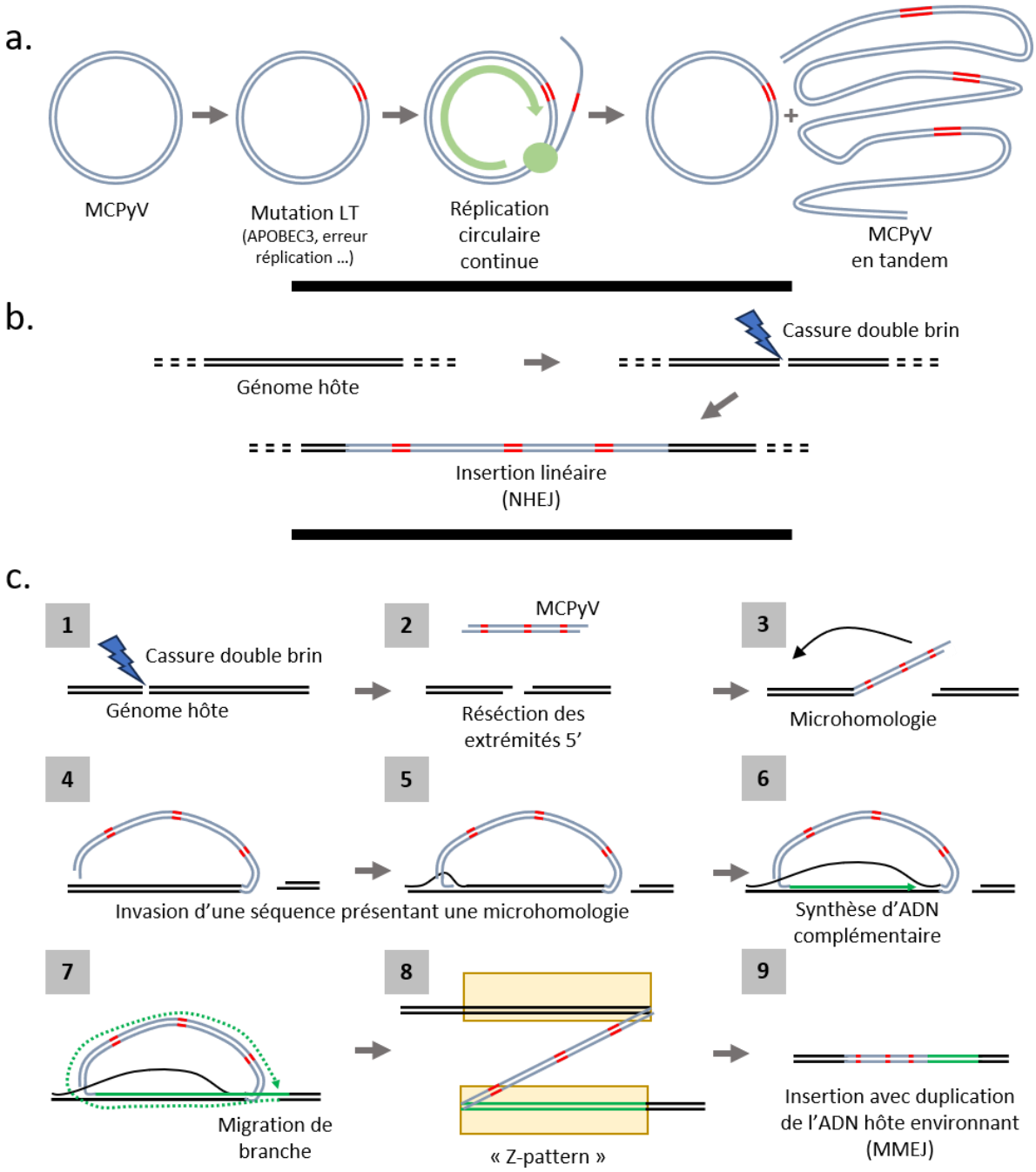


Figure 3 : Processus de mutation, de linéarisation et d'insertion du MCPyV par NHEJ et MMEJ.

***(A)** Les mutations de troncation de LT surviennent avant l'insertion par action des APOBECs, suite à une réparation défectueuse ou par erreur de réplication. Lors de la réplication du génome, une réplication circulaire continue se met probablement en place donnant lieu à une molécule d'ADN linéaire en tandem. **(B)** L'insertion du MCPyV via NHEJ est le mécanisme le plus simple. Le génome viral est incorporé au génome hôte lors de la réparation par NHEJ suite à une cassure double brin de l'ADN hôte. **(C)** Suite à une cassure double brin et à la résection des extrémités 5' (1 ;2), le MCPyV s'apparie par micro-homologie (3). Une invasion est réalisée par son extrémité 3' vers une séquence présentant une micro-homologie (4 ;5). S'ensuit une synthèse d'ADN avec la séquence ADN hôte pour matrice (6). Au niveau de la jonction hôte-virus, une migration de branche s'effectue et lie le brin néosynthétisé à l'ADN hôte en 3' de la cassure. La synthèse du brin complémentaire est réalisée prenant pour matrice l'ADN néosynthétisé (7). Le produit de cette intégration associée à la synthèse d'ADN est un « Z-pattern » avec une duplication de la séquence avoisinant le site d'insertion (cadre jaune) (8). Le génome viral est alors inséré au génome sous forme simple ou en tandem et flanqué d'une séquence hôte dupliquée (9).*

2.3 ONCOGENESE DES CARCINOMES A CELLULES DE MERKEL

Le carcinome à cellule de Merkel est, comme nous l'avons dit précédemment, un cancer rare et très agressif. La proportion de CCM lié à la présence du MCPyV est variable selon les études mais il est considéré qu'environ 80% d'entre eux sont associés au virus. Les études portant sur les mécanismes oncogéniques, l'origine cellulaire des CCM et les analyses descriptives de ces cancers tendent à montrer une forte différence entre les formes virus positives et virus négatives des CCM. La principale différence entre ces deux formes réside dans l'élément déclencheur du cancer. Alors que le MCPyV est l'agent oncogénique des formes virus positives, les formes virus négatives présentent une plus grande variabilité en reposant sur des mécanismes plus communs dans les cancers comme des mutations inactivant des gènes suppresseurs de tumeurs.

2.3.1 Carcinomes à cellules de Merkel virus positifs

Les formes virus positives des CCM représentent 80% des CCM. L'expression des antigènes T est un élément essentiel au développement et au maintien de ce cancer. Parmi les protéines codées par ce virus, seules deux sont connues comme étant impliquées dans l'oncogenèse des CCM à la fois par l'élucidation de mécanismes qui seront détaillés dans cette partie mais également par l'observation de l'effet de l'inhibition de ces protéines. En effet, ces protéines virales ont été ciblées par shRNA au sein de lignées CCM-VP. Les shRNA sont des petits ARN ayant une structure en épingle à cheveux (« short hairpin RNA », shRNA) dont une partie de la séquence est complémentaire à celle de l'ARNm (ARN messager) cible. L'appariement entre le shRNA et l'ARNm entraine une extinction de l'expression du gène cible soit en bloquant la traduction, soit en entrainant la dégradation de cette ARNm par la machinerie cellulaire. Ces ciblages par shRNA ont provoqué une forte réduction allant jusqu'à l'arrêt de la prolifération des cellules CCM-VP (Houben *et al.*, 2010, 2012; Shuda, Chang and Moore, 2014). L'oncogenèse des CCM-VP est donc dépendante de deux protéines virales, l'antigène grand T (LT) et le petit T (sT). Avec l'augmentation des informations sur ces protéines virales, une tendance quant à la répartition de leurs fonctions se dégage. Il semblerait en effet que bien que les deux protéines soient nécessaires à l'instauration et au maintien d'un CCM, leur importance est variable selon la phase de développement du cancer. Il a notamment été observé lors de l'inhibition de l'expression des antigènes T par shRNA que

la prolifération des cellules pouvait être relancée par l'expression d'une séquence de LT non sensible aux shRNA. Ainsi, dans une lignée cellulaire CCM-VP établie, la présence de sT n'est pas nécessaire à la prolifération des cellules. Le rôle de sT se situerait plus tôt dans l'oncogenèse des CCM. L'expression de sT est suffisante pour transformer des fibroblastes dermaux ce qui suggère que sT est un facteur important pour la transformation des cellules cancéreuses (Shuda *et al.*, 2011).

2.3.1.1 LT

Comme mentionné précédemment, l'une des caractéristiques de l'antigène grand T des CCM est qu'il n'est pas exprimé sous sa forme pleine taille dans les cellules cancéreuses. Le processus oncogénique des CCM nécessite la survenue de mutations du gène du LT qui entrainent la perte de la partie C-terminal portant l'activité ATPase/hélicase tout en conservant le domaine de liaison à RB1. C'est donc toujours une forme tronquée qui est exprimée au sein des CCM, cette forme sera désignée tLT (« truncated Large T antigen »).

2.3.1.1.1 RB1

L'interaction entre tLT et RB1 est l'un des premiers mécanismes oncogéniques à avoir été identifié pour le MCPyV. Cette interaction se fait via la séquence LXCXE présente dans le MUR de l'exon 2 du gène de LT. Le domaine d'interaction avec RB1 est une partie essentielle de la séquence de tLT. Il a été observé que la baisse de prolifération induite par l'inhibition des antigènes T pouvait être contrée par l'expression d'une forme de tLT insensible aux shRNA mais pas par une forme mutée sur la séquence LXCXE (Houben *et al.*, 2012). L'interaction entre tLT et RB1 a pour conséquence la séquestration de ce suppresseur de tumeur (**Figure 4a**). Ainsi, RB1 ne peut plus assurer son rôle dans le contrôle du cycle cellulaire. En condition physiologique, les protéines de la famille Rb (Retinoblastoma) tiennent un rôle suppresseur de tumeur. Elles agissent notamment en séquestrant les facteurs de transcription E2F. Ces facteurs sont impliqués dans la promotion du cycle cellulaire. L'inhibition de la séquestration des facteurs E2F se traduit donc par une perte de contrôle sur le cycle cellulaire et l'induction de la progression en phase S (Chellappan *et al.*, 1991). Cette interaction semble spécifique à la protéine RB1, tLT ne semble pas agir sur les autres protéines de la famille rétinoblastome

p107 et p130. En plus de la séquence LXCXE, il est envisageable que le domaine DnaJ présent à l'extrémité N-terminale de la protéine soit également impliqué dans l'inhibition de RB1. Ce domaine contient la séquence HPDKGG responsable de l'interaction avec les protéines chaperonnes de la famille Hsc70. Il a été observé dans d'autres polyomavirus oncogéniques chez l'animal comme le SV40 que l'activité chaperonne de Hsc70 était nécessaire pour permettre la dissociation du complexe RB1-E2F (Sullivan, Cantalupo and Pipas, 2000). La **Figure 4** comporte une représentation schématique du mécanisme oncogénique impliquant RB1 ainsi que les mécanismes décrit dans la suite de ce manuscrit.

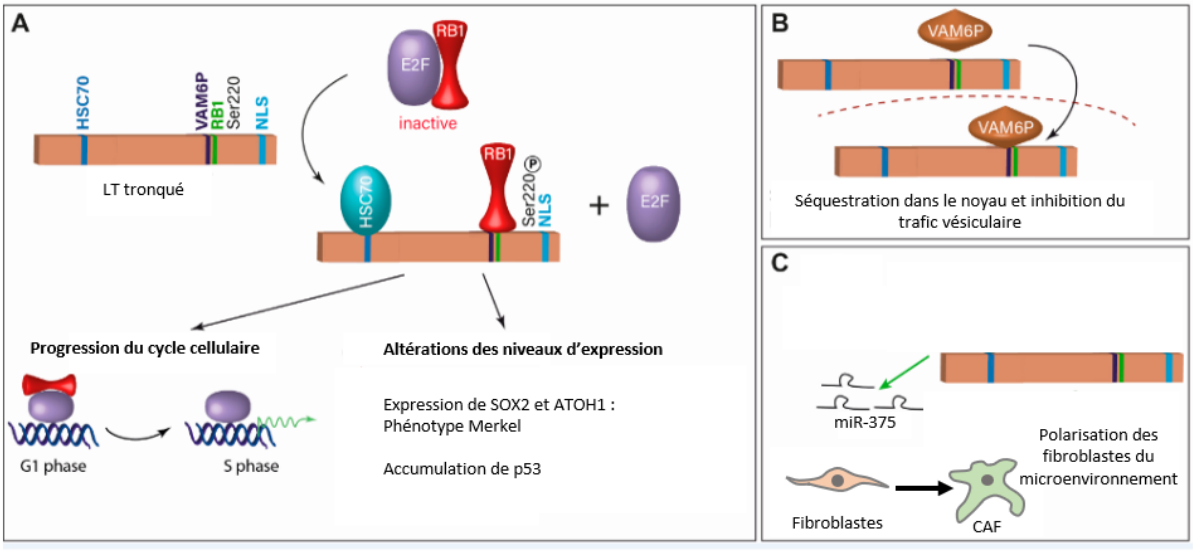


Figure 4 : Mécanismes oncogéniques associés au LT du MCPyV.

***(A)** La plupart des fonctions oncogéniques connues pour LT impliquent la séquestration de RB1. Le LT du MCPyV lie RB1 par son domaine d'interaction LXCXE et permet de relâcher les facteurs de transcriptions E2F avec l'aide des Hsc70 qui interagissent également avec LT. L'inhibition des fonctions de RB1 entrainent une perte de contrôle sur le cycle cellulaire et une progression des cellules en phase S. Cette inhibition se traduit également par l'expression de Arf, inhibiteur de MDM2 qui cause une baisse de la dégradation et donc une accumulation du suppresseur de tumeur p53. L'inhibition de RB1 est responsable de l'expression de SOX2 qui induit l'expression de ATOH1, ce qui permet l'acquisition d'un phénotype semblables aux cellules de Merkel. **(B)** LT permettrais, via son site d'interaction avec la protéine Vam6p, la séquestration de cette protéine au sein du noyau, l'empêchant de remplir sa fonction dans le trafic vésiculaire et la fusion vacuolaire. **(C)** L'expression de LT entraine l'expression du miR-375 qui est sécrété par les cellules CCM. Le miR-375 permet la polarisation des fibroblastes en CAF et forme ainsi un micro environnement pro-tumorale.*

(Adapté de Houben et al., 2023)

2.3.1.1.2 p53

La protéine la plus fréquemment mutée dans les cancers solides humains, p53, se trouve également affectée par certaines protéines oncovirales. La protéine p53 est un facteur de transcription qui joue un rôle central dans la régulation du cycle cellulaire. Elle est activée

suite à divers stress génotoxique et va, en réponse, induire la transcription de gènes impliqués dans l'arrêt du cycle cellulaire. Les gènes dont l'expression est induite par p53 codent également pour des protéines impliquées dans la réparation de l'ADN ou, dans le cas de dommages trop importants, dans l'apoptose (Hassin and Oren, 2023). Le LT du virus SV40 interagit avec la protéine p53 et est capable d'inactiver ce suppresseur de tumeur (Ludlow, 1993). C'est également le cas des protéines E6 et E6AP du papillomavirus humain. Ces deux protéines ne sont pas capables d'interagir seules avec p53. E6 et E6AP forment un complexe et c'est sous cette forme que les protéines virales sont capables d'interagir avec p53 et d'inhiber sa fonction de suppresseur de tumeur (Martinez-Zapien *et al.*, 2016). De la même manière, LT a été observé en interaction avec la protéine p53. Cette interaction n'a cependant pu être observée que pour la forme complète de la protéine. La forme tronquée tLT ne semble pas avoir conservé l'interaction avec ce suppresseur de tumeur (Borchert *et al.*, 2014). A l'inverse, l'expression de tLT dans les CCM semble même avoir un effet activateur de la voie p53. Park *et al.* ont montré que l'inhibition de RB1 mentionnée précédemment entrainait une augmentation de la protéine ARF. Cette protéine possède une fonction inhibitrice de la protéine MDM2 qui est la principale E3 ligase impliquée dans la dégradation de p53. Ainsi, l'expression du tLT de MCPyV au sein des CCM ne résulte pas en une inactivation de p53 mais au contraire en l'accumulation de cette protéine par inhibition de sa dégradation (Park *et al.*, 2018). Les antigènes T du MCPyV conservent cependant un contrôle sur la voie p53 grâce à la protéine sT. Ce contrôle est détaillé plus tard dans ce manuscrit. Comme vu précédemment, la forme complète de LT qui n'est pas exprimée dans les CCM entraine elle aussi une activation de la voie p53. En effet, l'expression de cette protéine virale sans troncation de la partie C-terminale déclenche les voies de réponses aux dommages de l'ADN. Cette activation se poursuit jusqu'à la phosphorylation activatrice de p53 (Li *et al.*, 2013).

2.3.1.1.3 SOX2 et ATOH1

Lors de test de ciblages des antigènes T au sein de lignées CCM-VP co-cultivée avec des kératinocytes, Harold *et al.* ont observé une diminution de l'expression des protéines ATOH1 et SOX2. Les protéines ATOH1 (« Atonal homolog 1 ») et SOX2 (« Sex-determoning region Y-box 2") sont des facteurs de transcription essentiels dans la différenciation des cellules de Merkel chez la souris. L'expression d'ATOH1 chez la souris entraine l'apparition de nouvelles cellules de Merkel. L'association de cette expression avec celle de sT a permis l'apparition

d'agrégats similaires à ce qui est observé dans les CCM (Ostrowski *et al.*, 2015; Verhaegen *et al.*, 2017). L'étude menée par Harold *et al.* a pu montrer l'apparition d'un phénotype neuronal pour des cellules CCM co-cultivée avec des kératinocytes suite à l'inhibition de l'expression des antigènes T. Ce changement phénotypique s'est accompagné d'une baisse des niveaux d'expression des facteur SOX2 et ATOH1 (Harold *et al.*, 2019). L'expression de l'une des formes de tLT de 339 acides aminés est capable d'induire une augmentation de l'expression de SOX2 qui induit à son tour l'expression de ATOH1. Cette promotion de l'expression est probablement due au lien entre tLT et RB1. En effet, le tLT muté sur son site d'interaction avec RB1 ne partage pas cet effet sur l'expression de SOX2 et ATOH1. L'expression de sT seul n'a pas non plus permis cette augmentation d'expression (Pietropaolo, Prezioso and Moens, 2020).

2.3.1.1.4 miR-375

La chaine d'activation initiée par tLT et son domaine de liaison à RB1 entraine l'activation de l'expression de SOX2 qui induit à son tour l'expression de ATHO1 (Harold *et al.*, 2019). Le facteur de transcription ATOH1 est responsable de l'activation transcriptionnelle au niveau de promoteurs présentant un site de liaison E47 ainsi qu'une E-box. Le promoteur du miR-375 rassemble donc les conditions pour être activé par ATOH1. Outre la compatibilité de son promoteur avec ATOH1, il a été montré que l'expression du tLT entrainait une augmentation de l'expression de ATOH1 et également du miR-375 (Fan, Gravemeyer, *et al.*, 2020). Comme pour ATOH1 et SOX2, l'expression de sT n'est pas suffisante pour augmenter l'expression du miR-375. Le miR-375 est un micro ARN ayant le potentiel de cibler les ARNm de protéines des voies de signalisations Hippo et de la transition épithelio-mesenchymateuse. Les micros ARN (miARN) sont de petites séquences ribonucléiques qui après maturation par l'enzyme Drosha et le complexe DICER mesurent de 21 à 23 nucléotides. Ces courtes séquences sont complémentaires d'ARNm qu'ils vont réguler. La régulation se fait via l'intégration au complexe RISC, les miARN orientent ce complexe vers ses ARNm cibles. La traduction des protéines codées par ces ARNm est alors inhibée par blocage de l'initiation de la traduction ou en induisant la dégradation de l'ARNm (O'Brien *et al.*, 2018). miR-375 est l'un des miARN les plus abondants dans les CCM et est même utilisé comme marqueur pour distinguer les CCM d'autres cancers de la peau. Son expression est dérégulée dans de nombreux cancer et est considéré comme oncogénique. Sa fonction dans le cadre des CCM a

été étudiée par Fan *et al.*. Ils ont observé que l'inhibition de l'expression de miR-375 n'avait pas d'effet sur la croissance de plusieurs lignées CCM-VP et ne modifiait que peu le transcriptome des cellules. Il semble donc que le miR-375 ne remplisse pas de rôle oncogénique clair au sein des cellules CCM (Fan, Zebisch, *et al.*, 2020).

Des travaux menés par la même équipe ont permis d'identifier la fonction de cette expression accrue dans les CCM. Le miR-375, bien que n'ayant pas de fonction évidente au sein des cellules CCM, se trouve être sécrété. Ils ont mis en évidence un lien entre la sécrétion de ce miRNA et la polarisation des fibroblastes de l'environnement tumoral. L'apport de miR-375 par les cellules CCM-VP aux fibroblastes entraine leur polarisation en CAF (« Cancer Associated Fibroblasts », fibroblastes associés au cancer) (Fan, Zebisch, *et al.*, 2020). Les CAF correspondent à des fibroblastes ayant acquis un phénotype pro-tumoral (**Figure 4c**). La polarisation est généralement déclenchée par la sécrétion des facteurs comme le TGF-β ou l'IL6 par les cellules tumorales (Ping *et al.*, 2021). Les CAF surexpriment des métalloprotéases matricielles et sécrètent des facteurs pro-angiogéniques. Ainsi, l'induction de la polarisation des fibroblastes en CAF permet aux cellules tumorales de créer un micro-environnement favorable à leur croissance et à dissémination. Les travaux de Fan *et al.* ont donc mis en évidence un rôle du miR-375 dans la polarisation des fibroblastes une fois sécrété par les cellules CCM-VP. Dans ce cadre, la polarisation semble être menée par l'inhibition de l'expression des protéines RBPJ (« Recombination Signal Binding Protein For Immunoglobulin Kappa J Region ») et p53 par le miR-375. Ces deux protéines sont fortement impliquées dans la régulation de la polarisation. La polarisation des fibroblastes exposés au miR-375 a été confirmée par des observation phénotypiques et par l'induction de CWCL2, IL-1β et α-SMA. Enfin ces observations réalisées en culture cellulaire ont été confirmées par hybridation *in situ* au sein d'échantillons cliniques CCM. Ces échantillons présentaient bien le miR-375 au sein de CAF, phénotype associé à une faible expression de RBPJ et p53 ainsi qu'une forte expression de α-SMA (Fan *et al.*, 2021). Une étude récente a rapporté une corrélation entre la présence de CAF dans le micro environnement tumoral et la survie sans progression de la maladie. La présence de CAF est ici associée à une plus faible infiltration immunitaire par les lymphocytes T CD8+ et à une plus faible survie sans progression (Zheng *et al.*, 2022).

En plus des effets oncogéniques au sein de la cellule, l'inhibition de RB1 par LT semble donc avoir également une influence sur le micro environnement tumorale. La chaine de

régulation passant par RB1, SOX2, ATOH1 et aboutissant à la sécrétion du miR-375 et la polarisation des fibroblastes en CAF permet aux CCM-VP de façonner un environnement pro-tumoral favorable au développement du cancer.

2.3.1.1.5 Vam6p

L'exon 2 de LT code un motif de liaison à la protéine Vam6p en amont de la séquence LXCXE nécessaire à l'interaction avec RB1. Vam6p est une protéine de tri cytoplasmique également appelée VPS39. Elle joue un rôle dans la fusion entre vacuoles et endosomes et également dans la fusion homotypique entre vacuoles. Cette protéine interagit avec les protéines Rab sur les membranes vacuolaires et régule l'assemblage du complexe SNARE permettant la fusion membranaire. Vam6p est ainsi importante pour le trafic vésiculaire notamment lysosomal et la fusion vacuolaire (Caplan *et al.*, 2001). A la fois LT pleine taille et tLT interagissent avec la protéine Vam6p (Liu *et al.*, 2011). Cette interaction a pour conséquence la séquestration de la protéine Vam6p par la LT/tLT dans le noyau (**Figure 4b**). Ainsi, cette protéine ne peut pas, en présence de LT/tLT, assurer son rôle dans le trafic vésiculaire. Parmi les nombreux points de troncation de tLT observés, certains entrainent la perte partielle ou totale du NLS. Dans les cas de troncation en amont du NLS, tLT peut être observé dans le cytoplasme mais est toujours détecté dans le noyau (Martel-Jantin *et al.*, 2012; Houben *et al.*, 2015). Il semble donc que la petite taille de tLT lui permette de pénétrer dans le noyau et de remplir ses fonctions. La capacité de ces formes dépourvues de NLS à séquestrer Vam6p dans le noyau n'est cependant pas connue. Le rôle de l'interaction avec Vam6p dans le cadre CCM n'est pas connu, il est envisageable que cette fonction de LT soit impliquée dans le cycle viral plutôt que l'oncogenèse.

2.3.1.2 sT

La protéine sT du MCPyV joue un rôle essentiel dans l'activité transformante du virus. Contrairement à la protéine LT qui est fréquemment mutée et tronquée dans les tumeurs, sT est généralement retrouvée intacte avec une séquence sauvage. sT partage sa région N-terminale avec LT mais possède en C-terminal une séquence unique de 106 acides aminés contenant des motifs critiques pour ses fonctions oncogéniques. Des études *in vitro* et sur modèles animaux suggèrent que sT serait plus directement impliquée que LT dans les processus de transformation tumorale, tandis que LT est plutôt requis pour la croissance des cellules tumorales. Verhaegen *et al.* ont en effet pu observer en modèle murin que la co-expression de sT et ATOH1 menait à l'acquisition d'un phénotype CCM indépendamment de la présence de tLT. L'expression de tLT seule n'a quant à elle pas permis la transformation des cellules (Verhaegen *et al.*, 2017). Il faut retenir ici que l'acquisition de phénotype CCM par l'expression de sT n'a pu être réalisée qu'avec induction concomitante d'ATOH1, mécanisme naturellement induit par tLT comme évoqué précédemment.

2.3.1.2.1 PP2A/PP4C

Le petit T possède dans sa région spécifique une séquence d'interaction avec les protéines phosphatases PP2A et PP4C. De nombreuses voies de signalisation sont régulées par phosphorylation de leurs différents facteurs. Ainsi, l'inhibition de phosphatases peut participer à l'émergence d'un cancer. Ces protéines sont d'ailleurs fréquemment les cibles de protéines virales. C'est notamment le cas de l'antigènes petit T du SV40 qui contribue à la transformation des cellules via l'inhibition de la phosphatases PP2A (Sablina and Hahn, 2008). De manière surprenante, cette fonction de sT ne semble pas se transposer dans les CCM-VP. En effet, il a été rapporté que la mutation du domaine d'interaction entre sT et PP2A n'affectait pas son potentiel transformant au sein de fibroblastes de rongeurs (Shuda *et al.*, 2011). L'interaction entre le sT du MCPyV et le complexe PP2A a néanmoins été observée. Cette interaction est traduite par une réduction de l'activité catalytique de PP2A. Contrairement au sT du SV40 qui bloque l'assemblage des sous unités B55α , B56α et B46ε, le sT du MCPyV n'empêche l'assemblage que de la sous unité B56α et pas des autres (Kwun *et al.*, 2015).

L'interaction entre le sT du MCPyV et PP4C semble en revanche avoir un effet au sein des CCM-VP. Selon une étude de Knight *et al.*, sT serait impliquée dans la promotion de la mobilité et migration cellulaire en déstabilisant les microtubules. Cette fonction serait médiée par son interaction avec PP4C. Ils ont observé une augmentation de la proportion de stathmine déphosphorylée en présence de sT. La stathmine est une protéine déstabilisatrice des microtubules dont l'activité est régulée par phosphorylation. La forme déphosphorylée de la stathmine est capable de séquestrer les dimères de tubuline (**Figure 5a**). Cette séquestration résulte en une stimulation du désassemblage de la tubuline polymérisée en microtubules. La phosphorylation de cette protéine entraine une fort perte d'affinité entre les dimères de tubuline (Di Paolo *et al.*, 1997). Cette hausse de proportion de la forme déphosphorylée a été liée à l'interaction entre sT et PP4C par utilisation de sT muté sur les sites d'interaction avec PP2A et PP4C. La mutation du site d'interaction avec PP2A n'a pas influencé la capacité de sT à déstabiliser les microtubules tandis que la mutation du site d'interaction avec PP4C a aboli cette fonction. Ainsi, sT stimule la mobilité et migration cellulaire en induisant déstabilisation des microtubules via la déphosphorylation de la stathmine (Knight *et al.*, 2015).

L'intégration entre sT et PP4C a également été reliée à la réponse immunitaire à travers la voie NFκB. La voie d'activation de la transcription des gènes cibles de NFκB est dépendante de la phosphorylation de IκB qui entraine sa dégradation et met fin à la séquestration de NFκB dans le cytoplasme. Ainsi, NFκB peut transloquer dans le noyau et activer la transcription de ses gènes cibles. Le sT de MCPyV est capable de s'associer avec PP4C et d'ainsi favoriser l'interaction entre le complexe NEMO (« NFκB Essential Modulator », Modulateur essentiel de NFκB) et le complexe PP4C-PP4R1. Ce recrutement entraine l'inhibition de la phosphorylation de IκB et empêche donc la translocation de NFκB vers le noyau (**Figure 5a**)(Griffiths *et al.*, 2013). Cette régulation de la voie NFκB par l'oncoprotéine sT peut être impliquée dans la régulation de la réponse immunitaire innée des cellules infectées. L'interférence de sT dans la voie NFκB est probablement liée à l'inhibition de la réponse anti-virale plutôt qu'à l'oncogenèse. En effet, une forte présence de NFκB a pu être observé dans le noyau des CCM-VP suggérant que cette inhibition de translocation n'est pas prépondérante en contexte néoplasique (Matsushita *et al.*, 2017). De plus, l'activité de NFκB a été reliée dans

les cancers à l'évasion immunitaire notamment par la stimulation de l'expression de PD-L1 qui est l'un des mécanismes principal retrouvé dans les CCM (Antonangeli *et al.*, 2020).

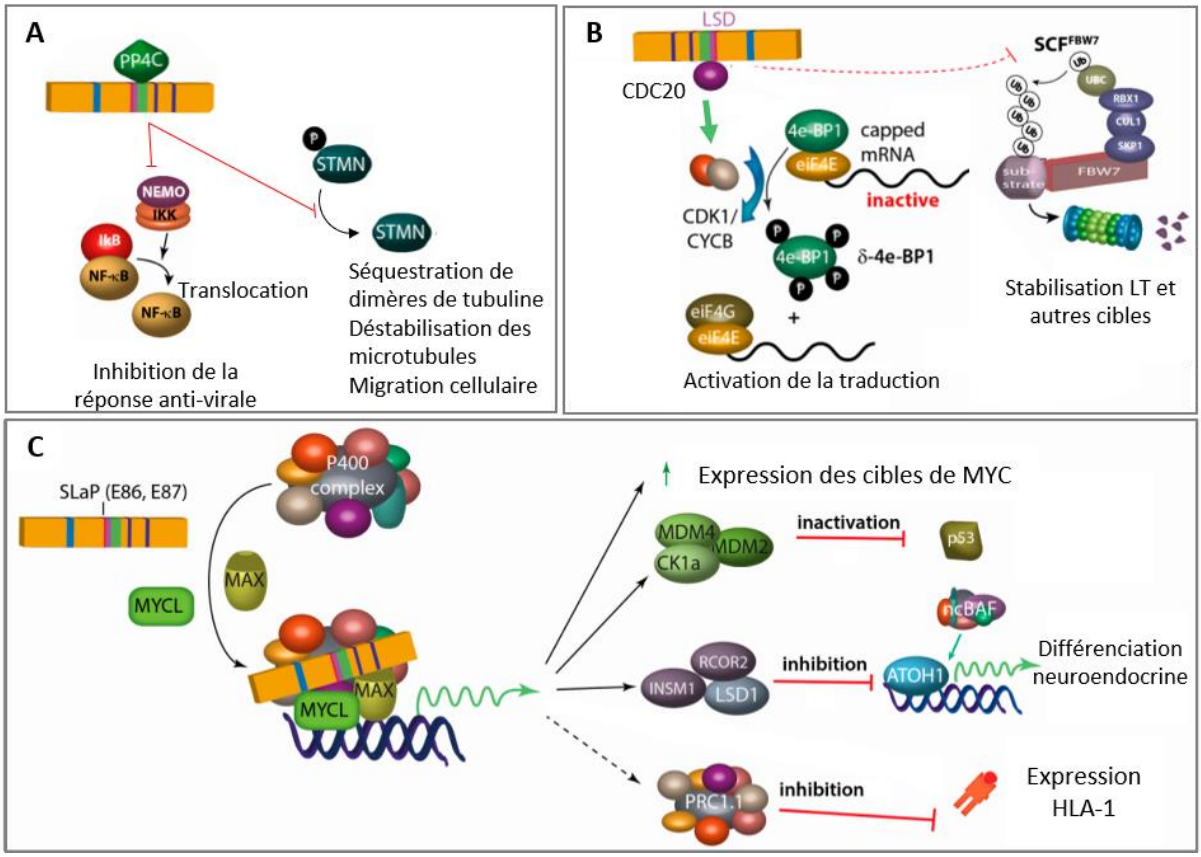


Figure 5 : Mécanismes oncogéniques connus impliquant le sT du MCPyV

***(A)** sT forme un complexe avec MYCL et EP400 et influence l'expression de leurs cibles. La transcription des cibles de MYC est ainsi activée. L'expression de MDM2 est également activée par le complexe ce qui entraine la dégradation de p53 et donc la voie de signalisation passant par p53 et menant à l'arrêt du cycle cellulaire. La lysine démethylase LSD1 est également surexprimée suite à l'interaction sT-MYCL-EP400. L'action de LSD1 entraine l'inhibition de l'expression des cibles de ATOH1 permettant d'éviter la différenciation des cellules. Enfin, l'action de sT au sein de ce complexe mène à une augmentation d'expression de composants du PRC1.1 qui est un régulateur négatif de l'expression du HLA-1 participant ainsi à l'évasion immunitaire. **(B)** Par son domaine de stabilisation de LT (LSD), sT est capable d'interagir avec diverses E3 ubiquitine ligases. L'interaction de sT avec FBXW7 est décrite comme inhibitrice et stabilisatrice de LT et d'autres cibles endogènes de FBXW7. Le LSD permet également une interaction avec CDC20 qui active le complexe CDK1/CyclineB1. L'activation de ce complexe entraine l'hyper-phosphorylation du facteur 4E-BP1. L'hyper-phosphorylation de 4E-BP1 empêche la séquestration de eiF4E et permet ainsi l'initiation de la traduction. **(C)** sT interagit avec la phosphatase PP4C et favorise son interaction avec le complexe NEMO ce qui inhibe la phosphorylation de IκB et empêche la libération de NFκB et sa translocation dans le noyau. L'interaction entre sT et PP4C mène également à une déphosphorylation de la stathmine. La forme déphosphorylée cette protéine lie les dimères de tubuline ce qui a pour effet de déstabiliser le cytosquelette. La déstabilisation des microtubules a un effet favorable pour la mobilité et la migration cellulaire.*

(adapté de Houben et al., 2023)

2.3.1.2.2 4E-BP1

Etant démontré que le potentiel transformant de sT n'était pas affecté par sa capacité à se lier à PP2A, il a été proposé que cette propriété de sT soit plutôt liée au facteur 4E-BP1 (« Eukaryotic Translation Initiation Factor 4E Binding Protein », Protéine liant le facteur d'initiation de la traduction eucaryote 4E). Shuda *et al.* ont ainsi identifié une interaction entre ce facteur d'initiation de la traduction et sT. Ils ont également mis en évidence la phosphorylation de 4E-BP1 en réponse à l'expression de sT mais pas celle de LT (Shuda *et al.*, 2011). Cette phosphorylation empêche la liaison de 4E-BP1 à eIF4E (« Eukaryotic Translation Initiation Factor 4E », Facteur d'initiation de la traduction eucaryote 4E). Ainsi, l'assemblage du complexe d'initiation n'est pas inhibé par 4E-BP1. Dans cette étude, la phosphorylation de 4E-BP1 induite par sT est reliée à son interaction avec PP4C (**Figure 5b**) (Shuda *et al.*, 2011). La suite de leurs travaux a cependant identifié un mécanisme différent régulant la phosphorylation de cette protéine. Le LSD (« LT Stabilization Domain », domaine de stabilisation de LT) de sT permet une interaction avec la protéine CDC20, ce qui active le complexe CDK1/CYCB1 (Cyclin dependant kinase 1 / Cyclin B1). C'est enfin CDK1 qui va hyper-phosphoryler 4E-BP1, empêchant ainsi l'inhibition de l'assemblage du complexe d'initiation de la traduction. La dérégulation de l'initiation de la traduction dépendante de la coiffe, comme c'est le cas ici, a déjà été observé comme mécanisme impliqué dans l'oncogenèse (Bjornsti and Houghton, 2004; Shuda *et al.*, 2015).

2.3.1.2.3 MycL-EP400

L'une des interactions majeures de sT est sa liaison au facteur de transcription MYCL. sT recrute MYCL au sein du super complexe protéique EP400, composé d'une quinzaine de protéines et incluant le partenaire habituel de MYC, MAX. sT est capable de modifier la composition du complexe de transcription EP400 entraînant des changements profonds de l'expression génique dans les cellules. L'importance de l'interaction sT-MYCL est suggérée par la perte de viabilité des cellules de carcinome à cellules de Merkel lors de l'inhibition de l'expression de MYCL ou de la formation d'hétérodimères MYCL-MAX. De plus, une forme mutante de sT incapable de lier MYCL perd son potentiel transformant (Cheng *et al.*, 2017).

Parmi les effets de sT médiés par MYCL-EP400, on peut citer l'activation transcriptionnelle de MDM2. MDM2 est une E3 ubiquitine ligase qui assure la dégradation de

la protéine p53 (**Figure 5c**). Ce complexe entraine également de manière indirecte l'activation de MDM4, une autre E3 ubiquitine ligase ciblant p53. Cette activation se fait par la promotion de la transcription du gène de la protéine kinase 1α qui va pouvoir phosphoryler et activer MDM4 (Pei, Zhang and Zheng, 2012). Ainsi, l'activité de sT permet par son action sur le complexe MYCL-EP400 de contrecarrer l'effet antitumoral de l'activation de la voie p53 par LT détaillée plus tôt. Ce mécanisme d'inhibition permet de contrer efficacement la voie p53. En effet, un niveau de p53 mesurable n'a été observé que dans 7% des CCM-VP lors d'une étude sur des patients Finnois réalisée par Waltari *et al.* Ils ont également constaté un niveau de p53 généralement supérieur dans les formes virus négative ainsi qu'un corrélation entre le nombre de copies virales intégrées et le niveau du niveau de p53 (Waltari *et al.*, 2011).

D'autre part, le complexe sT-MYCL-EP400 stimule également l'expression de la protéines LSD1 (« Lysine-specific histone demethylase 1A », déméthylase d'histone spécifique des lysines 1A). En retirant les mono et di- méthylation de la lysine 4 de l'histone H3, LSD1 est responsable d'une compaction de la chromatine aux *loci* ciblés. Ces modulations épigénétiques répriment la transcription de gènes impliqués dans la différenciation cellulaire dépendante du facteur de transcription ATOH1 (Park *et al.*, 2020). Ainsi, l'expression d'ATOH1 est nécessaire pour permettre à sT d'exercer son potentiel transformant et obtenir des cellules ayant un phénotype CCM. En revanche, afin d'éviter une différenciation complète en cellules de Merkel, sT est capable d'inhiber l'expression de certaines cibles de ATOH1 via l'expression de LSD1.

Enfin, une étude récente menée au sein de lignées CCM dérivées de patients visant à caractériser ces lignées nouvellement générées a rapporté une nouvelle conséquence de l'interaction de sT avec les protéines de ce complexe. Par une analyse des gains et pertes de fonctions provoqués par l'extinction de différents gènes à l'échelle du génome, les auteurs ont identifié MYCL et PRC1.1 (« non-canonical Polycomb Repressive Complex 1.1 ») comme étant régulateurs de l'expression du HLA-I (Lee *et al.*, 2022). Le HLA-I (« Human Leukocyte Antigen class I », antigène leucocytaire humain de classe I) est une protéine de surface ayant la fonction de présentation d'antigènes. Cette protéine permet aux lymphocytes T d'identifier les cellules infectées ou cancéreuses et de déclencher la réponse cytotoxique. L'expression de cette protéine est inhibée dans les CCM-VP, ce qui participe probablement à l'évasion immunitaire de ces cancers (Paulson *et al.*, 2014). Les auteurs de cette étude montrent que

l'inhibition de l'expression du HLA-I est médiée par sT et son interaction avec MYCL ainsi que par le PRC1.1. Ainsi, les auteurs proposent l'utilisation d'inhibiteurs d'USP7 afin de rétablir l'expression de HLA-I, ce qui pourrait permettre une meilleure prise en charge des CCM par le système immunitaire. La protéine USP7 stabilisant de nombreuses protéines, le degré d'implication du PRC1.1 demeure néanmoins à clarifier (Lee *et al.*, 2022).

2.3.1.2.4 FBXW7

Outre les sites d'interaction avec les protéines PP2A et PP4C, la partie spécifique à sT contient également le LSD. Ce domaine permet la stabilisation de l'autre oncoprotéine virale codée par la région précoce, LT. Le mécanisme par lequel sT stabilise LT par son LSD est controversé. Dans un premier temps, Kwun *et al.* ont identifié une interaction entre sT et l'E3 ubiquitine ligase FBXW7 (« F-Box And WD Repeat Domain Containing 7 »). Cette interaction sT-FBXW7 aurait un effet inhibiteur sur l'E3 ubiquitine ligase (Kwun, Chang and Moore, 2017). Ainsi, LT mais également les autres cibles de FBXW7 s'en trouveraient stabilisées (**Figure 5b**). FBXW7 est la sous-unité de ciblage, conférant la spécificité de substrat, du complexe SCF/C auquel il est associé. La machinerie d'ubiquitination est une cascade de modification post-traductionnelle régulée par l'action séquentielle de trois composants : E1, E2 et E3. E1 correspond à l'enzyme activatrice de l'ubiquitine et va comme son nom laisse le supposer, cliver le propeptide et fixer l'ubiquitine, l'activer et la transférer sur E2. Il existe ~40 Ub E2 différentes, qui sont des enzymes de conjugaison de l'ubiquitine et servent de relai pour l'ubiquitine reçue de la partie E1. Les E2 vont ensuite transférer l'ubiquitine sur les protéines cibles (substrats) pour induire -dans la majorité des cas- leurs dégradation protéasomale. On estime qu'il y a ~600 Ub E3 chez l'homme. Les Ub E3, appelées ubiquitine ligases par abus de langage, sont des protéines sans activité enzymatique intrinsèque qui assurent le lien entre une Ub E2 d'un côté et un substrat spécifique de l'autre. Chaque E3 reconnait donc un éventail de cible défini et va lier ses protéines substrats selon différents motifs de reconnaissances (Scheffner, Nuber and Huibregtse, 1995). Ici, la composante E3 est représentée par le complexe SCF/FBXW7 dans lequel FBXW7 est la protéine qui assure la spécificité de substrat. FBXW7 compte parmi ses cibles des proto-oncogènes majeurs comme c-Myc, c-Jun ou encore la cycline E. Les niveaux de ces trois protéines se trouvent augmentés suite à l'expression de sT dans des cellules murines (Verhaegen *et al.*, 2015). La même équipe a également identifié

d'autres E3 ubiquitine ligases qui interagiraient avec sT, cdc20 et β-TrCP (Kwun *et al.*, 2015; Kwun, Chang and Moore, 2017).

Une autre équipe a cependant mis en doute ce mécanisme potentiel. Dye *et al.* ont étudié l'interaction entre sT et FBXW7 et ne sont pas parvenus à l'observer par co-immunoprécipitation. Ils se sont également intéressés aux conséquences fonctionnelles d'une inhibition de FBXW7 par sT. Ils n'ont pas non plus identifié de stabilisation des cibles de FBXW7 comme c-Myc suite à l'expression de sT. La stabilisation de LT dépendante de l'expression de sT a cependant été confirmée par leurs expériences mais selon un mécanisme qui semble indépendant de FBXW7. De manière intéressante, les auteurs ont observé que l'expression de LT et non sT entrainait une légère diminution de la quantité de FBXW7 sans que la quantité de son ARNm ne soit affectée (Dye *et al.*, 2019). Il est donc envisageable que le MCPyV cible bel et bien la protéine FBXW7 selon un mécanisme non élucidé.

2.3.1.3 Effets du MCPyV sur la régulation de la chromatine

Les difficultés d'identification de la lignée d'origine des CCM viennent essentiellement de leur dérive phénotypique. L'expression de marqueurs de lignées différentes, leur morphologie ou encore leur métabolisme ne sont ni cohérents ni en majorité attribuable à un seul type cellulaire normal. Ceci est dû à un profil d'expression génique (donc protéique) profondément dérégulé prenant sa source dans des altérations de la régulation de la chromatine. Les antigènes T induisent donc une reprogrammation épigénétique dont les mécanismes restent mal définis. Nous avons jusqu'ici évoqué des mécanismes transformants discrets impliquant des protéines bien identifiées. Ici, nous présentons les altérations épigénétiques inhérentes à la présence de MCPyV dans les CCM.

2.3.1.3.1 Méthylation des histones

La méthylation des histones est une marque épigénétique majeure qui peut être responsable de répression comme d'activation transcriptionnel selon le résidu méthylé. Il s'agit ici de l'ajout d'un groupement méthyl sur un résidu lysine libre des histones. Les effets de la méthylation sont modulés par le résidu méthylé ainsi que par le nombre de groupement méthyle ajouté. Les lysines des histones peuvent être mono-, di- ou tri-methylées. Les enzymes régulant cette modification sont les histones méthyltransférases (HMT) et histones

déméthylases (HDM). La méthylation des histones est une composante de ce qui est appelé le code histone (Jenuwein and Allis, 2001). Il s'agit de l'ensemble des modifications post-traductionnelles observées sur ces protéines. La théorie du code histone implique que cet ensemble de modifications soit responsable de modulations de l'expression des gènes à différents *loci* de l'ADN. En substance, les différentes modifications post-traductionnelles vont impacter l'interaction entre les histones et l'ADN, entrainant soit une compaction soit une ouverture de la chromatine. La compaction de la chromatine aura un effet de répression transcriptionnel tandis que l'ouverture rendra la chromatine accessible pour les facteurs de transcription et les polymérases et entrainera la transcription et l'expression des gènes concernés. Ce code histone est impliqué dans la régulation épigénétique qui régit l'expression différentielle entre différents types cellulaires possédant le même génome. Schématiquement, il permet la distinction entre deux états de chromatine : l'euchromatine qui est constitué de régions décompactées et accessibles pour la transcription, et l'hétérochromatine qui est constitué de régions du génome condensées et qui ne sont pas transcrites. Le passage d'un état à l'autre est principalement régulé par les modifications post-traductionnelles des histones. Les acétylations des lysines 14 et 36 de l'histone H3 sont associées à l'euchromatine en induisant la décompaction de la chromatine et induisent donc une activation transcriptionnelle. A l'inverse, les méthylations des lysines 9 et 27 de l'histone H3 sont associées à la condensation de la chromatine en hétérochromatine, la rendant inaccessible pour la transcription, et sont donc associées à une répression transcriptionnelle. L'hétérochromatine peut être constitutive ou facultative, la première est la plus condensée, inclue des régions répétées non-codantes et des gènes qui ne sont pas exprimés. La seconde est moins condensée que la première et est plus dynamique, elle peut être condensée/décondensée selon les besoins de la cellule pour réguler l'expression génétique (Jenuwein and Allis, 2001).

L'action de certaines HMT est apparue comme importante dans le remodelage épigénétique observé dans les CCM. C'est le cas de EZH2 (« Enhancer of Zest Homolog 2 ») qui est la sous-unité catalytique du complexe PRC2 (« Polycomb Repressive Complex 2 »). Ce complexe est responsable de la trimethylation de la lysine 27 de l'histone H3 (H3K27me3). Le PRC2 est donc un répresseur transcriptionnel qui entraine la compaction de la chromatine. La protéine EZH2 est fréquemment mutée ou surexprimée dans les cancers et représente un

régulateur de l'expression génique d'intérêt majeur dans la biologie du cancer (Kim and Roberts, 2016). Une forte expression de cette protéine a été observée au sein des CCM (Harms *et al.*, 2017). Plusieurs études contradictoires ont été publiée sur les impacts de la présence de EZH2 et des niveaux de H3K27me3. Une étude a montré qu'une diminution de H3K27me3 était associée aux formes CCM-VP (Busam *et al.*, 2017). Cependant, il a été publié plus tard que les formes CCM-VN étaient associées à une diminution de H3K27me3 (Matsushita *et al.*, 2019). Enfin, une étude postérieure à celles-ci a observé une perte de H3K27me3 au sein des CCM indépendamment du statut viral de la pathologie (Gujar *et al.*, 2021). Récemment, nous avons participé à des travaux où il a été observé que l'expression de EZH2 était plus forte dans les formes virus positives que dans les formes virus négatives des CCM. De plus, une forte présence de EZHIP, une protéine inhibitrice de EZH2, a été mesurée au sein des CCM-VN. Le traitement de cellules CCM par le DZNEP, un inhibiteur chimique de la protéine EZH2 a entrainé une inhibition sélective de la croissance des CCM-VP (Durand *et al.*, 2023). La perte de viabilité des cellules causée par l'inhibition de EZH2 a été identifiée comme étant reliée à la perte de répression de l'expression de la protéine SIX1 (Gartin *et al.*, 2022).

2.3.2 Carcinomes à cellules de Merkel virus négatif, une autre entité pathologique ?

La découverte du MCPyV en 2008 a entrainé la séparation des CCM en deux catégories : les formes virus positives et négatives. La première étude à identifier ce virus l'a identifié chez 80% des patients. Cette prévalence a ensuite été confirmé par méta-analyse de 35 publications portants sur le sujet (Wijaya *et al.*, 2022). Cette différence d'étiologie entre ces deux catégories de CCM suggère des mécanismes oncogéniques différents entre ces deux formes. Alors que l'oncogenèse des formes CCM-VP est supportée par les antigènes T, les formes CCM-VN semblent plus dépendantes de mutations accumulées au fil du temps. En effet, les différentes études sur les mutations observées dans les CCM ont conclu à un faible fardeau mutationnel dans les formes CCM-VP contrastant avec le fort taux de mutations observé dans les formes CCM-VN. L'étude menée par Goh *et al.* sur les mutations au seins des CCM a révélé une séparation en deux extrêmes comparés aux autres cancer épithéliaux. Ils ont observé une dichotomie du nombre de SSNV par exome entre les CCM-VP et -VN (« somatic single nucléotide variation », mutations somatiques de nucléotide unique). Les SSNV correspondent à des mutations survenant au sein de cellules somatiques (non germinales) et sont généralement causées par des facteurs environnementaux comme

l'exposition à des produits chimiques, des radiations UV et/ou par des facteurs héréditaires. Ces mutations sont évaluées sur l'exome, qui correspond à l'ensemble des exons du génome. Il est ainsi possible d'identifier les mutations qui impacteront les séquences des protéines résultant de la traduction de ces exons. Cette distribution bimodale a mis en évidence deux groupes au sein des CCM, le premier présentant un nombre de SSNV par exome nettement inférieur à ce que l'on observe dans autres cancers épithéliaux. Le second groupe quant à lui présentait à l'inverse, un nombre de SSNV bien supérieur à ce qui a été observé au sein des autres cancers épithéliaux. Le premier groupe présentait un nombre médian de 12,5 SSNV par exome et correspondait aux formes CCM-VP, tandis que le second groupe était composé des CCM-VN et présentait un nombre médian de 1121 SSNV par exome. Cette même étude a également identifié des mutations canoniques dans les formes CCM-VN, avec RB1 et p53 fréquemment mutés, ce qui n'était jamais le cas dans les CCM-VP. Ils ont également relevé une forte concentration de mutations dans les gènes codant des protéines impliquées dans la modification de la chromatine et dans les voies de signalisation et réparation des dommages de l'ADN (Goh *et al.*, 2016).

Les mutations observées au sein des CCM-VN sont composées à 85% de transitions cytosine->thymine (Harms *et al.*, 2015). Ce type de mutation est caractéristique de l'expositions aux rayons UV. Au sein d'un dimère CC, CT ou TC, l'exposition aux rayons UV peut entrainer une déamination spontanée de la cytosine, la transformant ainsi en uracile. Lors de la réplication, une adénine sera associée à cette uracile et une transition C->T sera complètement effectuée après deux cycles de réplication, lorsqu'une thymine sera associée à cette adénine. Dans le cas des îlots CpG où la cytosine est méthylée, cette déamination entraine la formation de thymine directement et non d'uracile (Sugiyama and Chen, 2019).

Il apparait donc que là où les CCM-VP sont dépendants de la présence du MCPyV, les formes CCM-VN sont principalement issues de mutations induites par les rayons UV. L'absence de signature UV et le faible taux de mutations des formes CCM-VP, même chez des patients généralement âgés, suggère alors une cellule d'origine différente pour les deux formes. L'émergence d'une forme CCM-VN dépendante des UV ne peut ainsi se faire que dans une cellule exposée aux UV et donc dans les couches les plus superficielles de la peau (**Figure 6**). Cette localisation est incompatible avec les profils mutationnels observés pour les CCM-VP. Il a donc été mis en avant que les CCM-VN ont probablement une origine épidermale,

incompatible avec ce qui est observé pour les CCM-VP malgré les similarités cliniques et phénotypes des formes -VP et -VN (Sunshine *et al.*, 2018). Un changement de dénomination distinguant ces deux forme a même été proposé avec « Merkel Type sarcoma » désignant les formes CCM-VP et « Squamous cell carcinoma, Merkel type » désignant les formes CCM-VN (Nirenberg *et al.*, 2020).

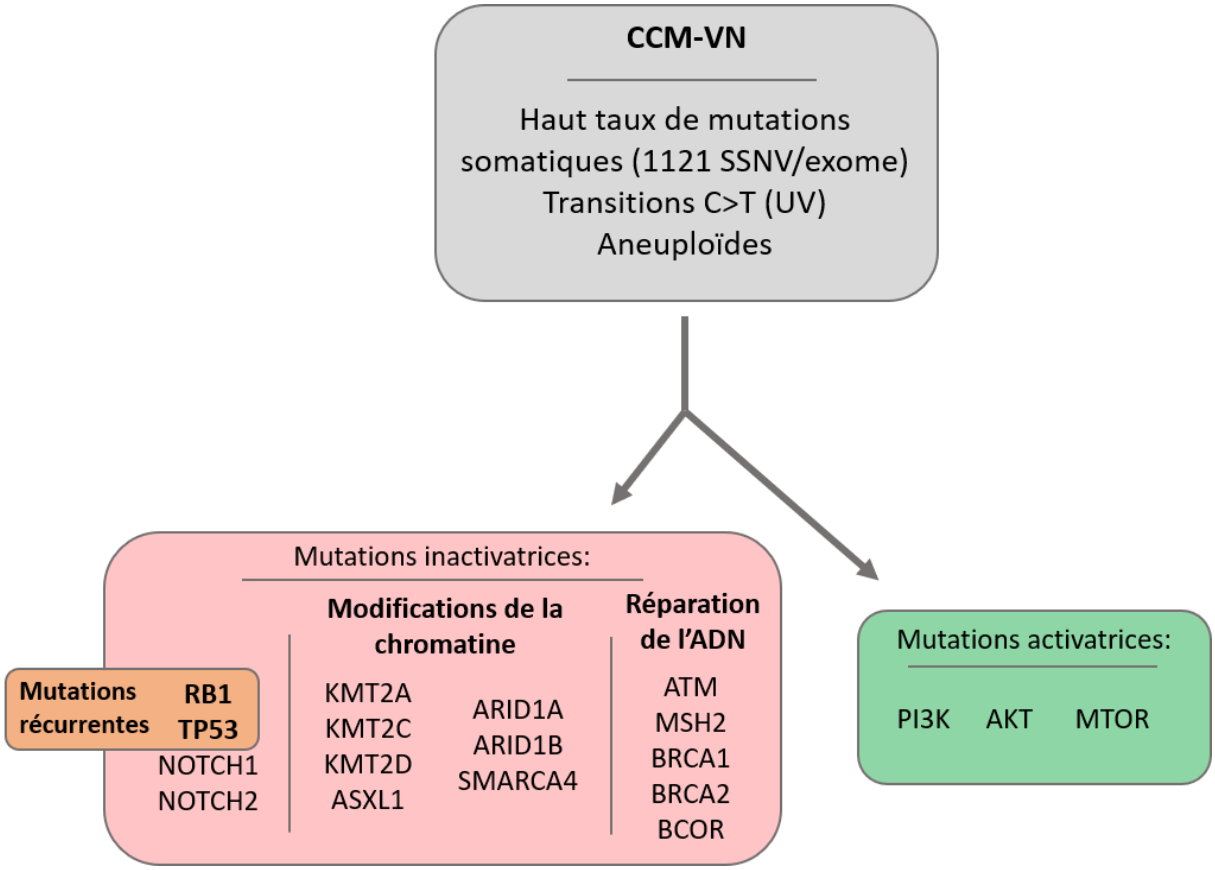


Figure 6 : Mécanismes oncogéniques au sein des CCM-CN

L'oncogenèse des CCM-VN repose sur une accumulation de mutations activatrices ou inactivatrices. Les mutations observées sont généralement des transitions C->T caractéristiques des mutations causées par l'exposition aux rayons UV. La majorité des CCM-VN présente des mutations inactivatrices sur les séquences de RB1 et TP53. Les protéines codées par les gènes fréquemment mutés peuvent être réparties en trois grandes catégories : les protéines influant sur la structure de la chromatine, les protéines impliquées dans la réponse aux dommages de l'ADN dont les gènes portent des mutations inactivatrices, et les protéines impliquées dans la prolifération cellulaire dont les gènes portent des mutations activatrices.

2.4 INTERACTOMIQUE PROTEIQUE

Le potentiel oncogénique du MCPyV repose en grande partie sur les interactions formées par les protéines virales exprimées. L'étude des interactions formées entre les protéines virales et les protéines de la cellule hôte est un moyen efficace d'identifier les mécanismes mis en place par ces protéines pour le cycle viral ou dans le cas des oncovirus, pour promouvoir l'émergence et le maintien d'un cancer. Les techniques d'études d'interactions protéiques se divisent en deux catégories. D'un côté les techniques supervisées où l'interaction entre deux protéines connues est testée, et de l'autre, les techniques semi-supervisées où les différentes interactions formées par une protéine d'intérêt sont étudiées, sans préjuger des partenaires à identifier. Les principes des différentes techniques d'étude supervisée et semi-supervisée sont représentées de manière schématique dans la **Figure 7**.

2.4.1 Interactomique supervisée

Les techniques dites supervisées d'étude d'interactions protéine-protéine (IPP) sont généralement utilisées pour confirmer une interaction entre deux protéines. Elles nécessitent un postulat de base sur une interaction entre deux partenaires connus.

2.4.1.1 PLA in situ

La PLA *in situ* (« Proximity Ligation Assay ») est un test d'interactions supervisé. Cette méthode permet de tester l'interaction entre deux protéines et dans un même temps de localiser cette interaction au sein de la cellule par microscopie confocale.

Cette technique permet l'observation d'interactions protéiques au sein de cellules fixées ou de coupes tissulaires. Il est donc possible de savoir si l'interaction entre les deux protéines d'intérêt est possible dans leur environnement naturel. Cette méthode repose sur l'utilisation d'anticorps primaires dirigés spécifiquement contre les deux protéines d'intérêt. Les anticorps primaires utilisés sont produits dans deux espèces différentes afin de pouvoir utiliser des anticorps secondaires différents. Ces anticorps secondaires sont couplés de manière covalente à des sondes ADN. Ces sondes sont de courtes séquences synthétiques qui sont utilisées pour la détection (Söderberg *et al.*, 2006).

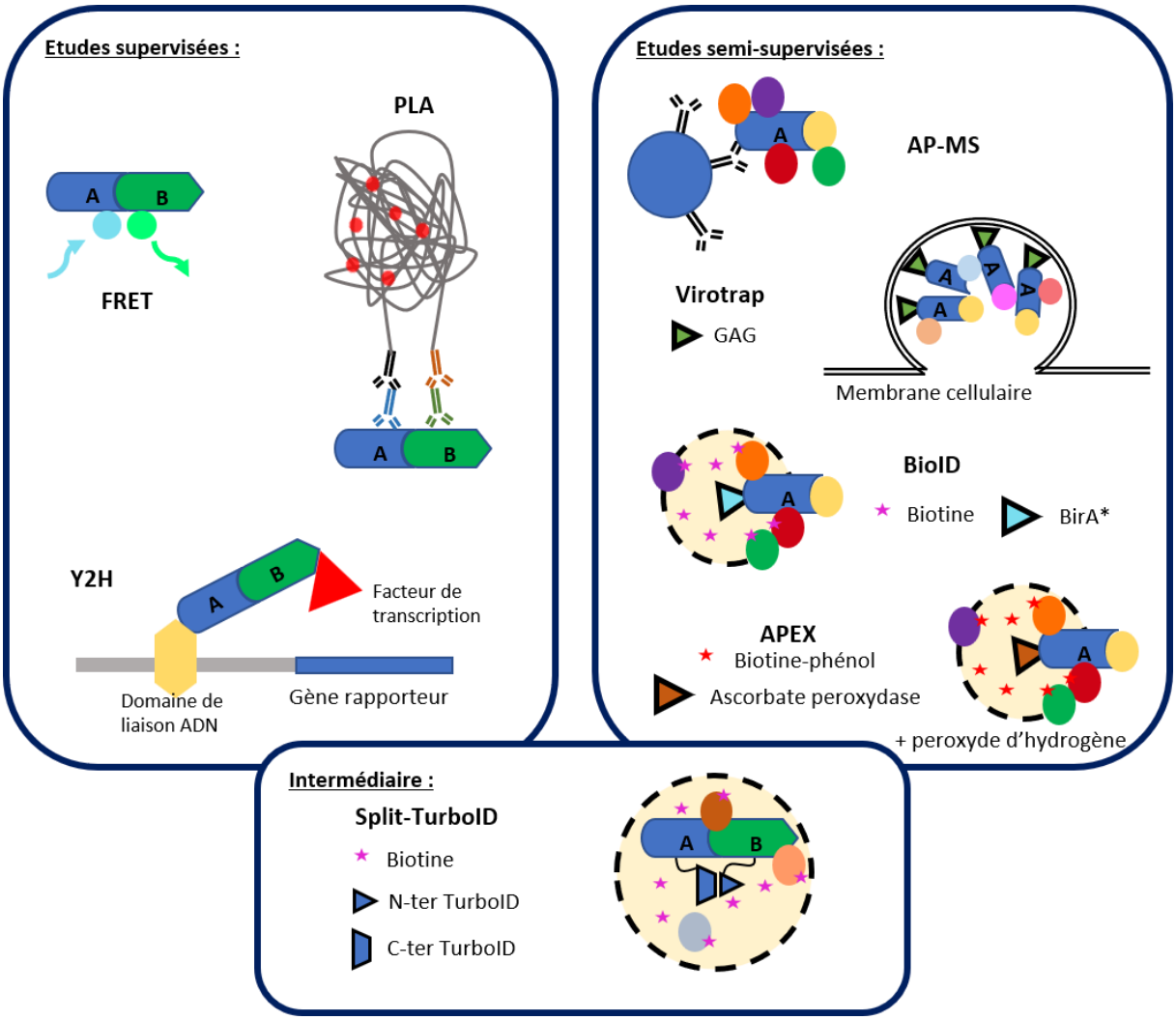


Figure 7 : Aperçu des techniques d'interactomique supervisées et semi-supervisées.

Les techniques d'études d'interactions supervisées permettent des tests d'interactions binaires entre deux protéines connues. Les principales techniques sont représentées par le FRET, la PLA et le Y2H. Le FRET repose sur l'émission de signal par un fluorophore accepteur après excitation du donneur si les deux sont suffisamment proches (3-6 nm). La PLA repose sur l'utilisation de sondes ADN couplées à des anticorps secondaires qui seront amplifiées et détectables si les protéines sont suffisamment proches (40 nm). Le Y2H se base sur l'expression d'un gène rapporteur activée par l'interaction entre deux protéines de fusions portant des domaines fonctionnels.

Les techniques d'études d'interactions semi-supervisées permettent une approche sans a priori pour identifier les interacteurs d'une protéine d'intérêt. La purification d'affinité est la plus conventionnelle et repose sur la captation d'une protéine d'intérêt avec ses interacteurs après une lyse douce préservant ces interactions. Le virotrap permet la formation de pseudo particules virales contenant une protéine d'intérêt et ses interacteurs qui sont purifiés sans besoin de lyser les cellules. Le BioID et l'APEX reposent sur des principes similaires avec biotinylation des protéines présentes dans un rayon de 10 à 20 nm de la protéine d'intérêt puis purification des protéines biotinylées après une lyse agressive.

Le Split-TurboID peut être considérée comme une technique d'étude entre le supervisé et le semi-supervisé. Elle repose sur la complémentation fonctionnelle d'une biotine ligase si les deux protéines d'intérêts sont suffisamment proches (supervisé). La biotinylation qui est alors engagée permet l'identification des protéines de l'environnement moléculaires du complexe ainsi formé (semi-supervisé).

Après fixation des cellules, l'hybridation des anticorps primaires à lieu (**Figure 8**). Les deux anticorps primaires utilisés vont reconnaitre les protéines cibles et fixer les épitopes. Le couple d'anticorps secondaires est ajouté dans un second temps. Chacun des anticorps secondaires va reconnaitre et s'hybrider à l'un des deux anticorps primaires. Ce sont, comme précisé précédemment, ces anticorps secondaires qui sont liés de manière covalente à des sondes d'ADN synthétique. Dans les cas où les épitopes reconnus par les anticorps primaires sont suffisamment proches, les sondes ADN pourront s'hybrider à des connecteurs et former une matrice d'ADN circulaire en présence d'une ligase. La distance doit pour cela être inférieure à 40 nm. Une polymérase et des oligonucléotides intégrant des bases modifiées par un fluorochrome sont ensuite ajoutés. La molécule d'ADN circulaire couplée aux anticorps va servir de matrice pour la RCA (« Rolling Circle Amplification », amplification en cercle roulant). Le produit est une longue molécule d'ADN en tandem, répétant la séquence circulaire comme pour une rotative, liée de manière covalente à l'un des deux anticorps (Söderberg *et al.*, 2008). Les oligonucléotides s'hybrident sur cet amplicon et la détection de l'interaction peut ensuite se faire par microscopie à fluorescence. Lorsque l'interaction protéique est confirmée, des *foci* de fluorescence correspondant à l'ADN amplifié sont observables au site d'interaction au sein de la cellule.

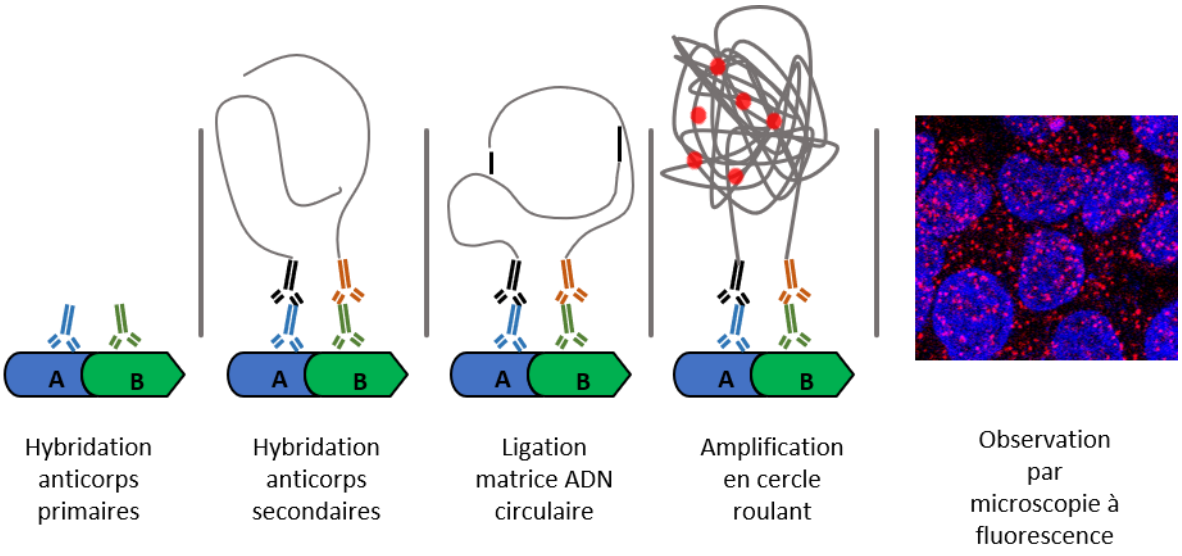


Figure 8 : Représentation schématique du principe de la PLA in situ.

Deux anticorps primaires issus d'espèces différentes s'hybrident sur les protéines d'intérêt. Les anticorps secondaires couplés à des sondes ADN s'hybrident sur les anticorps primaires. Les sondes ADN sont reliées par l'intermédiaire de connecteurs et forment une matrice d'ADN circulaire. La polymérase est ajoutée permettant la réalisation de l'amplification en cercle roulant. Des sondes ADN marquées s'hybrident sur les amplicons concatémériques. Le signal fluorescent et détecté par microscopie.

Cette technique est utilisée pour sa sensibilité et sa spécificité. Il est possible en adaptant les temps d'élongation de détecter une interaction rare entre deux protéines. Très peu de matériel cellulaire est nécessaire pour réaliser la PLA *in situ* (Liang and Yuan, 2022). De plus cette méthode permet la détection d'interaction dans l'environnement cellulaire natif des protéines, ce qui offre une visualisation pertinente des interactions formées *in vivo*. Contrairement à d'autres techniques détaillées ultérieurement, la PLA est utilisable pour étudier les interactions de tous types de protéines, et est donc applicable aux protéines cytosoliques comme aux protéines membranaires ou nucléaires (Bobrich *et al.*, 2013; Poulard *et al.*, 2021). Cette technique présente cependant des limites, elle repose notamment sur la disponibilité d'anticorps primaires spécifiques qui doivent être produits au sein d'espèces différentes. En cas de non disponibilité des anticorps, il est nécessaire d'exprimer une protéine tagguée ce qui retire la possibilité d'utilisation de tissus primaires. La PLA est difficilement applicable à haut débit et se limite généralement à la confirmation d'une interaction entre quelques couples de protéines d'intérêt.

2.4.1.2 Y2H

Le Y2H (« Yeast two hybrid ») est une technique permettant également l'étude des interactions entre deux protéines d'intérêt. Il s'agit d'une méthode mise au point et publiée en 1989 (Fields and Song, 1989). Cette technique utilise un système rapporteur composé de deux domaines complémentaires d'un facteur de transcription qui, ensemble, activent la transcription d'un gène rapporteur (**Figure 7**).

L'étude par Y2H passe par une construction génétique ou les gènes codant les protéines d'intérêt (A et B) sont fusionnés à des séquences codant des domaines de facteurs de transcription qui fonctionnent par complémentarité. Les domaines fusionnés sont le domaine de liaison à l'ADN qui reconnait la séquence du promoteur du gène rapporteur, et le domaine activateur qui induit l'expression du gène rapporteur en recrutant la machinerie transcriptionnelle. Les protéines de fusion sont co-exprimées dans *Saccharomyces cerevisiae*. La protéine A qui est couplée au domaine de liaison à l'ADN va pouvoir fixer le promoteur du gène rapporteur. Si la protéine A et la protéine B interagissent, le domaine d'activation qui est fusionné à la protéine B se trouvera alors à proximité du promoteur du gène rapporteur et va

pouvoir recruter la machinerie transcriptionnelle. Ainsi, lorsque les protéines A et B interagissent, le gène rapporteur est transcrit et la protéine traduite. Il existe une variété de gènes rapporteurs pouvant être utilisé comme le gène codant la GFP détectable par fluorescence ou le gène lacZ codant la β-galactosidase qui permet une coloration bleue suite à l'ajout de X-Gal quand elle est exprimée (Hamdi and Colas, 2012).

Cette technique peut être utilisée pour observer une interaction entre deux protéines mais elle peut également être utilisée par une approche matricielle. Une approche matricielle permet une étude d'interaction à haut débit. Dans ce type d'approche, le nombre de couples A et B testés simultanément peut être de quelques dizaines à plusieurs milliers (Ito *et al.*, 2001). Le développement d'approches combinant le Y2H à l'utilisation de BFG (« Barcode Fusion Genetics) et NGS(« Next-Generation Sequencing ») a permis d'augmenter le nombre de couples testés jusqu'à 2,5 millions simultanément (Yachie *et al.*, 2016). Ce type d'étude a notamment permis l'identification de certains mécanismes impliqués dans la maladie d'Huntington (Tourette *et al.*, 2014).

Le Y2H permet donc une détection d'interaction protéique *in vivo* bien que le système cellulaire soit constitué de levures. Cette technique permet des études à haut débit testant de nombreuses interactions. C'est également une technique relativement abordable pour les laboratoires équipés et elle ne nécessite pas l'existence d'anticorps spécifiques pour les protéines à tester (Ito *et al.*, 2001). Le Y2H présentent cependant certaines limitations, de nombreuses interactions peuvent ne pas être détectées à cause de l'utilisation de levures qui ne permet pas forcément le bon repliement de la protéine ou l'ajout des modifications post traductionnelles nécessaires à l'interaction (Brückner *et al.*, 2009). Certains cas de toxicité des protéines étudiées pour les levures ont également pu donner lieu à de faux négatifs. Cette technique limite également à l'étude des interactions binaires et uniquement pour les interactions directes. Elle est également inadaptée pour des protéines membranaires et plus généralement des protéines incapables de récapituler leurs fonctions quand leur localisation dans le noyau est forcée. Le Y2H ne permet pas l'identification de complexes de plus de deux protéines ou des interactions proximales.

2.4.1.3 FRET/BRET

Le FRET (« Förster Resonance Energy Transfer », Transfert d'énergie par résonnance de Förster) est une technique permettant d'identifier une interaction directe entre deux protéines en se basant comme pour le PLA, sur la distance entre deux sondes (Sekar and Periasamy, 2003). Contrairement à la PLA, le FRET nécessite l'utilisation de protéines de fusion. Pour étudier une interaction binaire entre une protéine A et une protéine B, il faut exprimer chacune de ces protéines fusionnées à un fluorophore (**Figure 7**). Les deux fluorophores utilisés doivent être choisi avec précaution. Ils représenteront un fluorophore donneur et un fluorophore accepteur. Ces deux fluorophores doivent avoir des longueurs d'ondes d'excitation et d'émission compatibles. En effet, la longueur d'onde d'émission du fluorophore donneur doit correspondre à la longueur d'onde d'excitation du fluorophore accepteur. Les protéines de fusion sont exprimées en cellules vivantes. Un stimulus lumineux est ensuite émis à la longueur d'onde adéquate pour exciter le fluorophore donneur. Si les protéines A et B sont en interaction, les deux fluorophores qui leurs sont fusionnés peuvent être suffisamment proches pour que le fluorophore transmette son énergie au fluorophore donneur par un couplage dipôle-dipôle intermoléculaire longue portée. Il faut pour cela que les deux fluorophores se trouve à une distance inferieure au rayon de Förster soit moins de 6 nm. Si les protéines sont en interaction et que le transfert d'énergie s'opère, le fluorophore accepteur va alors émettre une fluorescence à sa longueur d'onde d'émission. Ainsi, si un signal lumineux correspondant à la longueur d'onde d'émission est mesuré, le transfert d'énergie a été réalisé entre les fluorophores et les protéines A et B sont en interaction (Sekar and Periasamy, 2003).

Ce principe de transfert d'énergie par résonnance est également utilisé dans la technique de BRET (« Bioluminescence Resonance Energy Transfer », Transfert d'énergie de bioluminescence par résonance). Le principe ici est le même que pour le FRET à la différence que la première excitation ne se fait pas par émission de signal fluorescent. Dans le BRET, le fluorophore donneur est remplacé par une enzyme comme la luciférase. En présence de son substrat, la luciférase émet un signal bioluminescent. Si les protéines fusionnées à cette enzyme et au fluorophore accepteur sont en interaction, alors l'énergie de bioluminescence peut être transférée au fluorophore accepteur par résonnance comme dans le FRET (Xu, Piston and Johnson, 1999; Couturier and Deprez, 2012).

Ces techniques permettent donc l'étude d'interaction en cellules vivantes et de manière très spécifique. En effet, les conditions strictes à remplir pour avoir émission d'un signal par le fluorophore accepteur réduisent fortement le nombre de faux positifs. Cependant, contrairement au PLA, cette approche n'est pas applicable aux coupes tissulaires à cause de la nécessité de l'expression des protéines de fusion. L'auto-fluorescence et le photoblanchiment peuvent également représenter des limites du FRET.

2.4.2 Interactomique semi-supervisée

2.4.2.1 Purification par affinité couplée à la spectrométrie de masse

Conventionnellement, les techniques de purification par affinité représentent les moyens les plus utilisés pour les études d'interactomique. Dans cette catégorie, l'immunoprécipitation a notamment été largement utilisée (Lin and Lai, 2017). Cette technique repose sur la captation de la protéine d'intérêt par un anticorps spécifique couplé à une phase stationnaire comme par exemple des billes d'agarose (**Figure 7**). Pour réaliser une immunoprécipitation, les cellules subissent une lyse qui doit être douce. Une lyse trop agressive entrainerait la perte des interactions protéine-protéine. La phase stationnaire contenant l'anticorps spécifique de la protéine d'intérêt est ajoutée au lysat. Cet anticorps fixé va s'hybrider à sa protéine cible récupérant également les protéines associées à cette protéine d'intérêt. Plusieurs lavages sont nécessaires pour éliminer le bruit de fond de facteurs liés de façon non-spécifique. Historiquement, les protéines étaient ensuite éluées, séparées en SDS-PAGE et analysées par western-blot. Actuellement, les protéines purifiées sont ensuite digérées en peptides par une enzyme come la trypsine et sont identifiées par spectrométrie de masse couplée à la chromatographie en phase liquide.

Pour pallier à ces potentiels problèmes liés à l'utilisation d'anticorps dirigés contre la protéine d'intérêt, d'autres techniques utilisant la purification d'affinité ont vu le jour. L'ajout par rapport à l'immunoprécipitation consiste en l'utilisation de protéines portant une étiquette, un « tag ». Le fait d'utiliser un tag permet à la fois de purifier des protéines pour lesquelles aucun anticorps n'est disponible et également d'augmenter la sensibilité, la reproductibilité et la spécificité de la purification. Le FLAG-tag est l'un des tags les plus rependu en biologie. Dans ce cas de figure, la protéine d'intérêt fusionnée à un polypeptide (DYKDDDDK) est exprimée en cellule vivante. Ce FLAG-tag est reconnu de manière très

spécifique par un anticorps ce qui permet son utilisation pour purifier les protéines tagguées ainsi que les protéines interagissant avec celles-ci sans nécessiter un anticorps spécifique de la protéine d'intérêt (Hopp *et al.*, 1988). Afin d'augmenter encore la spécificité de détection des interacteurs protéiques, une méthode de « Tandem Affinity Purification » ou TAP-Tag a été développée (Puig *et al.*, 2001). Cette approche repose sur l'utilisation d'une protéine de fusion comportant plusieurs tags permettant la réalisation de deux étapes successives de purification biochimique. La protéine d'intérêt est génétiquement fusionnée à une séquence composée classiquement de deux domaines de liaisons aux IgG provenant de la protéine A de *Staphylococcus aureus*, suivi d'un peptide de liaison à la calmoduline CBP (« Calmoduline Binding Peptide »). Ces deux domaines sont séparés par un site de clivage par la protéase TEV (« Tobacco Etch Virus »). La protéine de fusion et ses partenaires sont purifiés dans un premier temps par affinité sur des billes couplées à des IgG. Après lavage, l'élution est ensuite réalisée par protéolyse au niveau du site reconnu par la TEV protéase. Une seconde étape de purification sur billes de calmoduline permet alors d'isoler la protéine d'intérêt et ses interacteurs les plus stables et spécifiques. Ces protéines sont ensuite digérées en peptides par une protéase comme la trypsine et identifiées en spectrométrie de masse. Cette approche TAP-Tag, en tandem, offre donc une spécificité accrue par rapport à une purification simple, et permet d'identifier des partenaires protéiques interagissant de façon très affine (Puig *et al.*, 2001). Cela dit, seules les interactions les plus stables sont détectables et ce protocole a un rendement faible, nécessitant une importante quantité de matériel au départ.

Ce type d'approche semi-supervisée permet l'étude sans *a priori* des interactions formées par une protéine d'intérêt et peut se montrer particulièrement efficace dans l'élucidation de mécanismes de protéines virales. La première description des interactions protéines-protéines formées entre un protéome viral pathogène et les protéines de la cellule hôte a été publiée en 2012 sur le VIH à l'aide de la purification par affinité couplée à la spectrométrie de masse (Jäger *et al.*, 2012). Cependant ces techniques de purification par affinité ne permettent pas l'observations d'interactions transitoires et ne conservent que les interactions les plus faibles durant le processus de lyse et de purification. De plus les contraintes relatives à l'utilisation d'une lyse douce pour conserver les interactions ne sont pas compatibles avec l'étude de protéines situées dans des compartiments peu solubles comme le cytosquelette, la membrane ou la chromatine.

2.4.2.2 Virotrap

L'une des limitations des techniques de purification par affinité est l'étape de lyse cellulaire associée à la nécessité de préserver les interactions durant le processus de lyse. Le Virotrap est une technique permettant de s'affranchir de ce processus de lyse pour récupérer les protéines d'intérêt ainsi que leurs interacteurs. Les inventeurs de cette techniques sont partis du postulat que l'incorporation de complexes protéiques au sein de VLP (« Virus-like Particules », pseudo-particules virales) permettrait de piéger la protéine d'intérêt avec ses interacteurs et les protègerait durant la purification de ces protéines (**Figure 7**)(Eyckerman *et al.*, 2016).

Cette technique se base sur la capacité de GAG à former des pseudo-particules virales s'échappant des cellules. La première étape consiste donc en l'expression d'une protéine de fusion rassemblant une protéine d'intérêt et la protéine GAG. Une fois exprimée, cette protéine de fusion va interagir avec les partenaires de la protéine d'intérêt et la partie GAG de cette protéine va mener à la formation de multimères à proximité de la membrane plasmique. Cette formation multimérique sur la membrane plasmique entraine un bourgeonnement de la membrane et la formation de VLP. Ainsi, la protéine de fusion, GAG-protéine d'intérêt et ses interacteurs se retrouvent piégés dans les VLP bourgeonnant de la cellule. Ces VLP peuvent ensuite être purifiés par ultra-centrifugations et le contenu des VLP analysé. Dans le but de simplifier le processus de purification, Eyckerman *et al.* ont coexprimé avec la protéine de fusion, la protéine VSV-G sous sa forme sauvage et sous une forme tagguée. Cette protéine est présente à la surface des VLP. Sa présence permet l'utilisation d'anticorps couplés à une phase stationnaire pour purifier les pseudo-particules virales sans avoir recours à l'ultracentrifugation. Les auteurs ont confirmé par Western-Blot, la capacité de cette méthode à identifier des interactions binaires connus comme l'interaction HRAS-RIF1. Sur un panel de 92 interactions connus, 30% ont été retrouvées par utilisation du Virotrap. Il est également possible d'utiliser cette technique de manière semi-supervisée en remplaçant l'étape de Western-Blot par une analyse par spectrométrie de masse en tandem (Eyckerman *et al.*, 2016).

Cette approche permet donc la détection d'interaction binaire et l'identification de nouvelles interactions formées dans leur environnement natif. La purification des protéines ne requiert pas la lyse des cellules et donc ne risque pas la perte des interactions. Cependant

cette méthode présente également des limites, le nombre de faux positifs peut être important suite à l'incorporation de protéines non reliées à la protéine d'intérêt au sein des VLP. Enfin le Virotrap est plus adapté à l'étude de protéine cytosolique, peu de données sont disponibles sur l'étude de protéines nucléaires ou d'organelles qui sont probablement plus difficilement transportées à la membrane plasmique avec leurs interacteurs.

2.4.2.3 BioID

Le développement d'une autre approche a également permis de s'affranchir de cette nécessité de conserver les interactions durant le processus de lyse. Il s'agit d'une technique d'étude d'interactions de proximité, le BioID. Cette technique se base sur l'utilisation d'une enzyme bactérienne provenant d'*Escherichi coli*, BirA. Cette protéine est une biotine ligase qui biotinyle l'une des sous unités de l'acetyl-CoA carboxylase. Cette enzyme est capable d'activer la biotine, elle va pour cela former l'espèce réactive biotinyl-5'-AMP (Adénosine MonoPhosphate) à partir de biotine et d'ATP (Adénosine TriPhosphate). Cette forme activée est retenue au niveau du site actif de l'enzyme qui va ensuite la transférer sur une lysine libre de la protéine ciblée. Dans le cadre du BioID, une forme BirA* mutée est utilisée. La mutation consiste en une substitution de l'arginine 118 en glycine. Ce variant de BirA* possède une affinité pour le biotinyl-AMP plus faible que la forme sauvage de BirA selon un mécanisme incertain impliquant probablement un réarrangement des liaisons hydrogène. Cette perte d'affinité entraine le relargage de cette forme activée de la biotine dans l'environnement proche de l'enzyme. Le biotinyl-AMP étant une forme très réactive et peu stable, elle va se lier de manière covalente à une Lysine libre dans un rayon de diffusion libre estimé à 10 à 20 nm autour de BirA* (**Figure 7**) (Roux *et al.*, 2012).

Afin d'appliquer cette propriété de BirA* aux études d'interactions, Roux *et al.* ont mis au point le BioID. Ici, une construction génétique codant pour une protéine de fusion entre la protéine d'intérêt et BirA* est formée (**Figure 9**). Cette protéine de fusion est exprimée au sein de cellules vivantes. De la biotine est supplémentée au milieu pour un marquage généralement d'une durée de 18 à 24h. Durant ce temps, la protéine d'intérêt va former des interactions avec ses partenaires protéiques, et BirA* qui lui est fusionnée va activer la biotine qui va se fixer de manière covalente aux protéines proximales. Le marquage à la biotine se fait donc sur les protéines, ou interacteurs proximaux, dans un rayon de 10-20 nm de la protéine

de fusion, qu'il s'agisse d'interactions stables ou transitoires. Ce type d'approche permet donc le marquage des interacteurs d'une protéine d'intérêt en cellule vivante dans son environnement natif. De plus, un avantage majeur de ce marquage covalent à la biotine est la possibilité de s'affranchir de la nécessité de conserver les interactions protéiques durant le processus de lyse et de purification. Le BioID permet l'utilisation d'une lyse agressive plus adaptée à la solubilisation de protéines de compartiments peu accessibles pour les purifications par affinité classiques comme le cytosquelette, les membranes ou la chromatine. Suite à la lyse des cellules, les protéines biotinylées sont purifiées à l'aide d'une phase stationnaire couplée à de la streptavidine. La streptavidine lie la biotine avec une grande spécificité et une grande sensibilité, l'affinité de cette protéine pour la biotine est supérieure à celle d'un anticorps pour sa cible. Il s'agit de l'une des interactions non covalentes les plus fortes présente dans la nature avec une constante de dissociation (Kd) d'environ $10^{-}14$ $mol.L^{-1}$ (Roux *et al.*, 2012).

Le BioID a déjà permis la cartographie d'importants réseaux d'interactions comme celui liant les 58 protéines présentes à l'interface des centrosomes et des cils (Gupta *et al.*, 2015). L'identification de l'implication du régulateur épigénétique EP300 dans la régulation effectuée par SOX2 sur la différenciation cellulaire et la croissance tumorale a également été permis par l'utilisation du BioID (Couzens *et al.*, 2013). Cette technique a même permis l'identification de nouveaux interacteurs du proto-oncogène c-Myc déjà bien étudié depuis les années 1980 (Dingar *et al.*, 2015). Le BioID est également adapté à l'étude des virus et a été utilisé pour identifier les interactions formées par les protéines virales du virus Zika (Coyaud *et al.*, 2018). Nous avons récemment contribué à la découverte par BioID d'interactions formées par les protéines virales d'un nouvel oncovirus, HPV42 , afin d'identifier les mécanismes dans lesquels elles sont impliquées (Leiendecker *et al.*, 2023). Cette technique permet donc l'identification des interacteurs proximaux d'une protéine au sein d'une cellule vivante. Le marquage covalent à la biotine permet l'utilisation d'une lyse agressive adaptée à l'identification d'interacteurs de compartiments peu solubles. La réalisation du BioID nécessite cependant un temps de marquage relativement long qui n'est pas compatible avec l'étude d'interactions ayant une cinétique rapide. De plus, il est envisageable que la taille de l'enzyme fusionnée à la protéine d'intérêt puisse poser problème quant au repliement correct de la protéine ou à la formation normale de ses interactions. Il est également important de

souligner le fait que les interacteurs identifiés par cette technique sont des interacteurs proximaux. Comme précisé plus tôt, les interactions directes sont identifiables qu'elles soient stables ou transitoires, et les protéines se trouvant dans l'environnement proche (< 10 nm) de la protéine d'intérêt seront également identifiées.

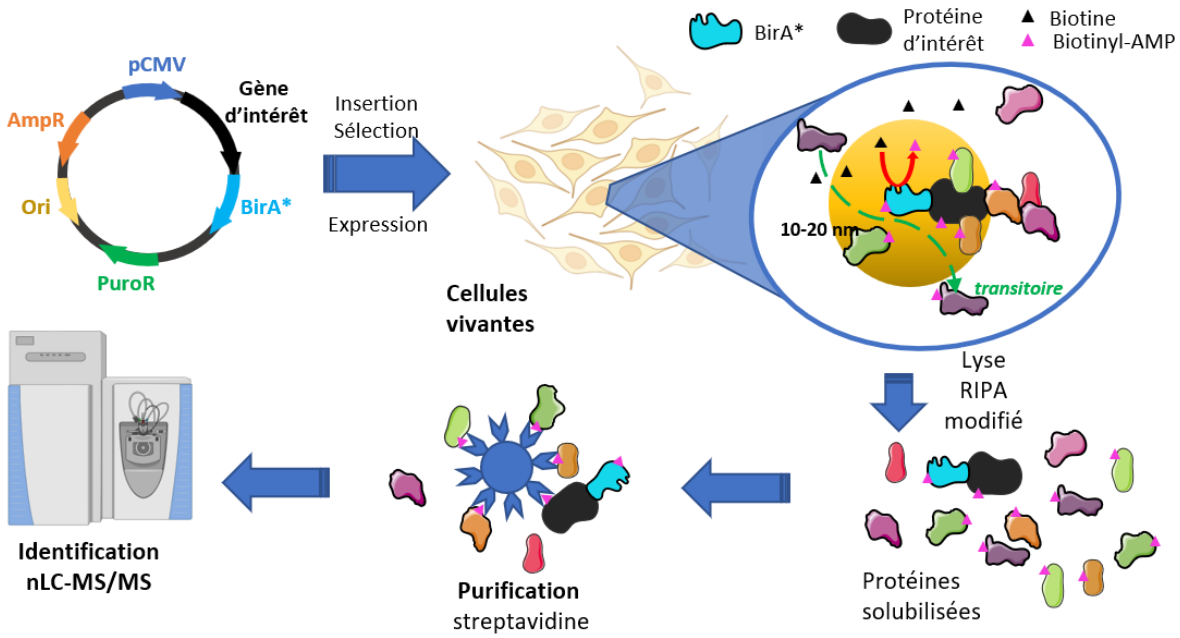


Figure 9 : Etude d'interaction protéique proximale par BioID

Le gène de la protéine d'intérêt est fusionné à celui de BirA, la construction est insérée de manière stable dans le génome de la cellule. Les cellules ayant intégré la construction sont sélectionnées par l'ajout d'antibiotiques. La protéine de fusion est exprimée de manière inductible et en présence de biotine. Les interacteurs proximaux sont marqués de manière covalente par la Biotine grâce à l'action de BirA*. Les cellules sont lysées par un tampon RIPA modifié et les interacteurs sont purifiés grâce à l'affinité de la streptavidine pour la biotine. Les protéines sont ensuite digérées à la trypsine et identifiées par LC-MS/MS*

Dans l'optique d'améliorer cette technique et de palier à certains défauts, d'autres variants de BirA* ont été développés. Branon *et al.* ont réalisés des expériences de mutagenèse évolutive sur cette enzyme. Ils ont ainsi conçu deux nouvelles formes de biotine ligase présentant des propriétés intéressantes pouvant être utilisées en interactomique : TurboID et miniTurboID. L'intérêt principal de ces deux formes est l'accroissement de l'activité catalytique. En effet, le temps de marquage à la biotine nécessaire pour l'obtention d'un signal est réduit à moins d'une heure avec un signal utilisable dès dix minutes de marquage tandis qu'un minimum de huit heures de marquage est requis dans la forme classique du BioID. Ainsi, ces deux formes permettent la réalisation d'études d'interactomiques plus rapidement et surtout l'observation de changements suite à l'ajout d'un stimulus ou à un moment précis du

cycle cellulaire sur cellules synchronisées. La forme minTurboID permet quant à elle une biotinylation plus importante que BirA* et présente également une taille qui est 20% inférieure à la forme utilisée en BioID (Branon *et al.*, 2018). La même équipe a cependant rapporté trois ans plus tard que la forme miniTurboID était moins stable que ses homologues (Qin *et al.*, 2021). Les auteurs à l'origine de l'élaboration du TurboID ont également observé que sa forte activité catalytique associée aux concentrations de biotine supplémentée au milieu de culture causait une forte réduction de viabilité cellulaire pour un temps de marquage prolongé. Une autre équipe s'est tournée vers la métagénomique et l'utilisation d'un algorithme de reconstruction d'enzyme ancestrale pour concevoir une nouvelle enzyme correspondant à un potentiel ancêtre de BirA, nommé AirID. Les tests menés par Kido *et al.* ont montré le potentiel de cette enzyme à biotinyler les partenaires connus d'une protéine d'intérêt (Kido *et al.*, 2020). Ils ont également observé une activité catalytique de l'enzyme supérieure à celle de BirA et inférieure à celle de TurboID. L'un des principaux intérêts de cette technique réside dans la non toxicité de l'utilisation de AirID même sur des temps prolongés. De plus les auteurs ont également rapporté la possibilité d'utiliser pour AirID des concentrations de biotines nettement inférieures à celles utilisées pour BioID ou TurboID (de 50 µM à 0,5 µM), voire même en utilisant la biotine endogène. Ainsi, la forme AirID serait plus adaptée pour des études nécessitant des marquages sur de longues périodes et des expériences *in vivo*, chez l'animal (Kido *et al.*, 2020).

2.4.2.4 Split-BioID

Dans la lancée du développement des techniques d'études d'interaction de proximité comme le BioID, une variante utilisant la complémentation fonctionnelle a été développée. Cette variante se base sur l'utilisation de deux protéines d'intérêt comme protéines « appâts » (De Munter *et al.*, 2017). La biotine ligase mutée BirA* est ici séparée en deux domaines fonctionnels qui sont inactifs individuellement. Pour la réalisation du Split-BioID, chacun de ces domaines est fusionné à l'une des protéines d'intérêt. En principe, aucune biotinylation n'est engendrée par l'expression de ces protéines individuellement. Ces deux protéines de fusion doivent être coexprimées en cellules vivantes. Lorsque les protéines d'intérêt interagissent entre elles, les deux domaines de BirA* peuvent s'assembler et l'activité biotine ligase est restaurée par complémentation. Ainsi, l'activation de la biotine en biotinyl-AMP peut avoir lieu comme lors du BioID classique (**Figure 7**). De la même manière, cette forme

réactive de la biotine va se lier de manière covalente à des résidus lysine de protéines dans un rayon de 10 à 20 nm. Les protéines biotinylées peuvent être purifiées et identifiées par le même procédé que pour le BioID. Ainsi, la présence de biotinylation est une confirmation de l'interaction entre deux protéines d'intérêt. De plus, cette approche permet l'étude de l'environnement moléculaire au site d'interaction entre ces deux protéines et peut permettre l'identification des partenaires interagissant avec ces protéines lorsqu'elles sont sous forme de complexe (De Munter *et al.*, 2017; Schopp *et al.*, 2017). Plus récemment, la technologie TurboID a également été adaptée à la complémentation fonctionnelle. De la même manière que pour le split-BioID, l'enzyme TurboID est séparée en deux domaines chacun fusionné à une protéine d'intérêt dont l'interaction entraine la complémentation des domaines de TurboID. Bien qu'inferieur à l'enzyme TurboID, l'activité catalytique accrue du split-TurboID permet une biotinylation efficace et rapide des protéines de l'environnement moléculaire d'une interaction. Cette technique a pu être utilisée pour étudier la composition protéique spécifiquement à l'interface entre le réticulum endoplasmique (RE) et la mitochondrie via l'utilisation de protéines spécifiques de la membrane du RE et de la mitochondrie. Afin d'avoir l'interaction nécessaire à la complémentation fonctionnelle, les auteurs ont ajouté des domaines FKBP et FRB permettant une dimérisation induite par la rapamycine (Cho *et al.*, 2020).

2.4.2.5 APEX

Une autre méthode utilisant un principe similaire à celui du BioID a été développée peu de temps après cette technique. Il s'agit de l'APEX (Rhee *et al.*, 2013). Cette technique requiert également un marquage covalent à la biotine mais sans utilisation d'une biotine ligase. C'est ici l'ascorbate peroxydase qui est utilisée. Cette enzyme permet l'oxydation de biotine-phénol en biotine-phénoxyl en présence de peroxyde d'hydrogène. Le biotine-phénoxyl est un composé très peu stable ayant une demi-vie très courte. Une fois oxydée, il va se lier de manière covalente à des tryptophanes, tyrosines, cystéines ou histidines libres des protéines à proximité. Dans ce cadre, le rayon de biotinylation et d'environ 20 nm autour du site d'oxydation du biotine-phénol. Pour l'utilisation de cette technique, la protéine d'intérêt est donc exprimée en fusion avec l'ascorbate peroxydase et en présence de biotine-phénol. La biotinylation des interacteurs proximaux est amorcée par l'ajout de peroxyde d'hydrogène (**Figure 7**). Ce composé est toxique pour les cellules et le temps de marquage

utilisé pour l'APEX est donc bien inférieur à celui de BioID (quelques minutes). La purification des protéines biotinylées s'effectue comme pour le BioID à l'aide de streptavidine liée à une phase stationnaire après une lyse agressive des cellules. Ces protéines sont ensuite digérées en peptides et identifiées par spectrométrie de masse. De la même manière que pour le BioID de nouvelles formes de cette enzyme ont été introduites par la suite augmentant les capacités catalytiques de l'ascorbate peroxydase au sein de l'APEX2. Cet accroissement de l'activité catalytique permet l'étude de protéines plus faiblement exprimées (Lam *et al.*, 2015). Bien que la méthode APEX permette d'identifier un ensemble d'interactions protéiques moins large que la méthode BioID, elle reste particulièrement adaptée pour la détection spécifique d'interactions protéiques transitoires sur une fenêtre de temps courte. Ainsi cette méthode est particulièrement adaptée à la caractérisation de voies de signalisations à activation très rapide (Béganton *et al.*, 2019).

L'APEX présente donc, comme le BioID, l'avantage de s'affranchir de la nécessiter de conserver les interactions durant le processus de lyse permettant une solubilisation d'une plus grande variété de protéines. La vitesse de biotinylation (quelques secondes à quelques minutes) permet l'étude de mécanismes à activation rapide ainsi que des interactions transitoires de manière efficace. Cette vitesse est cependant associée à une toxicité cellulaire due à l'ajout de biotine-phénol et de peroxyde d'hydrogène (Kushner, Papa and Marx, 2021). Le faible de temps de marquage entraine également une perte de sensibilité par rapport au BioID (Béganton *et al.*, 2019) et des biais de diffusion dans l'ensemble des compartiments de la biotine-phénol ont été empiriquement observé, le groupement phénol semblant enclin à se concentrer dans les régions riches en membranes.

2.4.3 Interactomique non supervisée, le Cross-Link-MS

Les méthodes présentées jusqu'ici, qu'elles soient supervisées ou semi-supervisées, ont comme point commun la nécessité d'une hypothèse de départ permettant de choisir une ou deux protéines. Elles permettent l'étude de l'interaction entre deux protéines connues ou l'identification de protéines interagissant avec une protéine d'intérêt. L'interactomique non supervisée vise à identifier les interactions et complexes formés au sein d'une cellule, sans *a priori*. Cette approche peut être réalisée par l'utilisation du pontage chimique (cross-linking) couplé à la spectrométrie de masse (XL-MS) (Tang *et al.*, 2005).

Le XL-MS repose sur la formation de liens covalents entre plusieurs protéines proches dans l'espace. Ces liens sont formés par l'insertion de petits composés chimiques bifonctionnels appelés cross-linkers entre deux chaînes latérales d'acides aminés. Ces chaînes latérales peuvent provenir d'une seule protéine (pontage intramoléculaire) ou de deux protéines en interaction (pontage intermoléculaire). Les cross-linkers ont une longueur de bras espaceur définie qui permet de déterminer la distance entre les acides aminés liés de manière covalente. Ces contraintes de distance apportent des informations structurales lors de l'analyse des complexes protéiques identifiées (Leitner *et al.*, 2016).

Concrètement, le cross-linker est incubé avec l'échantillon protéique d'intérêt, permettant la formation de ponts covalents entre protéines proches spatialement. Après arrêt de la réaction, l'échantillon est généralement digéré à la trypsine pour obtenir des peptides, linéaires ou pontés (Piersimoni *et al.*, 2022). Ces peptides sont ensuite analysés par chromatographie liquide couplée à la spectrométrie de masse en tandem (LC-MS/MS), où les peptides pontés sont fragmentés et identifiés grâce à des algorithmes dédiés. Les spectres MS/MS permettent de déterminer les acides aminés connectés par le cross-linker. Enfin, ces données sont interprétées en termes de distances inter-résidus, permettant la modélisation structurale des protéines et complexes étudiés. Ainsi, le XL-MS apporte des informations uniques sur l'organisation structurale des protéines, leur flexibilité et leurs interactions.

Différents types de cross-linkers existent, avec des groupements réactifs variés définissant leurs spécificités vis-à-vis des acides aminés ciblés (lysine, cystéine…etc). Certains cross-linkers possèdent également des caractéristiques particulières comme la clivabilité, propriété utile pour faciliter l'analyse en spectrométrie de masse. Les pontages inter- et intra-moléculaire peuvent être réalisés sur protéines purifiées, complexes protéiques simples, ou échantillons cellulaires/tissulaires complexes. Les cross-linker clivables gagnent en popularité notamment pour les échantillons complexes. En effet la clivabilité de ces composé présente l'avantage de générer des ions caractéristiques lors de la fragmentation en CID (« Collision Induced Dissociation ») utilisé pour la MS/MS (Sinz, 2017). Cette caractéristique permet l'obtention de peptides linéaires présentant un décalage de masse correspondant à la partie du cross-linker qui est restée fixée. L'identification de ces peptides linéaires est beaucoup moins complexe que dans le cas de peptides pontés (Götze *et al.*, 2015).

Le XL-MS permet donc l'identification d'interactions entre protéines sans a priori. L'identification des peptides sur lesquels s'effectuent les pontages permet également l'obtention d'informations structurales à la fois sur la géométrie de l'interaction mais également sur la structure d'une protéine donnée dans le cas de pontage intramoléculaire. La nature covalente du pontage permet également la détection d'interaction transitoires de faible affinité. Ce type d'approche nécessite néanmoins une optimisation du protocole de pontage et du choix des cross-linker qui peut être ardue. L'analyse bioinformatique des données générées par XL-MS est plus complexe que ce qui est réalisé dans d'autres méthodes d'interactomique. Enfin, l'interprétation des données spatiales, fournies grâce aux contrainte de distances est limitée par la dynamique structurale des protéines qui peut faire varier la distance entre deux résidus (Piersimoni *et al.*, 2022).

Dans cette partie introductive, nous avons exposés les différentes connaissances nécessaires à la compréhension des expériences qui suivent, visant donc à élucider les mécanismes oncogéniques en jeu dans les CCM-VP. Les travaux menés ici ont pour objectif l'identification des interactions formées par les protéines virales avec les protéines de la cellule hôte. Associé à l'étude des modifications du protéome induites par la présence des antigènes T, les hypothèses de mécanisme oncogénique sont alors testées de manière plus ciblée.

3 MATERIEL ET METHODES

3.1 CULTURE CELLULAIRE

Les HEK 293 Flp-In T-Rex ont été cultivées dans du DMEM (Dulbecco's Modified Eagle's Medium) (Gibco) supplémenté par du sérum de veau fœtal à 10% (Sigma), GlutaMAX et pénicilline/streptomycine 1:100 (Invitrogen). Les cellules CCM-VP et CCM-VN ont été cultivées dans du RPMI supplémenté par du sérum de veau fœtal à 10 %, de la L-Glutamine à 2mM (Invitrogen) et pénicilline/streptomycine 1:100 (Invitrogen). Les cellules ont été cultivées à 37°C, 5% CO_2 sous atmosphère humide.

3.2 CLONAGE

Les gènes codant les protéines virales ont été amplifiés par PCR à partir de séquences synthétiques pour les protéines virales et à partir d'ADNc de cellules 293 Flp-In T-REx pour les autres gènes clonés. Les amorces utilisées présentaient en 5' les séquences reconnues par les enzymes de restrictions AscI (amorce sens) et NotI (amorce antisens) dans le cas du BioID, du TurboID et du N-Split TurboID. Dans le cas du C-Split TurboID, les amorces utilisées présentaient en 5' les séquences reconnues par les enzymes de restriction BamHI (sens) et XhoI (antisens). La qualité de l'amplification PCR a été vérifiée par migration des produits sur gel d'agarose 1%. Une migration sur gel d'agarose 1% préparatif a ensuite été réalisée afin d'isoler la bande correspondant au gène d'intérêt amplifié par PCR. Cette bande a été découpée et l'ADN purifié à l'aide du kit Nucleospin® Gel and PCR Clean-up (Macherey-Nagel #740609.250). Les inserts purifiés ont ensuite été digérés par les enzymes de restriction Asc1 et Not1 durant 1h à 37°C au bain-marie et l'ADN une nouvelle fois purifié en utilisant le même kit. Ces gènes ont été clonés au sein de vecteurs pcDNA5-FRT/TO-3xFlag-BirA* et pcDNA5-FRT/TO-BirA*-3xFlag en utilisant les sites AscI/NotI afin de tagguer les protéines à leurs extrémités N-terminale et C-terminale respectivement. Les vecteurs ont préalablement été digérés par Asc1/Not1 1h à 37°C au bain-marie avant d'être déphosphorylés par l'ajout de CIP 2h à 37°C. Les vecteurs ont ensuite été purifiés à l'aide du kit Nucleospin® Gel and PCR Clean-up. Les gènes d'intérêts ont été insérés dans les vecteurs en utilisant la T4 ligase à 37°C sur la nuit. Dans le cas du TurboID et de l'expression d'une forme taggué Flag des protéines LT330

et PIN1, la même opération a été réalisée au sein des vecteurs pcDNA5-FRT/TO-3xflag-TurboID et pcDNA5-FRT/TO-3xFlag grâce aux sites AscI/NotI. Les séquences TurboID et 3xFlag sont situées en 5' de la séquence du gène cloné permettant l'expression d'une protéine portant le tag à son extrémité N-terminale.

Dans le cas du Split TurboID, le clonage est réalisé au sien d'un vecteur pcDNA5 FRT/TO 3xFlag-NSplit-IRES-3xHA-CSplit. Ce clonage a été réalisé en deux temps. Dans un premier temps la séquence codant PIN1 amplifiée à partir d'ADNc de 293 Flp-In T-REx a été insérée dans le premier site multiple de clonage en 3' de la séquence codant la partie N-terminale de la biotine ligase. Cette séquence a été insérée après digestion de l'insert et du vecteur par le couple d'enzyme AscI/NotI comme décrit précédemment. La séquence codant NEDD8 a été insérée suite à une digestion BamHI/XhoI dans le deuxième site multiple de clonage, en 3' de la séquence codant la partie C-terminale de la biotine ligase.

Ces constructions ont été amplifiées en bactéries DH5α subcloning efficiency. 2,5µL de la solution de ligation vecteur/insert ont été ajoutés à 20µL de DH5α sur glace. Un choc thermique a ensuite été réalisé en passant les bactéries 45sec à 42°C puis 5min sur glace. 100µL de LB ont ensuite été ajoutés et les bactéries incubées 30min à 37°C sous agitation. Les DH5α ont ensuite été déposées sur plaques de LB agar et incubées sur la nuit à 37°C en présence d'ampicilline. Les colonies s'étant développées ont été remise en culture en milieu liquide (LB + Ampicilline 50 µg/mL) à 37°C sur la nuit. Les vecteurs amplifiés ont ensuite été récupérés en utilisant le kit Nucleospin® Plasmid (Macherey-Nagel #740588.250). La présence des inserts a été vérifiées par migration sur gel d'agarose 1% suite à une digestion par Asc1/Not1 1h à 37°C au bain-marie. La bonne insertion des séquences d'intérêt ainsi que l'absence de décalage de phase entre la séquence d'intérêt et BirA* ont été vérifiées par séquençage Sanger.

Les vecteurs ont ensuite été transfectés avec du PolyJet (SignaGen Laboratories # SL100688-1ml) et le vecteur de la recombinase pOG44 au sein de la lignée Flp-In T-Rex 293. Les clones ayant inséré les constructions de manière stable grâce au système Flp-In ont été sélectionnés sous hygromycine B en conditions de culture classiques : DMEM 10% Sérum de Veau Fœtal et 1% pénicilline/streptomycine dans une atmosphère à 37°C, 5% CO_2 et saturée en humidité.

Les gènes codant pour sT et la forme tronquée de LT de 330 acides aminés ont été clonés dans un vecteur pGEx permettant l'expression en bactéries BL-21 des protéines de fusion GST-sT et GST-LT330. Après séquençage, des bactéries BL-21 ont été transformées avec ces constructions. Trois colonies ont été sélectionnées pour chaque construction et cultivées en LB + ampicilline 50 µg/mL. L'expression des protéines recombinantes a été induite par ajout d'IPTG à 0,5 mM, et analysée par SDS-PAGE et coloration au bleu de Coomassie.

3.3 PREPARATION DES ECHANTILLONS

3.3.1 BioID et TurboID

Dans le cas du BioID, les cellules sous-confluentes ont été incubées dans du DMEM complet supplémenté avec 1 µg/mL de tétracycline (Sigma) et 50 µM de biotine pendant 24h, récoltées par grattage, culottées, lavées avec du PBS froid, séchées et congelées.

Dans le cas du BioID réalisé sur EHMT2 en présence et absence de LT330, deux lignées cellulaires transfectées de manières stables et exprimant de manière inductible soit la forme LT330 portant un triple Flag sur son extrémité N-terminale, soit le triple Flag seul. Le vecteur pcDNA5-FRT/TO-3xFlag-BirA*-EHMT2 obtenu selon le procédé décrit dans la partie **Clonage** a été transfecté de manière transitoire au sein de ces deux lignées de manière analogue à la transfection stable et sans utilisation du vecteur pOG44 codant la recombinase.

Dans le cas du TurboID, la tetracycline a également été ajoutée à une concentration de 1µg/mL 24h avant la récolte et la biotine a été ajoutée à 50µM 1h avant la récolte.

Dans le cas du Split TurboID, la tetracycline a également été ajoutée à une concentration de 1µg/mL 24h avant la récolte et la biotine a été ajouté à 50µM 24h avant la récolte ou 4h avant la récolte.

Chaque culot cellulaire a été resuspendu dans 5 mL de tampon de lyse (50 mM Tris-HCl pH 7,5, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0,1% SDS, cocktail d'inhibiteurs de protéases (Sigma-Aldrich) au 1:500, Turbonucléase (Accelagen) au 1:1 000), incubé sur un agitateur rotatif à 4°C pendant 1 heure, brièvement soniqué pour dissocier les agrégats visibles, puis centrifugé à 40 000 x *g* pendant 30 minutes à 4°C. 25µl de résine Streptavidine Ultralink (Pierce) pré-équilibrée ont été ajoutés, et le mélange a été incubé

pendant 3 heures à 4°C sous agitation rotative. Les billes ont été culottées par centrifugation à 500 x *g* pendant 2 minutes et lavées une fois avec 1 mL de tampon de lyse et cinq fois avec 1 mL de bicarbonate d'ammonium 50 mM (pH = 8,3). La digestion trypsique a été réalisée en incubant les billes avec 1 µg de trypsine TPCK de grade MS (Promega) dissoute dans 200 µL de bicarbonate d'ammonium 50 mM (pH 8,3) sur la nuit à 37°C. Le lendemain matin, 0,5 µg de trypsine TPCK de grade MS a été ajouté aux billes et incubé 2 heures supplémentaires à 37°C. Après centrifugation à 500 x *g* pendant 2 minutes, le surnageant a été collecté et transféré dans un nouveau tube Eppendorf. Un lavage supplémentaire a été effectué avec 200 µL de bicarbonate d'ammonium 50 mM et ajouté au premier éluât. Les échantillons ont ensuite été lyophilisés et préparés pour la spectrométrie de masse.

3.3.2 GST Pull-down

Les clones présentant le meilleur niveau d'expression ont été amplifiés en BL-21 en culture liquide LB + 50 µg/mL ampicilline + 0,5 mM IPTG. Après lyse bactérienne en tampon SDS, des billes magnétiques Glutathione (Sigma-Aldrich, # G0924-5X1ML) ont été incubées dans le lysat afin de lier les protéines de fusion. Une quantification par SDS-PAGE/Coomassie a permis de normaliser les quantités de protéines de fusion immobilisées.

Des culots cellulaires de 75 million de cellules MKL-1 et WaGa ont été lysés dans un tampon de lyse douce (50mM Tris-HCl pH 7,5 ; 100mM KCl ; 2mM EDTA ; 0,1% NP40 ; 10% Glycérol ; 1mM PMSF ; 10mM NEM ; 1mM DTT ; cocktail d'inhibiteurs de protéases (Sigma-Aldrich) au 1:500, Turbonucléase (Accelagen) au 1:1 000) puis centrifugés à 40 000 x *g*, 30 minutes à 4°C. Le volume de billes a été adapté pour obtenir 2,5 µg de protéine de fusion par échantillon incubées avec les lysats cellulaires pendant 3h à 4°C. Après lavages dans du bicarbonate d'ammonium 50 mM, les protéines ont subi une digestion trypsique à 37°C sur la nuit. Les échantillons ont ensuite été lyophilisés.

3.3.3 Protéome total

Dans le cas de l'étude protéomique réalisée dans la partie **Délimitation des interactions fonctionnelles des antigènes T**, dix millions de cellules WaGa WT et WaGa shTA (expression inductible de shRNA ciblant les antigènes T) traitées avec de la doxycycline à 2µg/mL pendant 10 jours ont été utilisées.

Dans le cas de l'étude des variations du protéome dans la partie **Régulation des dommages de l'ADN par tLT et EHMT2** dix millions de cellules WaGa WT et WaGa shTA traitées 7 jours à la doxycycline puis 48h par du BRD4770 à 5µM ont été utilisées.

Les cellules ont été lysées en utilisant un tampon RIPA (50 mM Tris-HCl pH 7,5, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0,1% SDS, cocktail d'inhibiteurs de protéases (Sigma-Aldrich) au 1:100). Les lysats ont été brièvement soniqués et centrifugés à 16 000 x *g* pendant 20 minutes à 4°C. Les surnageants ont été récupérés et préparés en utilisant la préparation d'échantillon assistée par filtre (FASP) (Amicon 10kDA #UFC501024, Merck Millipore). La digestion des protéines a été réalisée avec les protéases trypsine et Lys-C (#V5073, Promega). Les échantillons ont ensuite été lyophilisés.

3.3.4 Préparation des échantillons pour la spectrométrie de masse

Les échantillons lyophilisés ont été suspendus dans une solution de TFA 0,1% et dessalés en utilisant une phase C18 (#ZTC18S096, Sigma-Aldrich). Une chromatographie liquide nano (easy1000, Thermo Scientific) couplée à un spectromètre de masse (Q-Exactive, Thermo Scientific) a été utilisée pour analyser les peptides. Le système a fonctionné en mode d'ionisation positive avec un gradient linéaire d'acétonitrile (2-30%) dans de l'acide formique 0,1% pendant 140 min à un débit de 300 nL/min. Une acquisition dépendante des données a été réalisée en sélectionnant les 10 ions MS1 les plus intenses pour la fragmentation et l'obtention de spectres MS2.

3.4 ANALYSE EN NLC-MS/MS

Les protéines ont été identifiées par comparaison des données MS/MS avec la base de données du protéome Homo Sapiens Uniprot (décembre 2021) à l'aide du logiciel MaxQuant (v1.6.10.43). Les paramètres ont été ajustés pour permettre jusqu'à deux clivages manqués par la trypsine. L'oxydation des méthionines et l'acétylation N-terminale des protéines ont été définies comme modifications variables. Les paramètres par défaut du logiciel ont été utilisés pour la quantification sans marquage (LFQ). Concernant la tolérance de masse, 6 ppm ont été choisis pour le mode MS et 20 ppm pour les données de fragmentation afin de correspondre à la tolérance MS/MS. L'identification des protéines et peptides a été réalisée avec un taux de fausse découverte (FDR) de 1%, nécessitant un minimum de deux peptides par protéine.

Dans le cas du BioID, les mesures LFQ des 20 expériences contrôles (échantillons BirA* seuls provenant de lignées cellulaires stables) ont été combinées pour obtenir les trois meilleurs scores pour chaque ID spécifique. Ce set de valeurs contrôles a servi de référence pour la comparaison avec les triplicats des protéines appâts virales. Les données ont été traitées à l'aide du logiciel Perseus (v1.6.15.0). En résumé, l'intensité LFQ pour chaque échantillon a été importée dans Perseus, suivie d'un filtrage de la matrice de données excluant les contaminants potentiels et les protéines identifiées uniquement par site. Les données ont ensuite été transformées en utilisant la fonction log2(x). Le t-test de Student a été réalisé pour comparer les triplicats biologiques des protéines appâts aux trois contrôles. Les protéines identifiées avec une q-value inférieure à 0,01, un log2 fold change supérieur à 1 et une valeur de LFQ expérimentale (détectées) dans chacun des trois réplicats ont été considérés comme des interacteurs de haute confiance. Les protéines pour lesquelles une valeur de LFQ expérimentale est retrouvée dans chacun des réplicats tandis qu'aucune valeur de LFQ n'est détectée dans les contrôles est également considérée comme interacteur de haute confiance.

Pour l'étude des interactions de EHMT2 par BioID en présence et absence de LT330, les paramètres d'analyses sont les mêmes que ceux décrit ci-dessus. En plus de l'analyse réalisée dans cette partie, les valeurs LFQ obtenues pour chaque interacteur identifié ont été comparées entre les conditions présence/absence de LT330. Les interacteurs ont été considérés comme significativement enrichis en présence de tLT lorsque la différence de Log2 était supérieure à 1 et la q-value inférieure à 0,05 (Test de Student).

L'analyse des résultats de TurboID est réalisée de manière analogue à l'analyse faite pour le BioID à la seule différence que les contrôles utilisés ici sont 8 échantillons où l'enzyme TurboID est exprimée seule au sein des cellules.

L'analyse des résultats de Split-TurboID a été réalisée de façon similaire à celle du BioID. Trois conditions contrôles ont été utilisées : 1) les deux parties de la biotine ligase BirA* exprimées ensemble sans protéine de fusion, 2) la partie N-terminale de BirA* fusionnée à PIN1 exprimée avec la partie C-terminale non fusionnée, et 3) la partie N-terminale non fusionnée exprimée avec la partie C-terminale fusionnée à NEDD8. Les 3 valeurs de LFQ maximales pour chaque identification ont été combinées en un triplicat contrôle. De la même manière, un t-test de Student a été réalisé pour comparer les triplicats biologiques des protéines appâts aux trois contrôles. Les protéines identifiées avec une q-value inférieure à

0,05, un log2 fold change supérieur à 1 et une valeur de LFQ valide dans les trois réplicats ont été considérés comme des interacteurs de haute confiance. De la même manière, lorsqu'une valeur de LFQ est mesurée dans chacun des réplicats et dans aucun contrôle, la protéine est considérée comme interacteur de haute confiance. Dans le cas du GST Pull-down, les données ont également été traitées avec le logiciel Perseus (v1.6.15.0). Les valeurs LFQ de chaque triplicat ont été comparées aux valeurs LFQ des triplicats contrôles (GST seul). Une protéine identifiée est considérée comme un interacteur lorsqu'une valeur de LFQ est mesurée dans les trois réplicats et dans aucun des réplicats contrôles ou que le Log2 fold change est supérieur à 1 et la q-value inférieure à 0,05.

L'analyse du protéome total a été réalisée en utilisant la même instrumentation et les mêmes logiciels que précédemment décrits. Pour chaque protéine identifiée, les valeurs de LFQ obtenues à partir des trois réplicats biologiques des deux conditions expérimentales ont été comparées à l'aide d'un t-test de Student. Une protéine a été considérée comme plus abondante dans une condition lorsque le log2 fold change était supérieur à 1, la q-value inférieure à 0,05 et qu'au minimum deux réplicats sur trois présentaient une valeur de LFQ valide dans les deux conditions. Une protéine a été considérée comme exclusive à une condition lorsque cette condition présentait une valeur de LFQ valide pour chacun des trois réplicats, et que la protéine était absente des trois réplicats de l'autre condition.

3.5 PLA

3.5.1 Préparation des cellules

Les cellules CCM-VP, MKL-1 et WaGa ont été cultivées dans du milieu RPMI supplémenté par du sérum de veau fœtal à 10 % (Sigma), de la L-Glutamine à 2mM (Invitrogen) et un mélange pénicilline/streptomycine 1 :100 (Invitrogen). Les HEK 293 Flp-In T-REx ont été cultivées dans du milieu DMEM (Gibco) supplémenté par du sérum de veau fœtal à 10% (Sigma)et un mélange pénicilline/streptomycine 1 :100 (Invitrogen). Une lignée HEK exprimant une forme tagguée Flag de PIN1 a été utilisée pour la réalisation de la PLA. Le processus de clonage est décrit dans la partie **Clonage**. Les cellules ont été cultivées à 37°C, 5% CO_2 sous atmosphère humide. Les cellules CCM-VP poussant en suspension ont été transférées dans des puits de labtek traités à la Poly-D-Lysine afin de faire adhérer les cellules.

Les cellules ont été fixée à l'aide de paraformaldéhyde 4% dilué dans du milieu de culture. Après lavage au PBS, les membranes cellulaires ont été perméabilisées par du PBS Triton X-100, 0,1%.

3.5.2 Réalisation de la PLA

Les cellules ont été incubées dans une solution de blocage PBS/ BSA 1% afin de saturer les sites de fixation aspécifiques. Les cellules ont ensuite été incubés à 4°C sur la nuit dans les différentes solutions contenant les anticorps primaires dilués dans la solution de resuspension fournie dans le kit (#ref). Les différents anticorps ont été utilisés aux dilutions suivantes : Anti-LT-MCPyV (Santa Cruz, #sc-136172, 1 :200), anti-PIN1 (Abcam, #ab12107, 1 :200), anti-EHMT2 (Cell Signaling Technology, #3306s, 1 :200), anti-Flag (Sigma-Aldrich, #F1804-50UG, 1 :500), Anti-NEDD8 (Abcam, #ab81264, 1 :200), anti-RBX1 (Cell Signaling Technology, #11922S, 1 :200). La réalisation de la PLA a été réalisé selon les instructions du fournisseur du kit Duolink (Sigma-Aldrich, #DUO92101-1KT).

L'acquisition d'images a été effectuée sur un microscope confocal Zeiss LSM700 couplé à un Axiovert 200 M Zeiss avec objectif à immersion 63x/1.30 ON (Carl Zeiss AG, Oberkochen, Germany). L'acquisition a été réalisée avec le logiciel Zeiss Zen 2 et les images ont été assemblées et analysées avec ImageJ (ImageJ, RRID:SCR_003070).

3.6 TISSUE MICRO ARRAY

3.6.1 Echantillons

Les cas de carcinomes à cellules de Merkel ont été sélectionnés à partir d'une cohorte historique et prospective multicentrique française de patients avec un diagnostic de carcinome à cellules de Merkel établi entre 1998 et 2017 (Comité d'éthique local de la recherche sur l'être humain, Tours, France ; no ID RCB2009-A01056-51). Les critères d'inclusion pour cette cohorte ont été décrits précédemment (Kervarrec *et al.*, 2018). En bref, les tumeurs avec des échantillons de tissus fixés au formol et inclus en paraffine ont été incluses comme carcinomes à cellules de Merkel si elles présentaient une morphologie compatible, avec la combinaison d'une positivité pour CK20 et d'au moins un marqueur neuroendocrine (synaptophysine et chromogranine A) ou en l'absence de positivité pour

CK20, l'expression d'au moins deux marqueurs neuroendocrines associée à l'absence de carcinome neuroendocrine profond confirmée par des examens d'imagerie.

3.6.2 Immuno histochimie

Tous les échantillons de carcinomes à cellules de Merkel et de carcinomes neuroendocrines ont été inclus dans un tissue microarray. Les zones tumorales centrales sans nécrose ont été sélectionnées sur les coupes colorées à l'hématoxyline phloxine safran (HPS) pour exclure un marquage non spécifique. Les zones sélectionnées ont été extraites sous forme de noyaux de tissu de 1 mm de diamètre, les noyaux ont été montés en triplicats sur le tissue microarray à l'aide d'un système de tissue array semi-motorisé (MTA booster OI v2.00, Alphelys). Le marquage immunohistochimique pour CK20, TTF-1, NF, PAX5, TdT et CD99 a été réalisé sur le BenchMark XT Platform selon les instructions du fabricant. Le marquage pour PIN1 a été effectué manuellement.

3.7 WESTERN-BLOT

Dans le cadre de l'analyse réalisée dans la partie **Régulation des dommages de l'ADN par tLT et EHMT2**, les cellules WaGa WT et shTA ont été traitées à la doxycycline 2µg/mL 7 jours puis au BRD4770 à 5 µM 2 jours. Les cellules ont été traitées à l'Hydroxyurée 2mM durant 4h suivies ou non d'une période de 18h de récupération sans hydroxyurée. Les cellules ont été lysées dans un tampon RIPA (100 mM NaCl ; 50 mM Tris ; 5 mM EGTA ; 5 mM EDTA ; 100 mM NAF ; 10 mM Sodium Pyrophosphate ; 1% NP-400 ; 1 mM PMSF) supplémenté avec un cocktail d'inhibiteurs de protéases (Sigma). Les lysats ont été soniqués 2*15s à 30% d'amplitude et centrifugés à 14 000 x *g* pendant 20 min à 4°C afin de culoter les débris, puis le surnageant a été récupéré et transféré dans un nouveau tube. La concentration protéique de chaque échantillon a été déterminée par dosage Bradford (Bio-Rad). La concentration des échantillons a été homogénéisée et du tampon Laemmli 4X a été ajouté. La même quantité de protéine a été déposée dans chaque puit et la migration a été réalisée sur gel de polyacrylamide. Le gel est séparé en un gel de concentration à 4% (4% acrylamide/bisacrylamide ; 0,125 M Tris-HCl pH 6,8 ; 0,1% SDS ; 0,06% APS ; 0,006% TEMED) et un gel de séparation à 12% (12% acrylamide/bisacrylamide ; 0,375 M Tris-HCl pH 8,8 ; 0,1% SDS ; 0,06% APS ; 0,006% TEMED). Dans le cas des analyses réalisées dans la partie **Régulation des dommages de l'ADN par tLT**

et EHMT2, les migrations ont été réalisées en utilisant des gels pré-coulés avec un gradient polyacrylamide 4-20% (Bio-Rad). La migration a été effectuée dans un tampon Tris-Glycine-SDS (192mM Glycine, 25mM Tris base, 0,1% SDS). Le transfert sur membrane de nitrocellulose a été réalisées par transfert semi-sec à l'aide du Trans-blot (BioRad) dans un tampon de transfert (192mM Glycine, 25mM Tris base, 20% methanol). Après lavages dans une solution de lavage TBS-T (0,1M Tris, 0,8M NaCl, 0,1% Tween-20), les membranes ont été incubées dans une solution de saturation TBS-T BSA 5%. Les membranes ont ensuite été incubées à 4°C sur la nuit en présence des anticorps primaires.

Les anticorps primaires et concentrations utilisés étaient : anti-CHK1 phosphorylé (Ser345) (Cell Signaling Technology, #2348, 1:1000), anti-CHK2 phosphorylé (Thr68) (Cell Signaling Technology, #2197, 1:1000), anti-ATR phosphorylé (Thr1989) (Cell Signaling Technology, #30632, 1:1000), anti-ATM phosphorylé (Ser1981) (Cell Signaling Technology, #5883, 1:1000), anti-p21 (Cell Signaling Technology, #2947, 1:1000), anti-EHMT2 (Cell Signaling Technology, #3306, 1:1000), anti-MCPyV-LT (Santa Cruz Biotechnology, #sc-136172, 1:2000), anti-PIN1 (Abcam, #ab12107, 1 :1000), anti-βActine (Cell Signaling Technilogy, #3700S, 1 :5000), anti-αTubuline (Abcam, #ab176560, 1 :5000).

3.8 IMMUNOFLUORESCENCE

En bref, les cellules CCM-VP ont été cultivées dans des chambres de culture Labtek préalablement recouvertes de polylysine, puis rincées avec du PBS. Elles ont ensuite été fixées avec du paraformaldéhyde à 4% pendant 15 minutes à température ambiante. Après fixation, les cellules ont été lavées trois fois avec du PBS et la perméabilisation des membranes a été réalisée avec du PBS contenant 0,1% de Triton X-100. Le blocage a été effectué avec une solution de PBS, 0,01% de Tween 20 et 5% de d'albumine sérum bovin (BSA) pendant 30 minutes à 4°C. Par la suite, les cellules ont été incubées une nuit à 4°C avec l'anticorps primaire anti-γH2Ax (dilution 1:1000, Abcam #ab2893). Après trois lavages de 10 minutes avec du PBS et 0,01% de Tween 20, les cellules ont été exposées à l'anticorps secondaire anti-lapin IgG couplé à l'Alexa Fluor 555 (dilution 1:2000, Thermo Fisher Scientific Cat# A-31572, RRID:AB_162543). Un rinçage ultérieur avec du PBS a été réalisé et les noyaux cellulaires ont été marqués avec du DAPI à 1 µg/ml (Thermo Fisher Scientific #62247) dans du PBS pendant 5 minutes. Après un dernier lavage et dessalage, les lames ont été montées avec du milieu de

montage fluorescent Dako (Agilent Dako, Santa Clara, CA, USA). L'acquisition d'images a été effectuée sur un microscope confocal Zeiss LSM700 couplé à un Axiovert 200 M Zeiss avec objectif à immersion 63x/1.30 ON (Carl Zeiss AG, Oberkochen, Germany). L'acquisition a été réalisée avec le logiciel Zeiss Zen 2 et les images ont été assemblées et analysées avec ImageJ (ImageJ, RRID:SCR_003070).

3.9 TEST DE PROLIFERATION

3.9.1 MTS

Les cellules traitées pendant 5 jours avec de la doxycycline ont été ensemencées en plaque 96 puits avant d'être traitées ou non avec du BRD4770 (MedChem Express, # HY-16705). Après 24h, 10 µL de CellTiter 96® AQueous One Solution Reagent (Promega, #G3580) ont été ajoutés et l'absorbance a été lue après 2,5h d'incubation à l'obscurité à 37°C.

3.9.2 Mixed culture assay

La croissance a été comparée entre des cellules WaGa avec expression inductible de shTA et des cellules WaGa avec expression constitutive de GFP. Les cellules WaGa-shTA ont été mélangées avec environ 30% de cellules WaGa-GFP. Les changements d'abondance relative entre les deux populations ont été quantifiées par cytométrie en flux. Les cellules ont été traitées ou non avec de la doxycycline au jour 0 et traitées ou non avec 5 µM de BRD4770 à partir du jour 6 de doxycycline ; le BRD4770 a été ajouté toutes les 48h après ce traitement initial.

3.10 COMET ASSAY

Des tests comet alcalins ont été réalisés en utilisant 2 x 10^5 cellules, en suivant des protocoles établis précédemment (Trapp-Fragnet *et al.*, 2014). Un microscope inversé à épifluorescence Axiovert 200 M (Zeiss) a été utilisé pour visualiser les comètes, et les photographies ont été capturées avec un dispositif Axiocam MRm (Zeiss). Au moins 50 comètes ont été examinées pour chaque échantillon à l'aide de l'outil CometScore (version 1.5 ; TriTek). Le paramètre du Olive Tail Moment (OTM) a été calculé sur la base de la longueur de la queue et du pourcentage d'ADN dans la queue. Les données sont représentées sous la forme de la moyenne des OTM plus ou moins l'écart-type (SD).

4 RESULTATS

L'objectif principal de ce projet est d'identifier des mécanismes oncogéniques impliquant les protéines virales du MCPyV. La méthodologie adoptée afin d'y parvenir consiste en un ensemble d'analyses exploratoires visant à identifier les partenaires et fonctions potentielles des antigènes T. Les hypothèses formulées à partir de ces données sont ensuite testées et approfondies afin de confirmer ou infirmer ces hypothèses et de mettre ainsi en évidence de nouveaux mécanismes impliqués dans l'oncogenèse des CCM.

4.1 DELIMITATION DES INTERACTIONS FONCTIONNELLES DES ANTIGENES T

L'approche choisie débute donc par une phase exploratoire visant à délimiter les protéines et mécanismes potentiels les plus intéressants et les plus vraisemblablement impliqués dans l'oncogenèse CCM dépendante du MCPyV. Cette pathologie est due à la présence du MCPyV et les cellules CCM-VP sont dépendantes de l'expression des antigènes T pour leur survie. Ainsi, il apparait évident que ces deux protéines virales constituent le point de départ de nos investigations. Nous nous sommes donc tournés dans un premier temps vers la délimitation de leurs interactomes protéiques en utilisant BioID.

4.1.1 Identification des interacteurs proximaux des antigènes T par BioID

L'objectif de cette étude est d'identifier les partenaires cellulaires des oncoprotéines virales. L'identification de l'environnement protéique des antigènes T permettra la formulation d'hypothèses quant à leurs fonctions.

Les antigènes T ont dans un premier temps été clonés au sein de vecteurs pcDNA5 FRT/TO FlagBirA* et pcDNA5 FRT/TO BirA*Flag. La construction FlagBirA* permet l'expression de la protéine d'intérêt fusionnée à son extrémité N-ter à la biotine ligase mutée portant un Flag tag. La construction BirA*Flag quant à elle permet l'expression de la protéine d'intérêt fusionnée en sa partie C-ter à BirA*Flag. Ces constructions ont été transfectées de manière

stable dans les cellules 293 Flp-In T-REx. Ces lignées stables permettent une expression inductible à la tétracycline des transgènes (Antigène T – BirA*) et sont utilisées afin d'étudier les interactions de ces protéines par BioID. En effet, la lignée 293 Flp-In T-REx possède une unique séquence FRT intégrée de manière clonale au génome, permettant la recombinaison homologue avec la séquence FRT présente sur le vecteur. Ainsi, un vecteur pOG 44 exprimant la Flp recombinase est co-transfecté avec le pcDNA5 FRT/TO, ce qui permet l'insertion du vecteur entier par recombinaison homologue dans le génome de la cellule. Cette technologie assure l'insertion d'une seule copie du vecteur dans une zone transcriptionnellement active du génome et sans mutagénèse insertionelle. Ces cellules expriment également un répresseur à la tétracycline (tet) qui va inhiber l'expression de la séquence codée par le vecteur grâce à la présence des tet-opérons en tandem en aval du promoteur CMV. L'ajout de tétracycline permet, en se liant de manière irréversible à ce répresseur, d'induire l'expression de la protéine d'intérêt par libération des tet-opérons. Chaque protéine virale a donc été fusionnée à la biotine ligase en C-Ter et en N-Ter. Des triplicats ont été réalisés pour chaque construction (soit pour une même protéine, 3 avec BirA* en N-ter et 3 avec BirA* en C-ter).

Les résultats obtenus ont été analysés en Label Free Quantification (LFQ) sur le logiciel Max Quant (Cox and Mann, 2008). Les abondances relatives ont été comparées à 10 contrôles FlagBirA*-only et 10 contrôles BirA*Flag-only afin d'éliminer le bruit de fond inhérent à BioID, c'est à dire la biotinylation endogène, le marquage non spécifique ou celui due aux interactions du tag BioID seul, et les protéines se fixant aux billes de manière aspécifique. En agrégeant toutes les données, cette étape a permis de définir un paysage protéique proximal de 553 facteurs (FDR<1%, p-value<0,01) (**ANNEXE I**). Nous retrouvons parmi nos données des interactions fonctionnellement importantes et déjà identifiées comme RB1-tLT ou EP-400. EP400 est retrouvé en interaction avec sT, comme décrit dans la littérature (Cheng *et al.*, 2017), mais également à proximité de tLT. Les TAs ont également été retrouvés en interaction avec un nombre important de chaperonnes/co-chaperonnes de la famille Hsc70, comme attendu avec la présence du domaine DnaJ. L'identification de ces interactions bien décrites illustre la pertinence de notre approche. La majorité des interactions identifiées sont formées par sT et LT330, peu d'interactions sont retrouvées avec la forme complète de LT, forme non exprimée dans les cancers. Nous retrouvons également peu d'interactions avec la forme

tronquée la plus courte étudiée ici, LT275. Les deux autres formes de tLT observées, LT330 et LT428 présentent une grande similarité d'interactomes, avec 196 interacteurs en commun.

4.1.1.1 sT

En fusionnant les listes d'interacteurs proximaux identifiés avec la forme FB-sT (Flag-BirA* en N-terminal) et avec la forme sT-BF (BirA*-Flag en C-terminal), 149 groupes protéiques uniques ont été identifiés pour cette protéine virale (**ANNEXE I**). Parmi les interacteurs ainsi identifiés, nous avons retrouvé les interactions bien connues formées par sT. La phosphatase PP4C a notamment été identifiée en interaction avec sT ainsi que les sous-unités catalytiques et régulatrices de la protéine phosphatase 2A, PPP2CA et PPP2R1A respectivement. Nous avons également pu observer la présence d'autres sous-unités de phosphatases à proximité de sT, comme PPP1R12A, PPP5C et PPP6R1. L'interaction majeure entre sT et EP400 est également retrouvée par BioID. Il est connu que cette interaction mène à l'inhibition de p53 en activant la transcription de MDM2. De manière surprenante, nous avons retrouvé ici une interaction entre sT et la protéines p53 ce qui n'est, à ce jour, pas rapportée dans la littérature.

Bien que possédant un domaine DnaJ, aucune fonction impliquant les protéines de choc thermique de la famille Hsc70 n'est connu pour sT. Nous avons pu identifier parmi les interacteurs proximaux de cette protéine virale 29 protéines impliquées dans le repliement des protéines. Nous retrouvons notamment différentes protéines de la famille Hsc70 comme HSPA1B, HSPA2, HSPA4, HSPA4L, HSPA6, HSPA8 et la protéine 1 de liaison à HSP70, HSPBP1. En plus de ces protéines de la famille Hsc70, sT compte également parmi ses interacteurs proximaux les protéines HSP90A A1, B1 et B2P et CDC37, la co-chaperonne de HSP90. Le domaine de fixation des protéines HSP90 n'est pas clairement identifié, mais il est supposé par analogie avec les homologues bactériennes de HSP70 et HSP90 que les protéines cibles puissent être fixées par les HSP70 via le domaine DnaJ puis remodelées par HSP90 qui se lie aux HSP70 (Genest, Wickner and Doyle, 2019). Nous retrouvons également plusieurs protéines HSP40, aussi appelées DNAJ, porteuses d'un domaine J homologue au domaine DnaJ retrouvé sur sT et responsable de l'interaction avec HSP70.

4.1.1.2 LT et tLT

4.1.1.2.1 LT275

Cette forme correspondant aux acides aminés 1-275 (des 817 de LT pleine taille) est l'une des plus courtes observées pour le tLT (LT tronqué). Contrairement aux formes plus longues, la troncation après l'acide aminé 275 entraine la perte du NLS. Cette forme de LT tronquée est celle qui présente le moins d'interactions en BioID. Le nombre d'interacteurs proximaux était particulièrement bas qu'il s'agisse de la forme portant BirA* en N-terminal ou C-terminal. Nous n'avons identifié que 36 interacteurs dont l'abondance observée en spectrométrie de masse était significativement supérieure à celle des contrôles (**ANNEXE I**).

Nous avons pu identifier parmi les quelques interacteurs de cette protéine, une part importante composée de protéines impliquées dans le repliement protéique. De manière attendue, nous retrouvons plusieurs protéines de la famille des Hsc70 comme HSPA4, A4L et HSPBP1 pour lesquelles LT possède un site d'interaction dans son domaine DnaJ. L'interaction bien documentée avec RB1 est également retrouvée. De manière surprenante, une interaction entre la protéine p53 et cette forme courte de LT a pu être identifiée.

4.1.1.2.2 LT330/428, modifications épigénétiques et réparation de l'ADN

Parmi les formes tronquées de LT, LT330 est la forme la plus récurrente chez les patients. Elle conserve son NLS et est présente dans plusieurs lignées cellulaires CCM-VP. Ce mutant est le plus souvent choisi pour l'étude du tLT et nous utiliserons cette forme pour le reste des études décrites dans ce manuscrit. LT428 est quant à elle une forme un peu plus longue et moins répandue. Ces deux formes de tLT sont celles qui ont permis l'identification du plus grand nombre d'interacteurs dans cette étude. Nous avons identifié 366 interacteurs de haute confiance pour LT330, et 266 pour LT428 (**ANNEXE I**). Les résultats obtenus pour ces deux formes sont très similaires, nous retrouvons en effet 196 interacteurs communs. Dans les deux cas nous retrouvons les principales interactions connues pour LT comme RB1 ou Vam6p. Nous retrouvons également plusieurs protéines de choc thermique de la famille Hsc70 comme HSPA1B, HSPA2, HSPA4, HSPA4L, HSPA6, HSPA8 (**Figure 10**). Ces protéines virales sont également retrouvées en interaction avec la protéine HSPBP1 liant HSP70 et avec une forme plus grande d'HSP, HSPH1. Nous retrouvons l'interaction décrite avec BRD4 supposément

impliqué dans la réplication du génome viral (Wang *et al.*, 2012). Les protéines BAG1/2/3/4/5 ont également été identifiées parmi les interactomes de LT330 et LT428. Ces protéines interagissent avec les Hsc70 et sont fortement impliquées dans la régulation de l'apoptose de manière Bcl-2 dépendante. L'interaction rapportée avec la sous unité 14 du complexe Mediator n'a pas été retrouvée, nous avons cependant identifier d'autres composants de ce complexe comme MED1 et MED15 (Rozenblatt-Rosen *et al.*, 2012). Enfin nous retrouvons également l'interaction connue avec la protéine USP7 impliquée comme Brd4 dans la réplication du génome virale, cette fois-ci en modulant l'interaction entre LT et l'origine de réplication (Czech-Sioli *et al.*, 2020). Nous avons identifié parmi les interacteurs de ces deux formes de tLT le facteur de transcription E2F3 qui est en condition physiologique séquestré par RB1. Cette interaction a également été observée par le passé et se produit probablement lors de la séparation du complexe RB1-E2F par tLT à l'aide des Hsc70 (Sullivan, Cantalupo and Pipas, 2000; Rozenblatt-Rosen *et al.*, 2012). Aucune interaction n'est connue entre le LT tronqué du MCPyV et p53 et contrairement à la forme LT275, p53 n'apparait pas dans les interactomes de LT330 et LT428. Nous avons cependant identifié plusieurs facteurs reliés à p53 comme TP53BP1, HDAC1 et SIRT1.

L'analyse d'enrichissement des processus biologiques dans lesquels les interacteurs ainsi identifiés sont impliqués a révélé plusieurs grands groupes de fonctions. Une large majorité de ces protéines se répartissent en effet entre les protéines intervenant dans la maturation des ARN, les protéines régulant la transcription, les protéines impliquées dans les modifications épigénétiques et les protéines de la réponse aux dommages de l'ADN. L'altération de la régulation épigénétique représente un moyen puissant pour un virus exprimant seulement deux protéines virales de prendre le contrôle de l'expression génique d'une cellule en absence de mutations somatiques. Les modifications épigénétiques ont déjà été rapportées dans les CCM. En effet, les CCM présentent un profil de méthylation de l'ADN « jeune » évoquant celui des cellules souches. Elles n'expriment cependant pas les marqueurs de pluripotence associés. Il est fort probable que ce profil de méthylation soit dû à une dérégulation provoquée par les TAs (Chteinberg *et al.*, 2020). Outre la méthylation de l'ADN, nous retrouvons un certain nombre de protéines impliquées dans la modification des histones. Un exemple intéressant est celui de BEND3 retrouvé en interaction avec LT330. Cette protéine est impliquée dans le recrutement du « Polycomb Repressive Complex 2 », un acteur majeur

de la régulation épigénétique responsable de la marque H3K27me3 entrainant une condensation de la chromatine. Nous avons contribué à une étude montrant que l'inhibition d'EZH2, l'un des composants de ce PRC2, entrainait une perte de viabilité des cellules CCM-VP (Gartin *et al.*, 2022; Durand *et al.*, 2023). Cette protéine BEND3 est également impliquée dans le recrutement d'un autre complexe épigénétique, le complexe NuRD. NuRD est également responsable de répression transcriptionnelle en entrainant la compaction de la chromatine, mais cette fois-ci par désacétylation des histones. La quasi-totalité des protéines composant ce complexe est retrouvée dans l'interactome proximal de LT330. Nous retrouvons notamment l'intégralité du sous complexe désacetylase, à savoir, HDAC1/2, MTA1/2 et RBBP4. RBBP7, qui fait également parti de ce complexe, a été identifié dans l'interactome de LT330 mais juste sous le seuil de significativité d'interacteur de haute confiance que nous avons fixés (ΔLog2(intensité LFQ) = 0.85 <1). Ces interactions ouvrent de nouvelles pistes expérimentales de caractérisation des mécanismes épigénétiques transformants associés à l'expression de tLT.

Parmi ces facteurs épigénétiques observés au sein de l'interactome proximal de tLT, nous avons retrouvées 65 protéines impliquées dans la modification des histones en interaction avec LT330. De manière intéressante, 22 d'entre elles sont plus spécifiquement impliquées dans la méthylation des histones. Les publications récentes concernant l'implication de EZH2 et LSD1 dans les CCM soulignent l'intérêt d'étudier cette modification post-traductionnelle en particulier, et de se pencher sur son rôle dans l'altération du profil épigénétique des CCM. Nous retrouvons plusieurs lysine démethylases, comme les protéines KDM1A/3B/6A. L'intégralité du complexe EHMT2-EHMT1-WIZ est identifié en interaction avec LT330 et LT428. Ce complexe est responsable des mono- et di-methylation de l'histone H3K9. La marque épigénétique H3K9me2 est associée à la compaction de la chromatine et donc à la répression transcriptionnelle. Cette méthylation réversible peut être retirée par la démethylase KDM3B, également en interaction avec ces deux formes de tLT. La protéine EHMT2 a récemment été associée à la prévention des dommages liés à l'ADN. Gaggiola *et al.* ont observé l'augmentation de la marque H3K9me2 au niveau des fourches bloquées en cas de stress réplicatif. La compaction de la chromatine qui est ainsi induite permettrait d'éviter la survenue de cassures de l'ADN (Gaggioli *et al.*, 2023).

De manière inattendue, l'analyse des interacteurs proximaux des antigènes T a révélé une forte présence de protéines impliquées dans la réponse aux dommages de l'ADN. Une analyse de « Gene Ontology » a révélé la présence de 69 protéines impliquées dans cette voie. Nous retrouvons notamment des acteurs bien connus de la voie de réponse aux dommages de l'ADN (DDR, « DNA Damage Response ») comme TP53BP1, TOPBP1, MDC1, RPA1/2, RFC3 ou XRCC4. Il est intéressant de noter que la méthyltransférase EHMT2, bien décrite pour son rôle de méthylation des histones, est également capable de méthyler des cibles non-histones. Il a notamment été observé que l'activité de recrutement de ATM par MDC1 au niveau des cassures doubles brins est dépendant de la méthylation de la protéine MDC1 par EHMT2 (Watanabe *et al.*, 2018). De la même manière, TOPBP1, qui est un cofacteur activateur de ATR dans des contextes de cassures simple brin et de stress réplicatif, est régulé par méthylation. La méthylation de cette protéine régule son affinité avec la forme phosphorylée de la protéine ATR. L'enzyme responsable de la déméthylation de TOPBP1 est PHF8 tandis que celle responsable de sa méthylation n'a pas été identifiée (Ma *et al.*, 2021).

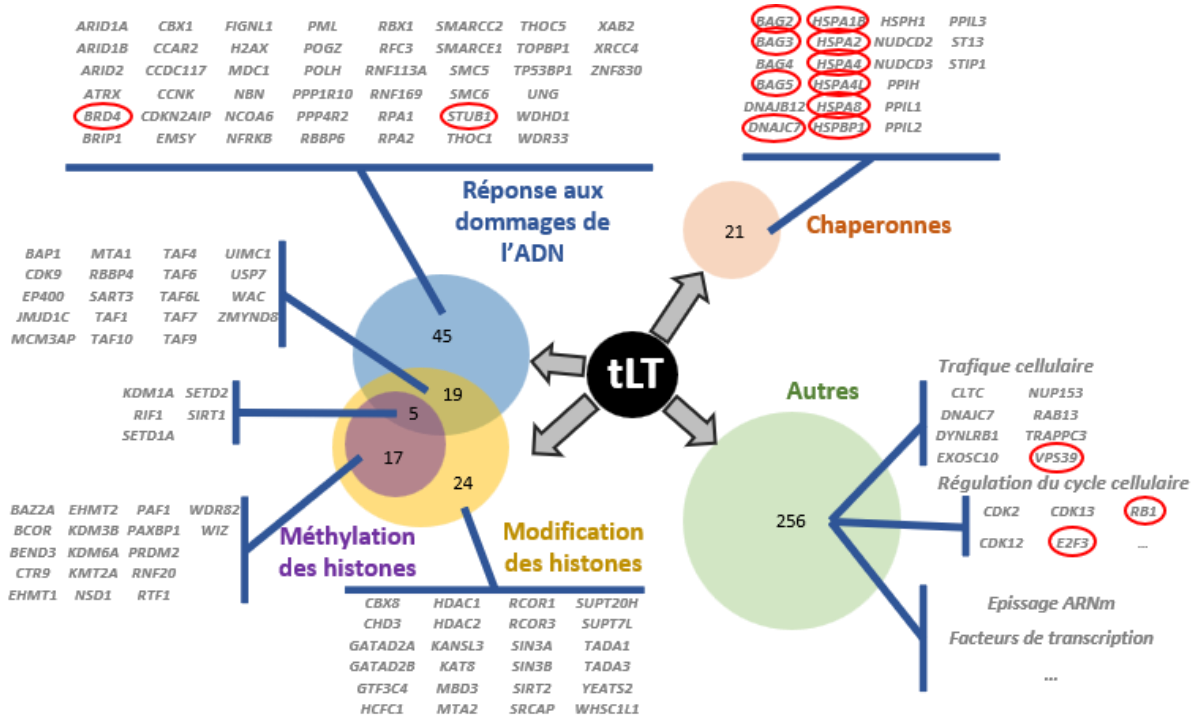


Figure 10 : Interacteurs proximaux du LT tronqué réparti selon les processus biologiques les impliquants.

Les interacteurs de la forme LT330 identifiés en BioID sont présentés selon les catégories de processus biologiques identifiés par analyse d'enrichissement de « Gene Ontology ». Les noms de protéines entourés d'un cercle rouge représentent les interacteurs déjà connus du LT du MCPyV.

4.1.1.2.3 LT pleine taille

La forme complète de l'antigène grand T a également été utilisée. Cette forme conserve sa partie C-terminal portant l'activité ATPase/hélicase. La réalisation du BioID sur cette forme n'a été possible qu'avec la fusion de BirA* à l'extrémité N-terminale de la protéine. En effet, la protéine de fusion présentant BirA* à l'extrémité C-terminale semblait manquer de stabilité et n'a pas permis l'acquisition de données probantes en spectrométrie de masse.

Une fois de plus, les principales interactions formées par LT sont retrouvées en interaction avec la forme complète de la protéine. Nous retrouvons notamment Vam6p, HSPA1B/2/4/4L/6/8, HSP90BP1, HSPH1 et HSPBP1 (**ANNEXE I)**. Nous retrouvons également plusieurs protéines BAG comme pour les formes plus courtes (BAG2-5) et DNAJ (B12 et C7). La protéine RB1 dont l'interaction avec LT est bien connue n'apparait pas parmi les partenaires de la forme complète. Cette protéine est cependant identifiée comme biotinylée mais la q-value étant de 0,02, la comparaison statistique avec les contrôles ne permet pas de considérer RB1 comme interacteur de haute confiance. La forme LT pleine taille n'étant jamais exprimée dans les CCM-VP, nous n'avons pas poursuivi ces investigations.

4.1.2 Identification des interacteurs des antigènes T par GST-Pull Down

Nous nous sommes également tournés vers une autre technique ne nécessitant pas la transfection des cellules, le GST-Pull Down. Cette technique repose sur la purification d'affinité et permet ainsi l'utilisation de lysats de cellules CCM-VP. Nous avons pour cela exprimé en bactérie BL-21 les protéines virales recombinantes en fusion avec la GST (Gluthatione S Transferase) codées en vecteur pGEx. Après lyse des bactéries, les protéines de fusion ont été fixées sur billes magnétiques recouvertes de gluthatione par affinité gluthatione-GST. Ces colonnes des protéines d'intérêt ont ensuite été utilisées pour capturer les protéines capables d'interagir avec les antigènes T dans des lysats de cellules CCM-VP. Nous pouvons ainsi identifier les protéines biochimiquement capables d'interagir avec une protéine d'intérêt *in vitro*. Nous avons réalisé les GST-Pull Down sur les protéines sT et LT tronqué. Nous avons ici utilisé la forme LT330 qui est retrouvée dans plusieurs lignées cellulaires issues de patients. Nous avons utilisé des lysats obtenus à partir de cellules MKL-1 et WaGa. Ces deux lignées sont des lignées CCM-VP exprimant sT et une forme tronquée de LT. Les cellules MKL-1

expriment la forme LT330 et les cellules WaGa expriment une forme plus courte de 273 acides aminés.

Suite à la soustraction du bruit de fond correspondant aux protéines capturées sur les colonnes GST seul (interactions non-spécifiques avec les billes ou avec la partie GST), nous avons identifié 102 protéines pouvant interagir spécifiquement avec LT330 au sein des lysats MKL-1 et 123 au sein des lysats WaGa. La majorité de ces interacteurs est composée de protéines ribosomales et spliceosomales. Ces protéines sont très abondantes dans la cellule et peuvent être surreprésentées dans ce type d'approche. Les protéines de choc thermique HSP90B1 et HSPH1 sont retrouvées parmi les interacteurs de tLT dans le lysat des WaGa. Ces protéines sont également retrouvées dans les échantillons MKL-1, mais suite à une variabilité inter-échantillon trop importante, la q-value obtenue ne permet pas leur qualification en tant qu'interacteurs de haute confiance. Parmi les protéines d'intérêt retrouvées en interaction avec tLT dans les lysats MKL-1 et WaGa, PIN1 est identifiée. PIN1 est une peptidyl-prolyl cis/trans isomérase et catalyse l'isomérisation des résidus Proline de ses substrats protéiques. L'impact de l'isomérisation varient selon la protéine touchée. En effet, ce changement conformationnel, comparable à un coude plié ou déplié d'un point de vue structural, peut résulter en une déstabilisation ou une stabilisation du substrat, régulant ainsi la demi-vie de la protéine. L'activité catalytique d'une enzyme ou sa localisation peuvent également être affectées par ce mécanisme. L'E3 Ubiquitine ligase FBXW7, supposément impliqué dans les CCM-VP, fait partie des cibles connues de PIN1. Il a été démontré qu'*in vitro*, l'isomérisation de FBXW7 par PIN1 entraine un changement conformationnel induisant son auto-ubiquitination conduisant ainsi à son adressage au protéasome (Min *et al.*, 2012). Il a été observé que l'expression de tLT entrainait une diminution de la quantité de FBXW7 sans que la quantité d'ARNm ne soit impactée (Dye *et al.*, 2019). Ainsi, il semblerait qu'il existe un mécanisme tLT-dépendant de régulation de cette protéine. Il est envisageable que l'interaction entre tLT et PIN1 puisse participer à cette régulation. Nous retrouvons également la protéine CAND1 en interaction avec tLT dans les lysats des deux lignées CCM-VP. CAND1 est recrutée aux SCF complexes en alternance avec la neddylation et est responsable de l'échange de la F-box, sous-unité de ciblage des complexes SCF, assurant de la spécificité de substrat. Cette interaction entre tLT et un régulateur des SCF complexes représente également une

possibilité pour les protéines virales d'influencer la dégradation protéique des cibles de ces complexes. Nous reviendrons plus loin sur l'étude de ce mécanisme potentiel.

Comme pour le BioID, plusieurs protéines impliquées dans la réparation de l'ADN sont retrouvées en interaction avec tLT. La protéine XRCC5 est notamment retrouvée en interaction avec tLT au sein des lysats MKL-1 et WaGa. Cette protéine fonctionne en hétérodimère avec XRCC4 que nous retrouvons en interaction avec LT330 en BioID. Ce complexe est essentiel dans la réparation de cassures double brin de l'ADN par recombinaison non homologue (NHEJ) (Tai *et al.*, 2000).

Plusieurs interactions avec des protéines impliquées dans la régulation de l'apoptose ont également été observées. La protéine pro-apoptotique BAX (Bcl-2 Associated X) a été identifiée en tant qu'interacteur de LT330 et sT dans les lysats WaGa. Cette protéine a également été retrouvée en abondance sur les billes sT incubées dans les lysats MKL-1, bien que la q-value de 0,09 ne permette pas de la considérer formellement comme un interacteur. Par ailleurs, la protéine sT a été identifiée en interaction avec BCLAF1 (Bcl-2 Associated transcription Factor 1), qui a été retrouvé en BioID en tant qu'interacteur de LT330 et LT428. De plus, nous avons identifié la protéine RBM25 (RNA Binding Protein 25), qui a été caractérisée comme interacteur de tLT par GST-Pull Down dans les lysats MKL-1 et WaGa, ainsi que par BioID avec LT330. Il a été montré que cette protéine est impliquée dans la régulation de l'apoptose via la promotion de l'épissage de formes pro-apoptotiques des ARNm codant les protéines Bcl-X (Zhou *et al.*, 2008). Ainsi, les antigènes T sont biochimiquement capables d'interagir avec plusieurs protéines reliées à la régulation de l'apoptose dépendante de Bcl-2. De manière intéressante, il a été observé lors d'un criblage de plusieurs lignées CCM-VP, CCM-VN et cellules non CCM, que l'inhibition de Bcl-2 par le Navitoclax permettait un arrêt de croissance de manière spécifique dans les CCM-VP (Garman *et al.*, 2021).

L'utilisation de lysats des lignées CCM-VP MKL-1 et WaGa permet l'observation d'interactions possibles avec des protéines exprimées en présence des antigènes T. Outre la confirmation de certaines interactions et l'observation des voies de régulation des ARNs ou de la réparation de l'ADN, le GST-Pull Down a également mis en évidence une interaction entre la protéine sT et la protéine tLT. En effet, il a été possible d'identifier la protéine tLT spécifiquement retenue par les billes présentant sT à leur surface. Il est donc envisageable que ces deux protéines virales puissent coopérer sous forme de complexe dans les CCM-VP.

4.1.3 Influence des antigènes T sur le protéome

Les protéines virales étudiées ici ne possèdent pas d'activité catalytique connues. L'antigène grand T à l'état sauvage possède une activité ATPase/hélicase au niveau de sa partie C-terminale mais la région portant cette activité est systématiquement perdue lors de la troncation requise à l'émergence d'un CCM-VP. La reprogrammation oncogénique médiée par ces protéines virales repose donc probablement sur l'altération des fonctions de leurs partenaires. Ces fonctions peuvent être impactées de différentes manières par la présence des antigènes T. Il peut s'agir d'une séquestration qui inhibe la fonction d'une protéine d'intérêt, comme dans le cas de l'interaction tLT-RB1. Il peut également s'agir d'un recrutement par un antigène T vers certaines cibles comme dans le cas de sT et MYCL-EP400 ou LT et HSC70. Les analyses d'enrichissement réalisées sur les interacteurs identifiés ont révélées de nombreuses protéines impliquées dans la régulation de l'expression génique. Ces protéines virales semblent donc agir à ce niveau par le biais de dérégulations épigénétiques et transcriptionnelles. Nous nous sommes intéressés à l'effet de l'expression des antigènes T sur le contenu protéique des cellules. Pour cela, des cellules WaGa présentant une inhibition des antigènes T par expression inductible de shRNA ont été utilisées. Ces cellules ont été traitées dix jours à la doxycycline afin d'induire l'inhibition des antigènes T, durée nécessaire pour que les antigènes T ne soient plus détectables au niveau protéique et que les modulations du protéome soient bien installées. Le contenu protéique a été analysé par spectrométrie de masse. Les abondances relatives de chaque protéine identifiée ont été comparées entre les cellules WaGa-shTA (WaGa présentant l'expression du shRNA ciblant les antigènes T) et les cellules WaGa WT.

Il a ainsi été possible d'identifier 72 protéines dont l'abondance est significativement augmentée (au minimum doublée) dans le cas de l'inhibition des antigènes T par rapport aux WT. En plus de ces 72 protéines surexprimées en présence du shRNA, 29 protéines ont été identifiées exclusivement dans les WaGa-shTA. A l'inverse, 110 protéines sont surexprimées et 40 exclusivement exprimées dans ces cellules WaGa WT. Au total, nous identifions 101 protéines plus/uniquement exprimées suite à l'inhibition des antigènes T et 150 protéines plus/uniquement exprimées en présence des antigènes T (**Figure 11 ; ANNEXE II**).

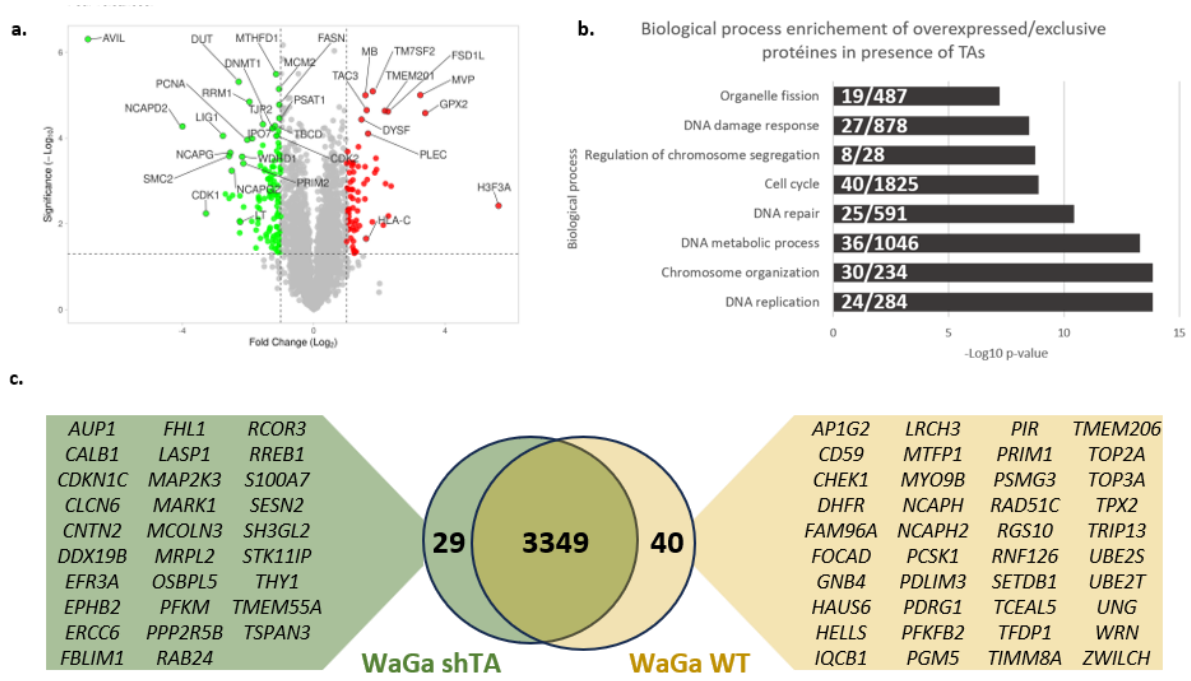


Figure 11 : Analyse de l'expression différentielle du protéome des cellules WaGa (CCM-VP) suivant l'inhibition de l'expression des antigènes T.

***(A)** Volcano plot représentant les variations d'abondances (Log2 fold change) mesurées par LFQ (Label Free Quantification) en fonction du -Log10 de la p-value. Les protéines dans le cadrant supérieur gauche sont plus abondantes dans les WaGa WT et les protéines dans le cadrant supérieur droit sont plus abondantes dans les WaGa shTA. **(B)** Analyse d'enrichissement des processus biologiques dans lesquels sont impliqués les protéines plus abondantes dans les WaGa WT. **(C)** représentation des protéines identifiées exclusivement dans l'une des deux conditions (valeurs valides dans 3 réplicats contre valeurs absente dans 3 réplicats).*

Parmi les protéines retrouvées exclusivement ou en plus forte abondance en présence des antigènes T, il a été possible d'identifier de nombreuses protéines impliquées dans la réplication et le contrôle du cycle cellulaire. Les protéines CDK1 et CDK2 sont significativement plus abondante en présence des protéines virales. A l'inverse, l'inhibiteur de CDK CDKN1C est exprimé exclusivement dans les cellules WaGa-shTAs. De manière intéressante, la protéine HELLS (« Lymphoid-specific helicase ») qui est un régulateur négatif de CDKN1C et est retrouvé en interaction avec la forme tronquée de LT, LT428, est retrouvée uniquement dans les cellules WaGa WT. A l'image des kinases dépendantes de cyclines, de nombreuses autres protéines impliquées dans la régulation du cycle cellulaire sont exprimées de manière différentielle entre les deux conditions. C'est notamment le cas de l'intégralité des protéines du complexe condensine. Chacune des protéines composant ce complexe est en effet significativement plus exprimée en présence des antigènes T. La totalité des protéines composant le complexe de réplication « replication factor C » sont également surexprimées

en présence des antigènes T. Ce complexe est essentiel au déroulement de la réplication et est également impliqué dans le maintien des télomères et la réparation de l'ADN, notamment en cas de mésappariement entre les brins. Dans le contexte de la réplication, les protéines RFC permettent l'ancrage de la protéine PCNA (« Proliferating cell nuclear antigen ») qui est responsable du recrutement des ADN polymérases (Kang *et al.*, 2019). Nous retrouvons également cette protéine PCNA et deux sous-unités de la polymérase delta surexprimées en présence de TAs. Les sous-unités exprimées différentiellement entre les conditions sont la 1 et la 2 (POLD1 et POLD2). POLD1 est la sous-unité catalytique de la polymérase delta, elle est impliquée dans la synthèse du brin retardé mais également dans la réparation de l'ADN dans le cadre de mésappariement ou de cassure double brin. La polymérase delta est considérée comme une enzyme majeure de la synthèse d'ADN dans les cas de dommages de l'ADN, notamment dans le mécanisme de recombinaison homologue (Tumini *et al.*, 2016). L'expression de la protéine RAD51C, également impliquée dans la recombinaison homologue, est détectée uniquement dans les cellules exprimant les antigènes T. Au total, l'analyse d'enrichissement a révélé 27 protéines impliquées dans la réponse aux dommages de l'ADN qui sont retrouvées surexprimées ou exclusivement en présence des TAs (WaGa WT). Parmi celles-ci, la polymérase delta et RAD51C ont déjà été citées. C'est également le cas de CHEK1 impliquée dans l'arrêt de cycle cellulaire et la transduction du signal pour la réponse aux dommages de l'ADN. De la même manière, RIF1 qui est retrouvée en interaction avec les formes LT330 et LT428 est surexprimée en présence des protéines virales. Cette protéine est un régulateur de TP53BP1, un autre interacteur de LT330 impliqué dans la réparation des cassures double brins.

La protéine RCOR3 est retrouvée exclusivement en absence des antigènes T ce qui suggère une inhibition de son expression par les protéines virales. Cette protéine a été identifiée comme inhibiteur de l'action de LSD1 sur la différenciation cellulaire (Upadhyay *et al.*, 2014). Il a été observé au sein des CCM-VP que l'expression de LSD1 médiée par sT et son action sur les cibles du facteur de transcription ATOH1 étaient essentielles au développement des CCM-VP (Park *et al.*, 2020). Ainsi, l'inhibition de l'expression de RCOR3 par les antigènes T permettrait de lever l'inhibition exercée sur LSD1 par cette protéine, favorisant l'activité oncogénique de LSD1 dans les CCM-VP.

Ces résultats indiquent également que la glycoprotéine CD59 est exclusivement détectée dans les cellules exprimant les antigènes T. Il a été montré que CD59 est le principal inhibiteur de la formation du complexe d'attaque membranaire, régulant ainsi l'activation du complément (Zhang *et al.*, 2018). Au niveau moléculaire, CD59 se lie au complexe C5b-8 et empêche le recrutement de C9, bloquant la formation du complexe d'attaque membranaire. L'induction de l'expression de CD59 médiée par les antigènes T pourrait permettre aux cellules tumorales d'échapper à la réponse immunitaire innée. De plus, une forte expression de CD59 dans le cancer du poumon est corrélée à une faible infiltration de macrophages M1 pro-inflammatoires et de lymphocytes T CD8+ antitumoraux (Shao *et al.*, 2022). En plus de l'induction de l'expression de CD59, la présence des TAs semble responsable de l'inhibition de l'expression de HLA-C qui a été détecté de manière plus abondante dans les cellules WaGa-shTAs. Il s'agit d'une protéine présentatrice d'antigène qui permet notamment l'activation des lymphocytes NK (« Natural Killer ») par la présentation de peptides antigéniques (Verheyden *et al.*, 2009). Par ailleurs, la protéine HLA-C impliquée dans la présentation antigénique est plus abondante en absence des antigènes T, suggérant une inhibition de son expression par les protéines virales. La présentation d'antigènes médiée par HLA-C est impliqué dans l'activation de lymphocytes NK (« Natural Killer ») qui ont un effet anti tumoral. Il semble donc que la down régulation de l'expression de HLA-C en présence des antigènes T puisse aussi participer à l'évasion immunitaire des CCM-VP. Il est également intéressant de préciser que lors de l'analyse, HLA-B ne répondait pas aux critères pour être considéré comme protéine exclusive à une condition mais une valeur valide de LFQ a pu être mesurée dans 2 des 3 réplicats des cellules présentant une inhibition de l'expression des TAs, tandis que cette protéine n'a pas été détectée dans les cellules WT.

La protéine Advillin (AVIL) a été quantifiée comme étant la protéine présentant la plus grande variation d'abondance entre les deux conditions, avec un fold change de 6,87 en échelle logarithmique en base 2 (Log2FC). Ainsi, le niveau de AVIL est plus de cent fois plus élevé en présence des antigènes T. Il a été montré que AVIL est exprimée dans les neurones dérivés des crêtes neurales et dans les cellules de Merkel (Hunter *et al.*, 2018). Nos résultats indiquent donc que l'expression de ce marqueur spécifique des cellules de Merkel est fortement influencée par la présence des protéines virales. Cette observation appuie

l'hypothèse selon laquelle les CCM-VP dériveraient d'un progéniteur épithélial, les antigènes T induisant l'acquisition d'un phénotype de type cellule de Merkel.

4.2 PIN1 COMME NOUVEAU REGULATEUR DES CRLs

Les complexes SCF, sous famille des CRLs (Cullin Ring Ligase), ont été mis en cause dans l'oncogenèse CCM. Il est supposé que la protéine virale sT interagisse avec certaines F-box des complexes SCF via sont LSD. Selon cette hypothèse, la présence de sT aurait un effet inhibiteur sur l'action de ces complexes d'ubiquitination. La stabilisation de la protéine tLT serait donc la résultante de l'inhibition de son ubiquitination par les complexes SCF. Comme détaillé précédemment, cette hypothèse a été remise en question par une étude réalisée *a posteriori*. En effet, les auteurs n'ont pas pu confirmer d'interactions entre sT et la protéine FBXW7. Ces auteurs ont néanmoins observé une diminution de la quantité de FBXW7 lors de l'expression de la protéine tLT. Cette diminution d'expression n'étant pas corrélée à une diminution de la quantité d'ARNm, c'est au niveau protéique que la régulation a lieu. Nous avons identifié dans la partie précédente une interaction entre la protéine tLT est la protéine PIN1. Cette protéine est un catalyseur de changement conformationnel par isomérisation des prolines. De manière intéressante, l'action de PIN1 est reliée à la stabilité de la protéine FBXW7. En effet, il a été observé que cette sous-unité de SCF/C pouvait subir une isomérisation catalysée par PIN1. Ce changement structural induit alors une auto-ubiquitination de FBXW7, l'adressant ainsi au protéasome. Il est alors envisageable que la régulation des SCF/C, et notamment de FBXW7, puisse impliquer PIN1 et tLT.

4.2.1 Implication de PIN1 dans les CCM

Les analyses exploratoires croisées aux informations disponibles dans la littérature ont mené à la suspicion d'un effet de PIN1 associé à tLT dans les CCM-VP. Afin d'étudier cette possibilité et d'élucider les mécanismes potentiels, il est nécessaire de confirmer l'interaction entre PIN1 et tLT.

4.2.1.1 Interaction tLT-PIN1

Cette interaction ayant été identifiée par GST pull-down, une méthode de validation orthogonale a été utilisée. L'interaction entre PIN1 et tLT a ainsi été étudiée par PLA (Proximity Ligation Assay) (**Figure 12**). Deux lignées CCM-VP ont pour cela été utilisées, les lignées MKL-1 et WaGa présentant des formes tronquées du LT de 330 et 278 acides aminés respectivement (Schrama *et al.*, 2019). L'utilisation des cellules MKL-1 permet la confirmation

de l'interaction entre PIN1 et la forme LT330 avec laquelle cette interaction a été identifié. L'utilisation des WaGa permet quant à elle de savoir si une forme plus courte du tLT est également capable d'interagir avec cette protéine. Les cellules non incubées avec des anticorps primaires ainsi que les cellules incubées avec chaque anticorps primaire seul ont été utilisées comme contrôles afin d'observer le signal aspécifique assimilé à du bruit de fond.

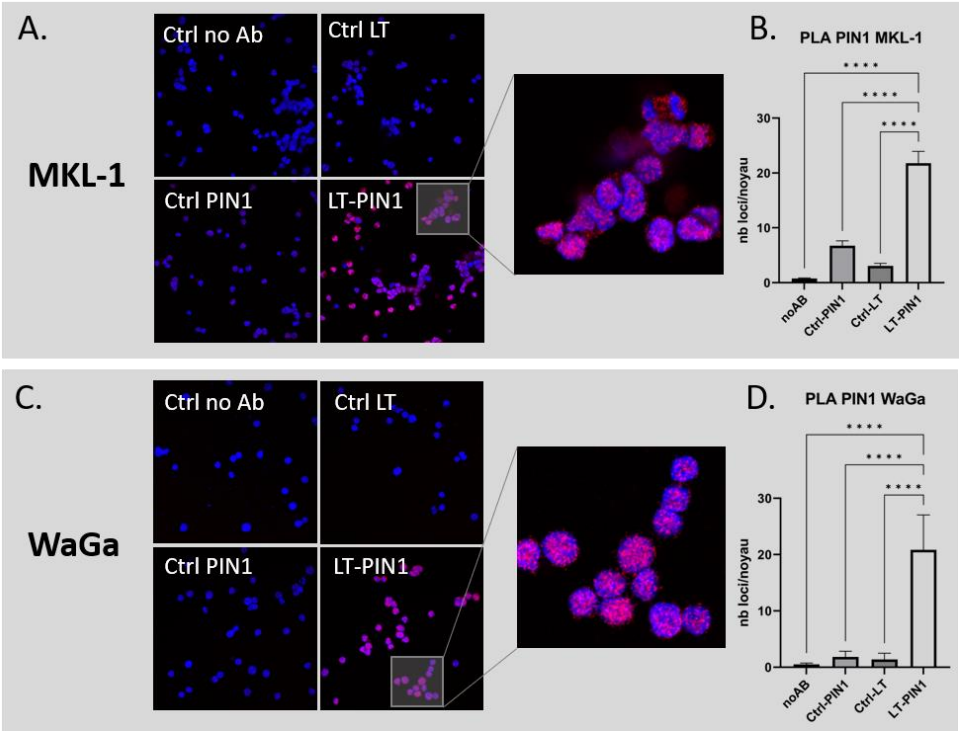


Figure 12 : Validation de l'interaction tLT-PIN1 par Proximity Ligation Assay

***(A)** Visualisation par microscopie confocale des signaux fluorescent de PLA dans les contrôles et la condition testée LT-PIN1 au sein de cellules MKL-1. **(B)** Représentation graphique du nombre de foci de fluorescence par noyau dans chaque condition au sein des cellules MKL-1. La significativité des différences est calculée par ANOVA. **(C)** Visualisation par microscopie confocale des signaux fluorescent de PLA dans les contrôles et la condition testée LT-PIN1 au sein de cellules WaGa. **(D)** Représentation graphique du nombre de foci de fluorescence par noyau dans chaque condition au sein des cellules WaGa. La significativité des différences est calculée par ANOVA.*

Il a ainsi été possible de confirmer l'interaction entre tLT et PIN1. Les deux formes tronquées de la protéine virale ont été retrouvées en interaction avec PIN1.

4.2.1.2 Expression de PIN1 dans les CCM

Dans le but d'évaluer la pertinence de l'hypothèse impliquant PIN1 et FBXW7, les niveaux de ces protéines ont été mesurés au sein de quatre lignées cellulaires. Les cellules utilisées pour cela sont une nouvelle fois les CCM-VP MKL-1 et les WaGa. Ces cellules ont été comparées aux cellules MCC13 et MCC26 qui sont deux lignées CCM-VN. Le niveau de la protéine Jun qui est une cible de FBXW7 a également été mesuré afin d'évaluer un possible déficit de son ciblage pour dégradation.

Pour les trois protéines d'intérêt, les niveaux d'expression observés sont similaires entre les lignées ayant le même statut viral. D'importantes différences sont cependant observées entre les CCM-VP et les CCM-VN (**Figure 13**). Nous observons une quantité de PIN1 plus importante au sein de cellules présentant l'expression des antigènes T. La présence des protéines virale semble donc influencer la quantité de PIN1 au niveau de son expression ou de sa stabilité. L'interaction entre tLT et PIN1 peut ainsi avoir une fonction de stabilisation. Il est également possible que cette expression accrue de PIN1 soit le résultat de l'interaction bien connue entre tLT et RB1. Il a en effet été observé dans le mélanome que l'inhibition de l'expression de RB1 entrainait une augmentation de l'expression de PIN1 par l'action des facteurs E2F (Kim *et al.*, 2018). Dans le cadre des CCM, la séquestration de facteurs E2F par RB1 est empêchée par tLT. Ainsi, les facteurs E2F libres peuvent induire l'expression de PIN1. Il est également envisageable que la quantité de PIN1 observée résulte d'une augmentation d'expression associée à une stabilisation par tLT. La quantité de la protéine FBXW7 est également similaire entre les cellules de même statut viral. Une nouvelle fois, cette quantité varie entre les CCM-VP et les CCM-VN. Les cellules exprimant les protéines virales présentent une quantité de FBXW7 inférieure à celle des CCM-VN. Ces résultats s'accordent avec ce qui a été observé par Dye *et al.*. Les auteurs ont observé une diminution de la quantité de FBXW7 sans diminution de la quantité de son ARNm en présence de tLT (Dye *et al.*, 2019). Les lignées cellulaires CCM-VP présentent donc une plus grande quantité de PIN1 et une plus faible quantité de FBXW7. Le niveau protéique de la protéine Jun a également été mesuré. Comme pour les deux autres protéines d'intérêt, les variations de niveaux dépendent du statut viral des cellules. La quantité de Jun observée est cependant plus importante dans les lignées CCM-VN. Cette protéine étant adressée au protéasome par FBXW7 qui est moins présent dans les CCM-VP, ce résultat est inattendu. Il est envisageable qu'une autre voie de dégradation soit

responsable de la plus faible quantité de JUN. Cette protéine étant un proto-oncogène connu, il est également possible que les lignées CCM-VN présentent une forte expression de cette protéine.

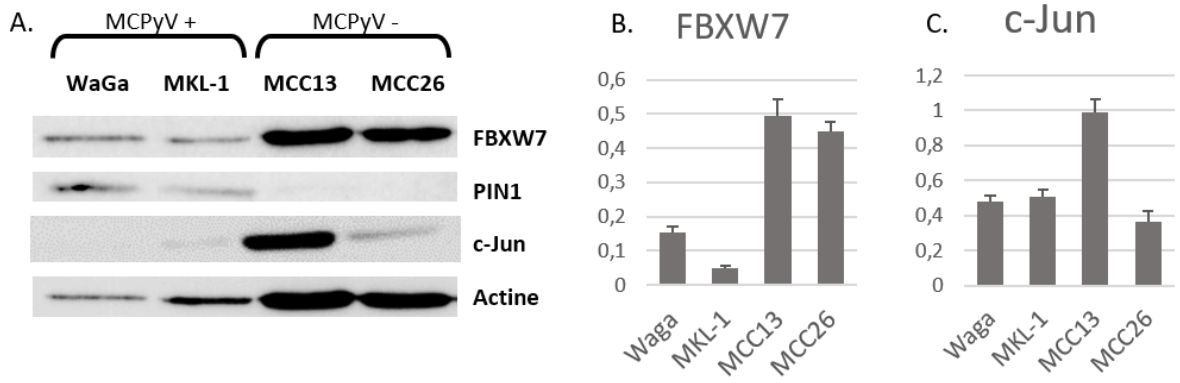


Figure 13 : Comparaison des niveaux de PIN1-FBXW7 et c-Jun au sein de cellules CCM -VP et -VN.

***(A)** Western blotting utilisant les cellules WaGa et MKL-1 qui sont des CCM-VP et MCC13 et MCC26 qui sont des CCM-VN. La normalisation est faite en utilisant la quantité de protéine déposée dans chaque puit. **(B)** Quantification de FBXW7 utilisant l'actine comme normalisateur. **(C)** Quantification de c-Jun en utilisant l'actine comme normalisateur.*

Ces observations par Western-Blot a également mis en lumière certains problèmes liés à la normalisation du signal. Dans le but de normaliser la quantité de protéine déposée dans chaque puit, la quantité de protéine présente dans chaque lysat a été quantifiée. La même quantité de protéine a été déposée pour chaque lignée cellulaire. L'actine est communément utilisée comme témoin de charge, l'immuno-blotting de l'actine a donc également été réalisée afin de doubler la normalisation. Il a alors été possible d'observer que l'intensité du signal d'actine ne correspondait pas entre les différents puits. Le signal mesuré pour les CCM-VN était plus intense que celui observé dans le cas des CCM-VP. Il semble donc que les deux types de normalisation ne correspondent pas. Plusieurs explications sont possibles pour cela. Il a été observé dans le cas de la dengue que la présence de protéine virale pouvait entrainer une modification du cytosquelette. Ainsi, l'utilisation de l'actine comme base pour la normalisation présente un biais dans les cas de comparaisons de cellules ayant un statut viral différent (Suttitheptumrong *et al.*, 2021). De plus, les cellules CCM-VN sont des cellules adhérentes tandis que les CCM-VP poussent en suspension, ce qui peut expliquer que la quantité d'actine varie autant entre ces types cellulaires.

L'interaction entre PIN1 a donc pu être validée avec plusieurs formes de LT tronqués. Des différences de niveau d'expression de cette protéine ont été observées entre les formes

virus positives et les formes virus négatives. La protéine PIN1 est ainsi présente en plus grande quantité dans les cellules présentant l'expression des antigènes T. Le niveau de FBXW7 potentiellement régulé par PIN1 varie également entre les cellules ayant un statut viral différent. La quantité de cette protéine est plus faible dans les CCM-VP que dans le cas des CCM-VN. Ainsi, dans le cadre du carcinome à cellules de Merkel, une plus faible quantité de FBXW7 semble être associée à une plus forte expression de PIN1 et à la présence des protéines virales.

Les niveaux de PIN1 observés sur lignée CCM-VP et -VN ont également été mesurés sur coupe de patients au sein d'un « tumor tissue array » (TMA) de 150 CCM par le Dr Thibault Kervarrec. Il a été possible de constater une expression plus importante de PIN1 chez les patients CCM-VP que dans le cas des patients CCM-VN (**Figure 14**).

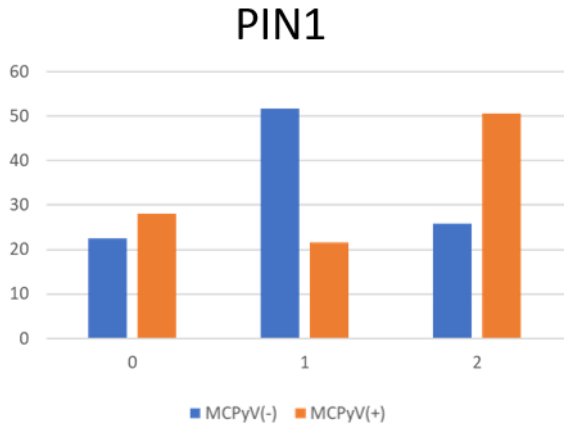


Figure 14 : Quantification de l'expression de PIN1 par Immuno Histochimie sur Tissue Micro Array.

L'expression protéique de PIN1 a été quantifiée sur coupe de patients CCM-VP (orange) et -VN (bleu). Les valeurs en abscisse correspondent à des niveaux d'expression (0 = non détecté/très faible ; 1 = expression faible ; 2 = expression forte). Les valeurs en ordonnées correspondent au pourcentage de patient exprimant PIN1 dans chacun des trois niveaux.

4.2.2 PIN1 en interaction avec des composants du système d'ubiquitinylation

Les interactions formées par la protéine PIN1 ont été identifiées par TurboID au sein des cellules 293 Flp-In T-REx. La version modifiée de la biotine ligase qui est utilisée en TurboID a été fusionnée à l'extrémité N-terminale de la protéine. Le vecteur codant la protéine de fusion a été inséré au génome de manière stable par système Flp-In, comme précédemment pour les BioID réalisés sur les antigènes T. La protéine de fusion a été exprimée pendant 24h et la durée de biotinylation utilisée était de 1h. Il a ainsi été possible d'identifier 406 protéines interagissant avec PIN1. Parmi ces interacteurs identifiés se trouvent plusieurs cibles connues de PIN1 comme Jun, RB1 ou NF-kB. L'analyse des partenaires de PIN1 a également révélé la présence de 27 protéines impliquées dans le système ubiquitine protéasome (UPS). Nous pouvons ainsi retrouver certains composants majeurs des SCF/C comme SKP1 et RBX1. Les

complexes SCF remplissent un rôle de E3 ubiquitine ligase. Ils permettent donc la fixation du substrat auquel sera lié l'ubiquitine transitant par la partie E2. Au sein de ce complexe, la protéine F-box est responsable de la spécificité de substrat. FBXW7 est une protéine F-box qui fait partie d'un SCF/C. Cette protéine déjà suspectée d'être impliquée dans l'oncogenèse CCM-VP a également été mise en relation avec PIN1. Ainsi, PIN1 est responsable de l'ubiquitinylation de FBXW7 menant à son adressage au protéasome (Min *et al.*, 2012). L'interaction PIN1-FBXW7 menant à la dégradation de cette dernière, il n'est pas surprenant de ne pas la détecter parmi les interacteurs identifiés en TurboID.

De nombreuses E3 ubiquitine ligase étant présente dans l'interactome de PIN1, il est envisageable que ce mode de régulation par changement conformationnel concerne de plus nombreux composants de la chaine d'ubiquitination. En effet, la géométrie des complexes SCF et cullin-ring ligases (CRL) en général suggère la possibilité de modulation de leur activation par une isomérisation. L'étude de la structure de ces complexes a révélé une distance d'environ 50Å entre l'E2 ubiquitine ligase et la protéine substrat (Wu *et al.*, 2003). Cette distance ne permettant pas le transfert de l'ubiquitine de l'E2 vers le substrat, un changement conformationnel est nécessaire au déroulement de ce processus. Il a été observé que la neddylation des cullins avait un effet activateur sur le fonctionnement de ces complexes d'ubiquitinylation. Un changement conformationnel suivant la neddylation permet le rapprochement entre l'E2 et le substrat dans le cas des complexes SCF (Saha and Deshaies, 2008). Les CRL sont également influencés par NEDD8 qui régule l'assemblage et l'activation de ce complexe par des cycle de neddylation/déneddylation (Deshaies, Emberley and Saha, 2010). La neddylation est une cascade enzymatique organisée comme celle de l'ubiquitination et permettant la conjugaison de NEDD8, une Ubiquitine-like protein, de manière covalente à ses substrats. Les mécanismes moléculaires induits par la neddylation restent à ce jour obscurs. Les Cullins sont les substrats les mieux décrits de NEDD8 (T *et al.*, 1999). Ces protéines sont des composants majeurs des complexes d'ubiquitinylation CRL. La Cullin 1 entre dans la composition des complexes SCF qui sont les plus étudiés des CRL. Comme dit ci-dessus, la neddylation de CUL1 dans le contexte SCF/C est un prérequis pour l'activation du complexe et induit des changements conformationnels. NEDD8 n'ayant pas d'activité enzymatique connue, il est envisageable qu'il s'agisse d'un signal de recrutement d'autres protéine capable d'induire ces changements de conformation. La peptidyl-prolyl cis/trans isomérase PIN1 est

retrouvée en interaction avec plusieurs composants des SCF complexes et des CRL de manière plus générale. La capacité de cette protéine à catalyser l'isomérisation de ses cibles en fait un bon candidat pour la régulation de l'activité de ces complexes d'ubiquitinylation.

4.2.3 Relation PIN1 et neddylation

Dans le but d'étudier la relation entre PIN1 et NEDD8, nous nous sommes tournés vers le Split-TurboID. Comme expliqué en introduction, cette technologie permet l'étude de l'environnement moléculaire au site d'une interaction entre deux protéines d'intérêt. Ces deux protéines sont exprimées en fusion avec chacune une partie de la biotine ligase TurboID. Lors de l'interaction, les deux parties inactives individuellement s'assemblent et l'activité biotine ligase est reconstituée par complémentation fonctionnelle. Lors du développement de cette technique, les séquences codant les protéines de fusions étaient portées par deux vecteurs différents co-transfectés de manière transitoire. Ainsi, lors de la transfection, les cellules en culture forment un ensemble hétérogène composé de cellules non-transfectées, de cellules ayant reçu l'un ou l'autre des vecteurs et de cellules ayant reçu les deux vecteurs. De plus, plusieurs copies du vecteur sont ainsi présentes au sein des cellules ce qui peut résulter en une surexpression pouvant donner lieu à la formation d'interactions artéfactuelles. Afin de pallier aux problèmes posés par la co-transfection, nous avons modifié un vecteur pcDNA5-FRT-TO dans le but d'y insérer les séquences codant les deux protéines de fusion en les séparant par une séquence IRES. L'IRES (« Internal Ribosome Entry Site ») est une séquence nucléotidique présente sur certains ARNm capable d'adopter une structure tige-boucle permettant l'attachement du ribosome et l'initiation de la traduction coiffe-indépendante. Ainsi, l'ajout de cette séquence permet l'obtention d'un vecteur codant un ARNm bicistronique. De cette manière, chaque vecteur transfecté contient les séquences codantes pour les deux protéines. En outre, l'utilisation d'un vecteur pcDNA5-FRT-TO au sein de cellules 293 TRe-X Flp-In permet l'insertion au sein du génome d'une seule copie du vecteur.

L'utilisation du Split-TurboID nous permet à la fois la confirmation d'une interaction et l'étude de l'environnement moléculaire au site d'interaction. Cette technique a été utilisée avec la partie N-terminal du Split-Turbo fusionnée à l'extrémité N-terminal de PIN1 et la partie C-terminale du Split-TurboID fusionnée à l'extrémité N-terminale de NEDD8. Plusieurs temps de biotinylation ont été utilisés. L'expression des protéines de fusions a été induite par l'ajout

de tétracycline 24h avant la récolte. La biotine a été ajouté au milieu de culture à deux temps différents. Le temps de biotinylation minimal utilisé ici en Split-TurboID est de 4h et le temps maximal est de 24h. Les conditions contrôle utilisées sont Split-seul (partie N-terminale et C-terminale sans protéine d'intérêt fusionnée), Split-PIN1 (PIN1 fusionné à la partie N-terminal coexprimé avec la partie C-terminal seule) et Split-NEDD8 (NEDD8 fusionné à la partie C-terminale co-exprimé avec la partie N-terminale seule).

Il a été possible d'identifier 43 protéines dont l'abondance mesurée en LFQ est au minimum doublée et significativement différente des contrôles (q-value ≤ 0,05). L'identification d'interacteurs spécifiques dans le cadre du Split-TurboID permet dans un premier temps de confirmer l'interaction entre les protéines étudiées (**Figure 15a**). PIN1 et NEDD8 sont suffisamment proches pour que l'activité Biotine ligase de l'enzyme soit reconstituée par complémentation fonctionnelle. La nature de ces interacteurs donne une indication sur l'environnement moléculaire au site d'interaction. Dans le cas de PIN1 et NEDD8, les deux principaux groupes identifiés sont constitués de la machinerie de neddylation et de composants des CRLs. Nous retrouvons notamment les cullin 1,2,3,4B et 5 au voisinage de ces deux protéines. La protéine RBX1 qui entre dans la composition de la majorité des CRLs et la F-box FBXO42 qui régit la spécificité de substrats de certains complexes SCF sont également retrouvés dans l'environnement de PIN1 et NEDD8. La proximité de ces deux protéines est donc confirmée et semble être localisée au niveau des CRLs.

Un temps de biotinylation de 4h semble être suffisant pour obtenir un signal convenable et identifier quelques protéines de l'environnement moléculaire. Dans le but d'obtenir une vision plus large de cet environnement, la biotine a également été ajouté en même temps que la tétracycline afin d'obtenir un temps de biotinylation de 24h. L'utilisation des mêmes conditions contrôles a permis l'identification de 508 interacteurs de haute confiance (**Figure 15b**). Nous retrouvons une nouvelle fois plusieurs composants des complexes SCF comme CUL1, et RBX1. En plus de ces deux protéines, la biotinylation de 24h montre également la présence de SKP1 ainsi que FBXO42 à proximité de PIN1-NEDD8. L'interaction entre ces deux protéines semble donc bien avoir lieu au niveau des complexes SCF. Ce contexte ne semble cependant pas être le seul dans lequel PIN1 et NEDD8 sont retrouvées à proximité. Comme lors de la biotinylation de 4h, de nombreux composants de CRLs de manière générale sont identifiés dans l'environnement moléculaire de cette

interaction. La quasi-totalité des cullins font partie des interacteurs de confiance (CUL1, 2, 3, 4A, 4B, 5). L'activation des CRLs est conditionnée par la neddylation des cullins et il est connu dans le cas des SCF que la distance pré-activation entre l'E2 et le substrat est trop grande pour avoir transfert de l'ubiquitine. L'identification des cullins ainsi que de plusieurs autres composants des complexes SCF dans l'environnement moléculaire de l'interaction PIN1-NEDD8 suggère la possibilité d'une action de cette protéine au niveau des complexes Ubiquitine ligase. En effet, la fonction catalytique de cette protéine permettrait un changement conformationnel de ce complexe réduisant l'écart entre l'E2 et le substrat. Il a été observé que l'activation des CRLs suivant la neddylation se faisait par un changement de conformation de RBX1. Cette protéine fait le lien entre la cullin et l'enzyme E2. Le site d'interaction entre RBX1 et la cullin se trouve à proximité du site de neddylation de cette dernière (Merlet *et al.*, 2009). Il est alors envisageable que la neddylation de la cullin soit un signal de recrutement de PIN1. Le recrutement de PIN1 au niveau des CRLs permettrait de catalyser l'isomérisation de RBX1 permettant ainsi le rapprochement entre l'E2 et le substrat.

Outre les composants de complexes d'ubiquitinylation, de nombreuses protéines ont été identifiées à proximité de l'interaction entre PIN1 et NEDD8. C'est notamment le cas de plusieurs composants du complexe d'initiation de la traduction. Les protéines eIF3I, eIF4B, G2, G3, H et eIF5 sont retrouvées dans l'environnement de PIN1-NEDD8. Il a été observé que l'inhibition de la neddylation empêchait l'assemblages de granules de stress impliquant ces facteurs d'initiation de la traduction. L'étude de ces granules de stress a révélé la nécessité de la présence de NEDD8 au sein des granules pour leur formation (Jayabalan *et al.*, 2016). Ainsi, en plus des CRL, les facteurs d'initiation de la traduction représentent un deuxième exemple de potentiel recrutement de PIN1 par la neddylation. Ces résultats résonnent bien au-delà du cas particulier du carcinome à cellules de Merkel.

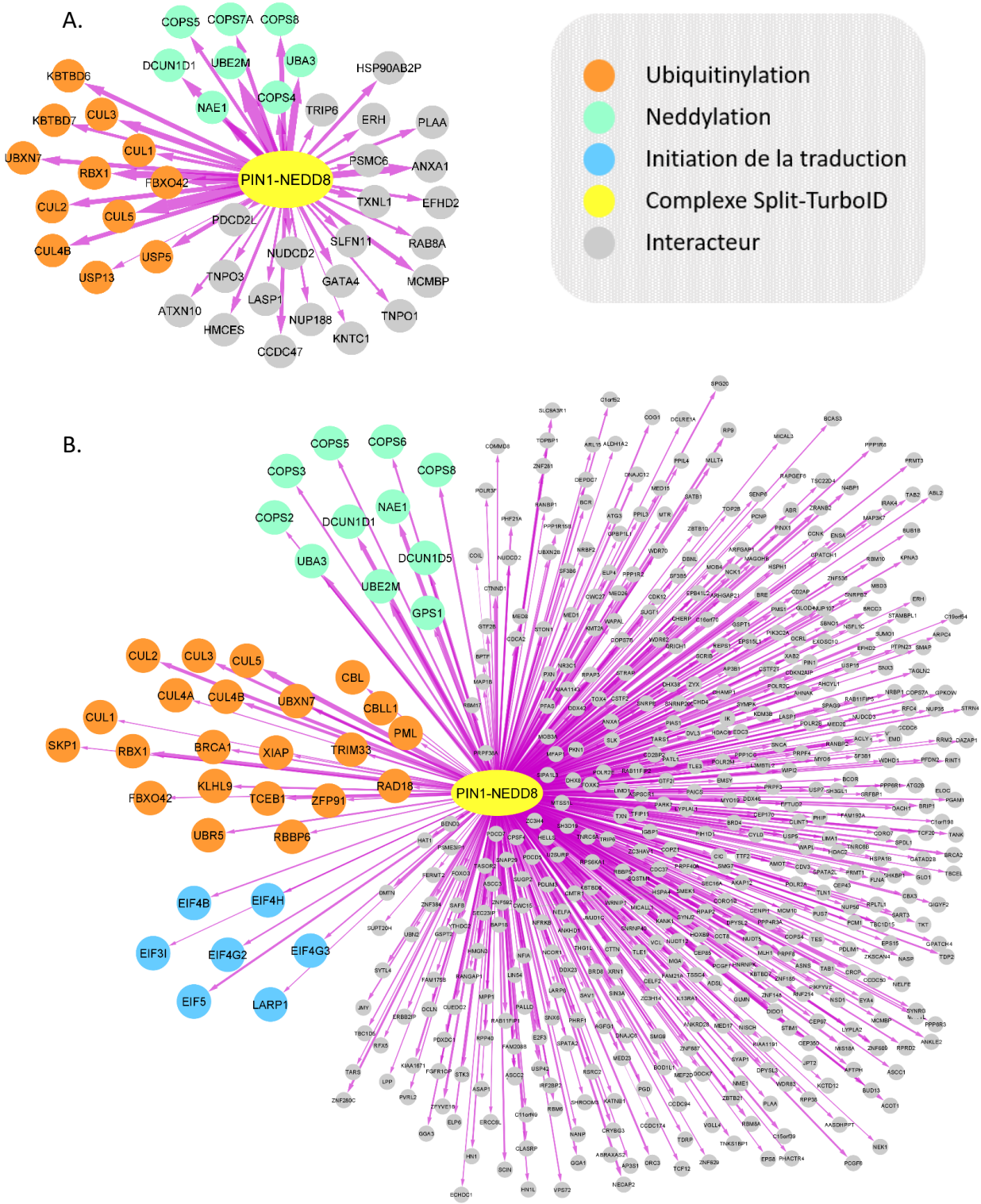


Figure 15 : Environnement moléculaire au site d'interaction de PIN1 et NEDD8

Les interacteurs représentés ont été identifiés au voisinage du site d'interaction entre PIN1 et NEDD8 par Split-TurboID (Log2fc (LFQ) ≥1 ; q-value ≤ 0,05). Ces interacteurs proximaux ont été identifiés en utilisant un temps de biotinylation de 4h ***(A)*** *et un temps de biotinylation de 24h* ***(B)****.*

4.2.4 Confirmation de la localisation de PIN1 au niveau des CRLs

Les analyses réalisées sur PIN1 dans le cadre des CCMs ont mené à la théorisation d'une régulation de l'activité des CRL par cette protéine dans un cadre plus général. L'étude de l'environnement moléculaire au site d'interaction entre PIN1 et NEDD8 indique une interaction entre ces protéines et suggère un recrutement de la peptidyl prolyl cis/trans isomérase au niveau des complexes dont l'activité est régulée par la neddylation. Afin de confirmer la proximité entre PIN1 et les CRL, l'interaction entre cette protéine et RBX1 a été étudiée par PLA. La protéine RBX1 est présente dans la plupart des complexes CRL et est supposément le composant subissant le changement conformationnel permettant le transfert de l'ubiquitine de l'E2 vers le substrat (Merlet *et al.*, 2009). Cette protéine a donc été sélectionnée pour la confirmation de la localisation de PIN1 à proximité des CRLs. L'interaction entre PIN1 et NEDD8 a également été évaluée par PLA (**Figure 16**).

Afin de disposer d'anticorps d'espèces différentes, pré-requis à la réalisation de la PLA, des lignées stables exprimant une forme tagguée Flag de PIN1 ont été mises en place. La PLA a été réalisée sur la forme taggué de PIN1 et les protéines NEDD8 et RBX1 endogènes. L'interaction entre PIN1 et NEDD8 a ainsi pu être observée par Split-TurboID et validée par PLA. De la même manière, l'interaction entre PIN1 et RBX1 qui est observée par TurboID et Split-TurboID PIN1-NEDD8 a également été confirmée par PLA.

Des approches orthogonales d'interatomique ont ainsi pu confirmer l'interaction entre les protéines PIN1 et NEDD8. La capacité de PIN1 a catalyser l'isomérisation cis/trans des prolines de ses cibles en fait un candidat idéal dans la régulation de l'activité des CRLs dont la structure de base ne permet pas le transfert de l'ubiquitine de l'E2 vers le substrat (Wu *et al.*, 2003). L'activité des CRL étant conditionnée par la neddylation, il est envisageable que l'interaction entre PIN1 et NEDD8 corresponde à un recrutement de l'isomérase vers les protéines neddylées comme les cullins. Le Split-TurboID a révélé la présence de la majorité des cullins et d'autres composants majeurs des CRLs comme RBX1 au site d'interaction entre PIN1 et NEDD8. L'interaction avec RBX1 a également été confirmée par PLA et représente un point commun entre les CRL potentiellement isomérisés dans le cadre de la régulation de l'activité de ces complexes d'ubiquitinylation (**Figure 16**) (Merlet *et al.*, 2009). Le mécanisme envisagé ici est un mécanisme de régulation de l'ubiquitinylation de manière générale. Il dépasse le contexte CCM et permettrait une meilleure compréhension de la régulation de

l'adressage au protéasome et du rôle de la neddylation dans ce processus. Ainsi, ces résultats dépassant le cadre de ce projet de thèse, ont donné lieu à la création d'un projet à part entière. Les travaux concernant le recrutement de PIN1 au niveau des CRLs par la neddylation ont été repris et sont poursuivis dans le cadre d'un projet post-doctoral.

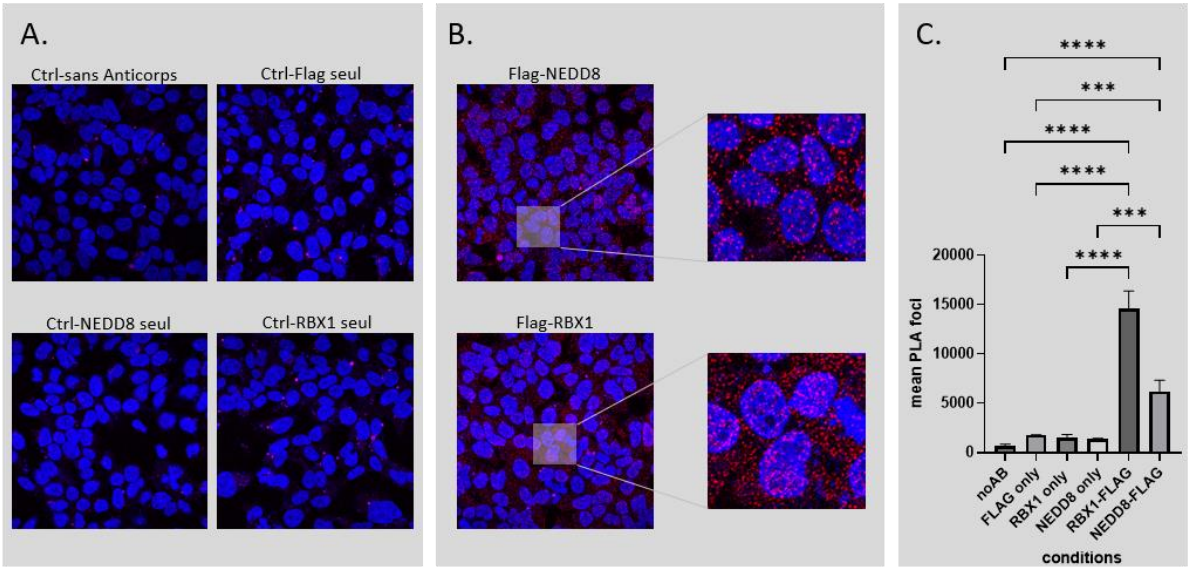


Figure 16 : Visualisation d'interaction PIN1-NEDD8 et PIN1-RBX1 par PLA.

Les cellules utilisées sont des HEK 293 Flp-In T-REx exprimant une forme tagguée par un Flag de la protéine PIN1. (A) Signal fluorescent observé dans les conditions contrôles sans anticorps primaires et avec chaque anticorps primaire seul. (B) Signal fluorescent observé site à la PLA en présence de l'anti-Flag ciblant la protéine PIN1 tagguée et les anticorps anti-NEDD8 et anti-RBX1. (C) Représentation graphique du nombre de foci observé et visualisation de la significativité statistique des différences entre les contrôles et les interactions testées (PIN1-NEDD8 et PIN1-RBX1).

4.3 REGULATION DES DOMMAGES DE L'ADN PAR TLT ET EHMT2

4.3.1 Relation entre CCM, DDR et EHMT2

Les résultats obtenus lors de la phase d'analyses exploratoires ont permis la formulation de plusieurs hypothèses mécanistiques. En plus de l'implication potentiel de PIN1 dans la régulation des CRLs, étudiée dans la partie précédente, les mécanismes épigénétiques et les mécanismes de réparation de l'ADN semblent occuper une place importante dans l'oncogenèse des CCM-VP. L'analyse des variations du protéome total a en effet permis l'identification de 27 protéines impliquées dans la réponse aux dommages de l'ADN surexprimées lorsque les antigènes T sont présents. Parmi ces protéines de la DDR (DNA Damage Response), nous retrouvons certains facteurs importants comme l'intégralité des composants du complexe RFC. Les protéines RIF1 et RAD51C impliquées dans la recombinaison homologue sont également identifiées en plus grande abondance. La protéine CHEK1 a été identifiée exclusivement en présence des antigènes T. Les cellules CCM-VP étant dépendante de l'expression des protéines virales, il n'est pas surprenant de retrouver une plus grande abondance de protéines impliquées dans la réplication (Houben *et al.*, 2012). Il est cependant plus inattendu de retrouver des facteurs impliqués dans la réponse aux dommages de l'ADN parmi ces protéines surexprimées. Cette présence accrue de facteurs de la DDR en présence du MCPyV suggère une meilleure capacité à prendre en charge ces dommages. Ainsi, cette observation mène potentiellement vers une explication pour la stabilité génétique et le peu de mutations observées au sein des CCM-VP.

Comme détaillé précédemment, les CCM-VP présentent très peu de mutations somatiques, sont génétiquement stables et sont dépendantes de l'expression des antigènes T pour leur prolifération. Dans le cas des lignées cancéreuses établies CCM-VP, l'arrêt de la croissance induit par l'inhibition de l'expression des antigènes T peut être complétement aboli par la réexpression de tLT seul (Angermeyer *et al.*, 2013). Il est donc probable que la majorité des effets liés à l'inhibition de l'expression des antigènes T par l'utilisation de shRNA soit liée à l'inhibition de tLT et non de sT dans ces modèles expérimentaux. Nous nous sommes donc concentrés ici sur les formes tronquées du LT.

Les données d'interactomique obtenues ont révélées une forte présence de facteurs liés au remodelage épigénétique comme les composants du complexe NuRD ou la protéine BEND3 qui représente un lien entre le complexe NuRD et le PRC2 déjà connu pour son implication dans les CCM-VP (Gartin *et al.*, 2022; Durand *et al.*, 2023). Plusieurs mécanismes impliquant la modification des histones ont été identifiés avec EP400, LSD1 ou EZH2. Il semble y avoir un lien étroit entre les modulations épigénétiques impliquant la modification des histones et les antigènes T. Comme décrit précédemment, 66 protéines modificatrices d'histones ont été identifiées en interaction avec tLT (LT330). Parmi elles, 22 sont impliquées dans la méthylation des histones (**Figure 10**). Notamment, le complexe EHMT2-EHMT1-WIZ qui catalyse la mono- et di-méthylation de la lysine 9 de l'histone H3 a été retrouvé en interaction avec tLT. La lysine déméthylase KDM3B, capable de retirer cette marque épigénétique, interagit également avec tLT. Ces observations suggèrent que la présence de tLT et son interaction avec EHMT2 pourraient moduler l'activité méthyltransférase de EHMT2 et ses cibles. En effet, en plus de méthyler H3K9, EHMT2 méthyle également H3K27, une marque associée à la répression transcriptionnelle (Tachibana *et al.*, 2002). De plus, EHMT2 a été localisée au niveau des fourches de réplication lors d'études d'interactome avec PCNA et par iPond, où elle méthyle H3K56, régulant ainsi la progression du cycle cellulaire (Sirbu *et al.*, 2013; Srivastava *et al.*, 2018).

EHMT2 méthyle également des cibles non-histones comme p53 (K373), inhibant ses fonctions dans la réponse aux dommages à l'ADN et l'apoptose (Huang *et al.*, 2010). EHMT2 méthyle aussi MDC1 (K45), augmentant son affinité pour ATM et le recrutement de facteurs de réparation de l'ADN (Watanabe *et al.*, 2018). Ainsi, l'interaction de tLT avec EHMT2 pourrait moduler ses activités de méthylation et impacter des processus clefs comme la régulation transcriptionnelle, la progression du cycle cellulaire et la réponse aux dommages à l'ADN.

Cette relation avec les dommages de l'ADN rejoint les observations faites sur les autres interacteurs de tLT. Comme indiqué dans la section BioID, tLT compte en effet parmi ses interacteurs proximaux de nombreux facteurs impliqués dans la DDR. L'interaction détectée en BioID entre EHMT2 et tLT a été testée par PLA au sein des lignées WaGa et MKL-1 exprimant de formes tronquées du LT de tailles différentes. Il a été possible de confirmer cette interaction dans la lignée MKL-1 exprimant la forme LT330 utilisée en BioID ainsi que dans la lignée WaGa exprimant une forme plus courte (**Figure 17a, b**).

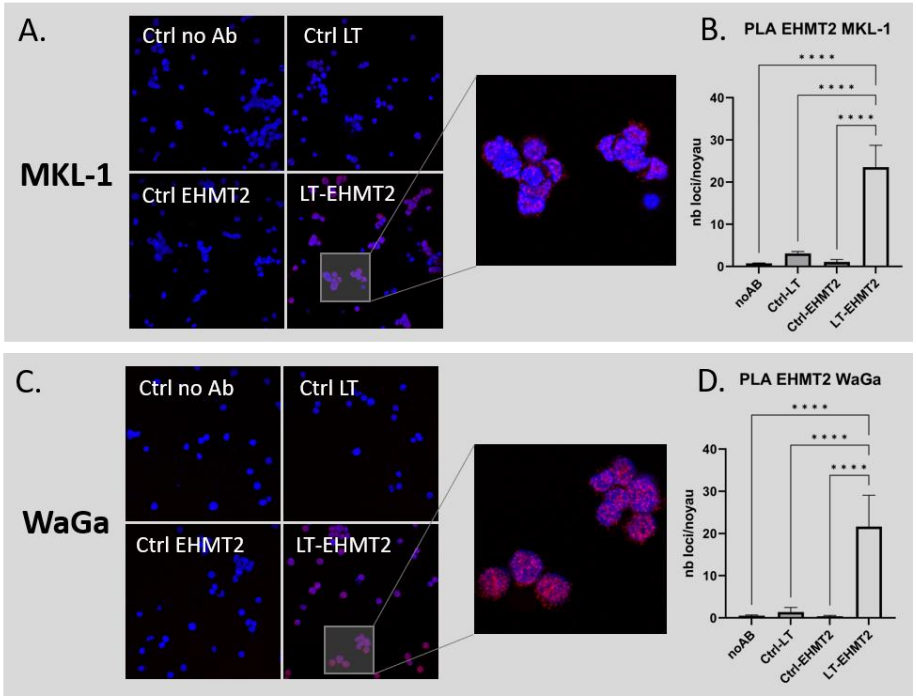


Figure 17 : Validation de l'interaction tLT-EHMT2

***(A)** Visualisation par microscopie confocale des signaux fluorescent de PLA dans les contrôles et la condition testée LT-PIN1 au sein de cellules MKL-1. **(B)** Représentation graphique du nombre de foci de fluorescence par noyau dans chaque condition au sein des cellules MKL-1. La significativité des différences est calculée par ANOVA. **(C)** Visualisation par microscopie confocale des signaux fluorescent de PLA dans les contrôles et la condition testée LT-PIN1 au sein de cellules WaGa. **(D)** Représentation graphique du nombre de foci de fluorescence par noyau dans chaque condition au sein des cellules WaGa. La significativité des différences est calculée par ANOVA.*

L'identification des interactions formées par les formes tronquées du LT du MCPyV suggère ainsi une certaine importance des processus épigénétiques dans le développement des CCM-VP. De la même manière, la forte présence parmi ces interacteurs de facteurs impliqués dans la DDR mets en avant la possibilité d'une régulation des mécanismes de réparation de l'ADN et de contrôle du cycle cellulaire qui y sont associés. Dans ce contexte, la protéine EHMT2 est particulièrement intéressante dans la mesure où elle se trouve à l'interface entre ces mécanismes en plus d'être retrouvée en interaction avec tLT.

4.3.2 Remodelage médié par EHMT2 et LT

Au sein de l'interactome de tLT, EHMT2 est la méthyltransférase retrouvée avec la plus grande différence d'abondance par rapport aux contrôles en se basant sur les valeurs LFQ. Il apparait donc intéressant d'étudier les conséquences de l'inhibition de cette protéine au sein de cellules CCM-VP. L'IC50 du BRD4770, un inhibiteur de EHMT2 couramment utilisé a été déterminé à environ 5µM pour les cellules CCM-VP. Une mesure de l'effet dose réponse sur

la viabilité cellulaire à 48h a pour cela été réalisée par MTS assay sur plusieurs lignées cellulaires. Deux lignées CCM-VP (MKL-1 et WaGa) et deux lignées non-CCM (A375 mélanome et HEK 293) ont été utilisées. De manière surprenante, les lignées CCM-VP étaient plus résistantes que les non-CCM à l'inhibition de EHMT2. Il est donc possible que la présence des antigènes T ait bel et bien une influence sur le rôle de EHMT2. Afin d'évaluer l'effet de la présence des antigènes T dans le contexte CCM-VP sur EHMT2 et son inhibition, la viabilité cellulaire a été mesurée après 48h d'inhibition de EHMT2 au sein de cellules présentant une inhibition de l'expression des antigènes T (**Figure 18a**). Les cellules ont été traitées cinq jours à la doxycycline avant de démarrer l'inhibition de EHMT2 afin de suffisamment diminuer l'expression des antigènes T. La baisse d'expression du tLT a été vérifiée par Western-Blot (**Figure 18c**). Conformément à ce qui est connu dans la littérature, une baisse importante de prolifération des cellules a été observée suite à l'expression du shRNA ciblant les antigènes T. L'inhibition de EHMT2 au sein des cellules WaGa WT a également eu pour effet la baisse de la prolifération des cellules. De manière radicale, l'arrêt de la prolifération causé par l'inhibition des TA est annulé par l'inhibition de EHMT2.

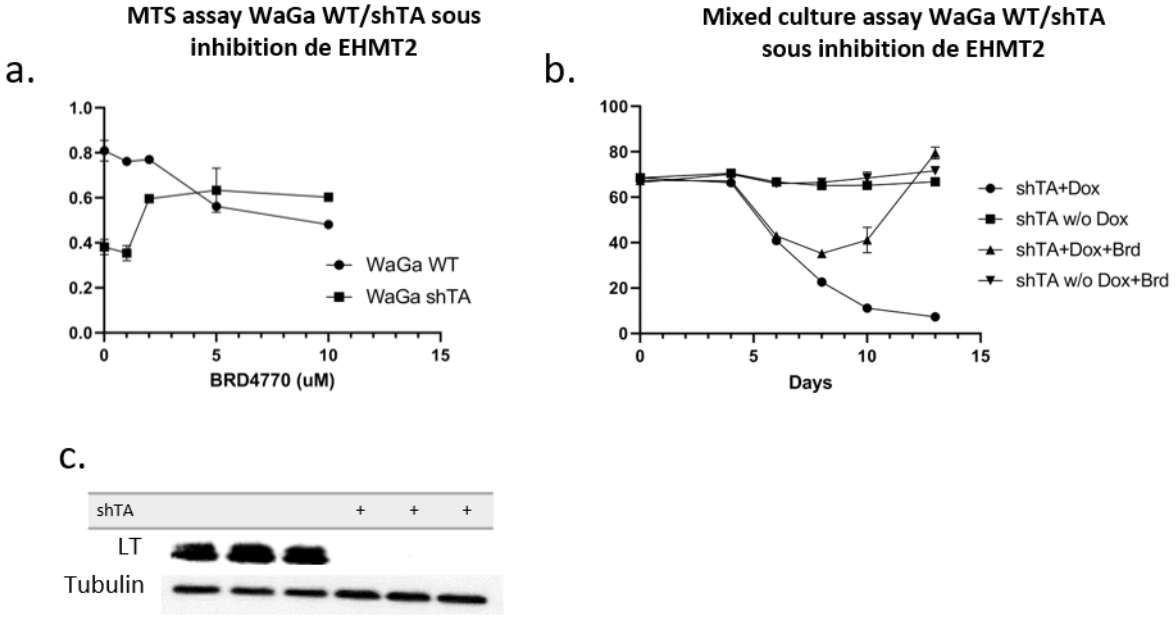


Figure 18 : Effet de l'inhibition de EHMT2 sur la prolifération de cellules CCM-VP en présence et absence des antigènes T.

***(A)** La prolifération des cellules WaGa WT et WaGa shTA a été mesurée par MTS assay après 48h de traitement au BRD4770 à différentes concentrations. **(B)** La prolifération des cellules a été mesurée par mixed culture assay à partir du traitement doxycycline permettant l'induction du shTA. Le BRD4770 a été ajouté au sixième jour de traitement à une concentration de 5µM. Les cellules WaGa shTA sont co-cultivées avec des cellules WaGa exprimant la GFP de manière constitutive en présence ou absence de doxycycline permettant l'induction du shTA. **(C)** Vérification par western blot du knock-down du tLT au sein des WaGa shTA.*

Ces résultats ont été confirmés par mixed culture assay réalisé par Pauline Gaboriaud de l'université de Tours (**Figure 18b**). Les cellules WaGa présentant une inhibition inductible des shTA ont été cultivées en co-culture avec des WaGa exprimant la GFP de manière constitutive. Ces cellules ont été traitées à la doxycycline afin d'induire l'expression du shRNA et le ratio WaGa-shTA/WaGa-GFP a été mesuré par cytométrie en flux. Une fois de plus, une inversion de l'arrêt de la croissance due au shTA a été observée suivant l'inhibition de EHMT2. La protéine EHMT2 et son interaction avec le tLT semble ainsi tenir un rôle central dans la prolifération des CCM-VP.

Dans le but d'identifier le rôle de EHMT2 dans les CCM-VP et son impact sur l'expression protéique en présence et en absence de des antigènes T, le protéome total des cellules WaGa WT et WaGa shTA a été analysé en condition basale et sous inhibition de EHMT2. Les cellules ont été traitées par 5µM de BRD4770 après 5 jours de traitement doxycycline permettant l'inhibition de l'expression des antigènes T. L'analyse du protéome par spectrométrie de masse et la comparaison des abondances relatives de chaque protéine identifiée n'a révélé que très peu de différence entre les conditions présentant ou non l'inhibition de EHMT2 (**Figure 19**).

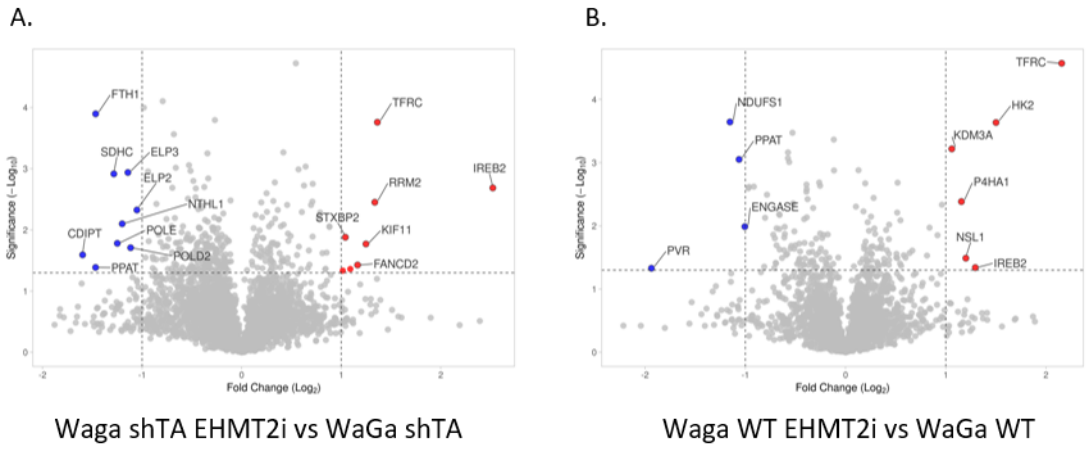


Figure 19 : Analyse protéomique de cellules CCM-VP sous inhibition chimique de EHMT2

***(A)** Représentation sous forme de volcano-plot de la différence d'abondance en LFQ des protéines identifiées dans les cellules WaGa avec inhibition des TA contre les WaGa avec inhibition des TA et de EHMT2 en fonction du -Log10 de la p-value. Les protéines dans le cadran supérieur gauche sont retrouvées de manière plus abondante sans inhibition de EHMT2 et les protéines dans le cadran supérieur droit sont retrouvées de manière plus abondante lorsqu'EHMT2 est inhibé. **(B)** Représentation sous forme de volcano-plot de la différence d'abondance en LFQ des protéines identifiées dans les cellules WaGa WT contre les WaGa WT avec inhibition de EHMT2 en fonction du -Log10 de la p-value. Les protéines dans le cadran supérieur gauche sont retrouvées de manière plus abondante sans inhibition de EHMT2 et les protéines dans le cadran supérieur droit sont retrouvées de manière plus abondante lorsque EHMT2 est inhibé.*

L'inhibition de EHMT2 semble donc avoir peu d'impact sur le protéome des cellules CCM-VP. De plus, la présence de tLT au sein des cellules n'a pas que peu d'impact sur l'influence de EHMT2 sur la composition du protéome. Ces résultats suggèrent que ni EHMT2, ni l'interaction entre EHMT2 et tLT ne jouent un rôle majeur dans la régulation de l'expression génétique au sein des cellules CCM, ce qui était pourtant attendu étant donné l'activité bien décrite de régulation de la chromatine via son activité histone methyltransferase. Cependant, l'activité de EHMT2 semble être centrale pour la prolifération des CCM-VP et conditionnée par la présence des antigènes T, comme montré par les résultats des études de prolifération par MTS et mixed culture assay. Dans le but d'obtenir plus d'informations concernant le rôle de tLT sur les fonctions de EHMT2, des analyses exploratoires ciblées ont été réalisées.

4.3.3 Influence de tLT sur les interactions formées par EHMT2

La fonction canonique de EHMT2 ne semblant pas occuper une place centrale dans le processus oncogénique CCM-VP, les interactions proximales formées par cette protéine ont été caractérisées. Afin de visualiser l'impact de la potentielle collaboration entre cette protéine et le tLT, l'interactome de EHMT2 a été étudié en présence et absence de la protéine virale. Une forme de EHMT2 fusionnée à BirA* à son extrémité N-terminale a pour cela été exprimée au sein de cellules 293 Flp-In T-REx présentant l'expression de manière stable inductible du Flag-tLT et dans une condition contrôle présentant l'expression du Flag seul. Il a été possible d'observer 136 interacteurs de haute confiance pour EHMT2 en condition basale (**ANNEXE III**). Les partenaires connus de EHMT2 comme EHMT1 et WIZ sont retrouvés parmi les interacteurs identifiés. Comme attendu, plusieurs autres protéines impliquées dans les modifications épigénétiques sont présentes dans l'environnement proximal de EHMT2. Il est notamment possible d'observer des facteurs impliqués dans la méthylation des histones comme SETD7, PRMT1 et PRMT7. Certaines protéines impliquées dans l'acétylation des histones sont identifiées en interaction avec EHMT2. C'est le cas de plusieurs composants du complexe de désacétylation NuRD comme GATAD2B, RBBP7, MTA1 et MTA2.

Les cibles non-histones connues de EHMT2 sont également retrouvées. Il est ainsi possible de retrouver MDC1, LIG1, MTA1 ou DNMT1 parmi les interacteurs identifiés. L'analyse de l'interactome proximal de EHMT2 a également révélé la présence de protéines

non-histone et non décrites comme cibles ou co-facteurs de la méthyltransférase. C'est notamment le cas de protéines impliquées dans la DDR comme XRCC5, XRCC6 ou MRE11A.

De manière intéressante, l'expression de tLT au sein des cellules n'a pas entrainé de perte d'interaction pour EHMT2. En effet, 89% des interacteurs identifiés en condition basale le sont également en présence de la protéine virale. La présence de tLT a cependant élargi le paysage moléculaire de EHMT2 de manière considérable. Au total, 324 protéines sont identifiées comme interacteur de haute confiance de EHMT2 en présence de tLT, dont tLT lui-même, confirmant une fois de plus la proximité entre les deux protéines (**ANNEXE III**). L'augmentation conséquente du nombre d'interacteur suggère une redistribution de EHMT2 à différents *loci* sur la chromatine médiée par son interaction avec tLT qui agirait ici comme une « scaffolding protein ».

Une partie des interacteurs de haute confiance identifiés en présence de tLT ne passent pas les seuils de significativité en condition basale, mais sont tout de même présents juste sous les limites de log2fc ou q-value. Les deux conditions ont donc été comparées entre elles fin d'identifier les protéines significativement plus présentes dans chacune (Log2Fc > 1 ; p-value ≤ 0,05). Il a ainsi été possible d'identifier 50 protéines enrichies de manière significatives ou retrouvées uniquement dans l'environnement de EHMT2 en présence de tLT (**Figure 20a**). De manière intéressante 25% de ces protéines sont impliquées dans la régulation de la DDR (**Figure 20b**).

La présence dans l'environnement proximal de EHMT2 de protéines comme MDC1, LIG1, XRCC5-6, MRE11A, RPA3 ou RIF1 situe cette méthyltransférase au niveau de la voie DDR et suggère un lien avec le stress réplicatif et les cassures de l'ADN. L'enrichissement de protéines impliquées dans la DDR en présence de tLT suggère un recrutement plus important de cette protéine au niveau des sites de DDR. Comme mentionné précédemment, EHMT2 a déjà été observé à proximité de ces sites de DDR et notamment des fourches de réplications bloquées par son identification dans l'interactome de PCNA (Srivastava *et al.*, 2018). Plusieurs cibles non-histones de EHMT2 impliquées dans la DDR sont détectées dans son interactome. C'est notamment le cas de MDC1 et LIG1. L'interaction entre EHMT2 et MDC1 est détectée en présence comme en absence de tLT et ne semble pas être affectée par sa présence. La régulation de MDC1 dépendante de EHMT2 est impliquée dans l'activation de ATM principalement dans le cadre de cassures double brins de l'ADN. Il apparait alors probable qu'il

existe un mécanisme similaire au sein de la voie ATR prenant en charge les cassures simples brins et les dommages liés au stress réplicatif.

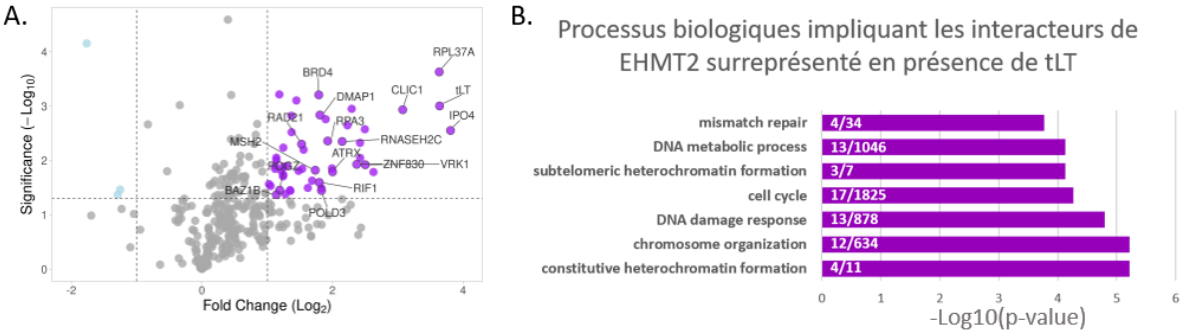


Figure 20 : Effet de l'expression de tLT sur les interactions formées par EHMT2

***(A)** Volcano plot représentant les interactions formées par EHMT2 en présence de tLT vs en absence de cette protéine virale. Les protéines dans le cadran supérieur droit sont significativement enrichies en présence de tLT (Log2 fold change ≥ 1 ; p-value ≤ 0,05). **(B)** Analyse d'enrichissement des termes de « Gene Onthology » montrant les processus biologiques impliquant les 50 protéines significativement enrichies dans l'interactome de EHMT2 en présence de tLT.*

Dans le but d'étudier la relation entre tLT et EHMT2, les effets de leurs inhibitions combinées ont été testées au sein de cellules CCM-VP.

4.3.4 Une coopération entre EHMT2 et tLT dans la régulation de la DDR

Un lien entre EHMT2 et la DDR a déjà pu être observé, il est notamment connu que son inhibition empêchait une réponse efficace face aux dommages de l'ADN. Il a été supposé que la di-méthylation de H3K9 puisse jouer un rôle dans la réparation de l'ADN (Gursoy-Yuzugullu *et al.*, 2017). Une relation entre EHMT2 et cassure double brins a également été observée par son implication dans le recrutement des RPA aux sites de cassures de manière dépendante de la méthylation (Yang *et al.*, 2017).

Afin d'observer la capacité des cellules CCM-VP à déclencher une réponse efficace aux dommages de l'ADN. Les niveaux de phosphorylation et la quantité de différents marqueurs de la DDR ont été mesurés suite à l'inhibition des antigènes T, de EHMT2 et en cas d'induction de stress réplicatif par traitement à l'hydroxyurée (HU) avec et sans temps de récupération après le traitement HU. L'HU est utilisée pour induire un stress réplicatif provoquant des cassures simple-brin de l'ADN (SSB) (Petermann *et al.*, 2010; Ercilla *et al.*, 2020). L'analyse avant récupération vise à évaluer la possibilité de déclencher l'activation de la DDR, tandis que l'analyse après récupération vise à évaluer la capacité des cellules à se rétablir des dommages consécutifs du stress réplicatif grâce à un système efficace de réparation de l'ADN. La réponse

des cellules aux cassures doubles brins a également été évaluée via le suivi des *foci* γH2Ax au sein de cellules exprimant ou non les TAs, sous inhibition de EHMT2 et suite à un traitement à l'etoposide, un inhibiteur de topoisomérase II. Les foyers de γH2AX sont couramment utilisés comme marqueur de la réponse aux dommages de l'ADN (DDR) et plus généralement des cassures de l'ADN. Ils ont été quantifiés par immunofluorescence afin d'évaluer l'impact de l'expression des antigènes T ainsi que de l'inhibition de l'activité méthyltransférase de EHMT2 sur l'intégrité du génome dans les cellules CCM-VP (**Figure 21**).

Nos résultats indiquent une augmentation significative du nombre de foyers γH2AX dans les cellules WaGa sauvages (WT) suite au traitement à l'étoposide. Cette augmentation s'est avérée exacerbée dans les cellules WaGa invalidées pour les antigènes T par shRNA inductible (shTA). L'inhibition pharmacologique de l'activité catalytique de EHMT2 par le composé BRD4770 à 5 µM a quant à elle révélé un effet dépendant de la présence des antigènes T. En conditions d'expression normale des antigènes T, l'inhibition de EHMT2 conduit à une élévation du nombre de foyers γH2AX ; tandis qu'en l'absence des protéines virales, une diminution est observée. Pris ensemble, ces résultats démontrent que la présence des oncoprotéines virales est fortement impliquée dans l'activation de la signalisation DDR médiée par l'inhibition de EHMT2, mais pas dans celle induite par l'étoposide, dans le contexte cellulaire CCM-VP. De façon notable, le traitement combiné étoposide et inhibition de EHMT2 a conduit à des résultats similaires à ceux obtenus avec l'inhibition de EHMT2 seule. Ainsi, le traitement à l'étoposide induit davantage de foyers γH2AX dans les cellules WaGa invalidées pour les antigènes T (shTA) comparativement aux cellules contrôles exprimant les oncoprotéines virales. Cependant, lorsque l'étoposide est associé à l'inhibition de l'activité méthyltransférase de EHMT2 par le BRD4770, on observe un nombre plus important de foyers γH2AX dans les cellules exprimant les antigènes T par rapport aux cellules shTA. Pris ensemble, ces observations suggèrent une coopération fonctionnelle entre les antigènes T viraux et EHMT2 dans la modulation de la signalisation DDR en réponse aux dommages de l'ADN, les antigènes T permettant de compenser partiellement l'effet délétère de l'étoposide.

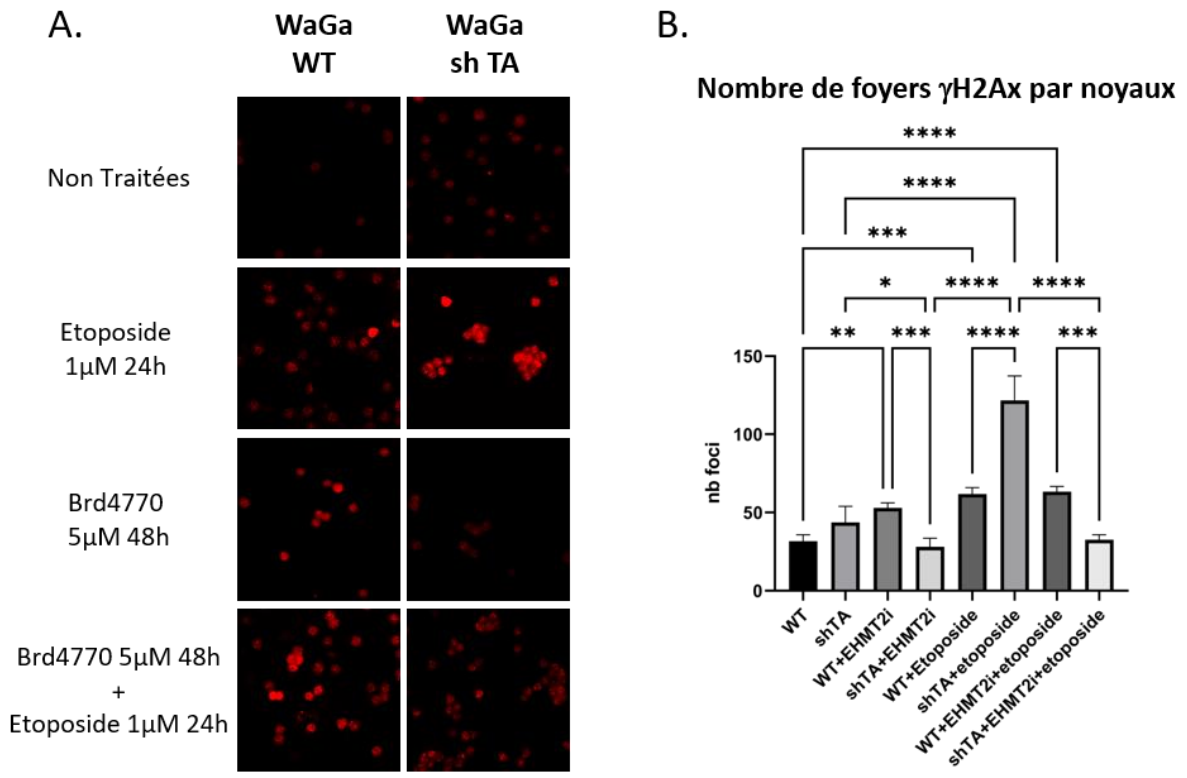


Figure 21 : Quantification des foyers de γH2Ax en relation avec l'inhibition de EHMT2 et des antigènes T.

***(A)** Visualisation par microscopie confocale des foyers γH2Ax. Les cellules ont été traitées à la doxycycline afin d'induire l'expression du shRNA ciblant les TAs. Les foyers γH2Ax ont été observés après traitement étoposide 1µM 24h, BRD4770 5µM 48h (noté EHMT2i dans B) ou traitement combiné BRD4770 et étoposide. **(B)** Représentation graphique de de la quantification du nombre de foyers par noyau. L'analyse statistique a été réalisée par one-way anova.*

L'analyse par Western blot de différents marqueurs de la DDR a révélé que l'inhibition pharmacologique de EHMT2 dans les WaGa WT entraîne une activation de la signalisation DDR (**Figure 22a**). Cette activation est illustrée par une augmentation des niveaux de protéines phosphorylées p-ATR, p-CHK1, p-CHK2 ainsi que de l'inhibiteur de CDK p21. Dans les cellules WaGa-shTA, nous observons également une élévation de ces marqueurs (à l'exception de p-CHK1) suite à l'inhibition de EHMT2. Toutefois, cette augmentation part de niveaux basaux plus faibles que ceux des cellules contrôles exprimant les oncoprotéines virales et n'atteint pas les niveaux observés dans ces dernières. De plus, l'inhibition de l'expression des antigènes T entraîne une diminution des taux protéiques de EHMT2. De manière intéressante, que les antigènes T soient exprimés ou non, l'inhibition de l'activité méthyltransférase de EHMT2 conduit à une accumulation de cette protéine, possiblement en tant que mécanisme compensatoire face à la perte de sa fonction enzymatique.

Comme attendu, le traitement à l'hydroxyurée (HU) induit un stress réplicatif et entraîne une augmentation des niveaux de tous les marqueurs de DDR analysés, reflétant son activation. Cependant, les niveaux atteints dans les cellules WaGa shTA restent inférieurs à ceux des cellules contrôles WaGa WT. De façon intéressante, nous observons également une accumulation de la protéine EHMT2 suite au stress réplicatif déclenché par l'HU. Le niveau le plus élevé de EHMT2 est détecté après 18h de récupération post-traitement HU. De plus, l'ajout d'un temps de récupération suivant l'exposition à l'HU révèle une forte réduction du niveau de p-CHK1 phosphorylée dans les cellules WaGa WT, alors que ce niveau reste élevé lorsque l'activité méthyltransférase de EHMT2 est inhibée par le BRD4770. Ainsi, l'inhibition pharmacologique de EHMT2 semble responsable du maintien prolongé de l'activation de la signalisation DDR en réponse au stress réplicatif, comme en atteste le niveau soutenu de CHK1 phosphorylée. Ce résultat suggère que l'activité enzymatique de EHMT2 est requise pour une réponse et une récupération appropriées suite aux dommages de l'ADN provoqués par un stress de réplication induit par l'HU. La phosphorylation prolongée de CHK1 lors de l'inhibition de EHMT2 pourrait refléter des défauts dans la réparation de l'ADN et une intégrité génomique compromise.

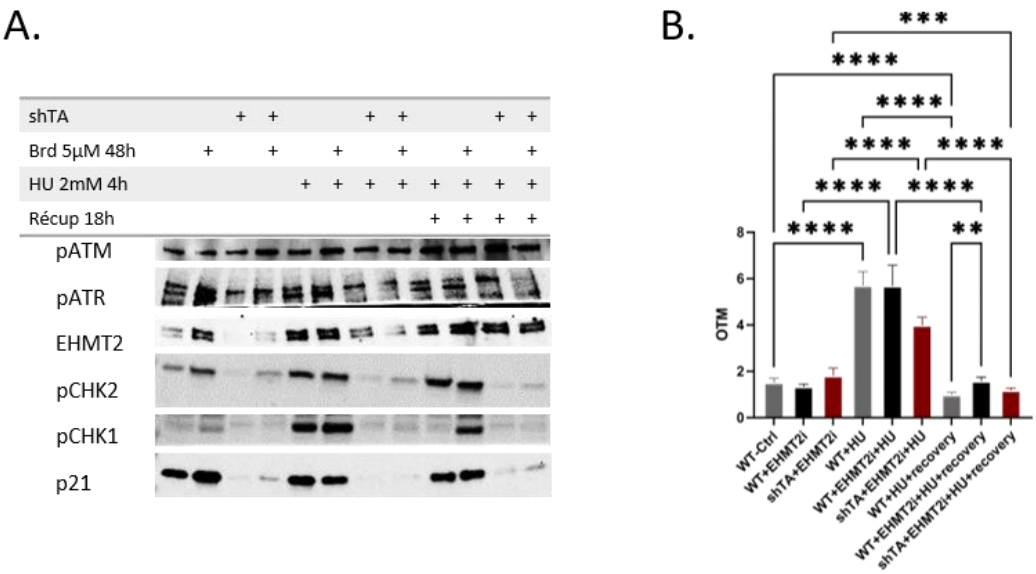


Figure 22 : Analyse des marqueurs de la DDR et des dommages réels de l'ADN.

(A) *Analyse Western Blot de plusieurs marqueurs de la DDR au sein de cellules WaGa WT et WaGa-shTA. Les cellules ont également été traitées ou non par 5µM de BRD4770 et 2mM d'hydroxyurée pendant 4h et ont été récoltée soit directement après le traitement HU soit après 18h de récupération.* ***(B)*** *Représentation graphique des mesures d'Olive tail Moment obtenues par Comet assay sur des cellules CCM-VP WaGa Wt et WaGa-shTA traitées ou non par 5µM de BRD4770 et 2mM d'hydroxyurée pendant 4h et ont été analysées soit directement après le traitement HU soit après 18h de récupération.*

L'analyse des marqueurs de DDR par Western blot permet d'évaluer l'activation de la signalisation de cette voie. Cependant, il est possible que cette signalisation soit découplée des dommages réels de l'ADN présents dans les cellules. Afin de tester ce découplage potentiel, des comet assays ont été réalisés pour observer directement l'état de dégradation de l'ADN par microscopie. Les comet assay ont été réalisés en collaboration avec le Dr Laetitia Trapp-Fragnet. Il a ainsi été possible d'observer une corrélation entre le niveau de CHK1 phosphorylée et le Olive Tail Moment (OTM), indicateur de cassures double et simple-brin de l'ADN (**Figure 22b**). De façon similaire aux niveaux de p-CHK1, le niveau de cassures de l'ADN est comparable entre les cellules WaGa WT traitées ou non au BRD4770 après exposition à l'HU. Toutefois, les mesures d'OTM révèlent que lorsque l'activité enzymatique de EHMT2 est inhibée, le temps de récupération post-HU se traduit par une réparation de l'ADN moins efficace. En effet, l'OTM mesuré est significativement plus élevé parmi les cellules traitées au BRD4770. Ainsi, bien que la voie de signalisation DDR soit activée en réponse au traitement HU comme le montre l'augmentation de p-CHK1, la réparation effective des dommages à l'ADN semble compromise lors de l'inhibition de l'activité méthyltransférase de EHMT2.

Afin de déterminer si l'impact observé de EHMT2 sur la DDR était restreint aux lignées cellulaires de CCM, les effets de l'inhibition de EHMT2 ont également été évalués dans les cellules U2OS parentales ainsi que dans les cellules U2OS transfectées avec tLT (**Figure 23**). Les cellules U2OS sont couramment utilisées pour les études sur la DDR car elles conservent la capacité à stopper leur cycle cellulaire suite à l'activation des points de contrôle du cycle cellulaire impliqués dans la DDR. Dans ce contexte, nous avons observé une augmentation des niveaux de pCHK1 suite au traitement à l'HU, indiquant une transmission efficace du signal DDR de la détection de cassures de l'ADN à la phosphorylation de CHK1. Nous avons également observé une augmentation des niveaux de pCHK1 suite à l'inhibition de EHMT2. De façon intéressante, contrairement aux cellules WaGa, l'augmentation de la phosphorylation de CHK1 ne corrèle pas avec une augmentation de l'expression de p21 lorsque EHMT2 est inhibé. Dans un contexte de stress réplicatif, c'est-à-dire suite au traitement à l'HU, l'inhibition de EHMT2 conduit à une forte augmentation des niveaux de pCHK1 mais également à une diminution du niveau total de p21. Dans cette lignée cellulaire, les niveaux de EHMT2 ne semblent pas affectés par son inhibition, le traitement à l'HU ou l'expression de tLT.

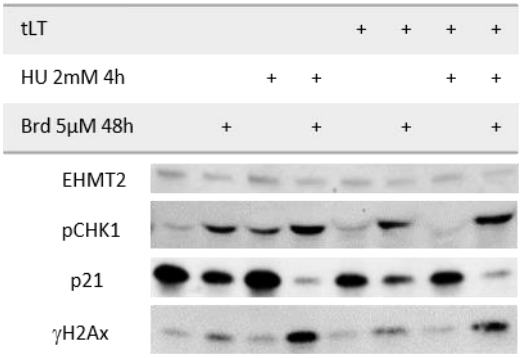


Figure 23 : Analyse de plusieurs marqueurs de la DDR

Western blotting réalisé au sein des cellules U2OS. Les cellules ont été traitées par 5µM de BRD4770 pendant 48h et 2mM de HU durant 4h. Les cellules utilisées expriment ou non la forme LT330 de tLT portant un Flag a son extrémité N-terminale.

Ainsi les résultats obtenus au sein des différentes lignées suggèrent une action importante de EHMT2 au niveau de la réponse aux dommages de l'ADN, qui pourrait se manifester différemment d'un type cellulaire à l'autre, probablement de manière dépendante à la fonctionnalité d'autres facteurs de DDR.

5 CONCLUSIONS/DISCUSSION GENERALES ET PERSPECTIVES

Cette thèse avait pour objectif d'étudier de manière systématique les interactions formées par les antigènes T du MCPyV et les perturbations associées à leur expression. Ces travaux avaient pour but l'identification de mécanismes oncogéniques dans les CCM-VP, et au-delà, la caractérisation de mécanismes potentiellement plus généraux dans les cancers par l'étude approfondie du système transformant minimal sT/LT. La caractérisation fine des mécanismes transformants est un préalable nécessaire à l'optimisation de traitements, ou au développement de thérapeutiques adjuvantes manquantes pour la prise en charge des CCM-VP, cancer de très mauvais pronostic. L'obtention des données d'interactomique visait à étudier l'action des protéines virales par le biais de leurs partenaires, permettant ainsi de délimiter les voies de transformation néoplasiques susceptibles de leur être associées. Ainsi la première partie de résultat consistait en un ensemble d'analyses exploratoires complémentaires qui, prises ensemble, ont permis la formulation d'hypothèses. De nombreux mécanismes potentiels se sont dégagés des données obtenues. Par recoupement de données de dimensions différentes et analyse de la littérature, deux d'entre eux se sont distingués et ont été étudiés plus en détail par la réalisation d'analyses et de tests ciblés.

Les études d'interactomique réalisées par BioID et GST-Pull Down ont ainsi permis d'identifier une population de protéines hôtes en relation avec les oncoprotéines virales. Une majorité des interactions connues pour ces protéines a été retrouvée par cette méthode, validant la pertinence de nos approches et des données obtenues. Nous avons également pu mettre en évidence un grand nombre de nouvelles interactions. Les interactions étudiées de manière plus détaillée ont pu être confirmées de manière orthogonale.

L'analyse des fonctions de ces interacteurs identifiés révèle que les antigènes T ciblent de façon préférentielle certains processus cellulaires clés. De nombreux facteurs impliqués dans la régulation de la transcription et les modifications épigénétiques ont notamment pu être mis en évidence. Ces protéines virales n'ayant pas d'activité enzymatique intrinsèque connues, elles agissent en altérant les fonctions de leurs partenaires cellulaires. Ainsi, la perturbation de mécanismes épigénétiques apparait comme un moyen efficace pour le virus de modifier profondément le programme d'expression génique. En effet, en l'absence de mutations somatiques, l'action d'uniquement deux protéines ectopiques, les antigènes T, est

capables de modifier le destin cellulaire en reprogrammant son schéma d'expression génique, donc protéique. De façon intéressante, les profils de méthylations de l'ADN observés dans les CCM-VP s'apparentent à ceux des cellules progénitrices, sans pour autant présenter l'expression de marqueurs de pluripotence (Chteinberg *et al.*, 2020). La forte présence des régulateurs épigénétiques, en particulier de protéines impliquées dans la méthylation/déméthylation de l'ADN, dans l'interactome des antigènes T représente un point de départ intéressant pour l'étude de ces reprogrammations. Il est envisageable que ces interactions aient un effet sur la fonction de ces facteurs épigénétiques ou sur leurs localisations sur la chromatine. Les antigènes T pourraient agir par recrutement préférentiel de sous-groupes d'interacteurs au niveau de certains *loci*, permettant par exemple la modification des marques épigénétiques des histones. Outre la méthylation de l'ADN, de nombreux facteurs impliqués dans la modification des histones ont également été identifiés en interaction avec les protéines virales. La méthylation des histones occupe notamment une part importante dans les fonctions de ces interacteurs. L'implication fonctionnelle de plusieurs régulateurs épigénétiques impliqués dans cette fonction a déjà été mis en lumière dans les CCM-VP. C'est notamment le cas de EZH2 et EP-400 (Cheng *et al.*, 2017; Gartin *et al.*, 2022; Durand *et al.*, 2023). Les interactions retrouvées ici avec des modificateurs épigénétiques impliqués dans la modification des histones permet la délimitation d'un nombre restreint de candidats potentiel pour l'étude de l'influence des antigènes T sur les processus épigénétiques.

Par ailleurs, la réponse aux dommages de l'ADN semble elle aussi affectée par la présence des oncoprotéines virales. De manière inattendue, celles-ci sont retrouvées en interaction avec de nombreux acteurs de la signalisation et de la réparation des cassures de l'ADN. Cette relation nouvelle entre antigènes T et réponse aux dommages de l'ADN peut intervenir dans un recrutement de ces facteurs pour améliorer la prise en charge de ces dommages. Il est également envisageable que ces interactions surviennent dans le cadre d'une inhibition des points de contrôles du cycle cellulaire faisant suite aux dommages, permettant aux cellules cancéreuses de contourner les mécanismes anti-tumoraux basés sur l'apoptose ou l'arrêt de croissance des cellules ayant un comportement anarchique.

L'analyse des variations du protéome rejoignent les constatations faites sur le protéome total. La présence des protéines virales se traduit bien par des variations

d'abondance protéiques pouvant être dues à des perturbations épigénétiques. Parmi les protéines dont l'abondance est affectée, il est possible de retrouver de nombreux proto-oncogènes et protéines impliquées dans la réplication et/ou la réponse aux dommages de l'ADN. Ces analyses ont également fourni des informations intéressantes sur de potentielles voies d'évasion immunitaire. En effet, nous avons pu observer que la protéine HLA-C impliquée dans la présentation antigénique est plus fortement exprimée lorsque l'expression des antigènes T est inhibée par shRNA. Ainsi il semblerait qu'un mécanisme de répression de son expression soit liée à la présence des protéines virales. L'identification de ce mécanisme pourrait permettre une meilleure compréhension des mécanismes d'évasion immunitaire des CCM-VP et mener au développement de thérapies adjuvantes pour potentialiser les immunothérapies. Une étude menée par le laboratoire d'Antoine Touzé avec lequel nous avons collaboré a été réalisée sur l'implication de plusieurs facteurs épigénétiques dans ce mécanisme et fournit des résultats encourageant sur l'utilisation potentielle de nouvelles thérapies adjuvantes. A l'inverse, la glycoprotéine CD59 est exprimée en présence des antigènes T et l'inhibition de leur expression entraine une perte de l'expression de cette protéine. La réexpression de HLA-C et l'inhibition de CD59 peuvent donc présenter un intérêt comme stratégie de sensibilisation des CCM-VP aux immunothérapies, potentiellement en promouvant l'immunogénicité de ces tumeurs et l'infiltration de lymphocytes T et NK au sein de la masse tumorale.

Parmi les résultats obtenus dans cette partie exploratoire, en plus des grandes fonctions d'intérêt identifiées comme l'épigénétique et la réponse aux dommages de l'ADN, certaines interactions ont retenu notre attention. C'est le cas de l'interaction entre PIN1 et tLT identifiée par GST-Pull Down qui présente un intérêt particulier. En effet, cette peptidylprolyl cis/trans isomérase est connue pour son rôle de régulation de nombreuses cibles protéiques par isomérisation de leurs prolines. Il est alors envisageable que PIN1 soit recrutée par tLT afin d'utiliser cette propriété dans l'oncogenèse. L'interaction entre tLT et PIN1 a pu être confirmée par PLA au sein de deux lignées CCM-VP exprimant différentes formes de LT tronqué. Il semble ainsi que cette interaction soit conservée au sein des différentes formes de MCPyV retrouvées.

En comparant les niveaux d'expression de PIN1 et FBXW7 au sein des lignées CCM-VP et CCM-VN, nous avons mis en évidence une surexpression de PIN1 associée à une sous-

expression de FBXW7 dans les cellules exprimant les antigènes T. Il a été rapporté que la stabilité de FBXW7 pouvait être régulée par l'action de PIN1, entrainant son adressage au protéasome. Ces résultats rejoignent ainsi l'hypothèse émise plus tôt sur un recrutement de PIN1 par tLT au niveau de certaines cibles, ici FBXW7, afin de réguler leurs fonctions ou stabilités. Ce type de mécanisme serait particulièrement intéressant au regard du rôle des CRLs et notamment du complexe SCF/FBXW7 dans la régulation de la dégradation de facteurs de transcription, proto-oncogènes et suppresseurs de tumeur. Ainsi, la diminution du niveau de FBXW7 dans les CCM-VP pourrait participer de manière importante à la reprogrammation du protéome observée dans ces cancers. L'éventualité d'une relation entre PIN1 et les CRLs est validée par l'analyse de l'interactome de PIN1. En effet, en plus de cibles connues de cette protéine qui valident les données, il est possible de retrouver plusieurs composants des complexes SCF et même de manière plus générale, des CRLs parmi ses interacteurs. L'activité des CRLs est régulée par neddylation qui entraine un changement conformationnel nécessaire au transfert de l'ubiquitine. En associant nos données à celles de la littérature, il apparait très probable que PIN1 soit responsable de ces changements conformationnels qui suivent la neddylation d'un complexe SCF. La neddylation agirait comme un signal de recrutement de PIN1, entrainant alors l'isomérisation d'un ou plusieurs composants du complexe et permettant de réduire l'écart de 50 Å entre l'E2 et le substrat, permettant *in fine* son ubiquitination.

L'étude de cette potentielle relation entre NEDD8 et PIN1 par Split-TurboID a dans un premier temps confirmé l'hypothèse d'interaction entre les deux protéines. Cette analyse a également situé cette interaction au niveau des CRL de manière générale, avec notamment dans leur environnement proximal la majorité des cullins, RBX1 et SKP1. L'utilisation d'un temps de biotinylation plus long a également montré que l'interaction PIN1-NEDD8 avait lieu à proximité de protéines impliquées dans un mécanisme influencé par la neddylation. Ainsi, ces résultats appuient l'hypothèse d'un recrutement de PIN1 par la neddylation. L'interaction entre PIN1 et NEDD8 a pu être également confirmée par PLA. De la même façon, la localisation de PIN1 au niveau des CRLs a été observée par PLA entre PIN1 et RBX1 composant majeur de ces complexes. Ainsi, l'étude de processus oncogéniques au sein des CCM-VP a mené à l'identification d'un potentiel nouveau mécanisme général et fondamental en biologie. La modulation du système ubiquitine-protéasome (UPS), notamment des complexes CRLs qui

ciblent de nombreux oncogènes et suppresseurs de tumeur, est un champ de recherche très prometteur en cancérologie (Zhou and Sun, 2021). La neddylation, système complet de modification post-traductionnel fonctionnant en cascade comme l'ubiquitine, est également ciblable et la compréhension de son rôle dans la régulation des CRLs représenterai une étape importante dans ce domaine.

Ces travaux ouvrent de nombreuses perspectives dans un cadre bien plus large que les CCM. Cette thématique est actuellement poursuivie par un projet postdoctoral. Il sera désormais intéressant de confirmer l'effet du recrutement de PIN1 sur les complexes SCF comme les complexes CRLs. Il semble également intéressant d'observer l'effet de l'inhibition de PIN1 sur l'activité de ces complexes. L'expression d'une forme mutée de PIN1 ne pouvant pas interagir avec NEDD8 peut également avoir son intérêt dans ce cadre.

Sur un plan technologique, l'étude de l'interaction PIN1-NEDD8 a également permis la mise en place d'outils Split-TurboID à mi-chemin entre interactomique supervisée et semi-supervisée. L'expression stable inductible des deux protéines de fusion à partir d'un seul vecteur comportant une séquence IRES n'avait pas encore été réalisé à ce jour. L'utilisation de ce type de vecteurs apporte une solution aux problèmes posés par la surexpression ou la nécessité de co-transfecter plusieurs vecteurs. Les résultats obtenus sur la paire PIN1-NEDD8 après 4h et 24h de biotinylation suggèrent que 4h est un temps suffisant pour l'obtention d'un résultat probant. L'allongement du temps de biotinylation permet l'obtention de résultats plus complets. Ce temps de biotinylation entraine cependant l'accroissement de nombre d'interacteurs d'un facteur 10, il est alors possible que le bruit de fond augmente avec la durée de biotinylation. Enfin, la durée de biotinylation semble nécessiter un ajustement au cas par cas, selon le type d'interaction étudiée et la stabilité de cette interaction (données non présentées ici).

L'analyse ciblée de mécanismes potentiels basée sur les analyses à l'échelle du système protéique nous a également orienté vers l'étude de la protéine EHMT2. La forte présence de régulateurs épigénétiques dans l'interactome des protéines virales associée aux dérégulations épigénétiques observées dans les CCM-VP laissait supposer un effet important de l'interaction tLT-EHMT2 sur sa fonction épigénétique.

L'analyse de l'interactome d'EHMT2 en présence ou non de tLT a montré un interactome proximal drastiquement augmenté lorsque la protéine virale était présente. Le nombre accru d'interactions suggérait une extension de la distribution d'EHMT2 sur et autour de la chromatine grâce à son interaction avec tLT. Compte tenu du fait que l'inhibition de l'activité méthyltransférase d'EHMT2 n'a pas changé de manière significative le contenu protéique des cellules, l'augmentation des interactions avec des protéines impliquées dans la réponse aux dommages de l'ADN a suggéré un rôle prépondérant de cette protéine dans le contrôle des dommages de l'ADN plutôt que dans la régulation épigénétique dans le contexte CCM-VP.

Le rétablissement spectaculaire de la prolifération cellulaire suite à l'inhibition de EHMT2, après déplétion des antigènes T dans les cellules CCM-VP, a démontré une dépendance à EHMT2 du mécanisme d'arrêt du cycle cellulaire normalement induit dans ces conditions. Outre son rôle dans la réponse aux dommages de l'ADN, il a été rapporté qu'EHMT2 était responsable de la répression transcriptionnelle de gènes liés au cycle cellulaire tels que RB1. L'inhibition de l'activité méthyltransférase d'EHMT2 s'accompagne d'une augmentation de l'expression de ce suppresseur de tumeur (Rao *et al.*, 2016). Si ces résultats étaient transposables au contexte des cellules CCM-VP après extinction des antigènes T par shRNA, l'inhibition d'EHMT2 devrait conduire à un arrêt strict du cycle cellulaire et ne devrait donc pas permettre un rétablissement de la croissance. Les résultats obtenus ici suggèrent un rôle plus complexe d'EHMT2 dans le contexte CCM-VP, impliquant possiblement sa relation avec la signalisation de la réponse aux dommages de l'ADN suite au stress réplicatif causé par la prolifération soutenue des cellules CCM-VP.

En outre, des travaux ont montré que EHMT2 était impliqué dans les mécanismes de réparation des cassures double-brin de l'ADN (Watanabe *et al.*, 2018). Nous montrons ici un effet dépendant de l'activité methyltransferase d'EHMT2 sur la réponse aux cassures simple-brin. Dans les lignées cellulaires CCM et non-CCM, l'inhibition d'EHMT2 conduit à des niveaux accrus de phospho-CHK1 qui transmet le signal suite aux cassures simple-brin. Ces résultats suggèrent que l'efficacité de la transmission du signal de DDR via la phosphorylation de CHK1 est indépendante de la fonction méthyltransférase d'EHTM2. Par contre, l'efficacité réduite de la réparation de l'ADN observée en comet assay lorsque EHTM2 est inhibé suggère une implication de son activité methyltransferase dans l'efficacité de la réparation des dommages

à l'ADN, soit par défaut de recrutement des facteurs de réparation, soit dans leur capacité à réparer les dommages. Les comptages de foyers γH2Ax dans les différentes conditions ont révélé une différence flagrante entre les cellules avec expression des antigènes T et celles après extinction des antigènes T par shRNA. Ces dernières présentaient plus de foyers γH2Ax que les WT lorsqu'elles étaient traitées avec de l'étoposide mais moins lorsqu'elles étaient traitées à la fois avec de l'étoposide et du BRD4770. Le nombre de foyers γH2Ax n'a pas été affecté par l'inhibition de EHMT2 dans les cellules exprimant les antigènes T. Ces différences de réaction au traitement combiné des cellules suggèrent une compensation de l'inhibition d'EHMT2 par la présence des antigènes T. Dans l'ensemble, nos travaux montrent pour la première fois une implication des antigènes T dans la gestion des dommages à l'ADN dans les CCM-VP. Plus précisément, nous décrivons la centralité d'un contrôle coordonné de la voie de réponse aux dommages de l'ADN par l'action coopérative de tLT et d'au moins EHMT2. Nos expériences dans les cellules U2OS ont révélé que l'activité d'EHMT2 sur la voie de réponse aux dommages de l'ADN semble non seulement dépendre de l'expression de tLT, mais aussi du contexte cellulaire. Cette dernière observation suggère fortement que des cofacteurs hôtes supplémentaires coopèrent pour moduler le type d'activité qu'EHMT2 aurait sur la voie de réponse aux dommages de l'ADN. Bien que des expériences supplémentaires soient nécessaires pour caractériser les interactions fonctionnelles de EHMT2 susceptibles de moduler son impact sur la réponse aux dommages de l'ADN, nos travaux révèlent : une forte interdépendance de tLT et EHMT2 dans le contrôle des dommages de l'ADN ; la centralité du contrôle des dommages de l'ADN dépendant de tLT dans le CCM-VP ; et l'implication jusqu'ici ignorée d'EHMT2 dans la signalisation des cassures simple-brin. L'ensemble de nos travaux suggèrent que l'utilisation de chimiothérapies induisant des dommages à l'ADN (ex : cisplatine, doxorubicine) pourrait être potentialisée par l'inhibition de EHMT2 dans les CCM-VP, empêchant ainsi la résolution efficace des dommages chimio-induits. Cependant, les recherches de thérapeutiques récentes ciblant des facteurs épigénétiques échouent le plus souvent et présentent de meilleurs résultats en combinaison avec d'autres traitements (Feng and De Carvalho, 2022). Nous pensons que notre connaissance des rôles des facteurs tels que EHMT2 sont encore très parcellaires. Nos travaux apportent donc également des nouveaux éléments au corpus de connaissance, qui *in fine* permettra de rationaliser le développement et l'utilisation de telles thérapeutiques.

Nos résultats révélant un rôle d'EHMT2 dans la signalisation et la réparation des dommages de l'ADN résonnent particulièrement bien avec une étude publiée dans Nature Cell Biology en juillet 2023 par Gaggioli *et al*. et impliquant EHMT2 dans la protection des fourches de réplication bloquées (Gaggioli *et al.*, 2023). Plus précisément, ils ont constaté qu'un stress réplicatif prolongé induit par HU conduisait à une triméthylation accrue de H3K9 par EHMT2 à proximité des fourches bloquées et ont décrit EHMT2 comme un nouveau point de contrôle du stress réplicatif. Cette augmentation des marques d'hétérochromatine H3K9me3 entraîne une compaction de la chromatine autour des fourches stressées, ce qui aide à maintenir l'intégrité de la fourche et à prévenir les dommages de l'ADN. Bien que nos données actuelles indiquent l'implication de EHMT2 dans les étapes en aval de la réponse aux dommages de l'ADN telles que la signalisation et la réparation, l'étude de Gaggioli *et al.* suggère que EHMT2 pourrait également agir en amont pendant le stress réplicatif pour induire un niveau de compaction protecteur de la chromatine, préservant la structure de la fourche bloquée. À la lumière de l'effet d'échaffaudage apparent de tLT au niveau de la fourche de réplication bloquée, les résultats de Gaggioli *et al.* suggèrent un effet protecteur dans les CCM-VP via un recrutement de EHMT2. En effet, nous pouvons faire l'hypothèse que la présence constitutive de EHMT2 au niveau des fourches bloquées protégerait les cellules CCM-VP de l'effondrement de la fourche et des cassures double-brin de l'ADN, contribuant ainsi à la remarquable stabilité génomique observée dans ces cancers. Ces résultats indiquent que EHMT2 pourrait participer à plusieurs étapes de la réponse au stress réplicatif, d'abord en favorisant des changements chromatiniens qui protègent la stabilité de la fourche, puis en facilitant une signalisation et une réparation efficaces des dommages de l'ADN. L'exploration plus approfondie des rôles complexes et dépendant du contexte des régulateurs épigénétiques comme EHMT2 à différentes phases de la réponse au stress réplicatif devrait fournir des informations précieuses sur la façon dont la dérégulation de ces mécanismes peut favoriser la réplication continue et la stabilité génomique dans le cancer. En somme, le cœur de nos travaux sur EHMT2 et la publication de cet article démontrant son rôle de checkpoint dans le stress réplicatif suggère très fortement que tLT serait capable d'exploiter deux checkpoints par ses interactions : en inhibant RB1 et permettant ainsi l'entrée en phase S, et en recrutant EHMT2 aux fourches de réplication, les protégeant d'un effondrement et en améliorant la résolution des dommages à l'ADN.

En plus d'EHMT2, d'autres histones méthyltransférases liées à l'oncogenèse des CCM sont impliquées dans la réponse aux dommages de l'ADN. Xu *et al.* ont montré dans le myélome multiple que l'inhibition d'EZH2 était responsable d'une sensibilisation des cellules aux dommages de l'ADN (Xu *et al.*, 2019). Le besoin de contrôle des dommages de l'ADN dans les CCM pourrait alors expliquer la cytotoxicité sélective de l'inhibition d'EZH2 dans les lignées cellulaires CCM-VP (Durand *et al.*, 2023).

Compte tenu de la place croissante dans le développement thérapeutique d'inhibiteurs ciblant des facteurs épigénétiques, l'implication de certains de ces facteurs dans la stabilité génétique et la résistance aux dommages de l'ADN est à considérer. En effet, l'inhibition de ces facteurs pourrait permettre de sensibiliser ces cellules aux dommages induits par leur rythme réplicatif.

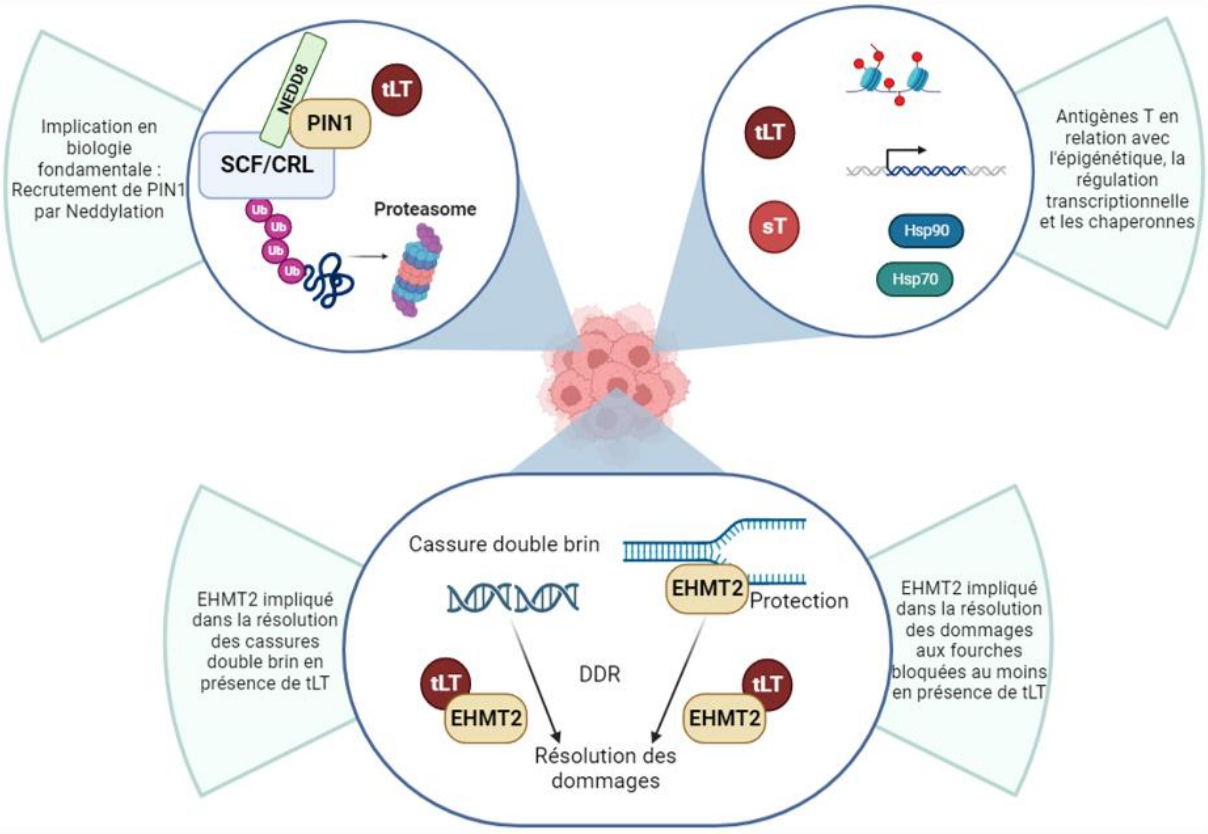


Figure 24 : Schéma récapitulatif des points importants des conclusions

Les conclusions de ces travaux sont synthétisées de manière schématique. Les mécanismes représentés ici présentent une implication en dehors du cadre CCM illustrant le potentiel de modèle d'étude des oncovirus. L'interaction entre tLT et PIN1 a permis l'identification d'une relation entre PIN1 et NEDD8 mettant en lumière un potentiel mécanisme de régulation des complexes SCF et CRL par isomérisation. L'interactomique a révélé une forte relation entre les antigènes T et la régulation épigénétique, transcriptionnel et la mise en conformation des protéines. Les études ciblées sur EHMT2 ont mis en lumière son implication dans la DDR et son effet sur la résolution des cassures.

6 BIBLIOGRAPHIE

Akhbari, P. *et al.* (2018) 'MCV-miR-M1 Targets the Host-Cell Immune Response Resulting in the Attenuation of Neutrophil Chemotaxis', *The Journal of Investigative Dermatology*, 138(11), pp. 2343–2354. Available at: https://doi.org/10.1016/j.jid.2018.03.1527.

Allen, P.J. *et al.* (2005) 'Merkel Cell Carcinoma: Prognosis and Treatment of Patients From a Single Institution', *J Clin Oncol*, 23, pp. 2300–2309. Available at: https://doi.org/10.1200/JCO.2005.02.329.

Al-Showbaki, L. *et al.* (2021) 'Network Meta-analysis Comparing Efficacy, Safety and Tolerability of Anti-PD-1/PD-L1 Antibodies in Solid Cancers', *Journal of Cancer*, 12(14), pp. 4372–4378. Available at: https://doi.org/10.7150/jca.57413.

Angermeyer, S. *et al.* (2013) 'Merkel cell polyomavirus-positive Merkel cell carcinoma cells do not require expression of the viral small T antigen', *The Journal of Investigative Dermatology*, 133(8), pp. 2059–2064. Available at: https://doi.org/10.1038/jid.2013.82.

Antonangeli, F. *et al.* (2020) 'Regulation of PD-L1 Expression by NF-κB in Cancer', *Frontiers in Immunology*, 11, p. 584626. Available at: https://doi.org/10.3389/fimmu.2020.584626.

Becker, J.C. *et al.* (2018) 'Epidemiology, biology and therapy of Merkel cell carcinoma: conclusions from the EU project IMMOMEC', *Cancer Immunology, Immunotherapy*, 67(3), pp. 341–351. Available at: https://doi.org/10.1007/s00262-017-2099-3.

Becker, M. *et al.* (2019) 'Infectious Entry of Merkel Cell Polyomavirus', *Journal of Virology*, 93(6), pp. e02004-18. Available at: https://doi.org/10.1128/JVI.02004-18.

Béganton, B. *et al.* (2019) 'Approches nouvelles pour l'étude des interactions protéine-protéine', *médecine/sciences*, 35(3), pp. 223–231. Available at: https://doi.org/10.1051/medsci/2019035.

Bjornsti, M.-A. and Houghton, P.J. (2004) 'Lost in translation: dysregulation of cap-dependent translation and cancer', *Cancer Cell*, 5(6), pp. 519–523. Available at: https://doi.org/10.1016/j.ccr.2004.05.027.

Bjursell, G. (1978) 'Effects of 2'-deoxy-2'-azidocytidine on polyoma virus DNA replication: evidence for rolling circle-type mechanism', *Journal of Virology*, 26(1), pp. 136–142. Available at: https://doi.org/10.1128/jvi.26.1.136-142.1978.

Blumberg, B.S. and Alter, H.J. (1965) 'A "New" Antigen in Leukemia Sera', *JAMA*, 191(7), pp. 541–546. Available at: https://doi.org/10.1001/jama.1965.03080070025007.

Bobrich, M.A. *et al.* (2013) 'PTPIP51: a new interaction partner of the insulin receptor and PKA in adipose tissue', *Journal of Obesity*, 2013, p. 476240. Available at: https://doi.org/10.1155/2013/476240.

Borchert, S. *et al.* (2014) 'High-affinity Rb binding, p53 inhibition, subcellular localization, and transformation by wild-type or tumor-derived shortened Merkel cell polyomavirus large T

antigens', *Journal of Virology*, 88(6), pp. 3144–3160. Available at: https://doi.org/10.1128/JVI.02916-13.

Boyerinas, B. *et al.* (2015) 'Antibody-Dependent Cellular Cytotoxicity Activity of a Novel Anti–PD-L1 Antibody Avelumab (MSB0010718C) on Human Tumor Cells', *Cancer Immunology Research*, 3(10), pp. 1148–1157. Available at: https://doi.org/10.1158/2326-6066.CIR-15-0059.

Branon, T.C. *et al.* (2018) 'Efficient proximity labeling in living cells and organisms with TurboID', *Nature Biotechnology*, 36(9), pp. 880–898. Available at: https://doi.org/10.1038/nbt.4201.

Brückner, A. *et al.* (2009) 'Yeast Two-Hybrid, a Powerful Tool for Systems Biology', *International Journal of Molecular Sciences*, 10(6), pp. 2763–2788. Available at: https://doi.org/10.3390/ijms10062763.

Busam, K.J. *et al.* (2017) 'Reduced H3K27me3 expression in Merkel cell polyoma virus-positive tumors', *Modern Pathology*, 30(6), pp. 877–883. Available at: https://doi.org/10.1038/modpathol.2017.8.

Caplan, S. *et al.* (2001) 'Human Vam6p promotes lysosome clustering and fusion in vivo', *Journal of Cell Biology*, 154(1), pp. 109–122. Available at: https://doi.org/10.1083/jcb.200102142.

Carter, J.J. *et al.* (2013) 'Identification of an overprinting gene in Merkel cell polyomavirus provides evolutionary insight into the birth of viral genes', *Proceedings of the National Academy of Sciences*, 110(31), pp. 12744–12749. Available at: https://doi.org/10.1073/pnas.1303526110.

Chang, Y. *et al.* (1994) 'Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma', *Science (New York, N.Y.)*, 266(5192), pp. 1865–1869. Available at: https://doi.org/10.1126/science.7997879.

Chellappan, S.P. *et al.* (1991) 'The E2F transcription factor is a cellular target for the RB protein', *Cell*, 65(6), pp. 1053–1061. Available at: https://doi.org/10.1016/0092-8674(91)90557-F.

Chen, T. *et al.* (2011) 'Serological evidence of Merkel cell polyomavirus primary infections in childhood', *Journal of Clinical Virology*, 50(2), pp. 125–129. Available at: https://doi.org/10.1016/j.jcv.2010.10.015.

Cheng, J. *et al.* (2017) 'Merkel cell polyomavirus recruits MYCL to the EP400 complex to promote oncogenesis', *PLoS Pathogens*, 13(10). Available at: https://doi.org/10.1371/journal.ppat.1006668.

Cho, K.F. *et al.* (2020) 'Split-TurboID enables contact-dependent proximity labeling in cells', *Proceedings of the National Academy of Sciences*, 117(22), pp. 12143–12154. Available at: https://doi.org/10.1073/PNAS.1919528117.

Choo, Q.L. *et al.* (1989) 'Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome', *Science (New York, N.Y.)*, 244(4902), pp. 359–362. Available at: https://doi.org/10.1126/science.2523562.

Chteinberg, E. *et al.* (2020) 'The curious case of Merkel cell carcinoma: epigenetic youth and lack of pluripotency', *https://doi.org/10.1080/15592294.2020.1773096*, 15(12), pp. 1319–1324. Available at: https://doi.org/10.1080/15592294.2020.1773096.

Couturier, C. and Deprez, B. (2012) 'Setting Up a Bioluminescence Resonance Energy Transfer High throughput Screening Assay to Search for Protein/Protein Interaction Inhibitors in Mammalian Cells', *Frontiers in Endocrinology*, 3. Available at: https://www.frontiersin.org/articles/10.3389/fendo.2012.00100 (Accessed: 5 September 2023).

Couzens, A.L. *et al.* (2013) 'Protein Interaction Network of the Mammalian Hippo Pathway Reveals Mechanisms of Kinase-Phosphatase Interactions', *Science Signaling*, 6(302), pp. rs15–rs15. Available at: https://doi.org/10.1126/scisignal.2004712.

Cox, J. and Mann, M. (2008) 'MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification', *Nature Biotechnology*, 26(12), pp. 1367–1372. Available at: https://doi.org/10.1038/nbt.1511.

Coyaud, E. *et al.* (2018) 'Global interactomics uncovers extensive organellar targeting by Zika Virus', *Molecular and Cellular Proteomics*, 17(11), pp. 2242–2255. Available at: https://doi.org/10.1074/mcp.TIR118.000800.

Czech-Sioli, M. *et al.* (2020) 'High-resolution analysis of Merkel Cell Polyomavirus in Merkel Cell Carcinoma reveals distinct integration patterns and suggests NHEJ and MMBIR as underlying mechanisms', *PLoS pathogens*, 16(8), p. e1008562. Available at: https://doi.org/10.1371/journal.ppat.1008562.

D'Angelo, S.P. *et al.* (2020) 'Avelumab in patients with previously treated metastatic Merkel cell carcinoma: long-term data and biomarker analyses from the single-arm phase 2 JAVELIN Merkel 200 trial', *Journal for Immunotherapy of Cancer*, 8(1), p. e000674. Available at: https://doi.org/10.1136/jitc-2020-000674.

De Munter, S. *et al.* (2017) 'Split-BioID: a proximity biotinylation assay for dimerization-dependent protein interactions', *FEBS letters*, 591(2), pp. 415–424. Available at: https://doi.org/10.1002/1873-3468.12548.

De Wolf-Peeters, C. *et al.* (1980) 'A Cutaneous APUDoma or Merkel Cell Tumor? A Morphologically Recognizable Tumor with a Biological and Histological Malignant Aspect in Contrast with Its Clinical Behavior PUDOMAS ARE TUMORS originating from cells', *Cancer*, 46, pp. 1810–1816. Available at: https://doi.org/10.1002/1097-0142.

Deshaies, R.J., Emberley, E.D. and Saha, A. (2010) 'Control of cullin-ring ubiquitin ligase activity by nedd8', *Sub-Cellular Biochemistry*, 54, pp. 41–56. Available at: https://doi.org/10.1007/978-1-4419-6676-6_4.

Di Paolo, G. *et al.* (1997) 'Phosphorylation regulates the microtubule-destabilizing activity of stathmin and its interaction with tubulin', *FEBS Letters*, 416(2), pp. 149–152. Available at: https://doi.org/10.1016/S0014-5793(97)01188-5.

Dingar, D. *et al.* (2015) 'BioID identifies novel c-MYC interacting partners in cultured cells and xenograft tumors', *Journal of Proteomics*, 118, pp. 95–111. Available at: https://doi.org/10.1016/j.jprot.2014.09.029.

Duncavage, E.J. *et al.* (2011) 'Hybrid capture and next-generation sequencing identify viral integration sites from formalin-fixed, paraffin-embedded tissue', *The Journal of molecular diagnostics: JMD*, 13(3), pp. 325–333. Available at: https://doi.org/10.1016/j.jmoldx.2011.01.006.

Durand, M.-A. *et al.* (2023) 'Distinct Regulation of EZH2 and its Repressive H3K27me3 Mark in Polyomavirus-Positive and -Negative Merkel Cell Carcinoma', *The Journal of Investigative Dermatology*, pp. S0022-202X(23)01973–5. Available at: https://doi.org/10.1016/j.jid.2023.02.038.

Dye, K.N. *et al.* (2019) 'Merkel cell polyomavirus Tumor antigens expressed in Merkel cell carcinoma function independently of the ubiquitin ligases Fbw7 and β-TrCP', *PLoS Pathogens*, 15(1). Available at: https://doi.org/10.1371/journal.ppat.1007543.

Edamoto, Y. *et al.* (2003) 'Alterations of RB1, p53 and Wnt pathways in hepatocellular carcinomas associated with hepatitis C, hepatitis B and alcoholic liver cirrhosis', *International Journal of Cancer*, 106(3), pp. 334–341. Available at: https://doi.org/10.1002/ijc.11254.

Epstein, M.A., Achong, B.G. and Barr, Y.M. (1964) 'VIRUS PARTICLES IN CULTURED LYMPHOBLASTS FROM BURKITT'S LYMPHOMA', *The Lancet*, 283(7335), pp. 702–703. Available at: https://doi.org/10.1016/S0140-6736(64)91524-7.

Ercilla, A. *et al.* (2020) 'Acute hydroxyurea-induced replication blockade results in replisome components disengagement from nascent DNA without causing fork collapse', *Cellular and molecular life sciences: CMLS*, 77(4), pp. 735–749. Available at: https://doi.org/10.1007/s00018-019-03206-1.

Eyckerman, S. *et al.* (2016) 'Trapping mammalian protein complexes in viral particles', *Nature Communications*, 7(1), p. 11416. Available at: https://doi.org/10.1038/ncomms11416.

Fan, K., Zebisch, A., *et al.* (2020) 'Highly Expressed miR-375 is not an Intracellular Oncogene in Merkel Cell Polyomavirus-Associated Merkel Cell Carcinoma', *Cancers*, 12(3), p. 529. Available at: https://doi.org/10.3390/cancers12030529.

Fan, K., Gravemeyer, J., *et al.* (2020) 'MCPyV Large T Antigen-Induced Atonal Homolog 1 Is a Lineage-Dependency Oncogene in Merkel Cell Carcinoma', *The Journal of Investigative Dermatology*, 140(1), pp. 56-65.e3. Available at: https://doi.org/10.1016/j.jid.2019.06.135.

Fan, K. *et al.* (2021) 'Merkel cell carcinoma-derived exosome-shuttle miR-375 induces fibroblast polarization by inhibition of RBPJ and p53', *Oncogene*, 40(5), pp. 980–996. Available at: https://doi.org/10.1038/s41388-020-01576-6.

Feng, H. *et al.* (2007) 'Human Transcriptome Subtraction by Using Short Sequence Tags To Search for Tumor Viruses in Conjunctival Carcinoma', *Journal of Virology*, 81(20), pp. 11332–11340. Available at: https://doi.org/10.1128/jvi.00875-07.

Feng, H. *et al.* (2008) 'Clonal integration of a polyomavirus in human Merkel cell carcinoma', *Science*, 319(5866), pp. 1096–1100. Available at: https://doi.org/10.1126/science.1152586.

Feng, S. and De Carvalho, D.D. (2022) 'Clinical advances in targeting epigenetics for cancer therapy', *The FEBS Journal*, 289(5), pp. 1214–1239. Available at: https://doi.org/10.1111/febs.15750.

Fields, S. and Song, O. (1989) 'A novel genetic system to detect protein–protein interactions', *Nature*, 340(6230), pp. 245–246. Available at: https://doi.org/10.1038/340245a0.

Foulongne, V. *et al.* (2011) 'Detection of Merkel cell polyomavirus on environmental surfaces', *Journal of Medical Virology*, 83(8), pp. 1435–1439. Available at: https://doi.org/10.1002/jmv.22110.

Gaggioli, V. *et al.* (2023) 'Dynamic de novo heterochromatin assembly and disassembly at replication forks ensures fork stability', *Nature Cell Biology*, 25(7), pp. 1017–1032. Available at: https://doi.org/10.1038/s41556-023-01167-z.

Garman, K. *et al.* (2021) 'Intersectional Screening Identifies Synergistic Combinatorial Therapies for Merkel Cell Carcinoma', *The FASEB Journal*, 35(S1). Available at: https://doi.org/10.1096/fasebj.2021.35.S1.03679.

Gartin, A.K. *et al.* (2022) 'Merkel Cell Carcinoma Sensitivity to EZH2 Inhibition Is Mediated by SIX1 Derepression', *Journal of Investigative Dermatology* [Preprint]. Available at: https://doi.org/10.1016/J.JID.2022.03.008.

Genest, O., Wickner, S. and Doyle, S.M. (2019) 'Hsp90 and Hsp70 chaperones: Collaborators in protein remodeling', *The Journal of Biological Chemistry*, 294(6), pp. 2109–2120. Available at: https://doi.org/10.1074/jbc.REV118.002806.

Geser, A. *et al.* (1982) 'Final case reporting from the ugandan prospective study of the relationship between ebv and burktit's lymphoma', *International Journal of Cancer*, 29(4), pp. 397–400. Available at: https://doi.org/10.1002/ijc.2910290406.

Giannecchini, S. (2020) 'Evidence of the Mechanism by Which Polyomaviruses Exploit the Extracellular Vesicle Delivery System during Infection', *Viruses*, 12(6), p. 585. Available at: https://doi.org/10.3390/v12060585.

Goh, G. *et al.* (2016) 'Mutational landscape of MCPyV-positive and MCPyV-negative merkel cell carcinomas with implications for immunotherapy', *Oncotarget*, 7(3), pp. 3403–3415. Available at: https://doi.org/10.18632/oncotarget.6494.

Götze, M. *et al.* (2015) 'Automated Assignment of MS/MS Cleavable Cross-Links in Protein 3D-Structure Analysis', *Journal of the American Society for Mass Spectrometry*, 26(1), pp. 83–97. Available at: https://doi.org/10.1007/s13361-014-1001-1.

Gravemeyer, J. *et al.* (2022) 'DNA-methylation patterns imply a common cellular origin of virus- and UV-associated Merkel cell carcinoma', *Oncogene*, 41(1), pp. 37–45. Available at: https://doi.org/10.1038/s41388-021-02064-1.

Griffiths, D.A. *et al.* (2013) 'Merkel cell polyomavirus small T antigen targets the NEMO adaptor protein to disrupt inflammatory signaling', *Journal of Virology*, 87(24), pp. 13853–13867. Available at: https://doi.org/10.1128/JVI.02159-13.

Gujar, H. *et al.* (2021) 'Characterizing DNA methylation signatures and their potential functional roles in Merkel cell carcinoma', *Genome Medicine 2021 13:1*, 13(1), pp. 1–17. Available at: https://doi.org/10.1186/S13073-021-00946-3.

Gupta, G.D. *et al.* (2015) 'A Dynamic Protein Interaction Landscape of the Human Centrosome-Cilium Interface', *Cell*, 163(6), pp. 1484–1499. Available at: https://doi.org/10.1016/j.cell.2015.10.065.

Gursoy-Yuzugullu, O. *et al.* (2017) 'Epigenetic therapy with inhibitors of histone methylation suppresses DNA damage signaling and increases glioma cell radiosensitivity', *Oncotarget*, 8(15), pp. 24518–24532. Available at: https://doi.org/10.18632/oncotarget.15543.

Hamdi, A. and Colas, P. (2012) 'Yeast two-hybrid methods and their applications in drug discovery', *Trends in Pharmacological Sciences*, 33(2), pp. 109–118. Available at: https://doi.org/10.1016/j.tips.2011.10.008.

Han, Y., Liu, D. and Li, L. (2020) 'PD-1/PD-L1 pathway: current researches in cancer', *American Journal of Cancer Research*, 10(3), pp. 727–742.

Harms, K.L. *et al.* (2017) 'Increased expression of EZH2 in Merkel cell carcinoma is associated with disease progression and poorer prognosis', *Human Pathology*, 67, pp. 78–84. Available at: https://doi.org/10.1016/j.humpath.2017.07.009.

Harms, P.W. *et al.* (2015) 'The Distinctive Mutational Spectra of Polyomavirus-Negative Merkel Cell Carcinoma', *Cancer Research*, 75(18), pp. 3720–3727. Available at: https://doi.org/10.1158/0008-5472.CAN-15-0702.

Harold, A. *et al.* (2019) 'Conversion of Sox2-dependent Merkel cell carcinoma to a differentiated neuron-like phenotype by T antigen inhibition', *Proceedings of the National Academy of Sciences of the United States of America*, 116(40), pp. 20104–20114. Available at: https://doi.org/10.1073/pnas.1907154116.

Harrington, C. and Kwan, W. (2014) 'Outcomes of Merkel cell carcinoma treated with radiotherapy without radical surgical excision', *Annals of Surgical Oncology*, 21(11), pp. 3401–3405. Available at: https://doi.org/10.1245/s10434-014-3757-8.

Hassin, O. and Oren, M. (2023) 'Drugging p53 in cancer: one protein, many targets', *Nature Reviews Drug Discovery*, 22(2), pp. 127–144. Available at: https://doi.org/10.1038/s41573-022-00571-8.

zur Hausen, A. *et al.* (2013) 'Early B-Cell Differentiation in Merkel Cell Carcinomas: Clues to Cellular Ancestry', *Cancer Research*, 73(16), pp. 4982–4987. Available at: https://doi.org/10.1158/0008-5472.CAN-13-0616.

zur Hausen, H. *et al.* (1974) 'Attempts to detect virus-secific DNA in human tumors. I. Nucleic acid hybridizations with complementary RNA of human wart virus', *International Journal of Cancer*, 13(5), pp. 650–656. Available at: https://doi.org/10.1002/ijc.2910130509.

Heath, M. *et al.* (2008) 'Clinical characteristics of Merkel cell carcinoma at diagnosis in 195 patients: the AEIOU features', *Journal of the American Academy of Dermatology*, 58(3), pp. 375–381. Available at: https://doi.org/10.1016/j.jaad.2007.11.020.

Hopp, T.P. *et al.* (1988) 'A short polypeptide marker sequence useful for recombinant protein identification and purification', *Bio/Technology*, 6(10), pp. 1204–1210. Available at: https://doi.org/10.1038/nbt1088-1204.

Houben, R. *et al.* (2010) 'Merkel cell polyomavirus-infected Merkel cell carcinoma cells require expression of viral T antigens', *Journal of Virology*, 84(14), pp. 7064–7072. Available at: https://doi.org/10.1128/JVI.02400-09.

Houben, R. *et al.* (2012) 'An intact retinoblastoma protein-binding site in Merkel cell polyomavirus large T antigen is required for promoting growth of Merkel cell carcinoma cells', *International Journal of Cancer*, 130(4), pp. 847–856. Available at: https://doi.org/10.1002/ijc.26076.

Houben, R. *et al.* (2015) 'Characterization of functional domains in the Merkel cell polyoma virus Large T antigen', *International Journal of Cancer*, 136(5), pp. E290-300. Available at: https://doi.org/10.1002/ijc.29200.

Huang, J. *et al.* (2010) 'G9a and Glp methylate lysine 373 in the tumor suppressor p53', *The Journal of Biological Chemistry*, 285(13), pp. 9636–9641. Available at: https://doi.org/10.1074/jbc.M109.062588.

Hunter, D.V. *et al.* (2018) 'Advillin Is Expressed in All Adult Neural Crest-Derived Neurons', *eNeuro*, 5(5), p. ENEURO.0077-18.2018. Available at: https://doi.org/10.1523/ENEURO.0077-18.2018.

Ito, T. *et al.* (2001) 'A comprehensive two-hybrid analysis to explore the yeast protein interactome', *Proceedings of the National Academy of Sciences*, 98(8), pp. 4569–4574. Available at: https://doi.org/10.1073/pnas.061034498.

Jäger, S. *et al.* (2012) 'Global landscape of HIV-human protein complexes', *Nature*, 481(7381), pp. 365–370. Available at: https://doi.org/10.1038/nature10719.

Jayabalan, A.K. *et al.* (2016) 'NEDDylation promotes stress granule assembly', *Nature Communications*, 7(1), p. 12125. Available at: https://doi.org/10.1038/ncomms12125.

Jenuwein, T. and Allis, C.D. (2001) 'Translating the Histone Code', *Science*, 293(5532), pp. 1074–1080. Available at: https://doi.org/10.1126/science.1063127.

Kang, M.-S. *et al.* (2019) 'Regulation of PCNA cycling on replicating DNA by RFC and RFC-like complexes', *Nature Communications*, 10(1), p. 2420. Available at: https://doi.org/10.1038/s41467-019-10376-w.

Kantola, K. *et al.* (2009) 'Merkel cell polyomavirus DNA in tumor-free tonsillar tissues and upper respiratory tract samples: implications for respiratory transmission and latency', *Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology*, 45(4), pp. 292–295. Available at: https://doi.org/10.1016/j.jcv.2009.04.008.

Kean, J.M. *et al.* (2009) 'Seroepidemiology of human polyomaviruses', *PLoS pathogens*, 5(3), p. e1000363. Available at: https://doi.org/10.1371/journal.ppat.1000363.

Kervarrec, T. *et al.* (2018) 'Merkel cell carcinomas infiltrated with CD33+ myeloid cells and CD8+ T cells are associated with improved outcome', *Journal of the American Academy of Dermatology*, 78(5), pp. 973-982.e8. Available at: https://doi.org/10.1016/j.jaad.2017.12.029.

Kervarrec, T., Samimi, M., *et al.* (2020) 'Merkel Cell Polyomavirus T Antigens Induce Merkel Cell-Like Differentiation in GLI1-Expressing Epithelial Cells', *Cancers*, 12(7), p. 1989. Available at: https://doi.org/10.3390/cancers12071989.

Kervarrec, T., Aljundi, M., *et al.* (2020) 'Polyomavirus-Positive Merkel Cell Carcinoma Derived from a Trichoblastoma Suggests an Epithelial Origin of this Merkel Cell Carcinoma', *The Journal of Investigative Dermatology*, 140(5), pp. 976–985. Available at: https://doi.org/10.1016/j.jid.2019.09.026.

Kido, K. *et al.* (2020) 'AirID, a novel proximity biotinylation enzyme, for analysis of protein–protein interactions', *eLife*. Edited by P.A. Cole et al., 9, p. e54983. Available at: https://doi.org/10.7554/eLife.54983.

Kim, D.I. *et al.* (2014) 'Probing nuclear pore complex architecture with proximity-dependent biotinylation', *Proceedings of the National Academy of Sciences of the United States of America*, 111(24). Available at: https://doi.org/10.1073/PNAS.1406459111.

Kim, G. *et al.* (2018) 'SUV39H1/DNMT3A-dependent methylation of the RB1 promoter stimulates PIN1 expression and melanoma development', *FASEB journal: official publication of the Federation of American Societies for Experimental Biology*, 32(10), pp. 5647–5660. Available at: https://doi.org/10.1096/fj.201700645RRRRR.

Kim, K.H. and Roberts, C.W.M. (2016) 'Targeting EZH2 in cancer', *Nature Medicine*, 22(2), pp. 128–134. Available at: https://doi.org/10.1038/nm.4036.

Kim, S. *et al.* (2022) 'Combined nivolumab and ipilimumab with or without stereotactic body radiation therapy for advanced Merkel cell carcinoma: a randomised, open label, phase 2 trial', *The Lancet*, 400(10357), pp. 1008–1019. Available at: https://doi.org/10.1016/S0140-6736(22)01659-2.

Knight, L.M. *et al.* (2015) 'Merkel cell polyomavirus small T antigen mediates microtubule destabilization to promote cell motility and migration', *Journal of Virology*, 89(1), pp. 35–47. Available at: https://doi.org/10.1128/JVI.02317-14.

Kodaz, H. *et al.* (2017) 'Frequency of Ras Mutations (Kras, Nras, Hras) in Human Solid Cancer', *EJMO*, 1(1), pp. 1–7. Available at: https://doi.org/10.14744/ejmo.2017.22931.

Kurokawa, M. *et al.* (2003) 'CD56: a useful marker for diagnosing Merkel cell carcinoma', *Journal of Dermatological Science*, 31(3), pp. 219–224. Available at: https://doi.org/10.1016/S0923-1811(03)00029-X.

Kushner, J., Papa, A. and Marx, S.O. (2021) 'Use of Proximity Labeling in Cardiovascular Research', *JACC: Basic to Translational Science*, 6(7), pp. 598–609. Available at: https://doi.org/10.1016/j.jacbts.2021.01.005.

Kwun, H.J. *et al.* (2009) 'The Minimum Replication Origin of Merkel Cell Polyomavirus Has a Unique Large T-Antigen Loading Architecture and Requires Small T-Antigen Expression for Optimal Replication', *Journal of Virology*, 83(23), pp. 12118–12128. Available at: https://doi.org/10.1128/jvi.01336-09.

Kwun, H.J. *et al.* (2013) 'Merkel cell polyomavirus small T antigen controls viral replication and oncoprotein expression by targeting the cellular ubiquitin ligase SCF Fbw7', *Cell Host and Microbe*, 14(2), pp. 125–135. Available at: https://doi.org/10.1016/j.chom.2013.06.008.

Kwun, H.J. *et al.* (2015) 'Restricted protein phosphatase 2A targeting by Merkel cell polyomavirus small T antigen', *Journal of Virology*, 89(8), pp. 4191–4200. Available at: https://doi.org/10.1128/JVI.00157-15.

Kwun, H.J., Chang, Y. and Moore, P.S. (2017) 'Protein-mediated viral latency is a novel mechanism for Merkel cell polyomavirus persistence', *Proceedings of the National Academy of Sciences of the United States of America*, 114(20), pp. E4040–E4047. Available at: https://doi.org/10.1073/pnas.1703879114.

Lam, S.S. *et al.* (2015) 'Directed evolution of APEX2 for electron microscopy and proximity labeling', *Nature Methods*, 12(1), pp. 51–54. Available at: https://doi.org/10.1038/nmeth.3179.

Lee, P.C. *et al.* (2022) 'Reversal of viral and epigenetic HLA class I repression in Merkel cell carcinoma', *The Journal of Clinical Investigation*, 132(13). Available at: https://doi.org/10.1172/JCI151666.

Leiendecker, L. *et al.* (2023) 'Human Papillomavirus 42 Drives Digital Papillary Adenocarcinoma and Elicits a Germ Cell–like Program Conserved in HPV-Positive Cancers', *Cancer Discovery*, 13(1), pp. 70–84. Available at: https://doi.org/10.1158/2159-8290.CD-22-0489.

Leitner, A. *et al.* (2016) 'Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines', *Trends in Biochemical Sciences*, 41(1), pp. 20–32. Available at: https://doi.org/10.1016/j.tibs.2015.10.008.

Leoni, C. *et al.* (2011) 'Genome walking in eukaryotes', *The FEBS Journal*, 278(21), pp. 3953–3977. Available at: https://doi.org/10.1111/j.1742-4658.2011.08307.x.

Lewis, C.W. *et al.* (2020) 'Patterns of distant metastases in 215 Merkel cell carcinoma patients: Implications for prognosis and surveillance', *Cancer Medicine*, 9(4), p. 1374. Available at: https://doi.org/10.1002/CAM4.2781.

Li, J. *et al.* (2013) 'Merkel Cell Polyomavirus Large T Antigen Disrupts Host Genomic Integrity and Inhibits Cellular Proliferation', *Journal of Virology*, 87(16), pp. 9173–9188. Available at: https://doi.org/10.1128/jvi.01216-13.

Li, T.-C. *et al.* (2015) 'Characterization of Self-Assembled Virus-Like Particles of Merkel Cell Polyomavirus', *PLOS ONE*, 10(2), p. e0115646. Available at: https://doi.org/10.1371/journal.pone.0115646.

Liang, W. and Yuan, J. (2022) 'Visualizing Endogenous Necrosomes in Necrosomes by In Situ Proximity Ligation Assay', *Current Protocols*, 2(2), p. e388. Available at: https://doi.org/10.1002/cpz1.388.

Lin, J.S. and Lai, E.M. (2017) 'Protein–protein interactions: Co-immunoprecipitation', in *Methods in Molecular Biology*. Humana Press Inc., pp. 211–219. Available at: https://doi.org/10.1007/978-1-4939-7033-9_17.

Liu, W. *et al.* (2016) 'Identifying the Target Cells and Mechanisms of Merkel Cell Polyomavirus Infection', *Cell Host and Microbe*, 19(6), pp. 775–787. Available at: https://doi.org/10.1016/j.chom.2016.04.024.

Liu, X. *et al.* (2011) 'Merkel cell polyomavirus large T antigen disrupts lysosome clustering by translocating human Vam6p from the cytoplasm to the nucleus', *The Journal of Biological Chemistry*, 286(19), pp. 17079–17090. Available at: https://doi.org/10.1074/jbc.M110.192856.

Loyo, M. *et al.* (2010) 'Quantitative detection of Merkel cell virus in human tissues and possible mode of transmission', *International Journal of Cancer*, 126(12), pp. 2991–2996. Available at: https://doi.org/10.1002/ijc.24737.

Ludlow, J.W. (1993) 'Interactions between SV40 large-tumor antigen and the growth suppressor proteins pRB and p53', *The FASEB Journal*, 7(10), pp. 866–871. Available at: https://doi.org/10.1096/fasebj.7.10.8344486.

Ma, S. *et al.* (2021) 'PHF8-promoted TOPBP1 demethylation drives ATR activation and preserves genome stability', *Science Advances*, 7(19), p. eabf7684. Available at: https://doi.org/10.1126/sciadv.abf7684.

Martel-Jantin, C. *et al.* (2012) 'Genetic variability and integration of Merkel cell polyomavirus in Merkel cell carcinoma', *Virology*, 426(2), pp. 134–142. Available at: https://doi.org/10.1016/j.virol.2012.01.018.

Martinez-Zapien, D. *et al.* (2016) 'Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53', *Nature*, 529(7587), pp. 541–545. Available at: https://doi.org/10.1038/nature16481.

Matsushita, M. *et al.* (2017) 'Higher Expression of Activation-induced Cytidine Deaminase Is Significantly Associated with Merkel Cell Polyomavirus-negative Merkel Cell Carcinomas', *Yonago Acta Medica*, 60(3), pp. 145–153.

Matsushita, M. *et al.* (2019) 'Decreased H3K27me3 Expression Is Associated With Merkel Cell Polyomavirus-negative Merkel Cell Carcinoma, Especially Combined With Cutaneous Squamous Cell Carcinoma', *Anticancer Research*, 39(10), pp. 5573–5579. Available at: https://doi.org/10.21873/anticanres.13751.

Merlet, J. *et al.* (2009) 'Regulation of cullin-RING E3 ubiquitin-ligases by neddylation and dimerization', *Cellular and Molecular Life Sciences*, 66(11–12), pp. 1924–1938. Available at: https://doi.org/10.1007/s00018-009-8712-7.

Mileo, A.M. *et al.* (2015) 'Hepatitis C virus core protein modulates pRb2/p130 expression in human hepatocellular carcinoma cell lines through promoter methylation', *Journal of Experimental & Clinical Cancer Research*, 34(1), p. 140. Available at: https://doi.org/10.1186/s13046-015-0255-1.

Min, S.-H. *et al.* (2012) 'Negative Regulation of the Stability and Tumor Suppressor Function of Fbw7 by the Pin1 Prolyl Isomerase', 46, pp. 771–783. Available at: https://doi.org/10.1016/j.molcel.2012.04.012.

Moll, I., Zieger, W. and Schmelz, M. (1996) 'Proliferative merkel cells were not detected in human skin', *Archives of Dermatological Research*, 288(4), pp. 184–187. Available at: https://doi.org/10.1007/BF02505222.

Moll, R. *et al.* (1992) 'Cytokeratin 20 in human carcinomas. A new histodiagnostic marker detected by monoclonal antibodies.', *The American Journal of Pathology*, 140(2), p. 427.

Morrison, K.M. *et al.* (2009) 'Mammalian Merkel cells are descended from the epidermal lineage', *Developmental Biology*, 336(1), pp. 76–83. Available at: https://doi.org/10.1016/j.ydbio.2009.09.032.

Munde, P.B. *et al.* (2013) 'Pathophysiology of merkel cell', *Journal of Oral and Maxillofacial Pathology : JOMFP*, 17(3), pp. 408–412. Available at: https://doi.org/10.4103/0973-029X.125208.

Murakami, I. *et al.* (2014) 'Immunoglobulin expressions are only associated with MCPyV-positive Merkel cell carcinomas but not with MCPyV-negative ones: comparison of prognosis', *The American Journal of Surgical Pathology*, 38(12), pp. 1627–1635. Available at: https://doi.org/10.1097/PAS.0000000000000279.

Nghiem, P.T. *et al.* (2016) 'PD-1 Blockade with Pembrolizumab in Advanced Merkel-Cell Carcinoma', *The New England Journal of Medicine*, 374(26), pp. 2542–2552. Available at: https://doi.org/10.1056/NEJMoa1603702.

Nirenberg, A. *et al.* (2020) 'Merkel cell carcinoma update: the case for two tumours', *Journal of the European Academy of Dermatology and Venereology*, 34(7), pp. 1425–1431. Available at: https://doi.org/10.1111/jdv.16158.

O'Brien, J. *et al.* (2018) 'Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation', *Frontiers in Endocrinology*, 9, p. 402. Available at: https://doi.org/10.3389/fendo.2018.00402.

Olsen, C.M., Pandeya, N. and Whiteman, D.C. (2021) 'International Increases in Merkel Cell Carcinoma Incidence Rates between 1997 and 2016', *The Journal of investigative dermatology*, 141(11), pp. 2596-2601.e1. Available at: https://doi.org/10.1016/J.JID.2021.04.007.

Ortonne, J.P. *et al.* (1988) 'Normal Merkel cells express a synaptophysin-like immunoreactivity', *Dermatologica*, 177(1), pp. 1–10. Available at: https://doi.org/10.1159/000248491.

Ostrowski, S.M. *et al.* (2015) 'Ectopic Atoh1 expression drives Merkel cell production in embryonic, postnatal and adult mouse epidermis', *Development*, 142(14), pp. 2533–2544. Available at: https://doi.org/10.1242/dev.123141.

Park, D.E. *et al.* (2018) 'Dual inhibition of MDM2 and MDM4 in virus-positive Merkel cell carcinoma enhances the p53 response'. Available at: https://doi.org/10.1073/pnas.1818798116.

Park, D.E. *et al.* (2020) 'Merkel cell polyomavirus activates LSD1-mediated blockade of non-canonical BAF to regulate transformation and tumorigenesis', *Nature Cell Biology*, 22(5), pp. 603–615. Available at: https://doi.org/10.1038/s41556-020-0503-2.

Parkin, D.M. (2006) 'The global health burden of infection-associated cancers in the year 2002', *International Journal of Cancer*, 118(12), pp. 3030–3044. Available at: https://doi.org/10.1002/ijc.21731.

Pastrana, D.V. *et al.* (2009) 'Quantitation of human seroresponsiveness to Merkel cell polyomavirus', *PLoS pathogens*, 5(9), p. e1000578. Available at: https://doi.org/10.1371/journal.ppat.1000578.

Paulson, K.G. *et al.* (2010) 'Antibodies to merkel cell polyomavirus T antigen oncoproteins reflect tumor burden in merkel cell carcinoma patients', *Cancer Research*, 70(21), pp. 8388–8397. Available at: https://doi.org/10.1158/0008-5472.CAN-10-2128.

Paulson, K.G. *et al.* (2014) 'Downregulation of MHC-I expression is prevalent but reversible in Merkel cell carcinoma', *Cancer Immunology Research*, 2(11), pp. 1071–1079. Available at: https://doi.org/10.1158/2326-6066.CIR-14-0005.

Pei, D., Zhang, Y. and Zheng, J. (2012) 'Regulation of p53: a collaboration between Mdm2 and Mdmx', *Oncotarget*, 3(3), pp. 228–235. Available at: https://doi.org/10.18632/oncotarget.443.

Perez, M.C. *et al.* (2018) 'Resection Margins in Merkel Cell Carcinoma: Is a 1-cm Margin Wide Enough?', *Annals of Surgical Oncology*, 25(11), pp. 3334–3340. Available at: https://doi.org/10.1245/s10434-018-6688-y.

Petermann, E. *et al.* (2010) 'Hydroxyurea-stalled replication forks become progressively inactivated and require two different RAD51-mediated pathways for restart and repair', *Molecular Cell*, 37(4), pp. 492–502. Available at: https://doi.org/10.1016/j.molcel.2010.01.021.

Piersimoni, L. *et al.* (2022) 'Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein–Protein Interactions–A Method for All Seasons', *Chemical Reviews*, 122(8), pp. 7500–7531. Available at: https://doi.org/10.1021/acs.chemrev.1c00786.

Pietropaolo, V., Prezioso, C. and Moens, U. (2020) 'Merkel cell polyomavirus and merkel cell carcinoma', *Cancers*, 12(7), pp. 1–38. Available at: https://doi.org/10.3390/cancers12071774.

Ping, Q. *et al.* (2021) 'Cancer-associated fibroblasts: overview, progress, challenges, and directions', *Cancer Gene Therapy*, 28(9), pp. 984–999. Available at: https://doi.org/10.1038/s41417-021-00318-4.

Poiesz, B.J. *et al.* (1980) 'Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma', *Proceedings of the National Academy of Sciences of the United States of America*, 77(12), pp. 7415–7419. Available at: https://doi.org/10.1073/pnas.77.12.7415.

Poulard, C. *et al.* (2021) 'Structure, Activity, and Function of the Protein Lysine Methyltransferase G9a', *Life*, 11(10), p. 1082. Available at: https://doi.org/10.3390/life11101082.

Puig, O. *et al.* (2001) 'The tandem affinity purification (TAP) method: A general procedure of protein complex purification', *Methods*, 24(3), pp. 218–229. Available at: https://doi.org/10.1006/meth.2001.1183.

Qin, W. *et al.* (2021) 'Deciphering molecular interactions by proximity labeling', *Nature Methods*, 18(2), pp. 133–143. Available at: https://doi.org/10.1038/s41592-020-01010-5.

Rao, V.K. *et al.* (2016) 'G9a promotes proliferation and inhibits cell cycle exit during myogenic differentiation', *Nucleic Acids Research*, 44(17), pp. 8129–8143. Available at: https://doi.org/10.1093/nar/gkw483.

Rhee, H.-W. *et al.* (2013) 'Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging', *Science*, 339(6125), pp. 1328–1331. Available at: https://doi.org/10.1126/science.1230593.

Rous, P. (1911) 'A SARCOMA OF THE FOWL TRANSMISSIBLE BY AN AGENT SEPARABLE FROM THE TUMOR CELLS', *The Journal of Experimental Medicine*, 13(4), pp. 397–411.

Roux, K.J. *et al.* (2012) 'A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells', *Journal of Cell Biology*, 196(6), pp. 801–810. Available at: https://doi.org/10.1083/jcb.201112098.

Rozenblatt-Rosen, O. *et al.* (2012) 'Interpreting cancer genomes using systematic host perturbations by tumour virus proteins', *Nature*, 487(7408), pp. 491–495. Available at: https://doi.org/10.1038/nature11288.

Sablina, A.A. and Hahn, W.C. (2008) 'SV40 small T antigen and PP2A phosphatase in cell transformation', *Cancer and Metastasis Reviews*, 27(2), pp. 137–146. Available at: https://doi.org/10.1007/s10555-008-9116-0.

Saha, A. and Deshaies, R.J. (2008) 'Multimodal Activation of the Ubiquitin Ligase SCF by Nedd8 Conjugation', *Molecular Cell*, 32(1), pp. 21–31. Available at: https://doi.org/10.1016/j.molcel.2008.08.021.

Sauer, C.M. *et al.* (2017) 'Reviewing the current evidence supporting early B-cells as the cellular origin of Merkel cell carcinoma', *Critical Reviews in Oncology/Hematology*, 116, pp. 99–105. Available at: https://doi.org/10.1016/j.critrevonc.2017.05.009.

Schadendorf, D. *et al.* (2017) 'Merkel cell carcinoma: Epidemiology, prognosis, therapy and unmet medical needs', *European Journal of Cancer*, 71, pp. 53–69. Available at: https://doi.org/10.1016/j.ejca.2016.10.022.

Scheffner, M., Nuber, U. and Huibregtse, J.M. (1995) 'Protein ubiquitination involving an E1–E2–E3 enzyme ubiquitin thioester cascade', *Nature*, 373(6509), pp. 81–83. Available at: https://doi.org/10.1038/373081a0.

Scholz, T., Rippmann, V. and Spilker, G. (2004) '[Merkel cell carcinoma -- a case report with regard to the current treatment concepts]', *Handchirurgie, Mikrochirurgie, plastische Chirurgie : Organ der Deutschsprachigen Arbeitsgemeinschaft fur Handchirurgie : Organ der Deutschsprachigen Arbeitsgemeinschaft fur Mikrochirurgie der Peripheren Nerven und Gefasse : Organ der V...*, 36(5), pp. 318–322. Available at: https://doi.org/10.1055/S-2004-821184.

Schopp, I.M. *et al.* (2017) 'Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes', *Nature Communications*, 8, p. 15690. Available at: https://doi.org/10.1038/ncomms15690.

Schowalter, R.M. and Buck, C.B. (2013) 'The Merkel cell polyomavirus minor capsid protein', *PLoS pathogens*, 9(8), p. e1003558. Available at: https://doi.org/10.1371/journal.ppat.1003558.

Schowalter, R.M., Pastrana, D.V. and Buck, C.B. (2011) 'Glycosaminoglycans and sialylated glycans sequentially facilitate Merkel cell polyomavirus infectious entry', *PLoS pathogens*, 7(7), p. e1002161. Available at: https://doi.org/10.1371/journal.ppat.1002161.

Schrama, D. *et al.* (2019) 'Characterization of six Merkel cell polyomavirus-positive Merkel cell carcinoma cell lines: Integration pattern suggest that large T antigen truncating events occur before or during integration', *International Journal of Cancer*, 145(4), pp. 1020–1032. Available at: https://doi.org/10.1002/ijc.32280.

Sekar, R.B. and Periasamy, A. (2003) 'Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations', *The Journal of Cell Biology*, 160(5), pp. 629–633. Available at: https://doi.org/10.1083/jcb.200210140.

Shao, F. *et al.* (2022) 'Silencing EGFR-upregulated expression of CD55 and CD59 activates the complement system and sensitizes lung cancer to checkpoint blockade', *Nature Cancer*, 3(10), pp. 1192–1210. Available at: https://doi.org/10.1038/s43018-022-00444-4.

Shuda, M. *et al.* (2008) 'T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus', *Proceedings of the National Academy of Sciences*, 105(42), pp. 16272–16277. Available at: https://doi.org/10.1073/pnas.0806526105.

Shuda, M. *et al.* (2011) 'Human Merkel cell polyomavirus small T antigen is an oncoprotein targeting the 4E-BP1 translation regulator', *Journal of Clinical Investigation*, 121(9), pp. 3623–3634. Available at: https://doi.org/10.1172/JCI46323.

Shuda, M. *et al.* (2015) 'CDK1 substitutes for mTOR kinase to activate mitotic cap-dependent protein translation', *Proceedings of the National Academy of Sciences of the United States of America*, 112(19), pp. 5875–5882. Available at: https://doi.org/10.1073/pnas.1505787112.

Shuda, M., Chang, Y. and Moore, P.S. (2014) 'Merkel cell polyomavirus-positive merkel cell carcinoma requires viral small T-antigen for cell proliferation', *Journal of Investigative Dermatology*, 134(5), pp. 1479–1481. Available at: https://doi.org/10.1038/jid.2013.483.

Sibley, R.K. *et al.* (1980) 'Neuroendocrine (Merkel cell) carcinoma of the skin. A histologic and ultrastructural study of two cases', *The American journal of surgical pathology*, 4(3), pp. 211–221. Available at: https://doi.org/10.1097/00000478-198006000-00001.

Sinz, A. (2017) 'Divide and conquer: cleavable cross-linkers to study protein conformation and protein-protein interactions', *Analytical and Bioanalytical Chemistry*, 409(1), pp. 33–44. Available at: https://doi.org/10.1007/s00216-016-9941-x.

Sirbu, B.M. *et al.* (2013) 'Identification of proteins at active, stalled, and collapsed replication forks using isolation of proteins on nascent DNA (iPOND) coupled with mass spectrometry', *The Journal of Biological Chemistry*, 288(44), pp. 31458–31467. Available at: https://doi.org/10.1074/jbc.M113.511337.

Siriwardena, S.U., Chen, K. and Bhagwat, A.S. (2016) 'Functions and Malfunctions of Mammalian DNA-Cytosine Deaminases', *Chemical Reviews*, 116(20), pp. 12688–12710. Available at: https://doi.org/10.1021/acs.chemrev.6b00296.

Söderberg, O. *et al.* (2006) 'Direct observation of individual endogenous protein complexes in situ by proximity ligation', *Nature Methods*, 3(12), pp. 995–1000. Available at: https://doi.org/10.1038/nmeth947.

Söderberg, O. *et al.* (2008) 'Characterizing proteins and their interactions in cells and tissues using the in situ proximity ligation assay', *Methods*, 45(3), pp. 227–232. Available at: https://doi.org/10.1016/j.ymeth.2008.06.014.

Soikkeli, A.I. *et al.* (2022) 'Oncogenic Merkel cell polyomavirus T antigen truncating mutations are mediated by APOBEC3 activity in Merkel cell carcinoma', *Cancer Research Communications* [Preprint]. Available at: https://doi.org/10.1158/2767-9764.CRC-22-0211.

Soliman, S.H.A., Orlacchio, A. and Verginelli, F. (2021) 'Viral Manipulation of the Host Epigenome as a Driver of Virus-Induced Oncogenesis', *Microorganisms*, 9(6), p. 1179. Available at: https://doi.org/10.3390/microorganisms9061179.

Spurgeon, M.E. and Lambert, P.F. (2013) 'Merkel Cell Polyomavirus: A Newly Discovered Human Virus with Oncogenic Potential', *Virology*, 435(1), pp. 118–130. Available at: https://doi.org/10.1016/j.virol.2012.09.029.

Srivastava, M. *et al.* (2018) 'Replisome Dynamics and Their Functional Relevance upon DNA Damage through the PCNA Interactome', *Cell reports*, 25(13), p. 3869. Available at: https://doi.org/10.1016/J.CELREP.2018.11.099.

Sugiyama, T. and Chen, Y. (2019) 'Biochemical reconstitution of UV-induced mutational processes', *Nucleic Acids Research*, 47(13), pp. 6769–6782. Available at: https://doi.org/10.1093/nar/gkz335.

Sullivan, C.S., Cantalupo, P. and Pipas, J.M. (2000) 'The molecular chaperone activity of simian virus 40 large T antigen is required to disrupt Rb-E2F family complexes by an ATP-dependent mechanism', *Molecular and Cellular Biology*, 20(17), pp. 6233–6243. Available at: https://doi.org/10.1128/MCB.20.17.6233-6243.2000.

Sunshine, J.C. *et al.* (2018) 'Are there multiple cells of origin of Merkel cell carcinoma?', *Oncogene*, 37(11), pp. 1409–1416. Available at: https://doi.org/10.1038/s41388-017-0073-3.

Suttitheptumrong, A. *et al.* (2021) 'Altered Moesin and Actin Cytoskeleton Protein Rearrangements Affect Transendothelial Permeability in Human Endothelial Cells upon Dengue Virus Infection and TNF-α Treatment', *Viruses*, 13(10), p. 2042. Available at: https://doi.org/10.3390/v13102042.

T, H. *et al.* (1999) 'Covalent modification of all members of human cullin family proteins by NEDD8', *Oncogene*, 18(48), pp. 6829–6834. Available at: https://doi.org/10.1038/SJ.ONC.1203093.

Tachibana, M. *et al.* (2002) 'G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis', *Genes & Development*, 16(14), pp. 1779–1791. Available at: https://doi.org/10.1101/gad.989402.

Tai, Y.T. *et al.* (2000) 'Ku86 variant expression and function in multiple myeloma cells is associated with increased sensitivity to DNA damage', *Journal of Immunology (Baltimore, Md.: 1950)*, 165(11), pp. 6347–6355. Available at: https://doi.org/10.4049/jimmunol.165.11.6347.

Tang, X. *et al.* (2005) 'Mass Spectrometry Identifiable Cross-Linking Strategy for Studying Protein−Protein Interactions', *Analytical Chemistry*, 77(1), pp. 311–318. Available at: https://doi.org/10.1021/ac0488762.

de-Thé, G. *et al.* (1978) 'Epidemiological evidence for causal relationship between Epstein-Barr virus and Burkitt's lymphoma from Ugandan prospective study', *Nature*, 274(5673), pp. 756–761. Available at: https://doi.org/10.1038/274756a0.

Theiss, J.M. *et al.* (2015) 'A Comprehensive Analysis of Replicating Merkel Cell Polyomavirus Genomes Delineates the Viral Transcription Program and Suggests a Role for mcv-miR-M1 in Episomal Persistence', *PLOS Pathogens*, 11(7), p. e1004974. Available at: https://doi.org/10.1371/journal.ppat.1004974.

Toker, C. (1972) 'Trabecular carcinoma of the skin.', *Archives of dermatology*, 105(1), pp. 107–110. Available at: https://doi.org/10.1001/ARCHDERM.1972.01620040075020.

Tolstov, Y.L. *et al.* (2009) 'Human Merkel cell polyomavirus infection II. MCV is a common human infection that can be detected by conformational capsid epitope immunoassays', *International Journal of Cancer*, 125(6), pp. 1250–1256. Available at: https://doi.org/10.1002/ijc.24509.

Tourette, C. *et al.* (2014) 'A large scale Huntingtin protein interaction network implicates Rho GTPase signaling pathways in Huntington disease', *The Journal of Biological Chemistry*, 289(10), pp. 6709–6726. Available at: https://doi.org/10.1074/jbc.M113.523696.

Trapp-Fragnet, L. *et al.* (2014) 'Cell Cycle Modulation by Marek's Disease Virus: The Tegument Protein VP22 Triggers S-Phase Arrest and DNA Damage in Proliferating Cells', *PLOS ONE*, 9(6), p. e100004. Available at: https://doi.org/10.1371/journal.pone.0100004.

Tsuchida, N., Murugan, A.K. and Grieco, M. (2016) 'Kirsten Ras* oncogene: Significance of its discovery in human cancer research ras for viral origin, RAS for human origin and Ras for both viral and human origins', *Oncotarget*, 7(29), pp. 46717–46733.

Tumini, E. *et al.* (2016) 'Roles of human POLD1 and POLD3 in genome stability', *Scientific Reports*, 6(1), p. 38873. Available at: https://doi.org/10.1038/srep38873.

Upadhyay, G. *et al.* (2014) 'Antagonistic actions of Rcor proteins regulate LSD1 activity and cellular differentiation', *Proceedings of the National Academy of Sciences*, 111(22), pp. 8071–8076. Available at: https://doi.org/10.1073/pnas.1404292111.

Van Keymeulen, A. *et al.* (2009) 'Epidermal progenitors give rise to Merkel cells during embryonic development and adult homeostasis', *Journal of Cell Biology*, 187(1), pp. 91–100. Available at: https://doi.org/10.1083/jcb.200907080.

Verhaegen, M.E. *et al.* (2015) 'Merkel cell polyomavirus small T antigen is oncogenic in transgenic mice', *The Journal of Investigative Dermatology*, 135(5), pp. 1415–1424. Available at: https://doi.org/10.1038/jid.2014.446.

Verhaegen, M.E. *et al.* (2017) 'Merkel Cell Polyomavirus Small T Antigen Initiates Merkel Cell Carcinoma-like Tumor Development in Mice', *Cancer Research*, 77(12), pp. 3151–3157. Available at: https://doi.org/10.1158/0008-5472.CAN-17-0035.

Verheyden, S. *et al.* (2009) 'Role of the inhibitory KIR ligand HLA-Bw4 and HLA-C expression levels in the recognition of leukemic cells by Natural Killer cells', *Cancer Immunology, Immunotherapy*, 58(6), pp. 855–865. Available at: https://doi.org/10.1007/s00262-008-0601-7.

Walker, J.W. *et al.* (2020) 'Efficacy and safety of avelumab treatment in patients with metastatic Merkel cell carcinoma: experience from a global expanded access program', *Journal for Immunotherapy of Cancer*, 8(1). Available at: https://doi.org/10.1136/JITC-2019-000313.

Waltari, M. *et al.* (2011) 'Association of Merkel cell polyomavirus infection with tumor p53, KIT, stem cell factor, PDGFR-alpha and survival in Merkel cell carcinoma', *International Journal of Cancer*, 129(3), pp. 619–628. Available at: https://doi.org/10.1002/ijc.25720.

Wang, X. *et al.* (2012) 'Bromodomain protein Brd4 plays a key role in Merkel cell polyomavirus DNA replication', *PLoS pathogens*, 8(11), p. e1003021. Available at: https://doi.org/10.1371/journal.ppat.1003021.

Watanabe, S. *et al.* (2018) 'MDC1 methylation mediated by lysine methyltransferases EHMT1 and EHMT2 regulates active ATM accumulation flanking DNA damage sites', *Scientific Reports*, 8(1). Available at: https://doi.org/10.1038/S41598-018-29239-3.

Wijaya, W.A. *et al.* (2022) 'Prevalence of Merkel Cell Polyomavirus in Normal and Lesional Skin: A Systematic Review and Meta-Analysis', *Frontiers in Oncology*, 12. Available at: https://doi.org/10.3389/fonc.2022.868781.

Wu, G. *et al.* (2003) 'Structure of a β-TrCP1-Skp1-β-Catenin Complex: Destruction Motif Binding and Lysine Specificity of the SCFβ-TrCP1 Ubiquitin Ligase', *Molecular Cell*, 11(6), pp. 1445–1456. Available at: https://doi.org/10.1016/S1097-2765(03)00234-X.

Xu, L. *et al.* (2019) 'Pharmacological inhibition of EZH2 combined with DNA-damaging agents interferes with the DNA damage response in MM cells', *Molecular Medicine Reports*, 19(5), pp. 4249–4255. Available at: https://doi.org/10.3892/mmr.2019.10075.

Xu, Y., Piston, D.W. and Johnson, C.H. (1999) 'A bioluminescence resonance energy transfer (BRET) system: Application to interacting circadian clock proteins', *Proceedings of the National Academy of Sciences*, 96(1), pp. 151–156. Available at: https://doi.org/10.1073/pnas.96.1.151.

Yachie, N. *et al.* (2016) 'Pooled-matrix protein interaction screens using Barcode Fusion Genetics', *Molecular Systems Biology*, 12(4), p. 863. Available at: https://doi.org/10.15252/msb.20156660.

Yang, Q. *et al.* (2017) 'G9a coordinates with the RPA complex to promote DNA damage repair and cell survival', *Proceedings of the National Academy of Sciences*, 114(30), pp. E6054–E6063. Available at: https://doi.org/10.1073/pnas.1700694114.

Yiengpruksawan, A. *et al.* (1991) 'Merkel cell carcinoma. Prognosis and management', *Archives of surgery (Chicago, Ill. : 1960)*, 126(12), pp. 1514–1519. Available at: https://doi.org/10.1001/ARCHSURG.1991.01410360088014.

Zhang, R. *et al.* (2018) 'CD59: a promising target for tumor immunotherapy', *Future Oncology*, 14(8), pp. 781–791. Available at: https://doi.org/10.2217/fon-2017-0498.

Zheng, Z. *et al.* (2022) 'Implication of IL6-positive Cancer-associated Fibroblasts in Merkel Cell Carcinoma Pathogenesis: A Possible Modulator of Immune Microenvironment', *Anticancer Research*, 42(9), pp. 4359–4369. Available at: https://doi.org/10.21873/anticanres.15936.

Zhou, A. *et al.* (2008) 'Novel splicing factor RBM25 modulates Bcl-x pre-mRNA 5' splice site selection', *Molecular and Cellular Biology*, 28(19), pp. 5924–5936. Available at: https://doi.org/10.1128/MCB.00560-08.

Zhou, X. and Sun, S.-C. (2021) 'Targeting ubiquitin signaling for cancer immunotherapy', *Signal Transduction and Targeted Therapy*, 6(1), pp. 1–15. Available at: https://doi.org/10.1038/s41392-020-00421-2.

Zwijnenburg, E.M. *et al.* (2021) 'Merkel Cell Carcinoma: New Trends', *Cancers*, 13(7), p. 1614. Available at: https://doi.org/10.3390/cancers13071614.

7 FORMATIONS ET PRODUCTIONS SCIENTIFIQUES

7.1 FORMATIONS

2014 : **Baccalauréat Scientifique Spécialité Mathématique mention européenne.**
Lycée Arthur Rimbaud, Sin le Noble.

2014-2016 : Première année commune aux études de santé.
Faculté de Médecine Henri Warembourg, Université de Lille.

2**018** : **Licence de Biochimie**.
Université Claude Bernard Lyon I.

2020 : Master Biotechnologie parcours Protéomique.
Université de Lille.

7.2 PRODUCTIONS SCIENTIFIQUE

7.2.1 Présentations orales

Journées Française de Spectrométrie de masse (14-24 juin 2021)
Organisé par la société française de spectrométrie de masse (SFSM) – en ligne.

13émes journées de l'AFIPP (30 septembre – 01 octobre 2021)
Organisé par l'association francophone de la recherche sur les infections Papillomavirus et Polyomavirus (AFIPP) – en ligne.

Journées de club jeune de la FPS (25-27 avril 2022)
Organisé par le club jeune de la French Proteomic Society – Bordeaux, France.

Analytics 2022 (05-08 septembre 2022)
Organisé par Réseau Francophone de Métabolomique et de Fluxomique, de la Société Française de Protéomique, de la Société Française de Spectrométrie de Masse et de l'Association Francophone des Sciences Séparatives – Nantes, France.

14èmes journées de l'AFIPP (28-30 septembre 2022)
Organisé par l'association francophone de la recherche sur les infections Papillomavirus et Polyomavirus (AFIPP) – Sedan, France.

Journée André Verbert (15 novembre 2022)
Organisé par l'école doctorale Biologie-Santé de Lille – Lille, France.

7.2.2 Présentations Poster

Journée SFR-TSM (22 mars 2022)
Organisé par la Structure Fédérative de Recherche – Technologies pour la Santé et Médicament – Lille, France

ONCOlille days (02-04 novembre 2022)
Organisé par l'institut de recherche ONCOlille – Lille, France.

➔ **Prix de meilleure présentation poster**

7.2.3 Publications

Borna Disease Virus 1 Phosphoprotein Forms a Tetramer and Interacts with Host Factors Involved in DNA Double-Strand Break Repair and mRNA Processing. Tarbouriech *et al*., 2022. (auteur associé)

Human Papillomavirus 42 Drives Digital Papillary Adenocarcinoma and Elicits a Germ Cell–like Program Conserved in HPV-Positive Cancers. Leiendecker *et al*., 2023. (auteur associé)

Distinct Regulation of EZH2 and its Repressive H3K27me3 Mark in Polyomavirus-Positive and -Negative Merkel Cell Carcinoma. Durand *et al*., 2023 (auteur associé)

DNA damage stress control is a tLT- and EHMT2-dependent central feature of Merkel Cell Carcinoma. Bachiri *et al*. en révision dans *Journal of Investigative Dermatology* au 14/03/2023 (premier auteur)

8 ANNEXES

ANNEXE I : *Les valeurs affichées correspondent au Log2 fold change des valeurs de LFQ mesurées. Seules les valeurs obtenues pour les interacteurs de haute confiance sont affichées (Log2fc > 1 ; q-value <0,01). La mention "Exclu" correspond aux protéines absentes des contrôles et identifiées dans chacun des trois réplicats.*

Uniprot ID	Gene names	Protein names	sT		LT275		LT330		LT428		LT
			N-ter	C-ter	N-ter	C-ter	N-ter	C-ter	N-ter	C-ter	N-ter
Q9NRG9	**AAAS**	*Aladin*	1,02								
Q9NY61	**AATF**	*Protein AATF*						1,31		2,29	
P24752	**ACAT1**	*Acetyl-CoA acetyltransferase, mitochondrial*									1,39
Q9H2P0	**ADNP**	*Activity-dependent neuroprotector homeobox protein*					1,04	1,50		1,98	
Q9UHB7	**AFF4**	*AF4/FMR2 family member 4*						1,99		1,46	
Q8WYP5	**AHCTF1**	*Protein ELYS*						1,51		1,79	
O00170	**AIP**	*AH receptor-interacting protein*	Exclu								
P54819	**AK2**	*Adenylate kinase 2, mitochondrial*	Exclu								Exclu
Q02040	**AKAP17A**	*A-kinase anchor protein 17A*					Exclu				
Q9BTT0	**ANP32E**	*Acidic leucine-rich nuclear phosphoprotein 32 family member E*								2,97	
O95782	**AP2A1**	*AP-2 complex subunit alpha-1*	1,87								1,68
Q96CW1	**AP2M1**	*AP-2 complex subunit mu*									2,22
Q8N6T3	**ARFGAP1**	*ADP-ribosylation factor GTPase-activating protein 1*		1,36							
P53367	**ARFIP1**	*Arfaptin-1*	3,61								
Q9NWB6	**ARGLU1**	*Arginine and glutamate-rich protein 1*						Exclu			
O14497	**ARID1A**	*AT-rich interactive domain-containing protein 1A*					1,74	1,77			
Q8NFD5	**ARID1B**	*AT-rich interactive domain-containing protein 1B*						Exclu			
Q68CP9	**ARID2**	*AT-rich interactive domain-containing protein 2*						1,87			
Q99856	**ARID3A**	*AT-rich interactive domain-containing protein 3A*						Exclu			
Q8IVW6	**ARID3B**	*AT-rich interactive domain-containing protein 3B*					1,95	1,86	1,08	2,01	
Q6VMQ6	**ATF7IP**	*Activating transcription factor 7-interacting protein 1*					1,00	1,57		1,40	
P24539	**ATP5F1**	*ATP synthase F(0) complex subunit B1, mitochondrial*									Exclu
P46100	**ATRX**	*Transcriptional regulator ATRX*						1,02		1,81	
Q9UBB4	**ATXN10**	*Ataxin-10*	1,62								1,26
Q96GD4	**AURKB**	*Aurora kinase B*		1,11							
Q99933	**BAG1**	*BAG family molecular chaperone regulator 1*							Exclu		
O95816	**BAG2**	*BAG family molecular chaperone regulator 2*	5,75		5,34		4,11	2,61	4,66		6,26
O95817	**BAG3**	*BAG family molecular chaperone regulator 3*	4,52		3,30		3,31		3,14		2,96
O95429	**BAG4**	*BAG family molecular chaperone regulator 4*	Exclu				Exclu		Exclu		Exclu
Q9UL15	**BAG5**	*BAG family molecular chaperone regulator 5*	Exclu		Exclu		Exclu		Exclu		Exclu
Q9UHR4	**BAIAP2L1**	*Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1*		1,11							
Q92560	**BAP1**	*Ubiquitin carboxyl-terminal hydrolase BAP1*					2,10				
Q8IXM2	**BAP18**	*Chromatin complexes subunit BAP18*						1,09		1,08	
Q9NRL2	**BAZ1A**	*Bromodomain adjacent to zinc finger domain protein 1A*								1,06	
Q9UIG0	**BAZ1B**	*Tyrosine-protein kinase BAZ1B*								1,94	
Q9UIF9	**BAZ2A**	*Bromodomain adjacent to zinc finger domain protein 2A*						1,15			
O75934	**BCAS2**	*Pre-mRNA-splicing factor SPF27*					1,22	1,33			
Q8WUZ0	**BCL7C**	*B-cell CLL/lymphoma 7 protein family member C*							Exclu		
Q9NYF8	**BCLAF1**	*Bcl-2-associated transcription factor 1*						1,21		1,02	
Q6W2J9	**BCOR**	*BCL-6 corepressor*					1,79	2,51		1,26	
Q5H9F3	**BCORL1**	*BCL-6 corepressor-like protein 1*					1,96		2,46		
Q5T5X7	**BEND3**	*BEN domain-containing protein 3*					Exclu	Exclu		Exclu	
Q14692	**BMS1**	*Ribosome biogenesis protein BMS1 homolog*								1,38	
Q14137	**BOP1**	*Ribosome biogenesis protein BOP1*						1,43		1,64	
Q12830	**BPTF**	*Nucleosome-remodeling factor subunit BPTF*					1,68	1,32		1,09	
P25440	**BRD2**	*Bromodomain-containing protein 2*								1,91	
Q15059	**BRD3**	*Bromodomain-containing protein 3*								1,76	
O60885	**BRD4**	*Bromodomain-containing protein 4*						1,13			
Q9BX63	**BRIP1**	*Fanconi anemia group J protein*						1,39			
Q53F19	**C17orf85**	*Uncharacterized protein C17orf85*					2,69	2,78		2,93	
Q9BQ61	**C19orf43**	*Uncharacterized protein C19orf43*								1,03	
P57076	**C21orf59**	*UPF0769 protein C21orf59*	1,27								
Q8N0T1	**C8orf59**	*Uncharacterized protein C8orf59*									2,63
Q9H8G2	**CAAP1**	*Caspase activity and apoptosis inhibitor 1*					Exclu			Exclu	
Q8WUQ7	**CACTIN**	*Cactin*					1,86	1,78	1,26	1,68	
Q9HB71	**CACYBP**	*Calcyclin-binding protein*	1,27								1,33
Q9P1Y5	**CAMSAP3**	*Calmodulin-regulated spectrin-associated protein 3*		3,15							
Q75N03	**CBLL1**	*E3 ubiquitin-protein ligase Hakai*						2,40			
P83916	**CBX1**	*Chromobox protein homolog 1*						1,11		1,61	
Q14781	**CBX2**	*Chromobox protein homolog 2*					2,29			2,55	
Q9HC52	**CBX8**	*Chromobox protein homolog 8*						1,01		1,31	
Q8IX12	**CCAR1**	*Cell division cycle and apoptosis regulator protein 1*					1,38				
Q8N163	**CCAR2**	*Cell cycle and apoptosis regulator protein 2*					1,85	1,30			
Q8IWD4	**CCDC117**	*Coiled-coil domain-containing protein 117*	Exclu		Exclu		Exclu		Exclu		
Q6PII3	**CCDC174**	*Coiled-coil domain-containing protein 174*					2,51	2,42		2,46	

Uniprot ID	Gene names	Protein names	sT		LT275		LT330		LT428		LT
			N-ter	C-ter	N-ter	C-ter	N-ter	C-ter	N-ter	C-ter	N-ter
Q96A33	**CCDC47**	*Coiled-coil domain-containing protein 47*				1,10					
Q16204	**CCDC6**	*Coiled-coil domain-containing protein 6*		2,68							
P14635	**CCNB1**	*G2/mitotic-specific cyclin-B1*		1,22							
O75909	**CCNK**	*Cyclin-K*						1,54			
Q96S94	**CCNL2**	*Cyclin-L2*						Exclu			
O60563	**CCNT1**	*Cyclin-T1*					1,81	2,00		1,36	
P78371	**CCT2**	*T-complex protein 1 subunit beta*	1,00	1,01							
P49368	**CCT3**	*T-complex protein 1 subunit gamma*		1,03							
O15446	**CD3EAP**	*DNA-directed RNA polymerase I subunit RPA34*								1,22	
Q16543	**CDC37**	*Hsp90 co-chaperone Cdc37*	3,52				1,51				2,13
O60508	**CDC40**	*Pre-mRNA-processing factor 17*								Exclu	
Q69YH5	**CDCA2**	*Cell division cycle-associated protein 2*						3,59		3,81	
Q9NYV4	**CDK12**	*Cyclin-dependent kinase 12*					1,76	1,24			
Q14004	**CDK13**	*Cyclin-dependent kinase 13*					Exclu				
P50750	**CDK9**	*Cyclin-dependent kinase 9*						1,14			
Q9NXV6	**CDKN2AIP**	*CDKN2A-interacting protein*						1,04			
Q03188	**CENPC**	*Centromere protein C*								1,35	
Q9H3R5	**CENPH**	*Centromere protein H*							2,82		
Q9BS16	**CENPK**	*Centromere protein K*					Exclu		Exclu		
Q96BT3	**CENPT**	*Centromere protein T*					Exclu	Exclu		Exclu	
O15182	**CETN3**	*Centrin-3*	Exclu								Exclu
Q96JM3	**CHAMP1**	*Chromosome alignment-maintaining phosphoprotein 1*					1,04				
Q12873	**CHD3**	*Chromodomain-helicase-DNA-binding protein 3*						Exclu		Exclu	
Q14839	**CHD4**	*Chromodomain-helicase-DNA-binding protein 4*								1,10	
Q9P2D1	**CHD7**	*Chromodomain-helicase-DNA-binding protein 7*					Exclu	Exclu		Exclu	
Q9HCK8	**CHD8**	*Chromodomain-helicase-DNA-binding protein 8*					2,22	1,88	1,87	1,40	
Q8IWX8	**CHERP**	*Calcium homeostasis endoplasmic reticulum protein*					1,02	1,52			
P0CG12	**CHTF8**	*Chromosome transmission fidelity protein 8 homolog isoform 2*		Exclu							Exclu
Q96RK0	**CIC**	*Protein capicua homolog*					1,62	2,03			
Q9ULV3	**CIZ1**	*Cip1-interacting zinc finger protein*						2,34		2,19	
P54105	**CLNS1A**	*Methylosome subunit pICln*				1,26					
Q00610	**CLTC**	*Clathrin heavy chain 1*	2,45				2,12		1,39		1,78
Q8N1G2	**CMTR1**	*Cap-specific mRNA (nucleoside-2-O-)-methyltransferase 1*						2,41		1,44	
Q9P003	**CNIH4**	*Protein cornichon homolog 4*						Exclu			
Q8NE01	**CNNM3**	*Metal transporter CNNM3*		Exclu							
Q05707	**COL14A1**	*Collagen alpha-1(XIV) chain*									Exclu
P35606	**COPB2**	*Coatomer subunit beta*	1,21								
P13073	**COX4I1**	*Cytochrome c oxidase subunit 4 isoform 1, mitochondrial*									Exclu
P14406	**COX7A2**	*Cytochrome c oxidase subunit 7A2, mitochondrial*									Exclu
Q86YQ8	**CPNE8**	*Copine-8*		1,49							
Q10570	**CPSF1**	*Cleavage and polyadenylation specificity factor subunit 1*						1,14			
O95639	**CPSF4**	*Cleavage and polyadenylation specificity factor subunit 4*					1,33	2,04		1,11	
Q9BZJ0	**CRNKL1**	*Crooked neck-like protein 1*					1,39	1,92		1,35	
P33240	**CSTF2**	*Cleavage stimulation factor subunit 2*	2,91		2,44		3,92		2,70		1,52
Q9H0L4	**CSTF2T**	*Cleavage stimulation factor subunit 2 tau variant*	2,06				1,93				
P49711	**CTCF**	*Transcriptional repressor CTCF*								1,37	
P32929	**CTH**	*Cystathionine gamma-lyase*		Exclu							
Q6PD62	**CTR9**	*RNA polymerase-associated protein CTR9 homolog*						Exclu			
P39880	**CUX1**	*Homeobox protein cut-like 1*					1,29	2,24		2,28	
Q9P013	**CWC15**	*Spliceosome-associated protein CWC15 homolog*						1,01			
Q6UX04	**CWC27**	*Peptidyl-prolyl cis-trans isomerase CWC27 homolog*								1,05	
Q13206	**DDX10**	*Probable ATP-dependent RNA helicase DDX10*								1,44	
Q86XP3	**DDX42**	*ATP-dependent RNA helicase DDX42*						1,46		1,30	
O75398	**DEAF1**	*Deformed epidermal autoregulatory factor 1 homolog*								1,73	
Q96DF8	**DGCR14**	*Protein DGCR14*					1,63	1,98		1,85	
Q8WYQ5	**DGCR8**	*Microprocessor complex subunit DGCR8*						1,25			
O60231	**DHX16**	*Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX16*						1,17		1,10	
Q9H5Z1	**DHX35**	*Probable ATP-dependent RNA helicase DHX35*								1,01	
Q9BTC0	**DIDO1**	*Death-inducer obliterator 1*					1,58	1,90		1,26	
Q9P265	**DIP2B**	*Disco-interacting protein 2 homolog B*	1,48								
Q8TDM6	**DLG5**	*Disks large homolog 5*		Exclu							
P31689	**DNAJA1**	*DnaJ homolog subfamily A member 1*		1,27							
O60884	**DNAJA2**	*DnaJ homolog subfamily A member 2*		1,10							
P25685	**DNAJB1**	*DnaJ homolog subfamily B member 1*	1,00								
Q9NXW2	**DNAJB12**	*DnaJ homolog subfamily B member 12*	Exclu				Exclu				Exclu
O75190	**DNAJB6**	*DnaJ homolog subfamily B member 6*		1,12					1,36		
Q9NVH1	**DNAJC11**	*DnaJ homolog subfamily C member 11*	1,23								
Q9NVM6	**DNAJC17**	*DnaJ homolog subfamily C member 17*						Exclu			

Uniprot ID	Gene names	Protein names	sT		LT275		LT330		LT428		LT
			N-ter	C-ter	N-ter	C-ter	N-ter	C-ter	N-ter	C-ter	N-ter
Q99615	**DNAJC7**	*DnaJ homolog subfamily C member 7*	6,01		3,94		4,40		3,31		3,66
Q9BQC3	**DPH2**	*Diphthamide biosynthesis protein 2*	Exclu				Exclu				Exclu
Q9Y4J8	**DTNA**	*Dystrobrevin alpha*	Exclu								
Q9NX74	**DUS2**	*tRNA-dihydrouridine(20) synthase [NAD(P)+]-like*	4,62				7,00		7,01		6,14
P51452	**DUSP3**	*Dual specificity protein phosphatase 3*	Exclu				Exclu				
Q9NP97	**DYNLRB1**	*Dynein light chain roadblock-type 1*	6,10		4,70		4,34		3,82		4,27
O00716	**E2F3**	*Transcription factor E2F3*					1,27		1,18		
O75461	**E2F6**	*Transcription factor E2F6*								1,53	
Q13011	**ECH1**	*Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial*		2,91			1,93				2,66
O00418	**EEF2K**	*Eukaryotic elongation factor 2 kinase*		2,71							
Q9H9B1	**EHMT1**	*Histone-lysine N-methyltransferase EHMT1*						1,26		1,90	
Q96KQ7	**EHMT2**	*Histone-lysine N-methyltransferase EHMT2*						2,42		2,80	
P78344	**EIF4G2**	*Eukaryotic translation initiation factor 4 gamma 2*		1,05							
P32519	**ELF1**	*ETS-related transcription factor Elf-1*					2,00		1,97		
Q15723	**ELF2**	*ETS-related transcription factor Elf-2*						1,23		1,78	
Q7Z589	**EMSY**	*Protein EMSY*					1,87	2,55		1,47	
Q96L91	**EP400**	*E1A-binding protein p400*	4,01	1,85			3,29	2,71		1,23	
Q8IUD2	**ERC1**	*ELKS/Rab6-interacting/CAST family member 1*		2,18							
Q86X53	**ERICH1**	*Glutamate-rich protein 1*						2,21			
Q9H501	**ESF1**	*ESF1 homolog*								1,66	
Q14674	**ESPL1**	*Separin*	1,61								
O60645	**EXOC3**	*Exocyst complex component 3*	2,78								
Q01780	**EXOSC10**	*Exosome component 10*						1,07			
Q13868	**EXOSC2**	*Exosome complex component RRP4*								1,04	
Q9NQT5	**EXOSC3**	*Exosome complex component RRP40*						1,12		1,26	
Q9UNN5	**FAF1**	*FAS-associated factor 1*									Exclu
Q9UK61	**FAM208A**	*Protein FAM208A*						2,14		1,03	
Q5VWN6	**FAM208B**	*Protein FAM208B*						1,19			
Q6PIW4	**FIGNL1**	*Fidgetin-like protein 1*						Exclu			
Q6UN15	**FIP1L1**	*Pre-mRNA 3-end-processing factor FIP1*					1,10	1,88			
Q06787	**FMR1**	*Fragile X mental retardation protein 1*	2,11								
Q8N3X1	**FNBP4**	*Formin-binding protein 4*					2,10				
P85037	**FOXK1**	*Forkhead box protein K1*					Exclu				
Q01167	**FOXK2**	*Forkhead box protein K2*						1,01			
Q12778	**FOXO1**	*Forkhead box protein O1*	Exclu								
P51114	**FXR1**	*Fragile X mental retardation syndrome-related protein 1*	5,00								2,21
P51116	**FXR2**	*Fragile X mental retardation syndrome-related protein 2*	5,15								2,14
P43694	**GATA4**	*Transcription factor GATA-4*					Exclu				
Q86YP4	**GATAD2A**	*Transcriptional repressor p66-alpha*					1,51	1,85		1,92	
Q8WXI9	**GATAD2B**	*Transcriptional repressor p66-beta*					1,94	1,96	1,09	1,96	
Q9NZM4	**GLTSCR1**	*Glioma tumor suppressor candidate region gene 1 protein*						Exclu			
O75496	**GMNN**	*Geminin*					Exclu		Exclu		
Q13823	**GNL2**	*Nucleolar GTP-binding protein 2*						1,04		1,63	
Q5T3I0	**GPATCH4**	*G patch domain-containing protein 4*								1,11	
Q92917	**GPKOW**	*G patch domain and KOW motifs-containing protein*						1,34			
Q14687	**GSE1**	*Genetic suppressor element 1*					2,48	1,55			
P35269	**GTF2F1**	*General transcription factor IIF subunit 1*					1,54	1,14	1,11	1,54	
P13984	**GTF2F2**	*General transcription factor IIF subunit 2*					1,07				
P78347	**GTF2I**	*General transcription factor II-I*						1,37		1,56	
Q12789	**GTF3C1**	*General transcription factor 3C polypeptide 1*						1,03		2,34	
Q9Y5Q9	**GTF3C3**	*General transcription factor 3C polypeptide 3*					1,02				
Q9UKN8	**GTF3C4**	*General transcription factor 3C polypeptide 4*					1,36	1,90	1,06	1,82	
Q9Y5Q8	**GTF3C5**	*General transcription factor 3C polypeptide 5*						1,36		1,36	
P16104	**H2AFX**	*Histone H2AX*						1,25			
O14929	**HAT1**	*Histone acetyltransferase type B catalytic subunit*								1,15	
P51610	**HCFC1**	*Host cell factor 1*					1,80	1,11	1,18		
Q13547	**HDAC1**	*Histone deacetylase 1*						1,65		1,03	
Q92769	**HDAC2**	*Histone deacetylase 2*						1,61		1,02	
Q9NRZ9	**HELLS**	*Lymphoid-specific helicase*							1,94		
P42694	**HELZ**	*Probable helicase with zinc finger domain*	1,63	1,57							
Q9BTY7	**HGH1**	*Protein HGH1 homolog*	Exclu								
P16403	**HIST1H1C**	*Histone H1.2*	2,67	2,28		1,91	1,69	1,77		2,00	1,69
P16402	**HIST1H1D**	*Histone H1.3*		Exclu							
Q93077	**HIST1H2AC**	*Histone H2A type 1-C*	Exclu	Exclu				Exclu			Exclu
P15822	**HIVEP1**	*Zinc finger protein 40*						Exclu		Exclu	
Q9UGU5	**HMGXB4**	*HMG domain-containing protein 4*								1,77	
P07900	**HSP90AA1**	*Heat shock protein HSP 90-alpha*	1,81								
P08238	**HSP90AB1**	*Heat shock protein HSP 90-beta*	2,04								1,12

Uniprot ID	Gene names	Protein names	sT		LT275		LT330		LT428		LT
			N-ter	C-ter	N-ter	C-ter	N-ter	C-ter	N-ter	C-ter	N-ter
Q58FF8	**HSP90AB2P**	*Putative heat shock protein HSP 90-beta 2*	1,46								
P0DMV9	**HSPA1B**	*Heat shock 70 kDa protein 1B*	18,55		15,90		17,86		17,42		17,39
P34931	**HSPA1L**	*Heat shock 70 kDa protein 1-like*	Exclu								
P54652	**HSPA2**	*Heat shock-related 70 kDa protein 2*	6,35		5,69		6,64		5,70		5,86
P34932	**HSPA4**	*Heat shock 70 kDa protein 4*	3,25		2,49		2,97		2,78		1,88
O95757	**HSPA4L**	*Heat shock 70 kDa protein 4L*	1,88		1,50		1,28		1,05		1,01
P17066	**HSPA6**	*Heat shock 70 kDa protein 6*	Exclu						Exclu		Exclu
P11142	**HSPA8**	*Heat shock cognate 71 kDa protein*	18,05				16,40	1,14	3,68		3,95
Q9NZL4	**HSPBP1**	*Hsp70-binding protein 1*	Exclu		Exclu		Exclu		Exclu		Exclu
Q92598	**HSPH1**	*Heat shock protein 105 kDa*	2,40		1,90		1,76		1,07		1,30
O43719	**HTATSF1**	*HIV Tat-specific factor 1*						1,41		1,45	
P78318	**IGBP1**	*Immunoglobulin-binding protein 1*	1,26								
Q13123	**IK**	*Protein Red*						1,41		1,45	
Q9NV88	**INTS9**	*Integrator complex subunit 9*					2,50		1,46		
Q7Z5L9	**IRF2BP2**	*Interferon regulatory factor 2-binding protein 2*					1,01				
Q9Y4H2	**IRS2**	*Insulin receptor substrate 2*		3,46							
Q9ULR0	**ISY1**	*Pre-mRNA-splicing factor ISY1 homolog*						1,05		1,14	
Q15652	**JMJD1C**	*Probable JmjC domain-containing histone demethylation protein 2C*		1,05			1,04	1,53		1,11	
P05412	**JUN**	*Transcription factor AP-1*						1,06		1,39	
Q9P2N6	**KANSL3**	*KAT8 regulatory NSL complex subunit 3*					Exclu				
Q9H7Z6	**KAT8**	*Histone acetyltransferase KAT8*						Exclu			
Q86V97	**KBTBD6**	*Kelch repeat and BTB domain-containing protein 6*								1,39	
O60341	**KDM1A**	*Lysine-specific histone demethylase 1A*					1,38	2,27		1,86	
Q7LBC6	**KDM3B**	*Lysine-specific demethylase 3B*					1,80	1,49		1,01	
O15550	**KDM6A**	*Lysine-specific demethylase 6A*						Exclu			
Q96EK5	**KIAA1279**	*KIF1-binding protein*	1,29			1,22					
Q8TCG1	**KIAA1524**	*Protein CIP2A*	5,64								3,94
O95239	**KIF4A**	*Chromosome-associated kinesin KIF4A*								1,25	
Q9BW19	**KIFC1**	*Kinesin-like protein KIFC1*								1,04	
Q9H0B6	**KLC2**	*Kinesin light chain 2*	1,04								
Q03164	**KMT2A**	*Histone-lysine N-methyltransferase 2A*						1,73		1,58	
Q8N9T8	**KRI1**	*Protein KRI1 homolog*								1,68	
Q9Y2U8	**LEMD3**	*Inner nuclear membrane protein Man1*								1,06	
Q96PV6	**LENG8**	*Leukocyte receptor cluster member 8*					1,26	2,27		1,70	
Q9UPQ0	**LIMCH1**	*LIM and calponin homology domains-containing protein 1*		1,74							
Q6MZP7	**LIN54**	*Protein lin-54 homolog*					1,86	1,79	1,40	1,13	
P50851	**LRBA**	*Lipopolysaccharide-responsive and beige-like anchor protein*	1,43								
P62310	**LSM3**	*U6 snRNA-associated Sm-like protein LSm3*					Exclu	Exclu		Exclu	
Q9Y4Z0	**LSM4**	*U6 snRNA-associated Sm-like protein LSm4*						Exclu			
A0A028	**LT**	*Large T antigen MCPyV*			Exclu	Exclu	Exclu	Exclu	Exclu	Exclu	Exclu
Q92585	**MAML1**	*Mastermind-like protein 1*						Exclu			
O95983	**MBD3**	*Methyl-CpG-binding domain protein 3*						1,56		1,44	
Q9HCC0	**MCCC2**	*Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial*		1,05							1,47
O60318	**MCM3AP**	*Germinal-center associated nuclear protein*						Exclu			
Q14676	**MDC1**	*Mediator of DNA damage checkpoint protein 1*					1,28	1,55		1,85	
Q15648	**MED1**	*Mediator of RNA polymerase II transcription subunit 1*					1,33	2,72		1,76	
Q96RN5	**MED15**	*Mediator of RNA polymerase II transcription subunit 15*					Exclu				
Q86W50	**METTL16**	*Methyltransferase-like protein 16*					Exclu		Exclu		
P55081	**MFAP1**	*Microfibrillar-associated protein 1*								1,59	
Q8IWI9	**MGA**	*MAX gene-associated protein*					1,27	2,78	1,30	3,37	
O60502	**MGEA5**	*Protein O-GlcNAcase*		2,96							
Q15773	**MLF2**	*Myeloid leukemia factor 2*	Exclu		Exclu		Exclu		Exclu		Exclu
Q9Y3A3	**MOB4**	*MOB-like protein phocein*		Exclu							
Q9Y6X9	**MORC2**	*MORC family CW-type zinc finger protein 2*								1,11	
Q9BQ48	**MRPL34**	*39S ribosomal protein L34, mitochondrial*									Exclu
Q6P1R3	**MSANTD2**	*Myb/SANT-like DNA-binding domain-containing protein 2*								1,42	
Q9BUK6	**MSTO1**	*Protein misato homolog 1*	1,15								
Q13330	**MTA1**	*Metastasis-associated protein MTA1*					1,32	1,82		1,30	
O94776	**MTA2**	*Metastasis-associated protein MTA2*					1,39	1,62		1,30	
Q9H019	**MTFR1L**	*Mitochondrial fission regulator 1-like*	Exclu				Exclu				
P10244	**MYBL2**	*Myb-related protein B*						Exclu			
Q5VVJ2	**MYSM1**	*Histone H2A deubiquitinase MYSM1*								1,01	
Q6NZ67	**MZT2B**	*Mitotic-spindle organizing protein 2B*	Exclu								
Q96RE7	**NACC1**	*Nucleus accumbens-associated protein 1*					1,63		1,31	1,35	
O60934	**NBN**	*Nibrin*					1,00	1,20			
Q9BPX3	**NCAPG**	*Condensin complex subunit 3*	2,85								
Q9NZQ3	**NCKIPSD**	*NCK-interacting protein with SH3 domain*	Exclu								
Q15596	**NCOA2**	*Nuclear receptor coactivator 2*					Exclu				

Uniprot ID	Gene names	Protein names	sT		LT275		LT330		LT428		LT
			N-ter	C-ter	N-ter	C-ter	N-ter	C-ter	N-ter	C-ter	N-ter
Q9HCD5	**NCOA5**	*Nuclear receptor coactivator 5*						1,06			
Q14686	**NCOA6**	*Nuclear receptor coactivator 6*					1,75	2,93		1,43	
O75376	**NCOR1**	*Nuclear receptor corepressor 1*					2,93	1,70			
Q9Y618	**NCOR2**	*Nuclear receptor corepressor 2*					Exclu				
P19404	**NDUFV2**	*NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial*								1,23	
Q9H3P2	**NELFA**	*Negative elongation factor A*						1,34		1,34	
P18615	**NELFE**	*Negative elongation factor E*						1,67		1,45	
Q8NCF5	**NFATC2IP**	*NFATC2-interacting protein*								1,49	
Q6P4R8	**NFRKB**	*Nuclear factor related to kappa-B-binding protein*					4,62	3,25	3,58	3,38	
Q13952	**NFYC**	*Nuclear transcription factor Y subunit gamma*					Exclu				
Q9NX24	**NHP2**	*H/ACA ribonucleoprotein complex subunit 2*						1,64		1,36	
Q9BVI4	**NOC4L**	*Nucleolar complex protein 4 homolog*						1,23		1,34	
P78316	**NOP14**	*Nucleolar protein 14*						1,27		1,53	
O75607	**NPM3**	*Nucleoplasmin-3*								1,12	
P49116	**NR2C2**	*Nuclear receptor subfamily 2 group C member 2*								2,36	
P24468	**NR2F2**	*COUP transcription factor 2*					Exclu				
Q96L73	**NSD1**	*Histone-lysine N-methyltransferase, H3 lysine-36 and H4 lysine-20 specific*					2,66	3,56		4,46	
Q9Y266	**NUDC**	*Nuclear migration protein nudC*	1,77								
Q8WVJ2	**NUDCD2**	*NudC domain-containing protein 2*	4,32		3,45		2,58		2,87		3,17
Q8IVD9	**NUDCD3**	*NudC domain-containing protein 3*	5,89		3,90		2,71		2,53		4,12
A8MXV4	**NUDT19**	*Nucleoside diphosphate-linked moiety X motif 19, mitochondrial*		Exclu							
Q14980	**NUMA1**	*Nuclear mitotic apparatus protein 1*						1,05		1,60	
P49790	**NUP153**	*Nuclear pore complex protein Nup153*					1,06	1,55			
Q9UKX7	**NUP50**	*Nuclear pore complex protein Nup50*					1,03	1,46			
Q99567	**NUP88**	*Nuclear pore complex protein Nup88*		1,03							
O15381	**NVL**	*Nuclear valosin-containing protein-like*						1,31		1,42	
Q96G74	**OTUD5**	*OTU domain-containing protein 5*					Exclu				
Q8N7H5	**PAF1**	*RNA polymerase II-associated factor 1 homolog*					1,47	2,36		1,87	
O95453	**PARN**	*Poly(A)-specific ribonuclease PARN*						1,11			
Q96RG2	**PASK**	*PAS domain-containing serine/threonine-protein kinase*		1,03							
Q9Y5B6	**PAXBP1**	*PAX3- and PAX7-binding protein 1*						1,67			
P11498	**PC**	*Pyruvate carboxylase, mitochondrial*		1,04							1,28
O94913	**PCF11**	*Pre-mRNA cleavage complex 2 protein Pcf11*						1,27			
Q29RF7	**PDS5A**	*Sister chromatid cohesion protein PDS5 homolog A*								1,42	
Q8NDX5	**PHC3**	*Polyhomeotic-like protein 3*						1,19		1,43	
Q8WUB8	**PHF10**	*PHD finger protein 10*								1,17	
O75151	**PHF2**	*Lysine-specific demethylase PHF2*								Exclu	
Q96BD5	**PHF21A**	*PHD finger protein 21A*					Exclu				
Q92576	**PHF3**	*PHD finger protein 3*						1,40		1,61	
Q9UPP1	**PHF8**	*Histone lysine demethylase PHF8*								1,51	
O43175	**PHGDH**	*D-3-phosphoglycerate dehydrogenase*	1,27								
Q9P1Y6	**PHRF1**	*PHD and RING finger domain-containing protein 1*						Exclu			
Q96BK5	**PINX1**	*PIN2/TERF1-interacting telomerase inhibitor 1*								1,29	
P29590	**PML**	*Protein PML*					1,55				
Q9H307	**PNN**	*Pinin*					1,26				
Q7Z3K3	**POGZ**	*Pogo transposable element with ZNF domain*					1,49	1,75	1,27	2,05	
Q9Y253	**POLH**	*DNA polymerase eta*						Exclu			
Q9BUI4	**POLR3C**	*DNA-directed RNA polymerase III subunit RPC3*						1,01			
P14859	**POU2F1**	*POU domain, class 2, transcription factor 1*					1,77				
Q8NEY8	**PPHLN1**	*Periphilin-1*					Exclu				
O43447	**PPIH**	*Peptidyl-prolyl cis-trans isomerase H*						Exclu			
Q9Y3C6	**PPIL1**	*Peptidyl-prolyl cis-trans isomerase-like 1*						1,62	1,72	1,32	
Q13356	**PPIL2**	*Peptidyl-prolyl cis-trans isomerase-like 2*						1,04		1,18	
Q9H2H8	**PPIL3**	*Peptidyl-prolyl cis-trans isomerase-like 3*						1,65		1,31	
Q8WUA2	**PPIL4**	*Peptidyl-prolyl cis-trans isomerase-like 4*						1,19		1,04	
O15355	**PPM1G**	*Protein phosphatase 1G*								1,01	
Q96QC0	**PPP1R10**	*Serine/threonine-protein phosphatase 1 regulatory subunit 10*					1,28	1,24			
O14974	**PPP1R12A**	*Protein phosphatase 1 regulatory subunit 12A*	2,98								
P67775	**PPP2CA**	*Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform*		2,46							
P30153	**PPP2R1A**	*Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform*	4,02	3,49							
P60510	**PPP4C**	*Serine/threonine-protein phosphatase 4 catalytic subunit*	1,77	2,02							
Q9NY27	**PPP4R2**	*Serine/threonine-protein phosphatase 4 regulatory subunit 2*					2,27	2,33		1,88	
P53041	**PPP5C**	*Serine/threonine-protein phosphatase 5*	6,56								3,62
Q9UPN7	**PPP6R1**	*Serine/threonine-protein phosphatase 6 regulatory subunit 1*	1,24			1,48					
O60828	**PQBP1**	*Polyglutamine-binding protein 1*					1,48	1,52		1,75	
Q92733	**PRCC**	*Proline-rich protein PRCC*					1,61	1,99		1,64	

Uniprot ID	Gene names	Protein names	sT		LT275		LT330		LT428		LT
			N-ter	C-ter	N-ter	C-ter	N-ter	C-ter	N-ter	C-ter	N-ter
Q13029	**PRDM2**	*PR domain zinc finger protein 2*						Exclu			
P49643	**PRIM2**	*DNA primase large subunit*									1,23
Q13131	**PRKAA1**	*5-AMP-activated protein kinase catalytic subunit alpha-1*	1,15								
Q05655	**PRKCD**	*Protein kinase C delta type*	Exclu								
O43395	**PRPF3**	*U4/U6 small nuclear ribonucleoprotein Prp3*					1,53	1,58			
Q8WWY3	**PRPF31**	*U4/U6 small nuclear ribonucleoprotein Prp31*						1,09		2,14	
Q8NAV1	**PRPF38A**	*Pre-mRNA-splicing factor 38A*						1,21		2,02	
O43172	**PRPF4**	*U4/U6 small nuclear ribonucleoprotein Prp4*					1,64	1,75		1,24	
O75400	**PRPF40A**	*Pre-mRNA-processing factor 40 homolog A*					2,24	1,26			
O94906	**PRPF6**	*Pre-mRNA-processing factor 6*						1,12			
Q5JSZ5	**PRRC2B**	*Protein PRRC2B*		2,56							
O14818	**PSMA7**	*Proteasome subunit alpha type-7*			1,69		1,55		1,89		
P20618	**PSMB1**	*Proteasome subunit beta type-1*					Exclu				
P61289	**PSME3**	*Proteasome activator complex subunit 3*						1,75			
Q6PGN9	**PSRC1**	*Proline/serine-rich coiled-coil protein 1*		1,24							
Q9UHX1	**PUF60**	*Poly(U)-binding-splicing factor PUF60*						1,18			
Q96C36	**PYCR2**	*Pyrroline-5-carboxylate reductase 2*	1,22								
Q2TAL8	**QRICH1**	*Glutamine-rich protein 1*					Exclu				
P51153	**RAB13**	*Ras-related protein Rab-13*					Exclu				
O60216	**RAD21**	*Double-strand-break repair protein rad21 homolog*								2,80	
Q7Z5J4	**RAI1**	*Retinoic acid-induced protein 1*						1,81		2,09	
Q8IY67	**RAVER1**	*Ribonucleoprotein PTB-binding 1*					1,03		1,35		
P06400	**RB1**	*Retinoblastoma-associated protein*			3,56	3,20	5,05	3,10	4,64	3,44	
Q09028	**RBBP4**	*Histone-binding protein RBBP4*						1,25		1,53	
Q7Z6E9	**RBBP6**	*E3 ubiquitin-protein ligase RBBP6*						1,32			
P98175	**RBM10**	*RNA-binding protein 10*					1,64	2,52		1,61	
P57052	**RBM11**	*Splicing regulator RBM11*						Exclu			
Q9NTZ6	**RBM12**	*RNA-binding protein 12*					1,03				
Q96PK6	**RBM14**	*RNA-binding protein 14*	1,08								
Q9Y4C8	**RBM19**	*Probable RNA-binding protein 19*								1,69	
Q9NW64	**RBM22**	*Pre-mRNA-splicing factor RBM22*					1,05				
P49756	**RBM25**	*RNA-binding protein 25*					1,03				
Q9Y580	**RBM7**	*RNA-binding protein 7*					1,61	1,92			
P62877	**RBX1**	*E3 ubiquitin-protein ligase RBX1*						1,37			
Q9UKL0	**RCOR1**	*REST corepressor 1*					1,45	1,98		1,78	
Q9P2K3	**RCOR3**	*REST corepressor 3*						Exclu			
P40938	**RFC3**	*Replication factor C subunit 3*					1,17				
P22670	**RFX1**	*MHC class II regulatory factor RFX1*						1,60		1,81	
Q5UIP0	**RIF1**	*Telomere-associated protein RIF1*						1,48		1,33	
O15541	**RNF113A**	*RING finger protein 113A*						1,01			
Q8NCN4	**RNF169**	*E3 ubiquitin-protein ligase RNF169*						Exclu		Exclu	
Q5VTR2	**RNF20**	*E3 ubiquitin-protein ligase BRE1A*					1,18				
Q15287	**RNPS1**	*RNA-binding protein with serine-rich domain 1*						1,12			
P27694	**RPA1**	*Replication protein A 70 kDa DNA-binding subunit*						1,21			
P15927	**RPA2**	*Replication protein A 32 kDa subunit*						1,00			
P46778	**RPL21**	*60S ribosomal protein L21*	1,14								1,46
Q59GN2	**RPL39P5**	*Putative 60S ribosomal protein L39-like 5*									1,35
P05387	**RPLP2**	*60S acidic ribosomal protein P2*	4,00				1,06				2,74
Q9NQG5	**RPRD1B**	*Regulation of nuclear pre-mRNA domain-containing protein 1B*						1,15			
Q5VT52	**RPRD2**	*Regulation of nuclear pre-mRNA domain-containing protein 2*					1,63	1,86			
Q92766	**RREB1**	*Ras-responsive element-binding protein 1*						Exclu		Exclu	
Q92541	**RTF1**	*RNA polymerase-associated protein RTF1 homolog*					1,54		1,16		
Q6SPF0	**SAMD1**	*Atherin*	Exclu		Exclu		Exclu		Exclu		Exclu
Q9NP81	**SARS2**	*Serine--tRNA ligase, mitochondrial*									1,05
O43290	**SART1**	*U4/U6.U5 tri-snRNP-associated protein 1*					1,12	1,32		1,33	
Q15020	**SART3**	*Squamous cell carcinoma antigen recognized by T-cells 3*					1,23	1,09			
A3KN83	**SBNO1**	*Protein strawberry notch homolog 1*					2,27	1,15			
Q9H7N4	**SCAF1**	*Splicing factor, arginine/serine-rich 19*								Exclu	
O95104	**SCAF4**	*Splicing factor, arginine/serine-rich 15*						1,03			
Q9UQR0	**SCML2**	*Sex comb on midleg-like protein 2*						1,05		1,31	
Q9H4L4	**SENP3**	*Sentrin-specific protease 3*						1,05		1,19	
Q9GZR1	**SENP6**	*Sentrin-specific protease 6*						1,71			
Q14141	**SEPT6**	*Septin-6*	1,86								
O15047	**SETD1A**	*Histone-lysine N-methyltransferase SETD1A*					1,50				
Q9BYW2	**SETD2**	*Histone-lysine N-methyltransferase SETD2*						Exclu			
Q15637	**SF1**	*Splicing factor 1*						1,09			
Q15459	**SF3A1**	*Splicing factor 3A subunit 1*						1,03			
Q15428	**SF3A2**	*Splicing factor 3A subunit 2*						1,21			

Uniprot ID	Gene names	Protein names	sT		LT275		LT330		LT428		LT
			N-ter	C-ter	N-ter	C-ter	N-ter	C-ter	N-ter	C-ter	N-ter
Q12874	**SF3A3**	*Splicing factor 3A subunit 3*					1,73	1,42			
O75533	**SF3B1**	*Splicing factor 3B subunit 1*						1,14			
Q13435	**SF3B2**	*Splicing factor 3B subunit 2*					1,08	1,22			
Q9Y3B4	**SF3B6**	*Splicing factor 3B subunit 6*						1,12			
Q12872	**SFSWAP**	*Splicing factor, suppressor of white-apricot homolog*					3,80	4,06		3,41	
O43765	**SGTA**	*Small glutamine-rich tetratricopeptide repeat-containing protein alpha*	Exclu				Exclu				Exclu
P29353	**SHC1**	*SHC-transforming protein 1*				1,26					
P34897	**SHMT2**	*Serine hydroxymethyltransferase, mitochondrial*								1,15	
Q96ST3	**SIN3A**	*Paired amphipathic helix protein Sin3a*						1,01		1,11	
O75182	**SIN3B**	*Paired amphipathic helix protein Sin3b*		1,23			1,66	2,70	1,44	2,14	
Q9P2F8	**SIPA1L2**	*Signal-induced proliferation-associated 1-like protein 2*		Exclu							
Q96EB6	**SIRT1**	*NAD-dependent protein deacetylase sirtuin-1*					Exclu	Exclu	Exclu	Exclu	
Q8IXJ6	**SIRT2**	*NAD-dependent protein deacetylase sirtuin-2*	Exclu	Exclu			Exclu				Exclu
Q9BWU0	**SLC4A1AP**	*Kanadaptin*						1,12			
O95391	**SLU7**	*Pre-mRNA-splicing factor SLU7*					1,51	2,37		2,04	
O00193	**SMAP**	*Small acidic protein*						2,22		2,62	
P28370	**SMARCA1**	*Probable global transcription activator SNF2L1*								1,36	
Q8TAQ2	**SMARCC2**	*SWI/SNF complex subunit SMARCC2*					1,23	1,23		1,22	
Q969G3	**SMARCE1**	*SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1*						1,09		1,04	1,27
Q8IY18	**SMC5**	*Structural maintenance of chromosomes protein 5*					Exclu	Exclu	Exclu		
Q96SB8	**SMC6**	*Structural maintenance of chromosomes protein 6*				2,04	4,14	2,79	4,05	3,26	
Q6IN85	**SMEK1**	*Serine/threonine-protein phosphatase 4 regulatory subunit 3A*						1,13			
Q16637	**SMN1**	*Survival motor neuron protein*		Exclu				Exclu			
O95295	**SNAPIN**	*SNARE-associated protein Snapin*	1,12								
Q96DI7	**SNRNP40**	*U5 small nuclear ribonucleoprotein 40 kDa protein*						1,06			
P09012	**SNRPA**	*U1 small nuclear ribonucleoprotein A*						1,09			
Q13573	**SNW1**	*SNW domain-containing protein 1*					1,40	1,25			
P18583	**SON**	*Protein SON*					1,03	1,16			
P08047	**SP1**	*Transcription factor Sp1*					Exclu				
Q9UBP0	**SPAST**	*Spastin*	Exclu								
Q9HBM1	**SPC25**	*Kinetochore protein Spc25*	Exclu				Exclu				Exclu
Q96EA4	**SPDL1**	*Protein Spindly*						2,16			
Q96T58	**SPEN**	*Msx2-interacting protein*					3,03	2,48	1,49	1,01	
Q6ZRS2	**SRCAP**	*Helicase SRCAP*					1,55		1,04		
Q05519	**SRSF11**	*Serine/arginine-rich splicing factor 11*					1,71				
A0A029	**sT**	*small T antigen*	Exclu	Exclu							
P50502	**ST13**	*Hsc70-interacting protein*	8,56		6,05		4,54		4,12		6,30
Q8N3U4	**STAG2**	*Cohesin subunit SA-2*								3,82	
P31948	**STIP1**	*Stress-induced-phosphoprotein 1*	8,75		6,64		6,37		5,83		7,48
Q5VSL9	**STRIP1**	*Striatin-interacting protein 1*	1,78								
O43815	**STRN**	*Striatin*		2,69							
Q9NRL3	**STRN4**	*Striatin-4*		2,47							
Q9UNE7	**STUB1**	*E3 ubiquitin-protein ligase CHIP*	6,97				4,52		4,03		5,34
Q8IWZ8	**SUGP1**	*SURP and G-patch domain-containing protein 1*					2,30	2,53	1,47	1,97	
Q8IX01	**SUGP2**	*SURP and G-patch domain-containing protein 2*					1,44	1,81		1,34	
Q9Y2Z0	**SUGT1**	*Suppressor of G2 allele of SKP1 homolog*	3,08		2,11		1,06		1,72		2,09
Q8NEM7	**SUPT20H**	*Transcription factor SPT20 homolog*						Exclu			
P63272	**SUPT4H1**	*Transcription elongation factor SPT4*					Exclu				
O00267	**SUPT5H**	*Transcription elongation factor SPT5*								1,37	
Q7KZ85	**SUPT6H**	*Transcription elongation factor SPT6*					4,03	3,54		3,80	
O94864	**SUPT7L**	*STAGA complex 65 subunit gamma*						Exclu			
Q92797	**SYMPK**	*Symplekin*						1,34			
O43426	**SYNJ1**	*Synaptojanin-1*		1,28							
Q9NYJ8	**TAB2**	*TGF-beta-activated kinase 1 and MAP3K7-binding protein 2*	Exclu								
Q8N5C8	**TAB3**	*TGF-beta-activated kinase 1 and MAP3K7-binding protein 3*		Exclu							
Q96BN2	**TADA1**	*Transcriptional adapter 1*					Exclu	Exclu			
O75528	**TADA3**	*Transcriptional adapter 3*						Exclu			
P21675	**TAF1**	*Transcription initiation factor TFIID subunit 1*					1,65	2,14		3,46	
Q12962	**TAF10**	*Transcription initiation factor TFIID subunit 10*						Exclu			
O00268	**TAF4**	*Transcription initiation factor TFIID subunit 4*					Exclu	Exclu	Exclu		
P49848	**TAF6**	*Transcription initiation factor TFIID subunit 6*					3,38	1,89		1,68	
Q9Y6J9	**TAF6L**	*TAF6-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kDa subunit 6L*						1,04			
Q15545	**TAF7**	*Transcription initiation factor TFIID subunit 7*					1,76		1,69		
Q16594	**TAF9**	*Transcription initiation factor TFIID subunit 9*						Exclu			
Q14241	**TCEB3**	*Transcription elongation factor B polypeptide 3*						1,25		1,51	
Q9UGU0	**TCF20**	*Transcription factor 20*					3,68	3,79		3,10	

Uniprot ID	Gene names	Protein names	sT		LT275		LT330		LT428		LT
			N-ter	C-ter	N-ter	C-ter	N-ter	C-ter	N-ter	C-ter	N-ter
Q14186	**TFDP1**	*Transcription factor Dp-1*								1,37	
Q9UBB9	**TFIP11**	*Tuftelin-interacting protein 11*					1,83	2,18	1,16	1,19	
Q96FV9	**THOC1**	*THO complex subunit 1*					1,27	1,30			
Q13769	**THOC5**	*THO complex subunit 5 homolog*						Exclu			
O14925	**TIMM23**	*Mitochondrial import inner membrane translocase subunit Tim23*					1,27			1,47	
Q96EY4	**TMA16**	*Translation machinery-associated protein 16*		1,90		1,21					
Q15785	**TOMM34**	*Mitochondrial import receptor subunit TOM34*	1,65	1,06		1,80					
Q92547	**TOPBP1**	*DNA topoisomerase 2-binding protein 1*					1,93				
O94842	**TOX4**	*TOX high mobility group box family member 4*						1,99		1,58	
P04637	**TP53**	*Cellular tumor antigen p53*		1,20		1,24					
Q12888	**TP53BP1**	*Tumor suppressor p53-binding protein 1*					1,57				
P12270	**TPR**	*Nucleoprotein TPR*					1,06				
Q9ULW0	**TPX2**	*Targeting protein for Xklp2*								1,27	
O43617	**TRAPPC3**	*Trafficking protein particle complex subunit 3*	2,91				1,65				1,58
Q9UPN9	**TRIM33**	*E3 ubiquitin-protein ligase TRIM33*					1,22	1,30			
Q99614	**TTC1**	*Tetratricopeptide repeat protein 1*	3,90								1,96
Q96AY4	**TTC28**	*Tetratricopeptide repeat protein 28*		1,14		1,08					
O95801	**TTC4**	*Tetratricopeptide repeat protein 4*	1,83								
Q8N5M4	**TTC9C**	*Tetratricopeptide repeat protein 9C*	3,16								
Q9BUF5	**TUBB6**	*Tubulin beta-6 chain*		1,74		1,43				1,35	1,01
O15042	**U2SURP**	*U2 snRNP-associated SURP motif-containing protein*					1,25	1,80		1,51	
Q9NPG3	**UBN1**	*Ubinuclein-1*						Exclu			
Q6ZU65	**UBN2**	*Ubinuclein-2*						Exclu			
Q9UHD9	**UBQLN2**	*Ubiquilin-2*	Exclu		Exclu						Exclu
P61960	**UFM1**	*Ubiquitin-fold modifier 1*	2,31								
Q96RL1	**UIMC1**	*BRCA1-A complex subunit RAP80*						1,70		1,98	
P11172	**UMPS**	*Uridine 5-monophosphate synthase*	1,07								
Q9H3U1	**UNC45A**	*Protein unc-45 homolog A*	1,60								
P13051	**UNG**	*Uracil-DNA glycosylase*						1,14		1,87	
Q93009	**USP7**	*Ubiquitin carboxyl-terminal hydrolase 7*						1,25			
O75691	**UTP20**	*Small subunit processome component 20 homolog*								2,15	
Q96JC1	**VPS39**	*Vam6/Vps39-like protein*			Exclu	Exclu	Exclu	Exclu	Exclu	Exclu	Exclu
Q9BTA9	**WAC**	*WW domain-containing adapter protein with coiled-coil*					1,03	1,22			
Q7Z5K2	**WAPAL**	*Wings apart-like protein homolog*								1,30	
Q9Y2W2	**WBP11**	*WW domain-binding protein 11*					1,85		1,52		
O75554	**WBP4**	*WW domain-binding protein 4*						1,15			
O75717	**WDHD1**	*WD repeat and HMG-box DNA-binding protein 1*					1,04	2,15		1,72	
Q9C0J8	**WDR33**	*pre-mRNA 3 end processing protein WDR33*						1,40			
Q9NW82	**WDR70**	*WD repeat-containing protein 70*						1,11		1,14	
Q6UXN9	**WDR82**	*WD repeat-containing protein 82*						1,05		1,02	
O96028	**WHSC1**	*Histone-lysine N-methyltransferase NSD2*								1,26	
Q9BZ95	**WHSC1L1**	*Histone-lysine N-methyltransferase NSD3*					1,13	1,24	1,01	2,04	
O95785	**WIZ**	*Protein Wiz*						1,23	1,02	2,44	
Q9HCS7	**XAB2**	*Pre-mRNA-splicing factor SYF1*					1,14	1,73		1,55	
P23025	**XPA**	*DNA repair protein complementing XP-A cells*								Exclu	
Q13426	**XRCC4**	*DNA repair protein XRCC4*					4,06		4,48		
P46937	**YAP1**	*Transcriptional coactivator YAP1*	4,00								
Q9ULM3	**YEATS2**	*YEATS domain-containing protein 2*					4,17	3,66	2,40	2,19	
Q9ULJ3	**ZBTB21**	*Zinc finger and BTB domain-containing protein 21*					1,97	2,24		2,35	
Q86T24	**ZBTB33**	*Transcriptional regulator Kaiso*						1,45		1,42	
Q96C00	**ZBTB9**	*Zinc finger and BTB domain-containing protein 9*								1,22	
Q6PJT7	**ZC3H14**	*Zinc finger CCCH domain-containing protein 14*					1,67	1,02			
Q86VM9	**ZC3H18**	*Zinc finger CCCH domain-containing protein 18*					1,05				
Q9UPT8	**ZC3H4**	*Zinc finger CCCH domain-containing protein 4*						1,56			
Q6NZY4	**ZCCHC8**	*Zinc finger CCHC domain-containing protein 8*		1,22			1,29	1,79		1,00	
Q9HCK1	**ZDBF2**	*DBF4-type zinc finger-containing protein 2*		3,30							
Q8N5A5	**ZGPAT**	*Zinc finger CCCH-type with G patch domain-containing protein*					Exclu				
Q9Y6X8	**ZHX2**	*Zinc fingers and homeoboxes protein 2*								Exclu	
Q9H4I2	**ZHX3**	*Zinc fingers and homeoboxes protein 3*						1,90		1,68	
Q969J2	**ZKSCAN4**	*Zinc finger protein with KRAB and SCAN domains 4*								1,26	
Q5SVZ6	**ZMYM1**	*Zinc finger MYM-type protein 1*						Exclu			
Q9UBW7	**ZMYM2**	*Zinc finger MYM-type protein 2*						1,43			
Q14202	**ZMYM3**	*Zinc finger MYM-type protein 3*						1,63		2,02	
Q9ULU4	**ZMYND8**	*Protein kinase C-binding protein 1*						1,68		2,24	
Q9H2Y7	**ZNF106**	*Zinc finger protein 106*						Exclu			
P52739	**ZNF131**	*Zinc finger protein 131*						2,25			
Q9UQR1	**ZNF148**	*Zinc finger protein 148*						1,76		2,96	
Q15697	**ZNF174**	*Zinc finger protein 174*								1,16	

Uniprot ID	Gene names	Protein names	sT		LT275		LT330		LT428		LT
			N-ter	C-ter	N-ter	C-ter	N-ter	C-ter	N-ter	C-ter	N-ter
P17028	**ZNF24**	*Zinc finger protein 24*						Exclu		Exclu	
Q8ND82	**ZNF280C**	*Zinc finger protein 280C*						1,28			
Q9Y2X9	**ZNF281**	*Zinc finger protein 281*					2,04	2,54		2,93	
Q5VUA4	**ZNF318**	*Zinc finger protein 318*					2,17	1,72		1,03	
Q5T0B9	**ZNF362**	*Zinc finger protein 362*						1,39		2,37	
Q8TF68	**ZNF384**	*Zinc finger protein 384*							Exclu		
Q96ME7	**ZNF512**	*Zinc finger protein 512*								1,38	
Q96KM6	**ZNF512B**	*Zinc finger protein 512B*								2,13	
Q92618	**ZNF516**	*Zinc finger protein 516*					Exclu	Exclu			
O15014	**ZNF609**	*Zinc finger protein 609*					1,36	1,00			
Q9UEG4	**ZNF629**	*Zinc finger protein 629*						1,68		2,30	
Q9H582	**ZNF644**	*Zinc finger protein 644*								Exclu	
Q8N1G0	**ZNF687**	*Zinc finger protein 687*						1,63		1,36	
Q9Y462	**ZNF711**	*Zinc finger protein 711*								1,23	
Q96NB3	**ZNF830**	*Zinc finger protein 830*						1,45		1,36	
Q8TBC5	**ZSCAN18**	*Zinc finger and SCAN domain-containing protein 18*						2,48			
Q9Y5A6	**ZSCAN21**	*Zinc finger and SCAN domain-containing protein 21*						Exclu		Exclu	
O43264	**ZW10**	*Centromere/kinetochore protein zw10 homolog*	Exclu								

ANNEXE II *: Les protéines réprésentées ici sont celles dont l'abondance varie de manière significative entre les conditions WT et shTA. Le Log2fc représence la différence de Log2(LFQ) entre les WaGa shTA et les WaGa WT. Les protéines marquées de vert sont plus exprimées en présence des antigènes T et les protéines marquées de rouge sont plus exprimées lorsque l'expression des antigènes T est inhibée. Une protéine est consdérée exclusive à une condition lorsqu'une valeur de LFQ est mesurée dans chacun des trois réplicats tandis qu'aucune valeur de LFQ n'est mesurée dans l'autre condition.*

Uniprot ID	Gene names	Protein names	WaGa wt-Waga shTA			
			WT l shTA	Log2fc	*q-value*	Exclusivité
Q5VST6	**ABHD17B**	*Alpha/beta hydrolase domain-containing protein 17B*		1,10	*0,01*	
P09110	**ACAA1**	*3-ketoacyl-CoA thiolase, peroxisomal*		1,01	*0,01*	
P45954	**ACADSB**	*Short/branched chain specific acyl-CoA dehydrogenase, mitochondrial*		1,06	*0,02*	
Q9BWD1	**ACAT2**	*Acetyl-CoA acetyltransferase, cytosolic*		-1,03	*0,00*	
P21399	**ACO1**	*Cytoplasmic aconitate hydratase*		-1,66	*0,01*	
P11117	**ACP2**	*Lysosomal acid phosphatase*		1,16	*0,00*	
P15309	**ACPP**	*Prostatic acid phosphatase*		-1,14	*0,03*	
Q4G176	**ACSF3**	*Acyl-CoA synthetase family member 3, mitochondrial*		1,19	*0,00*	
P27144	**AK4**	*Adenylate kinase 4, mitochondrial*		1,01	*0,01*	
Q92667	**AKAP1**	*A-kinase anchor protein 1, mitochondrial*		-1,09	*0,03*	
P13716	**ALAD**	*Delta-aminolevulinic acid dehydratase*		1,35	*0,00*	
O75843	**AP1G2**	*AP-1 complex subunit gamma-like 2*				**Exclusif WT**
P02649	**APOE**	*Apolipoprotein E*		2,35	*0,01*	
Q01814	**ATP2B2**	*Plasma membrane calcium-transporting ATPase 2*		1,35	*0,01*	
Q15904	**ATP6AP1**	*V-type proton ATPase subunit S1*		1,03	*0,02*	
Q9Y679	**AUP1**	*Ancient ubiquitous protein 1*				**Exclusif shTA**
O75366	**AVIL**	*Advillin*		-6,87	*0,00*	
Q9NRL2	**BAZ1A**	*Bromodomain adjacent to zinc finger domain protein 1A*		-1,37	*0,03*	
Q53F19	**C17orf85**	*Uncharacterized protein C17orf85*		1,11	*0,01*	
Q9HB71	**CACYBP**	*Calcyclin-binding protein*		-1,10	*0,01*	
P27708	**CAD**	*CAD protein*		-1,22	*0,00*	
P05937	**CALB1**	*Calbindin*				**Exclusif shTA**
P13987	**CD59**	*CD59 glycoprotein*				**Exclusif WT**
P49427	**CDC34**	*Ubiquitin-conjugating enzyme E2 R1*		1,19	*0,01*	
P06493	**CDK1**	*Cyclin-dependent kinase 1*		-3,28	*0,01*	
P24941	**CDK2**	*Cyclin-dependent kinase 2*		-1,14	*0,00*	
P49918	**CDKN1C**	*Cyclin-dependent kinase inhibitor 1C*				**Exclusif shTA**
Q5SZQ8	**CELF3**	*CUGBP Elav-like family member 3*		1,63	*0,00*	
O14757	**CHEK1**	*Serine/threonine-protein kinase Chk1*				**Exclusif WT**
P51797	**CLCN6**	*Chloride transport protein 6*				**Exclusif shTA**
Q15417	**CNN3**	*Calponin-3*		-1,69	*0,02*	
Q02246	**CNTN2**	*Contactin-2*				**Exclusif shTA**
O43405	**COCH**	*Cochlin*		-1,03	*0,02*	
Q9UQ03	**CORO2B**	*Coronin-2B*		1,78	*0,02*	
P29762	**CRABP1**	*Cellular retinoic acid-binding protein 1*		-1,31	*0,00*	
Q9UHQ9	**CYB5R1**	*NADH-cytochrome b5 reductase 1*		1,36	*0,00*	
Q9H773	**DCTPP1**	*dCTP pyrophosphatase 1*		-1,32	*0,00*	
O95865	**DDAH2**	*N(G),N(G)-dimethylarginine dimethylaminohydrolase 2*		-1,35	*0,01*	
Q92466	**DDB2**	*DNA damage-binding protein 2*		-1,21	*0,00*	
Q9UMR2	**DDX19B**	*ATP-dependent RNA helicase DDX19B*				**Exclusif shTA**
P00374	**DHFR**	*Dihydrofolate reductase*				**Exclusif WT**
O60610	**DIAPH1**	*Protein diaphanous homolog 1*		-1,71	*0,01*	
P26358	**DNMT1**	*DNA (cytosine-5)-methyltransferase 1*		-1,26	*0,00*	
Q16555	**DPYSL2**	*Dihydropyrimidinase-related protein 2*		-1,23	*0,01*	
P33316	**DUT**	*Deoxyuridine 5-triphosphate nucleotidohydrolase, mitochondrial*		-2,28	*0,00*	
O75923	**DYSF**	*Dysferlin*		1,45	*0,00*	
O43854	**EDIL3**	*EGF-like repeat and discoidin I-like domain-containing protein 3*		-1,33	*0,02*	
Q14156	**EFR3A**	*Protein EFR3 homolog A*				**Exclusif shTA**
Q9BY44	**EIF2A**	*Eukaryotic translation initiation factor 2A*		-1,08	*0,01*	
Q14576	**ELAVL3**	*ELAV-like protein 3*		-1,03	*0,03*	
Q8NFI3	**ENGASE**	*Cytosolic endo-beta-N-acetylglucosaminidase*		1,56	*0,00*	
P29323	**EPHB2**	*Ephrin type-B receptor 2*				**Exclusif shTA**
Q03468	**ERCC6**	*DNA excision repair protein ERCC-6*				**Exclusif shTA**

Uniprot ID	Gene names	Protein names	WaGa wt-Waga shTA WT l shTA	Log2fc	*q-value*	Exclusivité
Q9H5X1	**FAM96A**	*MIP18 family protein FAM96A*				**Exclusif WT**
P49327	**FASN**	*Fatty acid synthase*		**-1,05**	***0,00***	
Q8WUP2	**FBLIM1**	*Filamin-binding LIM protein 1*				**Exclusif shTA**
Q13642	**FHL1**	*Four and a half LIM domains protein 1*				**Exclusif shTA**
P21333	**FLNA**	*Filamin-A*		**-1,33**	***0,01***	
Q5VW36	**FOCAD**	*Focadhesin*				**Exclusif WT**
Q16658	**FSCN1**	*Fascin*		**-1,08**	***0,00***	
Q9BXM9	**FSD1L**	*FSD1-like protein*		**2,26**	***0,00***	
P60520	**GABARAPL2**	*Gamma-aminobutyric acid receptor-associated protein-like 2*		**-1,50**	***0,01***	
P34059	**GALNS**	*N-acetylgalactosamine-6-sulfatase*		**1,13**	***0,04***	
Q06210	**GFPT1**	*Glutamine--fructose-6-phosphate aminotransferase [isomerizing] 1*		**-1,23**	***0,00***	
Q86SX6	**GLRX5**	*Glutaredoxin-related protein 5, mitochondrial*		**-1,35**	***0,01***	
O94925	**GLS**	*Glutaminase kidney isoform, mitochondrial*		**-2,69**	***0,01***	
Q9HAV0	**GNB4**	*Guanine nucleotide-binding protein subunit beta-4*				**Exclusif WT**
Q8TD30	**GPT2**	*Alanine aminotransferase 2*		**-1,30**	***0,01***	
P07203	**GPX1**	*Glutathione peroxidase 1*		**1,89**	***0,00***	
P18283	**GPX2**	*Glutathione peroxidase 2*		**3,39**	***0,00***	
Q14687	**GSE1**	*Genetic suppressor element 1*		**1,06**	***0,02***	
Q9Y2Q3	**GSTK1**	*Glutathione S-transferase kappa 1*		**1,00**	***0,00***	
P07305	**H1F0**	*Histone H1.0*		**1,08**	***0,00***	
O75367	**H2AFY**	*Core histone macro-H2A.1*		**1,15**	***0,01***	
P84243	**H3F3A**	*Histone H3.3*		**5,62**	***0,01***	
O14929	**HAT1**	*Histone acetyltransferase type B catalytic subunit*		**-1,21**	***0,00***	
Q7Z4H7	**HAUS6**	*HAUS augmin-like complex subunit 6*				**Exclusif WT**
Q9BSH5	**HDHD3**	*Haloacid dehalogenase-like hydrolase domain-containing protein 3*		**1,04**	***0,01***	
Q9NRZ9	**HELLS**	*Lymphoid-specific helicase*				**Exclusif WT**
P10321	**HLA-C**	*HLA class I histocompatibility antigen, Cw-7 alpha chain*		**1,59**	***0,04***	
Q9UK76	**HN1**	*Hematological and neurological expressed 1 protein*		**-1,89**	***0,02***	
P37235	**HPCAL1**	*Hippocalcin-like protein 1*		**1,18**	***0,01***	
O43301	**HSPA12A**	*Heat shock 70 kDa protein 12A*		**1,17**	***0,02***	
P04792	**HSPB1**	*Heat shock protein beta-1*		**1,08**	***0,03***	
P51553	**IDH3G**	*Isocitrate dehydrogenase [NAD] subunit gamma, mitochondrial*		**1,05**	***0,01***	
P18065	**IGFBP2**	*Insulin-like growth factor-binding protein 2*		**1,04**	***0,03***	
P24593	**IGFBP5**	*Insulin-like growth factor-binding protein 5*		**2,18**	***0,00***	
Q969P0	**IGSF8**	*Immunoglobulin superfamily member 8*		**-1,34**	***0,03***	
O95373	**IPO7**	*Importin-7*		**-1,88**	***0,00***	
Q15051	**IQCB1**	*IQ calmodulin-binding motif-containing protein 1*				**Exclusif WT**
Q14573	**ITPR3**	*Inositol 1,4,5-trisphosphate receptor type 3*		**-1,68**	***0,01***	
P26440	**IVD**	*Isovaleryl-CoA dehydrogenase, mitochondrial*		**1,02**	***0,03***	
Q7Z4H8	**KDELC2**	*KDEL motif-containing protein 2*		**-2,47**	***0,01***	
P52292	**KPNA2**	*Importin subunit alpha-1*		**-2,14**	***0,01***	
P07942	**LAMB1**	*Laminin subunit beta-1*		**-1,40**	***0,00***	
Q14847	**LASP1**	*LIM and SH3 domain protein 1*				**Exclusif shTA**
Q6P1Q0	**LETMD1**	*LETM1 domain-containing protein 1*		**1,03**	***0,02***	
P17931	**LGALS3**	*Galectin-3*		**-1,70**	***0,02***	
P18858	**LIG1**	*DNA ligase 1*		**-2,77**	***0,00***	
Q96II8	**LRCH3**	*Leucine-rich repeat and calponin homology domain-containing protein 3*				**Exclusif WT**
A0A028	**LT**	*Large T Antigen MCPyV*		**-2,24**	***0,02***	
Q5U5X0	**LYRM7**	*Complex III assembly factor LYRM7*		**-1,05**	***0,04***	
Q8IV50	**LYSMD2**	*LysM and putative peptidoglycan-binding domain-containing protein 2*		**1,06**	***0,01***	
P43362	**MAGEA9**	*Melanoma-associated antigen 9*		**1,01**	***0,03***	
O00754	**MAN2B1**	*Lysosomal alpha-mannosidase*		**1,06**	***0,01***	
P46734	**MAP2K3**	*Dual specificity mitogen-activated protein kinase kinase 3*				**Exclusif shTA**
Q9P0L2	**MARK1**	*Serine/threonine-protein kinase MARK1*				**Exclusif shTA**
P02144	**MB**	*Myoglobin*		**1,57**	***0,00***	

Uniprot ID	Gene names	Protein names	WaGa wt-Waga shTA			
			WT l shTA	Log2fc	*q-value*	Exclusivité
Q9UIS9	**MBD1**	*Methyl-CpG-binding domain protein 1*		**1,00**	***0,04***	
P49736	**MCM2**	*DNA replication licensing factor MCM2*		**-1,06**	***0,00***	
Q14566	**MCM6**	*DNA replication licensing factor MCM6*		**-1,07**	***0,00***	
Q8TDD5	**MCOLN3**	*Mucolipin-3*				**Exclusif shTA**
Q9UBU8	**MORF4L1**	*Mortality factor 4-like protein 1*		**1,20**	***0,04***	
P29372	**MPG**	*DNA-3-methyladenine glycosylase*		**1,29**	***0,02***	
Q5T653	**MRPL2**	*39S ribosomal protein L2, mitochondrial*				**Exclusif shTA**
Q9UDX5	**MTFP1**	*Mitochondrial fission process protein 1*				**Exclusif WT**
P11586	**MTHFD1**	*C-1-tetrahydrofolate synthase, cytoplasmic*		**-1,15**	***0,00***	
P13995	**MTHFD2**	*Bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase, mitochondrial*		**-1,39**	***0,00***	
P22033	**MUT**	*Methylmalonyl-CoA mutase, mitochondrial*		**-1,65**	***0,01***	
Q14764	**MVP**	*Major vault protein*		**3,24**	***0,00***	
Q9P2K5	**MYEF2**	*Myelin expression factor 2*		**1,19**	***0,00***	
P35579	**MYH9**	*Myosin-9*		**-1,01**	***0,00***	
Q13459	**MYO9B**	*Unconventional myosin-IXb*				**Exclusif WT**
Q01538	**MYT1**	*Myelin transcription factor 1*		**1,07**	***0,01***	
P55209	**NAP1L1**	*Nucleosome assembly protein 1-like 1*		**-1,10**	***0,00***	
Q15021	**NCAPD2**	*Condensin complex subunit 1*		**-4,00**	***0,00***	
P42695	**NCAPD3**	*Condensin-2 complex subunit D3*		**-2,26**	***0,02***	
Q9BPX3	**NCAPG**	*Condensin complex subunit 3*		**-2,54**	***0,00***	
Q86XI2	**NCAPG2**	*Condensin-2 complex subunit G2*		**-2,50**	***0,00***	
Q15003	**NCAPH**	*Condensin complex subunit 2*				**Exclusif WT**
Q6IBW4	**NCAPH2**	*Condensin-2 complex subunit H2*				**Exclusif WT**
Q92597	**NDRG1**	*Protein NDRG1*		**2,27**	***0,02***	
Q86UT6	**NLRX1**	*NLR family member X1*		**1,33**	***0,01***	
Q13232	**NME3**	*Nucleoside diphosphate kinase 3*		**1,02**	***0,01***	
Q9UNW9	**NOVA2**	*RNA-binding protein Nova-2*		**1,19**	***0,00***	
Q86WQ0	**NR2C2AP**	*Nuclear receptor 2C2-associated protein*		**-1,17**	***0,01***	
Q5TFE4	**NT5DC1**	*5-nucleotidase domain-containing protein 1*		**-1,62**	***0,00***	
P80303	**NUCB2**	*Nucleobindin-2*		**-1,01**	***0,00***	
A8MXV4	**NUDT19**	*Nucleoside diphosphate-linked moiety X motif 19, mitochondrial*		**1,15**	***0,02***	
Q9Y6K5	**OAS3**	*2-5-oligoadenylate synthase 3*		**-1,10**	***0,02***	
Q8N543	**OGFOD1**	*Prolyl 3-hydroxylase OGFOD1*		**-1,01**	***0,01***	
Q9H0X9	**OSBPL5**	*Oxysterol-binding protein-related protein 5*				**Exclusif shTA**
Q9Y2J8	**PADI2**	*Protein-arginine deiminase type-2*		**1,34**	***0,01***	
Q96IZ0	**PAWR**	*PRKC apoptosis WT1 regulator protein*		**-1,44**	***0,03***	
P61457	**PCBD1**	*Pterin-4-alpha-carbinolamine dehydratase*		**1,07**	***0,00***	
P57723	**PCBP4**	*Poly(rC)-binding protein 4*		**1,19**	***0,00***	
P12004	**PCNA**	*Proliferating cell nuclear antigen*		**-2,03**	***0,00***	
Q9UKZ9	**PCOLCE2**	*Procollagen C-endopeptidase enhancer 2*		**1,25**	***0,02***	
P29120	**PCSK1**	*Neuroendocrine convertase 1*				**Exclusif WT**
O00151	**PDLIM1**	*PDZ and LIM domain protein 1*		**-1,14**	***0,01***	
Q53GG5	**PDLIM3**	*PDZ and LIM domain protein 3*				**Exclusif WT**
Q9NUG6	**PDRG1**	*p53 and DNA damage-regulated protein 1*				**Exclusif WT**
O60825	**PFKFB2**	*6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 2*				**Exclusif WT**
P08237	**PFKM**	*ATP-dependent 6-phosphofructokinase, muscle type*				**Exclusif shTA**
Q15124	**PGM5**	*Phosphoglucomutase-like protein 5*				**Exclusif WT**
Q8IWS0	**PHF6**	*PHD finger protein 6*		**-1,25**	***0,04***	
O00625	**PIR**	*Pirin*				**Exclusif WT**
Q8NHP8	**PLBD2**	*Putative phospholipase B-like 2*		**1,12**	***0,01***	
P19174	**PLCG1**	*1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-1*		**-1,26**	***0,00***	
Q15149	**PLEC**	*Plectin*		**1,65**	***0,00***	
O60568	**PLOD3**	*Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3*		**-1,20**	***0,00***	
Q8IYS1	**PM20D2**	*Peptidase M20 domain-containing protein 2*		**-1,16**	***0,04***	
P28340	**POLD1**	*DNA polymerase delta catalytic subunit*		**-1,30**	***0,00***	
P49005	**POLD2**	*DNA polymerase delta subunit 2*		**-1,14**	***0,02***	
Q06203	**PPAT**	*Amidophosphoribosyltransferase*		**-1,06**	***0,01***	
P30405	**PPIF**	*Peptidyl-prolyl cis-trans isomerase F, mitochondrial*		**-1,06**	***0,05***	

Uniprot ID	Gene names	Protein names	WaGa wt-Waga shTA			
			WT l shTA	Log2fc	*q-value*	Exclusivité
O43447	**PPIH**	*Peptidyl-prolyl cis-trans isomerase H*		1,30	*0,03*	
Q8WY54	**PPM1E**	*Protein phosphatase 1E*		-1,22	*0,02*	
Q15173	**PPP2R5B**	*Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit beta isoform*				**Exclusif shTA**
Q13162	**PRDX4**	*Peroxiredoxin-4*		-1,06	*0,00*	
P49642	**PRIM1**	*DNA primase small subunit*				**Exclusif WT**
P49643	**PRIM2**	*DNA primase large subunit*		-2,14	*0,00*	
P11908	**PRPS2**	*Ribose-phosphate pyrophosphokinase 2*		-1,03	*0,00*	
Q14558	**PRPSAP1**	*Phosphoribosyl pyrophosphate synthase-associated protein 1*		-1,11	*0,02*	
Q9Y617	**PSAT1**	*Phosphoserine aminotransferase*		-1,05	*0,00*	
Q9BT73	**PSMG3**	*Proteasome assembly chaperone 3*				**Exclusif WT**
Q8N8N7	**PTGR2**	*Prostaglandin reductase 2*		1,39	*0,02*	
Q9ULC3	**RAB23**	*Ras-related protein Rab-23*		-1,50	*0,00*	
Q969Q5	**RAB24**	*Ras-related protein Rab-24*				**Exclusif shTA**
P61006	**RAB8A**	*Ras-related protein Rab-8A*		-1,03	*0,03*	
O43502	**RAD51C**	*DNA repair protein RAD51 homolog 3*				**Exclusif WT**
P43487	**RANBP1**	*Ran-specific GTPase-activating protein*		-1,24	*0,04*	
O43374	**RASA4**	*Ras GTPase-activating protein 4*		1,86	*0,00*	
Q8IXT5	**RBM12B**	*RNA-binding protein 12B*		1,04	*0,02*	
Q96R05	**RBP7**	*Retinoid-binding protein 7*		-1,10	*0,01*	
Q9P2K3	**RCOR3**	*REST corepressor 3*				**Exclusif shTA**
Q8IUW5	**RELL1**	*RELT-like protein 1*		-2,25	*0,03*	
P35251	**RFC1**	*Replication factor C subunit 1*		-1,24	*0,01*	
P35250	**RFC2**	*Replication factor C subunit 2*		-1,07	*0,00*	
P40938	**RFC3**	*Replication factor C subunit 3*		-1,45	*0,03*	
P35249	**RFC4**	*Replication factor C subunit 4*		-1,26	*0,00*	
P40937	**RFC5**	*Replication factor C subunit 5*		-1,36	*0,00*	
O43665	**RGS10**	*Regulator of G-protein signaling 10*				**Exclusif WT**
Q5UIP0	**RIF1**	*Telomere-associated protein RIF1*		-1,13	*0,00*	
Q9BV68	**RNF126**	*E3 ubiquitin-protein ligase RNF126*				**Exclusif WT**
P78345	**RPP38**	*Ribonuclease P protein subunit p38*		1,18	*0,01*	
Q92766	**RREB1**	*Ras-responsive element-binding protein 1*				**Exclusif shTA**
P23921	**RRM1**	*Ribonucleoside-diphosphate reductase large subunit*		-1,96	*0,00*	
O95197	**RTN3**	*Reticulon-3*		-1,23	*0,03*	
Q9NQC3	**RTN4**	*Reticulon-4*		-1,07	*0,04*	
Q59EK9	**RUNDC3A**	*RUN domain-containing protein 3A*		1,91	*0,00*	
P31151	**S100A7**	*Protein S100-A7*				**Exclusif shTA**
Q13228	**SELENBP1**	*Selenium-binding protein 1*		1,53	*0,00*	
Q9NVA2	**SEPT11**	*Septin-11*		-1,00	*0,02*	
Q9UHD8	**SEPT9**	*Septin-9*		-1,40	*0,04*	
P58004	**SESN2**	*Sestrin-2*				**Exclusif shTA**
Q15047	**SETDB1**	*Histone-lysine N-methyltransferase SETDB1*				**Exclusif WT**
P31947	**SFN**	*14-3-3 protein sigma*		-1,63	*0,00*	
Q99962	**SH3GL2**	*Endophilin-A1*				**Exclusif shTA**
P34896	**SHMT1**	*Serine hydroxymethyltransferase, cytosolic*		-2,58	*0,01*	
P43007	**SLC1A4**	*Neutral amino acid transporter A*		-1,02	*0,00*	
Q8TBP6	**SLC25A40**	*Solute carrier family 25 member 40*		-1,29	*0,01*	
Q96QD8	**SLC38A2**	*Sodium-coupled neutral amino acid transporter 2*		-1,68	*0,01*	
Q9GZT3	**SLIRP**	*SRA stem-loop-interacting RNA-binding protein, mitochondrial*		-1,08	*0,01*	
O95347	**SMC2**	*Structural maintenance of chromosomes protein 2*		-2,58	*0,00*	
Q9NTJ3	**SMC4**	*Structural maintenance of chromosomes protein 4*		-2,25	*0,01*	
Q9NRA0	**SPHK2**	*Sphingosine kinase 2*		1,56	*0,01*	
Q8N1F8	**STK11IP**	*Serine/threonine-protein kinase 11-interacting protein*				**Exclusif shTA**
P16949	**STMN1**	*Stathmin*		-1,55	*0,00*	
Q9UHF0	**TAC3**	*Tachykinin-3*		1,61	*0,00*	
Q96BW9	**TAMM41**	*Phosphatidate cytidylyltransferase, mitochondrial*		1,03	*0,00*	
Q9BTW9	**TBCD**	*Tubulin-specific chaperone D*		-1,18	*0,00*	
Q969Z0	**TBRG4**	*Protein TBRG4*		-1,49	*0,00*	
Q5H9L2	**TCEAL5**	*Transcription elongation factor A protein-like 5*				**Exclusif WT**

Uniprot ID	Gene names	Protein names	WaGa wt-Waga shTA			
			WT l shTA	Log2fc	*q-value*	Exclusivité
Q14186	**TFDP1**	*Transcription factor Dp-1*				**Exclusif WT**
P04216	**THY1**	*Thy-1 membrane glycoprotein*				**Exclusif shTA**
O60220	**TIMM8A**	*Mitochondrial import inner membrane translocase subunit Tim8 A*				**Exclusif WT**
Q9UDY2	**TJP2**	*Tight junction protein ZO-2*		**-1,55**	***0,00***	
O76062	**TM7SF2**	*Delta(14)-sterol reductase*		**1,80**	***0,00***	
Q5SNT2	**TMEM201**	*Transmembrane protein 201*		**2,16**	***0,00***	
Q9H813	**TMEM206**	*Transmembrane protein 206*				**Exclusif WT**
Q8N4L2	**TMEM55A**	*Type 2 phosphatidylinositol 4,5-bisphosphate 4-phosphatase*				**Exclusif shTA**
O60784	**TOM1**	*Target of Myb protein 1*		**1,21**	***0,00***	
P11388	**TOP2A**	*DNA topoisomerase 2-alpha*				**Exclusif WT**
Q13472	**TOP3A**	*DNA topoisomerase 3-alpha*				**Exclusif WT**
P67936	**TPM4**	*Tropomyosin alpha-4 chain*		**-1,12**	***0,01***	
Q9ULW0	**TPX2**	*Targeting protein for Xklp2*				**Exclusif WT**
Q15645	**TRIP13**	*Pachytene checkpoint protein 2 homolog*				**Exclusif WT**
O60637	**TSPAN3**	*Tetraspanin-3*				**Exclusif shTA**
Q3ZCM7	**TUBB8**	*Tubulin beta-8 chain*		**-1,62**	***0,04***	
P23258	**TUBG1**	*Tubulin gamma-1 chain*		**-1,08**	***0,01***	
Q16763	**UBE2S**	*Ubiquitin-conjugating enzyme E2 S*				**Exclusif WT**
Q9NPD8	**UBE2T**	*Ubiquitin-conjugating enzyme E2 T*				**Exclusif WT**
Q8N806	**UBR7**	*Putative E3 ubiquitin-protein ligase UBR7*		**-1,12**	***0,01***	
Q9BZV1	**UBXN6**	*UBX domain-containing protein 6*		**1,22**	***0,03***	
Q96T88	**UHRF1**	*E3 ubiquitin-protein ligase UHRF1*		**-1,75**	***0,03***	
P13051	**UNG**	*Uracil-DNA glycosylase*				**Exclusif WT**
O75717	**WDHD1**	*WD repeat and HMG-box DNA-binding protein 1*		**-2,18**	***0,00***	
Q14191	**WRN**	*Werner syndrome ATP-dependent helicase*				**Exclusif WT**
Q9UQR1	**ZNF148**	*Zinc finger protein 148*		**2,12**	***0,02***	
Q9H900	**ZWILCH**	*Protein zwilch homolog*				**Exclusif WT**

ANNEXE III : *Interacteurs de haute confiance de EHMT2 (Log2fc > 1, q-value < 0,01). Les log2fc des interacteurs identifiés dans chaque condition sont en gras et sur fond vert. Les log2fc des interacteurs significativement enrichis en présence de tLT sont en gras sur fond bleu (Log2fc > 1, q-value < 0,05).*

Uniprot ID	Gene names	Protein name	FB-EHMT2 Flag-only		FB-EHMT2 Flag-LT330		in LT330 vs. empty	
			log2fc	q-value	log2fc	q-value	log2fc	q-value
Q13085	**ACACA**	*acetyl-CoA carboxylase alpha*	2,4	1,6E-02	**2,8**	**2,0E-03**	0,4	9,7E-02
Q9H2P0	**ADNP**	*activity-dependent neuroprotector homeobox protein*	**3,1**	**1,3E-03**	**3,6**	**1,6E-04**	0,4	1,2E-01
Q8WYP5	**AHCTF1**	*AT hook containing transcription factor 1*	**1,4**	**5,2E-03**	**1,9**	**8,0E-03**	0,5	4,1E-01
P54886	**ALDH18A1**	*aldehyde dehydrogenase 18 family member A1*	1,0	4,8E-06	**1,0**	**4,4E-04**	0,0	7,9E-01
Q6P6C2	**ALKBH5**	*alpha-ketoglutarate dependent dioxygenase alkB homolog 5*	**4,9**	**8,1E-03**	**4,6**	**1,8E-03**	-0,4	1,4E-01
Q9BTT0	**ANP32E**	*acidic nuclear phosphoprotein 32 family member E*	6,3	2,2E-02	**6,3**	**4,7E-03**	0,1	9,2E-02
Q9BZZ5	**API5**	*apoptosis inhibitor 5*	1,2	2,5E-02	**1,5**	**2,5E-03**	0,3	6,2E-02
O60306	**AQR**	*aquarius intron-binding splicing factor*	1,8	5,8E-02	**2,2**	**3,8E-03**	0,4	4,9E-01
Q8IVW6	**ARID3B**	*AT-rich interaction domain 3B*	2,8	1,9E-02	**3,6**	**1,7E-03**	0,8	4,9E-03
Q66PJ3	**ARL6IP4**	*ADP ribosylation factor like GTPase 6 interacting protein 4*	**2,3**	**8,6E-03**	**1,0**	**1,3E-03**	-1,3	4,3E-02
Q7Z6K5	**ARPIN**	*actin related protein 2/3 complex inhibitor*	1,0	4,2E-02	**1,4**	**2,9E-03**	0,4	6,2E-04
Q6VMQ6	**ATF7IP**	*activating transcription factor 7 interacting protein*	3,5	1,6E-02	**3,9**	**2,0E-03**	0,4	8,0E-02
P46100	**ATRX**	*alpha thalassemia/mental retardation syndrome X-linked*	1,3	1,9E-01	**3,3**	**8,1E-03**	**2,0**	**1,6E-02**
O95817	**BAG3**	*BCL2 associated athanogene 3*	**2,1**	**1,4E-03**	**2,1**	**4,3E-05**	0,0	8,6E-01
Q8IXM2	**BAP18**	*BRCA1 associated protein 1*	1,7	1,1E-01	**2,6**	**2,6E-05**	1,0	2,9E-01
Q9NRL2	**BAZ1A**	*bromodomain adjacent to zinc finger domain 1A*	**1,9**	**9,7E-03**	**2,3**	**8,1E-04**	0,4	1,9E-02
Q9UIG0	**BAZ1B**	*bromodomain adjacent to zinc finger domain 1B*	**3,3**	**3,6E-03**	**4,4**	**1,6E-04**	**1,1**	**4,3E-02**
Q9UIF9	**BAZ2A**	*bromodomain adjacent to zinc finger domain 2A*	2,1	6,5E-02	**2,4**	**9,0E-03**	0,3	5,3E-01
Q14692	**BMS1**	*BRCA1 mRNA splicing factor*	0,9	4,3E-02	**1,2**	**2,3E-03**	0,3	3,9E-01
Q15059	**BRD3**	*bromodomain containing 3*	**4,9**	**7,1E-04**	**5,4**	**3,4E-05**	0,5	1,5E-02
O60885	**BRD4**	*bromodomain containing 4*	0,9	2,6E-02	**2,6**	**1,2E-04**	**1,8**	**6,2E-04**
O43684	**BUB3**	*BUB3 mitotic checkpoint protein*	**1,0**	**8,8E-04**	0,7	3,4E-04	-0,3	9,6E-04
P41223	**BUD31**	*BUD31 homolog, splicing factor*	-0,5	6,1E-01	**1,1**	**7,4E-03**	1,6	1,7E-01
Q96C57	**C12orf43**	*chromosome 12 open reading frame 43*	1,6	3,1E-02	**1,7**	**2,0E-03**	0,2	5,9E-01
Q53F19	**C17orf85**	*chromosome 17 open reading frame 85*	2,7	1,5E-02	**3,1**	**1,7E-03**	0,5	2,1E-01
Q8N9M1	**C19orf47**	*chromosome 19 open reading frame 47*	2,2	4,3E-02	**2,5**	**3,9E-03**	0,4	3,8E-01
Q8N6N3	**C1orf52**	*chromosome 1 open reading frame 52*	**1,4**	**5,7E-03**	**1,9**	**2,4E-04**	0,4	4,1E-03
Q9BRJ6	**C7orf50**	*chromosome 7 open reading frame 50*	2,7	2,0E-02	**2,7**	**2,3E-03**	0,0	9,6E-01
Q9NZ63	**C9orf78**	*chromosome 9 open reading frame 78*	**5,2**	**3,8E-03**	**4,8**	**6,4E-04**	-0,3	9,9E-02
P83916	**CBX1**	*chromobox 1*	1,2	7,0E-02	**3,3**	**1,8E-03**	2,1	5,1E-02
Q14781	**CBX2**	*chromobox 2*	3,7	2,8E-02	**4,2**	**3,3E-03**	0,5	2,4E-01
Q13185	**CBX3**	*chromobox 3*	**2,6**	**4,4E-04**	**2,6**	**1,9E-07**	-0,1	6,8E-01
O00257	**CBX4**	*chromobox 4*	3,1	3,4E-02	**3,2**	**7,3E-03**	0,1	3,4E-01
Q9HC52	**CBX8**	*chromobox 8*	**4,7**	**2,2E-03**	**4,7**	**2,2E-04**	0,0	9,1E-01
Q8N163	**CCAR2**	*cell cycle and apoptosis regulator 2*	0,9	6,5E-03	**1,1**	**3,4E-04**	0,2	7,8E-02
Q6PII3	**CCDC174**	*coiled-coil domain containing 174*	1,7	1,1E-01	**3,3**	**1,7E-04**	1,6	1,1E-01
Q86WR0	**CCDC25**	*coiled-coil domain containing 25*	1,0	3,8E-02	**2,0**	**1,2E-04**	**1,0**	**2,7E-02**
O60563	**CCNT1**	*cyclin T1*	2,2	2,8E-02	**3,3**	**1,8E-03**	1,0	9,0E-02
Q96FF9	**CDCA5**	*cell division cycle associated 5*	3,2	2,2E-02	**3,2**	**4,8E-03**	0,0	7,5E-01
Q53HL2	**CDCA8**	*cell division cycle associated 8*	**3,8**	**8,1E-03**	**4,9**	**1,3E-04**	1,1	8,9E-02
Q9Y232	**CDYL**	*chromodomain Y like*	4,5	1,6E-02	**4,5**	**3,2E-03**	0,0	9,7E-01
Q03188	**CENPC**	*centromere protein C*	3,8	1,3E-02	**4,1**	**2,0E-03**	0,2	1,3E-02
Q7Z7K6	**CENPV**	*centromere protein V*	**1,6**	**5,5E-04**	0,8	1,6E-03	-0,8	2,2E-03
Q96JM3	**CHAMP1**	*chromosome alignment maintaining phosphoprotein 1*	**3,5**	**9,6E-03**	**4,4**	**6,3E-04**	0,8	4,6E-02
Q14839	**CHD4**	*chromodomain helicase DNA binding protein 4*	**1,4**	**2,0E-05**	**1,9**	**7,9E-05**	0,5	5,0E-02
Q9HCK8	**CHD8**	*chromodomain helicase DNA binding protein 8*	2,6	1,4E-02	**3,1**	**5,6E-04**	0,5	1,9E-01
O00299	**CLIC1**	*chloride intracellular channel 1*	0,7	3,3E-01	**3,8**	**9,8E-04**	**3,1**	**1,2E-03**
Q8NBJ5	**COLGALT1**	*collagen beta(1-O)galactosyltransferase 1*	2,9	3,8E-02	**4,3**	**2,3E-03**	**1,4**	**7,8E-04**
Q9NQ92	**COPRS**	*coproporphyrinogen oxidase*	1,8	1,7E-01	**2,7**	**6,0E-03**	0,9	3,9E-01
Q10570	**CPSF1**	*cleavage and polyadenylation specific factor 1*	**1,1**	**2,8E-03**	0,8	2,0E-03	-0,3	5,6E-02
Q16630	**CPSF6**	*cleavage and polyadenylation specific factor 6*	**1,6**	**8,3E-04**	**1,9**	**1,3E-05**	0,3	1,3E-01
Q8N684	**CPSF7**	*cleavage and polyadenylation specific factor 7*	**1,6**	**4,0E-03**	**1,3**	**1,4E-03**	-0,3	5,0E-02
P46108	**CRK**	*CRK proto-oncogene, adaptor protein*	0,7	3,0E-02	**2,1**	**1,4E-03**	**1,4**	**3,6E-02**
P46109	**CRKL**	*CRK like proto-oncogene, adaptor protein*	**1,4**	**5,6E-03**	**2,3**	**3,5E-04**	0,9	1,6E-02
O75390	**CS**	*citrate synthase*	2,7	1,1E-01	**4,4**	**2,4E-04**	1,7	2,5E-01
P48729	**CSNK1A1**	*casein kinase 1 alpha 1*	1,4	3,0E-02	**1,4**	**6,0E-03**	0,0	1,0E+00
P48730	**CSNK1D**	*casein kinase 1 delta*	**2,6**	**4,2E-04**	**2,6**	**3,1E-04**	0,0	9,7E-01
P19784	**CSNK2A2**	*casein kinase 2 alpha 2*	1,3	2,2E-02	**1,3**	**2,6E-03**	0,0	8,8E-01
Q69YN2	**CWF19L1**	*CWF19 like 1, cell cycle control*	**1,5**	**8,1E-03**	**1,6**	**6,0E-04**	0,2	1,3E-01
Q2TBE0	**CWF19L2**	*CWF19 like 2, cell cycle control*	2,0	6,3E-02	**2,4**	**6,1E-03**	0,4	3,7E-01
Q9H5V9	**CXorf56**	*chromosome X open reading frame 56*	**2,0**	**2,7E-03**	**2,1**	**7,3E-04**	0,0	9,2E-01
Q6NSI4	**CXorf57**	*chromosome X open reading frame 57*	**2,1**	**3,0E-04**	**3,0**	**3,4E-05**	0,8	1,2E-02
Q9UI36	**DACH1**	*dachshund family transcription factor 1*	2,9	4,8E-02	**3,3**	**9,5E-03**	0,4	2,2E-02
Q13206	**DDX10**	*DEAD-box helicase 10*	2,1	2,4E-02	**2,9**	**1,3E-04**	0,8	1,7E-01

Uniprot ID	Gene names	Protein name	FB-EHMT2 Flag-only		FB-EHMT2 Flag-LT330		in LT330 vs. empty	
			log2fc	q-value	log2fc	q-value	log2fc	q-value
Q9UHI6	DDX20	DEAD-box helicase 20	1,7	4,2E-02	2,6	1,6E-03	0,8	6,8E-02
Q9NR30	DDX21	DEAD-box helicase 21	1,0	3,9E-04	0,7	2,7E-03	-0,4	7,7E-02
Q9H8H2	DDX31	DEAD-box helicase 31	1,1	1,3E-01	3,0	3,1E-04	2,0	1,4E-02
Q86XP3	DDX42	DEAD-box helicase 42	1,7	1,1E-02	1,9	1,6E-03	0,1	2,8E-01
Q8TDD1	DDX54	DEAD-box helicase 54	1,5	1,6E-04	-0,3	6,6E-03	-1,8	7,0E-05
Q9NY93	DDX56	DEAD-box helicase 56	1,6	1,6E-02	1,2	5,7E-03	-0,4	1,3E-01
P35659	DEK	DEK proto-oncogene	1,4	8,6E-04	1,6	4,7E-05	0,2	9,1E-02
O60231	DHX16	DEAH-box helicase 16	1,0	3,2E-02	1,8	7,4E-04	0,8	4,3E-02
Q8IY37	DHX37	DEAH-box helicase 37	3,4	2,4E-02	3,7	2,0E-03	0,3	5,1E-01
Q92620	DHX38	DEAH-box helicase 38	1,3	9,5E-02	2,2	2,7E-03	1,0	7,4E-02
Q6P158	DHX57	DEAH-box helicase 57	1,4	5,2E-02	2,9	6,0E-04	1,6	6,3E-03
Q9BTC0	DIDO1	death inducer-obliterator 1	1,7	1,6E-03	2,4	8,6E-05	0,7	2,1E-02
Q14689	DIP2A	DIP2 disco-interacting protein 2 homolog A	1,1	6,0E-04	1,7	5,2E-03	0,5	2,8E-01
Q9Y2L1	DIS3	DIS3 exosome endoribonuclease and 3'-5' exoribonuclease	0,9	5,6E-04	1,3	8,3E-06	0,4	2,6E-05
Q15398	DLGAP5	disks large associated protein 5	1,1	4,7E-03	1,2	4,5E-04	0,1	7,7E-02
Q9NPF5	DMAP1	DNA methyltransferase 1 associated protein 1	0,8	2,5E-01	2,6	3,9E-03	1,8	1,5E-03
Q9NVM6	DNAJC17	DnaJ heat shock protein family (Hsp40) member C17	1,9	5,9E-02	2,4	7,1E-03	0,5	2,4E-01
P26358	DNMT1	DNA methyltransferase 1	1,1	6,2E-03	1,1	1,4E-03	0,0	8,0E-01
Q15029	EFTUD2	elongation factor Tu GTP binding domain containing 2	1,3	9,9E-04	1,5	5,8E-05	0,2	4,3E-02
Q9H9B1	EHMT1	euchromatic histone lysine methyltransferase 1	3,1	9,4E-04	4,1	3,1E-04	1,0	7,8E-02
Q96KQ7	EHMT2	euchromatic histone lysine methyltransferase 2	13,7	1,7E-03	14,1	1,4E-04	0,4	3,1E-02
P56537	EIF6	eukaryotic translation initiation factor 6	1,7	6,6E-02	2,6	5,5E-03	0,9	1,1E-02
P32519	ELF1	E74 like ETS transcription factor 1	1,3	2,5E-01	3,2	1,2E-03	1,9	1,0E-01
Q96L91	EP400	E1A binding protein p400	3,0	5,1E-03	3,5	3,6E-04	0,5	3,9E-02
Q9H501	ESF1	ESF1, nucleolar pre-rRNA processing protein	1,7	7,2E-02	2,7	4,3E-03	1,1	2,9E-02
Q01780	EXOSC10	exosome component 10	1,7	1,3E-03	2,2	1,3E-04	0,5	3,9E-02
Q13868	EXOSC2	exosome component 2	1,2	1,6E-02	1,8	4,0E-04	0,7	7,4E-02
Q15024	EXOSC7	exosome component 7	3,9	5,8E-03	4,7	3,9E-04	0,8	4,9E-03
Q9NZB2	FAM120A	family with sequence similarity 120A	1,1	2,1E-01	2,9	1,7E-04	1,8	7,3E-02
Q96E09	FAM122A	family with sequence similarity 122A	1,3	5,3E-02	1,9	3,5E-03	0,6	2,0E-02
Q9NSI2	FAM207A	family with sequence similarity 207 member A	2,6	4,7E-02	3,5	2,0E-03	0,9	1,5E-01
Q9UK61	FAM208A	family with sequence similarity 208 member A	2,3	1,3E-02	2,7	1,5E-03	0,4	4,1E-01
Q14320	FAM50A	family with sequence similarity 50 member A	1,3	1,6E-02	1,7	1,9E-04	0,4	1,7E-01
Q9NVI1	FANCI	FA complementation group I	1,6	1,4E-01	2,9	5,1E-03	1,4	6,6E-02
Q8NEZ5	FBXO22	F-box protein 22	1,7	9,4E-03	0,6	2,0E-01	-1,0	9,7E-02
P39748	FEN1	flap structure specific endonuclease 1	1,8	1,5E-03	1,9	1,9E-04	0,1	4,6E-01
Q01167	FOXK2	forkhead box K2	1,9	7,1E-02	2,8	1,1E-03	0,8	2,4E-01
Q70Z53	FRA10AC1	FRA10AC1 regulator of chromatin condensation 1	1,7	1,2E-02	2,5	2,3E-04	0,8	5,2E-02
P35637	FUS	FUS RNA binding protein	1,1	1,9E-04	0,9	5,7E-06	-0,2	6,9E-02
Q9NY12	GAR1	GAR1 ribonucleoprotein	1,7	4,4E-03	1,4	5,8E-02	-0,4	6,3E-01
Q8WXI9	GATAD2B	GATA zinc finger domain containing 2B	1,6	1,6E-03	2,3	1,1E-05	0,7	8,5E-03
O14908	GIPC1	GIPC PDZ domain containing family member 1	2,4	3,6E-02	3,1	1,5E-03	0,7	2,0E-01
P00367	GLUD1	glutamate dehydrogenase 1	1,7	7,0E-04	1,5	2,0E-04	-0,2	3,1E-01
Q49A26	GLYR1	glyoxylate reductase 1 homolog	1,6	3,4E-04	2,9	2,0E-04	1,2	2,0E-02
Q5JWF2	GNAS	GNAS complex locus	0,8	1,3E-01	1,8	2,9E-03	1,1	6,5E-02
Q13823	GNL2	G protein nucleolar 2	1,8	1,8E-04	1,9	8,9E-05	0,1	5,1E-01
Q9BVP2	GNL3	G protein nucleolar 3	1,0	1,8E-02	1,0	3,4E-03	0,1	6,3E-01
Q14687	GSE1	GSE1 coiled-coil protein	2,5	6,6E-03	1,4	1,7E-01	-1,1	4,0E-01
Q00403	GTF2B	general transcription factor IIB	2,8	4,5E-02	3,7	3,7E-03	1,0	3,9E-02
P29083	GTF2E1	general transcription factor IIE subunit 1	1,1	4,4E-01	2,8	2,5E-03	1,7	2,4E-01
P35269	GTF2F1	general transcription factor IIF subunit 1	3,3	1,1E-03	3,7	2,7E-06	0,4	1,7E-01
P78347	GTF2I	general transcription factor IIi	1,7	3,9E-04	2,0	1,5E-05	0,3	7,1E-02
Q12789	GTF3C1	general transcription factor IIIC subunit 1	2,3	5,4E-02	3,6	7,6E-04	1,3	1,5E-01
Q9UKN8	GTF3C4	general transcription factor IIIC subunit 4	3,4	5,7E-02	4,3	8,5E-03	0,9	6,8E-02
Q9Y5Q8	GTF3C5	general transcription factor IIIC subunit 5	1,2	2,4E-03	1,9	1,2E-04	0,7	2,6E-02
Q9NYZ3	GTSE1	G2 and S phase expressed 1	2,1	1,4E-01	2,9	7,3E-03	0,9	3,0E-01
Q5JVS0	HABP4	hyaluronan binding protein 4	1,7	7,7E-02	2,8	4,1E-03	1,1	8,9E-03
O14929	HAT1	histone acetyltransferase 1	3,2	1,0E-04	4,4	1,5E-05	1,1	9,8E-03
P51858	HDGF	hepatoma derived growth factor	1,5	8,1E-04	1,8	2,6E-04	0,3	1,4E-01
Q9NRZ9	HELLS	helicase, lymphoid specific	1,8	2,2E-02	3,0	1,2E-03	1,2	1,6E-02
P10412	HIST1H1E	histone cluster 1 H1 family member e	5,8	1,2E-01	5,1	4,8E-04	-0,6	8,3E-01
Q9NP66	HMG20A	high mobility group 20A	1,7	1,5E-01	3,6	6,2E-03	1,9	1,7E-03
Q9UGU5	HMGXB4	HMG box domain containing 4	1,9	4,8E-02	3,6	8,1E-04	1,7	2,4E-02
Q14103	HNRNPD	heterogeneous nuclear ribonucleoprotein D	0,9	1,2E-02	1,1	1,7E-04	0,2	3,6E-01
P31260	HOXA10	homeobox A10	3,3	5,7E-03	3,1	9,4E-04	-0,1	1,9E-01
Q5SSJ5	HP1BP3	heterochromatin protein 1 binding protein 3	2,0	8,5E-03	2,8	1,4E-04	0,8	1,0E-01

Uniprot ID	Gene names	Protein name	FB-EHMT2 Flag-only		FB-EHMT2 Flag-LT330		in LT330 vs. empty	
			log2fc	q-value	log2fc	q-value	log2fc	q-value
P0DMV9	**HSPA1B**	*heat shock protein family A (Hsp70) member 1B*	**1,8**	**2,7E-04**	**1,9**	**3,8E-06**	*0,1*	*1,8E-01*
P11021	**HSPA5**	*heat shock protein family A (Hsp70) member 5*	**1,3**	**9,5E-04**	**1,4**	**5,7E-05**	*0,1*	*2,9E-02*
P17066	**HSPA6**	*heat shock protein family A (Hsp70) member 6*	*3,2*	*4,7E-02*	**5,0**	**3,6E-05**	*1,8*	*1,5E-01*
P11142	**HSPA8**	*heat shock protein family A (Hsp70) member 8*	*0,9*	*4,0E-02*	**1,3**	**2,5E-03**	*0,5*	*2,4E-02*
Q9Y6M1	**IGF2BP2**	*insulin like growth factor 2 mRNA binding protein 2*	*0,7*	*2,9E-02*	**1,3**	**1,0E-03**	*0,5*	*3,7E-02*
Q13123	**IK**	*IK cytokine*	*2,5*	*3,6E-02*	**3,5**	**2,4E-04**	*1,0*	*2,4E-01*
Q8TEX9	**IPO4**	*importin 4*	*0,6*	*2,4E-01*	**4,4**	**2,5E-06**	**3,8**	**2,8E-03**
O95373	**IPO7**	*importin 7*	**3,1**	**4,8E-03**	**4,0**	**4,7E-04**	*0,8*	*7,4E-02*
Q96P70	**IPO9**	*importin 9*	*1,1*	*1,1E-01*	**2,9**	**4,1E-03**	**1,8**	**3,2E-02**
Q15652	**JMJD1C**	*Jumonji domain containing 1C*	*1,5*	*6,4E-02*	**2,3**	**1,1E-03**	*0,8*	*1,5E-01*
Q9BVA0	**KATNB1**	*katanin catalytic subunit B1*	*2,1*	*4,2E-02*	**2,9**	**2,6E-03**	*0,8*	*7,9E-02*
Q96AT1	**KIAA1143**	*KIAA1143*	*3,7*	*1,2E-01*	**6,2**	**2,0E-03**	*2,4*	*1,7E-01*
O00139	**KIF2A**	*kinesin family member 2A*	*1,4*	*1,1E-02*	**1,5**	**2,0E-03**	*0,1*	*4,7E-01*
O95239	**KIF4A**	*kinesin family member 4A*	**2,9**	**9,5E-03**	**3,7**	**7,5E-05**	*0,7*	*2,2E-01*
O60870	**KIN**	*KIN kinetochore scaffold protein*	*2,1*	*4,6E-02*	**3,3**	**2,7E-05**	*1,2*	*1,7E-01*
Q03164	**KMT2A**	*lysine methyltransferase 2A*	*2,7*	*3,4E-02*	**3,7**	**2,3E-03**	*1,0*	*7,0E-02*
Q96JM7	**L3MBTL3**	*L3MBTL3 methyl-CpG binding domain protein 3*	*2,7*	*3,6E-02*	**3,0**	**6,4E-03**	*0,3*	*1,1E-01*
P49916	**LIG3**	*DNA ligase 3*	**2,5**	**2,4E-03**	**2,8**	**2,2E-04**	*0,4*	*1,7E-01*
Q6MZP7	**LIN54**	*lin-54 DREAM MuvB core complex component*	*2,5*	*5,3E-02*	**3,0**	**2,9E-03**	*0,4*	*5,1E-01*
Q9UFC0	**LRWD1**	*leucine rich repeats and WD repeat domain containing 1*	**2,0**	**2,0E-04**	**2,6**	**6,9E-06**	*0,6*	*3,9E-03*
Q9Y333	**LSM2**	*associated*	**2,6**	**3,5E-03**	**3,4**	**1,4E-04**	*0,8*	*2,1E-03*
O95777	**LSM8**	*LSM8 homolog, U6 small nuclear RNA associated*	*2,0*	*7,7E-02*	**3,1**	**4,4E-03**	*1,0*	*9,1E-02*
A0A028	**LT330**	*MCPyV Large T Antigen*	*2,0*	*1,1E-02*	**5,6**	**6,0E-05**	**3,6**	**9,9E-04**
Q96A72	**MAGOHB**	*mago homolog B*	**3,3**	**7,8E-03**	**4,3**	**1,7E-05**	*1,0*	*1,8E-01*
Q9Y4K4	**MAP4K5**	*mitogen-activated protein kinase kinase kinase kinase 5*	**2,6**	**5,0E-03**	**2,9**	**3,9E-04**	*0,2*	*2,8E-01*
P45984	**MAPK9**	*mitogen-activated protein kinase 9*	*1,9*	*2,7E-02*	**3,3**	**7,8E-03**	*1,5*	*1,5E-01*
Q15691	**MAPRE1**	*microtubule associated protein RP/EB family member 1*	*1,3*	*2,2E-02*	**1,3**	**5,1E-03**	*0,0*	*6,9E-01*
Q9HCC0	**MCCC2**	*methylcrotonoyl-CoA carboxylase 2*	**2,4**	**1,7E-03**	**2,3**	**2,0E-04**	*0,0*	*6,8E-01*
Q7L590	**MCM10**	*minichromosome maintenance complex component 10*	*2,6*	*1,7E-02*	**2,2**	**7,1E-03**	*-0,4*	*7,5E-02*
Q14676	**MDC1**	*mediator of DNA damage checkpoint 1*	**1,4**	**3,4E-03**	**1,3**	**3,5E-03**	*-0,1*	*7,2E-01*
Q86U44	**METTL3**	*methyltransferase like 3*	*2,0*	*1,5E-01*	**3,8**	**1,3E-03**	*1,8*	*1,5E-01*
P55081	**MFAP1**	*microfibrillar associated protein 1*	**2,1**	**7,5E-04**	**2,7**	**6,7E-05**	*0,7*	*2,0E-02*
Q9NYP9	**MIS18A**	*MIS18 kinetochore protein A*	*0,9*	*3,7E-01*	**2,5**	**6,7E-03**	*1,6*	*7,5E-02*
P46013	**MKI67**	*marker of proliferation Ki-67*	**1,5**	**1,3E-03**	**1,6**	**9,4E-05**	*0,1*	*9,1E-02*
Q9ULH7	**MKL2**	*MKL/myocardin like 2*	*3,2*	*1,0E-02*	**3,9**	**3,4E-04**	*0,8*	*9,0E-02*
Q9UBU8	**MORF4L1**	*mortality factor 4 like 1*	**2,7**	**1,9E-03**	**2,6**	**1,1E-04**	*0,0*	*8,1E-01*
P29372	**MPG**	*N-methylpurine-DNA glycosylase*	*1,7*	*1,7E-02*	**1,5**	**6,2E-03**	*-0,2*	*7,1E-02*
P49959	**MRE11A**	*MRE11 homolog, double strand break repair nuclease*	**1,1**	**3,4E-03**	**1,1**	**6,2E-04**	*0,0*	*8,1E-01*
P43246	**MSH2**	*mutS homolog 2*	*-0,4*	*2,5E-01*	**1,3**	**2,4E-03**	**1,7**	**1,5E-02**
P52701	**MSH6**	*mutS homolog 6*	**1,8**	**2,2E-04**	**2,1**	**4,7E-05**	*0,3*	*1,0E-01*
Q13330	**MTA1**	*metastasis associated 1*	**1,8**	**7,4E-04**	**1,8**	**1,7E-04**	*0,0*	*9,4E-01*
O94776	**MTA2**	*metastasis associated 1 family member 2*	**2,2**	**1,9E-04**	**2,3**	**5,5E-06**	*0,2*	*2,7E-02*
O00159	**MYO1C**	*myosin IC*	*2,3*	*2,5E-02*	**2,4**	**2,4E-03**	*0,1*	*8,6E-01*
E9PAV3	**NACA**	*nascent polypeptide-associated complex subunit alpha*	**2,6**	**3,6E-03**	**3,2**	**1,7E-04**	*0,6*	*1,1E-02*
P55209	**NAP1L1**	*nucleosome assembly protein 1 like 1*	**2,8**	**1,3E-03**	**3,4**	**6,2E-05**	*0,6*	*1,8E-01*
Q99733	**NAP1L4**	*nucleosome assembly protein 1 like 4*	**3,5**	**1,5E-04**	**3,7**	**1,0E-05**	*0,2*	*2,7E-01*
P49321	**NASP**	*nuclear autoantigenic sperm protein*	**1,9**	**4,2E-03**	**2,3**	**5,0E-04**	*0,3*	*1,1E-01*
Q09161	**NCBP1**	*nuclear cap binding protein subunit 1*	*1,5*	*6,0E-02*	**3,8**	**5,1E-04**	**2,3**	**1,1E-03**
O75376	**NCOR1**	*nuclear receptor corepressor 1*	*1,9*	*5,0E-02*	**2,8**	**1,7E-03**	*0,9*	*9,8E-02*
P18615	**NELFE**	*negative elongation factor complex member E*	**2,5**	**2,5E-03**	**2,6**	**2,8E-04**	*0,1*	*5,5E-01*
Q12857	**NFIA**	*nuclear factor I A*	*5,4*	*1,2E-02*	**5,9**	**1,6E-03**	*0,4*	*1,7E-02*
Q9NX24	**NHP2**	*NHP2 ribonucleoprotein*	**1,9**	**3,7E-03**	*1,0*	*1,1E-01*	*-0,9*	*1,9E-01*
Q9BYG3	**NIFK**	*NIFK nucleolar protein*	*0,8*	*1,4E-01*	**1,3**	**7,9E-03**	*0,4*	*2,5E-01*
Q8N5F7	**NKAP**	*NFKB activating protein*	*0,4*	*6,3E-01*	**2,9**	**8,0E-03**	**2,4**	**4,7E-03**
Q9H2Z4	**NKX2-4**	*NK2 homeobox 4*	**4,2**	**7,8E-04**	**4,7**	**7,6E-05**	*0,4*	*1,4E-01*
P46087	**NOP2**	*NOP2 nucleolar protein*	**1,2**	**1,3E-04**	*0,7*	*1,2E-03*	*-0,5*	*1,6E-02*
Q9Y314	**NOSIP**	*nitric oxide synthase interacting protein*	*3,0*	*3,0E-02*	**2,9**	**7,8E-03**	*-0,1*	*5,1E-01*
O75607	**NPM3**	*nucleophosmin/nucleoplasmin 3*	*0,7*	*2,6E-02*	**1,4**	**2,8E-04**	*0,6*	*2,0E-02*
P49116	**NR2C2**	*nuclear receptor subfamily 2 group C member 2*	*1,8*	*1,6E-02*	**2,0**	**2,0E-03**	*0,2*	*3,0E-01*
Q5TFE4	**NT5DC1**	*5'-nucleotidase domain containing 1*	*2,3*	*3,2E-02*	**3,4**	**2,2E-03**	**1,1**	**1,4E-02**
O43809	**NUDT21**	*nudix hydrolase 21*	*1,2*	*2,4E-02*	**1,3**	**1,6E-04**	*0,1*	*7,8E-01*
Q14980	**NUMA1**	*nuclear mitotic apparatus protein 1*	*1,3*	*2,1E-02*	**1,6**	**8,1E-04**	*0,3*	*3,0E-01*
O15381	**NVL**	*nuclear VCP like*	*3,5*	*1,3E-02*	**4,1**	**1,6E-03**	*0,6*	*6,8E-02*
Q8N543	**OGFOD1**	*containing 1*	*2,2*	*1,2E-01*	**4,7**	**4,4E-03**	**2,4**	**9,0E-03**
Q13416	**ORC2**	*origin recognition complex subunit 2*	**2,3**	**4,6E-03**	**2,7**	**2,4E-04**	*0,4*	*4,9E-02*
P51003	**PAPOLA**	*poly(A) polymerase alpha*	*1,0*	*2,1E-02*	**1,6**	**5,8E-04**	*0,6*	*6,2E-02*

Uniprot ID	Gene names	Protein name	FB-EHMT2 Flag-only		FB-EHMT2 Flag-LT330		in LT330 vs. empty	
			log2fc	q-value	log2fc	q-value	log2fc	q-value
P09874	**PARP1**	*poly(ADP-ribose) polymerase 1*	**1,5**	**1,4E-03**	**1,7**	**2,2E-04**	*0,3*	*1,1E-01*
P05165	**PCCA**	*propionyl-CoA carboxylase alpha subunit*	*2,2*	*1,2E-02*	**2,3**	**1,9E-03**	*0,1*	*1,9E-01*
P05166	**PCCB**	*propionyl-CoA carboxylase beta subunit*	*2,6*	*1,5E-02*	**2,8**	**2,0E-03**	*0,3*	*1,5E-02*
Q9NTI5	**PDS5B**	*PDS5 cohesin associated factor B*	*4,0*	*1,4E-02*	**4,5**	**1,7E-03**	*0,5*	*4,5E-03*
P30086	**PEBP1**	*phosphatidylethanolamine binding protein 1*	*0,1*	*7,1E-01*	**1,0**	**2,9E-03**	*0,9*	*1,1E-01*
O00541	**PES1**	*pescadillo ribosomal biogenesis factor 1*	**1,5**	**2,1E-04**	**1,6**	**2,8E-05**	*0,1*	*2,5E-01*
O75928	**PIAS2**	*protein inhibitor of activated STAT 2*	*2,5*	*2,5E-02*	**2,9**	**2,8E-03**	*0,4*	*1,2E-01*
Q7Z3K3	**POGZ**	*pogo transposable element derived with ZNF domain*	**4,7**	**4,6E-03**	**5,8**	**1,6E-04**	**1,2**	**3,5E-02**
Q15054	**POLD3**	*DNA polymerase delta 3, accessory subunit*	*1,7*	*7,1E-02*	**3,5**	**2,1E-04**	**1,8**	**3,6E-02**
Q99575	**POP1**	*processing of precursor 1, ribonuclease P/MRP subunit*	**2,9**	**3,5E-03**	**2,9**	**3,9E-04**	*0,1*	*3,9E-01*
P14859	**POU2F1**	*POU class 2 homeobox 1*	*2,5*	*3,5E-02*	**3,7**	**6,2E-05**	*1,2*	*1,5E-01*
Q9Y3C6	**PPIL1**	*peptidylprolyl isomerase like 1*	**4,8**	**7,7E-04**	**5,1**	**1,2E-04**	*0,4*	*3,4E-01*
Q13356	**PPIL2**	*peptidylprolyl isomerase like 2*	*1,4*	*1,1E-02*	**1,7**	**9,5E-04**	*0,3*	*1,2E-02*
Q9H2H8	**PPIL3**	*peptidylprolyl isomerase like 3*	**4,3**	**1,3E-03**	**4,5**	**6,6E-05**	*0,3*	*1,5E-01*
Q8WUA2	**PPIL4**	*peptidylprolyl isomerase like 4*	**1,6**	**6,0E-04**	**2,2**	**6,6E-05**	*0,6*	*3,8E-02*
O43314	**PPIP5K2**	*diphosphoinositol pentakisphosphate kinase 2*	*1,3*	*1,3E-01*	**1,7**	**1,6E-03**	*0,4*	*5,8E-01*
P62140	**PPP1CB**	*protein phosphatase 1 catalytic subunit beta*	*4,4*	*2,6E-02*	**4,3**	**6,8E-03**	*-0,1*	*1,9E-01*
Q96QC0	**PPP1R10**	*protein phosphatase 1 regulatory subunit 10*	**1,1**	**3,3E-03**	**1,2**	**2,9E-04**	*0,1*	*1,7E-01*
Q96BP3	**PPWD1**	*PWP1 homolog, spliceosome associated protein*	*0,1*	*8,2E-01*	**1,5**	**9,7E-04**	*1,3*	*5,7E-02*
O60828	**PQBP1**	*polyglutamine binding protein 1*	*3,6*	*5,1E-02*	**4,5**	**3,4E-03**	*0,9*	*2,8E-01*
Q92733	**PRCC**	*papillary renal cell carcinoma*	*1,1*	*5,9E-01*	**3,5**	**8,1E-03**	*2,4*	*2,3E-01*
P30048	**PRDX3**	*peroxiredoxin 3*	*3,1*	*2,4E-02*	**3,4**	**2,7E-03**	*0,3*	*4,1E-01*
Q99873	**PRMT1**	*protein arginine methyltransferase 1*	**1,5**	**3,2E-03**	**2,3**	**3,3E-04**	*0,8*	*8,0E-02*
Q9NVM4	**PRMT7**	*protein arginine methyltransferase 7*	**3,2**	**1,4E-03**	**2,9**	**3,0E-04**	*-0,4*	*1,3E-01*
O43395	**PRPF3**	*pre-mRNA processing factor 3*	**2,2**	**1,6E-03**	**2,5**	**1,2E-04**	*0,4*	*1,8E-01*
Q8NAV1	**PRPF38A**	*pre-mRNA processing factor 38A*	**1,9**	**2,6E-03**	**2,3**	**1,1E-04**	*0,4*	*1,3E-01*
O43172	**PRPF4**	*pre-mRNA processing factor 4*	**1,9**	**1,3E-03**	**2,0**	**2,5E-04**	*0,1*	*6,7E-01*
P48634	**PRRC2A**	*proline rich coiled-coil 2A*	**1,4**	**1,3E-04**	**1,3**	**1,1E-05**	*-0,1*	*1,3E-01*
Q5JSZ5	**PRRC2B**	*proline rich coiled-coil 2B*	*1,8*	*5,5E-02*	**2,2**	**2,8E-03**	*0,4*	*4,1E-01*
P17706	**PTPN2**	*protein tyrosine phosphatase non-receptor type 2*	*2,2*	*1,5E-02*	**3,0**	**7,6E-04**	*0,8*	*4,1E-02*
Q96PZ0	**PUS7**	*pseudouridylate synthase 7*	**1,8**	**2,7E-03**	**1,9**	**3,4E-04**	*0,2*	*3,9E-01*
Q9H0K6	**PUS7L**	*pseudouridylate synthase 7 like*	*1,3*	*8,5E-02*	**2,8**	**1,5E-03**	**1,5**	**1,6E-02**
Q96C36	**PYCR2**	*pyrroline-5-carboxylate reductase 2*	*1,2*	*8,6E-02*	**2,8**	**7,4E-04**	**1,6**	**3,2E-02**
Q96PU8	**QKI**	*QKI, KH domain containing RNA binding*	*2,7*	*3,3E-02*	**2,9**	**7,0E-03**	*0,2*	*2,4E-01*
O60216	**RAD21**	*RAD21 cohesin complex component*	*1,6*	*1,3E-01*	**3,1**	**6,8E-03**	**1,5**	**5,0E-03**
Q7Z5J4	**RAI1**	*retinoic acid induced 1*	**3,1**	**1,6E-03**	**3,7**	**5,4E-05**	*0,6*	*8,2E-03*
Q9H6Z4	**RANBP3**	*RAN binding protein 3*	*0,7*	*7,4E-02*	**1,2**	**4,3E-03**	*0,5*	*2,3E-02*
Q8IY67	**RAVER1**	*ribonucleoprotein, PTB-binding 1*	*2,7*	*4,3E-02*	**3,2**	**1,0E-03**	*0,6*	*4,3E-01*
Q16576	**RBBP7**	*RB binding protein 7, chromatin remodeling factor*	**2,5**	**7,7E-03**	**3,5**	**7,2E-05**	*1,0*	*6,0E-02*
P98175	**RBM10**	*RNA binding motif protein 10*	**2,9**	**2,2E-03**	**2,9**	**4,3E-04**	*0,1*	*7,7E-01*
Q9NW64	**RBM22**	*RNA binding motif protein 22*	**2,1**	**2,6E-03**	**2,4**	**2,4E-04**	*0,3*	*8,3E-02*
Q96EV2	**RBM33**	*RNA binding motif protein 33*	*4,7*	*2,3E-02*	**4,8**	**4,7E-03**	*0,1*	*3,7E-01*
P62877	**RBX1**	*ring-box 1*	*1,7*	*3,3E-02*	**1,8**	**6,9E-03**	*0,1*	*3,2E-01*
P48382	**RFX5**	*regulatory factor X5*	*2,9*	*3,3E-02*	**2,8**	**8,7E-03**	*0,0*	*6,5E-01*
Q5UIP0	**RIF1**	*replication timing regulatory factor 1*	*2,6*	*1,9E-02*	**4,4**	**3,0E-04**	**1,8**	**2,5E-02**
Q9BRS2	**RIOK1**	*RIO kinase 1*	**2,7**	**4,0E-04**	**3,3**	**9,4E-05**	*0,6*	*7,9E-02*
Q8TDP1	**RNASEH2C**	*ribonuclease H2 subunit C*	*0,3*	*6,6E-01*	**2,4**	**6,2E-03**	**2,1**	**4,5E-03**
O15541	**RNF113A**	*ring finger protein 113A*	*4,9*	*2,7E-02*	**5,0**	**6,0E-03**	*0,1*	*5,7E-01*
Q99496	**RNF2**	*ring finger protein 2*	*2,7*	*7,6E-02*	**5,2**	**2,7E-03**	**2,5**	**2,7E-03**
O60942	**RNGTT**	*RNA guanylyltransferase and 5'-phosphatase*	*2,3*	*5,0E-02*	**3,2**	**2,8E-03**	*0,9*	*3,2E-01*
P35244	**RPA3**	*replication protein A3*	*1,3*	*2,6E-02*	**3,2**	**3,5E-04**	**1,9**	**4,4E-03**
P61513	**RPL37A**	*ribosomal protein L37a*	*1,7*	*2,2E-01*	**5,3**	**3,2E-03**	**3,6**	**2,3E-04**
P78346	**RPP30**	*ribonuclease P/MRP 30kDa subunit*	*1,0*	*4,8E-03*	**1,4**	**6,6E-04**	*0,4*	*8,9E-02*
P78345	**RPP38**	*ribonuclease P/MRP 38kDa subunit*	**1,7**	**6,9E-04**	*0,4*	*3,2E-01*	*-1,2*	*7,8E-02*
Q5VT52	**RPRD2**	*regulation of nuclear pre-mRNA domain containing 2*	**1,9**	**1,3E-03**	**3,0**	**1,7E-05**	**1,1**	**1,3E-02**
P62857	**RPS28**	*ribosomal protein S28*	**1,6**	**8,2E-03**	*-2,0*	*3,5E-01*	*-3,6*	*2,4E-01*
Q9Y3B9	**RRP15**	*ribosomal RNA processing 15 homolog*	*4,6*	*1,7E-02*	**4,9**	**2,8E-03**	*0,3*	*2,9E-02*
Q9UHR5	**SAP30BP**	*SAP30 binding protein*	**1,4**	**7,6E-03**	**1,8**	**5,3E-04**	*0,4*	*3,7E-02*
O43290	**SART1**	*squamous cell carcinoma antigen recognized by T-cells*	**1,2**	**6,2E-03**	**1,1**	**1,5E-03**	*-0,1*	*3,5E-01*
Q15020	**SART3**	*squamous cell carcinoma antigen recognized by T-cells 3*	**2,6**	**2,0E-03**	**3,1**	**1,9E-04**	*0,4*	*7,7E-02*
A3KN83	**SBNO1**	*strawberry notch 1*	*2,5*	*3,6E-02*	**3,9**	**2,0E-03**	**1,4**	**1,5E-03**
O95104	**SCAF4**	*SR-related CTD associated factor 4*	**1,1**	**2,9E-03**	*0,8*	*5,3E-03*	*-0,3*	*1,2E-01*
Q9UQR0	**SCML2**	*sex comb on midleg like 2*	**2,1**	**1,1E-03**	**2,4**	**7,1E-05**	*0,3*	*2,2E-02*
Q9NVU7	**SDAD1**	*SDA1 domain containing 1*	*2,3*	*7,7E-02*	**4,5**	**2,2E-03**	**2,2**	**2,3E-03**
Q9GZR1	**SENP6**	*sentrin specific peptidase 6*	*0,9*	*4,6E-03*	**2,1**	**4,3E-05**	**1,2**	**6,1E-04**
Q01105	**SET**	*SET nuclear proto-oncogene*	**2,3**	**7,4E-03**	**2,6**	**8,4E-04**	*0,2*	*5,9E-02*

Uniprot ID	Gene names	Protein name	FB-EHMT2 Flag-only		FB-EHMT2 Flag-LT330		in LT330 vs. empty	
			log2fc	q-value	log2fc	q-value	log2fc	q-value
Q8WTS6	**SETD7**	*SET domain containing 7*	**3,1**	**9,8E-03**	**3,6**	**2,5E-04**	*0,5*	*3,3E-01*
Q13435	**SF3B2**	*splicing factor 3b subunit 2*	**1,3**	**3,0E-03**	**1,5**	**4,3E-04**	*0,2*	*2,9E-01*
P29353	**SHC1**	*SHC adaptor protein 1*	**2,9**	**5,7E-03**	**1,7**	**1,1E-03**	*-1,3*	*3,4E-02*
P34897	**SHMT2**	*serine hydroxymethyltransferase 2*	**2,6**	**1,5E-04**	**2,8**	**4,8E-06**	*0,2*	*2,4E-02*
Q96ST3	**SIN3A**	*SIN3 transcription regulator family member A*	*0,6*	*5,6E-01*	**2,1**	**4,3E-03**	*1,5*	*1,9E-01*
Q9BWU0	**SLC4A1AP**	*solute carrier family 4 member 1 adaptor protein*	*1,6*	*9,1E-02*	**2,2**	**4,3E-03**	*0,6*	*3,6E-01*
O95391	**SLU7**	*pre-mRNA splicing factor SLU7*	**4,7**	**7,0E-03**	**4,6**	**1,1E-03**	*-0,1*	*4,3E-01*
O00193	**SMAP**	*small ArfGAP 1*	*0,6*	*1,9E-01*	**1,7**	**6,7E-03**	*1,0*	*1,5E-01*
P51532	**SMARCA4**	*of chromatin, subfamily a, member 4*	*4,5*	*3,9E-02*	**4,6**	**9,8E-03**	*0,1*	*6,4E-01*
O60264	**SMARCA5**	*of chromatin, subfamily a, member 5*	**1,8**	**2,8E-04**	**1,7**	**1,2E-04**	*0,0*	*9,0E-01*
Q969G3	**SMARCE1**	*of chromatin, subfamily e, member 1*	*1,2*	*1,3E-01*	**2,4**	**7,7E-04**	*1,2*	*8,9E-02*
A6NHR9	**SMCHD1**	*containing 1*	*1,5*	*1,4E-02*	**1,4**	**2,5E-03**	*-0,1*	*6,6E-01*
Q6IN85	**SMEK1**	*SMEK homolog 1, suppressor of mek1*	*-1,1*	*2,8E-01*	**1,2**	**2,6E-03**	*2,3*	*6,5E-02*
Q16637	**SMN1**	*survival of motor neuron 1, telomeric*	*1,3*	*8,5E-02*	**1,9**	**6,0E-03**	*0,7*	*6,9E-02*
O75940	**SMNDC1**	*survival motor neuron domain containing 1*	**1,5**	**3,5E-03**	**1,8**	**3,4E-04**	*0,4*	*6,1E-02*
O75643	**SNRNP200**	*small nuclear ribonucleoprotein U5 subunit 200*	*0,8*	*5,0E-02*	**1,1**	**5,5E-03**	*0,3*	*4,3E-02*
P09661	**SNRPA1**	*small nuclear ribonucleoprotein polypeptide A'*	**1,1**	**7,9E-04**	**1,1**	**7,0E-05**	*0,0*	*9,0E-01*
P62318	**SNRPD3**	*small nuclear ribonucleoprotein D3 polypeptide*	**1,1**	**6,9E-04**	**1,3**	**3,3E-05**	*0,2*	*4,6E-02*
Q13573	**SNW1**	*SNW domain containing 1*	**2,0**	**7,8E-04**	**2,4**	**4,8E-05**	*0,4*	*4,5E-02*
Q68D10	**SPTY2D1**	*SPT2, Suppressor of Ty, domain containing 1*	*1,4*	*1,7E-02*	**1,4**	**6,6E-03**	*0,0*	*9,2E-01*
Q8N5C6	**SRBD1**	*S1 RNA binding domain 1*	**3,1**	**3,8E-03**	**3,4**	**4,0E-04**	*0,3*	*1,5E-01*
P05455	**SSB**	*Sjogren syndrome antigen B*	*0,6*	*5,8E-02*	**1,1**	**6,7E-03**	*0,5*	*1,5E-01*
Q9Y3F4	**STRAP**	*serine/threonine kinase receptor associated protein*	*-0,3*	*3,8E-02*	**1,0**	**4,5E-03**	**1,3**	**1,3E-02**
P53999	**SUB1**	*SUB1 homolog, transcriptional regulator*	**1,2**	**9,1E-03**	**2,8**	**2,6E-04**	**1,5**	**1,4E-02**
Q8IWZ8	**SUGP1**	*SURP and G patch domain containing 1*	**3,7**	**8,1E-03**	**4,0**	**1,0E-03**	*0,2*	*6,6E-02*
O75528	**TADA3**	*transcriptional adaptor 3*	*1,6*	*3,4E-02*	**1,5**	**2,0E-03**	*0,0*	*9,6E-01*
Q9HBM6	**TAF9B**	*TATA-box binding protein associated factor 9b*	*0,1*	*9,1E-01*	**1,1**	**4,4E-03**	*1,0*	*2,9E-01*
Q12788	**TBL3**	*transducin beta like 3*	*0,8*	*1,9E-01*	**1,1**	**9,5E-03**	*0,3*	*6,4E-01*
P23193	**TCEA1**	*transcription elongation factor A1*	**2,9**	**6,7E-03**	*2,6*	*3,5E-02*	*-0,3*	*7,9E-01*
Q14241	**TCEB3**	*transcription elongation factor B polypeptide 3*	*0,7*	*9,0E-02*	**1,4**	**4,4E-03**	*0,8*	*4,0E-02*
Q9UGU0	**TCF20**	*transcription factor 20*	*3,4*	*3,9E-02*	**3,8**	**7,6E-03**	*0,4*	*2,6E-01*
Q96EY4	**TMA16**	*translation machinery associated 16 homolog*	*3,1*	*1,2E-01*	**3,7**	**8,4E-03**	*0,6*	*6,2E-01*
P42166	**TMPO**	*thymopoietin*	*1,0*	*1,2E-03*	**1,1**	**6,9E-04**	*0,1*	*5,8E-01*
Q02880	**TOP2B**	*DNA topoisomerase II beta*	*2,0*	*8,4E-02*	**3,2**	**7,1E-03**	**1,2**	**1,8E-02**
O94842	**TOX4**	*TOX high mobility group box family member 4*	*0,6*	*8,1E-02*	**1,1**	**8,8E-04**	*0,5*	*1,4E-01*
Q9ULW0	**TPX2**	*TPX2, microtubule nucleation factor*	**2,6**	**4,6E-03**	**2,6**	**5,5E-04**	*0,0*	*9,0E-01*
O15164	**TRIM24**	*tripartite motif containing 24*	**3,0**	**3,8E-03**	**3,7**	**3,9E-04**	*0,7*	*5,8E-02*
Q13263	**TRIM28**	*tripartite motif containing 28*	*0,7*	*4,5E-02*	**1,1**	**6,2E-03**	*0,4*	*9,8E-02*
Q9UPN9	**TRIM33**	*tripartite motif containing 33*	*1,5*	*2,8E-02*	**2,0**	**8,4E-04**	*0,5*	*3,2E-01*
Q9UI30	**TRMT112**	*tRNA methyltransferase 11-2 homolog*	**3,5**	**2,3E-03**	*1,8*	*4,5E-02*	*-1,7*	*1,0E-01*
Q13885	**TUBB2A**	*tubulin beta 2A class IIa*	*0,8*	*5,9E-02*	**1,1**	**6,1E-03**	*0,3*	*2,0E-01*
Q9BUF5	**TUBB6**	*tubulin beta 6 class V*	*1,2*	*8,6E-02*	**2,0**	**5,3E-03**	*0,8*	*3,4E-03*
Q8TBC4	**UBA3**	*ubiquitin like modifier activating enzyme 3*	*1,4*	*2,9E-02*	**2,7**	**7,2E-04**	**1,4**	**3,0E-03**
O94888	**UBXN7**	*UBX domain protein 7*	*1,4*	*7,5E-02*	**2,7**	**2,5E-03**	**1,3**	**3,5E-02**
Q92890	**UFD1L**	*ubiquitin recognition factor in ER associated degradation 1*	**1,3**	**6,6E-03**	**1,2**	**1,4E-03**	*-0,1*	*4,6E-01*
Q86UV5	**USP48**	*ubiquitin specific peptidase 48*	*3,7*	*6,5E-02*	**5,0**	**7,5E-04**	*1,4*	*3,0E-01*
Q9BVJ6	**UTP14A**	*UTP14, U3 small nucleolar ribonucleoprotein, homolog A*	*0,9*	*1,7E-01*	**1,5**	**9,3E-03**	*0,6*	*1,9E-01*
Q99986	**VRK1**	*vaccinia related kinase 1*	*-0,5*	*4,8E-01*	**2,0**	**4,9E-03**	**2,5**	**1,2E-02**
Q7Z5K2	**WAPAL**	*wings apart like protein homolog*	*1,9*	*1,1E-02*	**2,4**	**1,4E-03**	*0,5*	*1,0E-01*
O75554	**WBP4**	*WW domain binding protein 4*	*1,9*	*9,6E-02*	**2,9**	**8,0E-03**	*1,0*	*4,1E-02*
Q8NI36	**WDR36**	*WD repeat domain 36*	*0,9*	*8,4E-02*	**1,8**	**5,4E-04**	*0,8*	*9,4E-02*
Q9NW82	**WDR70**	*WD repeat domain 70*	**1,6**	**3,6E-03**	**2,1**	**2,0E-04**	*0,5*	*1,7E-02*
Q9BQA1	**WDR77**	*WD repeat domain 77*	*5,6*	*2,0E-02*	**6,0**	**3,3E-03**	*0,4*	*3,0E-02*
O96028	**WHSC1**	*Wolf-Hirschhorn syndrome candidate 1*	**4,2**	**2,8E-03**	**4,2**	**2,9E-04**	*0,0*	*8,7E-01*
Q9Y4P8	**WIPI2**	*WD repeat domain, phosphoinositide interacting 2*	*2,5*	*3,2E-02*	**5,1**	**2,4E-04**	**2,6**	**1,6E-02**
O95785	**WIZ**	*widely interspaced zinc finger motifs*	**7,1**	**4,3E-03**	**7,6**	**4,0E-04**	*0,5*	*2,0E-02*
P23025	**XPA**	*xeroderma pigmentosum complementation group A*	*2,2*	*3,1E-02*	**3,0**	**2,7E-04**	*0,8*	*2,0E-01*
P13010	**XRCC5**	*X-ray repair cross complementing 5*	*1,6*	*1,1E-02*	**1,6**	**1,7E-03**	*0,1*	*7,9E-01*
P12956	**XRCC6**	*X-ray repair cross complementing 6*	**1,3**	**1,2E-03**	**1,4**	**1,7E-04**	*0,0*	*7,9E-01*
Q7Z739	**YTHDF3**	*YTH N6-methyladenosine RNA binding protein 3*	**1,3**	**2,9E-03**	**1,3**	**3,0E-04**	*0,0*	*9,4E-01*
Q9ULJ3	**ZBTB21**	*zinc finger and BTB domain containing 21*	*2,6*	*2,0E-02*	**2,4**	**5,9E-03**	*-0,2*	*3,2E-01*
O75152	**ZC3H11A**	*zinc finger CCCH-type containing 11A*	**1,3**	**1,4E-03**	**1,2**	**2,1E-04**	*-0,1*	*2,5E-01*
Q96JP5	**ZFP91**	*ZFP91 zinc finger protein*	*2,2*	*1,1E-01*	**3,5**	**4,4E-03**	*1,2*	*1,7E-01*
Q969J2	**ZKSCAN4**	*zinc finger with KRAB and SCAN domains 4*	*2,6*	*2,8E-02*	**3,3**	**1,3E-03**	*0,7*	*2,0E-01*
Q9UBW7	**ZMYM2**	*zinc finger MYM-type containing 2*	*2,2*	*1,1E-02*	**3,4**	**4,1E-04**	**1,2**	**5,8E-03**
Q14202	**ZMYM3**	*zinc finger MYM-type containing 3*	*2,8*	*2,6E-02*	**2,8**	**6,3E-03**	*0,0*	*7,5E-01*

Uniprot ID	Gene names	Protein name	FB-EHMT2 Flag-only		FB-EHMT2 Flag-LT330		in LT330 vs. empty	
			log2fc	q-value	log2fc	q-value	log2fc	q-value
Q15697	**ZNF174**	*zinc finger protein 174*	*1,1*	*5,5E-02*	**1,5**	**6,8E-04**	*0,5*	*1,9E-01*
O43670	**ZNF207**	*zinc finger protein 207*	**1,4**	**7,4E-03**	**1,1**	**4,5E-03**	*-0,3*	*8,3E-02*
Q8ND82	**ZNF280C**	*zinc finger protein 280C*	*3,6*	*1,4E-02*	**4,1**	**1,8E-03**	*0,5*	*6,9E-02*
Q5T0B9	**ZNF362**	*zinc finger protein 362*	*2,7*	*3,8E-02*	**4,9**	**6,6E-04**	*2,2*	*8,0E-02*
Q8TF68	**ZNF384**	*zinc finger protein 384*	**6,0**	**7,9E-03**	**6,0**	**1,3E-03**	*0,0*	*8,0E-01*
Q96ME7	**ZNF512**	*zinc finger protein 512*	**3,0**	**1,9E-03**	**3,4**	**8,9E-05**	*0,4*	*1,0E-01*
Q96KM6	**ZNF512B**	*zinc finger protein 512B*	*4,5*	*1,6E-02*	**5,7**	**1,3E-03**	**1,2**	**1,3E-02**
Q9UEG4	**ZNF629**	*zinc finger protein 629*	*3,8*	*4,1E-02*	**4,2**	**6,4E-03**	*0,4*	*3,4E-01*
Q14966	**ZNF638**	*zinc finger protein 638*	*0,8*	*4,7E-02*	**1,6**	**2,3E-03**	*0,8*	*1,1E-01*
Q9H582	**ZNF644**	*zinc finger protein 644*	*1,9*	*6,1E-02*	**3,2**	**5,3E-03**	**1,3**	**4,2E-02**
P17098	**ZNF8**	*zinc finger protein 8*	*4,4*	*2,3E-02*	**4,4**	**4,7E-03**	*0,1*	*7,7E-01*
Q96NB3	**ZNF830**	*zinc finger protein 830*	*1,0*	*1,9E-01*	**3,4**	**4,2E-04**	**2,4**	**1,2E-02**

viruses

MDPI

Article

Borna Disease Virus 1 Phosphoprotein Forms a Tetramer and Interacts with Host Factors Involved in DNA Double-Strand Break Repair and mRNA Processing

Nicolas Tarbouriech [1], Florian Chenavier [1,†], Junna Kawasaki [2,3,†], Kamel Bachiri [4,†], Jean-Marie Bourhis [1], Pierre Legrand [5], Lily L. Freslon [1], Estelle M. N. Laurent [4], Elsa Suberbielle [6], Rob W. H. Ruigrok [1], Keizo Tomonaga [2,3,7], Daniel Gonzalez-Dunia [6], Masayuki Horie [8,9], Etienne Coyaud [4] and Thibaut Crépin [1,*]

1 Institut de Biologie Structurale (IBS), CEA, CNRS, Université Grenoble Alpes, 38058 Grenoble, France
2 Laboratory of RNA Viruses, Department of Virus Research, Institute for Life and Medical Sciences, Kyoto University, Kyoto 606-8507, Japan
3 Laboratory of RNA Viruses, Department of Mammalian Regulatory Network, Graduate School of Biostudies, Kyoto University, Kyoto 606-8507, Japan
4 Department of Biology, Univ. Lille, Inserm, CHU Lille, U1192—Protéomique Réponse Inflammatoire Spectrométrie de Masse—PRISM, 59000 Lille, France
5 Synchrotron SOLEIL, L'Orme des Merisiers, 91192 Gif-sur-Yvette, France
6 Institut Toulousain des Maladies Infectieuses et Inflammatoires (Infinity), Université de Toulouse, Inserm, CNRS, UPS, 31024 Toulouse, France
7 Department of Molecular Virology, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan
8 Laboratory of Veterinary Microbiology, Graduate School of Veterinary Science, Osaka Metropolitan University, Izumisano 598-8531, Japan
9 Osaka International Research Center for Infectious Diseases, Osaka Metropolitan University, Izumisano 598-8531, Japan
* Correspondence: thibaut.crepin@ibs.fr
† These authors contributed equally to this work.

check for updates

Citation: Tarbouriech, N.; Chenavier, F.; Kawasaki, J.; Bachiri, K.; Bourhis, J.-M.; Legrand, P.; Freslon, L.L.; Laurent, E.M.N.; Suberbielle, E.; Ruigrok, R.W.H.; et al. Borna Disease Virus 1 Phosphoprotein Forms a Tetramer and Interacts with Host Factors Involved in DNA Double-Strand Break Repair and mRNA Processing. *Viruses* **2022**, *14*, 2358. https://doi.org/10.3390/v14112358

Academic Editors: Kore Schlottau and Dennis Rubbenstroth

Received: 23 September 2022
Accepted: 24 October 2022
Published: 26 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract: Determining the structural organisation of viral replication complexes and unravelling the impact of infection on cellular homeostasis represent important challenges in virology. This may prove particularly useful when confronted with viruses that pose a significant threat to human health, that appear unique within their family, or for which knowledge is scarce. Among *Mononegavirales*, bornaviruses (family *Bornaviridae*) stand out due to their compact genomes and their nuclear localisation for replication. The recent recognition of the zoonotic potential of several orthobornaviruses has sparked a surge of interest in improving our knowledge on this viral family. In this work, we provide a complete analysis of the structural organisation of Borna disease virus 1 (BoDV-1) phosphoprotein (P), an important cofactor for polymerase activity. Using X-ray diffusion and diffraction experiments, we revealed that BoDV-1 P adopts a long coiled-coil α-helical structure split into two parts by an original β-strand twist motif, which is highly conserved across the members of whole *Orthobornavirus* genus and may regulate viral replication. In parallel, we used BioID to determine the proximal interactome of P in living cells. We confirmed previously known interactors and identified novel proteins linked to several biological processes such as DNA repair or mRNA metabolism. Altogether, our study provides important structure/function cues, which may improve our understanding of BoDV-1 pathogenesis.

Keywords: *Bornaviridae*; phosphoprotein; structure; interactomics

1. Introduction

Nonsegmented negative-strand RNA viruses (or *Mononegavirales*) are a class of viruses that infect plants and animals, with many of them causing significant diseases or even death in humans. These enveloped viruses share common strategies in viral replication

and/or gene expression [1]. Their genome, encapsulated by the nucleoprotein (N), is the biologically active template for both transcription and replication by the polymerase complex [2] resulting from the association of the viral RNA-dependent RNA polymerase (RdRp or L) with its essential cofactor, phosphoprotein (P). P is a highly versatile protein involved in numerous functions, with a high degree of structural flexibility. Through its modular architecture consisting of long disordered regions that alternate with structured domains, P interacts with the main proteins of the viral RNA synthesis machine. It self-associates in oligomers through a central oligomerisation domain (P_{OD}) [3–12] and bridges the RdRp [13–19] and the nucleocapsid [20]. It also chaperones N to maintain it in an RNA-free form needed for replication [21–28]. In addition, P is an essential actor of viral replication by recruiting a plethora of cellular partners with proviral activities [10,29–40]. The disordered regions of P seem to be of primary importance for these additional functions.

Bornaviruses (family *Bornaviridae*) occupy a special place within *Mononegavirales*. They infect the widest range of species of the animal kingdom, ranging from fishes to mammals [41–43]. While most mononegaviruses mainly perform cytoplasmic replication, bornaviruses replicate in the nucleus, where they assemble viral factories [44] designated as "viral Speckles Of Transcripts" (vSPOTs). It has been recently shown that some mammalian orthobornaviruses (*e.g.*, Borna disease virus 1 (BoDV-1) and variegated squirrel bornavirus 1 (VSBV-1)) can infect humans, predominantly affecting neuronal tissues and causing fatal encephalitis [45–47]. The genome of bornaviriruses consists of approximately 8900 nucleotides and represents the smallest known genome amongst *Mononegavirales*. BoDV-1 vSPOTs formed by P protein-driven liquid–liquid phase separation [48], are present in the nucleus, in close interaction with the chromatin, where they dock on neuronal DNA double-strand breaks (DSBs), thereby affecting neuronal epigenetics and activity [49]. In particular, both N and P seem to modulate epigenetic signalling in neurons [50,51].

BoDV-1 replicative machinery, a complex between the RdRp (L) and P, produces different mRNAs from the genome via polar sequential transcription, encoding for a total of six viral proteins. Except for the RNA-free N and the matrix protein [52,53], nothing is known concerning the structure of these viral proteins. Here, we used X-ray diffusion and diffraction experiments to provide a complete structural analysis of BoDV-1 P. We observed that P possesses a modular architecture, consisting of a central oligomerisation domain (P_{OD}) surrounded by two disordered regions, similar to other *Mononegavirales* Ps. We also showed that BoDV-1 P_{OD} has the most elongated multimerisation domain of all mononegaviruses, with one monomer composed of two consecutive α-helices linked by a short three-amino acid stretch, making a 140 Å long coiled-coil α-helical structure split into two parts by a β-strand twist motif. This particular motif appears to be important for viral replication and is highly conserved across the members of the whole *Orthobornavirus* genus. Moreover, we determined the BoDV-1 P proximal interactome in human cells, confirming previous interactors and identifying novel proteins linked to DNA breaks and mRNA metabolism. Interestingly, the strong overlap between the proximal interactomes in living cells of N- and C-terminal biotin ligase ($BirA_{R118G}$, or BirA*)-tagged BoDV-1 P data in living cell suggests that this structure is highly flexible at the β-strand twist motif and/or that the disordered domains confer enough freedom of movement to the BirA* moiety to biotinylate a highly similar protein environment regardless of its position in the primary protein sequence. Altogether, our results provide important advances in our understanding of the structural organisation of BoDV-1 P protein and strongly support the relevance of our structural model in a living cellular context.

2. Materials and Methods

2.1. Cloning and Plasmid Preparation

The bacterial expression-optimised DNA coding sequence of Borna disease virus 1 (BoDV-1) P was purchased from GeneArt (Thermofischer Scientific, Regensburg, Germany). Primers were designed for the amplification of the different constructs of BoDV-1 (residues 1 to 201 and 65 to 172), munia bornavirus 1 (MuBV-1; residues 65 to 172), and Gaboon viper

virus 1 (GaVV-1; residues 67 to 178). Each DNA fragment was cloned into pETM11 bacterial expression vector (EMBL) using 5′ end *Nco1* and 3′ end *Xho1* restriction sites introduced by PCR amplification. All constructs were expressed as N-terminally TEV-removable His-tagged proteins. Sequencing verifications were performed by Eurofins Genomics.

2.2. Protein Expression and Protein Purification

Plasmids with the corresponding constructs were transformed into *Escherichia coli* BL21 RIL (DE3) cells (Life Technologies, Thermo Fischer Scientific, Villebon-sur-Yvette, France). LB cultures were induced for 12 h with 0.3 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 18 °C, collected by centrifugation, and stored at −80 °C. Briefly, Bacterial pellets were resuspended in 50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 1 mM β-mercaptoethanol (β-ME), and cOmplete^{TM} ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor cocktail (Roche, Meylan, France). After lysis by sonication and centrifugation, the supernatant was complemented with 15 mM imidazole and the recombinant proteins were purified by nickel affinity chromatography (resin Ni-NTA; Qiagen, Courtabœuf, France). Based on SDS-PAGE analysis, fractions of interest were pooled together before the addition of 1% tobacco etch virus (TEV) protease and dialysed overnight in 20 mM HEPES (pH 7.5), 200 mM NaCl, and 1 mM β-ME. A second nickel affinity chromatography was performed to remove both the uncleaved protein of interest and TEV protease before the size-exclusion chromatography (SEC). The flow-through was concentrated by centrifugation (Millipore-Amicon Ultra centrifugal filters 10 K) and injected into a HiLoad^{TM} S200 16/600 column (Cytiva, Velizy-Villacoublay, France) with a NGC system (Bio-Rad, Marnes-La-Coquette, France). All proteins were in the same final buffer composed of 20 mM HEPES (pH 7.5), 150 mM NaCl, and 1mM β-ME.

2.3. Crystallisation and X-ray Structure Determination

Crystallisation experiments were initiated using the HTX platform (EMBL, Grenoble, France) before a final manual screening to obtain single crystals. BoDV-1 P_{OD} crystals were obtained from a 10 mg·mL^{-1} protein solution using a precipitant composed of 100 mM HEPES (pH 7.5), 14–18% polyethylene glycol (PEG) 8K, and 200 mM $CaCl_2$. MuBV-1 P_{OD} crystals were obtained from a 4 mg·mL^{-1} protein solution using a precipitant composed of 15–18% PEG 3350, 150 mM $Mg(NO^3)^2$, and 10 mM phenol. GaVV-1 P_{OD} crystals were obtained from a 5 mg·mL^{-1} protein solution using a precipitant composed of 100 mM HEPES (pH 7.5), 18–22% PEG 3350, and 200 mM NaSCN. Crystals were cryo-protected in their mother liquor solution containing 20% glycerol, mounted on cryoloops (Hampton Research, Aliso Viejo, CA, USA), and frozen in liquid nitrogen.

BoDV-1 P_{OD} diffraction experiments were undertaken on ID-29 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). Data were collected using a Dectris Pilatus 6M detector. Indexing and integration were performed using the XDS program suite [54,55]. Thereafter, data were processed using STARANISO [56], as implemented in autoPROC [57], which applies non-elliptical anisotropic limits based on a locally averaged mean I/σ(I) cut-off, performs a Bayesian estimation of structure amplitudes, and applies an anisotropic correction to the data. Detailed crystallographic statistics are provided in Table S1. Molecular replacement solutions were obtained with Phaser [58] using the measles virus P_{OD} structure [6]. The initial solution was then used through repeated cycles of autobuilding using Autobuild [59] within the Phenix suite [60]. The final building phase used both manual model building with COOT [61], together with Autobuild cycles and ArpWarp [62,63]. Refinement of the model was carried out using Refmac5 [64].

MuBV-1 P_{OD} diffraction experiments were undertaken on PROXIMA-1 at the SOLEIL Synchrotron (Saint-Aubin, France). Data were collected using a Dectris Eiger-X 16M detector. Indexing and integration were performed using the XDS program suite. Molecular replacement solutions were obtained with Phaser using the BoDV-1 P_{OD} tetrameric structure separated into two ensembles: one containing residues 86 to 125 and the other

containing residues 127 to 166. The final building phase used both a manual model building with COOT and refinement of the model using Refmac5.

GaVV-1 POD diffraction experiments were undertaken on PROXIMA-1 for the native dataset and on PROXIMA-2A for the Se-SAD dataset at the SOLEIL Synchrotron. PROXIMA-2A is equipped with a Dectris Eiger-X 9M detector. Indexing and integration were performed using the XDS program suite. Heavy atoms sites were determined using the SHELX suite [65], their positions were refined using PHASER in EP mode, and the phases were improved using PARROT [66] for density modification. The resulting electron density allowed the positioning of an AlphaFold2 [67] model, which was improved and refined against the native dataset with COOT and Refmac5.

The figures were drawn using PyMOL [68].

2.4. Small-Angle X-ray Scattering (SAXS) Analysis

All datasets were collected on the SWING beamline at the SOLEIL synchrotron. The collection parameters are listed in Table S2. The samples were loaded onto a Superdex^{TM} 200 increase 5/150 GL (Cytiva) SEC column and SAXS measurements were performed throughout elution, operating at 20 °C with a flow rate of 0.3 mL·min^{-1} and a mobile phase containing 20 mM HEPES (pH 7.5), 150 mM NaCl, and 5 mM β-ME. The data were integrated and subtracted with Foxtrot [69], providing the first estimation of I(0) and Rg for each frame of the SEC profile. This allowed the identification along the elution peak of the frames with constant Rg values. The corresponding frames were merged into a final scatter curve used for further analysis. Final Rg values were determined using ATSAS [70] and RAW [71]. Dmax values were determined using GNOM [72] in the ATSAS suite and compared to the values obtained with BIFT [73] implemented in RAW. Molecular weights of the proteins were estimated by using the Qr ratio derived from the invariant volume of correlation [74].

2.5. Ab Initio Molecular Shape Analysis of P_{OD}

Ab initio shape analyses were performed using DAMMIF [75] from the ATSAS package. Twenty independent DAMMIF calculations were performed both with and without four-fold (P4) symmetry. In each case, to create the most probable molecular shape, an average shape was determined with DAMAVER [76], with and without P4 symmetry. The filtered models were then reprocessed against the experimental data using DAMMIN [77].

2.6. Full Atom Modelling of BoDV-1 P_{FULL}

An initial pool of 10,000 random conformation models of BoDV P_{FULL} was generated using the Ensemble Optimisation Method (EOM 2.1 [78]). The models generated for the disordered N-terminal (residues 1 to 71) and C-terminal segments (residues 167 to 201) were attached to the P_{OD} X-ray structure. As the disordered parts generated by EOM are Cα, we used an in-house procedure to obtain the full atom structures of the pool. First, PD2 ca2main [79] was used to reconstruct the backbone, followed by SCWRL4 [80] to add the side chains. This pool was then used to select a final ensemble with the EOM algorithm based on the experimental scattering curve. This process was repeated three times.

2.7. Minireplicon Assays

Minireplicon assays were conducted as described previously [81]. Briefly, 293T cells seeded onto 48-well plates were transfected with 50 ng of minigenome plasmid containing Gaussia luciferase as a reporter gene [82], 50 ng of pCAGGS-N, 5 ng of pCAGGS-P (or point-mutated P), and 50 ng of pCAGGS-L using TransIT293 (Mirus) according to the manufacturer's instructions. The culture supernatants were collected at 72 h post transfection and subjected to luciferase assay using a Biolux Gaussia luciferase assay kit (New England BioLabs, Evry, France) according to the manufacturer's instructions. The assays were performed in technical duplicates for each experiment.

2.8. Determination of Proximal Host–Viral Protein Interactions in Living Cells (BioID)

These experiments were performed as in [83]. Briefly, Flp-In™ T-REx™ HEK293 cells (Life Technologies) cells were co-transfected with pOG44 (encoding the Flp recombinase) and pcDNA5 FRT/TO FlagBirA*- or BirA*Flag P_{FULL} (N- and C-ter tag, respectively) of BoDV-1. Cells were selected with Hygromycin B to generate new stable tetracycline-inducible cell lines expressing single copies of N- and C-terminal tagged P_{FULL}-BirA* fusion proteins. Cultures were expanded to ~100 million cells per sample, before being placed into a medium of 1 $\mu g \cdot mL^{-1}$ tetracycline to induce bait protein expression, along with 50 µM biotin. After 24 h, induction cells were collected, washed twice with PBS, spun down, and stored at −80 °C. Each pellet was resuspended in a 5 mL of modified RIPA buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.1% SDS, 1:500 protease inhibitor cocktail (Sigma-Aldrich, Saint-Quentin Fallavier, France), and 1:1000 Turbonuclease (BPS Bioscience, Nanterre, France)), incubated for one hour on an end-over-end rotator at 4 °C, sonicated until no visible aggregates remained, and centrifuged at 45,000× *g* for 20 min. Cleared supernatants containing all compartment solubilized proteins were transferred to new tubes and incubated with 30 µL of packed and pre-equilibrated streptavidin resin (Ultralink, Pierce, Thermo Fischer Scientific, Villebon-sur-Yvette, France; 2 h under rotation at 4 °C) to capture biotinylated species. Beads were pelleted (300 g, 1 min) and washed six times with 50 mM ammonium bicarbonate (AB). Trypsin (1 µg) was added to each 200 µL sample to digest on beads (overnight at 37 °C). Tryptic peptides were collected in the supernatant, lyophilized, and desalted on C18 ziptips. One-quarter of each sample was analysed by nLC MS-MS (Easy1000 in line with a Thermo Q-Exactive; see Supplemental Methods). The MS raw files were then analysed using MaxQuant (version 1.5.8.3) against the UniProt Reference database to generate the protein identification list with a 1% FDR. High confidence proximal interactors were defined using Perseus (version 1.6.15.0) with the following filters for each given ID: peptides detected in 3/3 bait protein runs; an average Log_2 LFQ abundance against average top 3/20 control over 1; and a corresponding q-value < 0.01 (Table S3).

3. Results

3.1. X-ray Structure of the Oligomerisation Domain of BoDV-1 Phosphoprotein

With *Mononegavirales* Ps being known as multimodular nonglobular molecules, we first performed a bioinformatic analysis of the BoDV-1 P amino acid sequence using 16 web servers, all of them predicting the location of disordered regions and secondary structure elements. We then combined all of the results in a single representation (D-score; Figure 1a). On this graph, residues with a D-score <0.50 were assigned as disordered, while residues with a D-score >0.5 were assigned as structured. We observed that BoDV-1 P presented an organisation similar to other *Mononegavirales* Ps, with central folded domains flanked by N- and C-terminal disordered regions. Structure prediction using AlphaFold2 is compatible with the D-score analysis (Figure S11). Based on this prediction, two constructs were designed (Figure 1b): the full-length P (P_{FULL}) and the central domain (from residue 65 to 172) that we hypothesized to be the oligomerisation domain (P_{OD}), even if its size represented nearly half of the protein length and was much longer than P_{OD} from other viruses.

BoDV-1 P_{OD} crystals (fine plates growing in urchins) were obtained in PEG 8000, only in the presence of calcium chloride. Crystals belonged to space group $P42_12$, with cell parameters a = b = 35 Å and c = 166 Å. The analysis of the asymmetric unit indicated that it could only contain a single-amino acid chain (Matthew coefficient of 2.1 $Å^3 \cdot Da^{-1}$ with a solvent content of 42% for a 108 amino acid-long chain), meaning that it was likely to contain a single narrow elongated molecule. This informed our choice to use a single coiled-coil chain of the measles virus tetrameric P_{OD} [6] as a starting model to solve the structure by molecular replacement. An initial solution was obtained from Phaser [58] and improved using Autobuild [59], before being classically built and refined. The final structure is shown in Figure 2a.

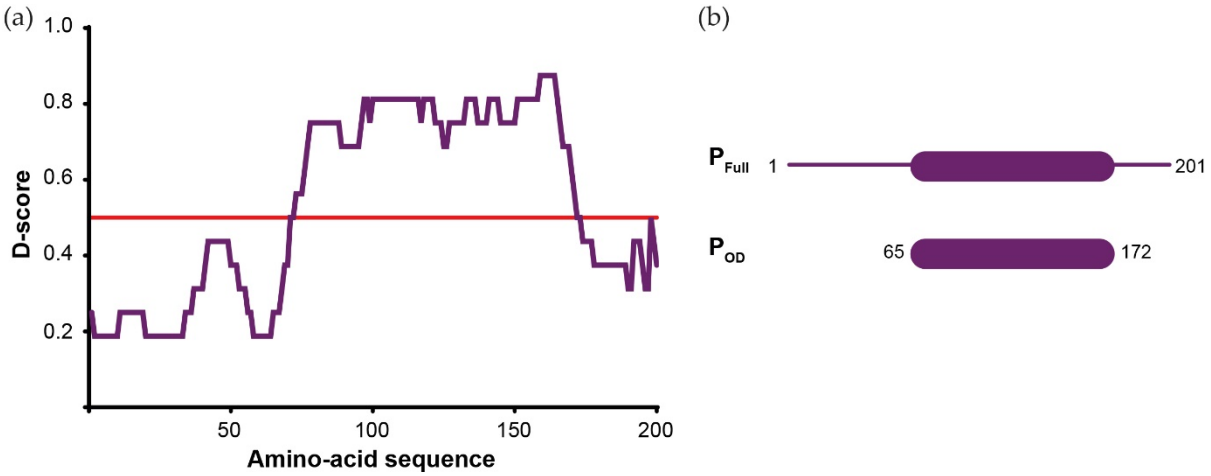


Figure 1. Predicted architecture of BoDV-1 phosphoprotein. (**a**) D-score (score for disorder as a function of residue) of BoDV-1 P. The prediction was based on 16 predictor web servers and the D-score was calculated by adding the values for each residue and dividing by the number of used algorithms. We arbitrarily defined a threshold level at 0.50. Residues with a D-score <0.50 were assigned as disordered [84]. (**b**) Scheme of the BoDV-1 P constructs designed for this work based on the bioinformatics analysis.

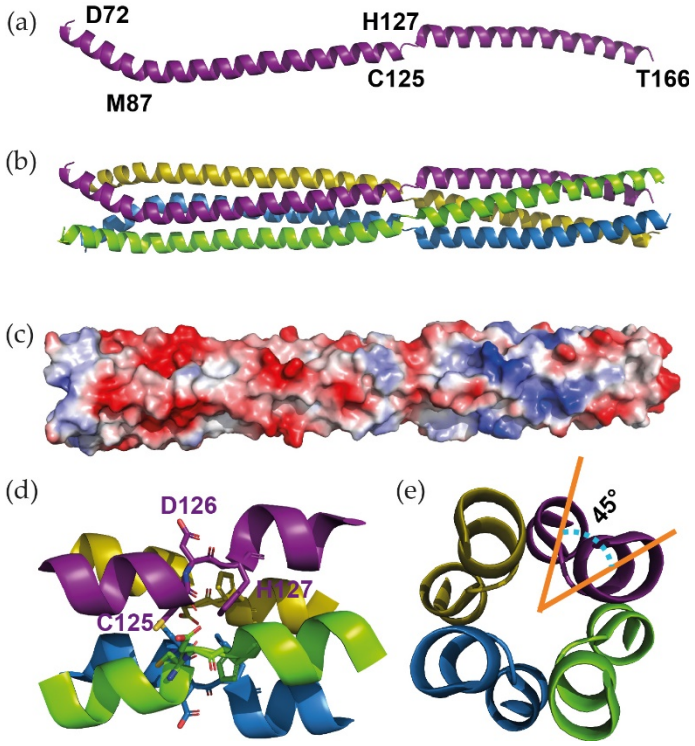


Figure 2. X-ray structure of BoDV-1 P_{OD}. (**a**) Structure of BoDV-1 monomeric P_{OD} as obtained in the asymmetric unit. (**b**) Structure of the tetrameric P_{OD} after applying the inherent symmetry operations of the tetragonal space group. The four parallel monomers are coloured in purple, yellow, blue, and green, respectively. (**c**) Electrostatic surface potential of P_{OD}. The potential scale ranges from −5 kT/e (red) to 5 kT/e (blue); the orientation is as in (**b**). (**d**) Detail of the β-strand twist motif that separates the two α-helical coiled-coils. (**e**) Orthogonal view of the β-strand twist motif as shown in (**d**), showing the 45° rotation of the helical axis around the super helical axis.

The monomeric P_{OD} is composed of two consecutive aligned α-helices (from residue E73 to Q123 and S128 to A164, respectively) linked by a short stretch of three residues,

making a ≈ 140 Å long molecule. As most *Mononegavirales* P assemble in oligomers through their central domain, we analysed P_{OD} by SAXS (Table S2) in order to compare its behaviour in the crystal and in solution. With a Dmax of 170 ± 9 Å, it confirmed one dimension of the protein. Furthermore, the volume of correlation (Vc) determination method [74] estimated a Mw of 40 kDa, corresponding to a tetramer (discussed later in Section 3.3). A tetrameric P_{OD} could then be reconstituted by applying the inherent 4-fold symmetry of this tetragonal space group (Figure 2b). It was then composed of two α-helical coiled-coils with all the monomers being parallel, with a break centred at position 126 in the middle of the whole structure (Figure 2d and Figure S9). Considering the electrostatic surface of the domain, the break seemed to separate the N-terminal acidic region from the C-terminal more neutral region (Figure 2c). A careful analysis of the Crick parameters [85,86] using CCCP [87] and TWISTER [88] demonstrated a decrease in the helical and superhelical radius together with an increase in the rise per residue and a decrease in the periodicity. In addition, the Ramachandran plot showed that residue D126 fell into the β-area, whereas neighbouring residues (C125 and H127) were in intermediate positions, between the α- and the β-areas (Figure S1a). Altogether, our data point towards the formation of a β-strand twist motif within the helices, as described for trimeric oligomerisation domains [89]. As opposed to the trimers, the tetrameric motif of P_{OD} identified herein did not form regular β-sheet main chain hydrogen bonds, and the shift in the helical axis was limited to 45° (Figure 2e) instead of the 120° observed in trimers.

3.2. The β-Strand Twist of BoDV-1 P Is Important for Viral Replication

In order to assess the involvement of the β-strand twist motif in BoDV-1 replication, we performed minireplicon assays using a series of P mutants, each one bearing a single amino acid substitution at the twist or in its proximal regions (from position 124 to 129). 293T cells were then transfected with a minigenome plasmid containing the Gaussia luciferase as a reporter gene, together with plasmids encoding for N, L, and different versions of P. The Gaussia luciferase activity levels were then measured (Figure 3a). Correct expression of each P mutant was confirmed by western blotting (Figure 3b). Interestingly, substitutions by alanine of R124 and D129, two proximal residues of the motif, did not affect the polymerase activity. Rather, the activity was abolished or markedly reduced by mutating C125, D126, or H127. Note that the quantity of protein for this mutant, as well as for the second mutant tested for this position (*i.e.*, C125D), appeared strongly reduced in the western blot analysis. Altogether, these results suggest that the β-strand twist motif is important for BoDV-1 replication, notably concerning D126 and H127 residues, whereas C125 would have an effect on the stability and/or the folding of P. Cysteine residues are known to form disulfide bridges and stabilise the structuration of secreted proteins. Surprisingly for a nuclear protein, such a disulfide bridge has been previously proposed at position 125 of BoDV-1 P [90]. Our experimental conditions did not allow us to detect them, especially in the crystal structure. However, compared to D126, which is strictly conserved (Figure 4j and Figure S8), the amino acid in position 125 is a cysteine only in mammalian 1 orthobornaviruses (*i.e.*, BoDV-1 and -2).

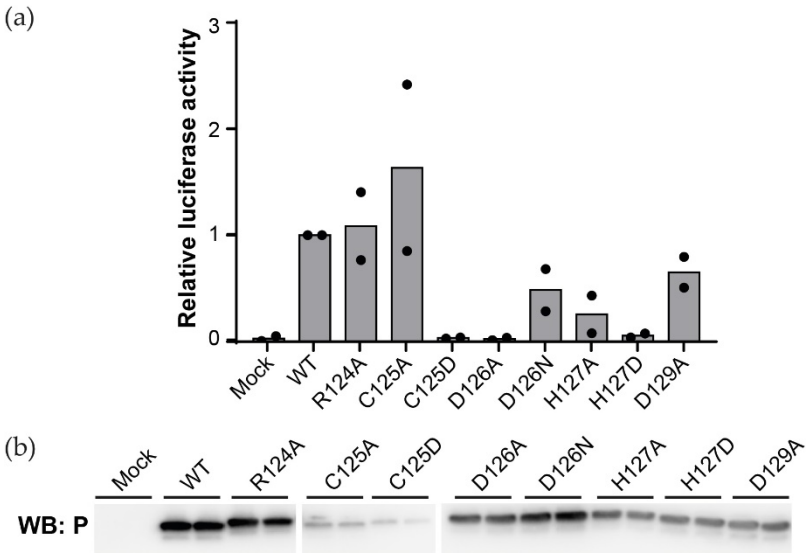


Figure 3. Mutations within the β-strand twist motif of BoDV-1 P modulates viral replication. (**a**) Minireplicon assay using a series of P mutants. The relative luciferase activities are shown as the ratios relative to wild-type P transfected cells (WT). The indicated BoDV-1 P mutants were used as helper plasmids. The dots show mean values (biological replicates; n = 2). (**b**) Detection of BoDV-1 P mutants by western blotting. The whole western blot is shown in Figure S7.

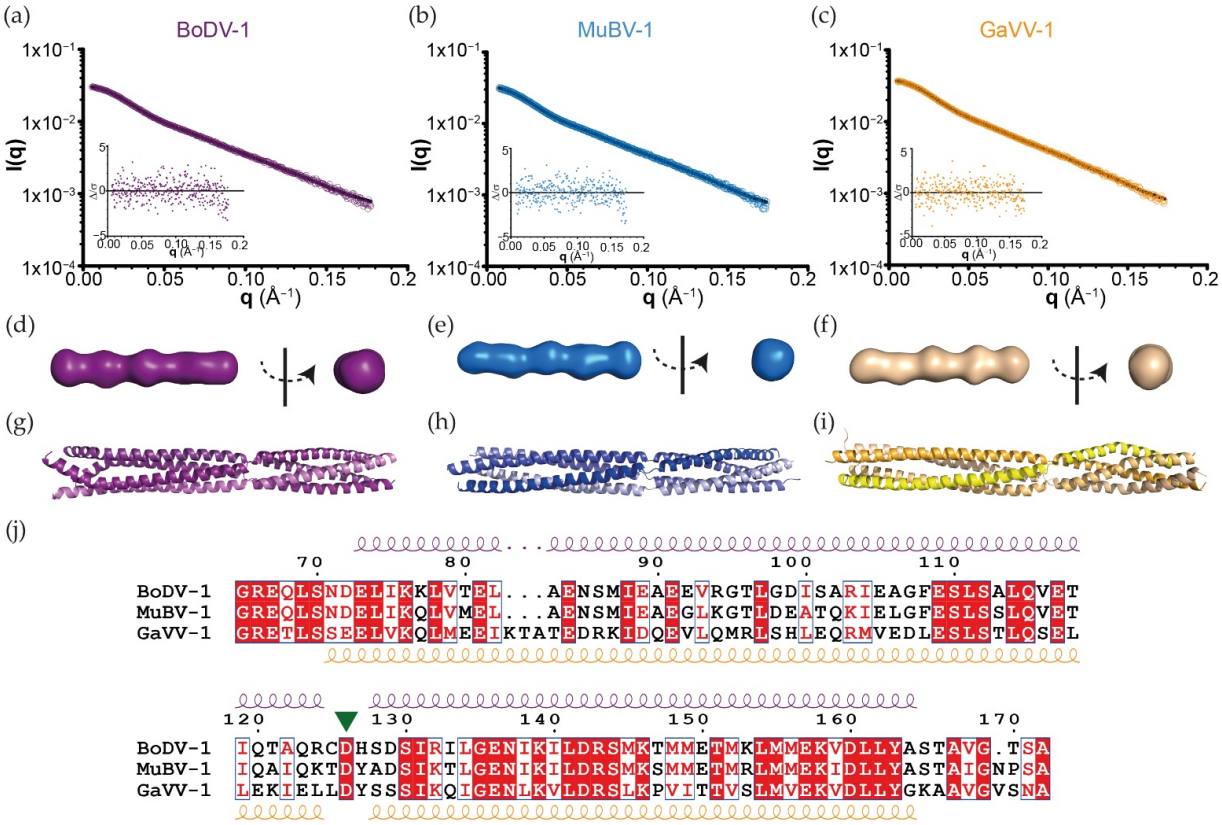


Figure 4. Structural analysis of orthobornaviral P_{OD}. Experimental SAXS diffusion curves (with residuals (small plots)) of (**a**) BoDV-1, (**b**) MuBV-1, and (**c**) GaVV-1 P_{OD} overlaid with the averaged SAXS curves (black) resulting from the *ab initio* models shown in panels **d**–**f**. Low-resolution envelops of (**d**) BoDV-1, (**e**) MuBV-1, and (**f**) GaVV-1 P_{OD} based on the *ab initio* modelling from the SAXS data. X-ray structure of (**g**) BoDV-1, (**h**) MuBV-1, and (**i**) GaVV-1 P_{OD} with the same orientation (*i.e.*, N-terminal on the left). The SAXS envelops (**d**–**f**) and the X-ray structures (**g**–**i**) are not at the same s-

cale. Figure S4 presents the dockings of the X-ray structures in the SAXS envelops. (**j**) Sequence alignment of the corresponding P_{OD} with the secondary structures of BoDV-1 and GaVV-1, shown above and below the sequence alignment, respectively. Residues in a solid red background are identical between the three sequences. The green triangle indicates the position of the conserved aspartate residue within the β-strand twist motif.

3.3. The β-Strand Twist Motif Is Conserved in Orthobornavirus P_{OD}

The comparison with tetrameric oligomerisation domains of P from other *Mononegavirales* (metapneumovirus, parainfluenza 5, measles, mumps, Nipah and murine respirovirus) showed that BoDV-1 P_{OD} was indeed the longest (the size of one monomer in the corresponding X-ray structures is 141, 100, and 93 Å long for BoDV-1, murine respirovirus, and measles P_{OD}, respectively). Above all, it is the only one with such a β-strand twist-like motif that breaks the continuity of its α-helical coiled-coil organisation. We therefore investigated its conservation in other members of the *Bornaviridae* family. Several P_{OD} were cloned, expressed, and purified. Crystals were obtained for the P_{OD} of munia bornavirus 1 (MuBV-1; 77% sequence identity with BoDV-1) and Gaboon viper virus 1 (GaVV-1; 41 and 43% sequence identity with BoDV-1 and MuBV-1, respectively). The complete SEC-SAXS and X-ray analysis of the three P_{OD} were performed (Figure 4).

In solution, the three P_{OD} presented the same behaviour. They were all eluted as a single straight symmetric peak from the SEC-SAXS column, without any aggregate signal. Using the Vc determination on the diffusion data, the MW were estimated between 40 and 45 kDa, 12–16% lower than the theoretical values (Table S2 and Figures S2–S4), all corresponding to tetramers. The Guinier transforms measured a hydrodynamic radius (Rg) between 44 and 46 Å and GNOM produced the pair distribution functions that fit with Dmax between 169 and 178 Å. Independently to the X-ray crystallography analysis, *ab initio* modelling was performed using twenty DAMMIF models averaged by DAMAVER. Averaged models' correlation with the diffusion curves were then checked using the damstart file as a starting envelope for DAMMIN (Figure 4a–c and Table S2). Each DAMAVER envelope appeared as an elongated tube (Figure 4d–f). In parallel, the X-ray structures of MuBV-1 and GaVV-1 were solved using different determination procedures (Table S1). The structure of MuBV-1 P_{OD} was solved by molecular replacement using BoDV-1 P_{OD} as a starting model, whereas GaVV-1 P_{OD} was obtained de novo using Se-Met derivative crystals. For both MuBV-1 and GaVV-1, the asymmetric units contained a tetrameric P_{OD}, whereas the tetramer for BoDV-1 could be obtained after applying the inherent symmetry operations of the tetragonal space group. Figure 4g–i shows the three biological P_{OD} (*i.e.*, as tetramers) in the same orientation. The three X-ray structures fit perfectly with the SAXS data. MuBV-1 and GaVV-1 P_{OD} presented similar architectures composed of two α-helical coiled-coils (the monomers being parallel) linked by a kink. Globally, the only differences concern the relative spatial position of the two coiled-coils, due to the inherent flexibility induced by such a break (Figure S10) and the crystal packing. MuBV-1 and GaVV-1 breaks were also composed of three residues, with a strictly conserved aspartate in the middle (Figure 4j and Figure S9). The Ramachandran analysis showed that all of them fall into the β-area, with neighbouring residues at intermediate positions (Figure S1); this confirms that this specificity is conserved in all Orthobornavirus P.

3.4. SAXS Analysis of Full-Length BoDV-1 Phosphoprotein

Having shown that P_{OD} assembles as a tetramer, we next analysed BoDV-1 P_{FULL} by SAXS in order to provide a complete view of the protein (Table S2). A P_{FULL} diffusion curve is shown in Figure 5a. Data processing (Figures 5b and S5) indicated that it is elongated for a 201 amino acid-long protein, with a Dmax value of 215 ± 10 Å, and Vc determination gives a value of 90 kDa, in accordance with a tetramer in solution, similar to the behaviour of P_{OD}. The main difference between P_{OD} and P_{FULL} became apparent when examining the Kratky plots. Indeed, while a nice bell-shaped (Gaussian) peak was observed for P_{OD}

(Figure S3b), we obtained a curve with a plateau that slowly decays for P_{FULL} (Figure 5c). These two profiles are characteristic of compact/globular and partially unfolded proteins, respectively, thereby validating our bioinformatic analysis. These data also indicate that the biological unit of BoDV-1 P forms a tetramer in solution, with a central domain responsible for this oligomerisation.

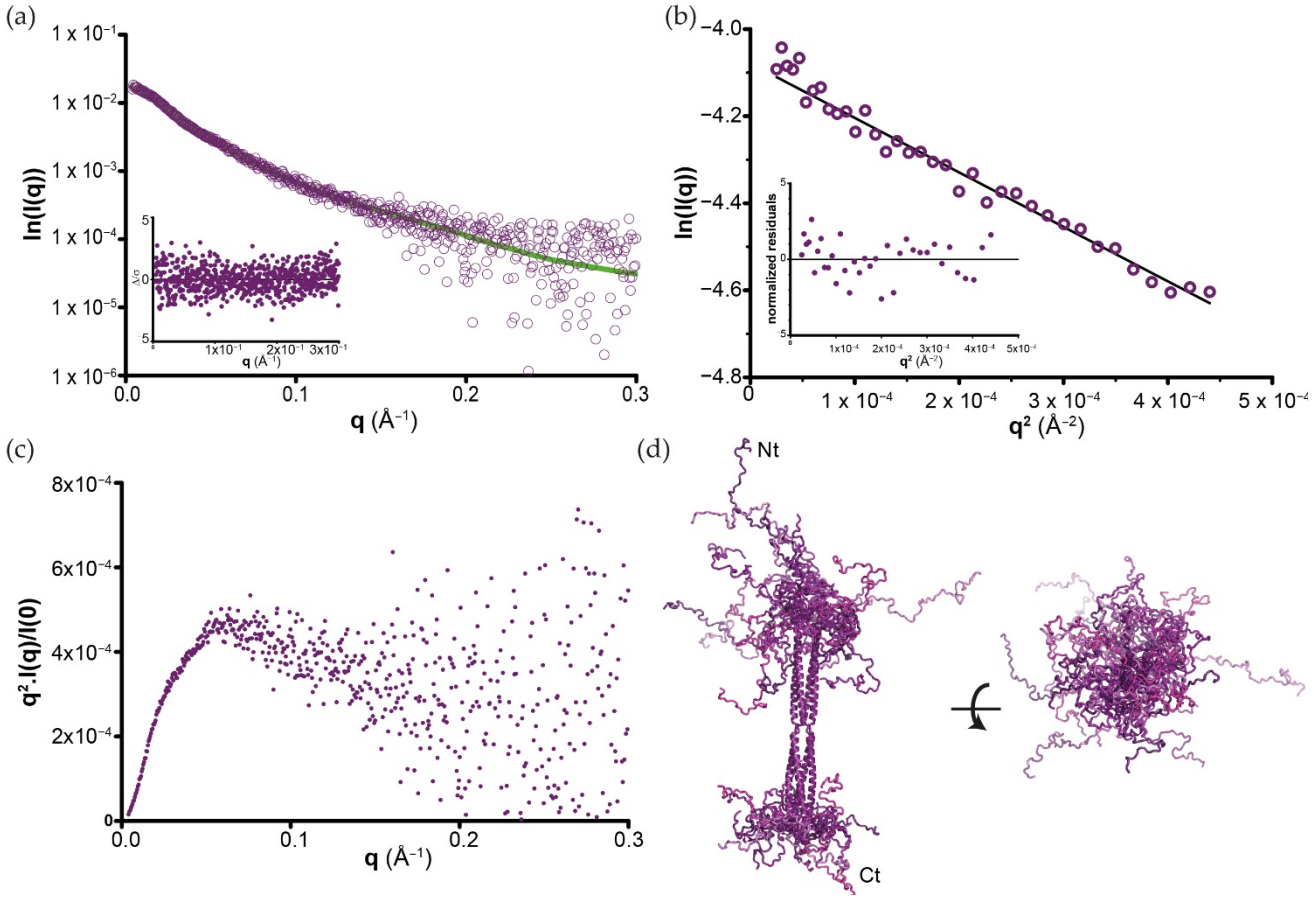


Figure 5. Flexibility analysis of BoDV-1 full-length P. (**a**) Experimental SAXS diffusion curve (deep-purple circles) of P_{FULL} overlaid with the averaged SAXS curves (green) calculated from the EOM ensemble shown in panel (**d**). (**b**) Guinier extrapolation for the determination of the Rg. The smallest plot corresponds to the normalised residual. (**c**) Kratky plot showing the partially unfolded propensity of P_{FULL} samples. (**d**) Conformational ensemble (6 of 10,000 conformers are shown; these models are perfectly representative of the whole population (Figure S6)) of P_{FULL} obtained from the modelling procedure based on EOM. P_{OD} X-ray structure is coloured deep purple, whereas the N- and C-terminal intrinsically disordered regions are coloured with six different shades of purples.

Being partially unfolded, the inherent flexibility of BoDV-1 P complicated the analysis of its structural organisation as a single particle. In order to progress in the complete structural analysis of P_{FULL}, we used a modelling procedure based on EOM [78] and the P_{OD} X-ray structure discussed in Section 3.2. A total of 10,000 Cα chains corresponding to the missing N- and C-terminal parts were selected to satisfy the SAXS data. Each protein backbone chain was reconstructed from these Cα models [79], and the missing side-chains were finally added with SCWRL4 [80]. Figure 5d shows an ensemble of 6 (out of the 10,000) independent models that fit the SAXS diffusion curves (black line in Figure 5a). This model revealed that despite its rigid central part, BoDV-1 P displays significant flexibility, as illustrated in the figure by the superimposition of the different conformers. However, transitory and/or stable small secondary-structure elements within both N- and C-terminal extremities cannot be excluded in our modelling.

3.5. Determination of BoDV-1 P–Host Proximal Interactomics in Living Cells

To gather insights in P_{FULL} behaviour in human living cells, we identified its proximity interactors using the BioID technique [91]. BioID uniquely provides information on physical and proximal protein-of-interest partners through the conjugation of biotin on the vicinal free ε-amine groups of surrounding proteins (within a 10 nm radius [92]). This covalent modification renders possible the thorough solubilisation of all cell compartments using harsh lysis conditions since protein interactions no longer need to be preserved during the purification process. Briefly, we performed BioID using stable and tetracycline-inducible N- and C-terminal BirA* tagged P_{FULL} HEK293 TREx Flp-In cells and purified the biotinylated proteins on streptavidin columns.

We then identified the proximal interactomes by nLC-MS/MS. Following background removal using a set of FlagBirA* alone samples (localising both in the cytoplasm and the nucleus [93]), we identified a total of 207 high-confidence interactors (Table S3a,b; 147 with FlagBirA*-P_{FULL} and 154 with P_{FULL}-BirA*Flag). As expected, we identified previously reported P_{FULL} interactions (*e.g.*, HMGB1 [44] and DNA break repair machineries), but the vast majority of our dataset was novel and links P_{FULL} to new mechanisms/complexes. In the set of more than 207 interactors (either nuclear or cytoplasmic), 123 were strictly nuclear, 55 could be in both compartments, and 6 were cytoplasmic (Table S3c). These findings are in agreement with the intrinsic ability of P_{FULL} to localise in the nucleus. The relevance of the non-nuclear interactors identified here is subject to caution and could be model-dependent, as discussed hereafter. Word clouds of Gene Ontology enrichment analyses indicated that P could associate with diverse nuclear subcompartments (Figure 6a; *e.g.*, telomeres, nuclear speck, kinetochores) and with factors involved in transcription and DNA repair processes (Figure 6b). Further exploring our data, P_{FULL} BioID identified 44 chromatin-associated partners, with 10 and 22 of them being linked to either replication forks or double-strand break repair, respectively (Table S3d; Figure 6c). P_{FULL} BioID also detected 52 partners linked to RNA metabolism, including 13 components of the U2-type spliceosomal complex and 6 members involved in mRNA polyadenylation. A comparison between N- and C-terminal BirA* tagged P_{FULL} (Table S3a,e; Figure 6d) revealed that N-terminal BirA* preferentially identified the HMGB1 and HMGB2 proteins, with HMGB3 being just below the threshold ($p = 0.014$) [94,95], as well as dynactin proteins (DCTN2 and DCTN4). This observation suggests two possibilities: (i) HMGB1/2 and DCTN2/4 proteins interact at the N-terminal region of the BoDV-1 phosphoprotein, with a putative structural "locking" of the twist or of the N-terminal disordered region; (ii) C-terminal tagging of P could hinder these interactions. Conversely, C-terminal tagged P_{FULL} preferentially identified two components of the Commander complex (COMMD3 and CCDC22), which are involved in the regulation of NFκB signalling regulation [96]. Similarly, it could indicate an association of these factors with the C-terminal region of P_{FULL}, potentially concomitant with an induced structural stiffness, or that the N-terminal tagging impairs their binding or proximity to P_{FULL}. Finally, several additional proximal partners were associated with the regulation of innate immune responses (*i.e.*, PQBP1 [97] and its paralog CACTIN [98], HEXIM1 [99], and METTL3 [100]), suggesting a putative function of P in BoDV-1-driven innate immune escape.

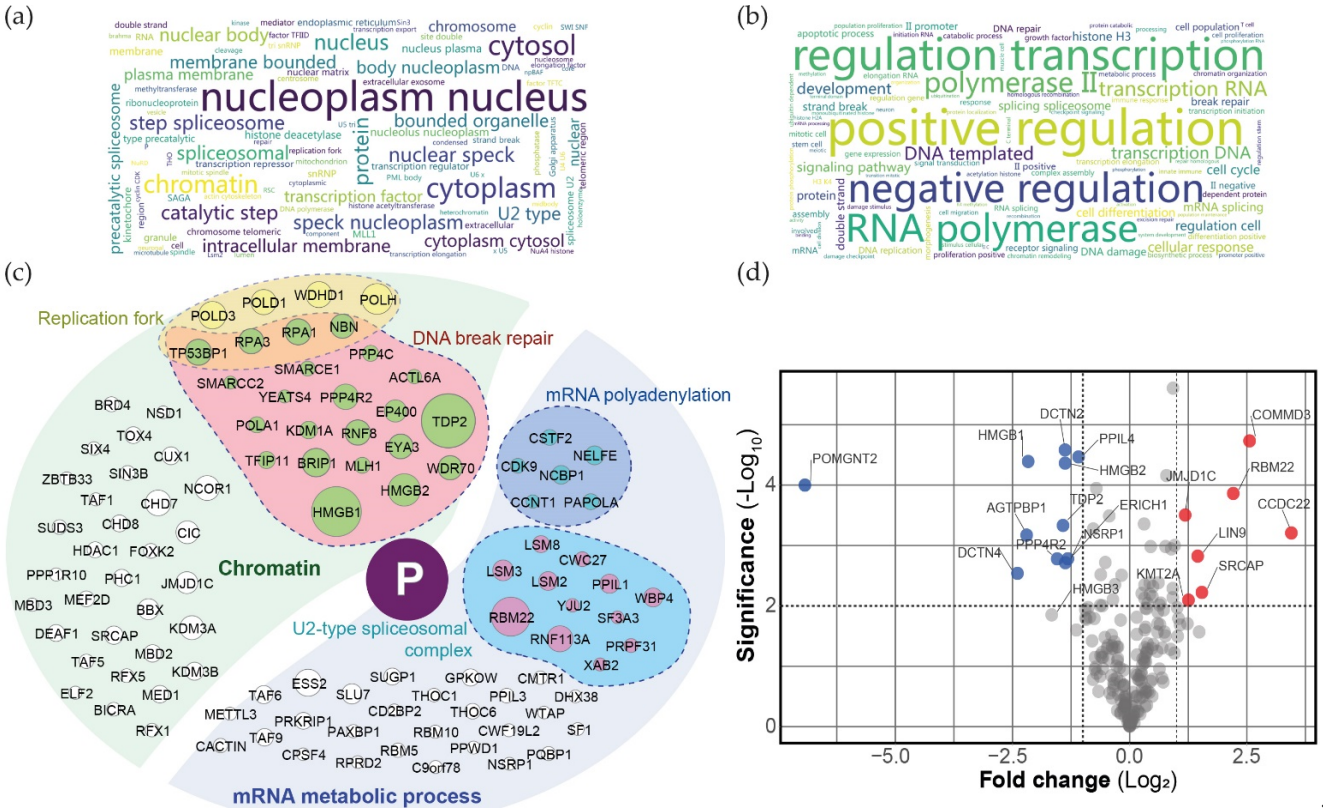


Figure 6. Proximal interactomes of P_{FULL} in human cells. Word clouds of Gene Ontology enrichment analysis (https://toppgene.cchmc.org/; accessed on 1 September 2022) of Cellular Component (**a**) and Biological Process (**b**) (assembled on http://www.bioinformatics.com.cn/plot_basic_wordcloud_118_en; accessed on 1 September 2022). (**c**) Illustration of a subset of P_{FULL} high-confidence interactors. Proteins are grouped and coloured by selected functions or compartments, as indicated on the diagram; node diameter is proportional to the Log_2 fold change against control. (**d**) Volcano plot comparing the N- and C-terminal BirA* tagged BioID data. Blue and red dots represent N- and C-terminal BirA* tagged preferential partners, respectively (p-value < 0.01 and Log_2 fold change > 1 between both fusions). Assembled using https://huygens.science.uva.nl/VolcaNoseR/; accessed on 1 September 2022). All data are based on three biological replicates and have been filtered for background against a set of 20 control runs (see Methods and Supplemental Methods for details).

4. Discussion

By interacting with all the proteins of the replication machinery and recruiting many cellular partners, *Mononegavirales* P can be considered as the cornerstone of viral proliferation. Their peculiar architecture consisting of long disordered regions that alternate with structured domains is largely related to all these functions. Our structural findings reveal that orthobornaviral P does not deviate from that model, with an extended central oligomerisation domain that represents half of the size of the protein, surrounded by N- and C-terminal flexible parts. In the past, it was speculated that BoDV-1 P assembled in oligomers, from dimers stabilized by an internal disulfide bridge to tetramers [90,101]. With our data, we confirm that orthobornaviral P forms tetramers in solution, which is in complete agreement with these previous works. There were uncertainties regarding the formation of a trimeric or a tetrameric form in these previous studies, which could be related to the fact that the qualitative analysis of such a protein with disordered propensities was performed using a Superose 12 column [101]. Indeed, disordered and globular proteins are known to be eluted differently from SEC, and the column used at that time was less suited to efficiently discriminate proteins with such small differences in composition. Because of our experimental procedures, in particular during the crystallisation step, we cannot

totally exclude the presence of a disulfide bridge at position 125 for a non-secreted protein [90]. However, the presence of a disulfide bridge at position 125 may not be of primary importance, since C125 is only conserved in mammalian 1 orthobornaviruses P (Figure S8). For the other orthobornaviruses, position 125 (or similar) is occupied by amino acids with different physico-chemical properties (from serine/threonine to leucine/isoleucine). In contrast, D126 is strictly conserved amongst the members of genus *Orthobornavirus*, and our minireplicon assays suggested that it plays an important role in viral replication, without any explanation yet. Future studies examining the role of D126 position in the assembly of the polymerase complex and vSPOT formation will help to better understand the role of this key position.

In *Mononegavirales*, *Bornaviridae* P is the shortest regarding amino acid composition. However, it bears the most elongated oligomerisation domain, although to date, this feature has only been proven in solution, far from the cellular context. Structural data on the *Bornaviridae* L-P complex may be particularly relevant for information about P structure within the replication complex. For instance, P may conserve its elongated structuration, similarly to PIV5, where the tetrameric P_{OD} protrudes from the globular RdRp [15]. Alternatively, the twist present in BoDV-1 P_{OD} may correspond to a hinge that would favour a structural rearrangement of the cofactor to reach an octopus-like organisation intertwined with L, as seen for RSV [14,102]. However, such a rearrangement would involve the destabilisation of the N- or the C-terminal coiled-coil, which would be energetically consuming. The L binding site for P has been described to be located between residues 135 and 183 [101]. This part of the P_{OD} corresponds to the most conserved of the two helical coiled-coils (the sequence conservations of the N- and C-terminal α-helixes are 30% and 55%, respectively). The C-terminal α-helix ending at position 162 in our crystal structure, the interaction between L and P could be only mediated by the end of the second α-helix and a stretch of flexible residues up to 183, without any major rearrangement of the cofactor.

Our results also show that BoDV-1 P presents significant inherent flexibility through its N- and C-terminal parts. This is stressed by the high overlap between the N- and C-terminal BioID tagged P proximal interactomes. Comparing the N- and C-terminal BioID tagged proximal interactomes, we observed that only 45% (94/207) of factors are in common. However, when comparing the relative abundance of the high-confidence interactors, over 90% (189/207) did not show significant differences depending on the tag positioning. This apparent discrepancy is due to stringent filtering used to define high-confidence proximal interactors, and the LFQ comparison between merged hits strongly supports the ability of the BirA* moiety to conjugate biotin on proteins located in similar places, whether it is positioned at the N- or C-terminal of P. Given that the biotinylation radius of BirA* has been estimated to be ~10 nm [92], and that that of P_{OD} alone is ~14 nm, we hypothesise that the disordered domains remain unstructured in living cells and confer to BirA* enough freedom to move around the P_{OD}, thereby biotinylating similar species. Our analysis of the proximal interactomes confirmed the intrinsic ability of BoDV-1 P to localise at DNA DSBs regions. The interactome analysis particularly pointed out a proximity with sensing factors (TP53BP1) or resolution by repair of the DSB (*i.e.*, NBN, BRIP, PPP4). It also supported its role in recruiting numerous mRNA processing factors, which could serve as complementation factors for vRNA production and maturation in the context of BoDV-1 infection (*e.g.*, mRNA polyadenylation machinery). Considering that BioID identifies proximal partners of the fusion bait protein over 24 h (labelling time roughly equal to one cell cycle in HEK293), our data hence represent a collection of all detectable proteins that were at some point in contact with or in the close vicinity of P_{FULL}. Considering that differentiated neurons (the cells targeted by BoDV-1) are postmitotic cells and thus do not enter in mitosis, it is important to stress that our approach must be seen as a discovery step with several limitations to keep in mind. Besides lacking the co-expression of N and L BoDV-1 proteins, which could trigger synergic interactions, labelling over a cell cycle time could explain the presence of a few non-nuclear factors that could interact with BoDV-1 P upon nuclear envelope breakdown (prophase). If this hypothesis

is true, non-nuclear interactors identified here should be considered with caution since P expressed alone would remain strictly nuclear. However, since P is translated in the cytoplasm, these interactions could represent P processing factors before its delivery to the nucleoplasm. In line with the previous remark, we also identified replication fork factors. In view of the tropism of BoDV-1 towards neurons, our data regarding the replication fork components must be considered with caution as well. Given the recruitment of DNA repair machineries on stalled forks during replication stress to repair DNA damage and resume replication ([103] for review), we consider the interactions with the replication fork factor to be model-dependent and not necessarily relevant in an infectious context. However, the frequent occurrence of DNA DSBs and the peculiar centrality of the DNA repair machineries in driving neurons somatic mosaicism and diverse neurodegenerative diseases are well documented (reviewed in [104]). We see our data as a discovery step sketching the ability of BoDV-1 phosphoprotein to interact with ubiquitous DNA repair factors. Moreover, these results may be particularly important in the context of novel findings showing that neuronal DSB responses play key roles in learning and memory [105,106]. For instance, our data could serve as a first step to supervised studies in infected neurons aiming at understanding which of these factors are essential for the positioning of vSPOTs at DNA DSBs. In addition, the identification of proximal partners involved in the regulation of innate immune responses suggests that P could interfere with the host cell's ability to defend against BoDV-1 infection.

Beyond supporting the relevance of our structural model in living cells, our proximal interactomics data represent—to our knowledge—the first published BioID study of any *Mononegavirales* P. Making sense of what is suspected in the field and providing dozens of nuclear ubiquitous factors, this work hence represents an unprecedented wealth of data, which could direct numerous studies aiming at deciphering BoDV-1 P functions in human cells. Repeating such experiments in an infectious context (*i.e.*, complementing with the other BoDV-1 proteins ideally in neurons) would provide information on the cooperation between viral proteins and better resolve the issue of vSPOT composition.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/v14112358/s1. Experimental procedures: generation of Bio-ID samples; Experimental procedures: BioID data acquisition; Experimental procedures: BioID data analysis; Figure S1: Ramachandran plot of the three P_{OD} X-ray structures; Figure S2: Guinier analysis of the P_{OD} SAXS analysis; Figure S3: P(R) functions and Krakty plots of the P_{OD} SAXS analysis; Figure S4: Orthobornaviral P_{OD} SAXS envelops; Figure S5: P_{FULL} P(r) function; Figure S6: Validity of the six P_{FULL} models presented in Figure 2d; Figure S7: Detection of BoDV-1 P mutants; Figure S8: Sequence alignment of P from all representative viruses of the genus *Orthobornavirus*; Figure S9: Detailed electron density maps; Figure S10: Orthobornaviral P_{OD} flexibility; Figure S11: BoDV-1 P structure prediction by AlphaFold2; Table S1: X-ray data collection and refinement parameters; Table S2: SAXS data collection and scattering-derived parameters; Table S3: BioID data and enrichment analyses. References [108–111] are cited in the supplementary materials.

Author Contributions: Conceptualization, T.C., M.H., D.G.-D. and E.C.; methodology, N.T., F.C., J.K., K.B., J.-M.B., E.M.N.L., P.L., L.L.F. and T.C.; formal analysis, N.T., J.-M.B., M.H., E.C. and T.C.; writing—original draft preparation, E.C., M.H. and T.C.; writing—review and editing, N.T., D.G.-D., M.H., E.C., E.S., R.W.H.R., K.T. and T.C.; funding acquisition, T.C., D.G.-D., E.C. and M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Agence Nationale de la Recherche (grant Bavarian; ANR-21-CE15-0026) and the Japan Society for the Promotion of Science (Grant-in-Aid for Scientific Research (B): JP21H01199). For this project, F.C. received funding from GRAL (Grenoble Alliance for Integrated Structural and Cell Biology, a program of the Chemistry Biology Health Graduate School of Université Grenoble Alpes (ANR-17-EURE-0003)), K.B. was co-funded by the Ministère de l'Enseignement Supérieur et de la Recherche and the Université de Lille Nord Europe (project MONET), and E.MN.L. was funded by the Métropole européenne de Lille ("Accueil de Talents" funding scheme).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The X-ray diffraction data are available in the protein databank (https://www.rcsb.org/; accessed on 1 September 2022) with the accession numbers 8B8A, 8B8B, and 8B8D for BoDV-1, MuBV-1, and GaVV-1 P_{OD}, respectively; the SAXS data can be found in the small-angle scattering biological databank (https://www.sasbdb.org/; accessed on 1 September 2022) with the accession numbers SASDQP5, SASDQQ5, SASDQR5, and SASDQS5 for BoDV-1 P_{FULL}, BoDV-1, MuBV-1, and GaVV-1 P_{OD}, respectively; the mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [107] partner repository with the dataset identifier PDX036750 at http://www.ebi.ac.uk/pride; accessed on 1 September 2022 (login: reviewer_pxd036750@ebi.ac.uk; pw: OXUWT9se).

Acknowledgments: We acknowledge Patrice Velicite, Alexandra Mousset, Ambre Valenti, Marwan Hamoudi, and Jean-Pascal Gimeno for technical support, with special thanks to Ada Roy for her involvement. We thank all members of our laboratories for their contributions and helpful discussions. This work used the platforms of the Grenoble Instruct-ERIC centre (ISBG; UAR 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-0005-02) and GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). We acknowledge SOLEIL (proposal numbers 20170827 and 20191314) and the ESRF (proposal number MX2285) for provision of synchrotron radiation facilities, and we would like to thank Aurélien Thureau, Pierre Montaville, Andrew Thompson, William Shepard, and Daniele De Sanctis for assistance in using the beamlines SWING, PROXIMA-1, PROXIMA-2A, and ID29, respectively.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Whelan, S.P.; Barr, J.N.; Wertz, G.W. Transcription and replication of nonsegmented negative-strand RNA viruses. *Curr. Top. Microbiol. Immunol.* **2004**, *283*, 61–119.
2. Kolakofsky, D.; Le Mercier, P.; Nishio, M.; Blackledge, M.; Crepin, T.; Ruigrok, R.W.H. Sendai virus and a unified model of mononegavirus RNA synthesis. *Viruses* **2021**, *13*, 2466. [CrossRef]
3. Tarbouriech, N.; Curran, J.; Ruigrok, R.W.; Burmeister, W.P. Tetrameric coiled-coil domain of Sendai virus phosphoprotein. *Nat. Struct. Biol.* **2000**, *7*, 777–781.
4. Ding, H.; Green, T.J.; Lu, S.; Luo, M. Crystal structure of the oligomerization domain of the phosphoprotein of vesicular stomatitis virus. *J. Virol.* **2006**, *80*, 2808–2814. [CrossRef] [PubMed]
5. Ivanov, I.; Crépin, T.; Jamin, M.; Ruigrok, R.W. Structure of the dimerization domain of the rabies virus phosphoprotein. *J. Virol.* **2010**, *84*, 3707–3710. [CrossRef]
6. Communie, G.; Crépin, T.; Maurin, D.; Jensen, M.R.; Blackledge, M.; Ruigrok, R.W. Structure of the tetramerization domain of measles virus phosphoprotein. *J. Virol.* **2013**, *87*, 7166–7169. [CrossRef] [PubMed]
7. Cox, R.; Green, T.J.; Purushotham, S.; Deivanayagam, C.; Bedwell, G.J.; Prevelige, P.E.; Luo, M. Structural and functional characterization of the mumps virus phosphoprotein. *J. Virol.* **2013**, *87*, 7558–7568. [CrossRef] [PubMed]
8. Bruhn, J.F.; Hotard, A.L.; Spiropoulou, C.F.; Lo, M.K.; Saphire, E.O. A conserved basic patch and central kink in the Nipah virus phosphoprotein multimerization domain are essential for polymerase function. *Structure* **2019**, *27*, 660–668 e4. [CrossRef]
9. Zinzula, L.; Nagy, I.; Orsini, M.; Weyher-Stingl, E.; Bracher, A.; Baumeister, W. Structures of Ebola and Reston virus VP35 oligomerization domains and comparative biophysical characterization in all ebolavirus species. *Structure* **2019**, *27*, 39–54 e6. [CrossRef]
10. Jensen, M.R.; Yabukarski, F.; Communie, G.; Condamine, E.; Mas, C.; Volchkova, V.; Tarbouriech, N.; Bourhis, J.M.; Volchkov, V.; Blackledge, M.; et al. Structural description of the Nipah virus phosphoprotein and its interaction with STAT1. *Biophys. J.* **2020**, *118*, 2470–2488. [CrossRef]
11. Wang, Z.X.; Liu, S.B.; Guan, H.; Lu, L.F.; Tu, J.G.; Ouyang, S.; Zhang, Y.A. Structural and functional characterization of the phosphoprotein central domain of spring viremia of carp virus. *J. Virol.* **2020**, 94. [CrossRef] [PubMed]
12. Cardone, C.; Caseau, C.M.; Bardiaux, B.; Thureaux, A.; Galloux, M.; Bajorek, M.; Eleouet, J.F.; Litaudon, M.; Bontems, F.; Sizun, C. A structural and dynamic analysis of the partially disordered polymerase-binding domain in RSV phosphoprotein. *Biomolecules* **2021**, *11*, 1225. [CrossRef] [PubMed]
13. Gilman, M.S.A.; Liu, C.; Fung, A.; Behera, I.; Jordan, P.; Rigaux, P.; Ysebaert, N.; Tcherniuk, S.; Sourimant, J.; Eleouet, J.F.; et al. Structure of the respiratory syncytial virus polymerase complex. *Cell* **2019**, *179*, 193–204.e14. [CrossRef] [PubMed]
14. Cao, D.; Gao, Y.; Roesler, C.; Rice, S.; D'Cunha, P.; Zhuang, L.; Slack, J.; Domke, M.; Antonova, A.; Romanelli, S.; et al. Cryo-EM structure of the respiratory syncytial virus RNA polymerase. *Nat. Commun.* **2020**, *11*, 368. [CrossRef]

15. Abdella, R.; Aggarwal, M.; Okura, T.; Lamb, R.A.; He, Y. Structure of a paramyxovirus polymerase complex reveals a unique methyltransferase-CTD conformation. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 4931–4941. [CrossRef]
16. Jenni, S.; Bloyet, L.M.; Diaz-Avalos, R.; Liang, B.; Whelan, S.P.J.; Grigorieff, N.; Harrison, S.C. Structure of the vesicular stomatitis virus L protein in complex with its phosphoprotein cofactor. *Cell Rep.* **2020**, *30*, 53–60.e5. [CrossRef]
17. Horwitz, J.A.; Jenni, S.; Harrison, S.C.; Whelan, S.P.J. Structure of a rabies virus polymerase complex from electron cryo-microscopy. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 2099–2107. [CrossRef]
18. Pan, J.; Qian, X.; Lattmann, S.; El Sahili, A.; Yeo, T.H.; Jia, H.; Cressey, T.; Ludeke, B.; Noton, S.; Kalocsay, M.; et al. Structure of the human metapneumovirus polymerase phosphoprotein complex. *Nature* **2020**, *577*, 275–279. [CrossRef]
19. Gould, J.R.; Qiu, S.; Shang, Q.; Ogino, T.; Prevelige, P.E., Jr.; Petit, C.M.; Green, T.J. The connector domain of vesicular stomatitis virus large protein interacts with the viral phosphoprotein. *J. Virol.* **2020**, 94. [CrossRef]
20. Green, T.J.; Luo, M. Structure of the vesicular stomatitis virus nucleocapsid in complex with the nucleocapsid-binding domain of the small polymerase cofactor, P. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11713–11718. [CrossRef]
21. Leyrat, C.; Yabukarski, F.; Tarbouriech, N.; Ribeiro, E.A., Jr.; Jensen, M.R.; Blackledge, M.; Ruigrok, R.W.; Jamin, M. Structure of the vesicular stomatitis virus N(0)-P complex. *PLoS Pathog.* **2011**, *7*, e1002248. [CrossRef] [PubMed]
22. Yabukarski, F.; Lawrence, P.; Tarbouriech, N.; Bourhis, J.M.; Delaforge, E.; Jensen, M.R.; Ruigrok, R.W.; Blackledge, M.; Volchkov, V.; Jamin, M. Structure of Nipah virus unassembled nucleoprotein in complex with its viral chaperone. *Nat. Struct. Mol. Biol.* **2014**, *21*, 754–759. [CrossRef] [PubMed]
23. Guryanov, S.G.; Liljeroos, L.; Kasaragod, P.; Kajander, T.; Butcher, S.J. Crystal structure of the measles virus nucleoprotein core in complex with an N-terminal region of phosphoprotein. *J. Virol.* **2015**, *90*, 2849–2857. [CrossRef] [PubMed]
24. Leung, D.W.; Borek, D.; Luthra, P.; Binning, J.M.; Anantpadma, M.; Liu, G.; Harvey, I.B.; Su, Z.; Endlich-Frazier, A.; Pan, J.; et al. An intrinsically disordered peptide from Ebola virus VP35 controls viral RNA synthesis by modulating nucleoprotein-RNA interactions. *Cell Rep.* **2015**, *11*, 376–389. [CrossRef]
25. Renner, M.; Bertinelli, M.; Leyrat, C.; Paesen, G.C.; Saraiva de Oliveira, L.F.; Huiskonen, J.T.; Grimes, J.M. Nucleocapsid assembly in pneumoviruses is regulated by conformational switching of the N protein. *Elife* **2016**, *5*, e12627. [CrossRef]
26. Zhu, T.; Song, H.; Peng, R.; Shi, Y.; Qi, J.; Gao, G.F. Crystal structure of the Marburg virus nucleoprotein core domain chaperoned by a VP35 peptide reveals a conserved drug target for Filovirus. *J. Virol.* **2017**, 91. [CrossRef]
27. Aggarwal, M.; Leser, G.P.; Kors, C.A.; Lamb, R.A. Structure of the paramyxovirus parainfluenza virus 5 nucleoprotein in complex with an amino-terminal peptide of the phosphoprotein. *J. Virol.* **2018**, 92. [CrossRef]
28. Dong, X.; Wang, X.; Xie, M.; Wu, W.; Chen, Z. Structural basis of human parainfluenza virus 3 unassembled nucleoprotein in complex with its viral chaperone. *J. Virol.* **2022**, *96*, e0164821. [CrossRef]
29. Barik, S.; Banerjee, A.K. Phosphorylation by cellular casein kinase II is essential for transcriptional activity of vesicular stomatitis virus phosphoprotein P. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 6570–6574. [CrossRef]
30. Schwemmle, M.; De, B.; Shi, L.; Banerjee, A.; Lipkin, W.I. Borna disease virus P-protein is phosphorylated by protein kinase Cepsilon and casein kinase II. *J. Biol. Chem.* **1997**, *272*, 21818–21823. [CrossRef]
31. Dupuy, L.C.; Dobson, S.; Bitko, V.; Barik, S. Casein kinase 2-mediated phosphorylation of respiratory syncytial virus phosphoprotein P is essential for the transcription elongation activity of the viral polymerase; phosphorylation by casein kinase 1 occurs mainly at Ser(215) and is without effect. *J. Virol.* **1999**, *73*, 8384–8392. [CrossRef] [PubMed]
32. Raux, H.; Flamand, A.; Blondel, D. Interaction of the rabies virus P protein with the LC8 dynein light chain. *J. Virol.* **2000**, *74*, 10212–10216. [CrossRef] [PubMed]
33. Sun, M.; Fuentes, S.M.; Timani, K.; Sun, D.; Murphy, C.; Lin, Y.; August, A.; Teng, M.N.; He, B. Akt plays a critical role in replication of nonsegmented negative-stranded RNA viruses. *J. Virol.* **2008**, *82*, 105–114. [CrossRef] [PubMed]
34. Kubota, T.; Matsuoka, M.; Chang, T.H.; Bray, M.; Jones, S.; Tashiro, M.; Kato, A.; Ozato, K. Ebolavirus VP35 interacts with the cytoplasmic dynein light chain 8. *J. Virol.* **2009**, *83*, 6952–6956. [CrossRef] [PubMed]
35. Leung, D.W.; Ginder, N.D.; Fulton, D.B.; Nix, J.; Basler, C.F.; Honzatko, R.B.; Amarasinghe, G.K. Structure of the Ebola VP35 interferon inhibitory domain. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 411–416. [CrossRef]
36. Lahaye, X.; Vidy, A.; Fouquet, B.; Blondel, D. Hsp70 protein positively regulates rabies virus infection. *J. Virol.* **2012**, *86*, 4743–4751. [CrossRef]
37. Fouquet, B.; Nikolic, J.; Larrous, F.; Bourhy, H.; Wirblich, C.; Lagaudrière-Gesbert, C.; Blondel, D. Focal adhesion kinase is involved in rabies virus infection through its interaction with viral phosphoprotein P. *J. Virol.* **2015**, *89*, 1640–1651. [CrossRef]
38. Jespersen, N.E.; Leyrat, C.; Gerard, F.C.; Bourhis, J.M.; Blondel, D.; Jamin, M.; Barbar, E. The LC8-RavP ensemble structure svinces a role for LC8 in regulating lyssavirus polymerase functionality. *J. Mol. Biol.* **2019**, *431*, 4959–4977. [CrossRef]
39. Briggs, K.; Wang, L.; Nagashima, K.; Zengel, J.; Tripp, R.A.; He, B. Regulation of mumps virus replication and transcription by kinase RPS6KB1. *J. Virol.* **2020**, 94. [CrossRef]
40. Lim, D.; Shin, H.C.; Choi, J.S.; Kim, S.J.; Ku, B. Crystal structure of human LC8 bound to a peptide from Ebola virus VP35. *J. Microbiol.* **2021**, *59*, 410–416. [CrossRef]
41. Kuhn, J.H.; Durrwald, R.; Bao, Y.; Briese, T.; Carbone, K.; Clawson, A.N.; deRisi, J.L.; Garten, W.; Jahrling, P.B.; Kolodziejek, J.; et al. Taxonomic reorganization of the family Bornaviridae. *Arch. Virol.* **2015**, *160*, 621–632. [CrossRef] [PubMed]

42. Hyndman, T.H.; Shilton, C.M.; Stenglein, M.D.; Wellehan, J.F.X., Jr. Divergent bornaviruses from Australian carpet pythons with neurological disease date the origin of extant Bornaviridae prior to the end-Cretaceous extinction. *PLoS Pathog.* **2018**, *14*, e1006881. [CrossRef] [PubMed]
43. Shi, M.; Lin, X.D.; Chen, X.; Tian, J.H.; Chen, L.J.; Li, K.; Wang, W.; Eden, J.S.; Shen, J.J.; Liu, L.; et al. The evolutionary history of vertebrate RNA viruses. *Nature* **2018**, *556*, 197–202. [CrossRef] [PubMed]
44. Matsumoto, Y.; Hayashi, Y.; Omori, H.; Honda, T.; Daito, T.; Horie, M.; Ikuta, K.; Fujino, K.; Nakamura, S.; Schneider, U.; et al. Bornavirus closely associates and segregates with host chromosomes to ensure persistent intranuclear infection. *Cell Host Microbe* **2012**, *11*, 492–503. [CrossRef]
45. Hoffmann, B.; Tappe, D.; Hoper, D.; Herden, C.; Boldt, A.; Mawrin, C.; Niederstrasser, O.; Muller, T.; Jenckel, M.; van der Grinten, E.; et al. A variegated squirrel bornavirus associated with fatal human encephalitis. *N. Engl. J. Med.* **2015**, *373*, 154–162. [CrossRef] [PubMed]
46. Korn, K.; Coras, R.; Bobinger, T.; Herzog, S.M.; Lucking, H.; Stohr, R.; Huttner, H.B.; Hartmann, A.; Ensser, A. Fatal encephalitis associated with Borna disease virus 1. *N. Engl. J. Med.* **2018**, *379*, 1375–1377. [CrossRef]
47. Cadar, D.; Allendorf, V.; Schulze, V.; Ulrich, R.G.; Schlottau, K.; Ebinger, A.; Hoffmann, B.; Hoffmann, D.; Rubbenstroth, D.; Ismer, G.; et al. Introduction and spread of variegated squirrel bornavirus 1 (VSBV-1) between exotic squirrels and spill-over infections to humans in Germany. *Emerg. Microbes Infect.* **2021**, *10*, 602–611. [CrossRef]
48. Hirai, Y.; Tomonaga, K.; Horie, M. Borna disease virus phosphoprotein triggers the organization of viral inclusion bodies by liquid-liquid phase separation. *Int. J. Biol. Macromol.* **2021**, *192*, 55–63. [CrossRef]
49. Marty, F.H.; Bettamin, L.; Thouard, A.; Bourgade, K.; Allart, S.; Larrieu, G.; Malnou, C.E.; Gonzalez-Dunia, D.; Suberbielle, E. Borna disease virus docks on neuronal DNA double-strand breaks to replicate and dampens neuronal activity. *iScience* **2022**, *25*, 103621. [CrossRef]
50. Suberbielle, E.; Stella, A.; Pont, F.; Monnet, C.; Mouton, E.; Lamouroux, L.; Monsarrat, B.; Gonzalez-Dunia, D. Proteomic analysis reveals selective impediment of neuronal remodeling upon Borna disease virus infection. *J. Virol.* **2008**, *82*, 12265–12279. [CrossRef]
51. Bonnaud, E.M.; Szelechowski, M.; Betourne, A.; Foret, C.; Thouard, A.; Gonzalez-Dunia, D.; Malnou, C.E. Borna disease virus phosphoprotein modulates epigenetic signaling in neurons to control viral replication. *J. Virol.* **2015**, *89*, 5996–6008. [CrossRef] [PubMed]
52. Rudolph, M.G.; Kraus, I.; Dickmanns, A.; Eickmann, M.; Garten, W.; Ficner, R. Crystal structure of the borna disease virus nucleoprotein. *Structure* **2003**, *11*, 1219–1226. [CrossRef] [PubMed]
53. Neumann, P.; Lieber, D.; Meyer, S.; Dautel, P.; Kerth, A.; Kraus, I.; Garten, W.; Stubbs, M.T. Crystal structure of the Borna disease virus matrix protein (BDV-M) reveals ssRNA binding properties. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 3710–3715. [CrossRef] [PubMed]
54. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **1993**, *26*, 795–800. [CrossRef]
55. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66 Pt 2*, 133–144. [CrossRef]
56. Tickle, I.J.; Flensburg, C.; Keller, P.; Paciorek, W.; Sharff, A.; Vonrhein, C.; Bricogne, G. *STARANISO*; Global Phasing Ltd.: Cambridge, UK, 2017.
57. Vonrhein, C.; Flensburg, C.; Keller, P.; Sharff, A.; Smart, O.; Paciorek, W.; Womack, T.; Bricogne, G. Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67 Pt 4*, 293–302. [CrossRef]
58. McCoy, A.J.; Grosse-Kunstleve, R.W.; Adams, P.D.; Winn, M.D.; Storoni, L.C.; Read, R.J. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40 Pt 4*, 658–674. [CrossRef]
59. Terwilliger, T.C.; Grosse-Kunstleve, R.W.; Afonine, P.V.; Moriarty, N.W.; Zwart, P.H.; Hung, L.W.; Read, R.J.; Adams, P.D. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D Biol. Crystallogr.* **2008**, *64 Pt 1*, 61–69. [CrossRef]
60. Adams, P.D.; Afonine, P.V.; Bunkoczi, G.; Chen, V.B.; Davis, I.W.; Echols, N.; Headd, J.J.; Hung, L.W.; Kapral, G.J.; Grosse-Kunstleve, R.W.; et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66 Pt 2*, 213–221. [CrossRef]
61. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66 Pt 4*, 486–501. [CrossRef]
62. Cohen, S.X.; Ben Jelloul, M.; Long, F.; Vagin, A.; Knipscheer, P.; Lebbink, J.; Sixma, T.K.; Lamzin, V.S.; Murshudov, G.N.; Perrakis, A. ARP/wARP and molecular replacement: The next generation. *Acta Crystallogr. D Biol. Crystallogr.* **2008**, *64 Pt 1*, 49–60. [CrossRef]
63. Langer, G.; Cohen, S.X.; Lamzin, V.S.; Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **2008**, *3*, 1171–1179. [CrossRef] [PubMed]
64. Murshudov, G.N. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **1997**, *53 Pt 3*, 240–255. [CrossRef]
65. Sheldrick, G.M. A short history of SHELX. *Acta Crystallogr. A* **2008**, *64 Pt 1*, 112–122. [CrossRef]

66. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66 Pt 4*, 470–478. [CrossRef]
67. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef]
68. Schrodinger, LLC. *The PyMOL Molecular Graphics System*, Version 1.8; Schrödinger, Inc.: New York, NY, USA, 2015.
69. Thureau, A.; Roblin, P.; Perez, J. BioSAXS on the SWING beamline at Synchrotron SOLEIL. *J. Appl. Crystallogr.* **2021**, *54*, 1698–1710. [CrossRef]
70. Manalastas-Cantos, K.; Konarev, P.V.; Hajizadeh, N.R.; Kikhney, A.G.; Petoukhov, M.V.; Molodenskiy, D.S.; Panjkovich, A.; Mertens, H.D.T.; Gruzinov, A.; Borges, C.; et al. ATSAS 3.0: Expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Crystallogr.* **2021**, *54*, 343–355. [CrossRef] [PubMed]
71. Hopkins, J.B.; Gillilan, R.E.; Skou, S. BioXTAS RAW: Improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J. Appl. Crystallogr.* **2017**, *50*, 1545–1553. [CrossRef] [PubMed]
72. Svergun, D. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **1992**, *25*, 495–503. [CrossRef]
73. Hansen, S. Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering. *J. Appl. Crystallogr.* **2000**, *33*, 1415–1421. [CrossRef]
74. Rambo, R.P.; Tainer, J.A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **2013**, *496*, 477–481. [CrossRef] [PubMed]
75. Franke, D.; Svergun, D.I. DAMMIF, a program for rapid *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **2009**, *42 Pt 2*, 342–346. [CrossRef]
76. Volkov, V.V.; Svergun, D.I. Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **2003**, *36*, 860–864. [CrossRef]
77. Svergun, D.I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **1999**, *76*, 2879–2886. [CrossRef]
78. Tria, G.; Mertens, H.D.; Kachala, M.; Svergun, D.I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2015**, *2 Pt 2*, 207–217. [CrossRef]
79. Moore, B.L.; Kelley, L.A.; Barber, J.; Murray, J.W.; MacDonald, J.T. High-quality protein backbone reconstruction from alpha carbons using Gaussian mixture models. *J. Comput. Chem.* **2013**, *34*, 1881–1889. [CrossRef]
80. Krivov, G.G.; Shapovalov, M.V.; Dunbrack, R.L., Jr. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **2009**, *77*, 778–795. [CrossRef]
81. Kanda, T.; Horie, M.; Komatsu, Y.; Tomonaga, K. The Borna disease virus 2 (BoDV-2) nucleoprotein is a conspecific protein that enhances BoDV-1 RNA-dependent RNA polymerase activity. *J. Virol.* **2021**, *95*, e0093621. [CrossRef] [PubMed]
82. Reuter, A.; Horie, M.; Höper, D.; Ohnemus, A.; Narr, A.; Rinder, M.; Beer, M.; Staeheli, P.; Rubbenstroth, D. Synergistic antiviral activity of ribavirin and interferon-α against parrot bornaviruses in avian cells. *J. Gen. Virol.* **2016**, *97*, 2096–2103. [CrossRef]
83. Tu, Y.X.I.; Sydor, A.M.; Coyaud, E.; Laurent, E.M.N.; Dyer, D.; Mellouk, N.; St-Germain, J.; Vernon, R.M.; Forman-Kay, J.D.; Li, T.; et al. Global proximity interactome of the human macroautophagy pathway. *Autophagy* **2022**, *18*, 1174–1186. [CrossRef] [PubMed]
84. Gerard, F.C.; Ribeiro Ede, A., Jr.; Leyrat, C.; Ivanov, I.; Blondel, D.; Longhi, S.; Ruigrok, R.W.; Jamin, M. Modular organization of rabies virus phosphoprotein. *J. Mol. Biol.* **2009**, *388*, 978–996. [CrossRef] [PubMed]
85. Crick, F.H.C. The Fourier transform of a coiled-coil. *Acta Crystallogr.* **1953**, *6*, 685–689. [CrossRef]
86. Crick, F.H.C. The packing of α-helices: Simple coiled-coils. *Acta Crystallogr.* **1953**, *6*, 689–697. [CrossRef]
87. Grigoryan, G.; Degrado, W.F. Probing designability via a generalized model of helical bundle geometry. *J. Mol. Biol.* **2011**, *405*, 1079–1100. [CrossRef]
88. Strelkov, S.V.; Burkhard, P. Analysis of alpha-helical coiled-coils with the program TWISTER reveals a structural mechanism for stutter compensation. *J. Struct. Biol.* **2002**, *137*, 54–64. [CrossRef]
89. Hartmann, M.D.; Mendler, C.T.; Bassler, J.; Karamichali, I.; Ridderbusch, O.; Lupas, A.N.; Hernandez Alvarez, B. alpha/beta coiled-coils. *Elife* **2016**, *5*, 351–358. [CrossRef]
90. Kliche, S.; Stitz, L.; Mangalam, H.; Shi, L.; Binz, T.; Niemann, H.; Briese, T.; Lipkin, W.I. Characterization of the Borna disease virus phosphoprotein, p23. *J. Virol.* **1996**, *70*, 8133–8137. [CrossRef]
91. Roux, K.J.; Kim, D.I.; Raida, M.; Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **2012**, *196*, 801–810. [CrossRef]
92. Kim, D.I.; Birendra, K.C.; Zhu, W.; Motamedchaboki, K.; Doye, V.; Roux, K.J. Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E2453–E2461. [CrossRef]
93. Mehus, A.A.; Anderson, R.H.; Roux, K.J. BioID Identification of Lamin-Associated Proteins. *Methods Enzymol.* **2016**, *569*, 3–22.
94. Kamitani, W.; Shoya, Y.; Kobayashi, T.; Watanabe, M.; Lee, B.J.; Zhang, G.; Tomonaga, K.; Ikuta, K. Borna disease virus phosphoprotein binds a neurite outgrowth factor, amphoterin/HMG-1. *J. Virol.* **2001**, *75*, 8742–8751. [CrossRef] [PubMed]
95. Zhang, G.; Kobayashi, T.; Kamitani, W.; Komoto, S.; Yamashita, M.; Baba, S.; Yanai, H.; Ikuta, K.; Tomonaga, K. Borna disease virus phosphoprotein represses p53-mediated transcriptional activity by interference with HMGB1. *J. Virol.* **2003**, *77*, 12243–12251. [CrossRef] [PubMed]

96. Starokadomskyy, P.; Gluck, N.; Li, H.; Chen, B.; Wallis, M.; Maine, G.N.; Mao, X.; Zaidi, I.W.; Hein, M.Y.; McDonald, F.J.; et al. CCDC22 deficiency in humans blunts activation of proinflammatory NF-kappaB signaling. *J. Clin. Investig.* **2013**, *123*, 2244–2256. [CrossRef] [PubMed]
97. Yoh, S.M.; Schneider, M.; Seifried, J.; Soonthornvacharin, S.; Akleh, R.E.; Olivieri, K.C.; De Jesus, P.D.; Ruan, C.; de Castro, E.; Ruiz, P.A.; et al. PQBP1 Is a proximal sensor of the cGAS-dependent innate response to HIV-1. *Cell* **2015**, *161*, 1293–1305. [CrossRef]
98. Atzei, P.; Gargan, S.; Curran, N.; Moynagh, P.N. Cactin targets the MHC class III protein IkappaB-like (IkappaBL) and inhibits NF-kappaB and interferon-regulatory factor signaling pathways. *J. Biol. Chem.* **2010**, *285*, 36804–36817. [CrossRef]
99. Morchikh, M.; Cribier, A.; Raffel, R.; Amraoui, S.; Cau, J.; Severac, D.; Dubois, E.; Schwartz, O.; Bennasser, Y.; Benkirane, M. HEXIM1 and NEAT1 long non-coding RNA form a multi-subunit complex that regulates DNA-mediated innate immune response. *Mol. Cell* **2017**, *67*, 387–399.e5. [CrossRef] [PubMed]
100. Qiu, W.; Zhang, Q.; Zhang, R.; Lu, Y.; Wang, X.; Tian, H.; Yang, Y.; Gu, Z.; Gao, Y.; Yang, X.; et al. N(6)-methyladenosine RNA modification suppresses antiviral innate sensing pathways via reshaping double-stranded RNA. *Nat. Commun.* **2021**, *12*, 1582. [CrossRef] [PubMed]
101. Schneider, U.; Blechschmidt, K.; Schwemmle, M.; Staeheli, P. Overlap of interaction domains indicates a central role of the P protein in assembly and regulation of the Borna disease virus polymerase complex. *J. Biol. Chem.* **2004**, *279*, 55290–55296. [CrossRef] [PubMed]
102. Gilman, M.S.A.; Furmanova-Hollenstein, P.; Pascual, G.; B van't Wout, A.; Langedijk, J.P.M.; McLellan, J.S. Transient opening of trimeric prefusion RSV F proteins. *Nat. Commun.* **2019**, *10*, 2105. [CrossRef]
103. Cortez, D. Replication-Coupled DNA Repair. *Mol. Cell* **2019**, *74*, 866–876. [CrossRef] [PubMed]
104. Qing, X.; Zhang, G.; Wang, Z.Q. DNA damage response in neurodevelopment and neuromaintenance. *FEBS J.* **2022**. [CrossRef] [PubMed]
105. Suberbielle, E.; Sanchez, P.E.; Kravitz, A.V.; Wang, X.; Ho, K.; Eilertson, K.; Devidze, N.; Kreitzer, A.C.; Mucke, L. Physiologic brain activity causes DNA double-strand breaks in neurons, with exacerbation by amyloid-beta. *Nat. Neurosci.* **2013**, *16*, 613–621. [CrossRef] [PubMed]
106. Suberbielle, E.; Djukic, B.; Evans, M.; Kim, D.H.; Taneja, P.; Wang, X.; Finucane, M.; Knox, J.; Ho, K.; Devidze, N.; et al. DNA repair factor BRCA1 depletion occurs in Alzheimer brains and impairs cognitive function in mice. *Nat. Commun.* **2015**, *6*, 8897. [CrossRef]
107. Perez-Riverol, Y.; Bai, J.; Bandla, C.; Garcia-Seisdedos, D.; Hewapathirana, S.; Kamatchinathan, S.; Kundu, D.J.; Prakash, A.; Frericks-Zipper, A.; Eisenacher, M.; et al. The PRIDE database resources in 2022: A hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **2022**, *50*, D543–D552. [CrossRef]
108. Goedhart, J.; Luijsterburg, M.S. VolcaNoseR is a web app for creating, exploring, labeling and sharing volcano plots. *Sci. Rep.* **2020**, *10*, 20560. [CrossRef] [PubMed]
109. Williams, C.J.; Headd, J.J.; Moriarty, N.W.; Prisant, M.G.; Videau, L.L.; Deis, L.N.; Verma, V.; Keedy, D.A.; Hintze, B.J.; Chen, V.B.; et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **2018**, *27*, 293–315. [CrossRef]
110. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [CrossRef]
111. Gouet, P.; Courcelle, E.; Stuart, D.I.; Metoz, F. ESPript: Analysis of multiple sequence alignments in PostScript. *Bioinformatics* **1999**, *15*, 305–308. [CrossRef]

Human Papillomavirus 42 Drives Digital Papillary Adenocarcinoma and Elicits a Germ Cell–like Program Conserved in HPV-Positive Cancers

Lukas Leiendecker[1,2], Tobias Neumann[1,2,3], Pauline S. Jung[1,2,4], Shona M. Cronin[1,2], Thomas L. Steinacker[5], Alexander Schleiffer[1], Michael Schutzbier[1,5,6], Karl Mechtler[1,5,6], Thibault Kervarrec[7], Estelle Laurent[8], Kamel Bachiri[8], Etienne Coyaud[8], Rajmohan Murali[9], Klaus J. Busam[9], Babak Itzinger-Monshi[10], Reinhard Kirnbauer[4], Lorenzo Cerroni[11], Eduardo Calonje[12], Arno Rütten[13], Frank Stubenrauch[14], Klaus G. Griewank[15], Thomas Wiesner[4,16], and Anna C. Obenauf[1]

Illustration by Bianca Dunn

ABSTRACT

The skin is exposed to viral pathogens, but whether they contribute to the oncogenesis of skin cancers has not been systematically explored. Here we investigated 19 skin tumor types by analyzing off-target reads from commonly available next-generation sequencing data for viral pathogens. We identified human papillomavirus 42 (HPV42) in 96% ($n = 45/47$) of digital papillary adenocarcinoma (DPA), an aggressive cancer occurring on the fingers and toes. We show that HPV42, so far considered a nononcogenic, "low-risk" HPV, recapitulates the molecular hallmarks of oncogenic, "high-risk" HPVs. Using machine learning, we find that HPV-driven transformation elicits a germ cell–like transcriptional program conserved throughout all HPV-driven cancers (DPA, cervical carcinoma, and head and neck cancer). We further show that this germ cell–like transcriptional program, even when reduced to the top two genes (*CDKN2A* and *SYCP2*), serves as a fingerprint of oncogenic HPVs with implications for early detection, diagnosis, and therapy of all HPV-driven cancers.

SIGNIFICANCE: We identify HPV42 as a uniform driver of DPA and add a new member to the short list of tumorigenic viruses in humans. We discover that all oncogenic HPVs evoke a germ cell–like transcriptional program with important implications for detecting, diagnosing, and treating all HPV-driven cancers.

See related commentary by Starrett et al., p. 17.

[1]Research Institute of Molecular Pathology (IMP), Vienna BioCenter (VBC), Vienna, Austria. [2]Vienna BioCenter PhD Program, Doctoral School of the University at Vienna and Medical University of Vienna, Vienna BioCenter (VBC), Vienna, Austria. [3]Quantro Therapeutics, Vienna, Austria. [4]Department of Dermatology, Medical University of Vienna, Vienna, Austria. [5]Institute of Molecular Biotechnology (IMBA), Vienna BioCenter (VBC), Vienna, Austria. [6]The Gregor Mendel Institute of Molecular Plant Biology of the Austrian Academy of Sciences (GMI), Vienna BioCenter (VBC), Vienna, Austria. [7]Department of Pathology, University Hospital Center of Tours, University of Tours, Tours, France. [8]PRISM INSERM U1192, Université de Lille, Villeneuve d'Ascq, France. [9]Department of Pathology and Laboratory Medicine, Memorial Sloan Kettering Cancer Center, New York, New York. [10]Department of Dermatology, Hospital Rudolfstiftung, Vienna, Austria. [11]Department of Dermatology, Medical University of Graz, Graz, Austria. [12]Department of Dermatopathology, St John's Institute of Dermatology, St Thomas' Hospital, London, United Kingdom. [13]Dermatopathology Friedrichshafen, Friedrichshafen, Germany. [14]University Hospital Tuebingen, Institute for Medical Virology and Epidemiology of Viral Diseases, Tuebingen, Germany. [15]Department of Dermatology, University Hospital Essen, University of Duisburg, German Cancer Consortium (DKTK), Partner Site, Essen, Germany. [16]Department of Pathology, Medical University of Vienna, Vienna, Austria.

Corresponding Authors: Anna C. Obenauf, Research Institute of Molecular Pathology, Campus-Vienna-Biocenter 1, 1030 Vienna, Austria. Phone: 0043-179-730; E-mail: anna.obenauf@imp.ac.at; and Thomas Wiesner, Department of Dermatology, Medical University of Vienna, Waehringer Guertel 18-20, 1090 Vienna, Austria. Phone: 0043-1404-0077-100; E-mail: thomas.wiesner@meduniwien.ac.at

Cancer Discov 2023;13:70–84

doi: 10.1158/2159-8290.CD-22-0489

This open access article is distributed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

©2022 The Authors; Published by the American Association for Cancer Research

INTRODUCTION

Up to 15% of human cancers are caused by microbial pathogens, most notably by viruses that interfere with signaling pathways maintaining cellular homeostasis (1). Prominent examples include cervical carcinoma and head and neck cancer caused by oncogenic human papillomavirus (HPV), Merkel cell carcinoma by Merkel cell polyomavirus (MCV), Kaposi sarcoma by human herpesvirus 8 (HHV8), or hepatocellular carcinoma associated with chronic hepatitis B/C virus infection (2). Many virus-associated cancers occur at the interface between the human body and the environment, such as the mucosal epithelium or skin (1), but the contribution of viral pathogens to skin cancers has not been systematically investigated.

RESULTS

HPV42 Is Associated with Digital Papillary Adenocarcinoma

By exploiting the substantial amount of off-target reads in targeted sequencing data (3) and utilizing a metagenomic classifier with a custom database (Viral DB) covering all viruses (>6,000) infecting vertebrate hosts (4), we screened 214 samples from 19 skin tumor types for viral pathogens (Fig. 1A). We did not find viral sequences in most analyzed skin tumors except for two tumor types: MCV in Merkel cell carcinoma, confirming the feasibility of our approach, and, notably, human papillomavirus 42 (HPV42) in all ($n = 11/11$) digital papillary adenocarcinomas (DPA; Fig. 1B; Supplementary Fig. S1A).

DPA is an aggressive, metastasizing skin cancer that occurs predominantly on the fingers and toes of young males (Supplementary Table S1). Although DPA is thought to arise from eccrine stem cells, the molecular drivers are thus far unknown (Supplementary Fig. S1A; ref. 5). To validate our discovery of HPV42 in DPA, we enriched viral nucleic acids from tumor samples using a hybridization-based sequence capture assay (ViroCap; ref. 4) and confirmed the presence of HPV42 RNA and DNA (Fig. 1C; Supplementary Fig. S1B and S1C; Supplementary Tables S2 and S3). In all but one DPA sample, which harbored a deletion of the late HPV region, the sequencing reads covered the complete HPV42 genome (Supplementary Fig. S1D). In an extended cohort, we confirmed the presence of HPV42 in 96% of DPA samples using qPCR of tumor DNA ($n = 45/47$; Supplementary Fig. S1C; Supplementary Table S1).

HPV42 is currently classified as a "low-risk" HPV type of the Alpha-1 species (genus *Alphapapillomavirus*), which has not been associated with oncogenic transformation (Supplementary Fig. S2A; ref. 6). Similar to other nononcogenic, "low-risk" HPV types (e.g., HPV6 or 11), HPV42 causes benign warts and is only rarely found in cancer (7, 8). In contrast, oncogenic, "high-risk" HPV types, such as HPV16 or 18, have an established role in tumorigenesis (Supplementary Fig. S2A; refs. 9, 10) and are frequently integrated into the host genome, facilitating the high expression of the viral oncogenes E6 and E7 (Supplementary Fig. S2B; ref. 11). Notably, we identified and confirmed HPV42 integrations in six DPAs ($n = 6/16$, 37%) by mapping HPV42–human genome breakpoints by targeted and Sanger sequencing (Fig. 1D; Supplementary Figs. S2C and S3A–S3D and Supplementary Table S4). To investigate the breakpoints in DPA tumors at the single-cell level, we designed a fluorescence *in situ* hybridization (FISH) experiment using probes targeting the HPV42 genome and the human genome adjacent to the integration breakpoint, inferred from sequencing data (Supplementary Table S5). The magenta viral probe was visible as a single spot in the tumor cells that overlapped with the yellow human probe (Fig. 1E; Supplementary Fig. S4A and S4B). These fusion signals suggest a clonal expansion of cells with the integrated HPV42 virus. The viral oncogenes E6 and E7 always remained intact during the integration event, and in 44% of cases ($n = 7/16$), we found a disruption of the viral genes E1 and E2, which is important for the derepression of E6 and E7 (Fig. 1D; Supplementary Tables S4, S6, and S7; ref. 12).

To confirm the expression of HPV42 E6 and E7 in DPA tumors, we utilized RNA *in situ* hybridization (RNA-ISH) and found HPV42 transcripts in 23 of 24 DPA tumors but not in the surrounding normal tissues (Fig. 1F; Supplementary Table S1). HPV42 transcripts were predominantly expressed from the early rather than late region of the HPV42 genome, consistent with transcriptional patterns observed for HPV16-induced cervical cancers and attributable to an incomplete differentiation state of the infected cell (ref. 13; Supplementary Fig. S2B). Altogether, these data suggest that HPV42 plays a central role in the pathogenesis of DPA.

HPV42, Previously Classified as a Nononcogenic HPV, Induces Transformation *In Vitro* and Tumor Growth *In Vivo*

A key feature of oncogenic HPVs is the ability to extend the proliferative life span of primary cells such as human primary keratinocytes (HPK) or mouse embryonic fibroblasts (MEF; refs. 14–16). To functionally assess the oncogenic activity of HPV42, we transduced HPKs with the full genome of nononcogenic HPV11, oncogenic HPV16, and HPV42. Although HPK-HPV11 cells stopped proliferating, HPK-HPV16 and HPK-HPV42 cells showed an extended proliferative life span (Fig. 2A; Supplementary Fig. S5A–S5C). Similarly, coexpression of only the E6/E7 oncoproteins of HPV42 and HPV16, but not of HPV11, promoted proliferation and prolonged the life span of HPKs and MEFs (Fig. 2B; Supplementary Fig. S5D–S5F).

To facilitate transformation, oncogenic HPVs utilize their E6 and E7 oncoproteins to alter cell-cycle progression (17, 18) and differentiation programs (19). The main target of the E6 protein of oncogenic HPVs is p53, which is polyubiquitinated and proteasomally degraded upon recruitment of the ubiquitin ligase E6AP (17, 20). Interestingly, we found that HPV42-E6 indeed strongly binds E6AP (Supplementary Fig. S5G and S5H; Supplementary Table S8), but in contrast to HPV16, p53 was not degraded and p53 targets remained, also after DNA damage induction, unchanged (Supplementary Fig. S5I–S5L). We next focused on the E7 protein, which, when derived from oncogenic HPVs, is sufficient to transform cells (21, 22). We expressed E7 of HPV11, 16, and 42 in HPK and MEF cells, and whereas empty vector (EV) or HPV11-E7–transduced cells entered senescence, cells with HPV42-E7 and HPV16-E7 showed an extended proliferative life span (Supplementary Fig. S6A–S6D).

To assess the interaction partners of HPV42-E7, we performed a proximity-labeling experiment identifying pRB as

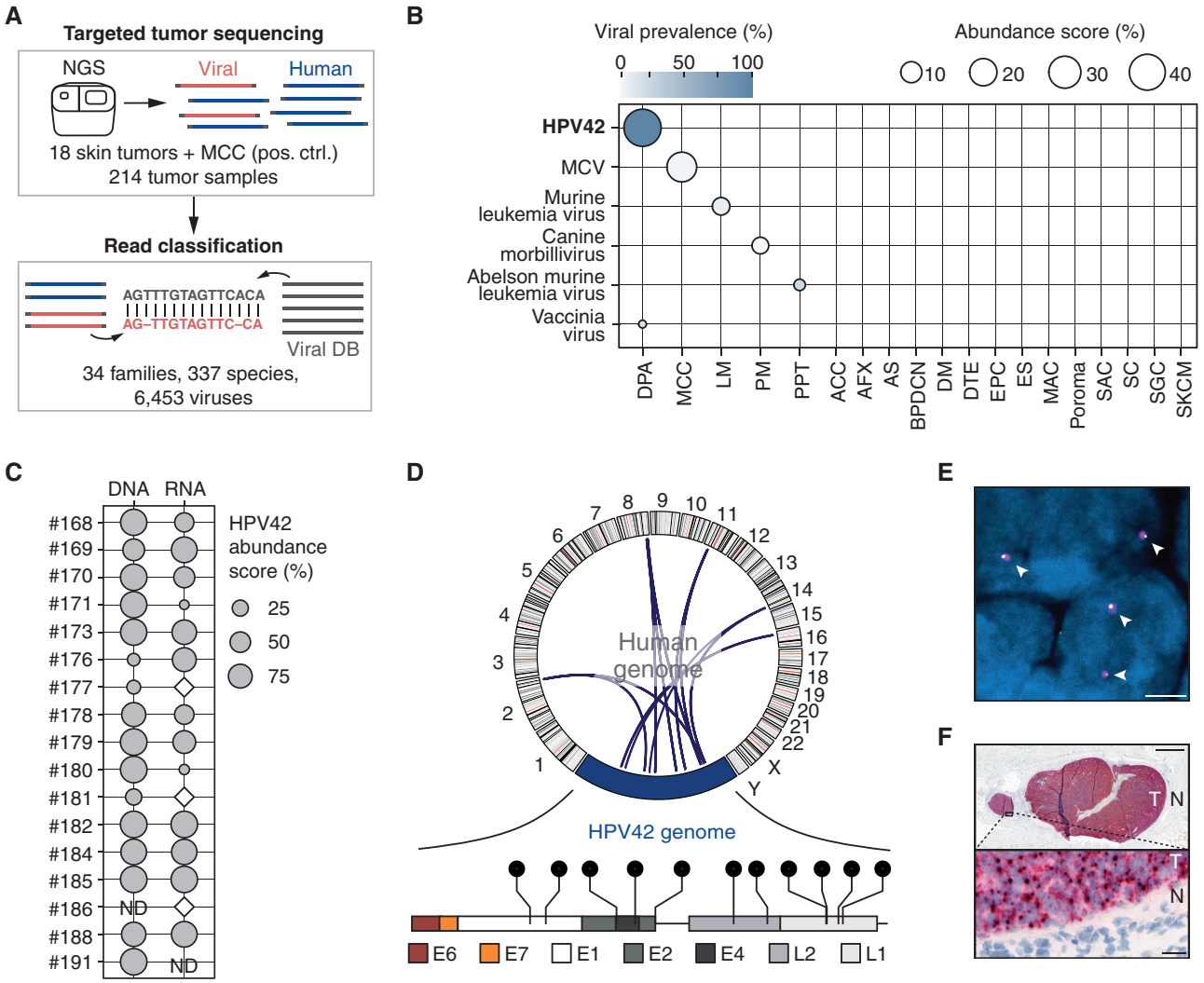


Figure 1. HPV42 is associated with DPA. **A,** Overview of unbiased screening approach for detection of vertebrate-infecting viruses in targeted next-generation sequencing (NGS) data. Targeted NGS data from 18 skin tumor types and MCC (positive control, pos. ctrl.) were analyzed with the Centrifuge metagenomic classifier utilizing a custom viral database (Viral DB). **B,** Virus abundance score in targeted NGS data of 214 human tumor samples. Detection of HPV42 in 100% ($n = 11/11$) of DPA samples. ACC, adenoid cystic carcinoma; AFX, atypical fibroxanthoma; AS, angiosarcoma; BPDCN, blastic plasmacytoid dendritic cell neoplasm; DM, desmoplastic melanoma; DTE, desmoplastic trichoepithelioma; EPC, eccrine porocarcinoma; ES, eccrine spiradenoma; LM, lentigo maligna melanoma; MAC, microcystic adnexal carcinoma; PM, pediatric melanoma; PPT, proliferating pilar cystic tumor; SAC, spiradenocarcinoma; SC, sebaceous carcinoma; SGC, sweat gland carcinoma; SKCM, cutaneous melanoma. Viral prevalence: proportion of samples with detected virus in any given tumor type. Abundance score: mean proportion of detected virus in any given tumor type. **C,** HPV42 abundance score in a cohort of 17 DPAs after sequence capture to DNA and RNA of 6,453 viruses. White diamond: HPV42 detected but below an abundance of 1%. ND, not determined. **D,** Top, Circos plot depicting HPV42–human genome breakpoints. Each arch represents one detected breakpoint. Bottom, lollipop presentation of integration breakpoints in the HPV42 genome. Each lollipop represents one breakpoint (Supplementary Table S4). **E,** FISH signal of HPV42 genome (magenta) and human genome integration site adjacent locus (yellow) in DPA nuclei (blue). White arrowheads indicate fusion signals. Scale bar, 2.5 μm. Individual channels and quantifications in Supplementary Fig. S4. **F,** Top, RNA-ISH staining for expression of early region HPV42 mRNA in lung metastasis of a DPA tumor (T) and the surrounding normal (N) tissue. Scale bar, 1 mm. Bottom, inset. Scale bar, 20 μm.

the top interaction partner together with several other proteins important for G_1–S phase transition, such as RBL1, RBL2, and E2F family proteins (Fig. 2C; Supplementary Table S9). The RB family proteins (pRB, RBL1, and RBL2) are bound by oncogenic E7 via a conserved LxCxE motif (18, 21, 23). Notably, we found that the aspartic acid (D) upstream of the LxCxE motif, which has been shown to enhance pRB binding in oncogenic HPVs, is conserved in HPV42 (ref. 23; Supplementary Fig. S6E). Compared with HPV11, E7 of HPV42 and HPV16 indeed bound more strongly to pRB (Supplementary Fig. S6F), leading to equally robust activation of E2F signaling, an important indicator of functional disruption of pRB (Fig. 2D). Transcriptomic profiling identified E2F targets as the top upregulated gene class in HPK cells transduced with E7, which is in line with the proposed release of E2F transcription factors from RB family proteins by E7 interference (ref. 24; Fig. 2E; Supplementary Fig. S6G).

The second strongest binding partner of HPV42-E7 was the tumor suppressor PTPN14, together with several PTPN14-regulated components of the HIPPO pathway (Fig. 2C; Sup-

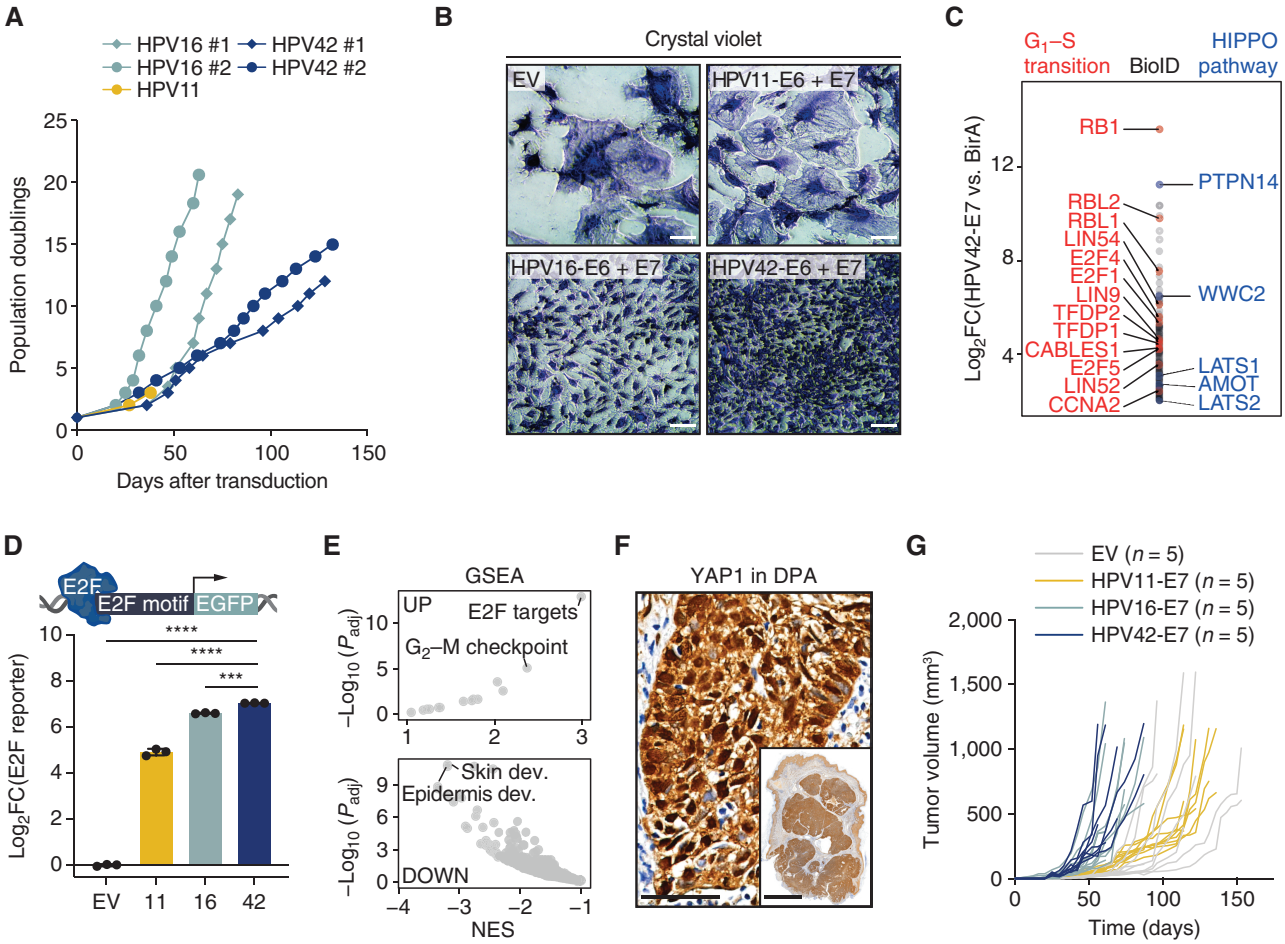


Figure 2. HPV42 induces transformation *in vitro* and tumor growth *in vivo*. **A,** Growth curves of HPKs transduced with the full genome of HPV11 (low-risk), HPV16 (high-risk), and HPV42. **B,** Crystal violet staining of MEF cells at day 14 after transduction. Scale bars, 100 μm. **C,** HPV42-E7 protein interaction partners identified in BioID proximity-labeling experiment. Pathway members of gene ontology terms G_1–S transition (red) and HIPPO pathway (blue) are labeled. FC, fold change. **D,** E2F reporter assay in U2OS cells transduced with E7 proteins of HPV11, 16, or 42. ***, $P < 0.001$; ****, $P < 0.0001$ (one-way ANOVA, Tukey multiple comparisons test). **E,** Gene set enrichment analysis (GSEA) of the Molecular Signatures Database (MSigDB) hallmark gene set (top) and MSigDB curated gene set (bottom) comparing HPV42-E7 versus EV-transduced HPK cells. dev., development; NES, normalized enrichment score. **F,** IHC for YAP1 in a DPA tumor. Scale bar, main: 50 μm; inset: 3 mm. **G,** Individual tumor growth of EV- or E7 of HPV11-, HPV16-, or HPV42-transduced HaCaT cells in NSG mice. Each group consists of 5 mice (i.e., 10 tumors).

plementary Table S9). We found that HPV42-E7, most likely involving the UBR4/p600 ubiquitin ligase (19), leads to the degradation of PTPN14, which promotes the persistence of dedifferentiated, HPV-infected cells in the basal layer of the skin (25) and is necessary for the transforming ability of oncogenic E7 proteins (ref. 26; Supplementary Fig. S6H). We found that PTPN14 loss leads to a downregulation of noncanonical HIPPO targets (Supplementary Fig. S6I), including skin/epidermis development–related pathways (Fig. 2E; Supplementary Fig. S6J), and downregulation of skin differentiation markers identified by quantitative mass spectrometry (MS) in HPV42-transduced cells (Supplementary Fig. S6K). We also observed predominantly nuclear localization of YAP1 in DPA tumors, indicating reduced noncanonical HIPPO signaling (Fig. 2F). Altogether, our data suggest that the inactivation of noncanonical HIPPO signaling by targeting PTPN14 for degradation plays a role in HPV42-driven DPAs.

To assess the oncogenic potential of HPV42-E7 *in vivo*, we expressed the E7 protein in HaCaT cells, a spontaneously immortalized human keratinocyte line (27, 28). We found that expression of E7 of HPV16 or HPV42 was sufficient to robustly induce tumor growth *in vivo* within 50 days, underlining the oncogenic potential of HPV42-E7 (Fig. 2G; Supplementary Fig. S6L). Collectively, our results suggest that HPV42 induces transformation *in vitro* and tumor growth *in vivo*.

HPV42 Is Specifically Found in DPA

Considering the high prevalence of HPV42 in DPA and its oncogenic properties, we systematically screened for HPV42 and other HPVs in common cancers. We found that only ~32% ($n = 62/193$) of HPVs and neither HPV42 nor other oncogenic HPVs (e.g., HPV45, 33, 52) are included in the NCBI Reference Sequence Database (RefSeq; Supplementary Fig. S7A–S7C), which is commonly used for metagenomic classification (1, 29). Thus, we generated an index based on the curated PapillomaVirus Episteme database (PaVE; ref. 30) and assessed raw RNA sequencing (RNA-seq) data of 10,078 tumor samples across 32 cancer types from The Cancer

Genome Atlas (TCGA) project. We detected HPVs in 3% of TCGA samples ($n = 361/10,078$; Supplementary Fig. S7D and Supplementary Table S10), including 93% of cervical cancer cases [cervical squamous cell carcinoma (CESC); $n = 282/301$] and 12% of head and neck cancer [head and neck squamous cell carcinoma (HNSCC); $n = 62/498$]. We also found HPV in rare cases of colon, urothelial, soft tissue, bladder, kidney, and lung cancers (Supplementary Fig. S7D and S7E). Notably, we did not identify HPV42 in any of the 10,078 TCGA tumors, indicating a cell lineage–specific role of HPV42-driven transformation in DPA (Supplementary Fig. S7D and S7E).

All Oncogenic HPVs Induce a Germ Cell–like Transcriptional Program Conserved throughout HPV-Driven Cancers

We next used a machine learning approach to investigate whether HPV-driven cancers share a common transcriptional program that reveals more about the biology of HPV-driven cancers and could be exploited diagnostically and therapeutically. We annotated expression data of CESC and HNSCC from TCGA with the HPV status and identified 489 differentially expressed protein-coding genes between HPV$^+$ and HPV$^-$ cases across both cancer types (Fig. 3A; Supplementary Table S11). Feature selection using a random forest (RF) algorithm on the differentially expressed genes revealed a signature of 12 protein-coding genes (RF12; Fig. 3B; Supplementary Fig. S8A). The RF12 signature was sufficient to separately cluster HPV$^+$ and HPV$^-$ tumors across CESC and HNSCC regardless of their tissue origin and the oncogenic HPV type (Fig. 3C). Importantly, RF12 was also conserved in DPA, which clustered with the HPV$^+$ tumors, but was absent in HPV-associated skin warts, which clustered with HPV$^-$ tumors (Fig. 3C and D; Supplementary Fig. S8B).

To assess the specificity and sensitivity of RF12 in identifying HPV-driven tumors, we trained four machine learning classifiers on a training set of randomly selected tumor samples ($n = 484$, 60%) from CESC, HNSCC, and DPA (Fig. 3A). By applying the models to a test set withheld from training ($n = 323$, 40%), we found that all models achieved excellent performance (AUC >0.99) in predicting HPV as the cancer driver (Supplementary Fig. S8C). RF12 identified HPV-driven tumors with an accuracy of 94% in CESC, 99% in HNSCC, and 100% in DPA (Supplementary Fig. S8D) and scored equally in independent cohorts of HNSCC and CESC tumors (Supplementary Fig. S8E). We found that RF12 also scored in precancerous lesions of the cervix [high-grade squamous intraepithelial lesion (HSIL)], but not in HPV-associated skin warts and matched normal tissue (Fig. 3D; Supplementary Fig. S8F). Collectively, RF12 serves as an HPV fingerprint specific for HPV-driven oncogenic transformation, which is present in DPA tumors and supports that HPV42 is the oncogenic driver of DPA.

In the clinic, the accurate identification of HPV in HNSCC and CESC tumors and its precancerous lesions plays an important role (5). IHC for upregulation of p16/*CDKN2A*, the gene with the highest predictive contribution within RF12, is an established surrogate marker for oncogenic HPV infection (Supplementary Fig. S8A and S8B). However, p16/CDKN2A IHC lacks specificity, leading to diagnostic and therapeutic shortcomings. For example, ~10% of patients with HPV$^-$ oropharyngeal squamous cell carcinoma, a common form of HNSCC, are p16/CDKN2A$^+$ but not HPV-driven. Therefore, these patients are incorrectly assigned to the more prognostically favorable group of HPV$^+$ patients, putting these patients at risk of undertreatment (31, 32).

To improve the accurate identification of HPV$^+$ tumors, we investigated *SYCP2*. *SYCP2*, a gene that in germ cells is important for meiotic prophase, has the second-highest predictive contribution within the RF12 signature and shows specific expression in HPV-driven tumors (Fig. 3B; Supplementary Fig. S8A). We found that the combination of *SYCP2* with *CDKN2A* improved HPV status assessment in HNSCC from 85% with *CDKN2A* alone to 97% accuracy (Fig. 3E; Supplementary Fig. S8G). Also, in DPAs, *CDKN2A* expression or p16/*CDKN2A* staining alone correctly identified HPV42-positive cases only in 70% to 75% of cases; however, the combination with *SYCP2* improved HPV status assessment to 100% accuracy (Supplementary Fig. S8G and S8H). Importantly, we confirmed in HPV$^+$ DPA and CESC tumors that IHC is sufficient to assess *SYCP2* expression (Supplementary Fig. S8I and S8J). Altogether, these data suggest that adding a single gene, *SYCP2*, to routine diagnostic IHC of p16/*CDKN2A* could provide a reliable, cost-effective, and easily accessible method for accurately assessing HPV status in the clinic.

When investigating RF12 in normal tissues retrieved from the Genotype-Tissue Expression portal (GTEx; ref. 33), we found that RF12 is absent in 30 normal tissues but strongly scored in testis (Fig. 3F). Interestingly, we found that the majority of RF12 genes are highly and exclusively expressed in cells of the germ cell lineage and 9 of 12 genes play a role in processes specific for germ cells, such as meiosis (Fig. 3G; Supplementary Fig. S9A and S9B). Notably, this germ cell–enriched expression pattern is not limited to RF12 but expands to the 197 protein-coding genes we found commonly upregulated in HPV$^+$ tumors (Supplementary Fig. S9B–S9D). We found that the E2F family transcription factor motif was the top enriched motif in the genes upregulated in HPV-driven tumors (Supplementary Fig. S9E–S9G). Altogether, our data suggest that the E7 oncoprotein of oncogenic HPVs interferes with the pRB–E2F axis in cancer cells and evokes a conserved germ cell–like program in HPV-driven tumors, which, even when reduced to the top two genes (*CDKN2A* and *SYCP2*), accurately distinguishes HPV-driven from mutation-driven tumors and outperforms established clinical markers (Fig. 3H).

DISCUSSION

In summary, our study reveals HPV42 infections as a uniform driver mechanism of DPA, an aggressive skin cancer with thus-far unknown etiology that lacks mechanism-based therapeutic options. By functionally characterizing HPV42, previously considered a "low-risk" virus, our study adds a new member to the short list of tumorigenic viruses in humans. Prophylactic HPV vaccination is a remarkably effective preventive measure for HPV-associated diseases, but current vaccines immunize against only the most common (2–9) HPV types (34). Our data provide a rationale for including additional HPV types in the development of new vaccines and highlight the potential of pan-HPV vaccines (35, 36).

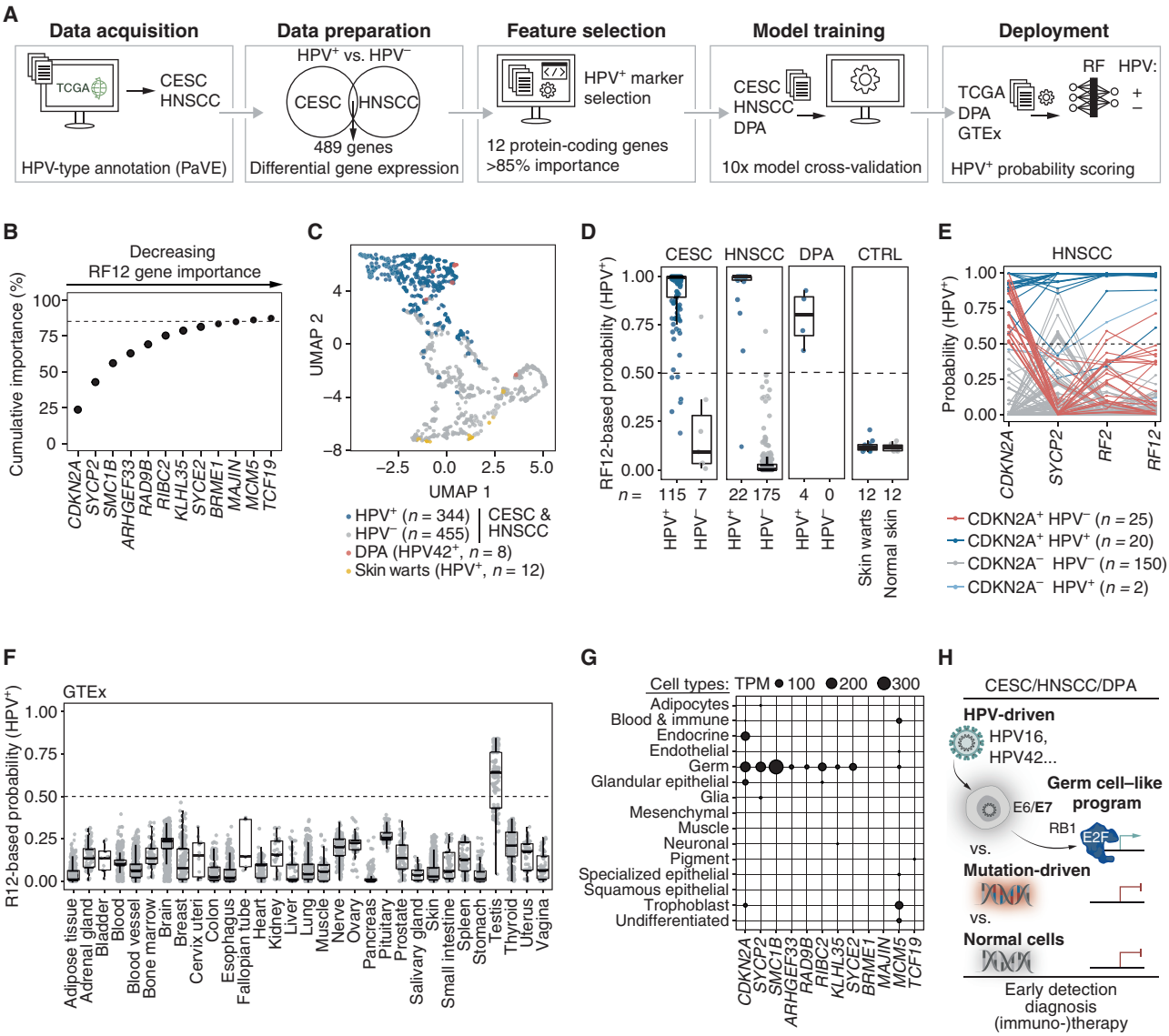


Figure 3. Oncogenic HPVs induce a germ cell–like transcriptional program conserved throughout HPV-driven cancers. **A,** Overview of the random forest (RF) feature selection and machine learning procedure to identify a transcriptional fingerprint of HPV-driven oncogenesis. CESC and HNSCC expression data were annotated with HPV status (data acquisition), commonly differentially expressed genes between HPV+ and HPV− were identified (data preparation), and 12 protein-coding signature genes (RF12) were identified (feature selection). Four different machine learning models were trained with RF12 on a subset of DPA, CESC, and HNSCC samples (model training); the model with the best hyperparameters was evaluated using the withheld sample subset and used to classify Genotype-Tissue Expression (GTEx), skin warts, or normal skin expression data (deployment). **B,** Contribution of RF12 signature genes to cumulative feature importance (%) to discriminate between HPV+ and HPV− tumor samples. **C,** Uniform manifold approximation and projection (UMAP) dimensionality reduction of RF12 expression in DPA, CESC, HNSCC, and skin warts. **D,** HPV+ probability scores were calculated by RF12 for CESC, HNSCC, DPA, skin warts, and normal skin. For CESC, HNSCC, and DPA, the test set samples are displayed. **E,** HPV+ probability scores for HNSCC calculated for *CDKN2A* and *SYCP2* alone, the combination of *CDKN2A* and *SYCP2* (RF2), and RF12. **F,** HPV+ probability scores were calculated by RF12 for 31 normal tissues obtained from the GTEx database. **G,** Expression levels of RF12 genes in 15 cell types obtained from Human Protein Atlas. **H,** Schematic model of the germ cell–like program in HPV-driven cancer.

It is notable that the cell context in which HPV42 is present may determine the full oncogenic potential of HPV42. In the keratinocyte cell lineage of the mucosal (ano-)genital region, HPV42 can only lead to warts (thereby recapitulating features of "low-risk" HPV types), whereas in the eccrine lineage (sweat glands), which are most abundant on the hands and the feet, it causes oncogenic transformation leading to DPAs (recapitulating features of "high-risk" HPV types). In addition to this cell lineage–specific oncogenic potential of HPV42, it is also intriguing that HPV16 and HPV18, which dominate cervical and head cancers, are absent in DPAs.

The HPV-induced germ cell–like program we find in all HPV-driven cancers not only has diagnostic implications but also represents a vulnerability that could be exploited immunotherapeutically, which has proven difficult in HPV-driven cancers but holds promise for durable tumor control if suc-

cessful. The top genes of the HPV-induced germ cell–like program have immunogenic properties, as they are only expressed in HPV$^+$ cancers and immune-privileged germ cells but not in healthy somatic tissues and are predicted to be presented on common MHC class I molecules (Supplementary Fig. S9H). These properties make them promising targets for therapeutic cancer vaccines to induce an antitumor T-cell response, particularly in combination with vaccines involving HPV antigens E6 and E7 that are in clinical trials (NCT04405349 and NCT04287868), or as targets for engineered T cells in combination with T cells targeting E7 (37).

METHODS

Targeted Sequencing of Rare Skin Tumors

Sequencing of IMPACT tumors (n = 208) and pooled normal samples was performed as previously described (38). For whole-exome sequencing (WES) of six DPA tumor samples (Supplementary Table S1), genomic DNA was extracted from 10-µm-thick sections of formalin-fixed, paraffin-embedded (FFPE) tumor tissues. Sections were deparaffinized and manually macrodissected, and genomic DNA was isolated using the QIAamp DNA Mini Kit (#56304, Qiagen) according to the manufacturer's instructions. WES libraries were generated with the SureSelectXT Library Preparation Kit following the SureSelectXT Target Enrichment System for Illumina B.2 protocol (April 2015). Samples were sequenced on an Illumina HiSeq4000 in 150-bp, paired-end mode. The study complies with the Declaration of Helsinki and was performed in accordance with the guidelines set forth by the ethics committee of the University of Duisburg-Essen under the Institutional Review Board number 15-6661-BO; written informed consent was obtained from each patient.

Rare Skin Cancer Virus Detection Pipeline

Unprocessed reads of targeted sequencing data from 214 skin tumor samples (WES, n = 6; IMPACT, n = 208; Supplementary Table S1) were analyzed with the Centrifuge metagenomics classifier (RRID:SCR_016665, v1.0.4) using a custom-build reference index comprising 6,453 viruses (Viral DB; ref. 4). Results tables were filtered for viruses with a minimum abundance of 1% and at least one unique genome aligning read. A collection of matched normal samples derived from the IMPACT cohort was analyzed in the same way, and viruses present in the normal sample cohort were excluded from further analysis in the tumor cohort. For each virus exclusive to the tumor cohort, a mean abundance score, as well as the viral prevalence, was calculated.

Virus Sequence Capture Assay (ViroCap)

Total RNA was extracted from FFPE sections using the Qiagen AllPrep DNA/RNA FFPE Kit (#80234, Qiagen), and rRNA was depleted using the NEBNext rRNA Depletion Kit (#E6310, NEB). Sequencing libraries were prepared with the NEBNext Ultra II RNA Library Kit (#E7760, NEB) and enriched for viral sequences using the Roche SeqCap EZ Choice Probes using the ViroCap targeted sequence capture panel (4).

Total DNA was extracted from FFPE sections using the Qiagen AllPrep DNA/RNA FFPE Kit (#80234, Qiagen). Sequencing libraries were prepared with the KAPA Hyper Prep Kit (#KK8504) and enriched for viral sequences using the Roche SeqCap EZ Choice Probes using the ViroCap targeted sequence capture panel (4). Samples were sequenced on an Illumina HiSeq4000 in 75-bp, paired-end mode. Unprocessed reads of virus sequence capture assay were analyzed using the Centrifuge metagenomics classifier (RRID:SCR_016665, v1.0.4) as described above. For each sample (DNA/RNA), abundance scores representing each detected viral species were calculated (Supplementary Tables S2 and S3).

Alpha Genus Phylogenetic Tree

L1, L2, E1, and E2 protein sequences of 65 human *Alphapapillomavirus* strains were downloaded from PaVE (RRID:SCR_016599), and each protein family was aligned with mafft (-linsi method; RRID:SCR_011811). For phylogenetic analysis, the aligned sequences from each strain were fused, and a maximum likelihood tree was inferred with iqtree2 (RRID:SCR_021163, v.2.1.3), with standard model selection using ModelFinder (39) and ultrafast bootstrap (UFBoot2) support values (40). The tree was visualized in iTOL (RRID:SCR_018174, v6). According to PaVE (RRID:SCR_016599), HPV177 is classified as Alpha-7 but is placed in the clade of the Alpha-11 group in the phylogenetic tree.

Viral Genome Analysis

Unprocessed reads from the virus sequence capture assay were processed with the Nextflow nf-core/viralrecon pipeline (https://nf-co.re/viralrecon, v2.4.1). Briefly, adapters were trimmed with fastp (RRID:SCR_016962, v0.23.2), reads aligned with Bowtie2 (RRID:SCR_016368, v2.4.4), and genomes sorted and indexed with SAMtools (RRID:SCR_002105, v1.14). Duplicates were marked with picard (RRID:SCR_006525, v2.26.10). Variant calling and consensus sequence generation were performed with BCFTools (RRID:SCR_002105, v1.14) and BEDTools (RRID:SCR_006646, v2.30.0; Supplementary Tables S6, S7, S12, and S13). Consensus genomes and reference genomes were compared using BLAST+ (ref. 41; v2.8.1; Supplementary Table S13). Viral genome coverage statistics were calculated using mosdepth (RRID:SCR_018929, v0.3.3; Supplementary Table S3). Genome coverage was visualized using the ggplot2 package (RRID:SCR_014601, v3.3.5) in R (https://www.r-project.org/, v.4.1.1).

HPV42–Human Genome Breakpoint Identification and Validation

For integration site calling, reads were aligned with bwa (RRID:SCR_010910, v.0.7.17) to a fusion genome of the human genome (hg38) and the HPV42 PaVE reference sequence. Duplicates were marked, alignments were sorted, and BAM files were indexed with sambamba (ref. 42; v0.6.6). Integration sites were called using the Manta structural variant caller (ref. 43; v1.6) in high-sensitivity calling mode (minEdgeObservations = 2, minCandidateSpanningCount = 2). Structural variants detected within the HPV42 genome are reported in Supplementary Table S14, and all HPV42–human genome breakpoints with a read support of at least two spanning- and two split-reads supporting the alternative over the reference allele were further analyzed and are reported in Supplementary Table S4. Based on assembled contig sequences covering the HPV42–human genome breakpoints, primers were designed for validation by PCR followed by Sanger sequencing (Supplementary Table S4). HPV42-human genome breakpoints within 500 kb were inferred as one integration event (44–46). HPV42–human genome breakpoints were visualized using the circlize (RRID:SCR_002141, v0.4.13), trackViewer (ref. 47; v1.32.1), sangerseqR (ref. 48; v1.32), and Gviz (ref. 49; v1.40.1) packages in R (v.4.1.1).

FISH

FISH was performed on 4-µm FFPE tumor sections using probes targeting HPV42 or a 10,000-bp region on the human genome adjacent to the integration site. Primary probes were designed using the OligoMiner package (50) using a fusion genome of HPV42 (PaVE) and human (hg38). Briefly, oligonucleotides were identified using the blockParse.py script (-l 36, -L 42 -t 42 -T 46 -S 2), candidates aligned using Bowtie2 (RRID:SCR_016368, v2.4.4; –no -hd -t -k 100 –very-sensitive-local) to the fusion genome, unique probes identified using outputClean.py (-T 42 -p 0.5), and high-abundance kmers filtered using kmerFilter.py (options -m 18 -k 5; Supplementary Table S5). Primary oligonucleotide probes containing handles for the annealing

of secondary fluorescent oligonucleotides were purchased from IDT as oPools Oligo Pools with a concentration of 10 pmol/oligonucleotide. Secondary oligonucleotide probes 5′-ATTO647-labeled CACCG ACGTCGCATAGAACGGAAGAGCGTGTG (HPV42), 5′-ATTO565-labeled CGAGCCAGGTCATCCTAGCCCATACGGCAATG (#170) and 5′-ATTO565-labeled TAGCGCAGGAGGTCCACGACGTGCAA GGGTGT (#184) were purchased from IDT.

Deparaffinized, protein-digested, and dehydrated tissue slides were preincubated for 20 minutes at 60°C in 2× SSCT (0.3 M NaCl, 0.03 M NaCitrate, 0.1% Tween-20) with 50% formamide (118143000, Sigma-Aldrich) and then incubated with a hybridization cocktail containing 2× SSCT, 50% formamide, 10% (w/v) dextran sulfate, 10 μg RNase A, 20 pmol/L primary oligonucleotide probe, and 40 pmol/L secondary oligonucleotide probe for 5 minutes at 95°C, followed by overnight incubation at 42°C in a humidified chamber as reported before (51). Slides were washed for 15 minutes at 60°C in 2× SCCT, followed by washes in 2× SSCT and 0.2× SSC for 10 minutes each at room temperature. Tissue sections were stained for 5 minutes in 0.2× SSC DAPI (1 μg/mL) solution in the dark, washed for an additional 3 minutes in 2× SSC, and mounted with Prolong Gold Antifade reagent (#P36970, Thermo Fisher Scientific).

FISH-stained tissue sections were acquired at room temperature on a Zeiss LSM980 microscope with the Airyscan2 detector, operated by the ZEN 3.3 (RRID:SCR_013672, blue edition) software. A 63 × 1.4NA oil-immersion DIC plan-apochromat objective (Zeiss) and 405/561/639-nm laser excitation (imaged sequentially with a track switch after every Z-Stack) were used to acquire Z-Stacks with 18 frames (33.7 μm × 33.7 μm) within a range of 3.2 μm (0.18 μm interval). Image sampling was conducted with a pixel size of 0.04 μm, a sampling speed of 0.83 μs per pixel, and the 8× averaging and calculation of the mean intensity per frame.

For quantitative analysis of fusion signals, 9 to 11 fields of view (FOV) per tissue slide and condition were recorded. Per FOV, a median of 5 (range, 3–9) nuclei with HPV42 and human signals in the same nucleus were analyzed using Fiji [(RRID:SCR_002285), ImageJ v2.3, RRID:SCR_003070]. Detection, colocalization, and quantification were performed using the ComDet (https://github.com/ekatrukha/ComDet, v.0.5.3) plug-in for ImageJ.

RNA-ISH

In situ detection of HPV42 mRNA was performed by the Vienna BioCenter Core Facilities (VBCF) Histology Facility using RNAscope 2.5 HD Assay-Red (#322360, Advanced Cell Diagnostics) using probe #548288 (RNAscope 2.5 LS Probe-V-HPV42-O1) according to the manufacturer's protocol. The 5-μm-thin FFPE sections were pretreated for 15 minutes with Target Retrieval Solution and for 30 minutes with Protease Plus. After hybridization, the mRNA signal was amplified and visualized with an alkaline phosphatase–based red substrate giving red dots as a positive signal. Housekeeping gene *PPIB* was used as a positive control to ensure mRNA quality, whereas the bacterial Diaminopimelate (dapB) gene was used as a negative control.

RT-qPCR Detection of HPV42

HPV42 was detected in cellular DNA by qPCR with amplicons in HPV42-E6 (F: ACAGCCACGCACATTATACCA; R: AAATGGTACG CGAGCACCTC) and HPV42-E7 (F: AACGTGTGAGACACCCATTGA; R: ATGTCCTGTTTGGCTTGGTCA). The genomic LINE1 locus (F: CGCTTTTCAGACCGGCTTAAG; R: AGATTCCGTGGGCGTAGGA) was used for data normalization. qPCRs were performed with the Luna Universal qPCR Master Mix (#M3003, New England BioLabs) on a Bio-Rad CFX384 Real-Time Cycler.

Expression Profiling of FFPE DPA Tumor Samples

Total RNA was extracted from FFPE DPA tumor samples using the MagMAX FFPE DNA/RNA Ultra Kit (#A31881, Life Technologies Limited) according to the manufacturer's instructions and eluted in nuclease-free water. Concentration was determined with the spectrophotometer/fluorometer DeNovix DS-11 Fx and assessed for RNA integrity on a Fragment Analyzer System (Agilent). RNA-seq (3′-end) mRNA sequencing libraries (QuantSeq) were prepared from 500 ng RNA using the QuantSeq 3′ mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen) and PCR Add-on Kit for Illumina kits (020.96, Lexogen) according to the manufacturer's instructions. Sequencing was performed on an Illumina NovaSeq 6000 in 100-bp, single-end read mode. Analysis of QuantSeq data was performed with an in-house pipeline: Briefly, adapter and polyA sequences were clipped (bbdk, RRID:SCR_016965, v38.06), abundant sequences were removed with bbmap (RRID:SCR_016965, v38.06), and cleaned reads were aligned against the genome (hg38) with STAR (RRID:SCR_004463, v2.6.0c). Raw reads were mapped to 3′ untranslated region (UTR) annotations of the same gene and collapsed to gene level by Entrez Gene ID with featureCounts (RRID:SCR_012919, v1.6.2). For visualization of HPV42 genome QuantSeq read coverage, reads were aligned with HISAT2 (RRID:SCR_015530, v2.1.0) to a fusion genome of hg38 and the HPV42 PaVE (RRID:SCR_016599) reference sequence; genome coverage tracks were calculated with BEDtools (RRID:SCR_006646, v2.27.1) and visualized with SparK (ref. 52; v2.6.2).

Generation of Full-Genome HPV-Transduced Keratinocytes

HPKs were isolated from human foreskin after circumcision upon written informed consent from the patients, which was approved by the ethics committee of the medical faculty of the University Tuebingen (6199/2018BO2) and performed according to the principles of the Declaration of Helsinki. HPKs were maintained in keratinocyte serum-free medium supplemented with recombinant human epidermal growth factor, bovine pituitary extract, and gentamicin (#15710064, Thermo Fisher Scientific). HPKs were transfected with 4 μg recircularized HPV genomes and 2 μg of pSV2-neo plasmid using Fugene HD (#E2311, Promega; ref. 53). Cells were selected with G418 (150 μg/mL, #ant-gn-1-1g, InvivoGen) in Dulbecco's modified Eagle's medium and Ham's F12 (3:1) supplemented with 5% (v/v) fetal bovine serum; 24 μg/L adenine, 0.4 ng/L hydrocortisone (#H0888, Sigma-Aldrich), 10 ng/L cholera toxin (#100B, List Biological Labs, Inc), 5 μg/L transferrin (#T1147, Sigma-Aldrich), 20 pmol/L 3,3′-5-triodo-L-thyronine, 5 ng/L epidermal growth factor (#E4127, Sigma-Aldrich) and 5 μg/L insulin (#91077, Sigma-Aldrich) and mitomycin C-treated NIH 3T3 J2 NHP until mock-transfected keratinocytes were dead (8–10 days; refs. 53, 54). G418-resistant colonies were expanded as pooled cultures. Cells were maintained continuously, and splitting dates and dilutions were recorded to generate growth curves. For short-term growth measurements, 1.5 × 105 HPV16 or 42-positive keratinocytes were seeded in 6-well dishes in the presence of growth-arrested NIH3T3 J2. At the indicated time points, fibroblasts were first removed by incubation with 0.5 mmol/L EDTA in PBS. Keratinocytes were detached by trypsinization, stained with trypan blue, and counted using an automated cell counter (Countess, Invitrogen).

Generation of HPV E6/E7-Transduced Keratinocytes

For lentivirus production, HEK-293T (Lenti-X) cells were cotransfected with the lentiviral backbone expressing the E6 and E7 proteins under the control of an SFFV promoter (pTwist-SFFV-E6-P2A-E7-T2A-mCherry-IRES-PURO-WPRE), VSVG (RRID:Addgene_12259) as enveloping plasmid, and the packaging plasmid PAX2 (RRID:Addgene_12260) in standard medium with polyethyleneimine as previously reported (55). Twenty-four hours after transfection, the culture medium was changed to 1% fetal bovine serum (FBS) medium, and viral supernatant was collected after 24 hours and filtered through a 0.4-μm mesh. For transduction of HPK cells, 5×10^6 HPK cells were seeded in a 60-mm plate and 24 hours after seeding

incubated with viral supernatant for 6 hours. Another 24 hours later, cells were expanded into 100-mm plates and selected with 2 μg/mL puromycin. On day 3 after transduction, 5×10^6 HPK cells were seeded in a 6-well plate. Cells were counted each third day and reseeded with a 1:3 split ratio. The remaining cells were plated for crystal violet (CV) staining. HEK-293T cells were purchased from Takara (Lenti-X 293T, #632180).

Generation of HPV E6/E7-Transduced MEFs

Eight-week-old C57BL/6 female mice were euthanized 13.5 to 14.5 days after the appearance of the copulation plug by cervical dislocation to remove the uterine horns. Embryos within the uterine horn were collected by extraction of the embryonic sacs, beheaded, red tissue removed, and the remaining tissue minced into pieces of 1- to 2-mm size and resuspended in PBS. For each 1 mL of tissue suspension, 1 mL 0.25% trypsin was added, and the tissue was digested at 37°C for 10 minutes. The tissue suspension was further homogenized using a 21G needle, and the digestion was quenched by the addition of 5 mL growth medium (DMEM + 10% FBS). Cells were spun at 400 g for 5 minutes, plated in a growth medium, expanded the following day, and frozen in freezing medium (10% DMSO + 90% FBS). MEF (passage 2) cells were used for counting and staining assays. Two-hundred-thousand MEF cells were seeded per well in a 6-well plate. Twenty-four hours after seeding, cells were incubated with lentivirus encoding for the E6 and E7 proteins under the control of an SFFV promoter (pTwist-SFFV-E6/E7-IRES-PURO-WPRE) in the presence of 8 μg/mL polybrene and spun at 800 × *g* for 1 hour at 35°C. Viral supernatant was removed 24 hours after spinfection and replaced with fresh medium. On day 3 after transduction, 100,000 MEFs were seeded at equal transduction efficiency in a 6-well plate. Cells were counted each third day and reseeded with a 1:3 split ratio. The remaining cells were plated for CV staining. Cells were tested for *Mycoplasma* regularly.

CV Staining

The cell culture medium was removed, and cells were washed twice with PBS and fixed for 5 minutes with 1.5% PFA in PBS. PFA was removed, and cells were washed 3× with PBS and stained with CV solution (0.5% CV, 1% formaldehyde, 1% MeOH) for 5 minutes at room temperature. Finally, the CV solution was removed, and cells were washed 3× with PBS.

Viral Copy-Number Determination and Exonuclease V-Resistance Assay

Total cellular DNA was isolated by proteinase K digestion, followed by phenol/chloroform/isoamyl alcohol extraction and ethanol precipitation. HPV copy numbers were quantified in cellular DNA by qPCR detecting amplicons in HPV16 (HPV16-E6 F: GAGAACTGC AATGTTTCAGGACC, HPV16-E6 R: TGTATAGTTGTTTGCAGCTCT GTGC), HPV42 (HPV42-E7 F: ATGACCAAGCCAAACAGGAC, HPV42-E7 R: CAATATCCAGTGTGCCCAAA), and the cellular *ACTB* gene (ACTB F: GGAAGGTCCGTGCGAGG, ACTB R: GAGACGGAG CAGGTCCCA) and copy-number standards using a LightCycler 480 and the LightCycler 480 SYBR green I master mix (Roche Applied Science). The exonuclease V resistance assay was adapted from (56) with minor modifications: Total cellular DNA (100 ng) was incubated in the presence or absence of 5U of exonuclease V (#M0345S, NEB) in 1× NEBuffer 4 supplemented with 1 mmol/L ATP for 60 minutes at 37°C. Then, the enzyme was inactivated for 10 minutes at 95°C. Finally, 10 ng of input DNA was measured by qPCR using primers for HPV16, HPV42, and *ACTB*.

In Vitro *p53* Degradation Assay

TP53 and N-terminally V5-tagged E6 proteins of HPV11, 16, and 42 were cloned under the control of an SP6 promoter flanked by Xenopus globin 5′ and 3′ UTRs. Plasmids were linearized and transcribed for 12 hours at 37°C (#AM1340, mMessage mMACHINE SP6 Transcription Kit). Template DNA was removed by the addition of 1μL DNAse (#AM2238, TURBO DNAse) for 15 minutes at 37°C. Transcribed mRNA was purified using RNeasy MinElute Cleanup Kit (#74204, Qiagen) according to the manufacturer's instructions. The E6 and p53 proteins were *in vitro* translated for 90 minutes at 30°C using an in-house rabbit reticulocyte system (57). The p53 degradation reactions were performed in 10-μL volumes by incubating 2 μL of p53 translation product with 5 μL of E6 translation products at 28°C for 2 hours as previously described (58) and analyzed by immunoblotting.

DNA Damage–p53 Response Assay

U2OS (RRID:CVCL_0042, generously provided by Egon Ogris, Max Perutz Labs, Vienna, Austria) and N/TERT-1 (RRID:CVCL_CW92, generously provided by James Rheinwald, Department of Dermatology, Brigham and Women's Hospital, Boston, MA; ref. 59) cells transduced with E6 genes of HPV11, HPV16, HPV42, or EV (pTwist-SFFV-E6-IRES-PURO-WPRE) were treated with 1 μg/mL mitomycin C or DMSO for 18 hours as previously described (60). Cells were harvested and analyzed by immunoblotting and qPCR using the following primers: TP53 (F: GAG GTTGGCTCTGACTGTACC, R: TCCGTCCCAGTAGATTACCAC), BAX (F: CCCGAGAGGTCTTTTTCCGAG, R: CCAGCCCATGATGGTTCT GAT), MDM2 (F: CAGTAGCAGTGAATCTACAGGGA, R: CTGATC CAACCAATCACCTGAAT), FAS (F: AGATTGTGTGATGAAGGACAT GG, R: TGTTGCTGGTGAGTGTGCATT), and HPRT1 (F: TGACACTG GCAAAACAATGCA, R: GGTCCTTTTCACCAGCAAGCT). Cells were kept at a low passage and tested for *Mycoplasma* regularly.

Immunoblotting

Cells were lysed with RIPA buffer (#9806, Cell Signaling Technology) supplemented with cOmplete Protease Inhibitor Cocktail (#5056489001, Sigma-Aldrich) and HALT phosphatase inhibitor (#78427, Thermo Fisher Scientific). Lysates were sonicated and cleared by centrifugation at 14,000 g for 10 minutes at 4°C. According to the manufacturer's instructions, protein concentrations were determined with the Pierce BCA Protein Assay Kit (#23225, Thermo Fisher Scientific). Immunoblotting was conducted according to standard protocols. The primary antibodies used for immunoblotting were as follows: anti-Vinculin (Sigma-Aldrich; cat. #V9131, RRID:AB_477629, 1:1,000), anti-E6AP (Santa Cruz Biotechnology; cat. #sc-166689, RRID:AB_2211807, 1:500), anti-p53 (Santa Cruz Biotechnology; cat. #sc-126, RRID:AB_628082, 1:500), anti-RPL3 (GeneTex; cat. #GTX124464, RRID:AB_11169700, 1:1,000), anti-RB1 (Santa Cruz Biotechnology; cat. #sc-102, RRID:AB_628209, 1:500), anti-PEZ (Santa Cruz Biotechnology; cat. #sc-373766, RRID:AB_10917236, 1:500), anti-p21 (Cell Signaling Technology; cat. #2947), anti-H3 (Abcam; cat. #ab1791, RRID:AB_302613, 1:5,000), anti–α-tubulin (Cell Signaling Technology; cat. #2125, RRID:AB_2619646), 1:1,000), anti-V5 (Sigma-Aldrich; cat. #V8012, RRID:AB_261888, 1:5,000), and anti-CDKN2A/p16INK4 (Abcam; cat. #ab108349, RRID:AB_10858268, 1:1,000). The secondary antibodies used were as follows: anti-rabbit IgG HRP-linked (Cell Signaling Technology; cat. #7074, RRID:AB_2099233, 1:5,000), anti-mouse IgG HRP-linked (Cell Signaling Technology; cat. #7076, RRID:AB_330924, 1:5,000), and anti-mouse IgG light chain-specific HRP-linked (Cell Signaling Technology; cat. #58802, RRID:AB_2799549, 1:5,000).

E6/E7 Coimmunoprecipitation

Coimmunoprecipitation (co-IP) of E6 and E7 proteins were performed as previously reported (61). U2OS cells (RRID:CVCL_0042) were transfected with plasmids encoding 3xV5-tagged E6 proteins and 3xMYC-tagged E7 proteins under the control of a constitutive SFFV promoter (pTwist-SFFV-E6-P2A-E7-T2A-mCherry-IRES-PURO-WPRE). For E6 protein co-IPs, U2OS cells were cotransfected with pLX313-TP53-WT (RRID:Addgene_118014). Twenty-four hours

after transfection, cells were treated for 24 hours with 1 μmol/L of the proteasome inhibitor MG-132 (#S2619, Selleck Chemicals). Cells were washed with ice-cold PBS and lysed with 1 × Cell Lysis Buffer (#9803, Cell Signaling Technology). Lysates were cleared by centrifugation at 14,000 × *g* for 10 minutes at 4°C, and the lysate concentration was determined with the Pierce BCA Protein Assay Kit (#23227, Thermo Fisher Scientific) according to the manufacturer's instructions. Lysates were incubated with α-V5 agarose beads (#A7345, Sigma-Aldrich) or α-MYC agarose beads (#A7470, Sigma-Aldrich) for 2 hours at 4°C and washed 3× in 1 × Cell Lysis Buffer. Confirmation of successful co-IP was conducted by immunoblotting according to standard procedures.

E2F Reporter

U2OS cells (RRID:CVCL_0042) were transduced with a lentiviral backbone expressing the E7 genes of HPV11, 16, or 42 or EV under the control of an SFFV promoter (pTwist-SFFV-E7-IRES-PURO-WPRE) and selected for puromycin resistance. Transduced cells were transfected with plasmids expressing destabilized GFP (GFPd2) under the control of an E2F DNA-binding site containing promoter (62). GFPd2 expression was assessed by flow cytometry 48 hours after transfection. Single-cell events (2×10^4) were acquired per sample on a FACS LSRFortessa cytometer, and GFP-positive fractions were determined with the FlowJo10 software (RRID:SCR_008520).

MS: Quantitative MS

Cells (1×10^6 cells) were washed twice with ice-cold PBS, and proteins were prepared for MS analysis using the iST-NHS kit (P.O.00030, PreOmics) together with the TMT10plex Isobaric Label Reagent Set (#90110, Thermo Fisher Scientific) according to the manufacturer's instructions.

nanoLC-MS/MS Analysis

Samples were analyzed on an UltiMate 3000 RSLC nano system (Thermo Fisher Scientific) coupled to an Exploris 480 mass spectrometer equipped with a FAIMS pro interface and a Nanospray Flex ion source (Thermo Fisher Scientific). Peptides were loaded onto a trap column (PepMap Acclaim C18, 5 mm × 300 μm ID, 5 μm particles, 100 Å pore size, Thermo Fisher Scientific) at a flow rate of 25 μL/minute using 0.1% TFA as mobile phase. After 10 minutes, the trap column was switched in line with the analytical column (PepMap Acclaim C18, 500 mm × 75 μm ID, 2 μm, 100 Å, Thermo Fisher Scientific). Peptides were eluted using a flow rate of 230 nL/minute, starting with the mobile phases 98% A (0.1% formic acid in water) and 2% B (80% acetonitrile, 0.1% formic acid) and linearly increasing to 35% B over the next 120 or 180 minutes. This was followed by a steep gradient to 90% B in 5 minutes, which stayed there for 5 minutes and ramped down in 2 minutes to the starting conditions of 98% A and 2% B for equilibration at 30°C.

For tandem mass tag (TMT) experiments, the Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific) was operated in data-dependent mode, performing a full scan (m/z range, 350–1,200, resolution 60,000, normalized AGC target 1×10^6) at three different compensation voltages (CV; 40, 55, 70), followed each by MS/MS scans of the most abundant ions for a cycle time of 1 second per CV. MS/MS spectra were acquired using a collision energy of 34, isolation width of 0.7 m/z, resolution of 45.000, fill time of 120 ms, normalized AGC target of 2×10^5, and intensity threshold of 1 × 104. Precursor ions selected for fragmentation (including charge states 2–6) were excluded for 45 seconds.

Data Processing Protocol for TMT Experiments

For peptide identification, the RAW files were loaded into Proteome Discoverer (RRID:SCR_014477, v2.5.0.400). All hereby created MS/MS spectra were searched using MSAmanda (ref. 63; v2.5.0.16129) against the human Uniprot reference database (2021-06-30, 20,531 sequences; 11,395,157 residues), supplemented with common contaminants. The following search parameters were used: Carbamidomethyl on cysteine was set as a fixed modification, oxidation on methionine, deamidation of asparagine and glutamine, carbamylation and 10-plex/16-plex TMT on lysine and on peptide-N-term were set as variable modifications. The peptide mass tolerance was set to ±10 ppm and the fragment mass tolerance to ±10 ppm. The maximal number of missed cleavages was set to 2, and the minimum peptide length was set to 7. The result was filtered to 1% false discovery rate (FDR) on PSM and protein level using the Percolator algorithm integrated in Thermo Proteome Discoverer. Additionally, a minimum MS Amanda score of 150 was applied on the PSM level, and proteins had to be identified by a minimum of two PSMs. The localization of the modification sites within the peptides was performed with the tool ptmRS, which is based on phosphoRS (64). Peptides were quantified based on Reporter Ion intensities extracted by the "Reporter Ions Quantifier"-node implemented in Proteome Discoverer. Proteins were quantified by summing unique and razor peptides. Protein-abundances-normalization was done using sum normalization. The statistical significance of differentially expressed proteins was determined using limma (RRID:SCR_010943).

BioID Sample Preparation

Briefly, Flp-In T-REx HEK293 cells were grown in DMEM (Gibco) supplemented with 10% FBS (Sigma-Aldrich), GlutaMAX, and penicillin–streptomycin (P/S, 1×). Using the Flp-In system (Invitrogen), Flp-In T-REx HEK293 stably expressing BirA* alone (for control samples) or N- and C-terminally tagged HPV16 and 42 E6 and E7 viral proteins were generated cotransfecting pOG44 with each pcDNA5 FRT/TO BirA*Flag-viral bait protein sequence plasmid. After selection (DMEM + 10% FBS + P/S + 200 μg/mL hygromycin B), three independent replicates of two 150 cm^2 plates of subconfluent (60%) cells were incubated for 24 hours in complete media supplemented with 1 μg/mL tetracycline (Sigma) and 50 μmol/L biotin (Thermo Fisher Scientific). Cells were collected and pelleted (2,000 rpm, 3 minutes) and washed twice with PBS, and dried pellets were snap-frozen. Each cell pellet was resuspended in 5 mL of lysis buffer [50 mmol/L Tris-HCl pH 7.5, 150 mmol/L NaCl, 1 mmol/L EDTA, 1 mmol/L EGTA, 1% Triton X-100, 0.1% SDS, 1:500 protease inhibitor cocktail (Sigma-Aldrich), 1:1,000 Turbonuclease (Accelagen)], incubated on an end-over-end rotator at 4°C for 1 hour, briefly sonicated to disrupt any visible aggregates, and then centrifuged at 45,000 × *g* for 30 minutes at 4°C. The supernatant was transferred to a fresh 15-mL conical tube. Twenty-five microliters of packed, preequilibrated Streptavidin Ultralink Resin (Pierce) were added, and the mixture was incubated for 3 hours at 4°C with end-over-end rotation. Beads were pelleted by centrifugation at 2,000 rpm for 2 minutes and transferred with 1 mL of lysis buffer to a fresh Eppendorf tube. Beads were washed once with 1 mL lysis buffer and twice with 1 mL of 50 mmol/L ammonium bicarbonate (pH = 8.3), and then transferred in ammonium bicarbonate to a fresh centrifuge tube and washed two more times with 1 mL ammonium bicarbonate buffer. Tryptic digestion was performed by incubating the beads with 1 μg MS-grade TPCK trypsin (Promega) dissolved in 200 μL of 50 mmol/L ammonium bicarbonate (pH 8.3) overnight at 37°C. The following morning, 0.5 μg MS-grade TPCK trypsin was added to the beads and incubated 2 additional hours at 37°C. Following centrifugation at 2,000 × *g* for 2 minutes, the supernatant was collected and transferred to a fresh Eppendorf tube. Two additional washes were performed with 150 μL of 50 mmol/L ammonium bicarbonate and pooled with the first eluate. The sample was lyophilized and resuspended in buffer A (2% ACN 0.1% formic acid). One third of each sample was analyzed per mass spectrometer run.

BioID Data Acquisition

Samples were separated by online reversed-phase chromatography using a Thermo Scientific Easy-nLC1000 system equipped with a Proxeon trap column (75 μm ID × 2 cm, 3 μm, Thermo Scientific) and a C18 packed-tip column (Acclaim PepMap, 75 μm ID × 50 cm, 2 μm, Thermo Scientific). The digested peptides were separated using an increasing amount of acetonitrile in 0.1% formic acid from 2% to 30% for 2 hours at a flow rate of 300 nL/minute. A voltage of 2.4 kV was applied by the liquid junction to electrospray the eluent using the nanospray source. A high-resolution mass spectrometer (Q-Exactive, Thermo Scientific) was coupled to the chromatography system to acquire the 10 most intense ions of MS1 analysis (Top 10) in data-dependent mode. The MS analyses were performed in positive mode at a resolving power of 70,000 FWHM, using an automatic gain control target of 3×10^6, and the default charge state was set at 2 and a maximum injection time at 120 ms. For full scan MS, the scan range was set between m/z 300 and 1,600. For ddMS², the scan range was between m/z 200 to 2000, 1 microscan was acquired at 17,500 FWHM, an AGC was set at 5×10^4 ions and an isolation window of m/z 4.0 was used.

BioID Data Analysis

The proteins were identified by comparing all MS/MS data with the Homo sapiens proteome database (Uniprot, release April 2021, Reviewed, comprising 20,360 entries + viral bait protein sequences added manually), using the MaxQuant software (RRID:SCR_014485, v1.5.8.3). The digestion parameters were defined using trypsin with two maximum missed cleavages. The oxidation of methionine and N-terminal protein acetylation were defined as variable modifications. The label-free quantification (LFQ) was done keeping the default parameters of the software. As for initial mass tolerance, 6 ppm was selected for MS mode, and 20 ppm was set for fragmentation data to match MS/MS tolerance. The identification parameters of the proteins and peptides were performed with an FDR at 1%, and a minimum of two unique peptides per protein. The LFQ values from the 20 control runs (BirA* alone samples from stable cell lines) were collapsed to the three highest values for each given ID. These three values were defined as the control group for comparison with viral bait proteins triplicates. The statistical analysis was done by Perseus software (RRID:SCR_015753, v1.6.2.3). Briefly, the LFQ intensity of each sample was downloaded in Perseus, and the data matrix was filtered by removing the potential contaminants, reversed, and only identified by site. The data were then transformed using the $\log_2(x)$ function. Before statistical analysis, eight (four viral bait proteins, N-ter, C-ter BirA* tag for each) were defined with three replicates per group. Only preys with detected values in all three replicates of a given viral bait protein were kept for further analysis. Missing values were then replaced from the normal distribution separately for each column. Two-sample Student *t* test was then performed comparing all three biological replicates of each bait and condition against the three controls runs. High-confidence proximal interactors were defined by permutation-based FDR with a cutoff of 0.01. Supplementary Tables S8 and S9 show the average $\log_2$-fold change against control and the corresponding q-values for each bait and condition.

RNA Extraction from HPK, N/TERT-1, and U2OS Cells

Total RNA isolation from cells was performed with an in-house paramagnetic bead-based purification protocol. Briefly, cells were lysed, incubated with functionalized paramagnetic beads, and washed to remove proteins. Beads were then treated with DNase I, and purified total RNA was eluted in nuclease-free water. RNA concentration was determined with the spectrophotometer/fluorometer DeNovix DS-11 Fx and assessed for integrity on a Fragment Analyzer System (Agilent).

Hematoxylin and Eosin and IHC Staining

Sections were stained with hematoxylin (#6765002, Thermo Scientific) and eosin (#6766008, Thermo Scientific; H&E) for histologic examination using the Thermo Scientific Gemini AS Automated Slide Stainer. For IHC, FFPE tissues were sectioned at 2 μm (HM355S Rotary Microtome), deparaffinized, and rehydrated. Antigens were retrieved using EDTA retrieval (Tris-EDTA, pH 8.5, 0.05% Tween-20) at 100°C for 30 minutes. Endogenous peroxidase was quenched for 10 minutes in 3% H_2O_2 diluted in water and slides blocked with 10% BSA and 10% goat serum in TBST for 1 hour at room temperature. IHC with anti-YAP1 and anti-SYCP2 was performed using DAB chromogen and hematoxylin counterstain. The following antibodies against human epitopes were used: anti-YAP1 (Cell Signaling Technology; cat. #14074, RRID:AB_2650491, 1:200 in 2% TBST, 1 hour) and anti-SYCP2 (Atlas Antibodies; cat. #HPA062401, RRID:AB_2684752, 1:200 in 2% TBST, 2 hours).

QuantSeq Analysis of HPV/E7-Transduced HPKs

HPKs were washed in PBS, and cell pellets were snap-frozen. RNA extraction was performed using an in-house total RNA magnetic beads–based purification protocol. RNA-seq (3′-end) mRNA sequencing libraries (QuantSeq) were prepared from 500 ng RNA using the QuantSeq 3′ mRNA-Seq Library Prep Kit FWD for Illumina (#015.96, Lexogen) and PCR Add-on Kit for Illumina kits (#020.96, Lexogen) according to the manufacturer's instructions. Sequencing was performed on an Illumina NovaSeq 6000 in 100-bp, single-end mode. Analysis of QuantSeq data was performed with an in-house pipeline. Briefly, adapter and polyA sequences were clipped, and abundant sequences were removed with bbmap (RRID:SCR_016965, v38.06). Cleaned reads were aligned against the genome (hg38) with STAR (RRID:SCR_004463, v2.6.0c). Raw reads were mapped to 3′ UTR annotations of the same gene and collapsed to gene level by Entrez Gene ID with featureCounts (RRID:SCR_012919, v1.6.2). Differentially expressed genes were calculated using DESeq2 (RRID:SCR_015687, v1.18.1). Pathway enrichment analysis was performed using g:Profiler (RRID:SCR_006809) and Metascape (RRID:SCR_016620). Gene set enrichment analysis (65, 66) was performed using the FGSEA (RRID:SCR_020938, v1.19.2). Gene-based motif analysis was performed using Homer (RRID:SCR_010881, v4.11) with the findMotifs.pl script and default settings.

In Vivo *Tumor Growth of E7-Transduced HaCaT Cells*

HaCaT cells (RRID:CVCL_0038) were transduced by spinfection (30 minutes, 800 × *g*, 37°C) with lentiviral backbones expressing the E7 gene of HPV11, HPV16, or HPV42 or EV under the control of an SFFV promoter. Transduced cells were selected with puromycin. HaCaT cells (2×10^6) mixed with an equal volume of Matrigel (#3562377, Corning) were injected subcutaneously into the left and right flanks of 5- to 8-week-old sex-randomized NSG mice. NSG (NOD-scid IL2R gamma null) mice were purchased from The Jackson Laboratory (strain 005557). Prior to tumor injection, the mice were injected intraperitoneally with analgesic ketamine (100 mg/kg), xylazine (10 mg/kg), and acepromazine (3 mg/kg). Tumor size was monitored by tumor measurement with an electronic caliper. Measured tumor size was determined by the formula [x (mm)² * y (mm)]/2 = volume (mm³), where x refers to the shortest measurement and y the longest. Mice were sacrificed by cervical dislocation when the total tumor volume on either side was larger than 1,000 mm³. All experiments using animals were performed in accordance with the protocol approved by the Austrian Ministry (BMBWF-66.015/0009-V/3b/2019 or GZ: 340118/2017/25). Survival curves were plotted with the software GraphPad PRISM 8 (RRID:SCR_002798, v8).

BLAST-Based Identification of HPV Genome Overlaps between Databases

A BLAST database was built from the reference fasta file using makeblastdb (BLAST+; ref. 41; v2.8.1), and query sequences were searched against this BLAST database with blastn (-outfmt 6 -max_target_seqs 1 -max_hsps 1). Query sequences with a sequence match ≥95% and sequence coverage ≥35% were considered as represented in the reference database.

TCGA Reanalysis for HPV Associations

A custom-built Papillomavirus Episteme (PaVE, RRID:SCR_016599) database–based index for the Centrifuge metagenomics classifier (RRID:SCR_016665, v1.0.4) consisting of the human hg38 genome and 376 papillomavirus genomes (obtained from PaVE on August 24, 2018) was generated. All subsequent processing was performed using the Nextflow pipeline virus-detection-nf found under https://github.com/ObenaufLab/virus-detection-nf. Briefly, raw paired/unpaired TCGA tumor bam files (32 tumor types with a total of 10,078 tumor samples) were converted to fastq files using samtools (RRID:SCR_002105, v1.9) and analyzed using virus-detection-nf with Centrifuge (RRID:SCR_016665, v1.0.4). For each species, we calculated reads-per-million (RPM) and median absolute deviations as z-score across all samples. To classify a given sample as positive for a given virus, we require an RPM ≥5 and a corresponding P value of the z-score ≤0.05 for that specific virus.

Identification of RF12 Gene Signature

TCGA-CESC and TCGA-HNSC expression data were downloaded from TCGA and annotated with HPV status based on our PaVE-based classification. Differentially expressed genes between HPV^+ and HPV^- samples were independently called on raw counts of TCGA-CESC and TCGA-HNSC using DESeq2 (RRID:SCR_015687, v1.22.2). Genes with ≥10 counts and an adjusted $P < 0.1$ were considered differentially expressed and resulted in a gene set of 942 genes, which are commonly differentially expressed in TCGA-CESC and TCGA-HNSC with $\log_2$-fold changes in the same direction between HPV^+ and HPV^- samples.

Expression data [transcripts per million (TPM)] of combined TCGA-CESC and TCGA-HNSC datasets were split into a training (60%) and a test (40%) set. Hyperparameters for an RF consisting of 1,000 trees were tuned, and a final RF with tuned parameters was fit to the training set using 10-fold cross-validation with the R packages tidymodels (https://www.tidymodels.org/, v0.1.4) and ranger (RRID:SCR_022521, v0.13.1). Importance scores for each gene were extracted, and the resulting gene list was filtered for protein-coding genes and ordered by decreasing importance. The top 12 genes covering a cumulative importance of 87% were picked as gene signature (RF12). Significantly enriched promoter motifs in RF12 signature genes were identified using Homer (RRID:SCR_010881, v4.11) with the findMotifs.pl script and default parameters. Enriched pathways in RF12 signature genes were identified using g:Profiler (RRID:SCR_018190) with default parameters. Cancer–testis genes were annotated based on a previous classification (67). RF12 gene expression in 76 cell types was obtained from the HPA project (68). Genes with a minimum expression level of 10 normalized transcripts per million (nTPM) were plotted. Cell type–specific enrichment analysis for RF12 genes was performed using PanglaoDB (RRID:SCR_022580). MHC-I antigens were predicted for RF12 genes, known cancer/testis antigens (MAGEA3, MAGEA4, MAGEA10, and NYESO1) and HPV16 and HPV42 oncogenes E6 and E7 using NetMHCPan (RRID:SCR_018182, v4.1) with a peptide length of 9. Strong predicted binders were all peptides with a predicted %Rank <0.5 and weak predicted binders with a predicted %Rank <2.

Machine Learning–Based Classification of HPV^+ Tumor Samples

Expression data (TPM) of TCGA-CESC and TCGA-HNSC were combined with DPA expression data (CPM) and annotated with HPV status based on our PaVE-based classification. The resulting sample set was split into a training (60%) and a test (40%) set. For classification, we tuned hyperparameters for logistic regression models with glmnet (RRID:SCR_015505, v2.0–16), support vector machines with kernlab (ref. 69; v0.9-29), RFs with 1,000 trees using ranger (RRID:SCR_022521, v0.13.1), and single-layer neural networks with five hidden units using nnet (70; v7.3-12) in the tidymodels (https://www.tidymodels.org/, v0.1.4) framework. Each model was fitted with tuned parameters to the training set using 10-fold cross-validation. AUCs were extracted for each model on the test set, which yielded the random forest as the best-performing classifier for detecting HPV^+ samples based on our RF12 gene signature. RF classification models based on CDKN2A, SYCP2, and RF2 (CDKN2A + SYCP2) were built as described above. GTEx normal tissue (obtained from the GTEx portal on May 11, 2021), HSIL (GSE150227), and normal skin/wart (GSE136347) expression data were classified with the indicated classification models.

Data Availability

The datasets used in this study are available in the following databases: ViroCap DNA- and RNA-seq targeted sequencing data have been deposited in the Sequence Read Archive under accession number PRJNA812756. QuantSeq data of HPV-E7–transduced HPKs (GSE196217), HPV full-genome transduced HPKs (GSE196216), and DPA tumors (GSE196215) have been deposited in the Gene Expression Omnibus (GEO) database. The MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the accession number PXD031764 for quantitative TMT-MS and PXD031438 for proximity-labeling MS data. Public RNA-seq data used in this study are available from GEO under the accession numbers GSE136347 (skin wart and matched normal skin RNA-seq data), GSE150227 [normal, cervical intraepithelial neoplasia (CIN), and CESC RNA-seq data], and GSE121906 (noncanonical HIPPO signature). GTEx expression data were obtained from https://gtexportal.org/home/, TCGA expression data were obtained from https://gdc.cancer.gov/, and TCGA raw sequencing data were accessed via the database of Genotypes and Phenotypes (dbGaP) accession number phs000178.v11.p8.

Authors' Disclosures

R. Murali reports grants from the NIH/NCI during the conduct of the study. A.C. Obenauf reports research support and funding from Boehringer Ingelheim. No disclosures were reported by the other authors.

Authors' Contributions

L. Leiendecker: Conceptualization, data curation, formal analysis, visualization, methodology, writing–original draft, writing–review and editing. **T. Neumann:** Data curation, software, investigation, visualization, writing–review and editing. **P.S. Jung:** Investigation, visualization, writing–review and editing. **S.M. Cronin:** Investigation, visualization, writing–review and editing. **T.L. Steinacker:** Investigation, writing–review and editing. **A. Schleiffer:** Investigation, visualization, writing–review and editing. **M. Schutzbier:** Investigation, writing–review and editing. **K. Mechtler:** Supervision, writing–review and editing. **T. Kervarrec:** Investigation, writing–review and editing. **E. Laurent:** Investigation, writing–review and editing. **K. Bachiri:** Investigation, writing–review and editing. **E. Coyaud:** Data curation, supervision, writing–review and editing. **R. Murali:** Resources, writing–review and editing. **K.J. Busam:** Resources. **B. Itzinger-Monshi:** Resources. **R. Kirnbauer:** Resources, writing–review and editing.

L. Cerroni: Resources, writing–review and editing. **E. Calonje:** Resources, writing–review and editing. **A. Rutten:** Resources, writing–review and editing. **F. Stubenrauch:** Resources, investigation, visualization, writing–review and editing. **K.G. Griewank:** Resources, investigation, writing–original draft, writing–review and editing. **T. Wiesner:** Conceptualization, resources, supervision, funding acquisition, investigation, visualization, writing–original draft, writing–review and editing. **A.C. Obenauf:** Conceptualization, supervision, funding acquisition, visualization, writing–original draft, project administration, writing–review and editing.

Acknowledgments

We are grateful to all members of the Obenauf laboratory and to Johannes Zuber, Egon Ogris, Francisco Balzarotti, Baki Akgül, Frank Rösl, Bojan Zagrovic, and Anton Polyansky for experimental advice and helpful discussions. We thank Izabela Krecioch, Ha Pham, Konstantina Georgakopoulou, and Hacer Ezgi Karakas Schüller for their technical support. QuantSeq sequencing was performed by the Next-Generation Sequencing Facility, and HPV42 RNA-ISH and YAP1 and SYCP2 IHC were performed by the Histology Facility at VBCF, member of the Vienna BioCenter (VBC), Austria. FISH was performed in collaboration with Leonhard Müllauer and the Molecular Pathology lab at the Department of Pathology, Medical University of Vienna, Austria. We thank Joakim Dillner for providing the HPV11, HPV16, and HPV42 reference clones obtained from the International HPV Reference Center, Karolinska Institute, Stockholm, Sweden. The results published here are in part based upon data generated by The Cancer Genome Atlas managed by the NCI and National Human Genome Research Institute (NHGRI). Information about TCGA can be found at http://cancergenome.nih.gov. The Genotype-Tissue Expression (GTEx) Project is supported by the Common Fund of the Office of the Director of the NIH and by the NCI, NHGRI, National Heart, Lung, and Blood Institute, National Institute on Drug Abuse, National Institute of Mental Health, and National Institute of Neurological Disorders and Stroke. A.C. Obenauf was supported by the Starting Grant of the European Research Council, ERC-StG-759590. A.C. Obenauf and T. Wiesner were supported by the Vienna Science and Technology fund, #LS16063. Research at the IMP is generously supported by Boehringer Ingelheim and the Austrian Research Promotion Agency (headquarter grant FFG-852936). R. Murali was supported by NIH/NCI Cancer Center Support Grant P30 CA008748 and K.G. Griewank by Deutsche Forschungs-Gemeinschaft DFG 36714-1.

Note

Supplementary data for this article are available at Cancer Discovery Online (http://cancerdiscovery.aacrjournals.org/).

Received April 29, 2022; revised August 24, 2022; accepted October 4, 2022; published first October 10, 2022.

REFERENCES

1. Zapatka M, Borozan I, Brewer DS, Iskar M, Grundhoff A, Alawi M, et al. The landscape of viral associations in human cancers. Nat Genet 2020;52:320–30.
2. Moore PS, Chang Y. Why do viruses cause cancer? Highlights of the first century of human tumour virology. Nat Rev Cancer 2010;10:878–89.
3. Guo Y, Long J, He J, Li CI, Cai Q, Shu XO, et al. Exome sequencing generates high-quality data in non-target regions. BMC Genomics 2012;13:194.
4. Wylie TN, Wylie KM, Herter BN, Storch GA. Enhanced virome sequencing using targeted sequence capture. Genome Res 2015;25:1910–20.
5. Elder DE, Massi D, Scolyer RA, Willemze R, editors. WHO classification of skin tumours: WHO classification of tumours. 4th ed. Vol. 11. Geneva (Switzerland): World Health Organization; 2018.
6. Philipp W, Honoré N, Sapp M, Cole ST, Streeck RE. Human papillomavirus type 42: new sequences, conserved genome organization. Virology 1992;186:331–4.
7. Arroyo Mühr LS, Lagheden C, Lei J, Eklund C, Nordqvist Kleppe S, Sparén P, et al. Deep sequencing detects human papillomavirus (HPV) in cervical cancers negative for HPV by PCR. Br J Cancer 2020;123:1790–5.
8. Muñoz N, Bosch FX, de Sanjosé S, Herrero R, Castellsagué X, Shah KV, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med 2003;348:518–27.
9. Cancer Genome Atlas Research Network. Integrated genomic and molecular characterization of cervical cancer. Nature 2017;543:378–84.
10. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature 2015;517:576–82.
11. Jeon S, Lambert PF. Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: implications for cervical carcinogenesis. Proc Natl Acad Sci U S A 1995;92:1654–8.
12. Romanczuk H, Howley PM. Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity. Proc Natl Acad Sci U S A 1992;89:3159–63.
13. Chen J, Xue Y, Poidinger M, Lim T, Chew SH, Pang CL, et al. Mapping of HPV transcripts in four human cervical lesions using RNAseq suggests quantitative rearrangements during carcinogenic progression. Virology 2014;462–463:14–24.
14. Storey A, Pim D, Murray A, Osborn K, Banks L, Crawford L. Comparison of the in vitro transforming activities of human papillomavirus types. EMBO J 1988;7:1815–20.
15. Dürst M, Dzarlieva-Petrusevska RT, Boukamp P, Fusenig NE, Gissmann L. Molecular and cytogenetic analysis of immortalized human primary keratinocytes obtained after transfection with human papillomavirus type 16 DNA. Oncogene 1987;1:251–6.
16. Hawley-Nelson P, Vousden KH, Hubbert NL, Lowy DR, Schiller JT. HPV16 E6 and E7 proteins cooperate to immortalize human foreskin keratinocytes. EMBO J 1989;8:3905–10.
17. Werness BA, Levine AJ, Howley PM. Association of human papillomavirus types 16 and 18 E6 proteins with p53. Science 1990;248:76–9.
18. Dyson N, Howley PM, Münger K, Harlow E. The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. Science 1989;243:934–7.
19. White EA, Münger K, Howley PM. High-risk human papillomavirus E7 proteins target PTPN14 for degradation. MBio 2016;7:e01530–16.
20. Scheffner M, Huibregtse JM, Vierstra RD, Howley PM. The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. Cell 1993;75:495–505.
21. Demers GW, Espling E, Harry JB, Etscheid BG, Galloway DA. Abrogation of growth arrest signals by human papillomavirus type 16 E7 is mediated by sequences required for transformation. J Virol 1996;70:6862–9.
22. Phelps WC, Münger K, Yee CL, Barnes JA, Howley PM. Structure-function analysis of the human papillomavirus type 16 E7 oncoprotein. J Virol 1992;66:2418–27.
23. Heck DV, Yee CL, Howley PM, Münger K. Efficiency of binding the retinoblastoma protein correlates with the transforming capacity of the E7 oncoproteins of the human papillomaviruses. Proc Natl Acad Sci U S A 1992;89:4442–6.
24. Phelps WC, Bagchi S, Barnes JA, Raychaudhuri P, Kraus V, Münger K, et al. Analysis of trans activation by human papillomavirus type 16 E7 and adenovirus 12S E1A suggests a common mechanism. J Virol 1991;65:6922–30.
25. Hatterschide J, Castagnino P, Kim HW, Sperry SM, Montone KT, Basu D, et al. YAP1 activation by human papillomavirus E7 promotes basal cell identity in squamous epithelia. Elife 2022;11:e75466.

26. Hatterschide J, Bohidar AE, Grace M, Nulton TJ, Kim HW, Windle B, et al. PTPN14 degradation by high-risk human papillomavirus E7 limits keratinocyte differentiation and contributes to HPV-mediated oncogenesis. Proc Natl Acad Sci U S A 2019;116:7033–42.
27. Boukamp P, Stanbridge EJ, Foo DY, Cerutti PA, Fusenig NE. c-Ha-ras oncogene expression in immortalized human keratinocytes (HaCaT) alters growth potential in vivo but lacks correlation with malignancy. Cancer Res 1990;50:2840–7.
28. Ren Q, Kari C, Quadros MRD, Burd R, McCue P, Dicker AP, et al. Malignant transformation of immortalized HaCaT keratinocytes through deregulated nuclear factor kappaB signaling. Cancer Res 2006; 66:5209–15.
29. Edgar RC, Taylor J, Lin V, Altman T, Barbera P, Meleshko D, et al. Petabase-scale sequence alignment catalyses viral discovery. Nature 2022;602:142–7.
30. Van Doorslaer K, Li Z, Xirasagar S, Maes P, Kaminsky D, Liou D, et al. The papillomavirus episteme: a major update to the papillomavirus sequence database. Nucleic Acids Res 2017;45:D499–506.
31. Albers AE, Qian X, Kaufmann AM, Coordes A. Meta analysis: HPV and p16 pattern determines survival in patients with HNSCC and identifies potential new biologic subtype. Sci Rep 2017;7:16715.
32. Marklund L, Holzhauser S, de Flon C, Zupancic M, Landin D, Kolev A, et al. Survival of patients with oropharyngeal squamous cell carcinomas (OPSCC) in relation to TNM 8 - Risk of incorrect downstaging of HPV-mediated non-tonsillar, non-base of tongue carcinomas. Eur J Cancer 2020;139:192–200.
33. GTEx Consortium. The Genotype-Tissue Expression (GTEx) Project. Nat Genet 2013;45:580–5.
34. Falcaro M, Castañon A, Ndlela B, Checchi M, Soldan K, Lopez-Bernal J, et al. The effects of the national HPV vaccination programme in England, UK, on cervical cancer and grade 3 cervical intraepithelial neoplasia incidence: a register-based observational study. Lancet 2021;398:2084–92.
35. Olczak P, Matsui K, Wong M, Alvarez J, Lambert P, Christensen ND, et al. RG2-VLP: a vaccine designed to broadly protect against anogenital and skin human papillomaviruses causing human cancer. J Virol 2022;96:e0056622.
36. Schiller JT, Müller M. Next generation prophylactic human papillomavirus vaccines. Lancet Oncol 2015;16:e217–25.
37. Nagarsheth NB, Norberg SM, Sinkoe AL, Adhikary S, Meyer TJ, Lack JB, et al. TCR-engineered T cells targeting E7 for patients with metastatic HPV-associated epithelial cancers. Nat Med 2021;27:419–25.
38. Cheng DT, Mitchell TN, Zehir A, Shah RH, Benayed R, Syed A, et al. 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 Diagn 2015;17:251–64.
39. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 2017;14:587–9.
40. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 2018; 35:518–22.
41. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinf 2009;10:421.
42. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. Bioinformatics 2015;31: 2032–4.
43. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 2016;32:1220–2.
44. Zhou L, Qiu Q, Zhou Q, Li J, Yu M, Li K, et al. Long-read sequencing unveils high-resolution HPV integration and its oncogenic progression in cervical cancer. Nat Commun 2022;13:2563.
45. Akagi K, Li J, Broutian TR, Padilla-Nash H, Xiao W, Jiang B, et al. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. Genome Res 2014;24:185–99.
46. Hu Z, Zhu D, Wang W, Li W, Jia W, Zeng X, et al. Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. Nat Genet 2015;47:158–63.
47. Ou J, Zhu LJ. trackViewer: a Bioconductor package for interactive and integrative visualization of multi-omics data. Nat Methods 2019; 16:453–4.
48. Hill JT, Demarest BL, Bisgrove BW, Su YC, Smith M, Yost HJ. Poly Peak Parser: method and software for identification of unknown indels using sanger sequencing of polymerase chain reaction products. Dev Dyn 2014;243:1632–6.
49. Hahne F, Ivanek R. Visualizing genomic data using Gviz and Bioconductor. In: Mathé E, Davis S, editors. Statistical genomics: methods and protocols. New York: Springer; 2016. p. 335–51.
50. Beliveau BJ, Kishi JY, Nir G, Sasaki HM, Saka SK, Nguyen SC, et al. OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. Proc Natl Acad Sci U S A 2018;115:E2183–92.
51. Beliveau BJ, Boettiger AN, Nir G, Bintu B, Yin P, Zhuang X, et al. In Situ super-resolution imaging of genomic DNA with OligoSTORM and OligoDNA-PAINT. In: Erfle H, editor. Super-resolution microscopy: methods and protocols. New York, NY: Springer; 2017. p. 231–52.
52. Kurtenbach S, William Harbour J. SparK: a publication-quality NGS visualization tool. BioRxiv 845529 [Preprint]. 2019 [cited 2022 Mar 17]. Available from: https://doi.org/10.1101/845529.
53. Frattini MG, Lim HB, Laimins LA. In vitro synthesis of oncogenic human papillomaviruses requires episomal genomes for differentiation-dependent late expression. Proc Natl Acad Sci U S A 1996;93:3062–7.
54. Straub E, Dreer M, Fertey J, Iftner T, Stubenrauch F. The viral E8^E2C repressor limits productive replication of human papillomavirus 16. J Virol 2014;88:937–47.
55. Muhar M, Ebert A, Neumann T, Umkehrer C, Jude J, Wieshofer C, et al. SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. Science 2018;360:800–5.
56. Myers JE, Guidry JT, Scott ML, Zwolinska K, Raikhy G, Prasai K, et al. Detecting episomal or integrated human papillomavirus 16 DNA using an exonuclease V-qPCR-based assay. Virology 2019;537: 149–56.
57. Leesch F, Lorenzo-Orts L, Pribitzer C, Grishkovskaya I, Matzinger M, Roitinger E, et al. A molecular network of conserved factors keeps ribosomes dormant in the egg. BioRxiv 2021.11.03.467131 [Preprint]. 2021 [cited 2022 Feb 8]. Available from: https://doi.org/10.1101/2021.11.03.467131.
58. Martinez-Zapien D, Ruiz FX, Poirson J, Mitschler A, Ramirez J, Forster A, et al. Structure of the E6/E6AP/p53 complex required for HPV-mediated degradation of p53. Nature 2016;529:541–5.
59. Dickson MA, Hahn WC, Ino Y, Ronfard V, Wu JY, Weinberg RA, et al. Human keratinocytes that express hTERT and also bypass a p16(INK4a)-enforced mechanism that limits life span become immortal yet retain normal growth and differentiation characteristics. Mol Cell Biol 2000;20:1436–47.
60. White EA, Walther J, Javanbakht H, Howley PM. Genus beta human papillomavirus E6 proteins vary in their effects on the transactivation of p53 target genes. J Virol 2014;88:8201–12.
61. Leiendecker L, Jung PS, Krecioch I, Neumann T, Schleiffer A, Mechtler K, et al. LSD1 inhibition induces differentiation and cell death in Merkel cell carcinoma. EMBO Mol Med 2020;12:e12525.
62. Kovesdi I, Reichel R, Nevins JR. Identification of a cellular transcription factor involved in E1A trans-activation. Cell 1986;45:219–28.
63. Dorfer V, Pichler P, Stranzl T, Stadlmann J, Taus T, Winkler S, et al. MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. J Proteome Res 2014;13:3679–84.
64. Taus T, Köcher T, Pichler P, Paschke C, Schmidt A, Henrich C, et al. Universal and confident phosphorylation site localization using phosphoRS. J Proteome Res 2011;10:5354–62.
65. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50.

66. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 2003;34:267–73.
67. Wang C, Gu Y, Zhang K, Xie K, Zhu M, Dai N, et al. Systematic identification of genes with a cancer-testis expression pattern in 19 cancer types. Nat Commun 2016;7:10499.
68. Karlsson M, Zhang C, Méar L, Zhong W, Digre A, Katona B, et al. A single-cell type transcriptomics map of human tissues. Sci Adv 2021;7:eabh2169.
69. Karatzoglou A, Smola A, Hornik K, Zeileis A. kernlab - An S4 package for kernel methods in R. J Stat Softw 2004;11:1–20.
70. Venables WN, Ripley BD. Modern applied statistics with S. 4th ed. New York: Springer Science + Business Media; 2003.

Distinct Regulation of EZH2 and its Repressive H3K27me3 Mark in Polyomavirus-Positive and -Negative Merkel Cell Carcinoma

Marie-Alice Durand[1], Aurélie Drouin[1,21], Alice Mouchard[1,2,21], Laurine Durand[1], Clara Esnault[1], Patricia Berthon[1], Anne Tallet[3], Yannick Le Corre[4], Ewa Hainaut-Wierzbicka[5], Astrid Blom[6,7], Philippe Saiag[6,7], Nathalie Beneton[8], Guido Bens[9], Charlee Nardin[10,11], François Aubin[10,11], Monica Dinulescu[12], Christine Collin[3], Gaëlle Fromont-Hankard[13], Bernard Cribier[14], Sara Laurent-Roussel[15], Kristof Cokelaere[16], Roland Houben[17], David Schrama[17], Paul Peixoto[18,19], Eric Hervouet[18,19], Kamel Bachiri[20], Diala Kantar[20], Etienne Coyaud[20], Serge Guyétant[13,1], Mahtab Samimi[1,2,22], Antoine Touzé[1,22] and Thibault Kervarrec[1,13,22]

Merkel cell carcinoma (MCC) is an aggressive skin cancer for which Merkel cell polyomavirus integration and expression of viral oncogenes small T and Large T have been identified as major oncogenic determinants. Recently, a component of the PRC2 complex, the histone methyltransferase enhancer of zeste homolog 2 (EZH2) that induces H3K27 trimethylation as a repressive mark has been proposed as a potential therapeutic target in MCC. Because divergent results have been reported for the levels of EZH2 and trimethylation of lysine 27 on histone 3, we analyzed these factors in a large MCC cohort to identify the molecular determinants of EZH2 activity in MCC and to establish MCC cell lines' sensitivity to EZH2 inhibitors. Immunohistochemical expression of EZH2 was observed in 92% of MCC tumors (156 of 170), with higher expression levels in virus-positive than virus-negative tumors ($P = 0.026$). For the latter, we showed overexpression of EZHIP, a negative regulator of the PRC2 complex. In vitro, ectopic expression of the large T antigen in fibroblasts led to the induction of EZH2 expression, whereas the knockdown of T antigens in MCC cell lines resulted in decreased EZH2 expression. EZH2 inhibition led to selective cytotoxicity on virus-positive MCC cell lines. This study highlights the distinct mechanisms of EZH2 induction between virus-negative and -positive MCC.

Journal of Investigative Dermatology (2023) **143,** 1937–1946; doi:10.1016/j.jid.2023.02.038

INTRODUCTION

Merkel cell carcinoma (MCC) is a primary cutaneous neuroendocrine carcinoma affecting the elderly and immunocompromised people (Becker et al., 2017). In 2008, Feng et al. (2008) revealed a clonal genomic integration of the Merkel cell polyomavirus (MCPyV) in about 80% of MCC cases, and subsequently, the oncogenic properties of the MCPyV oncoproteins, that is, small T and large T (LT), were shown (Hesbacher et al., 2016; Houben et al., 2010; Verhaegen et al., 2022, 2015). The 20% of virus-negative cases are associated with UV-induced genetic alterations (Goh et al., 2016; Harms et al., 2015; Knepper et al., 2019; Starrett et al., 2020) and can arise from squamous cell carcinoma (SCC) or its precursors (Harms et al., 2022; Kervarrec et al., 2022), frequently resulting in combined tumors with SCC and MCC components. Dysregulation of similar

[1]Team "Biologie des Infections à Polyomavirus", ISP UMR 1282, INRAE, University of Tours, Tours, France; [2]Department of Dermatology, CHRU of Tours, University of Tours, Chambray-lès-Tours, France; [3]Platform of Somatic Tumor Molecular Genetics, CHU of Tours, University of Tours, Tours, France; [4]Dermatology Department, CHU of Angers, LUNAM University, Angers, France; [5]Dermatology Department, CHU of Poitiers, University of Poitiers, Poitiers, France; [6]Department of General and Oncologic Dermatology, Ambroise-Paré Hospital, CARADERM Network, Boulogne-Billancourt, France; [7]Research unit EA 4340, University of Versailles-Saint-Quentin-en-Yvelines, Paris-Saclay University, Boulogne-Billancourt, France; [8]Dermatology Department, CHU of Le Mans, University of Le Mans, Le Mans, France; [9]Dermatology department, CHR Orleans, Orleans, France; [10]Dermatology, CHU Besançon, Besançon, France; [11]INSERM 1098, Université Bourgogne Franche-Comté, Besançon, France; [12]Dermatology department, CHU Rennes, Institut Dermatologique du Grand Ouest (IDGO), Rennes, France; [13]Pathology Department, CHU of Tours, University of Tours, Tours, France; [14]Dermatology Department, CHU of Strasbourg, University of Strasbourg, Strasbourg, France; [15]Paris - La roquette, National Dermatopathology Center, Paris, France; [16]Pathology Department, Regional Hospital Jan Yperman, Ieper, Belgium; [17]Department of Dermatology, Venerology and Allergology, University Hospital of Würzburg, Würzburg, Germany; [18]INSERM, EFS-BFC, UMR 1098 RIGHT, University Bourgogne-Franche-Comté, Besançon, France; [19]EPIgenetics and GENe Expression Technical Platform (EPIGENExp), University Bourgogne Franche-Comté, Besançon, France; and [20]Department of Biology, Inserm U1192, Protéomique Réponse Inflammatoire Spectrométrie de Masse-PRISM, CHU Lille, Université de Lille, Lille, France

[21]These authors contributed equally to this work.

[22]These authors contributed equally to this work.

Correspondence: Thibault Kervarrec, Department of Pathology, Hôpital Trousseau, CHRU de Tours, Tours 37044, France. E-mail: thibaultkervarrec@yahoo.fr

Abbreviations: DZNeP, 3-deazaneplanocin A; EZH2, enhancer of zeste homolog 2; FCS, fetal calf serum; H3K27me3, trimethylation of lysine 27 on histone 3; LT, large T; MCC, Merkel cell carcinoma; MCPyV, Merkel cell polyomavirus; SCC, squamous cell carcinoma; TA, T antigen

Received 28 June 2022; revised 15 February 2023; accepted 17 February 2023; accepted manuscript published online 8 April 2023; corrected proof published online 5 May 2023

© 2023 The Authors. Published by Elsevier, Inc. on behalf of the Society for Investigative Dermatology.

molecular pathways, notably p53 and pRb inactivation (Harms et al., 2015; Hesbacher et al., 2016), either through the function of MCPyV oncoproteins or somatic mutations is expected to account for the phenotypic similarities between both groups (Harms et al., 2018).

The involvement of epigenetic modulation has been recently proposed in MCC development (Gujar et al., 2021; Leiendecker et al., 2020; Park et al., 2020; Rotondo et al., 2021). Notably, expression of enhancer of zeste homolog 2 (EZH2), a histone methyltransferase belonging to the PRC2, leads to trimethylation of lysine 27 of histone H3 (H3K27me3) (Cao and Zhang, 2004). In tumors, EZH2 acts either as an activator or repressor of tumor suppressor genes (Gan et al., 2018). Importantly, the EZH2 allosteric inhibitor tazemetostat was recently approved by the United States Food and Drug Administration for the treatment of epithelioid sarcoma and follicular lymphoma, two neoplasms showing an oncogenic dependency on EZH2 (Gounder et al., 2020; Italiano et al., 2018; Morschhauser et al., 2020).

Gartin et al. (2022) recently showed the sensitivity of MCC cell lines and xenotransplant tumors to EZH2 inhibition through the derepression of the transcription factor SIX1. However, controversial data have been reported regarding EZH2 function and H3K27me3 levels in MCC tumors. Harms et al. (2017) reported high EZH2 expression levels in 85 MCC samples, such a feature being associated with tumor aggressiveness and reduced survival. However, low H3K27me3 levels were observed in a cohort of 40 MCCs, this reduction is associated with virus-positive status (Busam et al., 2017), whereas the opposite result was observed in another study (n = 41) (Matsushita et al., 2019). Furthermore, Gujar et al. (2021) showed global loss of H3K27me3 in both virus-positive and -negative MCC cases through the overexpression of EZHIP, a recently identified protein inhibiting the PRC2 complex. Noteworthy, PRC2 inhibition resulting in reduced H3K27me3 levels has been identified as a contributor to virus-negative tumor development (Harms et al., 2022; Tsai, 2021).

In this study, immunohistochemistry revealed EZH2 expression and high H3K27me3 levels in most MCC tumors. Therefore, we subsequently aimed (i) to identify the molecular determinants of EZH2 expression and H3K27me3 regulation in MCC and (ii) to evaluate whether EZH2 and its repressive mark contribute to MCC biology and therefore whether it might constitute a potential therapeutic target.

RESULTS

High expression of EZH2 is more frequent in MCC than in other skin cancers

To assess the potential contribution of EZH2 and H3K27me3 in MCC biology, we first aimed to evaluate EZH2 expression and H3K27me3 levels. To this end, we performed immunohistochemistry and used a semiquantitative score to analyze a previously described cohort of MCC tumor samples (Thanguturi et al., 2022) (Figure 1a).

Overall, 170 patients with MCC were included in this study, including 91 women (53%). The median age at diagnosis was 80.2 years (range = 49.3–96.7). American Joint Committee on Cancer stage at diagnosis was I (n = 41, 28%), II (n = 52, 35%), III (n = 50, 34%), or IV (n = 5, 3%) (missing data, n = 22). Immunosuppression (HIV positivity, organ transplant recipients, or hematological malignancy) was present in 26 cases (19%). Primary tumor location was the limbs (n = 74, 47%), head and neck (n = 59, 37%), trunk (n = 7, 4%), or occult (n = 18, 12%) with 12 inconclusive cases. Using a previously described multimodal approach (Moshiri et al., 2017), 111 tumors were classified as virus positive (74%), whereas 39 were virus negative (26%) (inconclusive cases, n = 20).

EZH2 expression was observed in 156 MCC cases (91.8% of included cases), with low/heterogeneous (score 1) and intense and diffuse (score 2) expression detected in 16.5% (n = 28) and 75.3% (n = 128) of the cases, respectively (Table 1 and Figure 1a). High levels of H3K27me3 (score 2) were detected in the majority of MCC samples (n = 121, 72.9%), whereas low levels (score 1) were observed in 25 (15.1%), and the absence of staining was observed in 20 (12%) samples (Table 1 and Figure 1a and b). Moreover, a positive association was observed between EZH2 expression and H3K27me3 levels, with H3K27me3 scores of 0, 1, and 2 observed in 8, 5, and 0 of the EZH2 nonexpressing cases (score 0) and 12, 20, and 122 of the EZH2-expressing cases (scores 1 and 2), respectively ($P < 0.0001$) (Figure 1b).

In other skin tumors (84 SCC, 48 basal cell carcinomas, and 100 other cutaneous benign or malignant epithelial tumors) (Table 1 and Supplementary Table S1 and Figure 1c), strong expression of EZH2 was observed in 42% of cases (n = 71), without an association with H3K27me3 levels, which remained detectable in almost all tumors (n = 162 of 169, 96%) (Supplementary Table S1). Interestingly, among non-MCC tumors, EZH2 high-expressing cases were mostly basal cell carcinoma (n = 29 of 71, 41%) and SCC (n = 32 of 71, 45%), whereas benign adnexal tumors had weak or absent EZH2 expression. These results might suggest that in skin cancers as a whole and in MCC especially, EZH2 may play a role in the development and/or progression of malignant tumors.

Absence of H3K27me3-repressive mark is restricted to virus-negative tumors

To identify the potential factors affecting EZH2 expression and activity as indicated by the H3K27me3 mark, the characteristics of EZH2-negative (score 0) and EZH2-positive (scores 1 and 2) MCC cases as well as H3K37me3-negative and H3K27me3-positive cases were compared. These analyses showed no association between clinical data and EZH2 expression levels (Supplementary Table S2) or H3K27me3 expression (Supplementary Table S3). However, lack of EZH2 expression was more frequent in virus-negative (n = 7 of 39, 19%) than in virus-positive (n = 5 of 111, 4.5%) tumors ($P = 0.0078$) (Supplementary Table S4 and Figure 1d). Accordingly, all of the 18 cases with a complete lack of H3K27me3 were virus-negative cases (Supplementary Table S4 and Figure 1d). Among these cases, seven (39%) showed a loss of EZH2 expression (Supplementary Table S4). Interestingly, among virus-negative cases, H3K27me3 loss was observed in a majority of combined cases harboring an SCC and an MCC component (n = 7 of 8, 87.5%), supporting the hypothesis that loss of the H3K27me3-repressive hallmark might contribute to SCC-to-MCC transformation (Harms et al.,

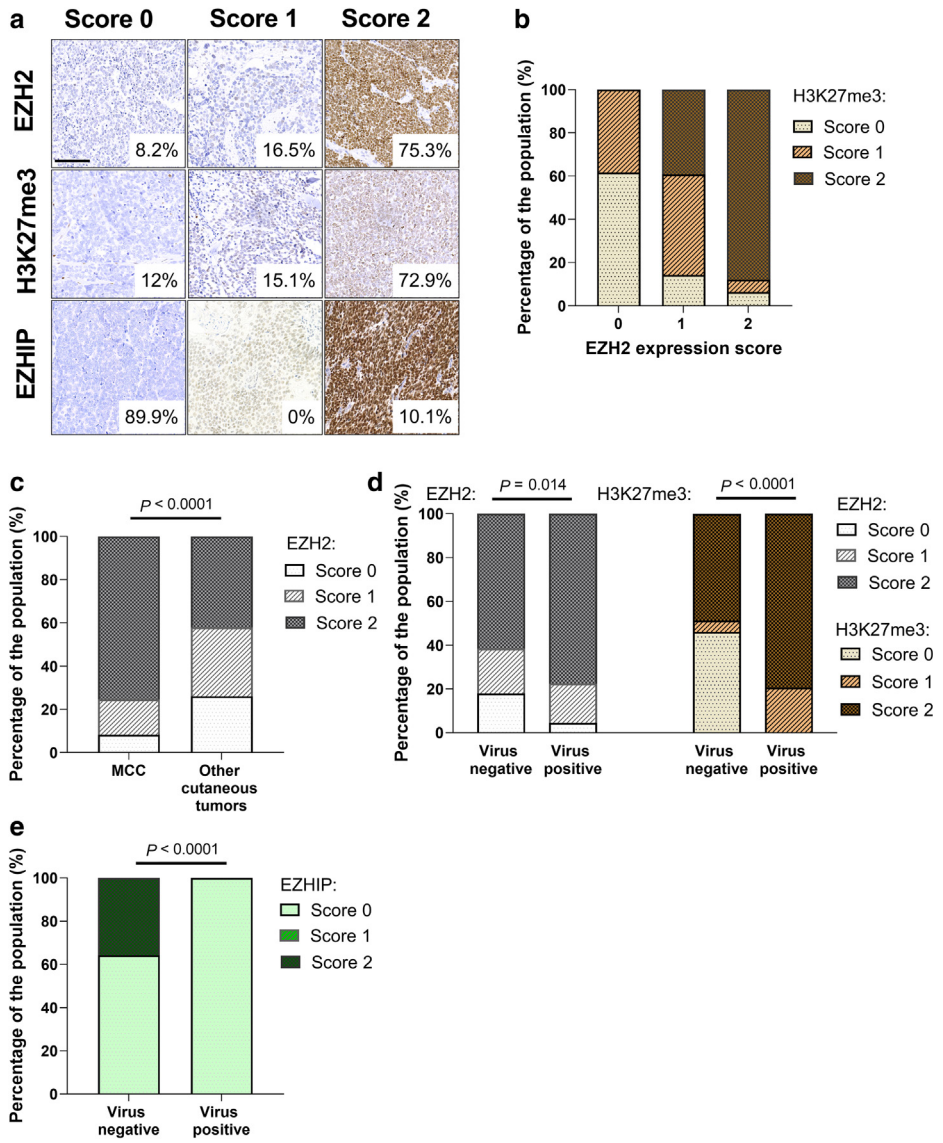


Figure 1. Expression of EZH2, EZHIP, and H3K27me3 in MCC tumors and other skin cancers. (**a**) Immunohistological staining of EZH2, H3K27me3, and EZHIP in MCC tumor samples. Representative MCC cases with no (score 0), low (score 1), and high (score 2) staining are shown (n = 170). (**b**) H3K27me3 expression score according to EZH2 score in the MCC cohort. (**c**) Comparison of EZH2 expression score among a cohort of MCC and a cohort of other cutaneous tumors. (**d**) EZH2 expression score, H3K27me3 score, and (**e**) EZHIP expression score according to MCPyV status in the MCC cohort. EZH2, enhancer of zeste homolog 2; H3K27me3, trimethylation of lysine 27 on histone 3; MCC, Merkel cell carcinoma; MCPyV, Merkel cell polyomavirus.

2022). Because only a minority of these cases have non-detectable EZH2 expression, this data further suggests that other factors contribute to EZH2 activity regulation in this subset.

Overexpression of EZHIP is restricted to virus-negative cases and is frequently observed in combined tumors

Besides EZH2, the H3K27me3-repressive mark is regulated by several proteins such as the SWI/SNF complex (Kim et al., 2015) and EZHIP (Gujar et al., 2021). Therefore, to identify the molecular mechanisms leading to H3K27me3 repression in a small subset of virus-negative MCC, we assessed the expression levels of SMARCB1 and SMARCA4, two subunits of the SWI/SNF complex, as well as of EZHIP in our cohort. SMARCB1 and SMARCA4 were highly expressed in all interpretable cases (n = 160 and 161 interpretable cases, respectively). By contrast, diffuse positivity for EZHIP was restricted to 14 virus-negative MCC cases (and two cases with undetermined viral status), including five of the eight combined tumors and all displaying a complete lack of H3K27me3 (Figure 1a and e and Table 1 and Supplementary Figure S1).

Notably, among the 18 H3K27me3-negative cases identified, 10 of the 11 discordant cases with EZH2 expression but the absence of H3K27me3 were positive for EZHIP. These results suggest that two distinct mechanisms regulate H3K27me3 mark loss in a subset of virus-negative MCC: inhibition of EZH2 expression and inhibition of EZH2 activity by EZHIP expression.

To further support these findings, we evaluated EZH2, H3K27me3, and EZHIP levels on complete slides on six additional combined MCC cases consisting of an SCC and an MCC part. For these cases, a lack of H3K27me3 and diffuse EZHIP positivity were observed in four cases (Figure 2 and Supplementary Table S5), with EZHIP expression being either restricted to the MCC component or throughout the combined tumor. Furthermore, analysis of EZHIP in non-MCC skin tumors revealed overexpression of EZHIP in seven cases, all being SCC and five of them showing a lack of H3K27me3 (Table 1 and Supplementary Table S1).

Table 1. Expression Score of EZH2, H3K27me3, and EZHIP among Skin Tumors

Immunohistochemical Features	Merkel Cell Carcinoma (n = 170)	Other Skin Tumors (n = 232)	*P*-Value
EZH2 expression score, n (%)			<0.0001
Score 0 (no expression)	14 (8.2)	44 (26.0)	
Score 1 (low expression)	28 (16.5)	54 (32.0)	
Score 2 (medium expression)	128 (75.3)	71 (42.0)	
Data not available	0	63	
H3K27me3 score, n (%)			0.0002
Score 0 (no expression)	20 (12.0)	7 (4.1)	
Score 1 (low expression)	25 (15.1)	12 (7.1)	
Score 2 (high expression)	121 (72.9)	150 (88.8)	
Data not available	4	63	
EZHIP expression score, n (%)			0.0013
Score 0 (no expression)	141 (89.9)	167 (91.8)	
Score 1 (low expression)	0 (0)	8 (4.4)	
Score 2 (high expression)	16 (10.1)	7 (3.8)	
Data not available	13	50	

Abbreviations: EZH2, enhancer of zeste homolog 2; H3K27me3, trimethylation of lysine 27 on histone 3.

Therefore, our data showed that EZHIP expression is observed in a small subset of SCC and in virus-negative cases, the latter frequently showing divergent differentiation and H3K27me3 loss. By contrast, EZH2 expression was observed in most MCC cases, with more frequent and higher expression levels in virus-positive cases suggesting a potential involvement of MCPyV proteins in EZH2 induction.

EZH2 expression in MCC is induced by viral oncoproteins

Because observation of MCC tumors might suggest a contribution of MCPyV to EZH2 expression, we next sought to assess whether T antigens (TAs) affect EZH2 expression levels in vitro. Thus, EZH2 expression and H3K27me3 levels were first assessed by western blot in three virus-positive MCC cell lines (WaGa, PeTa, and MKL-1), two virus-negative MCC cell lines (MCC13 and MCC26), and controls (primary fibroblasts, primary keratinocytes, and HaCaT). This analysis revealed intense EZH2 expression and high H3K27me3 levels in all the three virus-positive cell lines, whereas such features were absent or weak in the others cell lines (Figure 3a and b). Inducible knockdown of TA expression in virus-positive MCC cell lines WaGa and PeTa by short hairpin RNA interference resulted in reduced EZH2 expression at both mRNA and protein levels ($P = 0.007$) (Figure 3c–e), confirming an impact of TA on EZH2 expression. To further support these results and determine the impact of the individual TAs, small T, truncated LT, and TA were ectopically expressed in primary fibroblasts using lentiviral vectors. In addition, because pRb inactivation is considered the main oncogenic property of LT in MCC cells (Houben et al., 2012) and because EZH2 is a well-established downstream effector of the pRb pathway (Bracken et al., 2003), we also included a truncated LT mutant with a nonfunctional pRb-binding site (LT^{E216K}). After transduction, induction of EZH2 expression was observed in cells expressing TA and the functional LT but not for the other constructs (Figure 3f–h), confirming the role of LT and specifically its pRb-binding site in the induction of EZH2 expression.

Therefore, our results strongly suggest that EZH2 overexpression induced by LT constitutes a relevant therapeutic target in virus-positive MCC tumors.

EZH2 is a therapeutic target in virus-positive cell lines

To confirm the potential contribution of EZH2 to the growth of virus-positive MCC, we next evaluated the effect of EZH2 inhibition on MCC cell lines in vitro using two competitive chemical inhibitors (tazemetostat and 3-deazaneplanocin A [DZNeP]). After 2 days of drug exposure, flow cytometry on WaGa cell lines (Figure 4a) showed the reduction of H3K27me3 levels among the alive cells upon both tazemetostat ($P = 0.0187$) or DZNeP ($P = 0.0187$) treatment, confirming the on-target effect of both inhibitors. Dose–response curves were then determined by assessing cell viability after 4 days for DZNeP and 14 days of drug exposure for tazemetostat. As previously described (Gartin et al., 2022), the EZH2 inhibition reduced the growth of the three virus-positive MCC cell lines, whereas cell growth was only affected at high concentration levels in controls, suggesting that these chemical inhibitors impact the viability of virus-positive MCC cell lines predominantly through EZH2 inhibition (Figure 4b and c). Furthermore, the observed half-maximal inhibitory concentration values were lower for these virus-positive cell lines than for G-401 (Kurmasheva et al., 2017) and SW-1353 (Lhuissier et al., 2021), two non-MCC cell lines previously identified as being sensitive to tazemetostat and DZNeP, respectively (Supplementary Figure S2).

To further determine whether the effect of EZH2 inhibition on virus-positive MCC cell lines was related to proliferation arrest or cell death, cell cycle analysis was performed on WaGa and PeTa cells treated with tazemetostat for 7 days or DZNeP for 4 days (Figure 4d and e and Supplementary Figure S3). This analysis revealed an increase in the sub-G1 population ($P < 0.05$) upon treatment with the two inhibitors and a decrease in the G1 phase upon treatment with DZNeP, suggesting that EZH2 inhibition results in cell death rather than proliferation inhibition in MCC cell lines. Finally, because Gartin et al. (2022) recently reported that EZH2

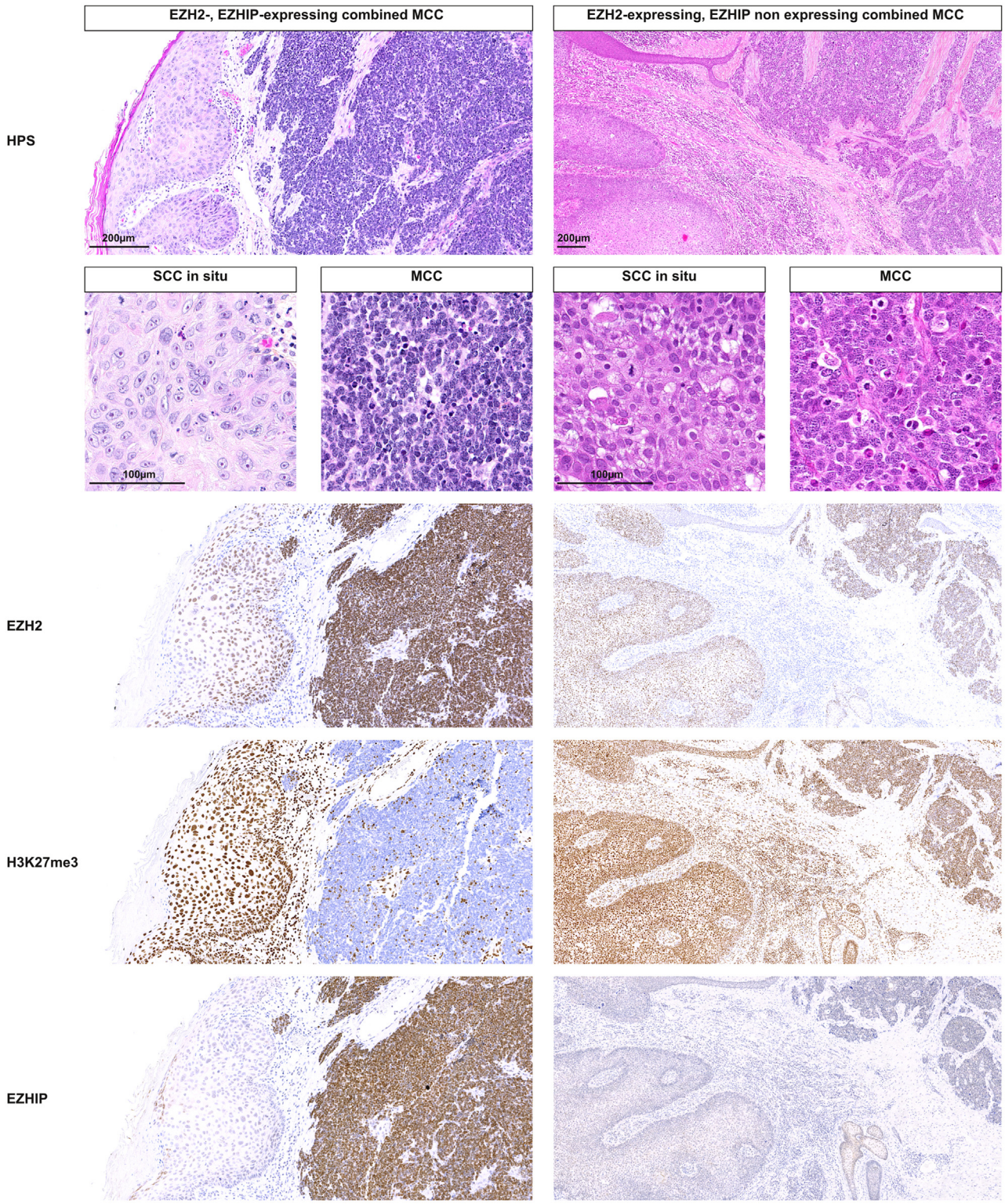


Figure 2. Expression of EZH2, its target H3K27me3, and EZHIP in combined tumors. A combined MCC/SCC tumor is depicted (case 1 and case 2, bar = 250 µm) with HPS, EZH2, H3K27me3, and EZHIP immunohistological staining. EZH2, enhancer of zeste homolog 2; H3K27me3, trimethylation of lysine 27 on histone 3; HPS, hematein, phloxin, saffran; MCC, Merkel cell carcinoma; SCC, squamous cell carcinoma.

inhibition reduces MCC cell survival through the derepression of SIX1 transcription, we then assessed SIX1 expression in our MCC cohort (n = 38) by immunohistochemistry and detected expression of SIX1 with low/heterogenous and high expression levels in 21 and 17 cases, respectively. Notably, increased SIX1 expression was observed in MCC tumors with H3K27me3 loss ($P < 0.0001$) (Supplementary Figure S4a and b). Furthermore, induction of SIX1 expression was observed in

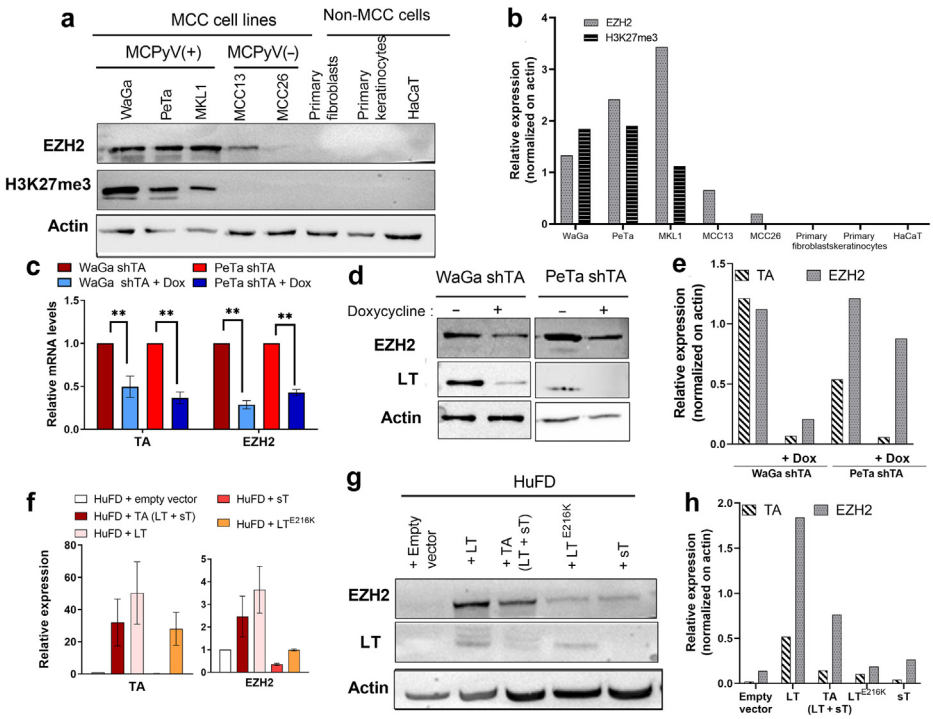


Figure 3. Expression of EZH2 in MCC is enhanced by viral oncoproteins. (**a**) Assessment of EZH2 expression in five MCC cell lines (WaGa, PeTa, MKL1, MCC13, and MCC26) and three other skin cell types (primary fibroblasts, primary keratinocytes, and HaCaT) and (**b**) quantification of EZH2 expression and H3K27me3 using western blot. Protein levels were quantified using ImageJ, and actin served as loading control. (**c**) Effect of doxycycline-inducible shRNA targeting TAs on *EZH2* mRNA expression in WaGa and PeTa evaluated by qPCR. Shown are relative expression to controls of *EZH2* and *TA* as means ± SEM, normalized to *RPLP0* expression. n = 5. Mann–Withney test was performed ($P = 0{,}007$; $^{***}P < 0.01$). (**d**) Effect of doxycycline-inducible shRNA targeting TAs on EZH2 protein expression in WaGa and PeTa assessed by western blot and (**e**) quantification of TA and EZH2 protein levels (n = 1). (**f**) Effect of ectopic expression of various constructs of TAs in HuFDs assessed by qPCR and (**g**) western blot and (**h**) quantification of proteins levels (n = 1). EZH2, enhancer of zeste homolog 2; H3K27me3, trimethylation of lysine 27 on histone 3; HuFD, human dermal fibroblast; LT, large T; MCC, Merkel cell carcinoma; shRNA, short hairpin RNA; shTA, T antigen–short hairpin RNA; sT, small T; TA, T antigen.

MCC cell lines after tazemetostat treatment (Supplementary Figure S4c). Together, these results suggest that pharmacological inhibition of EZH2 affects the survival of virus-positive MCC cell lines through the inhibition of H3K27me3 and SIX1 induction and might constitute a potential therapeutic approach in MCC.

DISCUSSION

EZH2 is the functional enzymatic component of the PRC2 complex, whose main target is lysine 27 of histone 3 leading to trimethylation as a repressive mark (Cao and Zhang, 2004; Gan et al., 2018). EZH2 is notably active in cell progenitors and stem cells repressing the transcription of genes involved in cell cycle arrest and differentiation (Margueron and Reinberg, 2011). Under pathological conditions, overexpression of EZH2 has been shown in many solid or hematological malignancies and is frequently associated with proliferation, progression, metastasis, drug resistance, and poor outcome (Jiang et al., 2016). The main EZH2 targets include *CDKN2A*, which encodes $p14^{ARF}$, $p16^{INK4A}$ (Sasaki et al., 2008; Zhong et al., 2013), *RUNX3* (Fujii et al., 2008), and *CDKN1A* (Xu et al., 2019).

Our results confirmed that EZH2 is expressed in 92% of MCC tumors, frequently at moderate and high levels (75% of the cases). Moreover, high EZH2 expression was associated with the repressive H3K27me3 mark in MCC tumors. Interestingly, a recent work underlines the therapeutic vulnerability of MCC to EZH2 inhibition using tazemetostat (Gartin et al., 2022). In this study, using EZH2 chemical inhibitors on MCC cell lines, we confirm that EZH2 inhibition might constitute a potential therapeutic option for virus-positive MCC as recently proposed for decitabine—another treatment targeting epigenetic modification—with and showing a selective effect restricted to virus-positive MCC (Harms et al., 2022).

EZH2 activity is counterbalanced by EZHIP and by the chromatin remodeling complex SWI/SNF (switch/non-fermentable sugar), two antagonists of the PRC2 complex (Italiano, 2016). Of note, immunohistochemical analysis of our cohort of MCC tumors did not show a loss of expression of SMARCB1 or SMACA4 subunits in line with a previous report (Gandhi et al., 2021). In addition, although several large sequencing studies on MCC have been published (Goh et al., 2016; Harms et al., 2015; Knepper et al., 2019; Starrett et al., 2020), an activating mutation in the *EZH2* gene was only detected in one MCC case (Harms et al., 2016). By contrast, we showed that pRb inactivation by the LT antigen results in EZH2 overexpression in virus-positive MCC cases. Indeed, among the pathways controlling EZH2 expression (Walker et al., 2020), one of the most important is the pRb/E2F axis. E2F binds to the *EZH2* promoter to activate transcription, this activity being counterbalanced by the sequestration of E2F by pRb (Bracken et al., 2003). Accordingly, overexpression of EZH2 both at transcriptional and at translational levels has been previously described in MCC, with

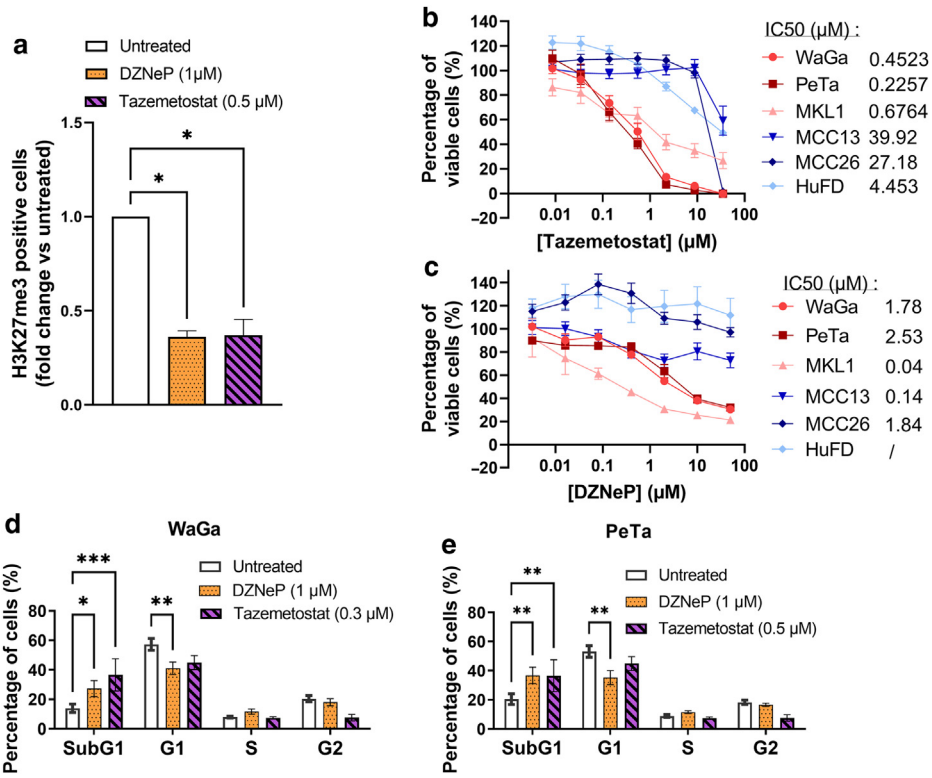


Figure 4. EZH2 promotes cell survival of MCC cells in vitro. (**a**) Flow cytometry to determine the percentage of H3K27me3-positive cells among viable WaGa cells after 2 days of treatment with the respective drug. n = 6. (**b**) Cytotoxicity assay of tazemetostat treatment during 14 days and (**c**) DZNeP treatment during 4 days on different cell lines: MCC (WaGa, PeTa, MKL1, MCC13) and primary fibroblasts (HuFD). Results are presented as a percentage of viable cells compared with that of untreated cells. n = 3. (**d**) Cell cycle analysis of the two EZH2 inhibitors on WaGa and (**e**) PeTa cell lines for 7 days for tazemetostat and 4 days for DZNeP. n = 5; mean ± SEM; two-way ANOVA tests, Friedman ANOVA test; $*P < 0.05$ and $**P < 0.01$. DZNeP, 3-deazaneplanocin A; EZH2, enhancer of zeste homolog 2; H3K27me3, trimethylation of lysine 27 on histone 3; HuFD, human dermal fibroblast; IC50, half-maximal inhibitory concentration; MCC, Merkel cell carcinoma.

high EZH2 expression levels more frequently observed in metastasis than in primary tumors (Harms et al., 2017; Veija et al., 2017). Furthermore, although Busam et al. (2017) observed reduced H3K27me3 levels in virus-positive MCC compared with the levels in virus-negative cases, the opposite results were reported by Matsushita et al. (2019) and further confirmed by the recent investigation of combined MCC/SCC cases (Harms et al., 2022). Notably, in this study, Harms et al. (2022) suggested that loss of H3K27me3 might contribute in combination with pRb inactivation to the transformation of SCC to MCC. Accordingly, in this work, the complete loss of the H3K27me3 mark was observed in a subset of virus-negative cases notably in combined cases (n = 7 of 8). Interestingly, in this setting, induction of SOX2 expression is observed in the MCC component (Kervarrec et al., 2020a). Moreover, previous studies showed that the inactivation of PRC2 in epidermal stem cells leads to SOX2 expression and Merkel cell development (Bardot et al., 2013; Perdigoto et al., 2016), an observation that can similarly be made in EZH2 inhibitor–treated MCC cells (Supplementary Figure S5). In conclusion, these results suggest that PRC2 inactivation could contribute to the neuroendocrine differentiation in virus-negative combined MCC.

Inhibition of the PRC2 complex has recently been identified as a genetic hallmark of some CNS tumors such as diffuse midline gliomas and posterior fossa type A ependymomas. Indeed, about 80% of the diffuse midline glioma harbored a mutation in the histone 3 gene, resulting in the loss of H3K27me3 (Wu et al., 2012). Moreover, overexpression of EZHIP, a negative regulator of the PRC2 complex with physiologic expression restricted to the oocyst, placenta, and testis, has been reported in posterior fossa type A ependymomas (Pajtler et al., 2018). In this study, we showed that H3K27me3 loss is frequently observed in virus-negative MCC cases and in a subset of SCC, these findings either being due to EZHIP overexpression or due to inhibition of EZH2 expression.

To conclude, we have provided evidence that EZH2 is overexpressed in MCC and that chemical inhibition of EZH2 using two chemical inhibitors was cytotoxic in MCC virus-positive cell lines, suggesting that EZH2 inhibition may be a therapeutic target for patients with metastatic virus-positive MCC. Moreover, this study shows EZH2 overexpression in virus-positive MCC and that MCPyV LT antigen impacts EZH2 expression in virus-positive MCC through pRb sequestration. By contrast, although we found indications of preserved EZH2 activity in about half of virus-negative MCC, EZH2 action appears to be repressed in others either through the downregulation of EZH2 expression or by EZHIP expression.

MATERIALS AND METHODS

Patients and samples

Tumor samples were selected from a prospective cohort of patients with MCC from 10 French hospital centers, and tumors were collected between March 1998 and October 2020 (Ethics Committee in Human Research of CHRU of Tours, Tours, France; identification number RCB2009-A01056-51). Patients provided written informed consent. Inclusion criteria for the cohort have been described previously (Gardair et al., 2015; Kervarrec et al., 2018; Thanguturi et al., 2022). Among the cohort, all cases with available formalin-fixed paraffin-embedded tumors were included in a tissue microarray as previously described (Kervarrec et al., 2019). Only cases with interpretable EZH2 immunohistochemical staining were included in this study. MCPyV status was determined using a multimodal approach, as previously reported (Moshiri et al., 2017).

Briefly, MCPyV DNA copies per cell were determined by real-time PCR (positive cut-off: 1.2 viral copy number per cell), and expression of TAs was determined by two immunohistochemical methods using the AB3 and CM2B4. MCC cases were classified as virus positive if at least two of these three tests were positive (Higaki-Mori et al., 2012; Moshiri et al., 2017). Age, sex, American Joint Committee on Cancer stage on diagnosis, immunodepression (organ transplant recipient, hematological malignancy, HIV, immunosuppressive drugs), and primary tumor location were collected from patient records.

Immunohistochemistry

Immunohistochemical staining for EZH2, H3K27me3, SOX2, SIX1, and EZHIP was performed manually on tissue microarray and/or combined tumor slides. Immunohistochemical staining of SMARCA4 and SMARCB1 on tissue microarray was performed using a BenchMark XT Platform as instructed. Antibodies and dilutions are provided in Supplementary Table S6. Microscopic evaluation was performed by a pathologist (TK) blinded to clinical data. EZH2, H3K27me3, SIX1, and EZHIP stainings were graded using a semiquantitative score: 0, no staining; 1, low or heterogeneous expression; and 2, moderate or strong and diffuse expression.

Cell lines and EZH2 inhibitors

The following MCC cell lines were used: WaGa (RRID: CVCL_E998), PeTa (RRID: CVCL_LC73), MKL-1 (RRID: CVCL_2600) (Rosen et al., 1987), and MCC13 (RRID: CVCL_2583) (Leonard et al., 1995). MCC cell lines were cultured in RPMI 1640 containing 10% fetal calf serum (FCS), 100 U/ml penicillin and 0.1 mg/ml streptomycin, and amphotericin B. HaCaT (RRID: CVCL_0038), human embryonic kidney 293T (RRID: CVCL_0063), primary keratinocytes, normal human dermal fibroblasts, (Promocell, Heidelberg, Germany), and SW1353 (provided by Catherine Baugé) (Lhuissier et al., 2021) were grown in DMEM supplemented with 10% FCS, 100 U/ml penicillin and 0.1 mg/ml streptomycin, and amphotericin B. G401 cells (CRL1442, ATCC, Manassas, VA) were grown in McCoy's 5A modified medium plus 10% FCS and 100 U/ml penicillin and 0.1 mg/ml streptomycin. All cells were grown at 37 °C and 5% carbon dioxide in a humidified incubator. EZH2 chemical inhibition was performed on MCC cell lines using two different inhibitors: tazemetostat (EPZ 6438, Clinisciences, Leipzig, Germany) at indicated concentration during 14 days for cell viability assay and during 7 days for other experiments and DZNeP (SelleckChemical, Houston, TX) at the indicated doses for 4 days. Control cells were treated with 2% of DMSO.

Plasmids and lentiviral transduction

Plasmid encoding TAs (*Pcdh-TA*), LT antigen (*Pcdh-LT*), small T antigen (*pcdh-sT*) (Houben et al., 2015), short hairpin RNA targeting TAs (induce *shTA),* and LT mutated at the pRb binding site (*pih-LTA*^{E216K}) (Houben et al., 2012) have been described previously. Lentiviral infections were performed as previously described (Houben et al., 2012).

mRNA expression assay

For RT-qPCR, total cellular RNA was isolated using the NucleoSpin RNA Plus kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions, followed by cDNA synthesis using the RT^2 first strand kit (Qiagen Benelux, Venlo, Netherlands). Total MCPyV *LT, EZH2*, or *RPLPO* transcripts as endogenous control were amplified using the primers listed in Supplementary Table S7. Quantitative real-time PCR reagents were Rox SYBR MasterMix blue dTTP and Takyon from Eurogentec (Seraing, Belgium), and real-time PCR was performed using the Step One Real-time PCR system (Applied Biosystem, Waltham, MA). Quantification was performed in five independent experiments.

Immunoblot

Immunoblotting was performed as previously described (Kervarrec et al., 2020b). Antibodies used in this study are described in Supplementary Table S8.

Viability assay

To assess cell viability after chemical inhibition of EZH2 by tazemetostat or DZNeP, XTT assays were performed. Cells were plated in six replicates in a 96-well plate at 25,000 cells/well for tazemetostat assay and at 50,000 cells/well for DZNeP assays. Increasing concentrations of inhibitors were incubated on the cells for 14 or 4 days, respectively (Chan-Penebre et al., 2017). Untreated cells were used as a viability reference, and cells treated with 1% Triton X-100 were used as a control for maximum cell death. After drug incubation time, 25 μl of XTT reagent at 1 mg/ml (Alfa Aesar, Thermo Fisher Scientific, Waltham, MA) with N-methyl-dibenzopyrazine-methyl-sulfate activator (25 μM) was added to each well. After 4 hours, absorbance was measured at 450 nm (Multiskan EX, Thermo Fisher Scientific). The absorbance at 620 nm was used as a reference. The experiment was performed in triplicate for each cell line.

Flow cytometry

For cell cycle analysis, cells were fixed in ice-cold 70% ethanol, treated with propidium iodide mixture (PBS + 1% FCS + 0.1 mg/ml propidium iodide + 0.1 mg/ml RNAse A) at 37 °C for 1 hour and analyzed by flow cytometry. All flow cytometry data were analyzed by Kaluza software (Beckman Coulter, Brea, CA).

For H3K27me3 staining of viable cells, WaGa dead cells were first labeled with Zombie Green Fixable Viability Kit (BioLegend, San Diego, CA) and washed as stated by the manufacturer's instructions. Cells were fixed with formaldehyde and permeabilized by 1% Triton-X-100 and washed with PBS + 1% FCS. Primary antibodies for H3K27me3 staining and secondary antibody were used at the indicated concentration and incubation time in Supplementary Table S9.

Statistical analysis

Continuous data are described with means and first and third quartiles (quartile 1−quartile 3). Categorical data are described with a number (percentage) of interpretable cases. Associations were assessed by the two-tailed Fisher's exact test for categorical data and Mann−Whitney or ANOVA tests for continuous data. $P < 0.05$ was considered statistically significant. Half-maximal inhibitory concentration values were determined using Prism 9 (GraphPad Software, San Diego, USA).

Data availability statement

No dataset was generated or analyzed during this study.

ORCIDs
Marie-Alice Durand: http://orcid.org/0000-0002-8159-7296
Aurélie Drouin: http://orcid.org/0000-0003-1004-9513
Alice Mouchard: http://orcid.org/0000-0002-4143-0766
Laurine Durand: http://orcid.org/0000-0002-7824-9952
Clara Esnault: http://orcid.org/0000-0003-4216-5711
Patricia Berthon: http://orcid.org/0000-0002-3491-7613
Anne Tallet: http://orcid.org/0000-0002-2601-2443
Yannick Le Corre: http://orcid.org/0000-0003-3089-3068
Ewa Hainault-Wierzbicka : http://orcid.org/0000-0002-6475-4018
Astrid Blom: http://orcid.org/0000-0002-2848-3724

Philippe Saiag: http://orcid.org/0000-0002-6500-3507
Nathalie Benethon: http://orcid.org/0000-0002-3052-2366
Guido Bens: http://orcid.org/0000-0002-7697-4392
Charlee Nardin: http://orcid.org/0000-0002-9902-5990
François Aubin: http://orcid.org/0000-0002-1421-4996
Monica Dinulescu : http://orcid.org/0000-0002-3950-4340
Christine Collin : http://orcid.org/0000-0002-7930-1855
Gaëlle Fromont: http://orcid.org/0000-0002-7730-2264
Serge Guyétant: http://orcid.org/0000-0003-2783-9722
Bernard Cribier: http://orcid.org/0000-0002-6306-770X
Sara Laurent-Roussel: http://orcid.org/0000-0003-0558-6680
Kristof Cokelaere: http://orcid.org/0000-0001-7608-9193
Roland Houben: http://orcid.org/0000-0003-4538-2324
David Schrama: http://orcid.org/0000-0002-6931-8194
Peixoto Paul: http://orcid.org/0000-0001-6302-7823
Hervouet Eric: http://orcid.org/0000-0002-4841-7812
Kamel Bachiri: http://orcid.org/0000-0003-0297-0521
Diala Kantar: http://orcid.org/0000-0001-9723-7204
Etienne Coyaud: http://orcid.org/0000-0002-5893-4557
Mahtab Samimi: http://orcid.org/0000-0001-6742-9088
Antoine Touzé: http://orcid.org/0000-0002-9856-9945
Thibault Kervarrec: http://orcid.org/0000-0002-2201-6914

CONFLICT OF INTEREST

EHW reports Bristol-Myers Squibb, Novartis, SUNpharma, Sanofi, and AbbVie. MS has received fees from Amgen and Novartis for speaking at an educational meeting; has received fees from Merck-Pfizer for advisory board; has received reimbursement for travel and accommodation expenses from AbbVie, Amgen SAS, Bristol-Myers Squibb, Celgene SAS, Galderma International, Lilly France SAS, Merck Sharp & Dohme France, and Pierre Fabre Médicaments France. The remaining authors state no conflict of interest.

ACKNOWLEDGMENTS

The authors thank Catherine Baugé for kindly providing the SW1353 cells, Comité inter-régional Grand Ouest de la ligue contre le cancer, and Société française de dermatologie. The local Ethics Committee in Human Research of Tours (Tours, France) approved the study (identification number RCB2009-A01056-51).

AUTHOR CONTRIBUTIONS

Conceptualization: DS, RH, AT, MS, TK; Formal Analysis: MAD, MS, TK; Funding Acquisition: TK, MS, AT; Investigation: MAD, AD, AM, CE, LD, PB, KB, DK, EC; Resources: YLC, EHW, AB, PS, NB, GB, CN, FA, MD, CC, GFH, BC, SLR, KC; Supervision: TK, MS, AT; Validation: MAD, AD, AM, CE, LD, PB, KB, DK, EC; Writing - Original Draft Preparation: MAD, AM, TK; Writing - Review and Editing: MAD, CE, DS, MS, TK

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at www.jidonline.org, and at https://doi.org/10.1016/j.jid.2023.02.038.

REFERENCES

Bardot ES, Valdes VJ, Zhang J, Perdigoto CN, Nicolis S, Hearn SA, et al. Polycomb subunits Ezh1 and Ezh2 regulate the Merkel cell differentiation program in skin stem cells. EMBO J 2013;32:1990–2000.

Becker JC, Stang A, DeCaprio JA, Cerroni L, Lebbé C, Veness M, et al. Merkel cell carcinoma. Nat Rev Dis Primers 2017;3:17077.

Bracken AP, Pasini D, Capra M, Prosperini E, Colli E, Helin K. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. EMBO J 2003;22:5323–35.

Busam KJ, Pulitzer MP, Coit DC, Arcila M, Leng D, Jungbluth AA, et al. Reduced H3K27me3 expression in Merkel cell polyoma virus-positive tumors. Mod Pathol 2017;30:877–83.

Cao R, Zhang Y. The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. Curr Opin Genet Dev 2004;14:155–64.

Chan-Penebre E, Armstrong K, Drew A, Grassian AR, Feldman I, Knutson SK, et al. Selective killing of SMARCA2- and SMARCA4-deficient small cell carcinoma of the ovary, hypercalcemic type cells by inhibition of EZH2: in vitro and in vivo preclinical models. Mol Cancer Ther 2017;16:850–60.

Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science 2008;319:1096–100.

Fujii S, Ito K, Ito Y, Ochiai A. Enhancer of zeste homologue 2 (EZH2) downregulates RUNX3 by increasing histone H3 methylation. J Biol Chem 2008;283:17324–32.

Gan L, Yang Y, Li Q, Feng Y, Liu T, Guo W. Epigenetic regulation of cancer progression by EZH2: from biological insights to therapeutic potential. Biomark Res 2018;6:10.

Gandhi JS, Alnoor F, Sadiq Q, Solares J, Gradowski JF. SMARCA4 (BRG1) and SMARCB1 (INI1) expression in TTF-1 negative neuroendocrine carcinomas including Merkel cell carcinoma. Pathol Res Pract 2021;219:153341.

Gardair C, Samimi M, Touzé A, Coursaget P, Lorette G, Caille A, et al. Somatostatin Receptors 2A and 5 are expressed in Merkel cell carcinoma with no association with disease severity. Neuroendocrinology 2015;101:223–35.

Gartin AK, Frost TC, Cushman CH, Leeper BA, Gokhale PC, DeCaprio JA. Merkel cell carcinoma sensitivity to EZH2 inhibition is mediated by SIX1 derepression. J Invest Dermatol 2022;142:2783–92.e15.

Goh G, Walradt T, Markarov V, Blom A, Riaz N, Doumani R, et al. Mutational landscape of MCPyV-positive and MCPyV-negative Merkel cell carcinomas with implications for immunotherapy. Oncotarget 2016;7:3403–15.

Gounder M, Schöffski P, Jones RL, Agulnik M, Cote GM, Villalobos VM, et al. Tazemetostat in advanced epithelioid sarcoma with loss of INI1/SMARCB1: an international, open-label, phase 2 basket study. Lancet Oncol 2020;21:1423–32.

Gujar H, Mehta A, Li HT, Tsai YC, Qiu X, Weisenberger DJ, et al. Characterizing DNA methylation signatures and their potential functional roles in Merkel cell carcinoma. Genome Med 2021;13:130.

Harms KL, Chubb H, Zhao L, Fullen DR, Bichakjian CK, Johnson TM, et al. Increased expression of EZH2 in Merkel cell carcinoma is associated with disease progression and poorer prognosis. Hum Pathol 2017;67:78–84.

Harms PW, Collie AMB, Hovelson DH, Cani AK, Verhaegen ME, Patel RM, et al. Next generation sequencing of cytokeratin 20-negative Merkel cell carcinoma reveals ultraviolet-signature mutations and recurrent TP53 and RB1 inactivation. Mod Pathol 2016;29:240–8.

Harms PW, Harms KL, Moore PS, DeCaprio JA, Nghiem P, Wong MKK, et al. The biology and treatment of Merkel cell carcinoma: current understanding and research priorities. Nat Rev Clin Oncol 2018;15:763–76.

Harms PW, Vats P, Verhaegen ME, Robinson DR, Wu YM, Dhanasekaran SM, et al. The distinctive mutational spectra of polyomavirus-negative Merkel cell carcinoma. Cancer Res 2015;75:3720–7.

Harms PW, Verhaegen ME, Hu K, Hrycaj SM, Chan MP, Liu CJ, et al. Genomic evidence suggests that cutaneous neuroendocrine carcinomas can arise from squamous dysplastic precursors. Mod Pathol 2022a;35:506–14.

Harms PW, Verhaegen ME, Vo JN, Tien JC, Pratt D, Su F, et al. Viral status predicts the patterns of genome methylation and decitabine response in Merkel cell carcinoma. J Invest Dermatol 2022b;142:641–52.

Hesbacher S, Pfitzer L, Wiedorfer K, Angermeyer S, Borst A, Haferkamp S, et al. RB1 is the crucial target of the Merkel cell Polyomavirus Large T antigen in Merkel cell carcinoma cells. Oncotarget 2016;7:32956–68.

Higaki-Mori H, Kuwamoto S, Iwasaki T, Kato M, Murakami I, Nagata K, et al. Association of Merkel cell polyomavirus infection with clinicopathological differences in Merkel cell carcinoma. Hum Pathol 2012;43:2282–91.

Houben R, Adam C, Baeurle A, Hesbacher S, Grimm J, Angermeyer S, et al. An intact retinoblastoma protein-binding site in Merkel cell polyomavirus large T antigen is required for promoting growth of Merkel cell carcinoma cells. Int J Cancer 2012;130:847–56.

Houben R, Angermeyer S, Haferkamp S, Aue A, Goebeler M, Schrama D, et al. Characterization of functional domains in the Merkel cell polyoma virus large T antigen. Int J Cancer 2015;136:E290–300.

Houben R, Shuda M, Weinkam R, Schrama D, Feng H, Chang Y, et al. Merkel cell polyomavirus-infected Merkel cell carcinoma cells require expression of viral T antigens. J Virol 2010;84:7064–72.

Italiano A. Role of the EZH2 histone methyltransferase as a therapeutic target in cancer. Pharmacol Ther 2016;165:26–31.

Italiano A, Soria JC, Toulmonde M, Michot JM, Lucchesi C, Varga A, et al. Tazemetostat, an EZH2 inhibitor, in relapsed or refractory B-cell non-Hodgkin lymphoma and advanced solid tumours: a first-in-human, open-label, phase 1 study. Lancet Oncol 2018;19:649–59.

Jiang T, Wang Y, Zhou F, Gao G, Ren S, Zhou C. Prognostic value of high EZH2 expression in patients with different types of cancer: a systematic review with meta-analysis. Oncotarget 2016;7:4584–97.

Kervarrec T, Aljundi M, Appenzeller S, Samimi M, Maubec E, Cribier B, et al. Polyomavirus-positive Merkel cell carcinoma derived from a

trichoblastoma suggests an epithelial origin of this Merkel cell carcinoma. J Invest Dermatol 2020a;140:976–85.

Kervarrec T, Appenzeller S, Samimi M, Sarma B, Sarosi EM, Berthon P, et al. Merkel cell polyomavirus–negative -Merkel cell carcinoma originating from in situ squamous cell carcinoma: a keratinocytic tumor with neuroendocrine differentiation. J Invest Dermatol 2022;142:516–27.

Kervarrec T, Gaboriaud P, Berthon P, Zaragoza J, Schrama D, Houben R, et al. Merkel cell carcinomas infiltrated with CD33+ myeloid cells and CD8+ T cells are associated with improved outcome. J Am Acad Dermatol 2018;78:973–82.e8.

Kervarrec T, Samimi M, Hesbacher S, Berthon P, Wobser M, Sallot A, et al. Merkel cell polyomavirus T antigens induce Merkel cell-like differentiation in GLI1-expressing epithelial cells. Cancers (Basel) 2020b;12:1989.

Kervarrec T, Tallet A, Miquelestorena-Standley E, Houben R, Schrama D, Gambichler T, et al. Diagnostic accuracy of a panel of immunohistochemical and molecular markers to distinguish Merkel cell carcinoma from other neuroendocrine carcinomas. Mod Pathol 2019;32:499–510.

Kim KH, Kim W, Howard TP, Vazquez F, Tsherniak A, Wu JN, et al. SWI/SNF-mutant cancers depend on catalytic and non-catalytic activity of EZH2. Nat Med 2015;21:1491–6.

Knepper TC, Montesion M, Russell JS, Sokol ES, Frampton GM, Miller VA, et al. The genomic landscape of Merkel cell carcinoma and clinicogenomic biomarkers of response to immune checkpoint inhibitor therapy. Clin Cancer Res 2019;25:5961–71.

Kurmasheva RT, Sammons M, Favours E, Wu J, Kurmashev D, Cosmopoulos K, et al. Initial testing (stage 1) of tazemetostat (EPZ-6438), a novel EZH2 inhibitor, by the Pediatric Preclinical Testing Program. Pediatr Blood Cancer 2017;64:10.1002/pbc.26218.

Leiendecker L, Jung PS, Krecioch I, Neumann T, Schleiffer A, Mechtler K, et al. LSD1 inhibition induces differentiation and cell death in Merkel cell carcinoma. EMBO Mol Med 2020;12:e12525.

Leonard JH, Dash P, Holland P, Kearsley JH, Bell JR. Characterisation of four Merkel cell carcinoma adherent cell lines. Int J Cancer 1995;60:100–7.

Lhuissier E, Aury-Landas J, Lenté M, Boumediene K, Baugé C. Co-treatment with the epigenetic drug, 3-deazaneplanocin A (DZNep) and cisplatin after DZNep priming enhances the response to platinum-based therapy in chondrosarcomas. Cancers (Basel) 2021;13:4648.

Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. Nature 2011;469:343–9.

Matsushita M, Iwasaki T, Wardhani LO, Kuwamoto S, Nonaka D, Nagata K, et al. Decreased H3K27me3 Expression Is Associated with Merkel cell Polyomavirus-negative Merkel cell Carcinoma, Especially Combined with cutaneous squamous cell carcinoma. Anticancer Res 2019;39:5573–9.

Morschhauser F, Tilly H, Chaidos A, McKay P, Phillips T, Assouline S, et al. Tazemetostat for patients with relapsed or refractory follicular lymphoma: an open-label, single-arm, multicentre, phase 2 trial. Lancet Oncol 2020;21:1433–42.

Moshiri AS, Doumani R, Yelistratova L, Blom A, Lachance K, Shinohara MM, et al. Polyomavirus-negative Merkel cell carcinoma: A more aggressive subtype based on analysis of 282 cases using multimodal tumor virus detection. J Invest Dermatol 2017;137:819–27.

Pajtler KW, Wen J, Sill M, Lin T, Orisme W, Tang B, et al. Molecular heterogeneity and CXorf67 alterations in posterior fossa group A (PFA) ependymomas. Acta Neuropathol 2018;136:211–26.

Park DE, Cheng J, McGrath JP, Lim MY, Cushman C, Swanson SK, et al. Merkel cell polyomavirus activates LSD1-mediated blockade of non-canonical BAF to regulate transformation and tumorigenesis [published correction appears in Nat Cell Biol 2020;22:752] Nat Cell Biol 2020;22:603–15.

Perdigoto CN, Dauber KL, Bar C, Tsai PC, Valdes VJ, Cohen I, et al. Polycomb-mediated repression and sonic hedgehog signaling interact to regulate Merkel cell specification during skin development. PLoS Genet 2016;12:e1006151.

Rosen ST, Gould VE, Salwen HR, Herst CV, Le Beau MM, Lee I, et al. Establishment and characterization of a neuroendocrine skin carcinoma cell line. Lab Invest 1987;56:302–12.

Rotondo JC, Mazziotta C, Lanzillotti C, Tognon M, Martini F. Epigenetic dysregulations in Merkel cell polyomavirus-driven Merkel cell carcinoma. Int J Mol Sci 2021;22:11464.

Sasaki M, Yamaguchi J, Itatsu K, Ikeda H, Nakanuma Y. Over-expression of polycomb group protein EZH2 relates to decreased expression of p16 INK4a in cholangiocarcinogenesis in hepatolithiasis. J Pathol 2008;215:175–83.

Starrett GJ, Thakuria M, Chen T, Marcelus C, Cheng J, Nomburg J, et al. Clinical and molecular characterization of virus-positive and virus-negative Merkel cell carcinoma. Genome Med 2020;12:30.

Thanguturi S, Tallet A, Miquelestorena-Standley E, Coco C, Le Corre Y, Hainaut-Wierzbicka E, et al. Investigation of the RB1-SOX2 axis constitutes a tool for viral status determination and diagnosis in Merkel cell carcinoma. Virchows Arch 2022;480:1239–54.

Tsai KY. The origins of Merkel cell carcinoma: defining paths to the neuroendocrine phenotype. J Investig Dermatol 2021;142:507–9.

Veija T, Koljonen V, Bohling T, Kero M, Knuutila S, Sarhadi VK. Aberrant expression of ALK and EZH2 in Merkel cell carcinoma. BMC Cancer 2017;17:236.

Verhaegen ME, Harms PW, Van Goor JJ, Arche J, Patrick MT, Wilbert D, et al. Direct cellular reprogramming enables development of viral T antigen–driven Merkel cell carcinoma in mice. J Clin Invest 2022;132:e152069.

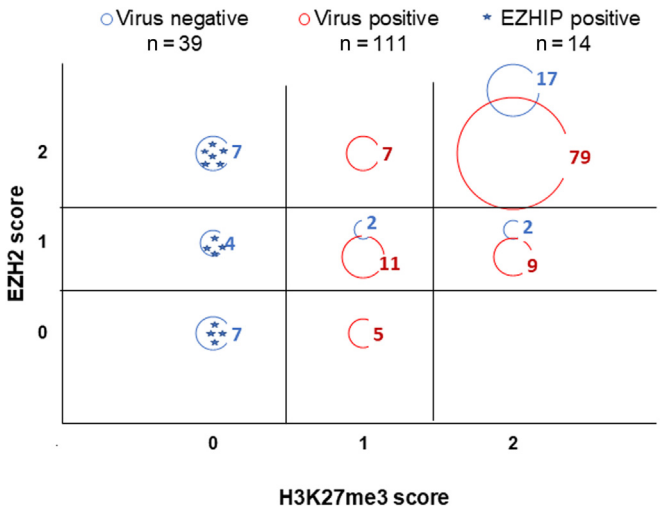
Verhaegen ME, Mangelberger D, Harms PW, Vozheiko TD, Weick JW, Wilbert DM, et al. Merkel cell polyomavirus small T antigen is oncogenic in transgenic mice. J Invest Dermatol 2015;135:1415–24.

Walker JW, Lebbé C, Grignani G, Nathan P, Dirix L, Fenig E, et al. Efficacy and safety of avelumab treatment in patients with metastatic Merkel cell carcinoma: experience from a global expanded access program. J Immunother Cancer 2020;8:e00031corr1.

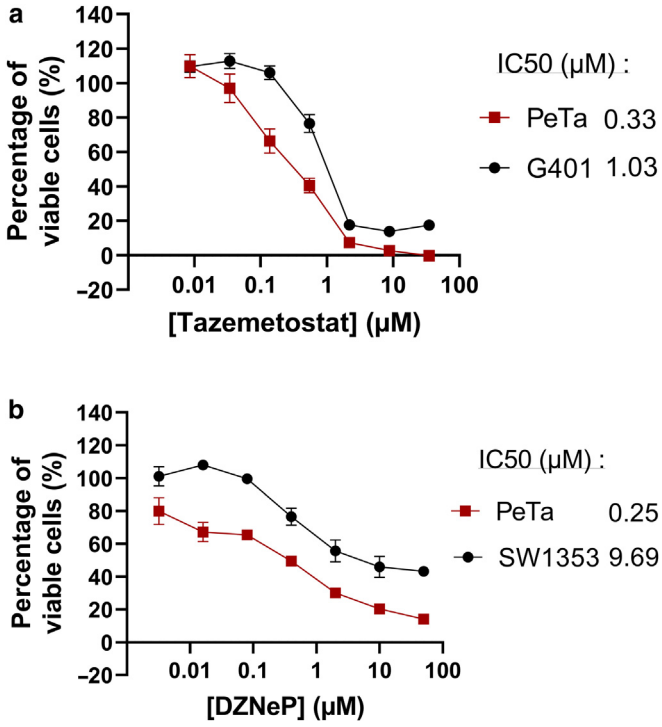
Wu G, Broniscer A, McEachron TA, Lu C, Paugh BS, Becksfort J, et al. Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. Nat Genet 2012;44:251–3.

Xu X, Gu J, Ding X, Ge G, Zang X, Ji R, et al. LINC00978 promotes the progression of hepatocellular carcinoma by regulating EZH2-mediated silencing of p21 and E-cadherin expression. Cell Death Dis 2019;10:752.

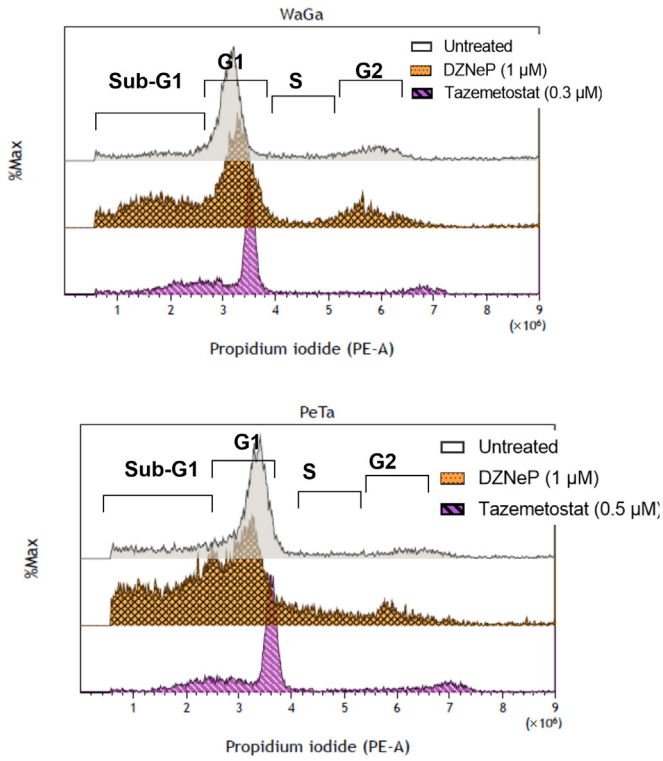
Zhong J, Min L, Huang H, Li L, Li D, Li J, et al. EZH2 regulates the expression of p16 in the nasopharyngeal cancer cells. Technol Cancer Res Treat 2013;12:269–74.



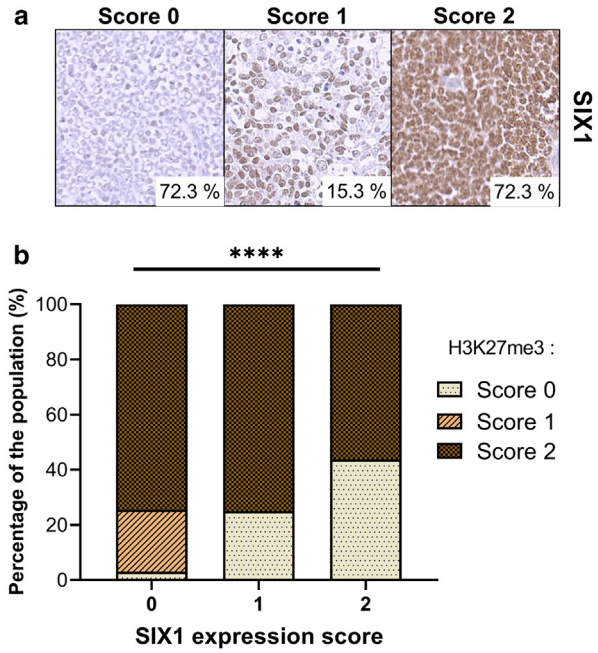
Supplementary Figure S1. Graphical representation of EZH2, H3K27me3, and EZHIP among virus-positive and virus-negative MCC. Only cases with interpretable EZH2 and H3K27me3 staining and interpretable viral status are compared. EZH2, enhancer of zeste homolog 2; H3K27me3, trimethylation of lysine 27 on histone 3; MCC, Merkel cell carcinoma.

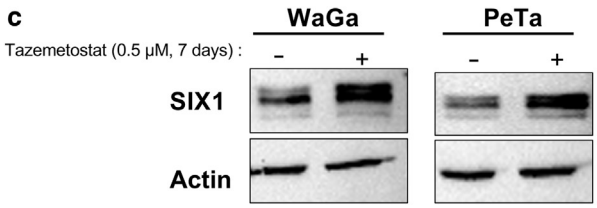


Supplementary Figure S2. Comparison of MCC and EZH2i-sensitive cell lines responses with EZH2is. (**a**) Cytotoxicity assay of tazemetostat treatment during 14 days on PeTa and G401 cells. (**b**) Cytotoxicity assay of DZNeP treatment during 14 days on PeTa and SW1353 cells. n = 3; mean ± SEM. DZNeP, 3-deazaneplanocin A; EZH2i, enhancer of zeste homolog 2 inhibitor; H3K27me3, trimethylation of lysine 27 on histone 3; IC50, half-maximal inhibitory concentration; MCC, Merkel cell carcinoma.

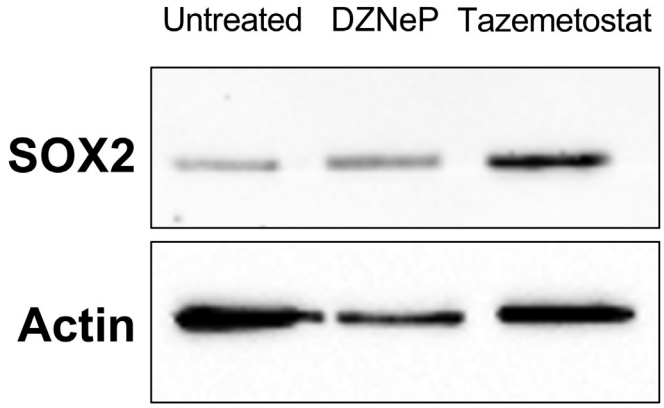


Supplementary Figure S3. Representative cell cycle of WaGa and PeTa MCC cell lines treated or not with tazemetostat for 7 days at 0.3 μM or DZNeP for 4 days at 1 μM. (**a**) WaGa MCC. (**b**) PeTa MCC. DZNeP, 3-deazaneplanocin A; Max, maximum; MCC, Merkel cell carcinoma.





Supplementary Figure S4. H3K27me3 represses SIX1 expression in MCC. (**a**) Immunological staining of SIX1 in MCC tumors samples. Representative MCC cases with no (score 0), low (score 1), and high (score 2) staining are shown (n = 137). (**b**) H3K27me3 expression score according to SIX1 score in the MCC cohort. (**c**) Immunoblot of SIX1 from WaGa and PeTa cells treated with 0.5 µM of tazemetostat for 7 days. Data are representative of two experiments. H3K27me3, trimethylation of lysine 27 on histone 3; MCC, Merkel cell carcinoma.



Supplementary Figure S5. Loss of H3K27me3 induces SOX2 expression. Immunoblot of SOX2 from WaGa cells treated with 0.5 µM of tazemetostat for 7 days or 1µM of DZNeP for 4 days. DZNeP, 3-deazaneplanocin A; H3K27me3, trimethylation of lysine 27 on histone 3.

Supplementary Table S1. EZH2, H3K27me3, and EZHIP Scores among Non-MCC Cutaneous Epithelial Tumors Cohort

	EZH2 Expression Score				H3K27me3 Score				EZHIP Expression Score			
Non-MCC Primary Epithelial Tumors (n = 232)	**0, n (%)**	**1, n (%)**	**2, n (%)**	**NR**	**0, n (%)**	**1, n (%)**	**2, n (%)**	**NR**	**0, n (%)**	**1, n (%)**	**2, n (%)**	**NR**
Squamous cell carcinoma (n = 84)	5 (9)	20 (35)	32 (56)	27	6 (10)	5 (8)	51 (82)	22	57 (84)	4 (6)	7 (10)	16
In situ/Bowen's disease (n = 20)	1 (7)	1 (7)	12 (86)	6	3 (23)	0	10 (77)	7	7 (54)	2 (15)	4 (31)	7
Well-differentiated (n = 26)	1 (6)	6 (35)	10 (59)	9	0	0	18 (100)	8	23 (100)	0	0	3
Moderately differentiated (n = 12)	0	3 (38)	5 (62)	4	0	1 (10)	9 (90)	2	9 (100)	0	0	3
Poorly differentiated (n = 26)	3 (17)	10 (56)	5 (27)	8	3 (14)	4 (19)	14 (67)	5	18 (78)	2 (9)	3 (13)	3
Basal cell carcinoma (n = 48)	0	7 (20)	29 (80)	12	0	0	34(100)	14	36(100)	0	0	12
Superficial (n = 5)	0	1 (33)	2 (66)	2	0	0	4 (100)	1	4 (100)	0	0	1
Nodular (n = 20)	0	3 (17)	15 (83)	2	0	0	14 (100)	6	14 (100)	0	0	6
Infiltrative (n = 22)	0	3 (21)	11 (79)	8	0	0	15 (100)	7	17 (100)	0	0	5
Sclerosing (n = 1)	0	0	1 (100)	0	0	0	1 (100)	0	1 (100)	0	0	0
Tumors with follicular différenciation (n = 32)	10 (42)	7 (29)	7 (29)	8	1 (4)	0	22 (96)	9	26(100)	0	0	6
Trichoblastoma (n = 10)	2 (25)	5 (63)	1 (12)	2	0	0	8 (100)	2	8 (100)	0	0	2
Trichoepithelioma (n = 7)	4 (100)	0	0	3	0	0	4 (100)	3	5 (100)	0	0	2
Trichilemmoma (n = 5)	3 (60)	2 (40)	0	0	0	0	4 (100)	1	4 (100)	0	0	1
Proliferating pilar tumor (n = 4)	0	0	3 (100)	1	0	0	3 (100)	1	4 (100)	0	0	0
Pilomatrixoma (n = 5)	0	0	3 (100)	2	0	0	3 (100)	2	4 (100)	0	0	1
Pilomatricial carcinoma (n = 1)	1 (100)	0	0	0	1 (100)	0	0	0	1 (100)	0	0	0
Tumors with sebaceous différenciation (n = 11)	6 (67)	3 (33)	0	2	0	1 (11)	8 (89)	2	9 (100)	0	0	2
Sebaceous adenoma (n = 9)	5 (71)	2 (29)	0	2	0	0	7 (100)	2	7 (100)	0	0	2
Sebaceous carcinoma (n = 2)	1 (50)	1 (50)	0	0	0	1 (50)	1 (50)	0	2 (100)	0	0	0
Tumors with eccrine/apocrine différenciation (n = 57)	23 (53)	17 (40)	3 (7)	14	0	6 (15)	35 (85)	16	39 (91)	4 (9)	0	14
Hidradenoma (n = 10)	5 (56)	4 (44)	0	1	0	2 (22)	7 (78)	1	5 (63)	3 (47)	0	2
Spiradenoma (n = 11)	2 (22)	7 (78)	0	2	0	1 (10)	9 (90)	1	11 100)	0	0	0
Cylindroma (n = 2)	2 (100)	0	0	0	0	0	2 (100)	0	2 (100)	0	0	0
Paget's disease (n = 6)	2 (40)	1 (20)	2 (40)	1	0	0	4 (100)	2	5 (100)	0	0	1
Poroma (n = 12)	4 (44)	4 (44)	1 (11)	3	0	2 (22)	7 (78)	3	7 (88)	1 (12)	0	4
Chondroid syringoma (n = 3)	1 (100)	0	0	2	0	0	1 (100)	2	2 (100)	0	0	1
Myoepithelioma (n = 3)	3 (100)	0	0	0	0	0	3 (100)	0	3 (10)	0	0	0
Porocarcinoma (n = 4)	1 (100)	0	0	3	0	1 (100)	0	3	1 (100)	0	0	3
Adenoid cystic carcinoma (n = 2)	1 (50)	1 (50)	0	0	0	0	2 (100)	0	2 (100)	0	0	0
Microcystic adnexal carcinoma (n = 1)	1 (100)	0	0	0	0	0	0	1	0	0	0	1
Sweat gland carcinoma NOS (n = 3)	1 (100)	0	0	2	0	0	1 (100)	2	1 (100)	0	0	2
Total	44 (26)	54 (32)	71 (42)	63	7 (4)	12 (7)	150 (89)	63	167 (92)	8 (4)	7 (4)	50

Abbreviations: EZH2, enhancer of zeste homolog 2; H3K27me3, trimethylation of lysine 27 on histone 3; MCC, Merkel cell carcinoma; NOS, not otherwise specified; NR, not reported.

Supplementary Table S2. Clinical and Virologic Characteristics of Patients with MCC according to EZH2 Expression Score

EZH2 Expression Score	Score 0 n = 14	Scores 1–2 n = 156	*P*-Value
Age, y			
Median, Q1–Q3	81.7, 75–85	80, 70–85	0.395
Data not available	1	17	
Sex, n (%)			
F/M	8 (62)/5 (38)	83 (56)/66 (44)	0.776
Data not available	1	7	
Localization, n (%)			0.474
Head and neck	6 (46)	53 (37)	
Limb	4 (31)	70 (48)	
Trunk	1 (8)	6 (4)	
Occult	2 (15)	16 (11)	
	1	11	
Immunosuppression, n (%)			
Yes/no	0 (0)/12 (100)	26 (21)/98 (79)	0.122
Data not available	8	26	
AJCC stage at diagnosis, n (%)			0.888
I	3 (25)	38 (28)	
II	5 (42)	47 (35)	
III	4 (33)	46 (34)	
IV	0 (0)	5 (4)	
Data not available	2	20	
Metastasis at diagnosis, n (%)			0.227
Yes/no	8 (67)/4 (33)	58 (45)/70 (55)	
Data not available	2	28	
Viral status, n (%)			0.014
Virus negative/positive	7 (58)/5 (42)	32 (23)/106 (77)	
Data not available	2	18	

Abbreviations: AJCC, American Joint Committee on Cancer; EZH2, enhancer of zeste homolog 2; F, female; M, male; MCC, Merkel cell carcinoma; Q1, first quartile; Q3, third quartile.

Supplementary Table S3. Clinical and Virologic Characteristics of Patients with MCC according to H3K27me3 Score

H3K27me3 Score	Score 0 n = 20	Scores 1–2 n = 146	*P*-Value
Age, y			0.395
Median, Q1–Q3	83, 76–86	80, 69–84	
Data not available	2	15	
Gender, n (%)			
F/M	13 (68)/6 (32)	77 (55)/62 (45)	0.331
Data not available	1	7	
Localization, n (%)			0.448
Head and neck	10 (55)	48 (35)	
Limb	6 (33)	66 (49)	
Trunc	1 (6)	6 (4)	
Occult	1 (6)	16 (12)	
	2	10	
Immunosuppression, n (%)			
Yes/no	1 (8)/12 (63)	24 (20)/95 (78)	0.461
Data not available	7	27	
AJCC stage at diagnosis, n (%)			0.662
I	5 (31)	34 (27)	
II	4 (25)	47 (37)	
III	7 (44)	43 (33)	
IV	0 (0)	4 (3)	
Data not available	4	18	
Metastasis at diagnosis, n (%)			
Yes/no	8 (62)/5 (38)	55 (45)/68 (55)	0.381
Data not available	7	23	
Viral status, n (%)			<0.0001
Virus negative/positive	18 (100)/0 (0)	21 (16)/111 (84)	
Data not available	2	14	

Abbreviations: AJCC, American Joint Committee on Cancer; F, female; H3K27me3, trimethylation of lysine 27 on histone 3; M, male; Q1, first quartile; Q3, third quartile.

A Fischer exact test was performed for qualitative variables. A Mann–Whitney test was performed for quantitative variables. F; M; Q1; Q3.

Supplementary Table S4. H3K27me3 Score according to H3K27me3 Score among Virus-Positive and Virus-Negative MCC Tumors

		MCPyV(+)			MCPyV(−)			
		EZH2			EZH2			
Markers		Score 0	Score 1	Score 2	Score 0	Score 1	Score 2	Total
H3K27me3	Score 0	0	0	0	7	4	7	18
	Score 1	5	11	7	0	2	0	25
	Score 2	0	9	79	0	2	17	107
Total		5	20	86	7	8	24	150
			111			39		

Abbreviations: EZH2, enhancer of zeste homolog 2; H3K27me3, trimethylation of lysine 27 on histone 3; MCC, Merkel cell carcinoma; MCPyV, Merkel cell polyomavirus.

Supplementary Table S5. EZH2, H3K27me3, and EZHIP Expression among the Six Additional Combined MCC Cases

	MCC Component			SCC Component		
	EZH2	H3K27me3	EZHIP	EZH2	H3K27me3	EZHIP
Case 1	2	0	2	2	2	0
Case 2	2	2	0	2	2	0
Case 3	2	2	0	2	2	0
Case 4	2	0	2	2	0	2
Case 5	2	0	2	2	0	2
Case 6	2	0	2	2	0	2

Abbreviations: EZH2, enhancer of zeste homolog 2; H3K27me3, trimethylation of lysine 27 on histone 3; MCC, Merkel cell carcinoma; SCC, squamous cell carcinoma.

Supplementary Table S6. Antibodies Used for Immunochemistry

Specificity	Reference	Dilution
Anti-EZH2	611267, BD Biosciences	1:200
Anti-H3K27me3	9733S, Cell Signaling Technology	1:400
Anti-SOX2	Ab92494, Abcam	1:50
Anti-SIX1	HPA001893, Sigma-Aldrich	1:800
Anti-EZHIP	HPA006128, Sigma-Aldrich	1:30
Anti-SMARCA4 (*BRG1*)	Ab110641, Abcam	1:100
Anti-SMARCB1 (*BAF47*)	Clone 25/BAF47, BD dilution	1:100

Abbreviations: EZH2, enhancer of zeste homolog 2; H3K27me3, trimethylation of lysine 27 on histone 3.

Supplementary Table S7. Sequences of Primers Used for RT-qPCR

Target Gene	Forward Primer	Reverse Primer
RPLP0	5′CCATCAGCACCACAGCCTTC3′	5′GGCGACCTGGAAGTCCAACT3′
TA	5′CCCCTCCTCAAGAAGCTC3′	5′AGAGCTTGCAGAGCTTCG3′
EZH2	5′AATCAGAGTAC ATGCGACTGAGA3′	5′GCTGTATCCTTCGCTGTTTCC 3′

Abbreviations: EZH2, enhancer of zeste homolog 2; TA, T antigen.

Supplementary Table S8. Antibodies Used for Immunoblot

Specificity	Reference	Dilution
Anti-EZH2	611267, BD Biosciences	1:1,000
Anti-H3K27me3	C15410069, Diagenode	1:2,000
Anti-SOX2	Ab92494, Abcam	1:1,000
Anti-SIX1	HPA001893, Sigma-Aldrich	1:1,000
Anti-LT	CM2B4, Santa Cruz Biotechnologies	1:1,000
Anti-actin	A1978; Sigma-Aldrich	1:1,000

Abbreviations: EZH2, enhancer of zeste homolog 2; H3K27me3, trimethylation of lysine 27 on histone 3; LT, large T.

Supplementary Table S9. Antibodies Used for Flow Cytometry

Specificity	Reference	Dilution
Anti-H3K27me3	9733S, Cell Signaling Technology	1:400

Abbreviation: H3K27me3, trimethylation of lysine 27 on histone 3.