UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE des Sciences de la Vie et de la Santé

IGBMC – CNRS UMR 7104 – Inserm U 1258

THÈSE présentée par :

Alexia VIDAL

Soutenue le : **9 décembre 2019**

Pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline : Sciences de la Vie et de la Santé

Spécialité : Aspects Moléculaires et Cellulaires de la Biologie

Rôle des mutations de l'histone variante H3.3 dans le développement des gliomes pédiatriques de haut grade

THÈSE dirigée par :
Dr HAMICHE Ali Directeur de Recherche CNRS, IGBMC, Illkirch

RAPPORTEURS :
Dr FRANCASTEL Claire Directeur de Recherche INSERM, Université Paris 7 Diderot
Dr TRONO Didier Professeur, EPFL, Lausanne

AUTRES MEMBRES DU JURY :
Dr ALMOUZNI Geneviève Directeur de Recherche CNRS, Institut Curie, Paris
Dr DAVIDSON Irwin Directeur de Recherche CNRS, IGBMC, Illkirch

UNIVERSITY OF STRASBOURG

GRADUATE SCHOOL 414

IGBMC – CNRS UMR 7104 – Inserm U 1258

THESIS presented by :

Alexia VIDAL

Defended on : **December 9, 2019**

For the degree of : **Doctor of Philosophy**

Discipline : Life and Health Sciences

Speciality : Molecular and Cellular Aspects of Biology

Role of histone variant H3.3 mutations in the development of pediatric high-grade glioma

THESIS directed by :
Dr HAMICHE Ali — Director of Research CNRS, IGBMC, Illkirch

RAPPORTEURS :
Dr FRANCASTEL Claire — Director of Research INSERM, Université Paris 7 Diderot
Dr TRONO Didier — Professor, EPFL, Lausanne

OTHER JURY MEMBERS :
Dr ALMOUZNI Geneviève — Director of Research CNRS, Institut Curie, Paris
Dr DAVIDSON Irwin — Director of Research CNRS, IGBMC, Illkirch

À mes parents,
À mes grands-parents,
À Thibaut.

« Je suis de ceux qui pensent que la science a une grande beauté… Un savant dans son laboratoire n'est pas seulement un technicien : c'est aussi un enfant placé en face des phénomènes naturels qui l'impressionnent comme un conte de fées. »

« On ne fait jamais attention à ce qui a été fait ;
on ne voit que ce qui reste à faire. »

Marie Curie

Thèse réalisée dans le cadre du dispositif CIFRE
Conventions Industrielles de Formation à la Recherche
(N°2016/0482)

Cette thèse a été élaborée :

À l'**Institut de Génétique et de Biologie Moléculaire et Cellulaire**
de l'Université de Strasbourg
« Génomique Fonctionnelle et Cancer – Chromatine et régulation épigénétique »
IGBMC - CNRS UMR 7104 - Inserm U 1258
1 rue Laurent Fries, 67404 Illkirch Cedex

et

À **Inoviem Scientific**
Bioparc 3, 850 boulevard Sébastien Brant, 67400 Illkirch-Graffenstaden

Remerciements

Je tiens tout d'abord à remercier mes rapporteurs ***Claire Francastel***, ***Didier Trono***, ainsi que ***Geneviève Almouzni*** et ***Irwin Davidson*** pour avoir accepté d'évaluer mon travail de thèse.

Merci à mon directeur de thèse, ***Ali,*** pour m'avoir donné la chance de rejoindre ce domaine passionnant. Merci de m'avoir fait confiance, de m'avoir donné la liberté et tous les moyens nécessaires pour mener à bien ce projet.

Je souhaite également remercier ***Pierre,*** sans qui je n'aurai pas eu la chance et le plaisir d'effectuer cette thèse sous le dispositif CIFRE avec Inoviem Scientific. Merci pour ta confiance et pour me pousser à sans cesse me remettre en question.

Un grand merci à ***Rachel*** pour ton soutien sans faille, ton sourire omniprésent et pour tous tes conseils. Tu as su me remotiver dans les moments les plus durs, me rassurer dans les moments de doutes, me guider vers de nouvelles pistes. C'est un vrai plaisir de travailler à tes côtés, merci pour tout.

Je tiens à remercier toute l'équipe d'Inoviem Scientific, ***Angélique, Daniel, Diane, Edwige, François, Jérôme, Judith, Léone, Pierre*** et ***Rachel***, sans oublier ***Isabel*** et ***Laura***. C'est un vrai plaisir de travailler avec vous, vous avez été une bouffée d'air frais dans mes semaines. Merci pour les nombreux afterworks, pour tous ces moments passés et à venir.

Un énorme merci à ***Christophe***, pour m'avoir fait découvrir le monde du genome-wide et des séquences répétées, pour m'avoir aiguillée tout au long de mon projet, pour toutes nos discussions scientifiques, pour tes conseils et ton soutien, pour ton calme à toute épreuve, pour tes conseils et corrections pour mon manuscrit. Désolée pour le saumon, mais je suis sûre que tu finiras par l'apprécier !

Merci ***Isabelle***, pour le partage de ton expertise technique et de ton expérience, ta disponibilité quotidienne, ta rigueur, ta gentillesse, et pour tous les bons moments passés autour d'un café ou à l'heure du déjeuner.

Je tiens à remercier ***Stéphanie*** pour toutes les analyses bio-informatiques en lien avec ce projet, c'est un réel plaisir d'échanger avec toi. Merci pour toutes les discussions passionnantes, ta motivation, ta disponibilité et toutes ces heures passées à analyser mes données.

Merci à l'ensemble de l'équipe du laboratoire, et particulièrement à ***Dimitra****,* ***Maria****,* ***Catherin****e* et ***Philippe*** avec qui j'ai pu partager le module. Un grand merci à ***Chrysa*** pour son accueil lors de mon arrivée au sein de l'équipe, son encadrement et sa disponibilité.

Ce travail de recherche n'aurait pu aboutir sans la participation de nombreux collaborateurs et plateformes au sein de l'IGBMC. Un merci tout particulier à toute l'équipe du service de culture cellulaire, et spéciale dédicace à ***Marion*** et ***Amélie*** pour m'avoir formée à la culture des cellules souches, et pour leur aide très précieuse sur mon projet. Vous avez toujours été là, motivées et persévérantes face aux très nombreux (et obscurs) problèmes, pour votre sourire et votre flexibilité, un grand merci. Merci à ***Marie*** pour la réalisation de la stratégie du modèle dans les cellules souches, pour tes conseils et ta précieuse aide. Merci à la team de la salle ES pour les nombreux week-end dans la bonne humeur et la musique, notamment à ***Vincent***, ***Paul****,* et ***Dominique*****.** Merci à toute l'équipe Godin pour vos retours sur l'aspect neuro, merci ***Juliette*** pour tes nombreux conseils et merci ***Peggy*** pour m'avoir formée à la dissection, à la coupe et à l'immunomarquage sur cerveaux de souris. Merci à l'équipe de la plateforme de séquençage Genomeast, et en particulier à ***Romain****.* C'était un plaisir de retravailler ensemble depuis le DUT. Merci à ***Michael*** et à toute l'équipe de l'animalerie pour avoir pris grand soin de mes souris.

Cette aventure a aussi été l'occasion de créer de belles amitiés. Un énorme merci à ***Arantxa, Bea, Katka, Jordi, Xenia, Paulo, Matej, Daniel*** et les autres, pour les déjeuners à toute heure - en français, en anglais, espagnol, tchèque ou que sais-je, les bières improvisées, les sorties randos, les week-ends, tous ces moments qui rendent la thèse encore plus belle. Merci ***Mario*** le geek, pour les chouettes moments à Roscoff, ton aide sur la bio-info, nos discussions-revendications, mais surtout pour celle belle amitié qui est née.

Outre les personnes formidables côtoyées au niveau professionnel, de nombreuses personnes dans ma vie personnelle ont participé de près ou de loin à l'accomplissement de ces trois années de thèse mettant fin à de longues et intenses années d'étude.

Je tiens tout d'abord à remercier de tout mon cœur mon compagnon de vie de ces 11 dernières années, ***Thibaut***. Merci pour ton soutien sans faille tout au long de mes études, pour ta présence quotidienne pendant ces trois dernières années. Tu as su me porter, et surtout me supporter. Cette thèse, c'est aussi la tienne. Pour toutes tes attentions, pour avoir toujours été présent avec les mots justes dans les moments les plus durs, pour avoir accepté et subit mon investissement dans ce projet très prenant, mais surtout pour tous les merveilleux moments partagés et à venir, milles mercis. Je suis heureuse de voir à quel point la relation qui nous unit est forte et j'ai hâte de continuer à écrire les pages de cette belle histoire.

Je souhaite également remercier mes amis de longue date dispersés aux quatre coins du globe. ***Baptiste***, mon pilote préféré, tu es celui qui comprend le mieux l'abnégation et la persévérance qu'une telle aventure peut demander. Pour ton soutien, même à distance, pour tous les moments partagés depuis la Croix Blanche, jusqu'à Jerez, toujours en Alsace et bientôt jusqu'à Dublin, pour cette amitié en or, merci. Merci ma ***Pauline,*** tant de belles aventures, des « babes » en or, on traverserait les continents avec ***Pierre*** et toi ! ***Giulia***, quelle chance j'ai eu de croiser ton chemin au pays du soleil levant. Depuis, c'est un vrai bonheur de te retrouver avec ***Gio*** pour des aventures toujours plus folles. ***Leslie, Bryan,*** merci pour toutes ces soirées jeu, restos, repas, sorties et ce beau voyage aux USA. Merci à la team des lyonnais, à ces plongeurs du ***GLUP (Aude-Line, Tatane, Pâquerette, Cricri, Marine, Oliv', Alex),*** toujours dispos pour profiter des bulles (sous l'eau ou dans la bière), ça fait toujours un bien fou de passer du temps avec vous. ***Diane***, bien plus qu'une collègue tu es devenue une amie. Merci pour ton soutien et vivement le prochain UBY ! Un grand merci au « **groupe** » pour les nombreuses et folles soirées, tous ces moments partagés et d'avoir accepté mes quelques siestes sur le canapé en soirée. Merci à notre chère docteur ***CamCam*** pour ton aide en neuro, merci ***Estelle, Julie, Marie-O, Dama, Thomas, Flo, Franck, Julien, Romain.*** Vous êtes des amis formidables.

Un grand merci à l'équipe interclub de l'AS Badminton Soufflenheim. Votre bonne humeur, votre présence hebdomadaire et votre jeu du tonnerre m'a aidé à me vider la tête et à conserver une santé de fer. Merci ***Marc-André*** pour ton soutien, ta motivation et tes encouragements, on a un président de compétition ! Un grand merci à ***Meumeu*** et ***Line*** pour les moments détentes et rigolades, toujours accompagnés de bonnes bulles. Merci à ***Valentin*** pour m'avoir proposé un nouveau challenge, avec beaucoup de chance et de talent notre double mixte finira en N3 dans 3 ans !

Merci aux amis musiciens de l'harmonie, ***Patrick, Fabi, Florence, JC,*** rien de mieux que de venir jouer pour recharger les batteries et se libérer l'esprit.

Enfin, je souhaite remercier ma famille, sans qui je n'en serai évidemment pas là, qui a su accepter mes choix d'étude et de vie, et qui m'a toujours soutenue. Je tiens également à remercier la famille de Thibaut qui m'a acceptée il y a de nombreuses années, merci pour vos encouragements et votre soutien.

Merci ***papa*** d'être toujours là pour moi. Même si on n'échange pas beaucoup de mots, je suis fière d'avoir héritée de ta ténacité. Merci ***maman*****,** merci pour notre complicité, merci pour ton éternel soutien, merci d'avoir fait la femme forte que je suis devenue. Merci à vous deux de m'avoir permis de suivre mes envies, de réaliser mes propres choix, merci pour tous les sacrifices réalisés pour financer mes études et mes activités. On ne choisit pas ses parents, mais les miens sont en or. Un grand merci à ***mamie*** et ***papi*** de m'avoir élevée et d'avoir sans cesse développé et cultivé ma curiosité. Je suis désolée d'avoir été si souvent loin ou occupée ces dernières années, mais je vous suis extrêmement reconnaissante de tout l'amour que vous m'avez apporté.

Le moment de clôturer ce manuscrit avec ces remerciements à sans doute été la partie la plus difficile émotionnellement. J'espère n'avoir oublier personne et je souhaite une fois de plus remercier l'ensemble des personnes ayant fait de ces trois dernières années une expérience humaine et scientifique intense, passionnante et tout simplement inoubliable.

Résumé en français

Les tumeurs cérébrales constituent le deuxième cancer pédiatrique le plus fréquent après les leucémies. Les enfants touchés développent ces tumeurs entre 7 et 13 ans et l'absence de traitement efficace se traduit par un pronostic sombre avec une espérance de vie inférieure à 2 ans. Les mutations de l'histone variante H3.3 K27M ou G34R ont été identifiées dans 30 % des gliomes pédiatriques de haut-grade et ont été décrites comme moteur de leur développement, mais le mécanisme sous-jacent reste à établir. L'histone variante H3.3 diffère des histones canoniques H3.1 et H3.2 par seulement 5 ou 4 résidus aminés, respectivement. H3.3 est codée par deux gènes (H3f3a et H3f3b) mais seul un allèle de H3f3a est touché par les mutations K27M ou G34R. Ces mutations sont somatiques et mutuellement exclusives. De plus, un antagonisme existe entre K27M et G34R concernant la localisation des tumeurs, l'âge de diagnostic ainsi que le pronostic associé. Le but de ma thèse a été de comprendre l'impact des mutations de H3.3 sur sa distribution génomique et sur l'activité du génome.

Pour ce faire, nous avons dans un premier temps développé un nouveau modèle murin exprimant H3.3 étiqueté et muté (K27M ou G34R) de façon conditionnelle sur un seul allèle du gène H3f3a. La conditionnalité de la mutation est nécessaire pour mimer le caractère somatique de la mutation et l'étiquette donne la possibilité de différencier H3.3 muté des copies sauvages et de pouvoir identifier sa dynamique d'incorporation au sein de la chromatine. La première partie de ma thèse présente la stratégie du développement du modèle murin ainsi que sa réalisation. Cependant, la conditionnalité de la stratégie s'est avérée défectueuse étant donné que la protéine mutante était déjà exprimée de façon constitutive dans les cellules souches embryonnaires (mESC) générées dans le cadre de ce projet. Par conséquent, nous avons décidé de modifier notre stratégie et d'utiliser les mESC contenant une forme étiquetée de H3.3 sauvage ou mutée comme modèle pour cette étude.

Après avoir vérifié que l'insertion des mutations ne perturbe ni le niveau de transcription du gène ni la quantité de protéine produite, nous avons mesuré l'impact des mutations de H3.3 sur son incorporation au sein de la chromatine. Pour cela, la distribution chromatinienne de H3.3 a été analysée par des expériences de ChIP-seq

(immuno-purification de la chromatine suivie de séquençage haut-débit). Mes données montrent que H3.3 est enrichit aux sites d'initiation de la transcription (TSS) et que son niveau d'enrichissement suit l'expression des gènes. H3.3 est également retrouvé au niveau des enhancers, régions définies par la présence de H3K4me1 et/ou H3K27ac et l'absence de H3K4me3 à une distance des promoteurs supérieure à 2 kb. Les données obtenues avec les formes mutées de H3.3 révèlent que les mutations K27M et G34R n'affectent pas son incorporation au sein de la chromatine active. En effet, les mutants des H3.3 marquent les mêmes régions que la forme sauvage avec des enrichissements équivalents. L'histone H3.3 marquant les promoteurs des gènes actifs, nous avons dans un second temps mesurer l'impact des mutations de H3.3 sur l'expression du génome par la technique de RNA-seq (séquençage haut-débit des ARN totaux). Une dérégulation modérée des gènes a été observée en présence de H3.3 K27M ou G34R (~ 200 et ~1150 gènes respectivement, dont 90 % avec un |log2FC|<2). L'analyse intégrée des données de ChIP-seq et de RNA-seq ne révèle aucun lien entre la densité en H3.3 au niveau des promoteurs et leur niveau de dérégulation. Nous concluons que la dérégulation transcriptionnelle observée n'est pas la conséquence d'un défaut d'incorporation et/ou d'éviction des formes mutées de H3.3 au sein des promoteurs.

De précédents travaux ont suggéré la présence de H3.3 au niveau de régions d'hétérochromatine et notamment de familles spécifiques de séquences répétées où elle jouerait un rôle répressif (Elsässer et al., 2015). Nous avons analysé l'enrichissement en H3.3 au niveau des séquences répétées, et montré que H3.3 est spécifiquement localisée au niveau des familles de rétrovirus endogènes (ERV) récemment intégrés dans le génome murin (qui correspondent aux retrovirus fonctionnels et potentiellement actifs). De façon similaire à ce qui a été observé au niveau de la chromatine active, notre étude révèle que les mutations de H3.3 n'ont pas d'incidence sur son incorporation au sein des régions hétérochromatiniennes. Cependant, les ERVs marqués par H3.3 sont spécifiquement surexprimés en présence des mutations K27M et G34R. Selon la littérature, les rétrovirus fonctionnels sont réprimés par différents mécanismes dans les cellules souches, notamment le système KRAB-ZFP/KAP1, la méthylation de l'ADN ou encore la déposition de la marque H3K9me3 (Ecco et al., 2017).

Afin de comprendre le mécanisme causant la surexpression des ERV en présence de H3.3 mutée, les complexes protéiques associées aux nucléosomes contenant H3.3 sauvage ou mutée ont été purifiées par double-immunoprécipitation et analysés par spectrométrie de masse. Le complexe répresseur associé à KAP1, les chaperonnes spécifiques de H3.3 ainsi que le complexe NuRD/HDAC interagissent avec H3.3. L'abondance de certains membres de ces complexes semble plus faible pour H3.3 mutée que pour H3.3 sauvage, notamment les enzymes à activité DNA-methyltransférases DNMT1 et DNMT3a ainsi que la protéine SETDB1 (responsable de la déposition de H3K9me3). Cela suggère un défaut de méthylation de l'ADN et/ou de déposition de la marque H3K9me3 au niveau des rétrovirus endogènes. L'analyse de la distribution génomique des marques 5-méthylcytosine, 5-hydroxyméthylcytosine et H3K9me3 au niveau des séquences répétées permettra de déterminer si la surexpression des ERV est due à la perte d'une ou de plusieurs de ces marques répressives. De précédents travaux ont montré que la machinerie répressive associée à KAP1 est recrutée par des protéines à doigt de zinc KRAB spécifiques pour la reconnaissance de certaines familles de séquences répétées. Une régulation négative de plusieurs protéines à doigt de zinc KRAB (KRAB-ZNP) connue pour se lier aux ERV récemment intégrés a été observée spécifiquement en présence de H3.3 mutée (e.g. Gm15446 et Zfp932). Cette baisse d'expression pourrait également être responsable de la surexpression des ERV en présence de H3.3 mutée.

Plusieurs études ont montré que les rétrovirus endogènes peuvent influés sur l'expression des gènes situés à leur proximité (Karimi et al., 2011 ; Jang et al., 2019). Nous nous sommes donc demandés si les ERV surexprimés sont situés à proximités des gènes dérégulés. Notre analyse statistique montre que les gènes qui sont physiquement proches des ERV surexprimés (⊠ 10 kb) sont d'avantage dérégulés. En résumé, H3.3 est enrichit au niveau des ERV fonctionnels. Ces derniers sont surexprimés en présence des mutants et sont situés à proximités des gènes dérégulés.

Le lien direct entre H3.3 et la surexpression des ERV a été confirmé cliniquement. En effet, certaines familles de rétrovirus endogènes humains sont surexprimées en présence de H3.3 K27M ou G34R dans les tumeurs ou lignées dérivées de patients. De plus, en ré-analysant des données publiées par Krug et al. (2019), nous avons montré que la régulation négative de H3.3 K27M (mais pas de

H3.3 sauvage) dans des xénogreffes dérivées de patient, est suffisante pour ralentir le développement de la tumeur. De manière intéressante, ce phénotype est associé à une répression des rétrovirus endogènes fonctionnels, suggérant une implication de ce mécanisme dans la pathogénicité des gliomes pédiatriques de haut-grade.

L'étude de la dynamique d'expression des ERV pendant la différentiation neurale dans un modèle in vivo serait d'un grand intérêt. Malgré l'émergence de certains modèles murins de gliomes pédiatriques de haut-grade K27M, aucun modèle n'a été développé à ce jour pour la mutation G34R/V. Afin de faire le lien entre nos résultats obtenus à partir de cellules souches et ceux obtenus dans les tissus de patients, j'ai analysé la capacité des mESC à se différencier en cellules souches neuronales (NSC), en utilisant un protocole expérimental caractérisé par une étape de différentiation intermédiaire appelée corps embryonnaires (embryonic bodies, EB). Alors que les cellules souches exprimant la forme sauvage de H3.3 ont été différenciées avec succès en NSC, des défauts de différentiation ont pu être observés pour les mESC exprimant H3.3 K27M ou G34R. L'analyse de la dynamique de transcription au cours de la différenciation montre que les gènes dérégulés en présence des mutations de H3.3 sont impliqués dans les processus de différentiation et de neurogénèse. Par ailleurs, alors que les éléments répétés exprimés dans les cellules souches s'éteignent rapidement au cours de la différentiation, les ERV enrichis en H3.3 échappent à cette règle, suggérant que la fonction essentielle de H3.3 au cours de la différentiation est étroitement liée au contrôle de l'expression des éléments transposables. En accord avec cette hypothèse, nous observons une dérégulation spécifique des ERVs marqués par H3.3 au cours de la différentiation lorsque H3.3 est mutée.

En conclusion, mon travail montre que H3.3 marque les rétrovirus endogènes fonctionnels, et que ces rétroéléments ont une dynamique d'expression spécifique dans notre modèle de différentiation neuronale in vitro. Les mutations de H3.3 provoquent une dérégulation globale des rétrotransposons, qui aboutit à un défaut de différentiation. Le lien direct entre H3.3 et les ERVs a été confirmé cliniquement, ce qui suggère que ce mécanisme est impliqué dans la pathogénicité des gliomes pédiatriques de haut-grade.

Table of contents

List of Abbreviations

ACVR1	ACtiVin A Receptor, type 1
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
ATP	Adenosine TriPhosphate
aHGG	adult High-Grade Glioma
ATRX	Alpha-Thalassemia/mental Retardation syndrome, X-linked
AutoN	Autoinhibitory N-terminal
BER	Base Excision Repair
BSA	Bovine Serum Albumin
C-ter	C-terminal
Cabin1	Calcineurin-binding protein 1
CNS	Central Nervous System
CENP-A	CENtromeric Protein A
CAF-1	Chromatin Assembly Factor 1
ChIP-seq	Chromatin Immunoprecipitation - sequencing
CHD	CHromodomain-helicase-DNA-binding protein
cds	coding DNA sequence
cDNA	complementary DeoxyRibonucleic Acid
cKI	conditional Knock-In
CCND	Cyclin D1
CDK	Cyclin-Dependent Kinase
DAXX	Death domain-Associated protein 6
DNA	DeoxyRibonucleic Acid
DIPG	Diffuse Intrinsic Pontine Glioma
DBD	DNA-Binding Domain
DNMT	DNA-MethylTransferase
EB	Embryonic Bodies
ES	Embryonic Stem Cell
ESC	Embryonic Stem Cell
ERV	Endogenous RetroVirus
EZH1/2	Enhancer of Zeste Homolog 1 or 2
EGFR	Epidermal Growth Factor Receptor
EMT	Epithelial-Mesenchymal Transition
FBS	Fetal Bovine Serum
FCS	Fetal Calf Serum
FGF	Fibroblast Growth Factor
HSS	HAND-SANT-SLIDE
HAS	Helicase/SANT-ASsociated
HP1	Heterochromatin Protein 1

HGG	High-Grade Glioma
HAT	Histone AcetylTransferase
HIRA	Histone cell cycle RegulAtor 1
HDAC	Histone DeAcetylTransferase
HKDM	Histone Lysine DeMethylase
KDM4	Histone Lysine DeMethylase 4
HKMT	Histone Lysine MethylTransferase
HRP	HorseRadish Peroxidase
ISWI	Imitation SWItch
IgG	Immunoglobulin G
IP	ImmunoPrecipitation
INO80	Inositol requiring 80
IARC	International Agency for Research on Cancer
IICC	International Incidence of Childhood Cancer
IAP	Intracisternal A-Particle
IDH	Isocitrate DeHydrogenase
KSR	Knockout-Serum Replacement
KAP1	KRAB-Associated Protein 1
KRAB-ZFP	KRüpell-Associated Box- containing Zinc Finger Protein
LIF	Leukemia Inhibitory Factor
LTR	Log Terminal Repeats
LINE	Long Interspersed Nuclear Elements
LGG	Low-Grade Glioma
MRI	Magnetic resonance imaging
mRNA	messenger RiboNucleic Acid
MBD	Methyl-CpG-binding domain protein
MMR	MisMatch Repair
MEF	Mouse Embryonic Fibroblasts
mESC	mouse Embryonic Stem Cell
MYCN	N-myc proto-oncogene protein
N-ter	N-terminal
NegC	Negative regulator of Coupling
NSC	Neural Stem Cell
NF2	NeuroFibromatosis type II
NuRD/HDAC	Nucleosome Remodeling and Deacetylase/Histone DeACetylase
MGMT	O-MethylGuanine-DNA-MethylTransferase
ORF	Open Reading Frame
PDX	Patient Derived Xenograft
pHGG	pediatric High-Grade Glioma
pNBS	pediatric Non-BrainStem
PI3K	PhospatidylInositide 3-Kinase
PTEN	Phosphatase and TENsin homolog
PBS	Phosphate Buffer Saline

PHD	Plant HomeoDomain
PDGFRA	Platelet-Derived Growth Factor Receptor Alpha
PDGFB	Platelet-Derived Growth Factor subunit B
PRC2	Polycomb Repressive Complex 2
PVDF	PolyVinyliDene Fluoride
PTM	Post-Translational Modification
PML	ProMyelocytic Leukemia
PRMT	Protein Arginine MethylTransferase
RPM	Reads Per Million
RTK	Receptor Tyrosine Kinase
RB	RetinoBlastoma
RT-PCR	Reverse Transcription - Polymerase Chain Reaction
RNA	RiboNucleic Acid
rRNA	ribosomal RiboNucleic Acid
RT	Room Temperature
B-RAF	Serine/threonine-protein kinase B-raf
SETDB1	SET Domain Bifurcated Histone Lysine Methyltransferase 1
SETD2	SET Domain Containing 2, Histone Lysine Methyltransferase
shRNA	short hairpin RNA
SINE	Short Interspersed Nuclear Elements
SVA	SINE-VNTR-Alus
SnAC	Snf2 ATP coupling
SDS-PAGE	Sodium Dodecyl Sulfate–PolyAcrylamide Gel Electrophoresis
SET	Su(var)3-9, Enhancer-of-zeste and Trithorax
SWI/SNF	SWItch/Sucrose Non-Fermentable
SNS	Sympathetic Nervous System
TET	Ten Eleven Translocase
TDG	Thymine DNA Glycosylase
TSS	Transcriptional Start Site
TGF- β	Transforming Growth Factor-beta
TE	Transposable Elements
TP53	Tumor Protein 53
UBN1/2	UBiNuclein 1/2
UTR	Untranslated Transcribed Region
VEGF	Vascular Endothelial Growth Factor
WT	wildtype
WT1	Wilms Tumor 1
WHO	World Health Organization

List of Figures and Tables

I. Introduction

Chapter 1: Pediatric high-grade glioma

1.1. Physiological bases of tumorigenesis

Tumors can arise in any tissue or organ of the body and have the potential to destroy a healthy tissue. The conversion of a healthy into a tumor cell is due to a succession of genetic alterations providing new properties to the cell and leading to the gradual tumor conversion. These alterations can result in oncogenes expression (gain of function) or in tumor suppressor genes inactivation (loss of function). The current knowledge about cancer biology is far from the original picture of cancer seen as a homogeneous group of cells growing out-of-control. Hanahan and Weinberg have defined ten hallmarks of cancer (**Figure 1**) as acquired functional capabilities allowing cancer cells to survive, proliferate and disseminate (Hanahan and Weinberg, 2011).

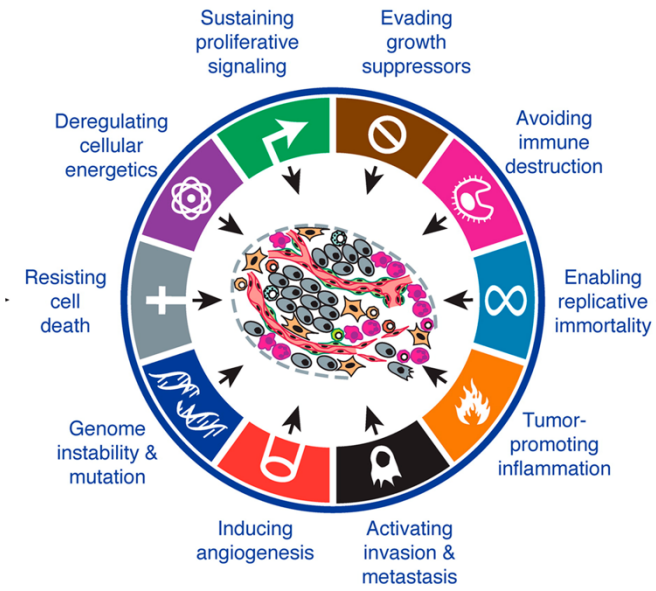


Figure 1: The Hallmarks from Cancer (adapted from Hanahan and Weinberg, 2011)

Each hallmark is overviewed below. The combination of several of these hallmarks allows cancer cells to disseminate.

1.1.1. Sustaining proliferative signaling

The ability to sustain proliferation in undoubtedly one of the most fundamental trait of cancer cells. Several alternative ways are used by cancer cells to maintain a chronic proliferation ability. Cancer cells are able to produce growth factors (autocrine proliferative stimulation) or to stimulate normal cells for growth factor production. Somatic mutations can also lead to growth factor independence by constitutive activation of downstream pathways (e.g. B-raf or PI3-kinase mutations; Davies and Samuels, 2010; Yuan and Cantley, 2008). Another way of enhancing proliferation is to disrupt negative-feedback mechanisms aimed at attenuating proliferative signaling (e.g. loss of PTEN expression; Jiang and Liu, 2009; Yuan and Cantley, 2008).

1.1.2. Evading growth suppressors

In addition to boosting proliferation, cancer cells circumvent programs that negatively regulate cell proliferation by several strategies. One strategy is to inactivate tumor suppressor genes (e.g. RB and p53) which operate as gatekeeper of cell-cycle progression and as central control nodes in the decision between cell proliferation and senescence or apoptotic programs activation (Wang et al., 2018a). Several studies have demonstrated the ability of normal cells to stop proliferating once they reach confluence. Hence, another strategy of cancer cell is to evade this “contact inhibition”, for example with the loss of NF2 (Curto et al., 2007).

1.1.3. Evading immune destruction and tumor-promoting inflammation

The role of the immune system in resisting or eradicating formation and progression of tumors is still an unresolved issue. Several of the immune system components can either eradicate cancer cells or promote their proliferation. Some cancer cells seem to evade immune destruction by impairing components of the immune system (e.g. disablement of cytotoxic T lymphocytes and natural killer by TGF-β secretion; Yang et al., 2010). In some other cases, inflammation is capable of promoting the development of incipient neoplasia into full-blown cancers through the release of highly mutagenic chemicals (e.g. reactive oxygen species) by inflammatory cells (Qian and Pollard, 2010).

1.1.4. Enabling replicative immortality

Telomeres are protecting the ends of the chromosomes and can be compared as a clocking device which determines the limited replicative potential of normal cells. They consist thus in a limit to overcome for cancer cells (Hanahan and Weinberg, 2011). Telomerase, the DNA polymerase in charge of the telomere addition at the ends of telomeric DNA, is almost absent in non-immortalized cells. Normal cell aging leads to a progressive truncation of telomeres ends, finally activating apoptosis or senescence processes. However, telomerase is more strongly expressed in cancer cells providing them the capability for unlimited proliferation. The acquisition of the telomerase function is often delayed in tumor progression. Cancer cells undergo first a phase of telomeres shortening with genomic instability and generation of tumor-promoting mutations. The latter are then stabilized by unlimited replication capacity thanks to the telomerase activation.

1.1.5. Activating invasion & metastasis

Metastasis has long been described as the final step in tumor progression but there are evidences showing that cells can disseminate at earlier steps. (Hu et al., 2017). Metastasis can be divided in two main phases: the physical dissemination of cancer cells from the primary tumor followed by the adaptation to foreign tissue microenvironments. A well-studied example of invasion potentiator is the loss of E-cadherin, a key cell-to-cell junction molecule (Berx and van Roy, 2009). The “epithelial-mesenchymal transition” (EMT) is a developmental regulatory program which has been implicated as a means by which transformed cells become invaders, apoptotic-resistant and capable of dissemination (Derynck and Weinberg, 2019). This invasion program is plastic and metastases may use the reverse process (mesenchymal-epithelial transition) resulting in the formation of new tumor colonies.

1.1.6. Inducing angiogenesis

To sustain their proliferation rate, tumors need high quantities of nutrients and oxygen as well as a way to evacuate metabolic wastes. For this reason, tumors promote angiogenesis in order to develop an associated neovasculature. Upregulation of angiogenesis activators like VEGF or FGF has been described in several tumors (Ferrara, 2009; Baeriswyl and Christofori, 2009). Moreover, angiogenesis has been

shown to be induced early during the multistage development of invasive cancers (Raica et al., 2009).

1.1.7. Genome instability & mutation

Acquisition of most of the described hallmarks depends mainly on tumor cell genomic alterations. The tumor progression is often described as a multistep process with sequential clonal expansions and enabling mutant genotypes. p53 has a central role, as its loss leads to a compromised surveillance of genomic integrity. p53 is indeed part of a bigger family named the “caretakers” of the genome (Kinzler and Vogelstein, 1997). Defects in caretakers can result in several impairments: in DNA damage detection and activation of the repair machinery, in the recruitment of the repair machinery itself or in interception of the mutagenic molecules before any DNA damage occurs. Loss of caretaker’s function occurs during tumor progression through inactivating mutations or epigenetic repression. As described previously, loss of telomeric DNA also generates karyotypic instability and chromosomal aberration in tumors (Artandi and DePinho, 2010).

1.1.8. Resisting cell death

Tumor cells have developed many strategies to limit or prevent apoptosis. One of the most studied strategy is the loss of p53 tumor suppressor which has a main role in sensing critical damages and activating apoptosis circuitry. Another strategy consists in the downregulation of proapoptotic factors leading to the overexpression of antiapoptotic regulators or survival signals. Besides, autophagy may have a dual role in cancer development, causing either tumor cell death or survival depending on the conditions. Nutrient starvation, radiotherapy, and certain drugs can indeed induce elevated levels of autophagy which seem to protect the tumor cells instead of killing them (White and DiPaola, 2009; Apel et al., 2009). Tumors have also been shown to tolerate some degree of necrotic cell death which helps recruiting tumor-promoting inflammatory cells (Grivennikov et al., 2010).

1.1.9. Deregulating cellular energetics

Energy is the fuel for proliferation, thus a fundamental need for cancer cells. Normal cells growing under aerobic conditions process glucose into pyruvate through glycolysis in the cytosol. Pyruvate is then further transformed in carbon dioxide in the

mitochondria. Under anaerobic conditions, glycolysis is favored with almost no pyruvate forwarded to the mitochondria. Even in aerobic conditions, cancer cells use almost exclusively glycolysis (known as “aerobic glycolysis” – the Warburg effect), without sending pyruvate toward mitochondria. To compensate the lower efficiency of energy production by glycolysis compared to mitochondrial oxidative phosphorylation, cancer cells increase glucose uptake by upregulating glucose transporters (e.g. GLUT1; Jones and Thompson, 2009). The potential advantage of favoring glycolysis might be the diversion of glycolytic intermediates into pathways for generation of nucleosides and amino acids, facilitating in turn the synthesis of macromolecules and organelles, thus the assembling of new cells.

The above described hallmarks are a general view of the different capabilities that cancer cells can develop to survive, proliferate and disseminate. Targeting those capabilities is of great interest for therapeutic development. However, the affected and modified pathways are under a delicate equilibrium and pathway redundancy makes mechanism-based targeted therapies challenging.

1.2. Epidemiology of pediatric cancers

Childhood cancers are rare and represent about 1 % of all cancers diagnosed each year in the world. Cancer constitutes the second leading cause of death for children and adolescents after domestic deaths in developed countries and about 1 in 7,000 children is diagnosed with cancer each year (Saletta et al., 2014). The International Agency for Research on Cancer (IARC) has led a worldwide collaborative project to determine incidence of cancer in children: International Incidence of Childhood Cancer (IICC). The two first studies have been published in 1988 and reported the worldwide incidence of cancer in children aged 0-14 years in the 1970s and 1980s, IICC-1 and IICC-2 respectively. In 2017, IARC published a third study covering the worldwide incidence of cancer in children aged 0-19 years over the period 2000-2010, IICC-3. The top three most common pediatric cancer types are leukemia, central nervous system (CNS) tumors and lymphomas and constitute more than 60 % of the cases (**Figure 2**).

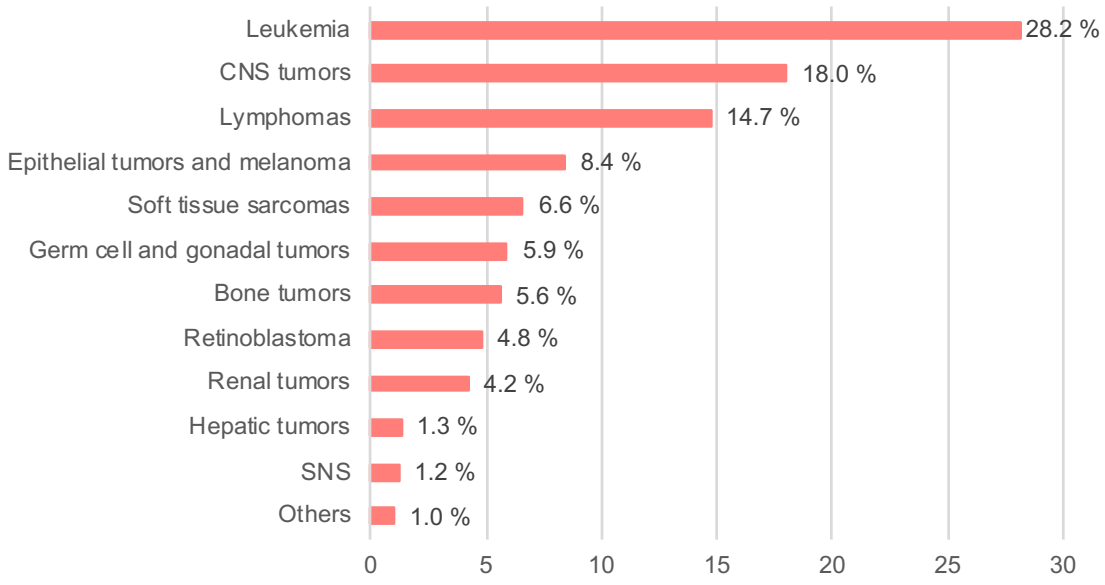


Figure 2: Worldwide distribution of cancer types among children aged 0-19 years, 2001-2010

According to the IICC-3 study and Steliarova-Foucher et al., 2017. Number of total referenced cases: 381,137. CNS: Central Nervous System, SNS: Sympathetic Nervous System

The range of tumor types varies with age. For instance, leukemia is more prevalent in younger children (0-4 years) while lymphoma prevalence is more prevalent in young adults (15-19 years). In addition, incidence rates as well as survival rates follow considerable geographic variations (Kaatsch, 2010). With a global 5-year survival rate having reached 80 %, four out of five diagnosed children can be cured of cancer with current therapies. However, there remain several pediatric cancers for which no effective treatment is available, for instance pediatric high-grade glioma. A better understanding of the genetics of childhood cancer, together with targeted therapeutics development constitute a priority and a challenge to overcome the remaining incurable pediatric cancers.

1.3. Pediatric High-Grade Glioma

1.3.1. Definition and location

Normal brain is composed of several types of cells including neurons, the functional unit of the nervous system, and glia which play a supportive role to the neurons. Glia is further subdivided in different cell types harboring different functions, for instance astrocytes which perform functions such as recycling the excess of neurotransmitters or creating the blood-brain barrier and oligodendrocytes which cover the axons of neurons. Gliomas are brain tumors arising from glial cells, particularly astrocytes, oligodendrocytes or their precursors. Gliomas are ranked from low-grade gliomas (LGG, grades I and II) to high-grade gliomas (HGG, grades III and IV) according to the World Health Organization (WHO) classification. While adult HGG arise predominantly in the cerebral cortex, pediatric HGG (pHGG) have a broader spectrum of locations (Wu et al., 2014). pHGG can further be divided according to their brain location in diffusive intrinsic pontine glioma (DIPG) and in pediatric non-brainstem (pNBS) HGGs. As indicated by its name, DIPG occurs in the brainstem. pNBS-HGGs can arise either in midline structures (thalamus or cerebellum) or in the cortex hemispheres (**Figure 3**).

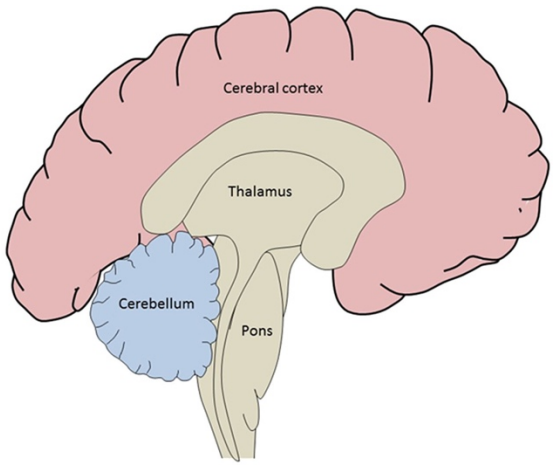


Figure 3: Anatomic representation of the brain structures from which a pHGG can originate (modified from Juratli et al., 2018)

1.3.2. Epidemiology and prevalence

Gliomas are the most common pediatric brain tumor and represent the first cause of cancer-related deaths in children (<14 years old). pHGG have an incidence rate of 3.51 per 100,000 and constitute about 7 % of all pediatric CNS tumors (Ostrom et al., 2018). With five-year survival rates of less than 20 % for pHGG and about 2 % for DIPG (Braunstein et al., 2017; Hoffman et al., 2018), pHGG show a median overall survival of 9-15 months (Jones et al., 2012). The epidemiology statistics are slightly changing depending the studies considered but they clearly highlight the pHGG aggressiveness and the necessity of a proper treatment for those untreatable cancers.

1.3.3. Diagnostic

The signs and symptoms of pHGG greatly vary depending on patient age, tumor location and aggressiveness. Impairment of recent memory, persistent headache awakening children during night, nausea and vomiting, irritability or change in feeding pattern, are examples of symptoms that should be considered as potential signs of brain tumors.

1.3.3.1. Magnetic resonance imaging

Magnetic resonance imaging (MRI) is an essential tool in the diagnosis of brain tumor. While performing computerized tomography is usually quicker than MRI for facility accessibility reasons, MRI is gradually replacing computerized tomography in children diagnosis in order to minimize radiation exposure. Unlike computerized tomography, MRI doesn't use X-rays but rather variation of strong magnetic fields to generate images of the organs. In addition, MRI provides higher sensitivity in differentiating tumor from normal brain tissue, especially for the brainstem and cerebellum (Braunstein et al., 2017). Areas of high density and which enhances with contrast correspond to actively dividing regions and are often sign of proliferating tumor cells. The use of MRI is thus essential in brain tumor diagnosis and give insights in the location, size, density, shape and borders of the tumor (**Figure 4**).

Figure 4: MRI sagittal view of a brainstem tumor from a 9-years-old female (shown by a red arrow, from Nazarian et al., 2016). T1 sequence, tumor appearing in hyposignal.

1.3.3.2. Biopsy

After MRI, a biopsy of the tumor is performed in order to classify the tumor and to select the most suitable treatment. When maximal safe surgical resection is possible, the biopsy is performed during the removal surgery. In case of DIPG or other invasive tumors, safe surgical resection is often impossible and a stereotactic needle biopsy is then performed. The biopsy is essential for tumor classification through histopathology and molecular classification according to the 2016 WHO classification of tumors of the CNS.

1.3.3.3. Classification (World Health Organization)

In 2016, the WHO has performed a major change in CNS tumors classification. Previously only classified by histology and malignant grade, CNS tumors are now also classified according to specific molecular characteristics. This change enables to better separate adult and pediatric diffuse gliomas that were grouped together due to their histological similarities while behaving in a very different way (Louis et al., 2016).

1.3.3.3.1. Histopathology

The classification starts with a typing and grading step through the histomorphology characterization of the biopsy. Together with the growth pattern (from MRI images) and the lineage determination, the tumor grade is defined according to the nuclear atypia, the cellular polymorphism, the number of mitoses, the micro-

vasculature and the presence of necrosis (**Figure 5**). With this first histomorphology step, the tumor can be pre-classified as WHO grade II-III or WHO grade IV. Due to the high histological similarities but strong differences in behavior, glial tumors need an additional molecular classification.

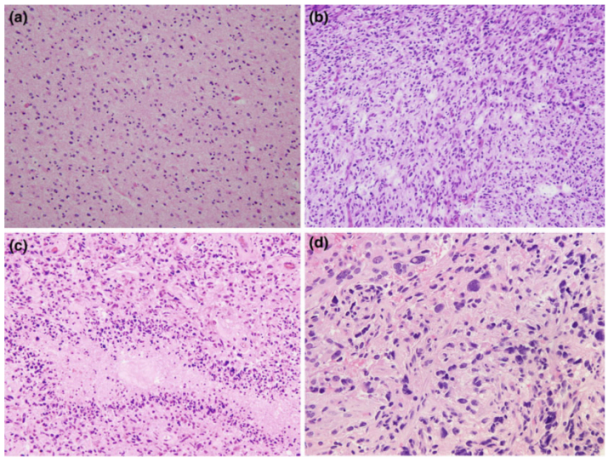


Figure 5: Morphologic appearance of gliomas (sections stained with hematoxylin and eosin), from Huttner, 2017.

a. Diffuse astrocytoma, WHO grade II, characterized by low cellularity and mild nuclear pleomorphism (variability in size and shape); b. Anaplastic astrocytoma, WHO grade III with increased cellularity and anaplastic nuclei; c. Glioblastoma, WHO grade IV with pleomorphic tumor cells, mitoses, and pseudo-palisading necrosis; d. Diffuse midline glioma with a high degree of pleomorphism.

1.3.3.3.2. Molecular classification

A molecular testing is performed on the tumor biopsy to identify several genetic alterations such as mutation of isocitrate dehydrogenase (IDH) or alpha-thalassemia/mental retardation syndrome, X-linked (ATRX), the chromosome arms 1p19q codeletion or the H3-K27M mutation (**Figure 6**). Additional molecular alterations are usually searched to have a more detailed view on the tumor landscape and a better clinical characterization. The rapid advances and cost reductions of high-throughput sequencing solutions lead to a progressive generalization of genome-wide identification of molecular alterations.

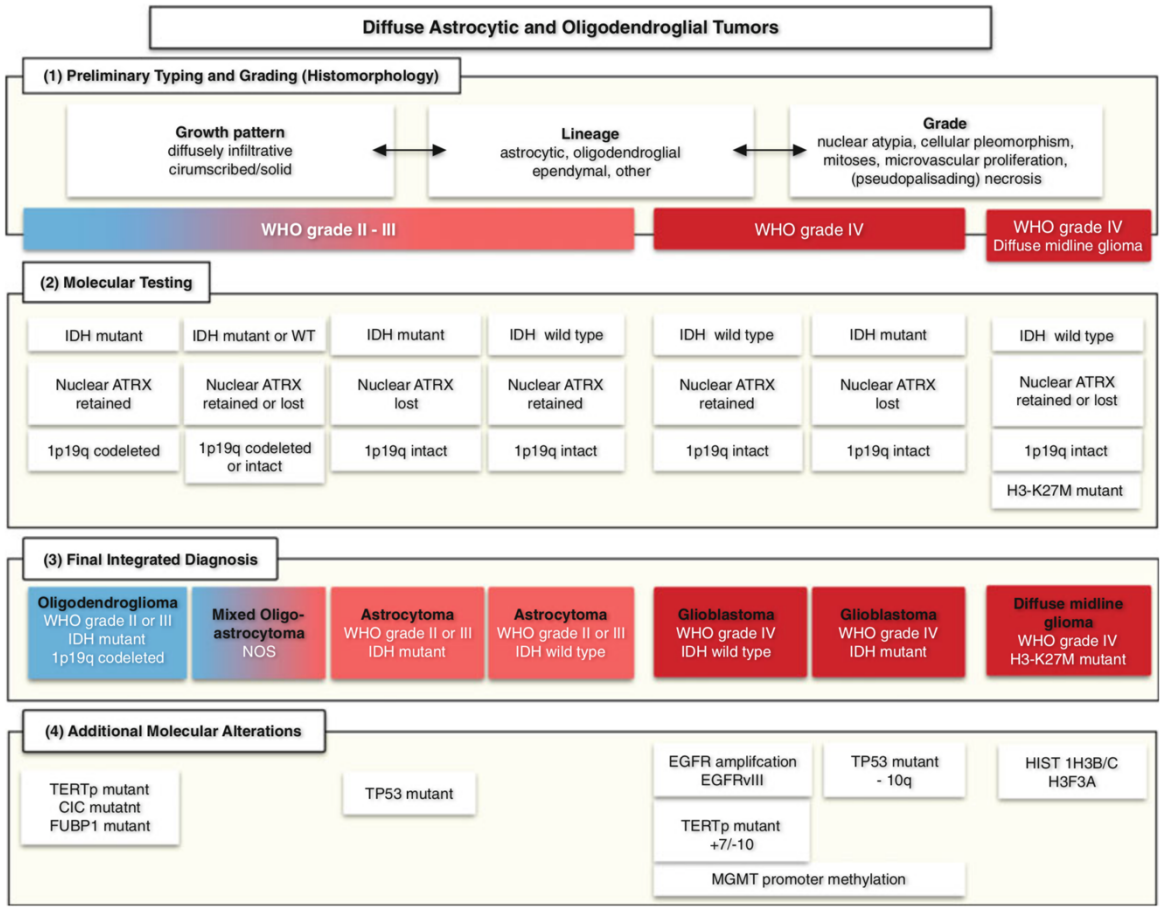


Figure 6: Example of the steps for the diagnosis of diffuse gliomas according to the 2016 WHO classification (Huttner, 2017).

1.3.4. Genetic alteration linked to pHGG

1.3.4.1. Main somatic mutations in pHGG

Adult HGG (aHGG) are characterized by the disruption of three core pathways, namely the receptor kinase-Ras-phosphatidylinositide 3-kinase (RTK-RAS-PI3K), the p53 and the retinoblastoma (RB) networks. pHGG also show disruption of one or more of those pathways, but with different effectors altered. For instance, the epidermal growth factor receptor (EGFR) is the most commonly altered receptor tyrosine kinase in aHGG but is rarely altered in pHGG. On the other hand, the platelet-derived growth factor receptor alpha (PDGFRA) has been found to be the most commonly altered receptor tyrosine kinase in pHGG (Diaz and Baker, 2014). Both EGFR and PDGFRA are members of the PI3K cascade and their amplification leads to PI3K over-activation

in HGG, thus sustaining proliferative signaling (see 1.1.1). Alterations of other members of the RTK-RAS-PI3K pathway have been found in pHGG, like MET, NTRK1/2/3, PIK3CA, PIK3R1, BRAF, Akt or NF1 (Wu et al., 2014, Pollack et al., 2010).

The RB pathway is commonly dysregulated in both aHGG and pHGG. Amplification in CCND1/2/3, CDK4/6 is found in about 14 % of pHGG (with a prevalence in DIPG) and leads to cell cycle progression. Besides, CDKN2A/CDKN2B locus encoding CDK4/6 inhibitors are specifically deleted in about 30 % of pNBS-HGG and cause loss of cell cycle checkpoints (Wu et al., 2014, Diaz and Baker, 2014). While the p53 pathway was found to be altered in 85 % of aHGG (Brennan et al., 2013), only 42 % of pHGG had p53 mutated (Wu et al., 2014).

Taken together, pHGG have been found to harbor fewer aberrations of RTK-RAS-PI3K/p53/RB pathways than aHGG, with a proportion of pHGG showing a “stable genome” devoid of copy number alterations (Bax et al., 2010). Recent studies have described specific alterations of actors in chromatin and transcriptional regulation in pHGG.

1.3.4.2. Histone H3.3 and its mutations

In 2012, a unique feature of pHGG has emerged with the discovery of specific point mutations in histone H3 family members (Schwartzentruber et al., 2012; Wu et al., 2012), namely K27M or G34R/V mutations. Those mutations are somatic and occur on a single allele (dominant negative effect). They affect mainly histone variant H3.3 (and more rarely H3.1), and arise at high frequency in pHGG. Moreover, they show distinct and specific tumor location and age span of development. All those attributes have led to the mutant histones being considered as potential driver of pHGG and have raised a great interest in understanding their underlying biology. In Chapter 2, we will develop the current knowledge about H3 family members and the implication of their mutation in pHGG development.

1.3.5. Treatment

Current treatments for pHGG include surgery (when possible), followed by radiation therapy combined or not with chemotherapy but the outcome remains dismal. Surgery is especially challenging, if not impossible, in DIPG. Radiation therapy is currently considered as a palliative treatment and only allows to increase the survival by several months. For many years, patient with pHGG were treated following aHGG treatment programs without convincing results. Despite several hundreds of clinical trials, no efficient treatment is yet available and pHGG prognosis remains really poor (Bailey et al., 2018; Lapin et al., 2017). This is not surprising considering the important molecular differences between pHGG and aHGG. Another limitation to overcome is the drug limited delivery to the brain when given systemically due to the blood-brain barrier.

Thus, the current challenges consist in developing drug delivery methods and in a better understanding of the biology of pHGG in order to develop targeted and efficient therapies.

Chapter 2: H3.3, a key player in pHGG development

2.1. DNA organization in chromatin

The eukaryotic genome is packaged in the nucleus into a structure composed of DNA associated to proteins and called chromatin. With a structural and functional role, chromatin enables to compact our genome of 3 billion base pairs in the nucleus. Each of our cell can thus pack about 2 meters of DNA in a compartment ~10 µm in diameter. This structure is not static but highly dynamic and its variations convey the epigenetic information defining the identity of each cell in the organism.

2.1.1. Structure of chromatin

2.1.1.1. History

The term ‘chromatin’ (from the Greek ‘khroma’ which stands for ‘colored’) has been introduced by Walther Flemming at the end of the 19th century due to its ability to retain colorants. Chromatin is a nucleoprotein composed of DNA, histones and non-histone proteins and is the guardian of genetic information. It’s only one century after the discovery of histones that the fundamental unit of chromatin, the nucleosome, has been described (**Figure 7**). From the second half of the 19th century onward, intense structural and functional studies were performed on chromatin.

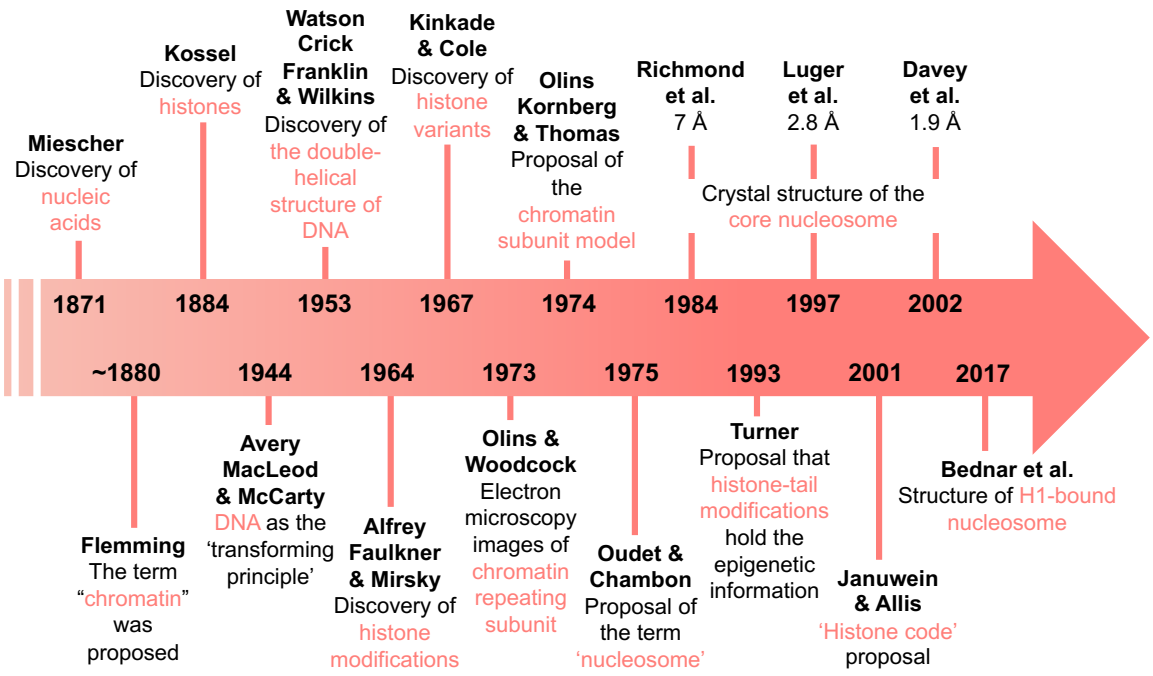


Figure 7: Main scientific advances in chromatin history (adapted from Olins & Olins, 2003)

Starting in 1944, DNA has been identified as the genetic information keeper (Avery et al., 1944). With the help of Franklin's work, Watson and Crick discovered the DNA double-helical structure a few years later (Watson and Crick, 1953). In 1973, the 'beads on a string' structure of chromatin was imaged for the first time by Olins & Woodcock (**Figure 8**), followed by the proposal of the term 'nucleosome' as the fundamental unit of chromatin repetitive structure (Kornberg, 1974; Oudet et al., 1975).

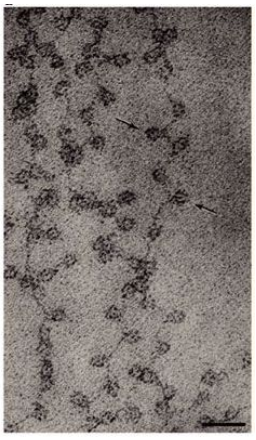


Figure 8: Electron-microscopy image of chromatin spread giving rise to the 'bead on a string' model (Olins & Olins, 2003)

2.1.1.2. The nucleosome, core unit of chromatin

The most basic level of chromatin is the nucleosome which consist of a histone octamer wrapped inside 146 base pairs (bp) of DNA. The histone octamer comprises two copies of the four core histones H2A, H2B, H3 and H4 (**Figure 9**). The nucleosome forms a cylinder with a diameter of 11 nm and 5.5 nm in height and is the first level of DNA compaction.

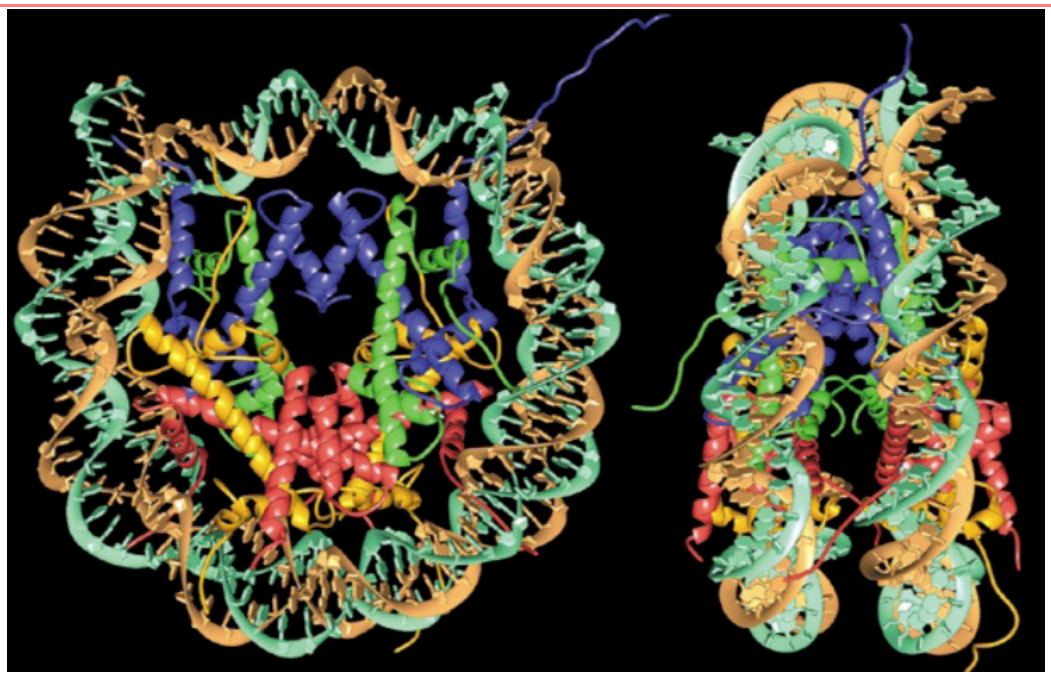


Figure 9: Crystallographic structure of the core nucleosome at 2.8 Å (adapted from Luger et al., 1997).

146-bp DNA phosphodiester backbones (brown and turquoise) and eight histone proteins (blue: H3; green: H4; yellow: H2A; red: H2B). View down the DNA superhelix axis on the left and perpendicular to it on the right.

2.1.1.3. Core histones

Histones are evolutionarily conserved DNA-binding proteins. Considered as scaffolding molecules, they take part in DNA packaging regulation into the nucleus of eukaryotic cells. Also used as docking units, they play a role in the recruitment of the transcriptional machinery. Canonical histones H2A, H2B, H3 and H4 are small proteins with a molecular weight comprised between 11 and 20 kDa. Their primary structure is very rich in basic amino acids. Histones are composed of two structural and functional distinct domains: the N-terminal (N-ter) tail and the histone fold domain. The histone fold domain is a highly conserved globular domain and is involved in dimerization of

histones H2A-H2B and H3-H4 through a 'handshake' motif (**Figure 10**). In physiological conditions and in the absence of DNA, two H3-H4 dimers associate and form a $(H3\text{-}H4)^2$ tetramer on which two H2A-H2B dimers will bind to form the canonical histone octamer. The histone N-ter tail is unstructured and is floating outside of the nucleosome (Khorasanizadeh, 2004). These regions are highly accessible and their amino acid residues are targets for specific post-translational modifications. Chemical modifications on the N-ter tail play a role in different biological processes by modifying chromatin structure or by serving as specific docking motifs. Canonical histones are encoded by replication-dependent genes and are massively expressed during S-phase to provide a sufficient supply for DNA replication. They have dedicated chaperones that are required for proper nucleosome assembly. Indeed, histone chaperones are key proteins binding to histones and involved in histone storage, deposition or eviction from the nucleosome.

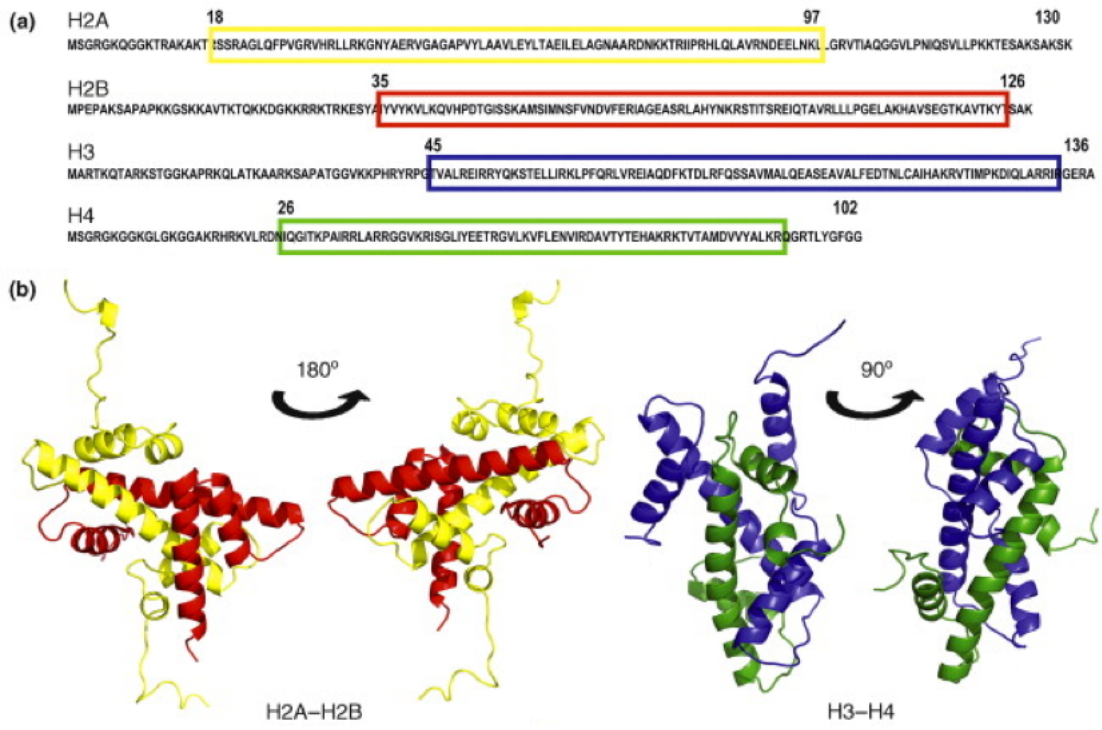


Figure 10: Structure of the histone 'handshake' dimerization (adapted from Das et al., 2010).
(a) Primary structure of H2A (yellow), H2B (red), H3 (blue) and H4 (green). Colored boxes indicate the folded domains in the context of the nucleosome. (b) Two views of the structures of H2A-H2B dimer (yellow and red) and H3-H4 dimer (blue and green).

Histone H1 also named 'linker' histone is composed of a globular domain surrounded by two unstructured tails (N-ter and C-ter). H1 differs from the other histones H2A, H2B, H3 and H4 in the sense that it is not included in the nucleosome but rather binds the two nucleosome linkers and rigidify nucleosome structure (**Figure 11**, Bednar et al., 2017).

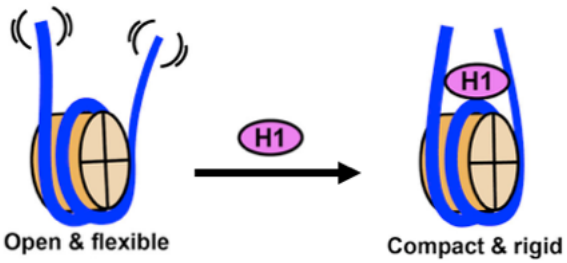


Figure 11: Histone 1 binding induces the nucleosome to adopt a more compact and rigid conformation (adapted from Bednar et al., 2017)

2.1.1.4. Higher levels of chromatin organization

In order to compact the 2 meters of DNA of a human cell in the nucleus, the chromatin is organized at several levels of compaction. The first level of compaction is the nucleofilament or 'beads on a string' which is a fiber of 11 nm. Following the longstanding compaction model, the 11 nm fiber would further fold into 30 nm fibers that would further fold into 120 nm chromonema, 300 to 700 nm chromatids, and finally mitotic chromosomes (**Figure 12a**). This compaction model is however only based on *in vitro* observations and no *in vivo* experiments have enabled to validate this model over the 30 nm fibers. A recent study resolved the 3D organization of chromatin in interphase and mitotic cells using ChromEMT (Ou et al., 2017). 30 nm chromatin fibers could be detected while no higher-order fibers were identified *in situ*. In contradiction with the previous compaction model, they conclude that chromatin is a flexible and disordered 5 to 24 nm diameter granular chain which is packed together at different concentration densities during the cell cycle (**Figure 12b**). The latter model suggests that the global accessibility and activity of DNA is obtained thanks to the assembly of 3D domains in the nucleus with different chromatin concentrations, rather than higher-order folding of the chromatin.

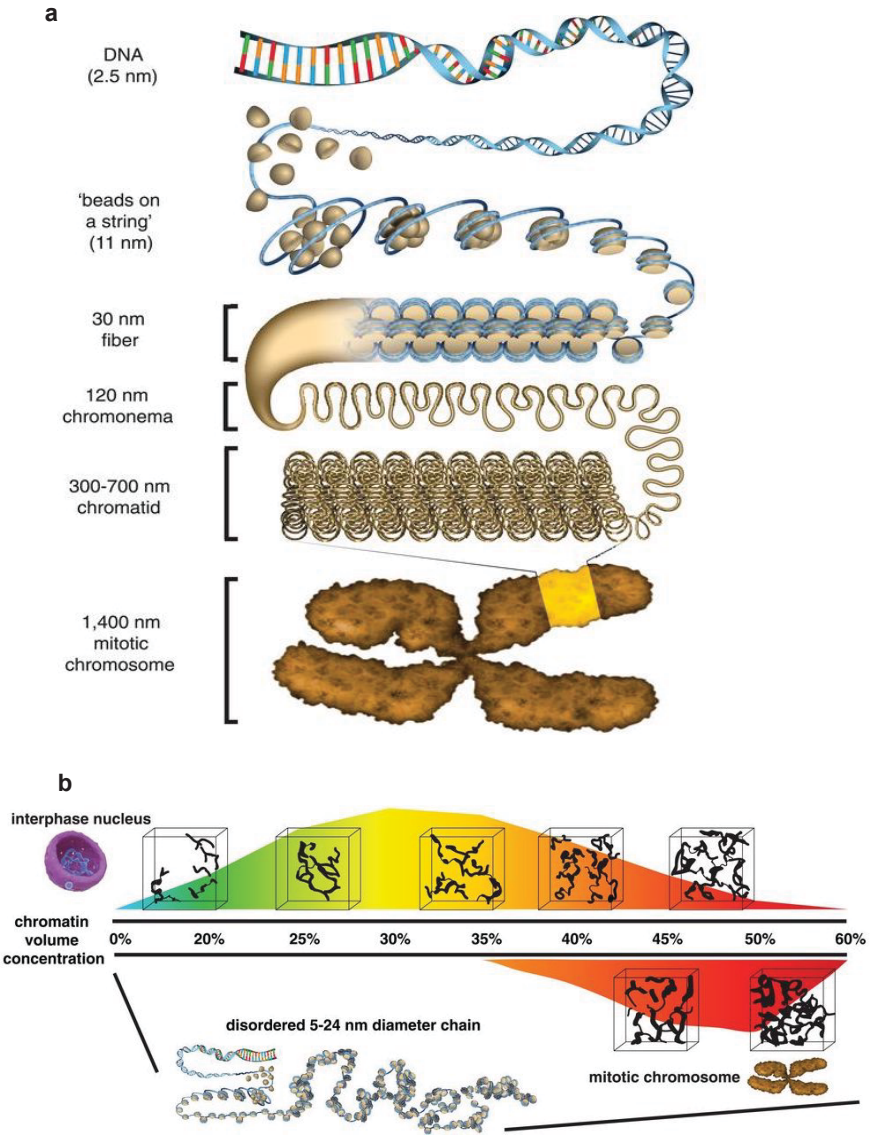


Figure 12: Models for chromatin higher levels of organization
(adapted from Ou et al., 2017)

a. Hierarchical longstanding chromatin-folding model. **b.** Higher-disorder 3D chromatin packaging model. Chromatin is a flexible and disordered granular chain that is packaged at different 3D volume concentration density distributions in interphase nuclei and mitotic chromosomes.

In addition to its structural role of DNA compaction, the chromatin has an important functional role thanks to its highly dynamic nature. Thus, it plays a major role in controlling DNA accessibility during processes as replication, DNA-repair or transcription.

2.1.2. Modifications of the structure of chromatin

The organization of chromatin structure presented above is flexible thanks to variations in the chromatin components. Eukaryotic cells have developed mechanisms to modulate chromatin structure in order to render genetic information more or less accessible. In fact, chromatin can be compacted at different levels and dynamically remodeled through different processes, namely the use of chromatin remodeling factors, the covalent modifications of DNA and histones, the replacement of canonical histones by histone variants. These chromatin modifications are performed by several actors in the cell and enable the integration of environmental and developmental signals to further control essential cell processes such as cell cycle regulation, DNA-repair or transcription regulation (**Figure 13**). Chromatin remodeling is performed by specific chromatin remodeling complexes. Histones and DNA chemical modifications are deposited by specific proteins commonly named 'writers', and removed by their 'erasers' counterparts. The accessible regions and chromatin modifications are further recognized by the 'readers'.

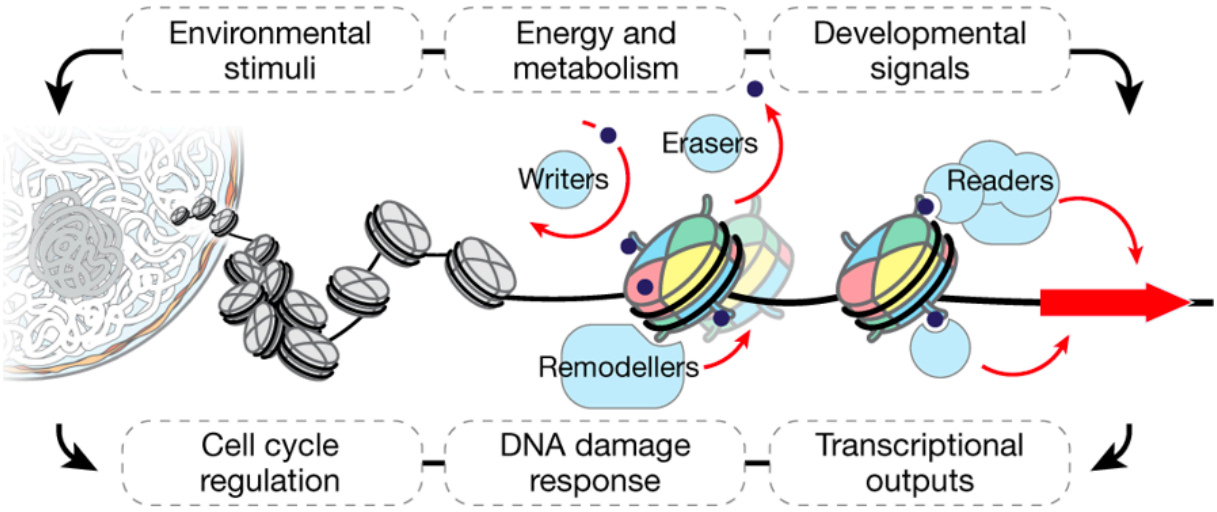


Figure 13: Integrators and effectors of chromatin (Nacev et al., 2019)

2.1.2.1. Compaction and remodeling

The level of chromatin compaction or packaging has a direct influence on genome accessibility and thus expression. A longstanding distinction has been made between euchromatin, or 'open' chromatin – where active genes are predominantly

located – from heterochromatin or ‘closed’ chromatin. Heterochromatin is further subdivided in constitutive and facultative heterochromatin. Constitutive heterochromatin is composed of the strongly silenced and packaged chromatin, rich in DNA repetitive elements, and located at telomeres, centromeres and peri-centromeric regions (Saksouk et al., 2015). The differential compaction state of chromatin can be obtained thanks to the coordination of several actors as histone H1, architectural proteins and RNA (e.g. HP1, pericentric RNA) and ATP-dependent chromatin-remodeling complexes.

Chromatin remodeling is an ATP-dependent mechanism through which interaction between histones and DNA is altered. ATP-dependent chromatin-remodeling complexes (remodelers) are specialized in one of the following functions: nucleosome assembly and organization, chromatin access and nucleosome editing (deposition or eviction of histone variants) (**Figure 14a**, Clapier et al., 2017). Chromatin remodelers are grouped in four different families defined by their ATPase domain: the ISWI family, the CHD family, the SWI/SNF family and the INO80 family (**Figure 14b**). They are all specialized in specific cellular processes thanks to a unique subunit composition and ATPase domain (**Figure 12b**).

Assembly remodelers, such as ISWI and CHD subfamilies, help the initial histone-DNA complexes to mature into canonical nucleosomes and further space nucleosomes at specific distances apart. This assembly and spacing process takes part both during replication and during transcription. Chromatin access remodelers, mainly composed of the SWI/SNF subfamily, are sliding the nucleosomes along the DNA of ejecting part- or full nucleosomes in order to make the chromatin more accessible to proteins (e.g. transcription factors) and RNA. Assembly remodelers are mostly implicated in gene silencing through tight packing of chromatin, while access remodelers are rather implicated in gene expression through chromatin opening. Nucleosome editing remodelers (e.g. INO80 subfamily), are able to replace canonical histones by replication-independent histone variants (see 2.1.2.4).

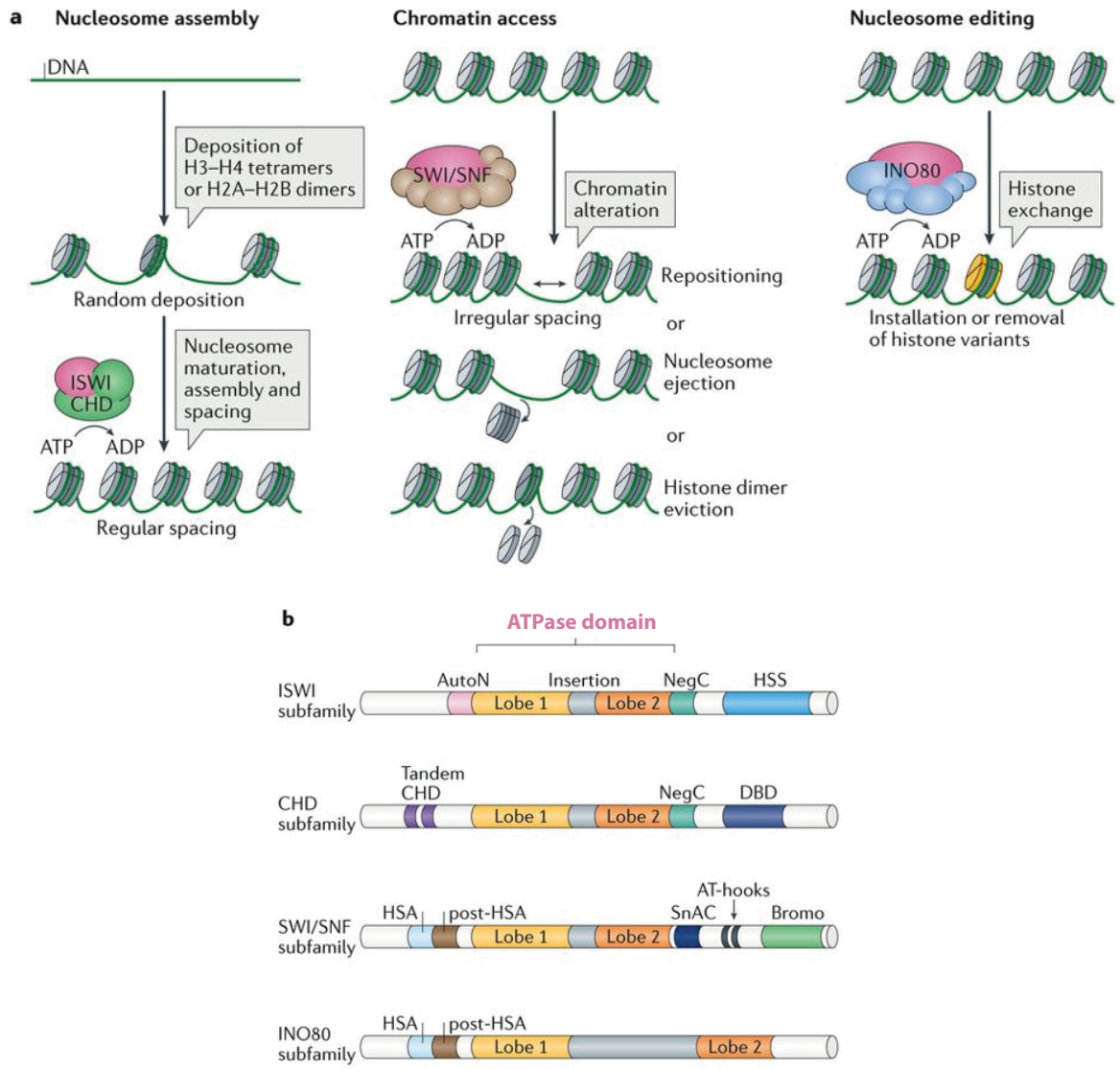


Figure 14: Functions and domain organization of chromatin remodelers (adapted from Clapier et al., 2017)
a. Functional classification of remodelers. The ATPase subunit of all remodelers is depicted in pink. **b.** Domain organization of remodeler subfamilies. The ATPase domain is composed of two lobes separated by a short or long insertion (grey). Remodelers are classified into the four subfamilies based on the length and function of the insertion and on their domain organization. AutoN, autoinhibitory N-terminal; Bromo, bromodomain; CHD, Chromodomain Helicase DNA binding protein; DBD, DNA-binding domain; HAS, helicase/SANT-associated; HSS, HAND-SANT-SLIDE; INO80, Inositol requiring 80; ISWI, Imitation Switch; NegC, negative regulator of coupling; SnAC, Snf2 ATP coupling; SWI/SNF, Switch/Sucrose Non-Fermentable.

Another level of modulation of chromatin structure is performed thanks to DNA methylation.

2.1.2.2. DNA methylation

DNA methylation can be seen as an epigenetic annotation system which provides instruction to the cell as to how and when to read the genetic information. DNA methylation is essential for mammalian development (Okano et al., 1999). Unlike genome sequence which is inherited, DNA methylation patterns are established throughout development in a time and tissue specific manner and remain plastic during life.

Methylation of the fifth position of cytosine is one of the most studied and understood epigenetic mark and is mainly restricted to the CpG context in mammals. CpG methylation is distributed all over the genome excepted in CpG rich regions called CpG islands which mainly remain unmethylated (Bird, 1986). 5-methylcytosine (5mC) deposition is performed by three conserved DNA methyl-transferases (DNMT): DNMT1 also known as the maintenance DNMT, and DNMT3a and DNMT3b which perform *de novo* methylation of both unmethylated or hemi-methylated DNA to assist the maintenance (Liao et al., 2015). The Ten eleven translocation (TET1, 2 and 3) enzymes can oxidize the 5mC into 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) (Ito et al., 2010 and 2011). Moreover, 5fC and 5caC can further be excised to regenerate unmodified cytosines by the action of thymine DNA glycosylase (TDG) together with the base excision repair (BER) enzymes (Cortellino et al, 2011; He et al., 2011; Maiti and Drohat, 2011) (**Figure 15**).

From a general point of view, DNA methylation is often linked to transcriptional repression in mammals and plants (Suzuki and Bird, 2008). However, the link between DNA methylation and transcription reveals to be far more complicated. For example, 5mC in gene transcription start site (TSS) vicinity blocks its expression while its presence in gene body might stimulate gene elongation (Jones, 2012). Finally, recent evidences of non-CpG methylation are emerging but the exact mechanisms are yet poorly understood (Jang et al., 2017).

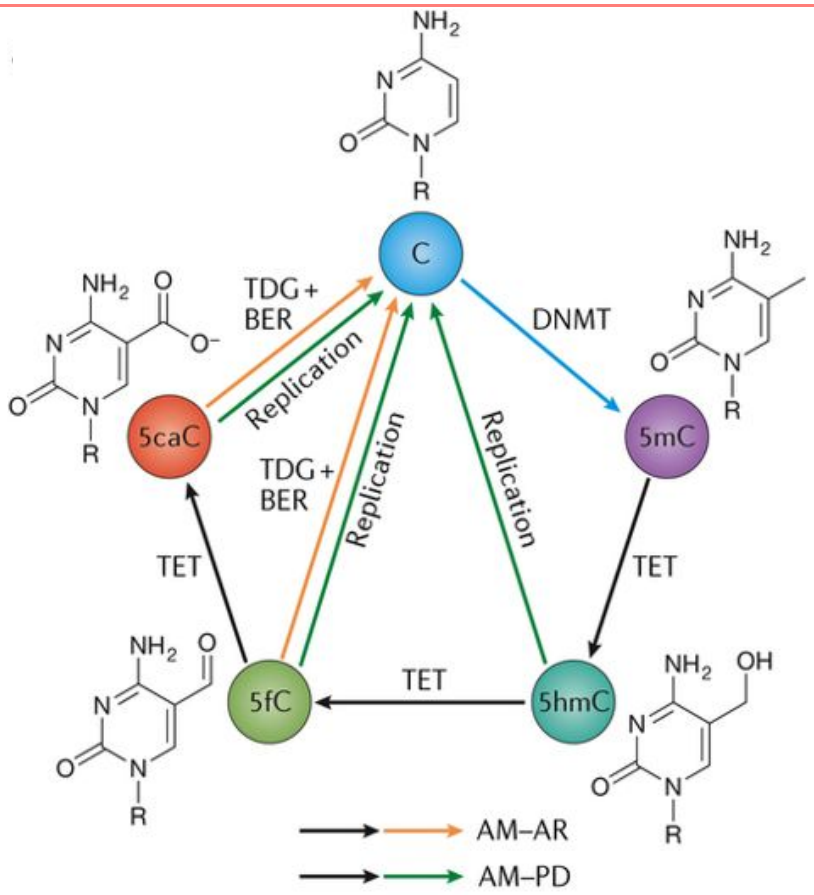


Figure 15: The cycle of active DNA demethylation (from Wu and Zhang, 2017).

DNA methyltransferases (DNMTs) convert unmodified cytosine to 5mC (5-methylcytosine). 5mC can further be reconverted to unmodified cytosine by TET-mediated oxidation to 5hmC (5-hydroxymethylcytosine), 5fC (5-formylcytosine) and 5caC (5-carboxylcytosine) followed by 5fC or 5caC excision mediated by thymidine DNA glycosylase (TDG) coupled with base excision repair (BER). AM–AR, active modification–active removal; AM–PD, active modification–passive dilution.

2.1.2.3. Covalent modification of histones

DNA is not the only chromatin component to be chemically modified. Canonical and linker histones are also covalently marked by post-translational modifications (Huang et al., 2014). The N-terminal tails of the core histones protrude out from the nucleosome and are subject to an important array of post-translational modifications. Change in PTM leads to modifications of chromatin structure and dynamics (Campos and Reinberg, 2009). Histone post-translational modifications (PTM) have been linked to several cellular processes including transcription, DNA replication or DNA repair. A wide variety of histone PTM exists, such as acetylation, sumoylation and ubiquitination of lysine, methylation of arginine and lysine, phosphorylation or serine, threonine and

tyrosine, and several others (For an exhaustive review of the currently documented histone PTM, see Zhao and Garcia, 2015) (**Figure 16**). The first discovered and most studied histone PTM are located in the N-ter tail but more and more PTM located in the globular domain are described.

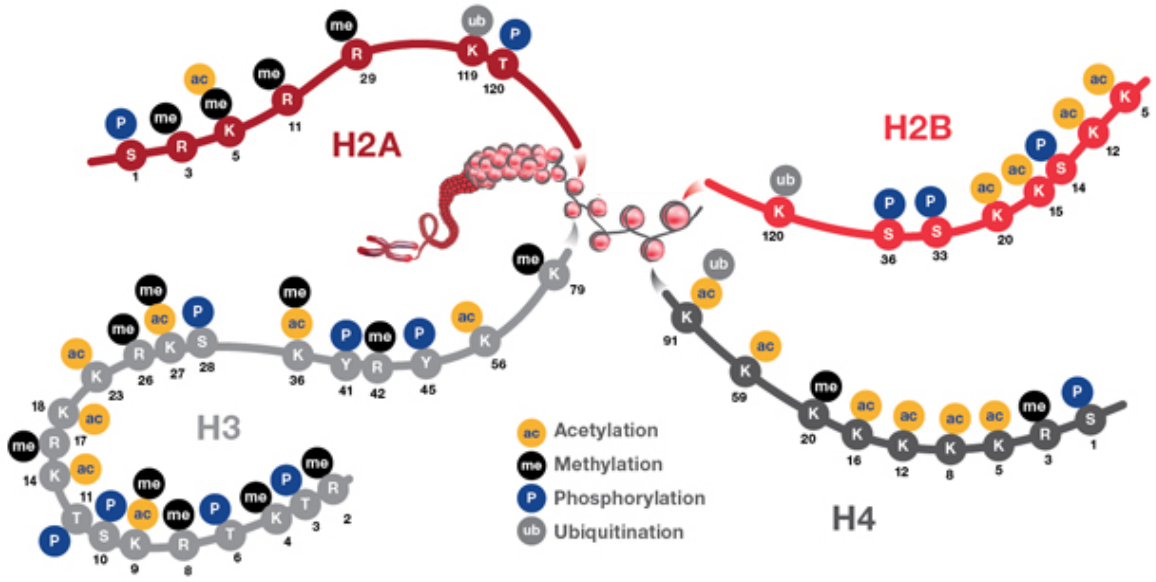


Figure 16: An example of histone canonical PTM (adapted from thermofisher.com)

How histone PTM function in chromatin regulation remains unclear, but the recent advances in the identification of the protein machineries that incorporate (write), remove (erase) and bind PTM give the first flavors of an exciting field aiming at understanding the 'histone code' (**Figure 17**). The epigenetic writers and erasers chemically modify histones and have a direct effect on chromatin structure (e.g. changing the charge of the amino acid residue as for lysine acetylation). The histone PTM landscape also serve as a binding platform through the selective recruitment of readers directing specific downstream chromatin changes (**Figure 17** and Rothbart & Strahl, 2014). Thanks to the reversibility of histone PTM, the histone code shows an outstanding plasticity and enables a fast modulation of epigenetic information in response to environmental changes. However, the interpretation of the histone code solely based on the PTM is not sufficient for a good understanding of chromatin structure and functional changes. The replacement of canonical histones by histone variants brings another level of complexity in the interpretation of the histone code.

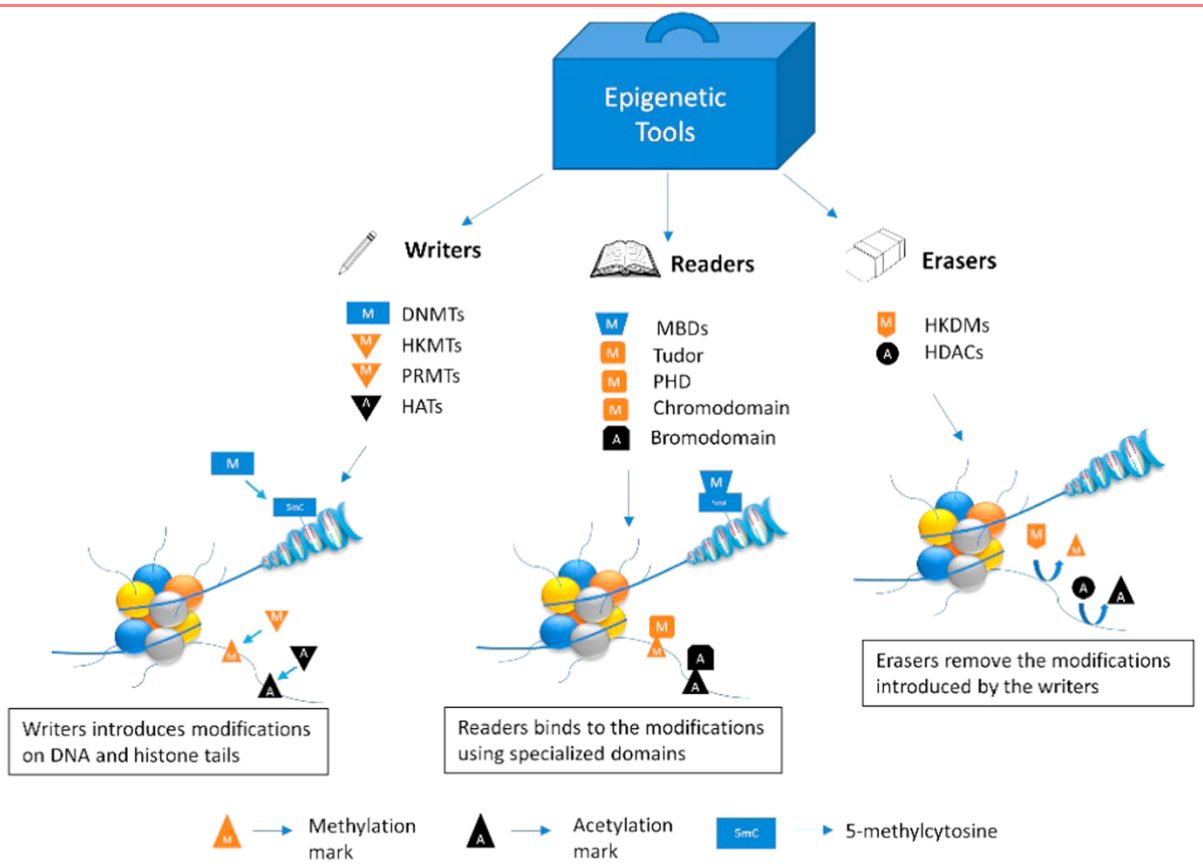


Figure 17: A representation of epigenetic writers, readers and erasers (adapted from Biswas & Rao, 2018).

DNMTs, DNA methyltransferases; HATs, Histone acetyltransferases; HDACs, Histone deacetylases; HKDMs, Histone lysine demethylases; HKMTs, Histone lysine methyltransferases; MBDs, Methyl-CpG-binding domains; PHD, Plant homeodomain; PRMTs, Protein arginine methyltransferases.

2.1.2.4. Incorporation of histone variants

Histone variants are non-allelic isoforms of canonical histones, sharing an overall similar structure while having a relatively different primary structure (33-97 % homology in sequence, see **Figure 18**). Unlike the canonical histones, histone variants are incorporated in chromatin throughout the cell-cycle in a replication-independent deposition. In higher eukaryotes, all canonical histones except H4 have been found to show variant counterparts (**Figure 18** and Talbert et al., 2012). In some cases, the variants differ in only a few amino acids from the canonical histones (e.g. H3.3 and H3.1) whereas other variants can show larger sequence dissimilarities (e.g. H2A variants).

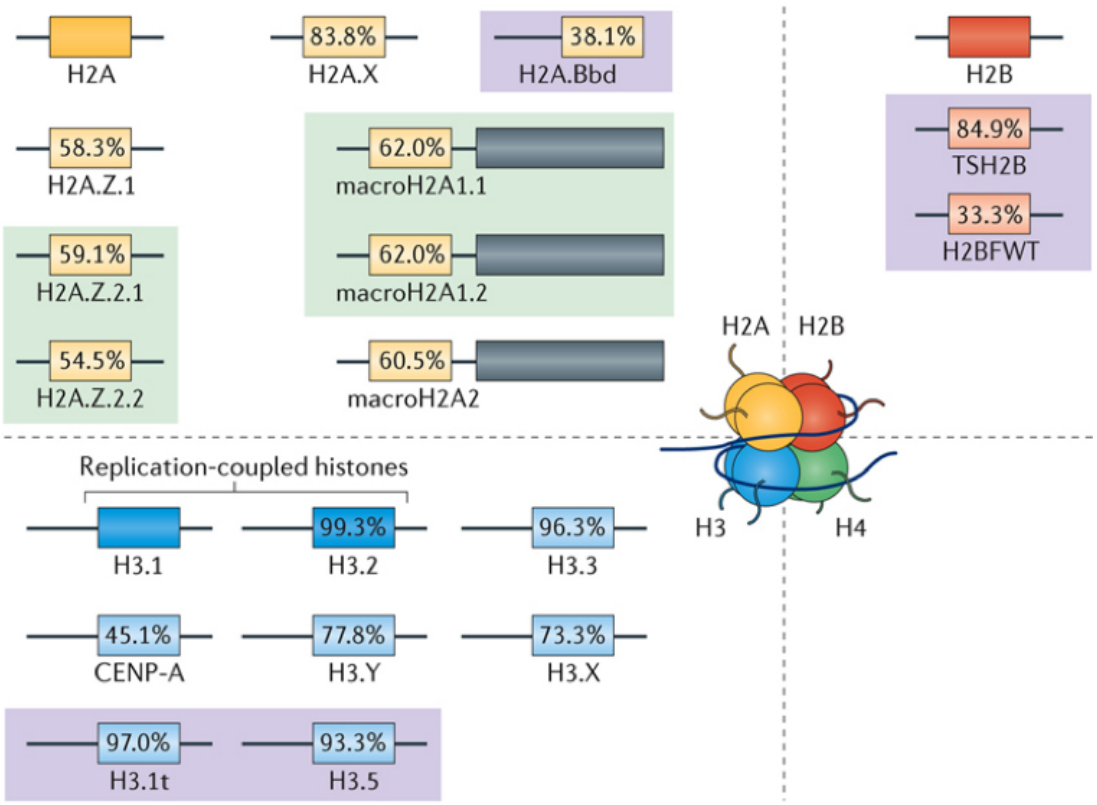


Figure 18: Conservation of currently described human histone variants (from Buschbeck & Hake, 2017).
Canonical histones H2A (yellow), H2B (red) and H3 (blue) and their associated variants in pale yellow, pale red and pale blue respectively. Rectangles represent core regions, and lines represent flexible histone tails. No variants of H4 (green) have yet been discovered in humans. Tissue-specific (testis) histone variants are highlighted by light purple boxes and alternative splice isoforms by light green boxes. Percentages indicate total amino acid sequence conservation (% sequence identity) of the variants relative to their replication-coupled counterparts. CENP-A, histone H3-like centromeric protein A; H2BFWT, histone H2B type WT; TSH2B, testis-specific histone H2B.

In addition to a difference in deposition timing, histone variants are deposited in specific chromatin location differing from their canonical counterparts. This spatio-temporal regulation is performed thanks to specific chaperones dedicated to variants deposition and eviction (For review, see Gurard-Levin et al., 2014). Indeed, each dedicated chaperone recognizes specifically a histone variant and escort it in different way. Thus, chaperones can regulate histone variant supply and dynamics for chromatin assembly and disassembly. They can also participate in defining distinct chromatin landscape by targeting histone variant distribution at specific genomic

location. For example, the canonical histones H3.1/2 are deposited genome-wide by the CAF-1 complex while their variant CENP-A is specifically deposited at centromeres by the chaperone HJURP. On the other hand, H3.3 is recognized by two different complexes (HIRA and DAXX/ATRX) leading to its deposition at different genomic location (**Figure 19**). We will further develop H3.3 regulation in 2.2.

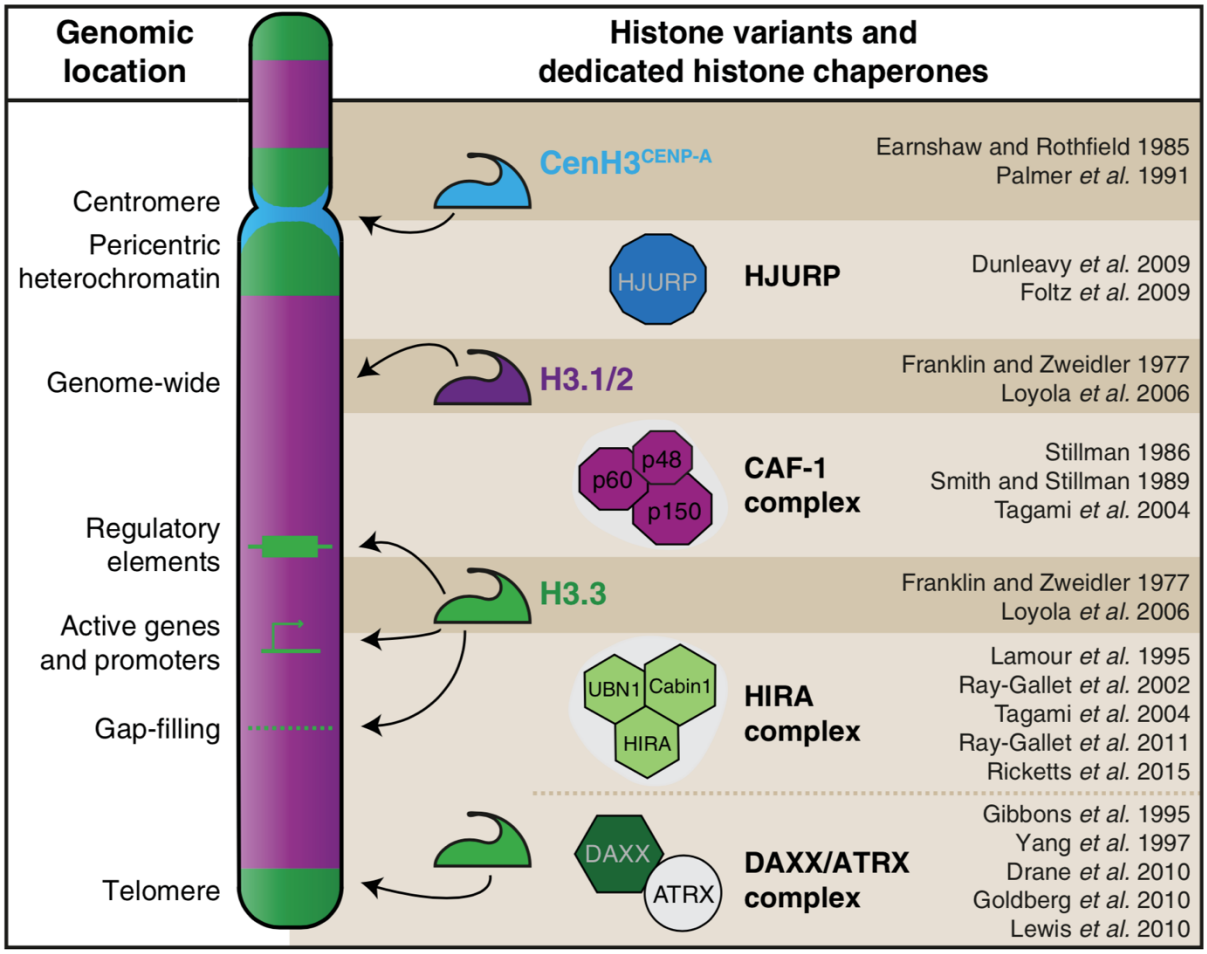


Figure 19: Histone variants and their dedicated chaperones for H3 family (from Sitbon et al., 2017).

CENP-A is incorporated at centromeres by HJURP. H3.3 is incorporated at regulatory elements and gene bodies by the HIRA complex and at telomeres and pericentric heterochromatin by the DAXX/ATRX complex (see 2.2.2 for further details). Canonical H3.1/2 is incorporated genome-wide by the CAF-1 complex.

2.2. H3.3: a multi-faced variant

2.2.1. Structure

In *Drosophila*, mouse and human, H3.3 is encoded by two distinct genes, namely *H3f3a* and *H3f3b*. Both genes are encoding the exact same protein sequence but are located on different chromosomes (1 and 17 respectively) and have distinct untranslated regions (Krimer et al., 1993 and Frank et al., 2003). H3.3 is one of the most conserved protein in eukaryotes and shows only five amino acid differences with H3.1 (at positions 31, 87, 89, 90 and 96) and four with H3.2 (same positions than H3.1 except 96) (Szenker et al., 2011). The position 31 is located in the N-terminal tail which protrude out of the nucleosome while the 3 or 4 other positions are located in the histone fold domain (**Figure 20**). These variations in sequence don't seem to affect the nucleosome structure *in vitro* (Tachiwana et al., 2011).

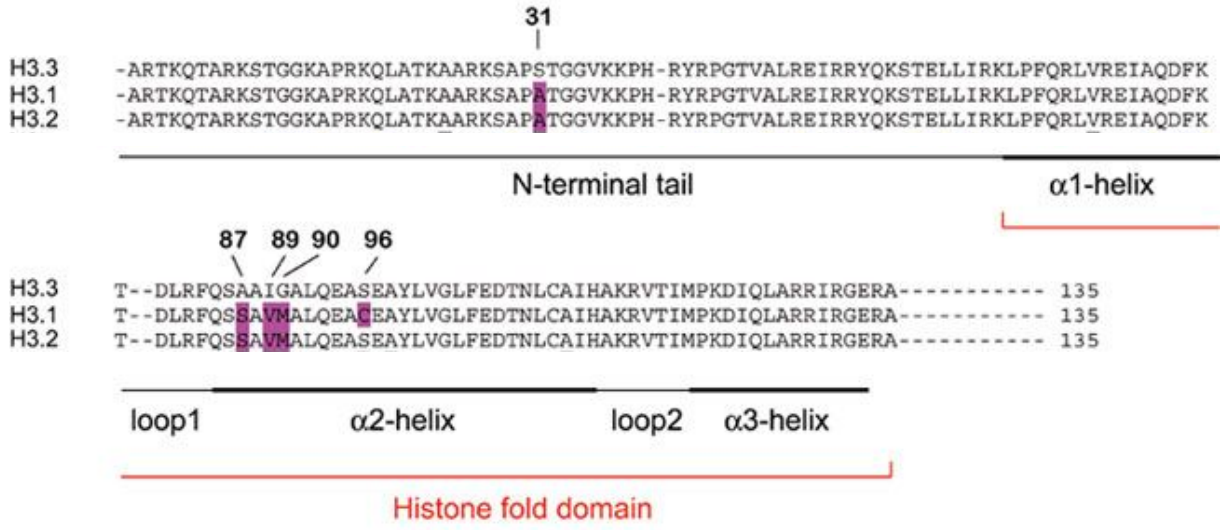


Figure 20: Amino acid sequence alignment between histone variant H3.3 and canonical histones H3.1 and H3.2 (adapted from Szenker et al., 2011)

The variation in amino acid residues between H3.3 and H3.1/2 are highlighted in purple. The secondary structure is composed of 3 alpha helixes separated by 2 loops. The N-ter tail in unstructured. The histone fold domain is indicated in red.

Despite the high sequence homology between H3.3 and H3.1/2, H3.3 is specifically recognized and escorted by dedicated chaperones leading to its enrichment at particular genome sites.

2.2.2. Chaperones

Canonical histones H3.1 and H3.2 are deposited in chromatin in a replication-dependent manner by the CAF-1 complex (see **2.1.2.4** and **Figure 19**). On the other hand, histone variant H3.3 deposition occurs throughout the cell cycle. Even though its deposition is replication independent, H3.3 plays a role in this cellular process by marking early-replication chromatin (Clément et al., 2018). The specific handling of H3.3 over its canonical counterparts is made possible thanks to dedicated chaperones.

To date, two dedicated chaperone complexes have been described for H3.3 handling. The histone cell cycle regulator 1 complex (HIRA complex) takes part in the assembly of H3.3-H4 dimers into nucleosomes. The HIRA complex is composed of HIRA, calcineurin-binding protein 1 (Cabin1), ubinuclein 1 (UBN1) or ubinuclein 2 (UBN2) (Ricketts and Marmorstein, 2017; Xiong et al., 2018) and proceeds to H3.3 deposition at gene bodies, promoters and regulatory elements (Ray-Gallet et al., 2002; Tagami et al., 2004). The mechanism by which HIRA specifically recognizes and deposits H3.3 remains elusive. A recent study suggested that Replication Protein A binds to gene regulatory elements and enables HIRA-mediated newly synthesized H3.3 deposition (Zhang et al., 2017). HIRA subunit trimerization has recently been shown to be required for the functional activity of the HIRA complex as a H3.3 chaperone (Ray-Gallet et al., 2018). Another study showed the specific binding of UBN1 and UBN2 to H3.3 and its concerted deposition with HIRA toward *cis*-regulatory regions in mESC (active promoters and enhancers). The specificity of interaction between UBN1 and H3.3 has also been shown to be mediated by H3.3 positions Ala87 and Gly90 (Xiong et al., 2018).

The death domain-associated protein (DAXX) assisted by ATRX forms the DAXX/ATRX complex and is responsible of H3.3 deposition at heterochromatin, including peri-centromeric regions and telomeres (Drané et al., 2010; Goldberg et al., 2010). Specific interaction between DAXX and H3.3 ‘AAIG’ motif (positions 87-90) has been shown (Lewis et al., 2010). In the absence of DAXX, H3.3 has been found associated with the CAF-1 complex, suggesting an alternative mechanism of deposition (Drané et al., 2010).

Moreover, the Ser31 replacement of H3.3 by an Alanine had no effect on H3.3 deposition, suggesting that Ser31 and its phosphorylation has no role in H3.3

recognition by its chaperones (Hake et al., 2005). In conclusion, the 'AAIG' motif of H3.3 (positions 87-90, **Figure 20**) lead to the specific recognition of H3.3 by its dedicated chaperones, namely the HIRA complexes and DAXX/ATRX complexes.

2.2.3. Biological functions

Despite the high homology between histone variant H3.3 and its canonical counterparts, H3.1/2 cannot rescue H3.3 loss without deleterious defects in mammals (Couldrey et al., 1999; Bush et al., 2013; Tang et al., 2015). The available data suggest that H3.3 is implicated in several essential cell processes, including transcription, DNA repair, mitosis and reprogramming. However, its functional roles remain mainly unknown. We will thus gather the current knowledges about H3.3 biological functions.

2.2.3.1. H3.3 role in transcription

Histone variant H3.3 is found at transcriptionally active chromatin and is enriched in active-associated PTM (Ahmad and Henikoff, 2002; McKittrick et al., 2004). H3.3 deposition has been further shown to be coupled to transcription in *Drosophila* (Schwartz and Ahmad, 2005). Several studies have depicted H3.3 as an active player in the maintenance of accessible chromatin structures in enhancer and transcribed regions. Indeed, nucleosomes harboring H3.3 have been shown to be intrinsically unstable and to promote gene activation, especially with simultaneous presence of histone variant H2A.Z (Jin and Felsenfeld 2007; Jin et al. 2009; Chen et al. 2013). H3.3 is enriched at transcription start sites (TSS) of both active and repressed CG-rich promoters, and in the gene bodies and transcriptional end site of active sequences. The latter has been shown to be proportional to transcriptional activity (Goldberg et al., 2010). HIRA-dependent deposition of H3.3 has been proposed to promote transcription recovery following genotoxic stress (Adam et al., 2013). Transcription-coupled H3.3 dynamics have also been described following stimulation by interferon, through H3.3 deposition on interferon stimulated genes (Bachu et al., 2019).

Due to H3.3 presence at promoters and its higher enrichment at TSS and gene bodies of highly expressed genes (Goldberg et al. 2010), a strong association between H3.3 and active transcription has been made. However, whether H3.3 deposition really drives transcription or simply reflects it is still under debate.

Several studies showed that the absence of H3.3 has a very mild effect on transcription (Bush et al., 2013; Jang et al., 2015, Ors et al., 2017). This leads to the conclusion that H3.3 might not play a pivotal role in transcriptional regulation. However, depletion of H3.3 shows strong defects in development.

2.2.3.2. H3.3 during development

In order to decipher the role and importance of a protein, a common practice is to delete or inactivate the gene coding for the protein of interest. Inactivation of H3.3 has been performed in several organisms and enabled a better understanding of its roles in development. In *Drosophila,* loss of H3.3 does not impair embryonic and postnatal development but leads to fertility defects (Hodl and Basler 2009; Sakai et al. 2009), which can be rescued by ectopic expression of H3.2 (Hodl and Basler, 2012). On the other hand, overexpression of H3.3 in S-phase can also rescue growth defects in H3.2-null flies (Hodl and Basler, 2012). Thus, only the overall level of H3 histones is important for fly development and not any specific variant. Complete loss of H3.3 don't have any effect on *Caenorhabditis elegans* viability or fertility (Piazzesi et al., 2016). In *Xenopus*, H3.3 loss leads to gastrulation defects that cannot be rescued by H3.2 overexpression (Szenker et al., 2012). In zebrafish, H3.3 is important for the proper cranial neural crest cell differentiation (Cox et al., 2012).

In mice, several studies have reported single knockout or knockdown of *H3f3a* and *H3f3b*. Partial or complete loss of both *H3f3a* alleles led to a reduced viability and subfertility or infertility of males (Couldrey et al., 1999; Tang et al., 2013). Mice lacking *H3f3b* had also reduced viability and both homozygous and heterozygous were infertile (Bush et al., 2013; Tang et al., 2013 and 2015; Yuen et al., 2014). In mouse zygotes, defect in nuclear envelope formation and change in chromosome condensation were observed after H3.3 knockdown (Lin et al., 2013; Inoue and Zhang, 2014). On the other hand, mouse models with a single knock-out of *H3f3a* or *H3f3b* have been described as normal and fertile in both sexes by Jang et al., 2015. But when both *H3f3a* and *H3f3b* were depleted, developmental retardation and embryonic lethality was observed. The difference in mice under H3.3 inactivation can be very much explained by the difference in genetic background used.

H3.3 also plays an important role in early development. H3.3R26 and H3.3K27 residues and their PTM have been shown to be essential for proper oogenesis and

good partitioning of the cells to the inner cell mass of the early embryo (Zhou et al., 2017). H3.3 also plays an important role in cell fate transition, first by maintaining the parental cell identities during reprogramming and then by its deposition on genes implicated in the lineage reprogramming (Fang et al., 2018b). Developmentally regulated genes in embryonic stem cells (ESC) harbor promoters decorated both by the activation-associated H3K4me3 mark and by the repression-associated H3K27me3 mark (Bernstein et al., 2006). Those domains are named 'bivalent' and H3.3 is required at the developmentally regulated bivalent promoters the recruitment of the Polycomb-repressive complex 2 (PRC2) recruitment in mouse ESC (mESC) (Banaszynski et al. 2013). Moreover, as suggested with the sterile-phenotype specifically observed in H3.3-lacking male, H3.3 is involved in spermatogenesis. Indeed, H3.3 is exclusively incorporated during mammalian meiotic sex chromosome inactivation and required for gene silencing in the male germ line (van der Heijden, 2007). After oocyte activation, H3 is replaced by maternal-derived H3.3 in the donor nucleus. HIRA-dependent H3.3 deposition has been shown to be required for transcriptional reprogramming following nuclear transfer to Xenopus or mouse oocytes, thus H3.3 plays an important role in reprogramming (Jullien et al., 2012; Wen et al., 2014a and 2014b). Right after fertilization, maternal H3.3 invades paternal chromatin by replacing protamines (Loppin et al., 2005; Torres-Padilla et al., 2006). In addition, HIRA-mediated H3.3 deposition is also required for rRNA transcription in addition to being essential for parental genome reprogramming (Lin et al., 2014).

Besides its role in development, H3.3 has specific contributions in differentiated tissues, as for example in the brain.

2.2.3.3. H3.3 in the brain

H3.3 accumulates in chromatin with age and is thus enriched in differentiated cells. First evidences depicted the accumulation of H3.3 with age in rodent brain (Pina and Suau, 1987). In mouse neurons, H3.3 represents less than 30 % of H3 pool at E16.5 while it constitutes more than 94 % in 2-years-old mice brain. This observation is consistent with H3.3 abundance in human brain, with 31 % of H3 pool in fetal brain and more than 93 % of H3 pool at 14 years old, staying stable until 72 years old (Maze et al., 2015). H3.3 knock down led to decreased numbers of dendritic spines in mouse neurons. Moreover, neuronal activity has been shown to promote HIRA-dependent

H3.3 turnover. Disruption of H3.3 turnover impairs learning and memory processes, hence H3.3 is required for proper neuronal function and brain plasticity (Maze et al., 2015). Moreover, another study showed that H3.3 loss reduces NSC proliferation and leads to premature neuronal differentiation. H3.3 is thus required for correct NSC proliferation and differentiation (Xia and Jiao, 2017). Studies have focused on the relative abundance of H3.3 according to H3.1/2 but little is known about the relative expression between *H3f3a* and *H3f3b* in mouse brain development. In human, *H3f3a* expression stands only for 1 % of H3.3 transcripts in the brain, independently of the stages of development considered (Ren and van Nocker, 2016).

2.2.3.4. H3.3 and mitotic progression

In addition to its role in development, H3.3 is important for mitotic progression. Depletion of H3.3 in mESC and mouse embryonic fibroblasts (MEF) caused mitotic defects such as anaphase bridges and lagging chromosomes (Jang et al., 2015). A recent study confirmed that H3.3 regulates mitotic progression in MEF, thus confirming its important role in the maintenance of genomic integrity (Ors et al., 2017). Moreover, cleavage of H3.3 tail at residue 21 plays a role in senescence by locking the cell in the senescent process (Duarte et al., 2014). H3.3 is also involved in the maintenance of the replication fork and in the transcription restart after UV damage (Adam et al., 2013; Frey et al., 2014).

2.2.3.5. Heterochromatin

In addition to active chromatin loci, H3.3 is deposited at heterochromatin by the DAXX/ATRX complex, more specifically at peri-centromeric regions and telomeres (Drané et al., 2010; Goldberg et al., 2010). H3.3 is involved in heterochromatin compaction. In H3.3-lacking MEF, decondensation of telomeres, centromeres and pericentromeric chromatin has been reported (Jang et al., 2015).

H3.3 has been found enriched at the telomeric (TTAGGG)n repeat, specifically at interphase telomeres, and its enrichment is dependent on ATRX in mESC (Wong et al., 2009; Goldberg et al., 2010). H3.3 deposition at telomeres relies on several factors. Its ATRX-mediated deposition is dependent on DEK (Ivanauskiene et al., 2014). Moreover, the specific interaction between ATRX, CBX5 (HP1) and H3.3 is important in the maintenance of telomere structural integrity (Wong et al., 2010). Tri-methylation

of H3.3K9 serves as an ATRX docking site and is essential for telomere transcriptional repression (Udugama et al., 2015).

H3.3 also contributes to the maintenance of the chromatin assembly at other heterochromatin loci. H3.3 deposition is deposited at peri-centromeric loci by the ATRX/DAXX complex in a PML-dependent manner (Delbarre et al., 2017). Moreover, H3.3 recruitment by DAXX into the PML nuclear bodies has been shown to be essential for the transcriptional regulation of pericentromeric satellite repeats in both mouse and human (Morozov et al., 2012).

Recently, H3.3 has also been found enriched at endogenous retroviral elements (ERV) containing long terminal repeats (LTR). ERV are a subset of transposable elements (see 3.1), and their expression is repressed by the presence of the H3K9me3 mark and a co-repressor complex containing KAP1 in mESC (Rowe et al., 2010; Rowe et al., 2013b). The recruitment of DAXX, H3.3 and KAP1 at ERV has been are co-dependent and occur before H3K9me3 deposition (see 3.2.2). When H3.3 is depleted, ERV-associated H3K9me3 mark decreases and ERV transcription is de-repressed (Elsässer et al., 2015). Thus, H3.3 seems to have an important role in the establishment and maintenance of ERV silencing (see 3.2.2.4).

In conclusion, H3.3 plays an important yet underexplored role in the maintenance of heterochromatin stability and integrity.

2.2.4. Post-translational modifications

As depicted in 2.1.2.3, histone PTM bring an extra layer of flavors in epigenetic regulation. Whether H3.3 has unique modifications compared to canonical H3.1/2 remains unclear, but they might share a vast majority of them, especially in their highly homologous N-terminal tail (**Figure 20**). H3.3-specific serine 31 is phosphorylated specifically in mitosis and is found enriched at telomeres in ESC and embryonic germ cells and at pericentromeric hetereochromatin in non-ESC types (Hake et al., 2005 ; Wong et al., 2009). As shown in 2.2.3, H3.3 is incorporated both at transcriptionally active and repressed loci. Thus, H3.3 can accumulate either active-related PTM or repressed-related PTM (**Figure 21**). H3.3 has first and mainly been described as

enriched in marks representative of a 'transcriptionally active' state. For instance, methylation on K4, K36, and K79 (histone core) and acetylation on K9, K14 and K27 are associated with active transcription of open chromatin while K9, K27 and K64 methylations are repressive marks (McKittrick et al. 2004; Hake et al., 2006; Loyola and Almouzni 2007). Some marks also have a dual role depending the place and time of enrichment. H3K9me3 is known as a repressive transcription mark at heterochromatin, but also associated with transcriptionally active genes (Vakoc et al., 2005). In addition, methylation of H3R26 has been usually associated with elongation of transcription, but is also with transcriptional initiation suppression (Carrozza et al., 2005).

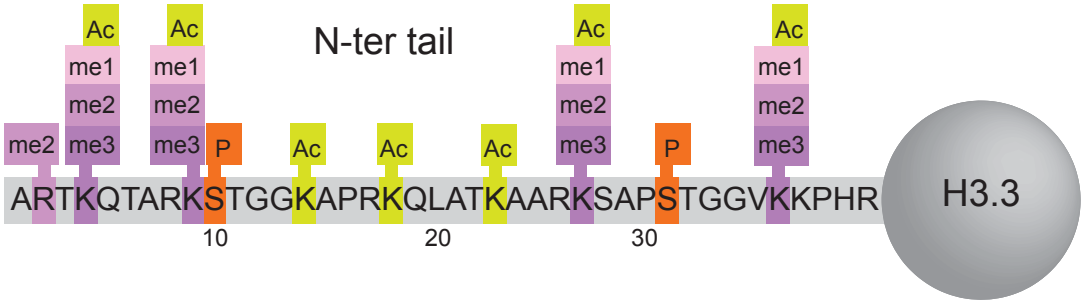


Figure 21: Main post-translational modifications on H3.3 N-terminal tail (adapted from Lowe et al., 2019).

Ac, acetylation; me1, mono-methylation; me2, di-methylation; me3, tri-methylation; P, phosphorylation.

Whether H3.3 PTM have dedicated readers is under intense study. ZMYND11 (also named BS69) has been described as a specific reader of H3.3K36me3 and regulator of transcription elongation and intron retention (Guo et al., 2014b; Wen et al., 2014c). However, the model that link one histone PTM to a biological function (e.g transcription activation or repression) is outdated. More and more evidences show that the histone code should rather be seen as a global PTM landscape than individual PTM. For instance, the association of the activation associated H3K4me3 mark with the repression-association H3K27me3 mark are associated to 'bivalent' promoters (see 2.2.3.2).

Mutations of H3.3 residue prone to PTM or close to modified residues have been shown in cancer and especially in pHGG.

2.2.5. H3.3 mutations in cancer

The first somatic mutations of H3 proteins in cancer have been discovered in 2012 in pHGG. Wu et al. (2012) have sequenced more than 70 pHGG (DIPG and NBS-pHGG) and have found that *H3f3a* or *Hist1h3b* (H3.1) were mutated in 78 % of DIPG and 22 % of NBS-pHGG at the K27 position (K27M) and *H3f3a* was mutated at position G34 (G34R/V) in 14 % of NBS-pHGG. The same year, Schwartzentruber et al. (2012) sequenced nearly 50 pHGG with 31 % of them harboring *H3f3a* mutation (K27M, G34R or G34V). Additional mutations of H3.3 chaperones ATRX or DAXX were found in 31 % of the overall pHGG, with 100 % of G34-mutated tumors. 86 % of *H3f3a* tumors also had p53 somatic mutations (**Table 1**). Moreover, *H3f3a* mutations have been shown to be specific to HGG and highly prevalent in children and young adults, and were thus proposed as drivers in pHGG development (Schwartzentruber et al., 2012). Those mutations appear to be somatic, at high frequency in pHGG and occur in a single gene (dominant negative effect).

H3.3 mutation	pHGG type	Prevalence	Additional mutations			MGMT methylation
			ATRX/DAXX	p53	Others	
K27M	DIPG	78%	30%	86%	PTEN, RB1, ACVR1, MYC, FGFR1, PDGFRA, ALT, NF1	Variable
	NBS-pHGG	22%				
G34R/V	NBS-pHGG	14%	100%	100%	PDGFRA, ALT, CDKN2A, EGFR	

Table 1: pHGG types harboring H3.3 mutation, additional mutations and prevalence.

Further studies have extended H3 mutations to other cancer types, including high frequency mutations in chondroblastoma (K36M in *H3f3b*), giant cell tumors of the bone (G34W and G34L in *H3f3a*) and low frequency mutations in pediatric soft tissue sarcoma (K36M in H3.1), head and neck squamous cell carcinoma (K36M in H3.1/2/3) and leukemia (K27M and K27I in H3.1) (Behjati et al., 2013; Lu et al., 2016; Lehnertz et al., 2017; Papillon-Cavanagh et al., 2017).

Somatic point mutations in the N-ter tail of H3 family members and various mutations in H3.3 dedicated chaperones might thus play an important role in human cancer and are very likely to participate in tumor development (**Figure 22**).

Mutated protein	Mutation	Cancer tissue	Function
H3.3 and replication-coupled H3.1	K27M	Paediatric brain tumours and juvenile bone tumours	Inhibits the K27 methyltransferase activity of Polycomb repressive complex 2
	G34R, G34V and G34L	Paediatric brain tumours and juvenile bone tumours	Unknown function, but tight association with ATRX and DAXX mutations
	K36M	Chondroblastoma	Inhibits K36 methyltransferases
ATRX	Various along coding sequence	Cancers of the peripheral and central nervous system; less frequent in various other types of cancer	Mostly inactivating
DAXX	Various	Various; most frequent in the nervous system	Considered to be inactivating; less frequent than ATRX mutations

Figure 22: Point mutations in H3 family members and their chaperones in human cancer (adapted from Buschbeck and Hake, 2017).

H3 point mutations and their potential role as cancer drivers have paved the way in the importance of histones in cancer development. Indeed, recent efforts in extensive re-analysis of tumor genomic sequencing have shed light on several histone mutations, both in the N-ter tails and in the core of the four canonical histones and their variant counterparts (Nacev et al., 2019).

In addition to K27, G34, K36 positions, several other point mutations have been discovered in H3 genes (**Figure 23**). Several of these mutations seem to be passenger while more and more evidences place K27/G34 mutations as drivers, especially in pHGG (see 2.3). Indeed, despite the heterogeneity and subclonality of pHGG demonstrated by single-cell RNA sequencing, H3.3 mutations are always found as a ‘root’ mutation in the tumor (Nikbakht et al., 2016; Vinci et al., 2018).

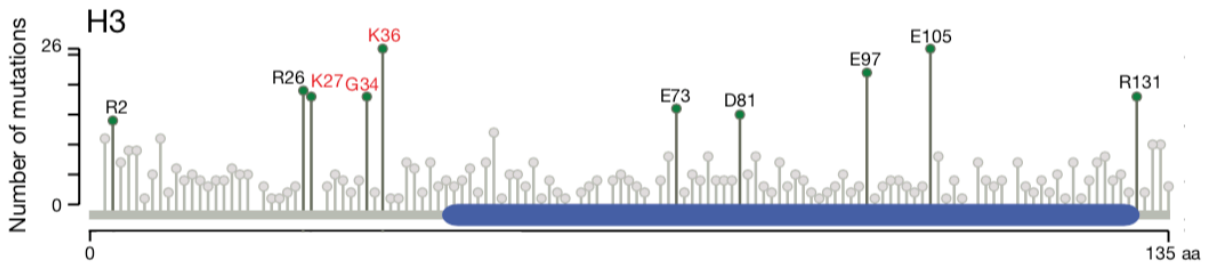


Figure 23: Identification of H3 point mutations in cancer (adapted from Nacev et al., 2019).
The ten most frequently mutated positions are shown in green circles. Established oncohistone mutations are indicated in red. The blue bar represents the globular domain.

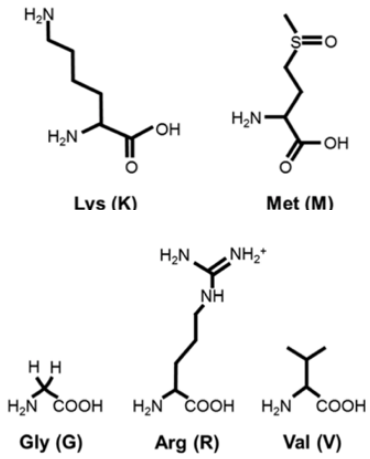
Only one allele of H3.3 (*H3f3a*) or H3.1 (*Hist1h3b* and *Hist1h3c*) is somatically mutated in pHGG. As *H3f3a* and *H3f3b* are coding for the exact same protein, the reason for the exclusiveness of *H3f3a* mutations over *H3f3b* in pHGG remains a mystery. The prevalence of *H3f3a* (or H3.1) K27M mutation might be explained by differences in the codon usage (**Figure 24**). For example, a single point mutation is sufficient for the mutation of lysine 27 in methionine for *H3f3a* and *Hist1h3b/c*, but not for *H3f3b.* However, the codon usage cannot explain the reason why only *H3f3a* is found mutated at G34 position and not *H3f3b*.

		K27	K36		G34
H3.3	*H3F3A*	AAG	AAG		GGG
	H3F3B	AAA	AAG		GGG
H3.1		AAG (6/10) AAA (4/10)	AAG (8/10) AAA (2/10)		GGC (8/10) GCT (2/10)
Mutation	Met	ATG	ATG	Val	GTG GTC GTT
				Arg	CGG CGA CGT CGC
				Trp	TGG
				Leu	TTG TTA CTT CTC CTA CTG

Figure 24: Codon usage in histone H3.3 and H3.1 genes at the sites of histone mutation (from Kallappagoudar et al., 2015).

Genes in which mutations are prevalent for each amino acid substitution is marked in orange.

H3 mutations at K27 and G34 have a different impact on the biochemistry of the residue depending the type of mutation (**Figure 25**). The lysine 27 which is basic, is replaced by a hydrophobic methionine. Moreover, this mutation abrogates the post-translational K27 site. On the other side, the hydrophobic glycine 34 is replaced either by a basic arginine that can be methylated, or a hydrophobic valine.



Amino Acid	Properties	PTM
Lys (K)	Basic	Me, Ac, Ub
Met (M)	Hydrophobic	
Gly (G)	Hydrophobic	
Arg (R)	Basic	Me
Val (V)	Hydrophobic	

Figure 25: Properties of the amino acids substituting K27 and G34 in pHGG and their possible PTM (adapted from Kallappagoudar et al., 2015). Me, methylation; Ac, acetylation; Ub, ubiquitination

Even though both K27 and G34 are in close vincinity on H3.3 N-ter tail, an important antagonism is observed between those two mutation sites in pHGG. The next part will focus on the antagonism between K27M and G34R/V at the level of the age of diagnostic, tumor location and prognosis (2.3.1). We will also detail the current knowledge on the impact of those mutations in pHGG development (2.3.2/2.3.3).

2.3. H3.3 mutations: driver in pHGG development

As described in 2.2.5, G34R/V mutations and the majority of K27M mutations occur specifically in the gene *H3f3a* coding for H3.3. These mutations are exclusive to high-grade tumors and predominant in children. The mutated residues are in the N-ter tail, a region rich in PTM (**Figure 26**). Even though the two mutation sites are very close, tumors harboring either K27M or G34R/V mutation differ on their age of diagnosis, location and prognosis.

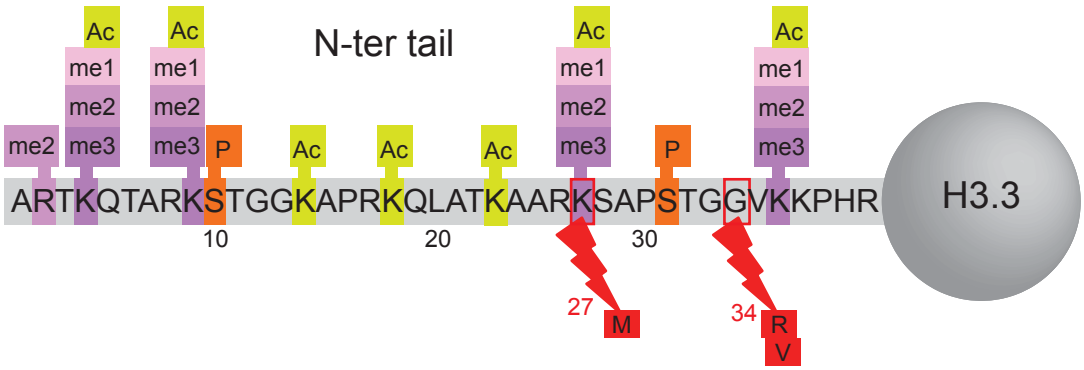


Figure 26: PTM environment of K27 and G34 mutations on H3.3 tail (adapted from Lowe et al., 2019).

Ac, acetylation; me1, mono-methylation; me2, di-methylation; me3, tri-methylation; P, phosphorylation.

2.3.1. Antagonism between K27M and G34R/V

2.3.1.1. Age of diagnostic

pHGG age of diagnostic greatly differs between H3.3 mutation sites. The age incidence profile of K27M pHGG peaks at 7 years while G34R/V pHGG peaks at 14 years (**Figure 27**). H3.3K27M pHGG can further be divided in NBS-pHGG with a median of 10 years and DIPG with a median of 6.5 years. H3.1K27M pHGG are globally found in younger patients than H3.3K27M with a median age at 5 years (Mackay et al., 2017).

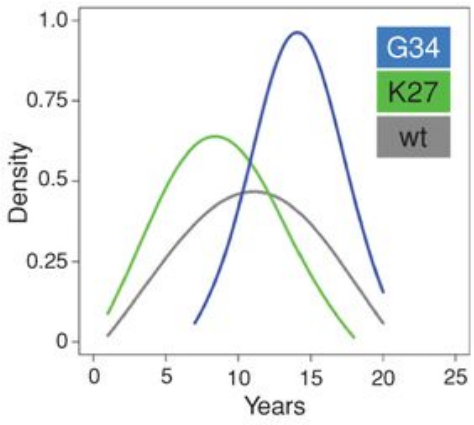


Figure 27: Different distribution of the age of diagnostic depending on H3.3 mutation (adapted from Bjerke et al., 2013).

2.3.1.2. Location

In addition to a different distribution of the age of diagnostic, the anatomical location differs between pHGG harboring K27M and G34R/V. H3.3K27M tumors are found throughout the midline structures, such as the brainstem, cerebellum, thalamus, and spine. On the other hand, H3.3G34R/V tumors are exclusively occurring in cerebral hemispheres (nonmidline supratentorial areas) (**Figure 28**). To note, H3.1K27M pHGG are restricted to the pons (Taylor et al., 2014).

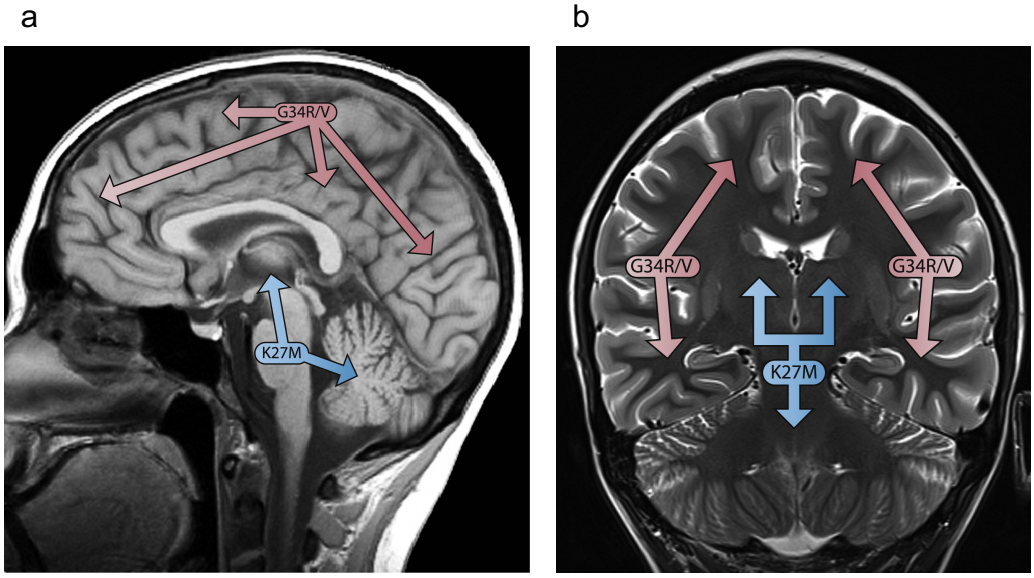


Figure 28: K27M and G34R/V pHGG differ in the anatomical distribution (from Diaz and Baker, 2014).

2.3.1.3. Prognosis

As well as the location, the clinical outcome is different depending on H3.3 mutation (**Figure 29**). H3.3G34R/V pHGG have been associated with a longer overall survival compared to K27M pHGG (median at 18 months and 27 % of 2-year overall survival). H3.3K27M pHGG have been depicted as more aggressive tumors with a median of survival of 11 months and less than 5 % of 2-year overall survival. In addition, H3.1K27M pHGG have been associated with a slightly longer survival than H3.3K27M, with a median of 15 months (Mackay et al., 2017).

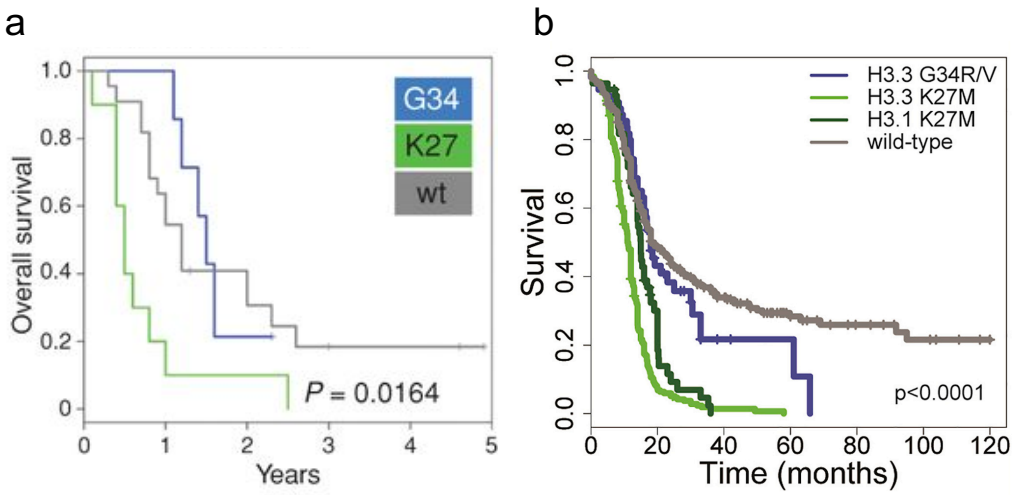


Figure 29: Overall survival of pHGG patients is defined by H3.3 mutations.

Kaplan–Meier plot for overall survival stratified by H3.3/1 mutations **a**. from Bjerke et al., 2013 **b.** from Mackay et al., 2017

2.3.1.4. Transcriptional and DNA methylation signatures

pHGG can be further classified and separated at the transcriptional level according to H3.3 mutation (Paugh et al., 2010; Schwartzentruber et al., 2012; Bjerke et al., 2013; Castel et al., 2018). As expected, the transcriptional profiling follows the tumor localization (K27M-midline vs. G34R/V-hemispheres). K27M and G34R/V tumors can also be separated according to their DNA methylation profile (Castel et al., 2018).

Since their discovery, H3.3 mutations have been described as tumor drivers and they have been intensively studied in order to attempt finding a cure to those deadly cancers. The main focus has been made on K27M while G34R/V mutations remain poorly studied.

2.3.2. G34R/V: an understudied mutation

2.3.2.1. Prevalence

G34R/V tumors account for 16% of cerebral hemisphere tumors. G34R mutation is more frequent than G34V and both are exclusive to H3.3. All G34R/V pHGG have been shown to concomitantly harbor ATRX loss of function (Mackay et al., 2017).

2.3.2.2. General impact of G34 mutation

G34R/V mutations have been associated to global DNA hypomethylation, especially at telomeric regions (Sturm et al., 2012). Moreover, the concomitant ATRX mutation led to an alteration of H3.3 loading at telomeres, disrupting the telomeric heterochromatin and thus inducing alternative lengthening of telomeres (Schwartzentruber et al., 2012). G34R/V tumors are the only pHGG harboring frequent MGMT promoter methylation (Korshunov et al., 2015). MGMT is coding for a DNA repair enzyme and its expression is associated with aHGG resistance to treatment with alkylating agents such as temozolomide. Hypermethylation of MGMT promoter in G34R/V is in favor of MGMT repression, thus avoiding temozolomide resistance (Hegi et al., 2005).

2.3.2.3. Impact of G34 mutation on PTM

G34 mutations occur on a residue which is not targeted by PTM. Nevertheless, G34 mutations happen in an environment very rich in PTM, such as K27, P31 and K36 modifications (**Figure 26**). Bjerke et al. (2013) highlighted a differential distribution of the active mark H3K36me3 between the patient derived KNS42 cell line harboring H3.3G34V mutation and the patient derived H3.3WT SF188 cell line. A significant correlation between H3K36me3 and polymerase II binding has been validated for ~150 differentially enriched genes involved in forebrain and cortex developmental processes. They concluded that H3K36me3 differential binding led to an upregulation of genes implicated in cell fate decisions. Moreover, they described an upregulation of MYCN under G34V expression, which could play an oncogenic role.

Another study from Lewis et al. (2013) described no overall H3K36me3 differences between cell lines carrying H3.3 G34R/V or WT, but they showed a diminution of SETD2 (SET domain containing 2)-mediated H3K36 methylation specifically on G34R/V mononucleosomes *in vitro*. This suggest that G34 mutation

inhibits SETD2-mediated H3K36me3 on the mutated histone. Another *in vitro* study in a G34 mutant background (G34L/W found in giant cell tumors of the bone) validated the alteration of K36 methylation through SETD2 inhibition (Shi et al., 2018). Recent data were also obtained in fission yeast which contain only one form of histone H3. A single enzyme performs H3K36 methylation - Set2 (homologous to human SETD2) in fission yeast. Expression of H3-G34R led to H3K36me3 and H3K36ac reduction, but not H3K36me2 (Yadav et al., 2017). The reduction of H3K36me3 was due to a decrease in Set2 activity which could not be rescued by an overexpression of Set2. In another *in vitro* study, G34R/V mutation has been described as impairing ZMYND11 (see **2.2.4**) binding to H3.3K36me3 (Wen et al., 2014c).

In addition to the specific inhibition of SETD2 methyltransferase, H3.3 G34R has been shown to inhibit the histone lysine demethylases KDM4 (Voon et al., 2018). KDM4 family is composed of the three lysine demethylases KDM4 A/B/C which are responsible for specific K9 and K36 demethylation. By the introduction of a single-copy G34R mutation of *H3f3a* in mESC, Voon et al. (2018) described an overall gain of H3K36me3 across the genome. Their work has pointed out the preferential binding of KDM4 for H3.3 G34R over WT leading to simultaneous inhibition of KDM4 activity. Thanks to structural studies, the G34 residue has indeed been shown to play an important role for KDM4 positioning relative to H3K36me3 (Couture et al., 2007). The latter indicate that a small amino acid is required at the -2 positions of K9 and K36 (A7 and G34) for proper catalytic activity. Thus, G34 mutation in arginine leads to the replacement of the small glycine by the bigger arginine (**Figure 25**) and could thus interfere with KDM4 catalytic activity. In addition to H3K36me3, Voon et al. (2018) have also reported that expression of G34R led to H3K9me3 gain in a KDM4-associated manner. Their model of G34R mutant had similar H3K9/K36me3 phenotypes than KMD4-triple knockout mESC, thereby demonstrating that expression of a single copy of G34R is sufficient to cause chromatin and transcriptional changes similar to those observed in KDM4-triple knockout. The major effects of H3.3 G34R is thus thought to be mediated through the KDM4 pathway (Voon et al., 2018). An increase of H3K36me3 has also been described in H3.3G34V pHGG, proposing that G34V mutation may also inhibit KDM4 (Schwartzentruber et al., 2012; Bjerke et al., 2013).

Nevertheless, special care should be taken in the interpretation of PTM recognition by antibodies while comparing mutant with wildtype proteins. For example,

a previous work demonstrated that the recognition of H3K36me2 was diminished by 10-fold in the presence of G34R mutation with two different antibodies (Yadav et al., 2017).

2.3.2.4. Impact of G34 mutation of genomic stability

H3-G34R expression in fission yeast has led to genomic instability with enhanced chromosome loss and segregation defects (Yadav et al., 2017). Homologous recombination has been found impaired under H3-G34R expression and checkpoint signaling was delayed though clearly functional. The replicative delay observed in H3-G34R expressing cells was an important source of genomic instability and could potentially represent a source of facilitation for tumor development. Interpretation of those results should be done with special attention as the whole fission yeast H3 population was mutant.

Recent studies have emphasized the importance of H3K36me3 in DNA repair, notably DNA mismatch repair (MMR) (Jha and Strahl, 2014; Pai et al., 2014; Li et al., 2013; Pfister et al., 2014). MMR has the role of correcting mispairs during DNA replication and is thus an important machinery for genome maintenance. G34R/V mutations have been shown to inhibit the MMR system leading to genomic defects (Fang et al., 2018c). Indeed, H3K36me3 is essential for the recruitment of the MMR component MutSα. The disruption of H3K36me3 (e.g. in G34R/V context) leads to MMR defects and genome instability. Fang et al. (2018c) showed that G34R/V mutation leads to a decrease of H3K36me3 and H3K36me2 in human cells, thus a decrease in MutSα recruitment. They also described the impediment in H3K36me3-MutSα binding caused by the large side chain of G34 mutant. In both cases, an impairment of MutSα recruitment to chromatin is observed, leading to MMR deficiency and potentially promoting tumorigenesis.

In conclusion, G34R/V is playing diverse roles in the context of pHGG (**Figure 30**). However, G34R/V impacts are different depending the study and the model used. For instance, some studies describe a global change in K36me3 while others depict only local changes. Moreover, G34R/V has been reported as potential inhibitor of KDM4 and SETB2 but those two actors are known to have antagonist roles (e.g. concerning K36 methylation). Additional studies will be needed to have a clearer view on the role of H3.3 G34R/V in the context of pHGG development.

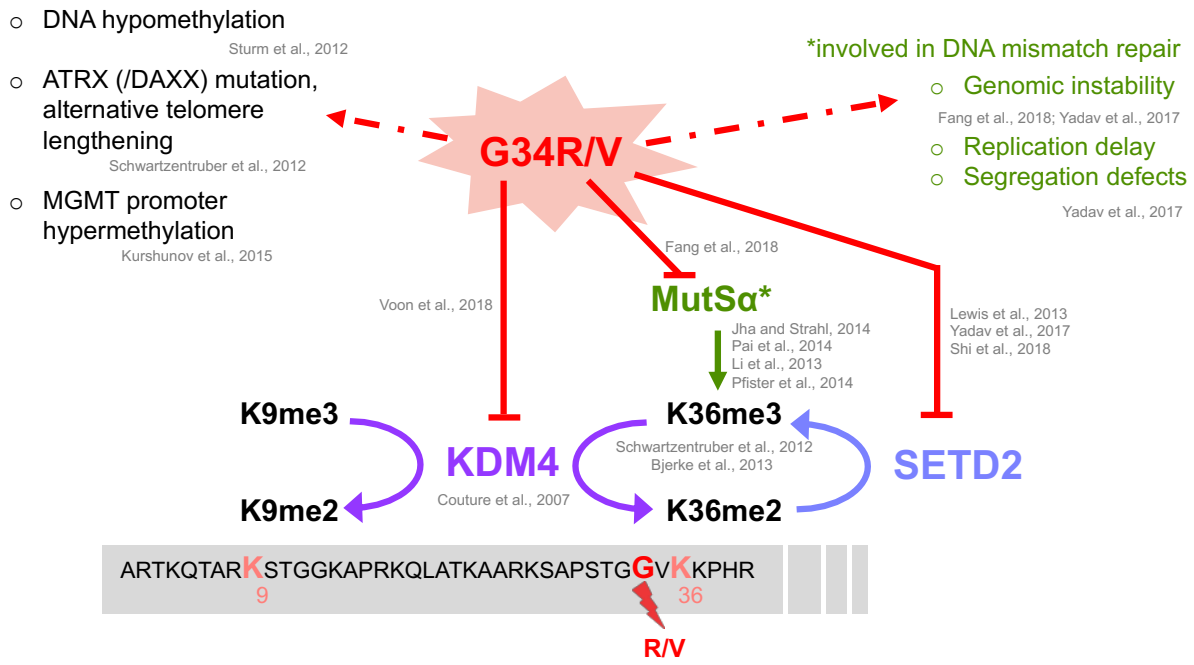


Figure 30: Summary of the current view of G34R/V impact on chromatin and cellular processes.

2.3.3. K27M: the mutation in the spotlight

2.3.3.1. Prevalence

H3K27M has been the first mutation identified and is found in about 30 % of pHGG. K27M mutation occurs in *H3f3a* for 70 % and more rarely in genes encoding H3.1/2 (*Hist1h3b, Hist1h3c, Hist2h3c*) (Lowe et al., 2019). H3.1K27M has been shown to co-segregate with ACVR1 (activing A receptor, type 1) mutation and p53 wildtype, to occur in younger patient and to correlate with better survival (Buczkowicz et al., 2014; Fontebasso et al., 2014; Taylor et al., 2014; Wu et al., 2014). Moreover, 63 % of DIPG and 60 % of NBS-pHGG were found to harbor H3.3K27M mutation (Mackay et al., 2017).

2.3.3.2. H3.3 K27M, a driver of tumorigenesis?

The driver characteristic of K27M remains under debate as K27M expression is not sufficient to induce proliferation in undifferentiated human ESC or primary human astrocytes (Funato et al., 2014) and even represses proliferation in immortalized human astrocytes (Buczkowicz et al., 2014). On the other hand, coupled expression of

H3.3K27M and mutant p53 induced proliferation but not glioma development in nestin expressing progenitor cells of mice neonatal brainstem (Lewis et al., 2013). Another study in human derived neural progenitor cells has depicted a synergy between K27M expression, p53 loss and PDGFRA activation in neoplastic transformation (Funato et al., 2014).

At the mouse level, Pathania et al. (2017) showed that H3.3K27M expression is lethal in postzygotic cells, while its expression under *Nes* or *Gfap* promoters - which are active in neural precursor cells during embryonic and post-natal development - failed to induce tumors, even in combination with p53 loss (**Table 2**). Using *in utero* electroporation, they generated a mice model by introduction of H3.3K27M or H3.3WT under an ubiquitous promoter combined with p53 loss. H3.3K27M but not H3.3WT led to the development of brain tumors which were slowly progressing toward high-grade lesions by 6 months of age. Addition of ATRX knockdown and WT PDGFRA overexpression shortened tumor latency with first ectopic proliferations seen at P21 followed by extensive tumors at 4 months. Another mice model has been developed by Mohammad et al. (2017) and consisted of overexpression of platelet derived growth factor subunit B (PDGFB) and H3.3K27M in mouse neural stem cells further injected in the mouse pons (**Table 2**). Combination of PDGFB with H3.3K27M led to faster tumor formation than with H3.3WT. Recently, Larson et al. (2019) developed the first endogenous inducible mice model for DIPG consisting of knock-in H3.3K27M on *H3f3a* endogenous allele, p53 loss and PDGFRA mutant (**Table 2**). Postnatally induction gave rise to brainstem high grade glioma. This mice model recapitulated the gene expression signature in DIPG with K27M mutation. Their results also demonstrated that H3.3K27M alone was able to enhance self-renewal of NSCs.

All together, these studies highlight the importance of the model used when studying H3.3 mutations. Indeed, the level of H3.3K27M mutant expression, the context and targeted cells as well as the timing of expression are of major importance and are directly impacting the study outcomes. Depending the model used, H3.3K27M possesses or not characteristics of a tumor driver. The establishment of models expressing H3.3K27M in a controlled and non-tumor background is important in order to determine H3.3K27M role and if it is really sufficient for tumorigenesis induction. To date, the results seem to agree on the fact that H3.3K27M alone is insufficient to lead to brain tumor in mice, but lead to higher tumor progression when put together in the

context of p53 loss. The use of diversified models helps to better understand and validate H3.3K27M mutation impacts.

Study	Background	Mutation	Strategy	Observation
Pathania et al., 2017	C57BL/6J	H3.3 K27M	Endogenous mutation of *H3f3a* by CRISPR	Postzygotic lethality
			Expression under *Nes / Gfap* Promoter w/ /wo p53 loss	No tumor development
		H3.3 K27M p53 KD	In utero electroporation (IUE) using piggyBac transposon-base vectors targeting neural progenitor cell at E12.5-13.5	Neoplastic transformation after 6-8 months
		H3.3 K27M p53 KD ATRX KD		Neoplastic transformation after 4 months, high-grade glioma after 9 months
		H3.3 K27M p53 KD ATRX KD PDGFRA amplification		Ectopic proliferation at P21, extensive proliferative regions at 4 months, important tumor penetrance and progression up to 9 months
Mohammad et al., 2017	Balb/C	H3.3 K27M p53 KO	Injection of NSC harboring the mutations in mouse pons	No tumor development
		H3.3 K27M PDGFB amplification		Tumor formation in the pons
Larson et al., 2019	C57BL/6J	cKI-H3.3 K27M	Induction of H3.3 K27M expression under *Nestin* promoter at post-natal days P0 and P1	No tumor development
		cKI-H3.3 K27M cKO-p53	Induction of H3.3 K27M expression under *Nestin* promoter and excision of p53 at post-natal days P0 and P1	Medulloblastoma formation with extensive infiltration of adjacent cerebral tissues
		cKI-H3.3 K27M cKO-p53 cKI-PDGFRA(V544ins)	Induction of H3.3 K27M expression under *Nestin* promoterexcision of p53 and overactivation of PDGFRA at post-natal days P0 and P1	Brainstem and supratentorial high-grade glioma with high penetrance

Table 2: Mouse models for pHGG harboring H3.3 K27M mutation.

(KD, Knock-Down; KO, Knock-Out; cKO, conditional Knock-Out; cKI, conditional Knock-In; Nes, Nestin; NSC, Neural Stem Cells)

2.3.3.3. *Impact of K27 mutation*

H3.3K27M-containing nucleosome have been found to retain the wildtype molecular architecture and stability, indicating that K27M did not lead to rearrangement in the nucleosome core particle structure. In addition, the diffusion kinetics was similar between H3.3K27M and H3.3WT-containing nucleosome in live human cells (Hetey et al., 2017). By studying patient-derived cell lines, a reduction in DNA methylation on oncogenic regions of the genome has been described in the K27M context and this potentially participates in the tumor phenotype stabilization (Bender et al., 2013). In addition, a recent study has described the activation of the expression of multiple cancer/testis antigens in K27M-pHGG patient derived cell lines (Deng et al., 2018). The most upregulated was the primate-specific VCX/Y family. K27M-driven global reduction of H3K27me3 and DNA hypomethylation has been proposed to be at the origin of cancer/testis antigens activation which potentially takes part in gliomagenesis (see **2.3.3.4** for H3K27me3 phenotype). The impact of K27M on the global chromatin landscape remains largely unknown, but many efforts have been made to understand the impact on K27 post-translation modifications.

2.3.3.4. *Effect on K27 post-translational modifications*

H3K27 residue can be acetylated (H3K27ac) or mono- (H3K27me1), di- (H3K27me2), or tri-methylated (H3K27me3) (**Figure 21**). H3K27 methylation is regulated by the evolutionarily conserved Polycomb group proteins, in particular the Polycomb Repressive Complex 2 (PRC2). The Polycomb group proteins are highly conserved and play an essential role in development, for instance in X-chromosome inactivation or in maintenance of stem cells identity. PRC2 is composed of the two methyltransferases Enhancer of Zeste Homolog 1 or 2 (EZH1 and EZH2). The active site of EZH1/2 consists in a S-adenosyl methionine-dependent methyltransferase SET domain.

A global decrease of H3K27me3 has been described in the context of K27M which has been suggested to play a dominant role in blocking the accumulation of this repressive mark (Bender et al., 2013; Chan et al., 2013a; Lewis et al., 2013; Venneti et al., 2013). The expression of H3.3K27M transgenes have been sufficient to reduce H3K27me3 *in vitro* and *in vivo* (Lewis et al., 2013). In K27M harboring tumors, a reduction of H3K27me3 but also H3K27me2 has been globally observed (Chan et al.,

2013a). The loss of H3K27me3 from gene promoters has been associated with transcriptional upregulation, notably the neural transcription factor OLIG2 which potentially promotes glioma development. Nevertheless, H3K27me3 has been found locally enriched together with EZH2 at hundreds of gene loci. H3K27me3 enrichment has been associated with gene silencing including tumor suppressor genes like p16/INK4A (Cordero et al., 2017; Mohammad et al., 2017; Piunti et al., 2017). The global decrease of H3K27me3 with localized enrichments has been confirmed by Bender et al. (2013). According to those studies, changes in H3K27me3 reprogram epigenetic landscape and gene expression and play a potential role in tumorigenesis.

Recent work by Larson et al. (2019) has described a global change in H3K27me3 as well as H3K27ac but a rather limited change in gene expression, restricted in genes involved in neural development. H3K27me3 loss has been further shown to be responsible for the activation of genes controlled by a bivalent promoter marked by concomitant H3K27me3 and H3K4me3.

To explain the global loss of H3K27me3 under K27M expression, a sequestration of PRC2 by K27M has been proposed with several theories and models.

2.3.3.5. The sequestration model: several theories

Under physiological conditions, H3K27me3 is deposited by the specific methyltransferase EZH2 (Margueron and Reinberg, 2011) and H3K27me3 has been shown to stimulate PRC2 methyltransferase activity (Margueron et al., 2009; Xu et al., 2010). K27M mutant has been shown to sequestrate PRC2 leading to a global decrease of H3K27me3 mark (Bender et al., 2013; Chan et al., 2013b; Lewis et al., 2013; Venneti et al., 2013; see **Figure 31a**). In this model, PRC2 is aberrantly recruited and stabilized at K27M site, and is thus unable to fully perform its K27 methyltransferase activity on wildtype H3, leading to a genome-wide decrease of K27me3 (Lewis and Allis, 2013). In addition, the preferential binding of PRC2 to H3.3K27M over H3.3WT has been shown *in vitro* and *in vivo*. *In vitro*, a H3K27M peptide is sufficient to block PRC2 activity and K27M has been shown to inactivate PRC2 via a specific hydrophobic interaction between methionine and the active site of EZH2 *in vitro* and *in vivo* (Lewis et al., 2013; Brown et al., 2014). K27M is sufficient to block the SET catalytic domain of EZH2 through a gain-of-function mechanism and thereby inhibits its enzymatic activity (Bender et al., 2013; Lewis et al., 2013). Affinity

of EZH2 has been shown to be 16-fold higher for H3.3K27M than H3.3WT (Jiao et al., 2015; Justin et al., 2016). In addition, Chan et al. (2013a) demonstrated that K27M led to the disruption of the positive feedback loop regulating PRC2 by its sequestration and this study confirmed a global loss of H3K27me3. But this first sequestration model cannot explain all the data published. Indeed, some genes retain high level of H3K27me3 and EZH2 at loci where K27M is absent (Bender et al., 2013; Chan et al., 2013a; Mohammad et al., 2017; Larson et al., 2018). By contrast with previous work, recent *in vitro* nucleosome binding assays showed that PRC2 has the same binding affinity for both H3.3K27M and H3.3WT containing nucleosomes (Wang et al., 2017). Indeed, PRC2 has been shown to bind both wildtype and K27M-containing nucleosome with similar nanomolar affinity *in vitro*. This is in line with another recent study suggesting that H3.3K27M is not involved in the recruitment and sequestration of PRC2 *in vivo* (Piunti et al., 2017). Indeed, Piunti et al. (2017) have shown that H3.3K27M colocalizes with transcriptionally active chromatin and active H3K27ac mark while being excluded from regions harboring PRC2 and H3K27me3 in human DIPG. Due to the mutually exclusive chromatin localization of K27M and PRC2, K27M does not seem to recruit and sequester PRC2 in DIPG cells (see **Figure 31c**). Other studies have confirmed that PRC2 is excluded from regions containing H3.3K27M (Mohammad et al., 2017; Fang et al., 2018a). These results are inconsistent with the theory of PRC2 sequestration by K27M. Fang et al. (2018a) reported a sequestration of PRC2 at poised enhancer (marked by H3K4me1 and low H3K27me3) and not at active promoters enriched for H3.3K27M (see **Figure 31b**). In addition to the H3K27me3-mediated repression of the tumor suppressor p16, they have shown that the tumor suppressor Wilms Tumor 1 (WT1) is specifically repressed in K27M but not WT context. They suggest that H3K27me3 and PRC2 are likely to silence several tumor suppressor genes including p16 and WT1 for the initiation and maintenance of DIPG tumors. Mohammad et al. (2017) further demonstrated that DIPG proliferation is dependent on PRC2. K27M has been shown to inhibit EZH2 auto-methylation and thus diminish EZH2 activation (Wang et al., 2019; Lee et al., 2019). In line with this study and by using live-cell single molecule tracking in mESC and DIPG cells, Tatavosian et al. (2018) suggested that H3.3 K27M stabilizes EZH2 on chromatin and prolongs its search process and residence time on chromatin. Stafford et al (2018) work is also in contradiction with the sequestration model. They indeed showed that PRC2 is only transiently recruited to H3.3K27M-containing chromatin and not sequestrated. In

addition, they highlighted a persistent inhibition of PRC2 even after its dissociation from H3.3K27M, likely by inhibition of PRC2 allosteric activation. This work depicted a potential 'memory' of PRC2 from its previous association with K27M (see **Figure 31d**). A recent study by Harutyunyan et al. (2019) in primary H3.3K27M tumor lines highlighted a defective spread of H3K27me3 rather than an inhibition of PRC2. PRC2 recruitment to chromatin and methyltransferase activity seem to be unaffected by H3.3K27M but its spread from the binding site toward the neighboring regions to be silenced is impaired. In addition, they pointed out the importance of PRC2/K27M ratio which is highly variable depending the model considered (from 10 to 100-fold excess) and could partly explain the variations observed in the phenotypes. On the other hand, H3.3K27M has been directly linked to the loss of global H3K27me3 *in vivo* by Silveira et al. (2019). Indeed, knockdown of *H3f3a* in DIPG-derived cell lines restored H3K27me3 levels which were lost under H3.3K27M expression.

In summary, H3.3K27M does not have the same effects in all cell types or models and is not always strongly interacting with PRC2. Indeed, even if PRC2 has shown strong affinity for K27M peptides in some studies, PRC2 and H3.3K27M are often mutually excluded from chromatin in DIPG. Thus, the PRC2 sequestration model by K27M on chromatin is not sufficient to explain the complexity and diversity observed and further study are needed to confirm and further decipher the role of H3.3 mutations in pHGG development.

2.3.3.6. K27M essentiality in tumor maintenance

A long-lasting question is whether H3.3K27M is only essential for tumor initiation or if it plays a role in tumor maintenance. Harutyunyan et al. (2019) demonstrated the essentiality of H3.3K27M in glioma tumorigenesis as its removal abolished the capacity of patient derived cells to form tumors in mice. In addition, Silveira et al. (2019) performed H3.3 knockdown experiments in DIPG xenograft harboring H3.3 WT or K27M. They showed that knockdown of H3.3K27M but not H3.3WT was sufficient to delay tumor growth in DIPG-derived xenografts. H3.3K27M plays thus an important role in tumor maintenance.

2.3.3.7. De-repression of endogenous retroviruses in a K27M tumor context

A recent study by Krug et al. (2019) highlighted a new role for K27M in pHGG biology. By generating tumor-derived isogenic mouse models bearing K27M mutation, they pointed out a global H3K27me3 loss followed by a pervasive H3K27ac deposition. The latter has been shown to induce baseline expression of normally silenced repetitive elements, notably endogenous retroviruses. This work give rise to potential novel treatment strategies and pave the way toward a new field of study in the role of H3.3 mutations in pHGG tumorigenesis.

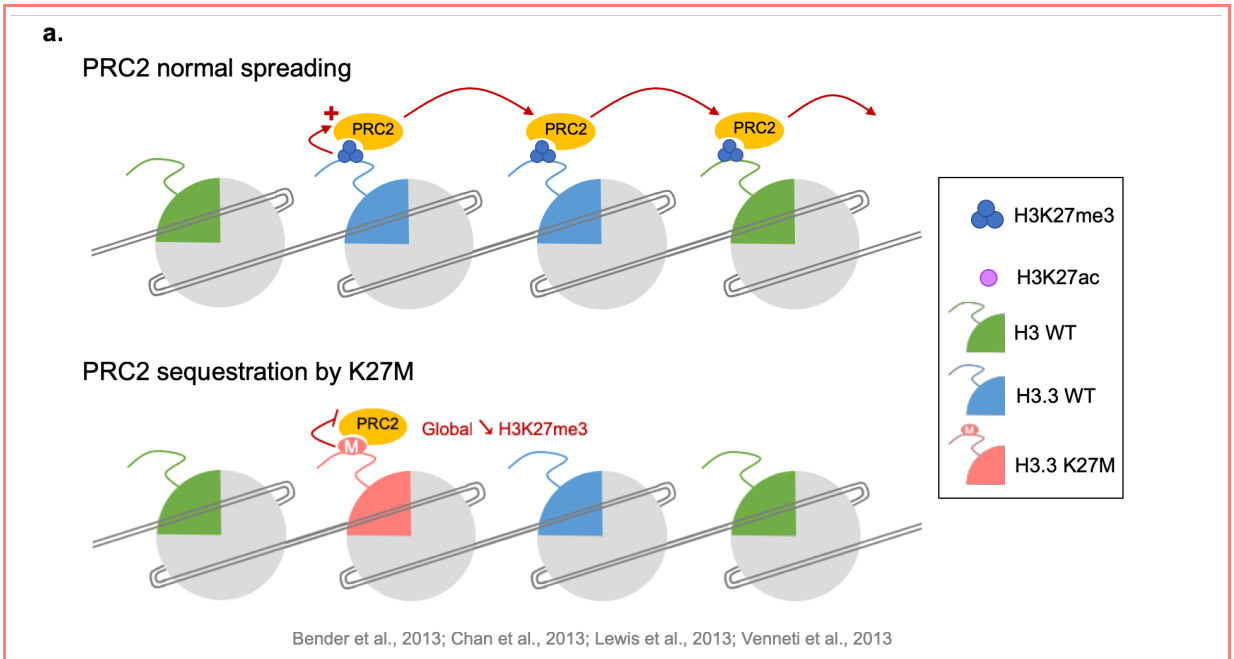


Figure 31 continues on next page…

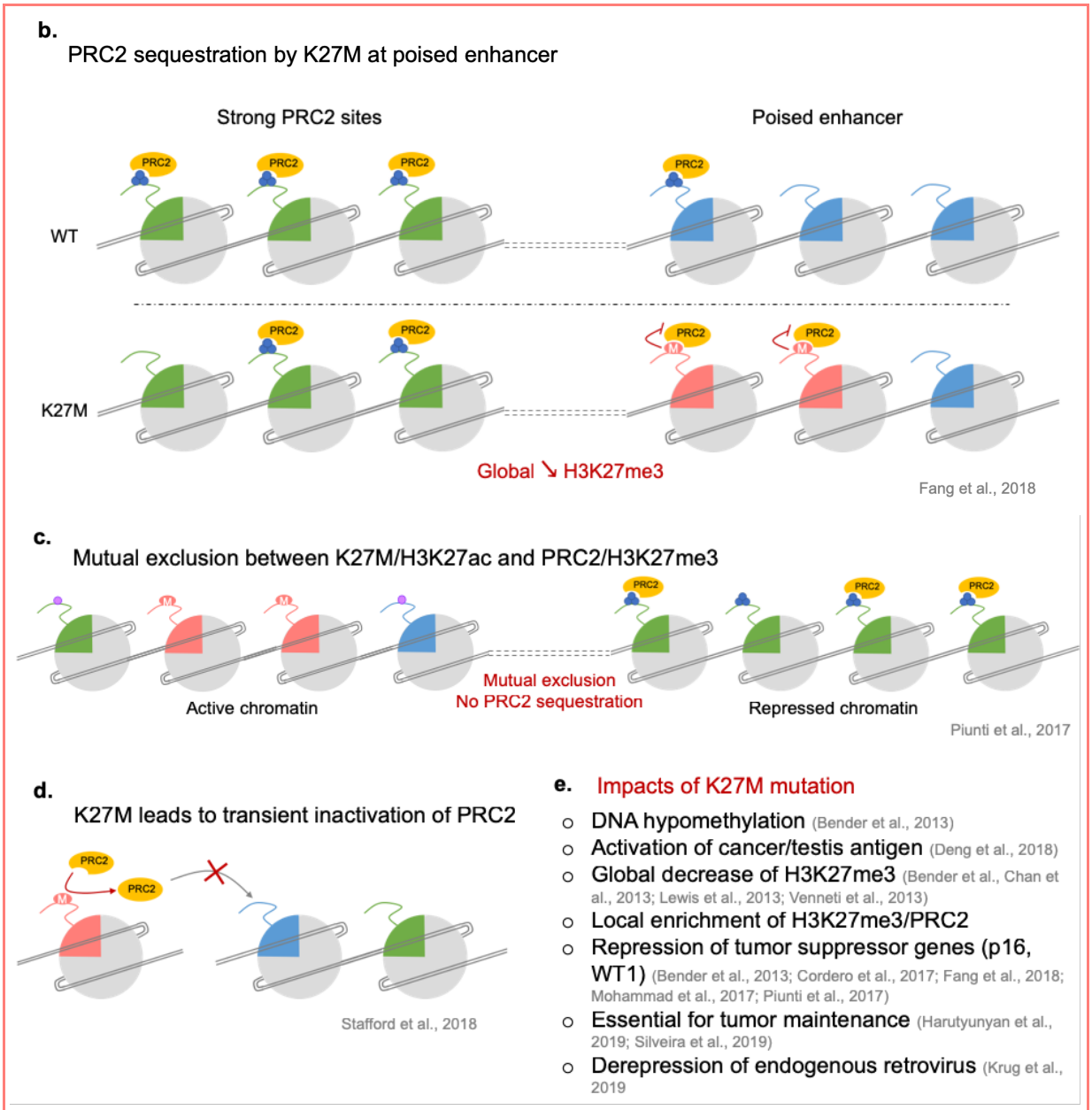


Figure 31: Current views of K27M mutation impacts: one mutation but several theories.

a. Original sequestration model of PRC2 by K27M. PRC2 is aberrantly recruited and stabilized at K27M site, and is thus unable to fully perform its K27 methyltransferase activity on wildtype H3 leading to a genome-wide decrease of K27me3. **b.** Other sequestration model of PRC2 by K27M but only at poised enhancers. This sequestration leads to a delocalization of PRC2 away from its target sites, thus inducing a global H3K27me3 decrease. **c.** Model of mutual exclusion between K27M localized at H3K27ac rich regions and PRC2 recruited at H3K27me3 rich regions (no PRC2 sequestration). **d.** Model of transient allosteric inactivation of PRC2 enzymatic activity. PRC2 is then unable to trimethylate other H3K27WT. **e**. Summary of currently described K27M biological impacts.

Chapter 3: Endogenous retroviruses, a potential awakening in cancer?

3.1. Transposable elements, a balance between threat and benefit

With 70 years of research since the first transposable element was described, the non-coding part of the genome previously considered as 'junk' or 'selfish' DNA (Orgel and Crick 1980) has become a field of intense study. The human genome is composed of ~ 3 billion base pairs with only 1,5 % of which are coding for proteins, commonly called the exome. On the other hand, more than half of the human genome is composed of transposable elements (Bannert & Kurth, 2004). The distribution is comparable in the mouse genome (Mouse Genome Sequencing Consortium et al., 2002). Transposable elements are considered as essential motors of evolution thanks to their ability to modify genomic architecture or gene expression. These latter led them to be also considered as potential threat for genomic stability, thus needing a finely tuned regulatory balance. This chapter will give a brief overview on transposable elements classification and role in evolution and will then focus on endogenous retroviruses, their regulation and proposed role in cancer.

3.1.1. Classification

The dogma considering genomes as static entities was challenged in the early 1950s when transposable elements (TE) were discovered in maize by Barbara McClintock (McClintock, 1950). TE are DNA sequences that have the ability to change their position in the genome. Due to the high homology between TE and viruses, the sum of the TEs in a genome is often named 'endovirome' (Friedli and Trono, 2015). TE classification has been established following their mechanism of mobilization and sequence homology, and is composed of two main classes: DNA transposons and retrotransposons (Wicker et al., 2007). In addition to those two classes, the satellites enlarge the DNA repetitive elements family (**Figure 32**).

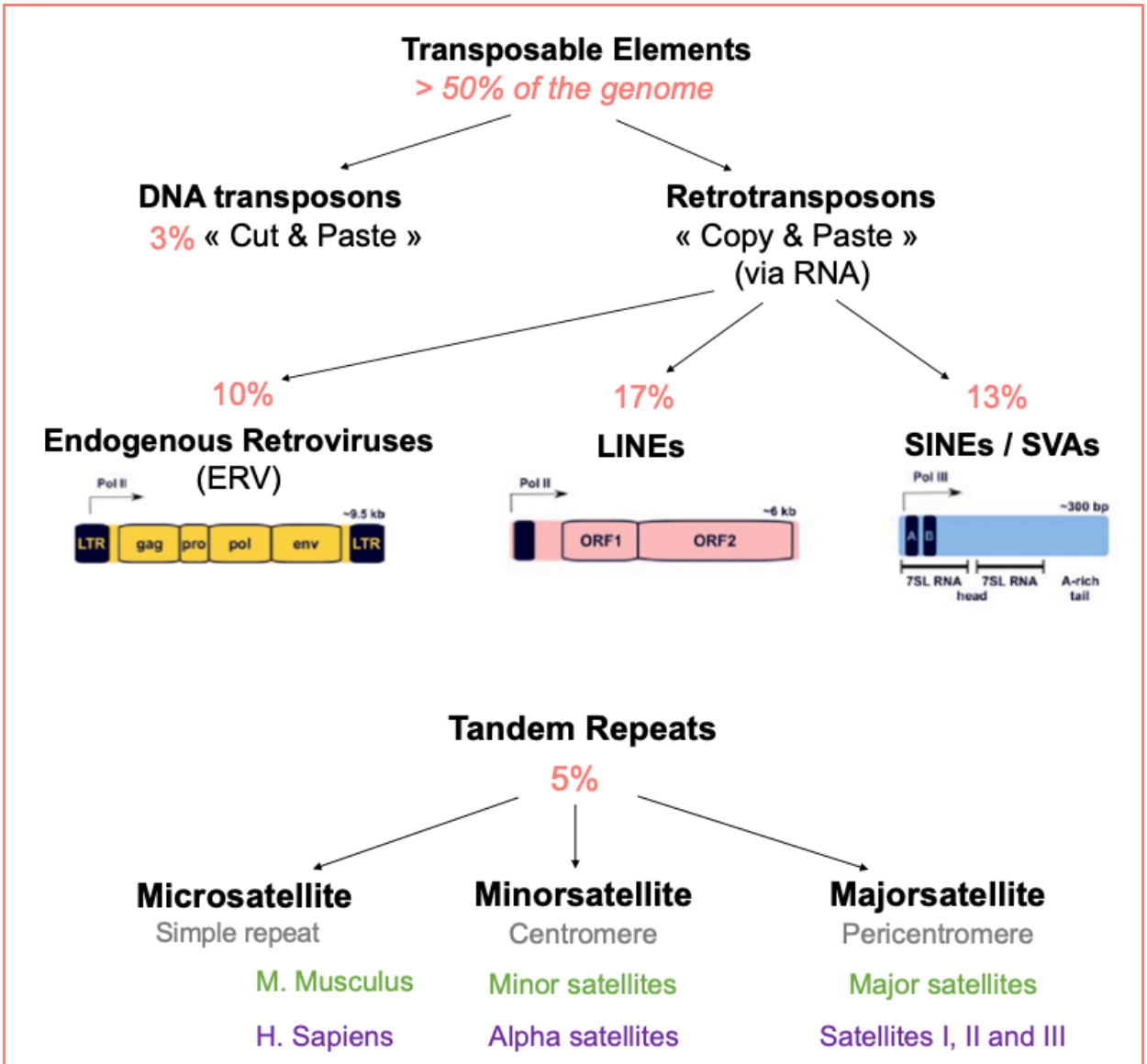
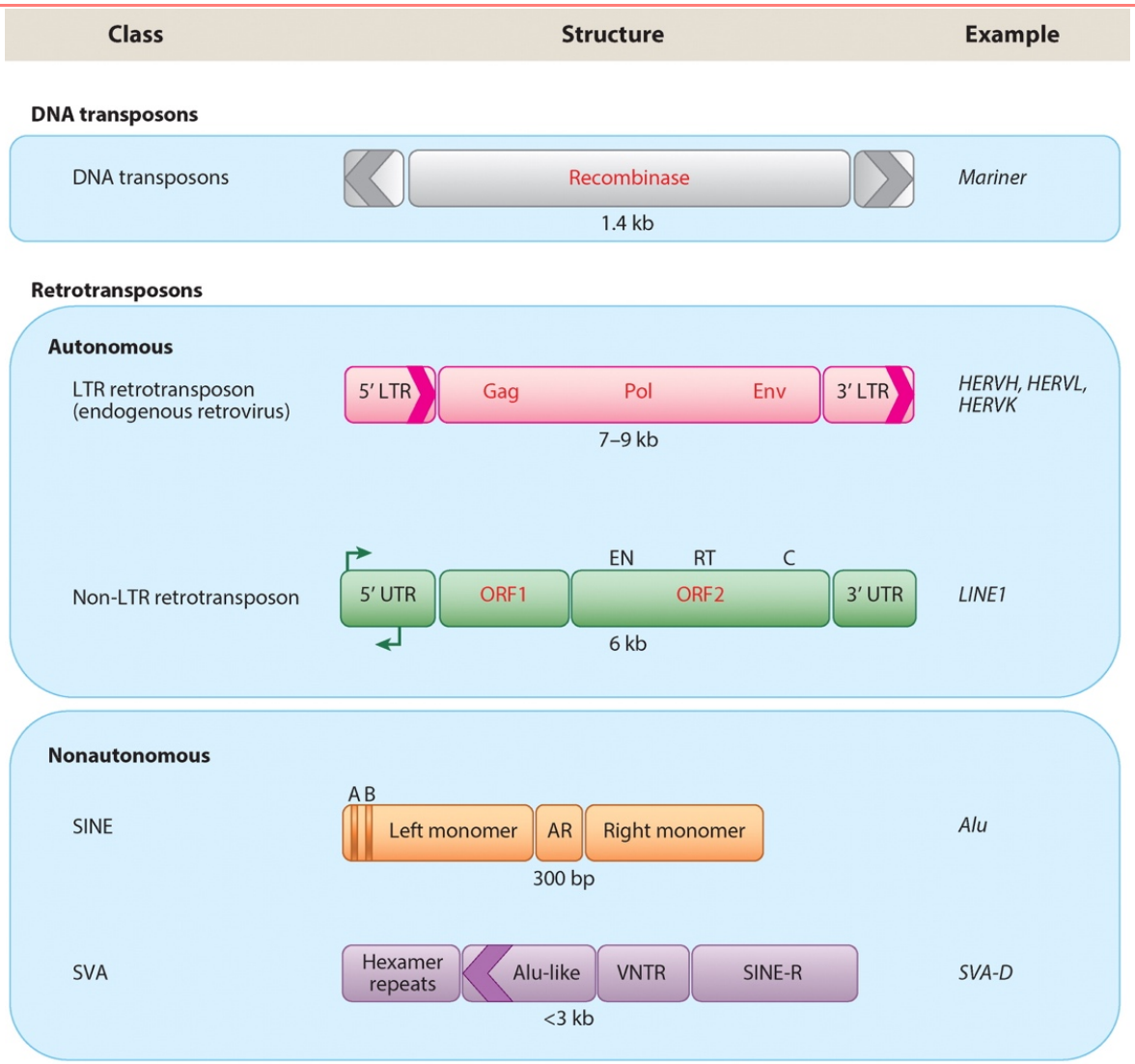


Figure 32: Classification of transposable elements.

a. Overall classification of human transposable elements. Percentage of the overall genome is given in orange. LTR, long-terminal repeat; ERV, endogenous retroviruses; SINE, short-interspersed nucleotide elements; SVA, SINE-VNTR-*Alus*; LINE, long-interspersed nucleotide elements. **b.** Chromosomal location of DNA repetitive elements and satellites classification for mouse and human (adapted from Crichton et al., 2014).

DNA transposons accounts for less than 3 % of the genome and performs transposition via a ‘cut-and-paste’ mechanism without RNA intermediate. On the other hand, retrotransposons amplify via a ‘copy-and-paste’ mechanism and rely on a RNA transcript which is retro-transcribed by a reverse transcriptase and further integrated into the genome (Stocking and Kozak, 2008). Retrotransposons can be further divided in ‘LTR-‘ and ‘Non-LTR’-containing retrotransposons (**Figure 32**). The Non-LTR retrotransposons are the only class known to be still active in human and might lead to one new germ line integrant every 50 births (Castro-diaz et al., 2015; Beck et al., 2011). The Non-LTR family is sub-composed of autonomous and non-autonomous retrotransposons. The autonomous Non-LTR retrotransposons include LINEs (long-interspersed nucleotide elements). LINEs encode a reverse transcriptase and a nuclease essential for transposition (ORF1 and 2, see **Figure 33**). The Non-autonomous are composed of the SINEs (short-interspersed nucleotide elements) and the hominid specific SVAs (SINE-VNTR-Alus). As their name indicate, they rely on other retro-transcription machineries (e.g. coded by LINEs). The LTR-retrotransposons are commonly called endogenous retroviruses (ERVs) as they are derived from ancient exogenous retrovirus infections (see **3.2**) and they are coding for the viral proteins *gag, pol* and *env* which makes them autonomous for retrotransposition (**Figure 33**). To note, only a small proportion of DNA sequences are coding for full length TEs and have thus the potential to transpose.



Friedli M, Trono D. 2015.
Annu. Rev. Cell Dev. Biol. 31:429–51

Figure 33: Main structural characteristics of the transposable elements (from Friedli and Trono, 2015).

An example of each class of transposable elements is given together with the average genome length. DNA transposons, LTR- and Non-LTR retrotransposons are autonomous as they encode the proteins necessary for their transposition (shown in red). LINE, long interspersed nuclear element; LTR, long terminal repeat; ORF, open reading frame; SINE, short interspersed nuclear element; SVA, SINE–VNTR–Alu; UTR, untranslated region.

3.1.2. Role in evolution

With their ability to transpose, TEs constitute a threat for the genome architecture and stability. For instance, as shown in **Figure 34**, TEs can have a plethora of impacts on the genome, such as disrupting existing genes, triggering deletions or duplications. As a result, TEs can lead to diseases. For example, more than 60 pathogenic mutations have been attributed to LINE-1 retrotranspositions (Belancio et al., 2008, Goodier and Kazazian, 2008) and about 30 monogenic disorders have been attributed to non-autonomous retrotransposons (Hancks and Kazazian, 2012). In mice, the ERV insertion in the agouti locus has become famous by leading to a change of fur color, by causing diabetes and obesity syndrome (Duhl et al., 1994).

Nevertheless, TEs are now recognized as essential contributors to evolution. Through the same mechanisms displayed in **Figure 34**, TEs can indeed lead to chromosomal rearrangement constituting adaptive benefits for the species (e.g. gene shuffling, genomic recombination, modulation of transcription). A well-known case of ERVs having participated in evolution is syncytin, a protein essential for placenta development. Indeed, both mice and human have two genes coding for syncytin, derived from the *env* gene of ERVs (Dupressoir et al., 2012). To note, even if the mouse and human syncytin have a similar function, they are not at orthologous positions, suggesting that the ERVs of origin were independently co-opted. In addition, TEs have also been proposed to have additional roles in development such as in neuronal diversity during brain development (Muotri et al., 2005) or in cell-fate regulation in placental mammals (Macfarlan et al., 2012)

In summary, TEs constitute a powerful and essential motor of evolution and hosts co-evolve with them through a fine balance between threat and benefit.

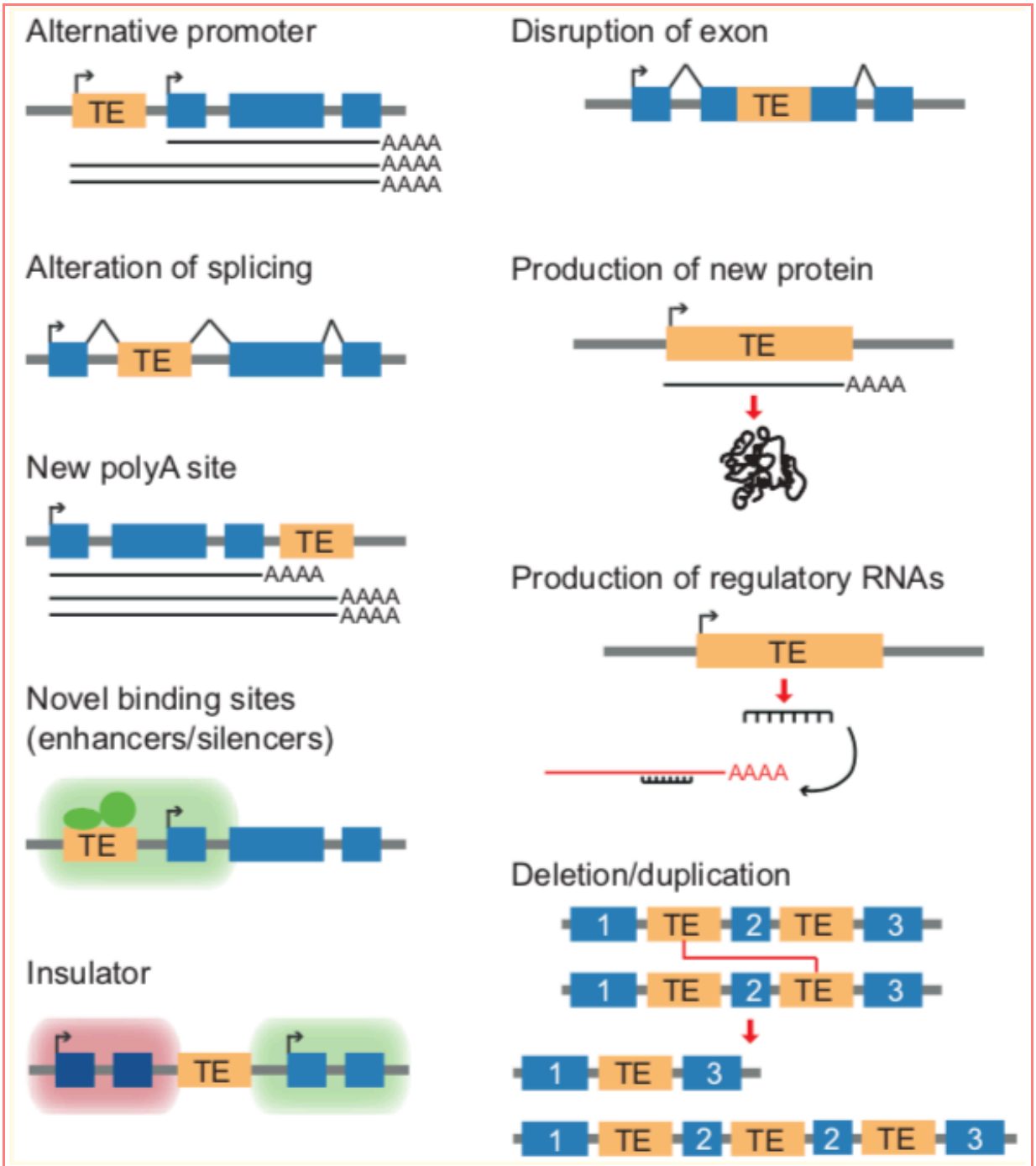


Figure 34: TE impacts on the host genomes (from Ecco et al., 2017)
TE often bear promoters, enhancers, suppressors, insulators, splice sites or transcriptional stop signals and can thus disrupt genes (via alternative splicing, truncation or insertion of new exons) or modify their expression (via promoter, enhancer or repressor effects). Due to their highly repetitive nature, TEs can also provoke recombination events that can lead to deletions, duplications, rearrangements or translocations. In addition, they can alter genome architecture via insulator sequences, long-range interaction modifications or they can provide entirely novel open reading frames.

3.2. Endogenous Retroviruses

The first ERVs were discovered in the late 1960s, beginning 1970s. The mouse genome is composed of 10 % of ERV sequences (8 % for human), mainly derived from ancient germ line infections from exogenous retroviruses (Boeke et al., 1997). Some ERVs have kept their ability of expression and replication even after millions of years within the host genome. Several ERVs are still active in mouse genome while almost all are extinct in human genome (potential exception of HERV-K), whereas mouse genomes still have many active ERVs (Friedli and Trono, 2015).

3.2.1. Three main families

ERVs are divided in 3 classes according to their similarities, notably of the reverse transcriptase, with modern exogenous retroviruses (**Figure 35**). The class I is composed of ERVs clustering with gamma- and epsilon-retrovirus. ERVs clustering with lentivirus, alpha-, beta-, and delta-retroviruses are termed Class II, and those that cluster with spumaviruses are termed Class III (Stocking and Kozak, 2008).

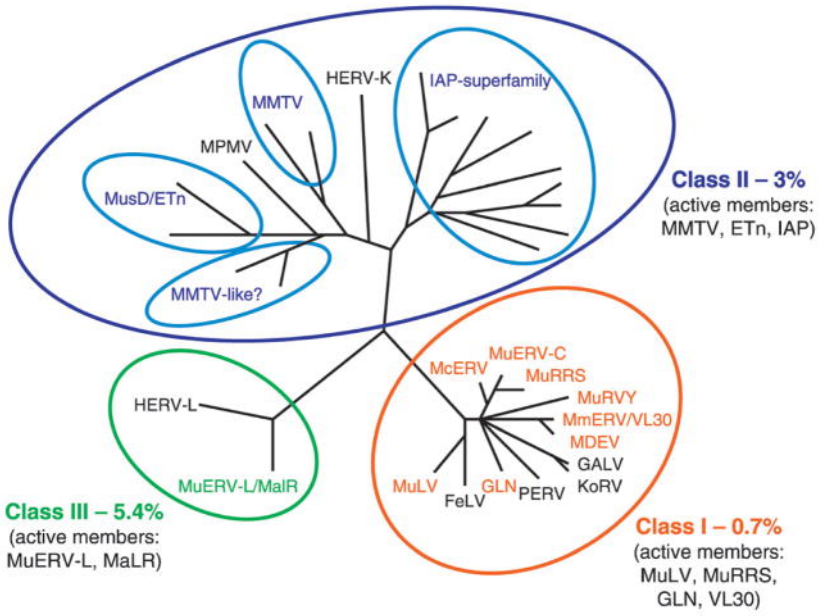


Figure 35: Phylogenetic analysis of mouse ERV reverse transcriptase leading to 3 ERV classes (from Stocking and Kozak, 2008).
ERVs from species other than mouse are indicated in black letters. Non-autonomous elements, such as VL30 (Class I), ETn (Class II), and MaLR (Class III) are listed with their presumed parental ERVs, as they do not contain RT domains.

ERVs are further classified in families (e.g. IAP - intracisternal A-particle). Even though homology can be found between mouse and human ERV families (**Figure 35**), greatest care should be taken while drawing conclusions. Indeed, mouse and human share a common ancestor approximately 100 million years ago. Since then, ERVs activities have evolved on different paths, leading to different ERVs families, located at different genomic loci, thus different classification in mouse and human. In addition, considerable confusion exists while naming ERV families and individual loci (Blomberg et al., 2009). ERVs are composed of repetitive sequences, which makes the sequencing alignment challenging. Accurate and locus specific analysis are thus difficult. Moreover, the lack of consensus in a precise classification led to different denominations for the same elements among different studies.

3.2.2. Mechanism of regulation

ERVs can be highly detrimental to their host, but are also functional components of the genome, so mechanisms of regulation are of main importance to maintain the fine balance of expression between the need of somatic and genomic diversity and the risk of disease and mutations. In mouse, several studies have shown that ERVs expression is finely tuned in early development with waves of activation and repression, through specific yet diverse mechanisms.

3.2.2.1. KRAB-ZFP/KAP1 system

TEs are recognized by a large group of tetrapod specific protein repressors: the KRAB-ZFPs (Krüpell-associated box- containing zinc finger proteins). KRAB-ZFPs are encoded by hundreds of genes in mice and humans, with more than 350 members (Huntley et al., 2006). They bind TEs thanks to their C-terminal zinc fingers domain and recruit the KRAB-associated protein 1 (KAP1, also named TRIM28) by their N-terminal domain. KAP1 further constitutes a scaffold for the recruitment of a heterochromatin-inducing machinery composed of the histone methyltransferase SETDB1 (also known as ESET), the histone deacetylase-containing complex NuRD, the heterochromatin protein 1 (HP1), and the DNA methyltransferases (**Figure 36**, Ecco et al., 2017).

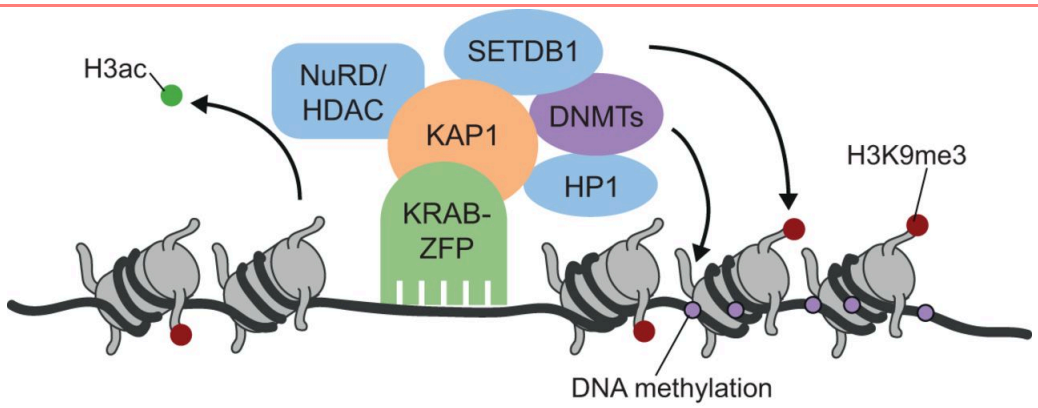


Figure 36: The KRAB-ZFP/KAP1 repressor complex (from Ecco et al., 2017)

KRAB-ZFPs (green) bind to DNA via their zinc fingers and recruit KAP1 (orange) via their KRAB domain. KAP1 then recruits a repressor complex, leading to heterochromatin formation through histone methylation (H3K9me3), DNA methylation, histone deacetylation (H3ac), and transcriptional silencing. DNMT, DNA methyltransferase; H3ac, acetylated histone H3; HDAC, histone deacetylase; HP1, heterochromatin protein 1; KAP1, Krüppel-associated box (KRAB)-associated protein 1; KRAB-ZFP, KRAB-zinc finger protein; NuRD, nucleosome remodeling deacetylase complex; SETDB1, SET domain bifurcated 1.

A first model emerged assuming that the KRAB/KAP1 system was responsible of the irreversible silencing of TEs during early development. Indeed, KRAB/KAP1-mediated TEs repression has been described in mouse and human embryonic stem cells as well as in early embryos (Yang et al., 2017; Guo et al., 2017; Theunissen et al., 2016; Wolf et al., 2015; Göke et al., 2015; Guo et al., 2014a; Smith et al., 2014; Turelli et al., 2014; Castro-Diaz et al., 2014; Rowe and Trono, 2011; Matsui et al., 2010; Rowe et al., 2010; Wolf and Goff, 2009). Indeed, KAP1 deletion led to ERVs overexpression in mESC, notably IAP, MusD, MEVL and MERVK. IAPs were also highly overexpressed in KAP1 knockout embryos, but not in embryonic fibroblasts (Rowe et al., 2010). Importantly, KRAB/KAP1 recruitment has been shown to result in TEs repression through DNA methylation during early development, while this repression was rather mediated by histone modifications in differentiated tissues (e.g. H3K9me3; Quenneville et al., 2011 and 2012). A recent work reported ERVs de-repression under KAP1 deletion in neural progenitor cells, together with a loss of H3K9me3 (see **3.2.2.3**) and activation of adjacent genes (Fasching et al., 2015). Coluccio et al. (2018) also demonstrated that the KZFP/KAP1 system is important to

preserve H3K9me3 and DNA methylation at TEs in the demethylated landscape of naïve mESC.

However, recent works rather suggest that KRAB-ZFPs play a far more elaborated role by contributing to the domestication of TE rather than being only in charge of their definitive silencing (Imbeault et al., 2017; Ecco et al., 2016). For example, KRAB/KAP1 plays a role in gene expression regulation in adult tissue by using TE-based platforms. The KRAB-ZFPs/TEs interplay is thus seen as a potential regulator of gene expression in somatic tissue. Interestingly, elevated levels of TE activity as well as high expression of KRAB-ZFPs have been reported in human brain (Erwin et al., 2014; Imbeault et al., 2017). KRAB-ZFPs would not only constitute arms aimed at TE silencing but rather elaborated instruments for their domestication, and a powerful tool for selective adaptation and turnover of transcriptional networks (Ecco et al., 2016; Pontis et al., 2019).

In addition to its role in silencing the endovirome, KAP1 has recently been shown to contribute to genome stability by taking part in DNA repair or in the maintenance of heterochromatin during DNA replication (Jang et al., 2018).

In addition to the KZFP/KAP1 system, DNA methylation has also been shown to play an important role in ERVs regulation.

3.2.2.2. Implication of DNA methylation

Several evidences reported the importance of DNA methylation in ERV silencing. Dnmt1 loss in mouse early development led to IAP loss of DNA methylation and strong upregulation (Kurihara et al., 2008; Walsh et al., 1998). However, DNMT triple knockout did not lead to ERV de-repression in mESC (Hutnick et al., 2010; Matsui et al., 2010; Tsumura et al., 2006), rather suggesting that DNA methylation is not important for ERV silencing in early development (mESC). In addition, Papin et al. (2017) reported a highly dynamic methylation of CG-rich LTR-retrotransposons (e.g. IAP, ERVK) during differentiation. They proposed a model in which IAP/ERVK methylation is highly dynamic and dependent on TET/TDG activities in mESC and become fully and stably methylated in mouse embryonic fibroblasts. Besides, many evidences were in line with the importance of DNA methylation for ERV silencing in somatic cells (Rowe et al., 2013a). DNA methyltransferase inhibitors have been shown

to induce ERV demethylation and upregulation in tumor cells (Chiappinelli et al., 2015). In summary, DNA methylation at ERV is highly dynamic upon differentiation and is thought to participate in ERV silencing in somatic cells. On the other hand, ERV silencing in early development might rather be controlled by SETDB1-mediated H3K9me3.

3.2.2.3. Role of H3K9me3

ERVs are marked by H3K9me3 and H4K20me3 in mESC. Matsui et al. (2010) showed that SETDB1 H3K9 methyltransferase and KAP1 are required for H3K9me3-mediated silencing of ERVs in mESC. They proposed that SETDB1-mediate H3K9me3 deposition is required for silencing ERVs during early embryogenesis, period during which DNA methylation is under high reprogrammation. In accordance with this study, Rowe et al. (2013b) showed an upregulation of IAPs following KAP1 or SETB1 removal in mESC without change in their DNA methylation status, suggesting that H3K9me3 is primarily responsible for ERV silencing in mESC. In addition, Deniz et al. (2018) demonstrated that SETDB1 has a more prominent role in ERV silencing in naïve mESC compared to primed mESC. Indeed, naïve mESC (cultures in 2 kinase inhibitors) have a globally hypomethylated genome that closely resemble the one of inner cell mass cells in pre-implantation blastocysts while primed mESC (serum grown) have a hypermethylated genome similar to the one of post-implantation embryos (Wu and Zhang, 2014). This highlight the effect and thus importance of the type of culture used for mESC (2i vs. serum) and underline the interplay between DNA methylation and H3K9me3 roles. Indeed, ERVs are strongly deregulated in the absence of H3K9me3 in the hypomethylated background of the naïve mESC.

Recently, KAP1 phosphorylation on serine 473 has been reported to be required for the maintenance of H3K9 methylation (Jang et al., 2018). Besides, H3K27 methylation has been shown to overlap with H3K9me3 in mESC (Mikkelsen et al., 2007) and this is consistent with data implicating Polycomb group proteins in ERV silencing (Leeb et al., 2010).

Recent work has described a reactivation of ERVs in several SETDB1-knockout somatic cells, suggesting a more general role of H3K9me3 in ERV silencing in differentiated cells (Collins et al., 2015; Wolf et al., 2015; Kato et al., 2018).

Taking into accounts the previous studies, ERV silencing can be mediated by both DNA methylation and histone modification and the balance between those two mechanisms is mostly depending on the context (e.g. time of development).

3.2.2.4. Role of H3.3

Histone variant H3.3 has been shown enriched at class I and class II ERVs, especially at early transposon (ETn)/MusD family and IAPs (Elsässer et al., 2015). Recruitment of DAXX, H3.3 and KAP1 to ERVs has been reported to be co-dependent and occurring upstream of SETDB1 recruitment. Upon H3.3 depletion, ERV-associated H3K9me3 is reduced, suggesting a link between H3.3 and H3K9me3 at ERVs. In addition, H3.3 deletion led to IAP de-repression and dysregulation of adjacent genes and to a reduction of KAP1 and DAXX recruitment at ERVs. Elsässer et al. (2015) thus proposed H3.3 as a key player in the control of ERV expression in mESC. In line with previous studies, they also showed that H3.3 enrichment at IAPs and ETns is lost together with H3K9me3 during differentiation from mESC to neural progenitor cells.

This work suggesting a role of H3.3 in ERVs regulation has opened an active debate in the field (Wolf et al., 2017; Elsässer et al., 2017) and will need further confirmation.

3.2.3. ERV in cancer

Recent evidences reported ERVs deregulation in cancer. ERVs have indeed been shown to be activated in transformed cells of various cancers (Hancks and Kazazian, 2012; Criscione et al., 2014; Bannert et al., 2018). But the role and consequences of ERVs activation remains poorly understood and whether this deregulation is driver or passenger is an open question. In addition, as presented in **2.3.3.7**, the introduction of H3.3 K27M mutation led to a de-repression of ERVs in a pHGG mouse model (Krug et al., 2019). The latter is potentially confirming the importance of H3.3 in ERV silencing and the implication or ERV de-repression in cancer development.

Aims of the study

Pediatric high-grade gliomas are a very aggressive and deadly cancer for which no treatment is currently available. Since their discovery in 2012, H3.3 point mutations K27M and G34R/V have been highlighted as driver for a subset of the pHGG and highly studied. Nevertheless, no clear consensus has emerged about the molecular impact of H3.3 mutations on chromatin landscape and transcription. Due to the heterogeneity in the tumor backgrounds between patients and the lack of proper clinical controls, understanding the role of H3.3 mutations using tumor tissue is highly challenging. Thus, the development of models to study the role of H3.3 K27M and G34R/V in tumorigenesis and their direct impact on chromatin is of main importance. In addition, most of the studies have focused on the more prevalent K27M mutation leaving G34R/V mutations quite unstudied.

The aims of my PhD project were first to understand the impact of H3.3 K27M and G34R mutations on its enrichment in chromatin and on global transcription. To this end, we designed an endogenously tagged H3.3 WT, K27M and G34R mESC model further used to develop a conditional knock-in mouse model.

The first chapter of this manuscript will present the strategy for a conditional knock-in tagged H3.3 WT, K27M or G34R mouse model and the characterization of the mESC designed in this project. As the conditionality of the mutation was not obtained in mESC, we further derived a constitutively tagged H3.3 WT, K27M or G34R mESC model.

The second chapter will address the impact of H3.3 mutations on its enrichment at active chromatin and at DNA repetitive elements and on global transcription in mESC.

In a third chapter, I investigated the role of H3.3 and the impact of its mutations on neural early differentiation.

II. Material and Methods

1. Experimental Models and Subject Details

1.1. Mouse strains

Both H3F3A^{Tag} and H3F3B^{Tag} cKO-KI mouse lines were established at the Phenomin-iCS (Phenomin – Institut Clinique de la Souris, Illkirch) and were a kind gift from Dr. Stefan Dimitrov (IAB Grenoble). Both lines are harboring a FLAG–FLAG–HA epitope sequence inserted in frame with the N-terminus of either H3.3A of H3.3B. The mice used in the experiments exhibit C57BL/6 genetic background. Mice were housed in the mouse facility of the Institute of Genetics and Molecular and Cellular Biology (IGBMC, Illkirch, France, agreement number C6721837, registered protocol APAFIS#15388-2018060709154166).

1.2. Cell lines

1.2.1. Cell lines development

Wildtype and endogenously tagged (FLAG-FLAG-HA)-H3.3A mouse Embryonic Stem Cell (mESC) lines were derived from an already characterized H3.3A mouse line (kind gift from Dr. Stefan Dimitrov, IAB Grenoble, see protein detailed sequence below) by Phenomin (Institut Clinique de la Souris, Illkirch). Briefly, heterozygous tagged H3f3a mice (H3F3A$^{Tag/+}$) were crossed in order to derivate H3F3A$^{+/+}$, H3F3A$^{Tag/+}$ and H3F3A$^{Tag/Tag}$ mESC from the same litter. Embryonic day 3.5 (E3.5) blastocysts were isolated and Inner Cell Mass expansion was performed, followed by ESC expansion. Endogenously tagged (FLAG-HA) mutant (K27M or G34R) H3.3A expressing mESC lines were developed by Phenomin (ICS, Strasbourg) by homologous recombination in wildtype mESC (see protein detailed sequence below). All mESC lines have the same C57BL/6 background. All the mESC lines were genotyped, tested negative for mycoplasma contamination (PlasmoTest™, InvivoGen) and karyotyped by ddPCR as described in Codner et al. (2016).

H3.3 WT

MDYKDDDDKGTDYKDDDDKADYDIPTTARNYENLYFQGELQYPYDVPDYAGGAAR
TKQTARKSTGGKAPRKQLATKAARKSAPSTGGVKKPHRYRPGTVALREIRRYQKST
ELLIRKLPFQRLVREIAQDFKTDLRFQSAAIGALQEASEAYLVGLFEDTNLCAIHAKRV
TIMPKDIQLARRIRGERA

H3.3 K27M

MDYKDDDDKGGYPYDVPDYAARTKQTARKSTGGKAPRKQLATKAARMSAPSTGG
VKKPHRYRPGTVALREIRRYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSAAIGAL
QEASEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA

H3.3 G34R

MDYKDDDDKGGYPYDVPDYAARTKQTARKSTGGKAPRKQLATKAARKSAPSTGR
VKKPHRYRPGTVALREIRRYQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSAAIGAL
QEASEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA

FLAG tag is shown in orange and HA tag in green.

1.2.2. mESC Cell culture

mESC were routinely maintained on a layer of mitomycin-inactivated mouse embryonic fibroblasts (feeder cells) in a daily-changed stem cell medium containing KnockOut-DMEM (Gibco™, Thermo Fisher Scientific, Ref 10829-018), 15 % KnockOut™ Serum Replacement (KSR, Gibco™, 10828-028), 1x Leukemia Inhibitory Factor (IGBMC Cell culture facility), 1x GlutaMAX™ (Gibco™, 35050-38), 1x MEM Non-Essential Amino Acids (NEAA, Gibco™, 11140050), 0,1 mM mercaptoethanol and 40 µg/mL gentamicin (Kos medium). Dissociation was performed with a trypsin-EDTA solution (Thermo Fisher Scientific, 25200-072). mESC were cultivated at 37°C and 5 % carbon dioxide.

All lines were regularly tested negative for mycoplasma contamination, and karyotyped after amplification by metaphase spreading and Giemsa coloration (protocol from Phenomin ICS, Strasbourg).

1.3. Differentiation of mESC in Neural Stem Cell

The differentiation protocol has been adapted from Colombo et al., 2006. Undifferentiated mES cells cultured in Kos HG medium (high-glucose-DMEM containing 15 % ES cell tested-FCS, 1 % GlutaMAX, 1x NNEA, 0,1 mM mercaptoethanol, 1x LIF, 40 µg/mL gentamicin) on p60 feeder layers were incubated in a dispase solution (Merck, SCM133) for 30 min at 37°C to allow the detachment of cells as whole colonies. The entire colonies were collected, pelleted by gravity and resuspended into differentiation medium (high-glucose-DMEM containing 15 % KSR, 1 % GlutaMAX, 0,1 mM mercaptoethanol, 1x NNEA, 40 µg/mL gentamicin). The colonies were transferred into an untreated 60 mm dish (Nunc™ - Thermo Fisher, 150340). Colonies rounded up forming embryonic bodies (EB) and the differentiation medium was changed after 48h. After 4 days in untreated 60 mm dish, EB were harvested, pelleted by gravity and resuspended in growth medium for Neural Stem Cells (NSC) consisting of EUROMED-N MEDIA (Euroclone, ECM0883L) containing 2 mM L-glutamine, 0.6 % glucose, 9.6 µg/ml putrescine (Sigma Aldrich, P5780), 6.3 ng/ml progesterone (Sigma Aldrich, P8783), 5.2 ng/ml sodium selenite, 0.025 mg/ml insulin, 0.1 mg/ml apo-transferrin, 0.2 % BSA, 20 ng/ml Epidermal Growth Factor (EGF, Sigma Aldrich, E1257), 10 ng/mL FGF-2 (Peprotech, 100-18B) and replated on 35 mm Matrigel™-coated culture dishes. EB attached and progressively spread into the plate. After 4 days, cells were treated 5 min with a trypsin solution and replated in a 60 mm Matrigel™-coated culture dish. mESC-derived NSC were validated by anti-nestin immunofluorescence and by RT-PCR for specific markers. Karyotyping was also verified for each mESC-derived NSC clone.

1.4. Patient derived cell lines

All patient material was collected after informed consent and subject to local research ethics committee approval. Total RNA from H3.3G34R pHGG (n=3), H3.3G34V pHGG (n=1) and H3.3WT pHGG (n=4) derived cell lines were obtained from Dr. Chris Jones (The Institute of Cancer Research, London, UK). A full description of all the samples included are provided hereafter:

Cell line	Histone mutation	Location	Diagnosis	Gender	Age	*TP53*	Others	Others
CXJ001	WT	Hemispheric	GBM	M	3,9	p.R282fs	PTEN p.N276D	MYCN amp
CXJ008	WT	Hemispheric	GBM	M	14	p.R209fs	NF1 p.L1246fs	CDKN2A/B HOM DELETION
HSJD-GBM-001	WT	Hemispheric	GBM	F	10,9	p.G245S		PDGFRA_gain; CDKN2A_loss
QCTB-R006	WT	Hemispheric	GBM	M	9,5			CDKN2A/B HOM DELETION
CHOP-GBM-0001	H3.3_G34R	Hemispheric	supratentorial PNET	M	17			
HSJD-GBM-002	H3.3_G34R	Hemispheric	GBM	M	14	p.P278T	ATRX p.R666* & splice site c.5787-8_5787-5delGTTT	PDGFRA p.D842V
KNS42	H3.3_G34V	Hemispheric	GBM	M	14	p.R342*	ATRX p.Q891E (SNP DB rs3088074)	
OPBG-GBM-001	H3.3_G34R	Hemispheric	GBM	M	12	p.G245S		CDKN2A/B HET DELETION

1.5. Use of published datasets

To validate our transcriptional results in patients harboring H3.3 K27M mutation, we reanalyzed the previously published data sets deposited in GEO under accession number GSE128745 (Krug et al., 2019). Total RNA-seq from 32 pHGG from patients harboring H3.3 WT or K27M were reanalyzed.

We also reanalyzed total RNA-seq data from Silveira et al. (2019) deposited in GEO under accession number GSE115875. Total RNA-seq from 2 patient-derived xenografts (X37 and XSUVI) under shH3f3a or shCtrl treatments were analyzed.

Both these data sets were reanalyzed independently as described in the “repeat analysis” section.

2. Method details

2.1. Antibodies

Primary antibodies. Rat monoclonal anti-HA (Sigma Aldrich, 11867423001 – western blotting (WB) 1:1,000 / Immunofluorescence (IF) 1:200); Mouse monoclonal anti-β-Actin (Sigma Aldrich, A2228 – WB 1:10,000); Mouse monoclonal anti-Nestin (Abcam, ab6142 – IF 1:200).

Secondary antibodies. Goat anti-rat IgG Alexa Fluor 488 (Invitrogen, A11006 – IF 1:500); Goat anti-mouse IgG Alexa Fluor 568 (Invitrogen, A11004 – IF 1:500); Goat anti-rat IgG-Peroxidase (Sigma Aldrich, A9037 – WB 1:10,000); Goat anti-mouse HRP (Sigma Aldrich, A2304 - WB 1:10,000).

2.2. Immunofluorescence

mESC cells were cultured on coverslips in 6-wells of 24-wells plate for at least 18 hours. Cells were washed twice 5 min with cold DPBS (HyClone, GE Healthcare) and fixed with 4 % paraformaldehyde solution in PBS (Electron Microscopy Sciences) for 20 min at room temperature (RT). The following steps were all performed at RT, and each wash performed for 5 min. After 2 washes with PBS, 2 washes with PBS-0.1 M Glycine pH 8.5, 1 wash with PBS- 1 % BSA- 1 % FBS, cells were permeabilized with PBS-0.05 % Triton X-100 for 15 min. After 2 washes with cold PBS, blocking was performed in PBS- 1 % BSA- 1 % FBS for 1h, followed by primary antibody incubation in PBS- 1 % BSA- 1 % FBS, 3 washes with PBS- 1 % BSA- 1 % FBS and secondary antibody incubation in PBS- 1 % BSA- 1 % FBS for 1h. After 3 washes with PBS- 1 % BSA- 1 % FBS and 1 wash with PBS, coverslips were incubated with DAPI 1 µg/mL in PBS for 10 min and washed with PBS before mounting on microscope slides using Aqua-Poly/Mount mounting medium (Polysciences, Inc.). Samples were examined and photographed using a Leica DM 4000 B microscope equipped with a Photometrics CoolSNAP HQ2 camera. No labelling was observed in the absence of primary antibodies (control samples) and no evidence of cross-reactivity was observed.

2.3. Western Blotting

Samples were separated by SDS-PAGE on 12 % gels and transferred to an Immobilon-P PVDF membrane (Merck). Membrane blocking was performed with 5 % skimmed milk in PBS supplemented with 0.1 % Tween 20 (PBST) overnight at 4°C. Membranes were incubated with primary antibody (see Antibodies section) in 5 % skimmed milk in PBST 1h at RT. Membranes were washed 3 times in PBST, and incubated with HRP-conjugated secondary antibody 1h at RT (see Antibodies section). Membranes were then washed 3 times in PBST, one time in PBS, and the signal was resolved with Immobilon® Forte Western HRP Substrate (Merck) and detected using Amersham Hyperfilm™ ECL (GE Healthcare) on a Kodak X-OMAT 3000RA Processor.

2.4. Preparation of Cytosolic, Nuclear Soluble and Nuclear Insoluble extracts

Extracts were prepared using a modification of the Dignam protocol (Dignam 1990), as described previously (Drané et al., 2010). Briefly, mESC were lysed in hypotonic buffer (10 mM Tris-HCl at pH 7.65, 1.5 mM MgCl2, 10 mM KCl) and disrupted by Dounce homogenizer. The cytosolic fraction was separated from the pellet by centrifugation at 4°C. The pellet was resuspended in sucrose buffer (20 mM Tris-HCl at pH 7.65, 60 mM NaCl, 15 mM KCl, 0.34 mM sucrose, 0.15 mM spermine, 0.5 mM spermidine). The nuclear-soluble fraction was obtained by addition of high-salt buffer (to get a final NaCl concentration of 300 mM: 20 mM Tris-HCl at pH 7.65, 1.5 mM MgCl2, 0.2 mM EDTA pH 8, 900 mM NaCl, 25 % Glycerol). After 30 min of incubation at 4°C under rotation, the nuclear-soluble fraction was separated from the pellet by centrifugation at 4°C. The pellet was resuspended in sucrose buffer and treated with micrococcal nuclease (Sigma Aldrich, N3755 – 2.5 U/g of cells) 10 min at 37°C. Digestion was stopped by addition of 4 mM EDTA pH 8 and incubation on ice for 5 min. The digested sample was then sonicated 3 x 1 min on ice (MSE Soniprep 150 Plus, amplitude 10). Cytosolic, nuclear soluble and nuclear insoluble extracts were supplemented with protease inhibitor (Roche, cOmplete™ Protease Inhibitor Cocktail) and ultra-centrifugated at 35,000 rpm 1h at 4°C (Beckman Optima L-90K ultracentrifuge, Rotor SW60 Ti). Lipid layer was discarded and extracts were used for immunoprecipitation. Chromatin concentration was determined (overnight RNAse

treatment, 1h of Proteinase K treatment, phenol/chloroform extraction followed by precipitation with sodium acetate/ethanol).

2.5. Tandem affinity purification

Tagged proteins were immunoprecipitated with anti-FLAG® M2 Affinity gel (Sigma Aldrich, A2220), eluted with FLAG peptide (0.5 mg/mL, IGBMC synthesis platform), further affinity-purified with anti-HA agarose (Sigma Aldrich, A2095), and eluted with HA peptide (1 mg/mL, Millipore, I2149). The HA and FLAG peptides were first buffered with 50 mM Tris-Cl (pH 8.5), then diluted to 4 mg/mL in TGEN 150 buffer (20 mM Tris at pH 7.65, 150 mM NaCl, 3 mM MgCl2, 0.1 mM EDTA pH 8, 10 % glycerol, 0.01 % NP40), and stored at −20°C until use. Between each step, beads were washed in TGEN 150 buffer.

2.6. Mass spectrometry analysis

Cells were crosslinked with 0.4 % paraformaldehyde in PBS for 10 min at RT. The reaction was stopped by incubation in 0.15 M glycine pH8 in PBS for 10 min at RT. Cells were rinsed twice with 1x PBS, scraped, pelleted and subjected to tandem affinity purification (see section **2.5**). Complexes were resolved by SDS-PAGE and stained using the Silver Quest kit (Invitrogen), and analyzed by microcapillary LC/MS/MS by the Taplin Biological Mass Spectrometry Facility (Harvard Medical School, USA).

2.7. Native Chromatin Immunoprecipitation and Sequencing (ChIP-seq)

For FLAG-HA native ChIP-seq, DNA was purified from the elution of the tandem affinity purification (see section **2.5**) on the nuclear insoluble fractions. DNA was purified by phenol-chloroform extraction followed by ethanol precipitation.

2.8. ChIP-seq Library Preparation and Sequencing

ChIP-seq libraries were performed using Diagenode MicroPlex Library Preparation kit v2 (Diagenode) following Instruction Manual (version v.2 02.15). The library was sequenced on Illumina Hiseq 4000 sequencer as Single-Read 50 base reads following Illumina's instructions. Image analysis and base calling were performed using RTA (version 2.7.7) and bcl2fastq (version 2.17.1.14). Adapter dimer reads were removed using DimerRemover (https://sourceforge.net/projects/dimerremover/). Read quality was checked with FastQC (version 0.11.2) and putative contamination assessed by FastQScreen (version 0.5.1). Alignment was performed onto the mm9 assembly of *Mus musculus* genome using Bowtie v1.0.0, followed by quality control using FastQC (version 0.11.2) and FastQScreen (version 0.5.1).

2.9. RNA-seq Library Preparation and Sequencing

Total RNA was extracted from subconfluent cells or tissue using TRI Reagent® (MRC, TR118) following manufacturer's protocol. The library was created using the TruSeq Stranded Total RNA SamplePrep kit (Illumina) and sequenced on Illumina Hiseq 4000 sequencer as Single-Read 50 base reads following Illumina's instructions. Reads were preprocessed in order to remove adapters, polyA and low-quality sequences (Phred < 20) and reads shorter than 40 bases were discarded for further analysis (cutadapt version 1.10). Reads were mapped to spike sequences using bowtie2 (version 2.2.8) and reads mapping to spike sequences were removed for further analysis. Reads were mapped onto the mm9 assembly of *Mus musculus* genome using STAR (version 2.5.3A). Quality control on the reads was performed with FastQC (version 0.11.5) and quality control on the alignments with RSeqQC (version 2.6.4). To note, total RNA from mESC+feeders were compared to total RNA from mESC subjected to differential plating in order to eliminate the feeders. Except the response to the dissociation stress, both methods gave the same results so all the replicates were merged for the differential expression analysis.

2.10. RT-PCR

Total RNAs were purified from subconfluent cells using TRI Reagent® protocol. 3 µg of RNA were subjected to reverse transcription using oligodT primers (Promega) and the SuperScript IV reverse transcriptase (Invitrogen). PCR were performed with the following oligonucleotide pairs:

Name	Sequence 5'->3'
1-mutFH_Fw	5'-ACTACAAAGACGATGACGACAAGGGAGGCT-3'
1-mutFH_Rv	5'-AGGCCTCACTTGCCTCCTGCAAAGCACCAAT-3'
2-cDNA-WT_Fw	5'-ATGGCTCGTACAAAGCAGACTGCCCGCAAA-3'
2-cDNA-WTpolyA_Rw	5'-AAATGTGGTATGGCTGATTATGATCCTCT-3'
3-Stop-3UTR_Rv	5'-TGAAATGTTTCCCCTCATAGTGGACTCTTA-3'

Name	Sequence 5'->3'	Reference	Size (bp)
ActB_Fw	5'-CTCTGGCTCCTAGCACCATGAAGA-3'	Stephens et al., 2011	200
ActB_Rv	5'-GTAAAACGCAGCTCAGTAACAGTCCG-3'		
Oct4_Fw	5'-CTCGAACCACATCCTTCTCT-3'	Colombo et al., 2006	313
Oct4_Rv	5'-GGCGTTCTCTTTGGAAAGGTG-3'		
Foxg1_Fw	5'-ACAAGAAGAACGGCAAGTACG-3'	Watanabe et al., 2016	108
Foxg1_Rv	5'-CATAGATGCCATTGAGCGTCA-3'		
Exm2_Fw	5'-GTCCCAGCTTTTAAGGCTAGAG-3'	Colombo et al., 2006	151
Exm2_Rv	5'-CTTTTGCCTTTTGAATTTCGTTC-3'		
Sox2_Fw	5'-TAGAGCTAGACTCCGGGCGATGA-3'	Huang et al., 2011	297
Sox2_Rv	5'-TTGCCTTAAACAAGACCACGAAA-3'		
Pax6_Fw	5'-CAGCTTCAGTACCAGTGTCT-3'	Colombo et al., 2006	461
Pax6_Rv	5'-GTCATTGGCAGAGTGAACACA-3'		

3. Quantification and Statistical Analysis

3.1. Analysis of ChIP-seq Data

Heatmaps and enrichment comparisons of the ChIP-seq data were performed using seqMINER v 1.3.3g (Ye et al. 2011), using data sets representing 20 million uniquely mapped reads of the pool of replicates. As reference coordinates, we used the RepeatMasker (rmsk) track (for repetitive elements) downloaded from the UCSC table browser or the Ensembl 67 database (limited to coding genes) of the mouse genome (mm9). Enhancer coordinates were determined as the distal peaks of H3.3 (>2kb from TSS) from the H3.3 ChIP-seq published by Chronis et al. (2017). The pooled samples densities were normalized in reads per million mapped reads (rpm) or normalized read counts. For visualization, either WIG files were generated using an in-house script, or H3K4me1/H3K4me3 ES data tracks, as part of the ENCODE (LICR) project, were displayed directly in the genome browser.

3.2. Analysis of RNA-seq Data

Gene expression quantification was performed from uniquely aligned reads using HTSeq-count (version 0.6.1p1), with annotations from Ensembl version 67 and “union” mode. Only non-ambiguously assigned reads have been retained for further analyses. Read counts were normalized across samples with the median-of-ratios method proposed by Anders and Huber (2010) to make these counts comparable between samples. Differential expression comparisons were performed using the Wald test proposed by Love et al. (2014) and implemented in the Bioconductor package DESeq2 version 1.16.1. P-values were adjusted for multiple testing using the Benjamini and Hochberg method (Benjamini and Hochberg, 1995).

3.3. Repeat analysis

Repeat analyses were performed as described in Papin et al. (2017) for both total RNA-seq and ChIP-seq data. Although, slight modifications were added for RNA-seq data in order to discriminate between the signal coming from genes expression and the one coming from repeat expression. For this analysis, reads were aligned to repetitive elements in two steps. In the first step, reads were aligned to the non-masked

mouse reference genome (NCBI37/mm9) using BWA v.0.6.2 (Li and Durbin 2009). For RNA-seq repeat analysis, reads which sense was of the same sense as overlapping transcript were removed. Prior to this step, genomic coordinates of transcripts were extended 3Kb upstream of TSS and 10Kb downstream of TTS to remove reads arising from transcriptional readthrough. Positions of the reads mapped uniquely to the mouse genome were cross-compared with the positions of the repeats extracted from UCSC (rmsk table in the UCSC data- base for mouse genome mm9), and reads overlapping a repeat sequence were annotated with the repeat family. In the second step, reads not mapped or multimapped to the mouse genome in the previous step were aligned to RepBase v.18.07 (Jurka et al. 2005) repeat sequences for rodent. Reads mapped to a unique repeat family were annotated with their corresponding family name. Finally, we summed up the read counts per repeat family of the two annotation steps. Data were normalized based upon library size. Differential analysis of repeat families was performed using the Wald test proposed by Love et al. (2014) and implemented in the Bioconductor package DESeq2 version 1.16.1. P-values were adjusted for multiple testing using the Benjamini and Hochberg method (Benjamini and Hochberg, 1995). For ChIP-seq samples, ChIP IP was compared to corresponding input sample(s). To avoid over- or underestimating fold enrichments due to low sequence representation, repeat families with less than 500 mapped reads per RNA or ChIP samples were excluded from further analysis. In this analysis, fold changes were computed as the log2 of normalized read counts of ChIP samples per repeat family divided by normalized read counts of matched input samples (average of the two or three replicates).

3.4. Full length LTR – closest gene association

To test whether LTR deregulation had an effect on the closest gene expression, the closest gene was associated to each full length LTR (>4 kb) by considering the smaller distance between the gene and LTR boundaries. Adjusted p-value for the gene differential expression between H3.3 mutant and WT was associated as well as the adjusted p-value from the LTR family differential expression. For each mutant, LTR were classified in deregulated (adjusted p-value ≤ 0.05) or not deregulated (adjusted p-value>0.05). They were then further classified according to the gene-LTR distance in the two categories: proximal (≤ 10 kb) or the others (>10 kb). The adjusted p-value for the gene differential expression analysis was plotted for each of the four categories

(LTR deregulated proximal or others, and LTR not deregulated proximal or others) and a Wilcoxon test was performed to determine whether the medians were significantly different.

3.5. Analysis of Me-DIP-seq Data

Data sets deposited in GEO under accession number GSE42250 were reanalyzed to determine 5-mC and 5-hmC enrichment in mESC on DNA repetitive elements (see 'repeat analysis' section **3.3**, Shen et al. 2013).

3.6. Timeseries analysis over differentiation

To test for differences in genes/repeat expression upon differentiation in WT cells, we tested read counts per gene/repeat using the likelihood ratio test implemented in the DESeq2 Bioconductor library (DESeq2 v1.6.3). Prior to the test, read counts per gene were normalized using the method implemented in DESeq2 and read counts per repeat family were normalized so that they would all have the same number of reads (normalized to the sample with the lowest raw read number). P-values were adjusted for multiple testing using the Benjamini and Hochberg method (Benjamini and Hochberg, 1995). Genes/repeats significantly changing over time (adjusted p-value ≤ 0.01) were then clustered with the mFuzz Bioconductor library v2.26.0 (Futschik and Carlisle, 2005).

III. Results and Discussion

Chapter 1: Mouse and mESC models for pHGG

1.1. cKI-H3.3 K27M or G34R mouse model

1.1.1. Strategy

As described earlier in the introduction, the choice of the model for pHGG study is of great importance. In collaboration with Phenomin (ICS, Strasbourg), we put in place a strategy to develop a mouse model expressing a conditional H3.3 K27M or G34R knock-in mutant. In order to closely recapitulate the patient context, only one allele of *H3f3a* gene was modified (**Figure 37**). The point mutation K27M (A**A**G>A**T**G) or G34R (**G**GG>**C**GG) was inserted in the first coding exon together with a FLAG-HA tag at the N-terminal side (Ex2). In addition, an untagged wildtype H3.3 cDNA with a 3' polyA terminator and flanked by two LoxP sites has been inserted in the first intron (between Ex1 and modified Ex2). Before any Cre recombinase expression, the mRNA splicing is predicted to be performed between non-coding Exon1 and WT H3.3 cDNA. The mRNA produced would thus be coding for WT untagged H3.3. After Cre recombinase expression, the WT cDNA-polyA sequence would be excised and the splicing would further be performed between non-coding Exon1, modified Exon 2, Exon 3 and Exon 4-3' UTR. This transcript would code for FLAG-HA-H3.3 K27M or G34R. This construct would allow to induce H3.3 mutant expression on a single *H3f3a* allele in a time and tissue specific manner using an inducible Cre recombinase under a tissue-specific promoter. Moreover, the expression and deposition of mutant H3.3 in chromatin would be specifically trackable thanks to the FLAG-HA tag. After successful homologous recombination, I characterized the validated mESC clones for blastocyst injection.

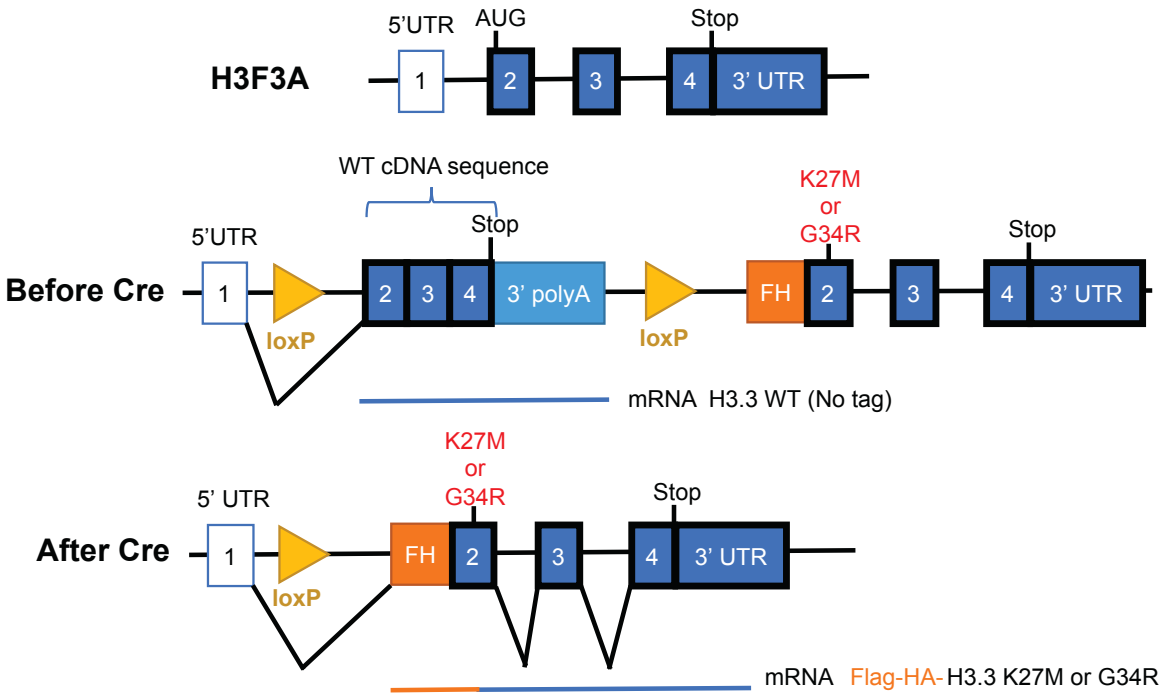
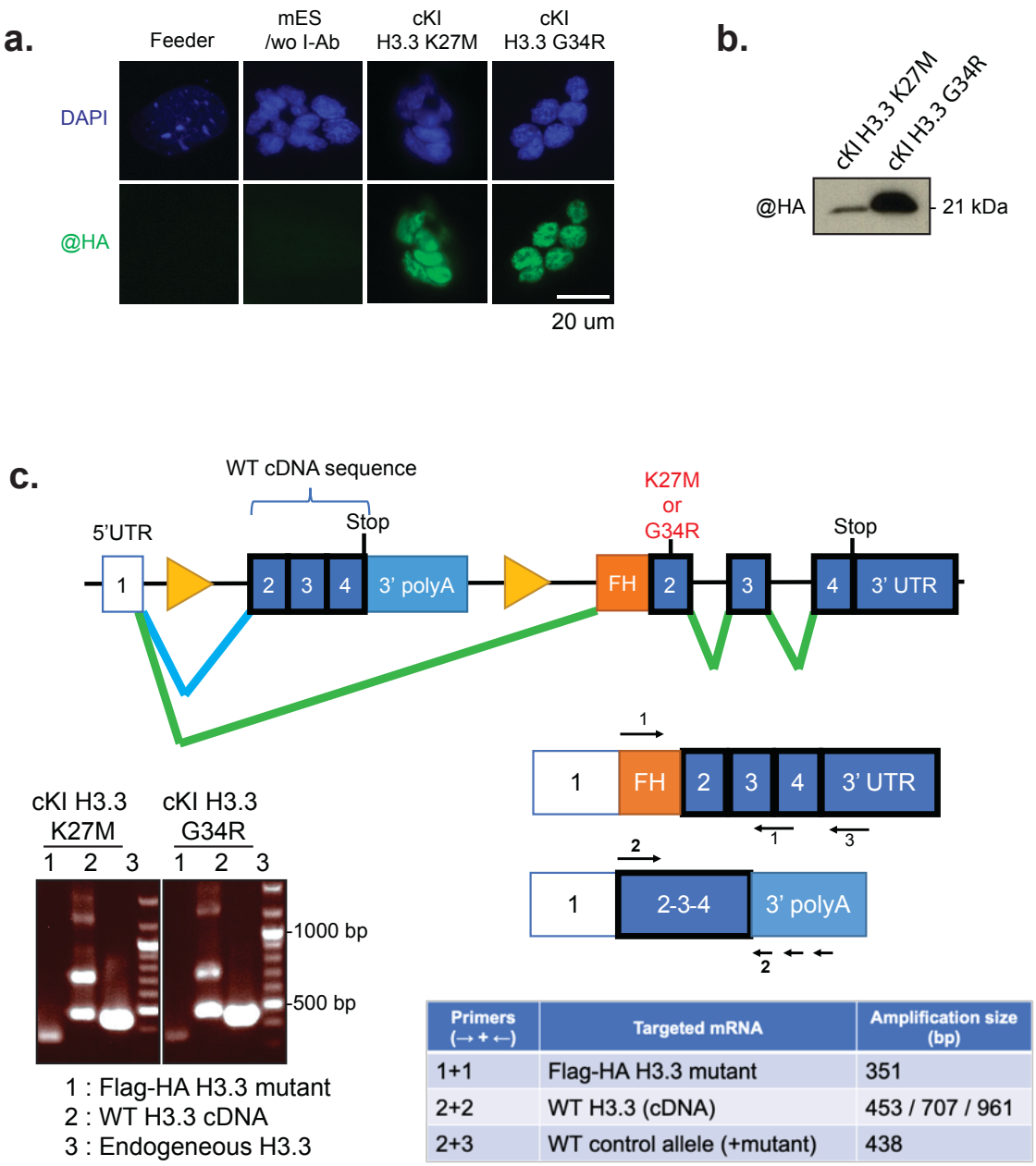


Figure 37: Schematic representation of the mouse model strategy for cKI-H3.3 mutant expression.
H3f3a genomic sequence is composed of 2 non-coding exons (5' UTR/Exon 1 and 3' UTR) and 3 coding exons (2-3-4) (top). The construct comprises the introduction of a wildtype H3.3 cDNA with a 3' polyA terminator flanked by two LoxP sites (in yellow) in the first intron. Before Cre recombinase expression, mRNA splicing would be performed between the 5'UTR and the WT cDNA further coding for a WT and untagged H3.3 (middle). After Cre recombinase expression, the WT cDNA sequence would be excised and the mRNA splicing would be performed between the remaining exons and would code for FLAG-HA-mutant H3.3 K27M or G34R (bottom).

1.1.2. Invalidation of the conditionality of the strategy

The mESC used for blastocyst injection showed H3.3 mutant expression before any Cre recombinase expression (**Figure 38a and b**), thus invalidating the conditionality of the strategy proposed by Phenomin. To determine whether the mRNA splicing was completely or only partially leaking toward the mutant expression, RT-PCR was performed on mESC total RNA. Both the WT cDNA and the FLAG-HA-mutant transcripts were expressed (**Figure 38c**). Thus, the construct led to a leakage of the splicing between the 5'UTR and the modified Exon2 in addition to the predicted splicing between the 5'UTR and WT cDNA.



Primers (→ + ←)	Targeted mRNA	Amplification size (bp)
1+1	Flag-HA H3.3 mutant	351
2+2	WT H3.3 (cDNA)	453 / 707 / 961
2+3	WT control allele (+mutant)	438

Figure 38: Mutant H3.3 is expressed in the mESC before Cre recombination due to an unexpected alternative splicing of the modified construct.

a. HA-immunofluorescence in cKI H3.3 K27M or G34R mESC. Controls were performed on feeders and on mESC without primary antibody (/wo I-Ab). Scale bar 20 μm. **b.** Total mESC extracts were probed for HA by Western blot. The amount of proteins loaded was not verified, thus the difference of signal between the sample cannot explain a difference in abundance. **c.** Identification of the cKI construct mRNA transcripts by RT-PCR in total RNA extract. Representative scheme of the potential splices on the cKI construct (top). Green splicing would give rise to FLAG-HA H3.3 mutant mRNA probed with 1+1 primer pair (1). Blue splicing would give rise to WT H3.3 cDNA mRNA probed with 2+2 primer pair (2). 2+3 primer pair would both amplify the WT and the mutant alleles (3).

1.1.3. Blastocyst injection results

Despite the invalidation of the conditionality of the strategy, the mESC clones were injected in blastocysts and gave rise to 12 and 5 male chimera founders for K27M and G34R respectively (**Figure 39**). The latter had a percentage of chimerism ranging from 15 to 80 % and were further bred with BALB/cN females for germline transmission. A wide majority of the F1 offspring had a wildtype fur color (agouti), while only 5 and 10 F1 animals had the construct fur color (black) for K27M and G34R respectively. Nevertheless, none of the black F1 animals had the cKI genotype. No animals with the cKI genotype were obtained.

a.

Mutation	mESC Clones	Nb of Injections	Nb of Chimera	% chimerism	X Agouti	✓ Black	cKI genotype
K27M	2	4	12	35-80 %	365	5	0
G34R	3	4	5	15-75 %	84	10	0

b. G34R cKI chimera

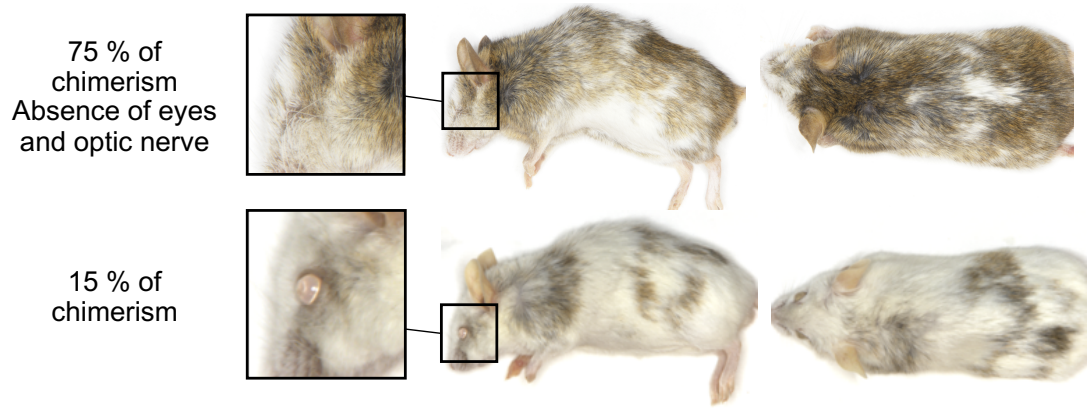


Figure 39: F0 chimera and germline transmission statistics.

a. Summary of the number of mESC clones validated for K27M and G34R constructs, number of blastocysts injection, number of F0 chimeras obtained with the associated range of chimerism, the number of agouti (WT) and black (WT or mutant) F1 animals and the number of cKI genotype in F1 black animals. **b.** Pictures of two F0 G34R chimeras at 6 months of age. The top one showed 75 % of chimerism with an absence of eyes and optic nerves and was sterile. The bottom one showed 15 % of chimerism and had no major phenotype.

In addition to the general failure of germline transmission, a G34R F0 animal with 75 % chimerism showed a complete absence of eyes and optic nerves and was sterile. A K27M F0 animal with 50 % of chimerism was also sterile and three other K27M F0 animals (45/55/80 % chimerism) prematurely died before 5 weeks of age. We first hypothesized that the WT cDNA of the construct could trigger an overexpression of WT H3.3 due to the non-endogenous regulation of the 3'-polyA terminator. However, the F0 chimeras were crossed with female expressing constitutively Cre-recombinase, and no offspring harboring the mutant construct was obtained, suggesting that the WT cDNA was not the cause of the germline transmission failure, but rather is the mutant expression leakage.

Thus, the leaking mutant expression led to the failure of germline transmission and to a dominant negative effect with phenotypes such as the absence of eyes and optic nerves for one G34R chimera and for the premature deaths of three K27M chimeras. H3.3 mutant K27M expression has been previously shown to lead to lethality when expressed in postzygotic cells, so H3.3 mutant expression might be the leading cause of the deleterious effects observed in our F0 chimera animals (Pathania et al., 2017) and probably to the observed embryonic lethality. Other mouse studies have only documented tissue specific expression of H3.3 K27M later during embryonic development or post-natally (Pathania et al., 2017, Larson et al., 2019). The failure of germline transmission could arise from a problem of spermatogenesis in the FO male chimera and/or from embryonic lethality. Indeed, H3.3 has been proposed to play an important role in spermatogenesis as male lacking H3.3 have been found to be sterile (van der Heijden, 2007). The leaking expression of H3.3 (K27M or G34R) mutants could hence interfere with spermatogenesis leading to the failure of germline transmission.

In summary, our attempt to develop a mouse model for conditional expression of mutant H3.3 failed. The conditionality of the strategy was invalidated as the mutant is already expressed in the mESC injected in blastocysts. Some F0 chimera animals showed deleterious phenotypes and this was correlated with high level of chimerism. The germline transmission failed either due to problem during spermatogenesis or to embryonic lethality. In conclusion, H3.3 mutant expression seems to have a dominant deleterious effect even if only one out of the four alleles is mutated. The strategy has thus to be revised to respect the conditionality of the mutant expression as the latter is

primordial for the viability and the germline transmission. The conditional allelic replacement developed by Mann team (Tang et al., 2013) could be adapted for K27M or G34R conditional expression. The endogenous coding sequence is replaced by a H3.3 cds which can further be excised allowing the expression of a downstream mutant coding sequence. This strategy would avoid alternative splicing between exons but the deletion of exons/introns might also interfere in the wildtype regulation of this coding sequence.

1.2. Characterization of the mESC model

Following the failure of the mouse model strategy, we decided to use the mESCs generated during the course of this project to study the role of H3.3 oncomutations. Toward this goal, WT cDNA was excised from mESC clones by electroporation of a Cre recombinase and clones were further selected and validated (**Figure 40**). Control mESCs were derived from an in-house mouse model expressing wildtype tagged *H3f3a*.

mESCs expressing H3.3 WT, K27M and G34R showed same level of tagged H3.3 expression (**Figure 40b and c**). The distribution of transcripts for H3 family members was determined by total RNA-seq (**Figure 40c, d and e**). Canonical histone H3 (H3.1 and H3.2) represented 80 % of the H3 family transcripts while H3.3A and H3.3B represented respectively 1 and 19 %. This distribution was maintained under mutant expression (**Figure 40e**). H3.3A transcription was not deregulated under the mutant expression (**Figure 40f**). Of note, H3.3A represented only 1 % of the overall H3 family transcripts, and only one of the *H3f3a* allele is harboring the tag/mutation.

In summary, there is no difference in the expression level of the tagged H3.3A between wildtype and mutants in the generated mESCs. The next chapter will detail the impact of H3.3 mutations on its genomic distribution and transcriptional repertoire in mESCs.

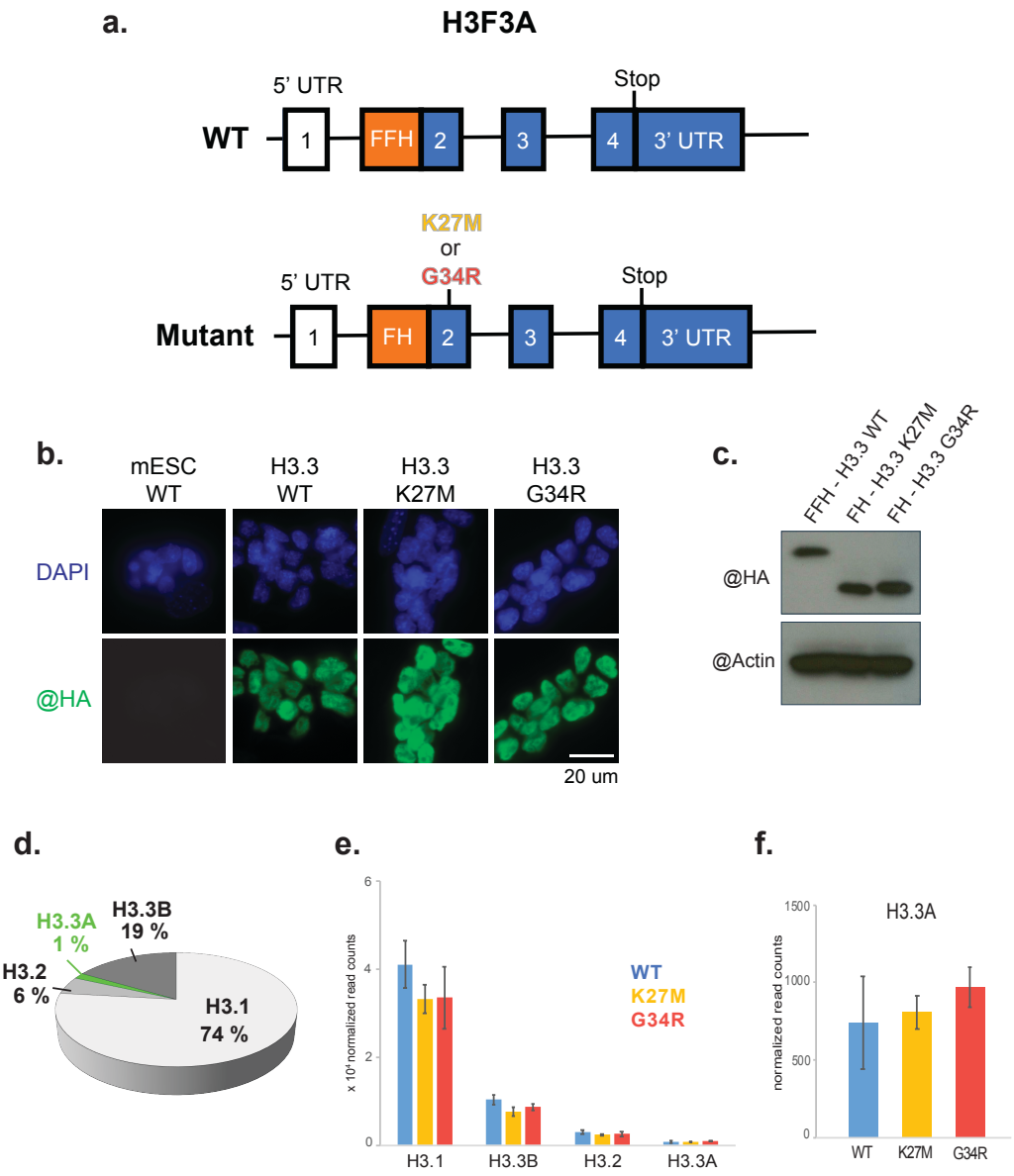


Figure 40: mESC model for tagged H3.3 WT, K27M or G34R expression

a. Schematic representation of the tagged H3.3 WT or mutant constructs. One allele of *H3f3a* is modified. The WT construct have a FLAG-FLAG-HA tag whereas the mutant one has a FLAG-HA tag. **b.** HA-immunofluorescence in mESC WT or expressing tagged H3.3 WT, K27M or G34R. Scale bar 20 μm. **c.** Total mESC extracts were probed for HA and Actin by Western blot. The difference of size between H3.3 WT and mutants is explained by the presence of an additional FLAG in the tag of the WT. **d-f** H3 family members expression determined by RNA-seq (RiboZero). **d.** Pie chart representing the proportion of total H3 transcripts for H3.1, H3.2, H3.3A and H3.3B for H3.3 WT mESC. **e.** Proportion of H3 family members transcripts for H3.3 WT or mutant mESC (WT n=12; K27M n=6; G34R n=8), no significant differences between WT and mutants. **f.** Zoom on *H3f3a* transcripts for H3.3 WT, K27M and G34R mESC (WT n=12; K27M n=6; G34R n=8), no significant differences between WT and mutants.

Chapter 2: H3.3 mutations cause major deregulation of endogenous retroviral elements

How H3.3 mutations affect its deposition or perturb the chromatin landscape has not been addressed so far, especially for G34R/V mutations. Indeed, H3.3 mutation could lead to a mislocalization of the mutant or to differential enrichment compared to the wildtype. Thanks to our endogenous tagged mESC models, wildtype and mutant H3.3 could be specifically tracked among the pool of untagged H3.3.

2.1. Impact of H3.3 mutations at active chromatin

H3.3 is found enriched at transcriptional active chromatin (see **I.2.2.3.1**). Whether H3.3 mutations impair its enrichment at transcriptional start sites (TSS) and enhancers was first investigated. Native ChIP-seq was performed on MNase-digested mononucleosomes purified by tandem affinity purification FLAG-HA from mESC

2.1.1. Enrichment at active chromatin: promoter and enhancers

H3.3 WT was found enriched at TSS and its enrichment follows transcription, with highest enrichment at the TSS of the most transcribed genes (**Figure 41a and b**). H3.3 K27M and G34R were found enriched at the same level than H3.3 WT at TSS (**Figure 41a, c and d**). In addition, H3.3 WT and mutants were also found enriched at similar level at enhancers, identified by peaks distal to promoters (>2 kb) lacking H3K4me3 and with presence of H3K4me1 and/or H3K27ac marks (**Figure 41d, e and f**).

In accordance with a recent study from Nagaraja et al. (2019), I confirm that H3.3K27M is enriched at the same level than H3.3WT at active chromatin. In addition, I show for the first time that H3.3G34R is also located and enriched at similar level than H3.3WT at active chromatin. Trafficking of the oncohistone seem thus not to be altered.

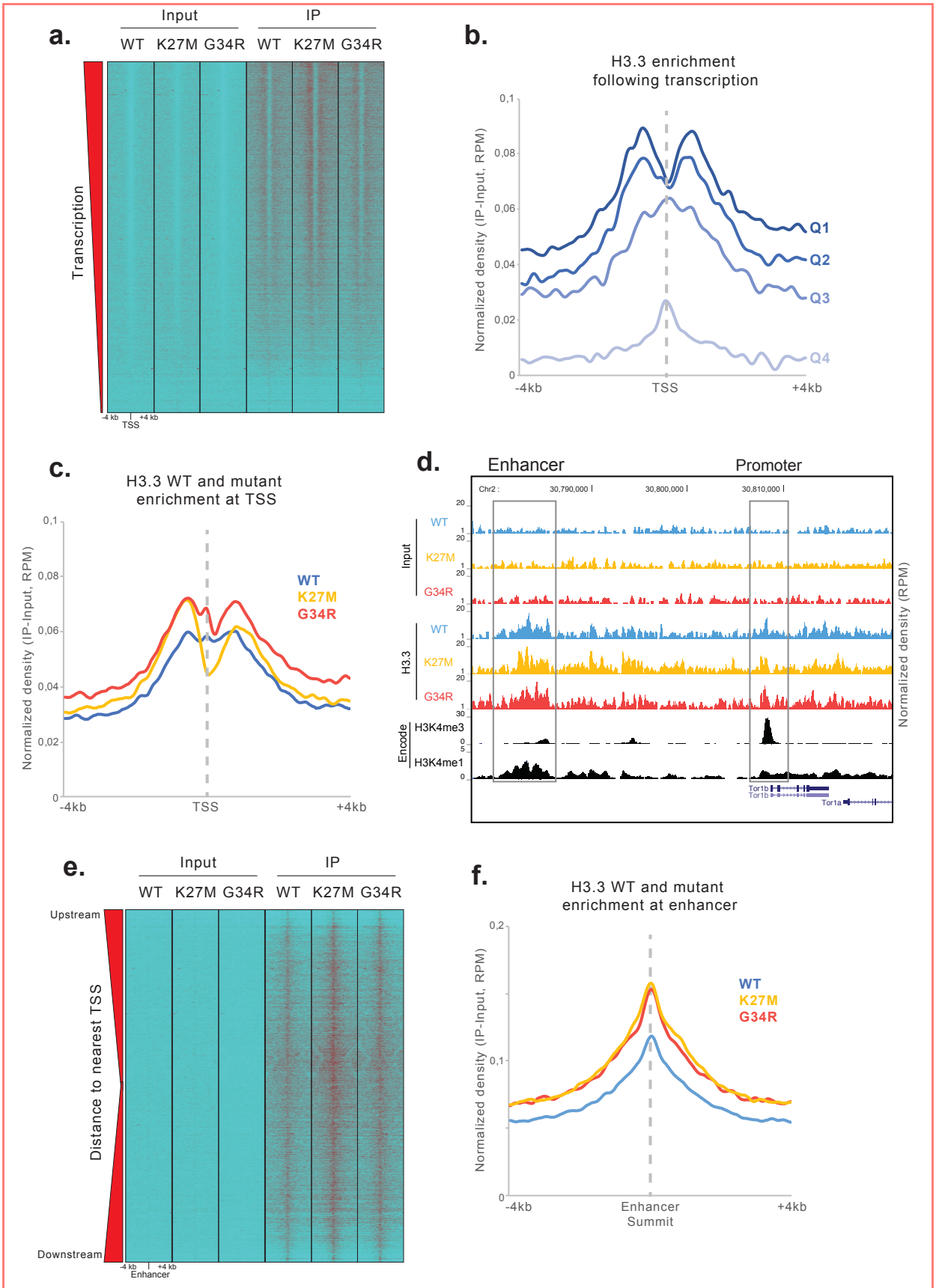


Figure 41: H3.3 mutants show similar level of enrichment than H3.3 WT at active chromatin in mESC. (Legend continues on next page)

a. Heatmaps of native ChIP-seq FLAG-HA inputs and IP for H3.3 WT, K27M and G34R densities around TSS (+/- 4 kb). As reference, the TSS coordinates from the Ensembl v67 database restricted to protein-coding genes (n= 20,229) were used. Promoters were sorted according to the transcriptional level found in H3.3 WT harboring mESC. **b.** Normalized density (IP-Input, RPM) of FLAG-HA-H3.3 WT deposition across TSS (+/-4 kb) following transcription level (by quartile from higher to lower transcription: Q1 to Q4). **c.** Mean normalized density (IP-Input, RPM) of FLAG-HA-H3.3 WT, K27M or G34R deposition across TSS (+/- 4 kb). **d.** Genome browser view showing the distribution of input and IP (normalized density, RPM) for H3.3 WT and mutant native ChIP-seq at a representative locus on chromosome 2, containing promoter and enhancer enriched in H3.3. H3K4me3 and H3K4me1 marks distribution (from Encode) are shown for promoter and enhancer identification, respectively. **e.** Heatmaps of native ChIP-seq FLAG-HA inputs and IP for H3.3 WT, K27M and G34R densities around enhancers (+/- 4 kb). As reference, the H3.3 distal peaks coordinates from Chronis et al. (2017) were used (>2 kb, n= 9,302). **f.** Mean normalized density (IP-Input, RPM) of FLAG-HA-H3.3 WT, K27M or G34R deposition at enhancers (+/- 4 kb).

H3.3 enrichment at TSS is proportional to transcription, but whether this association have an effect on transcription is under debate. In order to decipher the impact of H3.3 mutations on gene transcription, we performed a global Ribozero RNA-seq analysis followed by differential expression profiling between H3.3 WT and mutants (see Material and Methods).

2.1.2. Genes differential expression analysis

Endogenous expression of tagged H3.3 K27M or G34R led to mild transcriptional deregulation, respectively 209 and 1152 deregulated genes with 97 % and 89 % of them showing 1<|log2FC|<2 (**Figure 42**). No pathway was significantly enriched while performing functional annotation clustering on the deregulated genes, thus no specific function seemed to be affected.

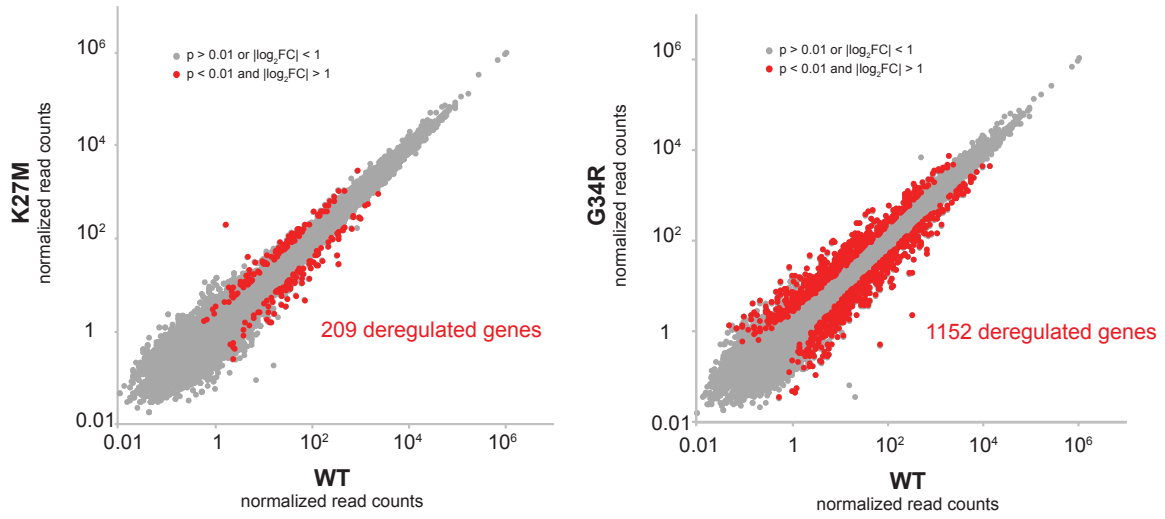


Figure 42: H3.3 mutant leads to mild gene deregulation in mESC.

Gene differential expression analysis for H3.3K27M (n=6) vs. H3.3WT (n=12) (left) and H3.3G34R (n=8) vs. H3.3WT (n=12) (right). Differentially expressed genes are shown in red (adjusted p-value < 0.01 and |log2FC|>1). 209 and 1152 genes are deregulated in H3.3K27M and H3.3G34R respectively.

Knowing that H3.3 enrichment tends to follow transcription and having shown that H3.3 mutants were globally enriched at similar levels than wildtype H3.3, I investigated whether H3.3 mutants were differentially enriched than the wildtype at deregulated genes.

2.1.3. Link between H3.3 enrichment at TSS and gene deregulation

No significant difference in H3.3 enrichment was observed between wildtype and mutants at promoters of differentially expressed genes (**Figure 43**). Moreover, the mutants enrichments were not following transcription for deregulated genes (e.g. no higher enrichment at upregulated genes' promoters).

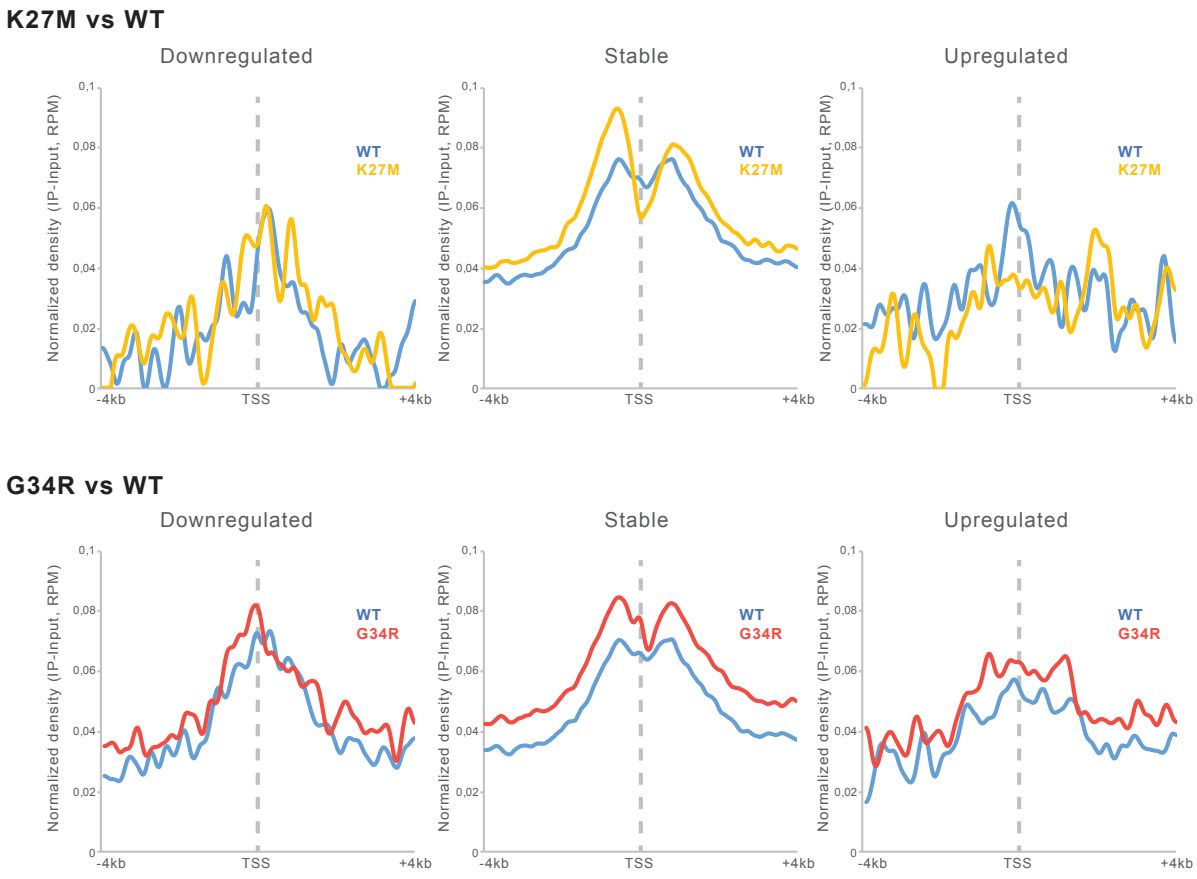


Figure 43: H3.3 mutants are not differentially enriched at TSS of deregulated genes.

Mean normalized density (IP-Input, RPM) of FLAG-HA-H3.3 WT (blue), K27M (yellow) or G34R (red) deposition across TSS (+/- 4 kb). Enrichment of H3.3 mutant and WT are shown for downregulated, stable and upregulated genes (from left to right) for K27M (top) and G34R (bottom).

These results suggest that the gene deregulation observed is not due to a differential enrichment of H3.3 mutant over wildtype at their promoters. H3.3 mutations do not lead to a differential enrichment at active chromatin (promoters and enhancers) and its enrichment at these loci is not the lead of gene deregulation.

2.2. Impact of H3.3 mutations at DNA repetitive elements

In addition to active chromatin, H3.3 has also been shown to be deposited at heterochromatin and at specific sets of DNA repetitive elements (**I.2.2.3.5**). We thereby asked whether H3.3 mutant enrichment at repetitive elements could be altered. DNA repetitive elements constitute a challenge in genome-wide analysis due to their very poor mappability. To counteract the alignment problems when using only the uniquely mapped reads, the native ChIP-seq FLAG-HA H3.3 WT, K27M and G34R were reanalyzed by using both uniquely and multimapped reads. Instead of assigning the read to a genomic location, it was assigned to a repetitive element family (see Material and Methods).

2.2.1. Enrichment at repetitive elements

The enrichment of H3.3 WT was first assessed among DNA repetitive elements in our mESC model. H3.3 WT has been found enriched at recently integrated CG-rich endogenous retroviruses (ERVs, **Figure 44**). Indeed, H3.3 WT has been found enriched at 21 DNA repetitive elements comprised of 13 from ERVK family, 5 from ERV1 family (LTRIS2, LTRIS3, RLTR1B and RLTR4_Mm/RLTR4_MM-int), one satellite (ZP3AR) and two simple repeats ((CTGTC)n and GC_rich) (**Figure 44a**). The differential enrichment of H3.3 mutant against wildtype on DNA repetitive elements was then assessed. No difference of enrichment was observed for H3.3K27M and G34R at DNA repetitive elements. Thus, H3.3 mutant is located at the same subset of recently integrated ERV families than the wildtype (**Figure 44b-e**). By visualizing the uniquely mapped reads in a genome browser, H3.3 enrichment can only be observed on the border of the repeat because of mappability issues (**Figure 44e**). This supports the need of using both uniquely and multimmaped reads when looking at repetitive elements. To note, H3.3 was also found enriched at telomeres repeats (TTAGGG)n (adjusted p-value = 0.018). In our mESC model, H3.3 is enriched at recently integrated and potentially functional CG-rich ERVs (Papin et al., 2017) and H3.3 mutations have no impact on this enrichment.

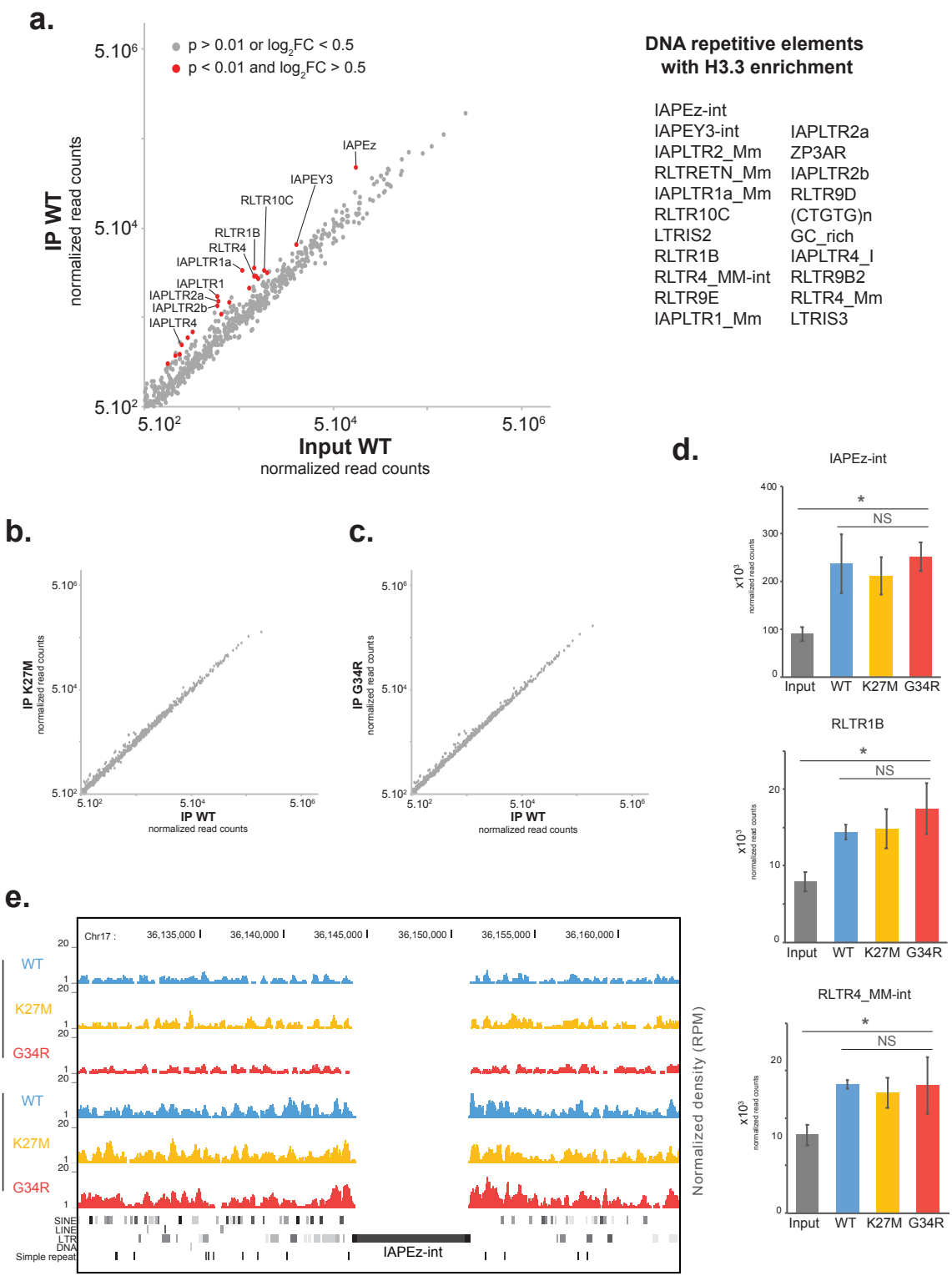


Figure 44: Both H3.3 WT and mutants are enriched at the same level at recently integrated and potentially functional endogenous retroviruses (Legend continues on next page).

a. Scatter plot showing native ChIP-seq FLAG-HA enrichment for H3.3 WT (IP vs. Input, n=2). DNA repetitive elements families enriched in H3.3 are shown in red (adjusted p-value < 0.01 and log2FC>0.5) and their full names are listed on the right. **b-c.** Scatter plot showing native ChIP-seq FLAG-HA enrichment for mutant H3.3 vs WT H3.3, respectively K27M (n=3) and G34R (n=2) (IP mutant vs IP WT). No differential enrichment was observed. **d.** Example of ERV families enriched by H3.3 WT and mutants. Histogram of ChIP-seq FLAG-HA input and IP (WT, K27M, G34R) for IAPEz-int, RTLTR1B and RLTR4_MM-int (from top to bottom). * adjusted p-value < 0.01 and log2FC > 0.5; NS: not significant. **e.** Genome browser view showing the distribution of input and IP (normalized density, RPM) for H3.3 WT and mutant native ChIP-seq at a representative locus on chromosome 17, containing a full length IAPEz-int element enriched in H3.3. 'SINE, LINE, LTR, DNA, Simple repeat' consist in the RepeatMasker track and is used for repetitive elements identification.

2.2.2. Repetitive elements differential expression analysis

H3.3 has been proposed to play a role in the repression of ERVs in mESC (Elsässer et al., 2015; **I.2.2.3.5** and **I.3.2.2.4**), so we investigated the impact of H3.3 mutations on ERV expression. Under H3.3 mutant expression, overexpression of 13 and 26 DNA repetitive element families was found for K27M and G34R respectively (**Figure 45a**). The 13 families overexpressed in K27M were composed of 9 LTR families (ERVK/ERV1, **Figure 45b left**) and 5 LINE families (L1: Lx3C, Lx3_Mus, L1_Mus4, L1MA4 and Lx7). For G34R, the 26 families overexpressed were composed of 25 LTR (21 ERVK/ERV1, 3 ERVL and 1 MalR, **Figure 45b right**) and 1 satellite (SYNREP_MM). The strong upregulation (4.5-fold change) of the minor satellites located at centromeres was specific of G34R (**Figure 45c left**). A trend to upregulation of major satellites located at pericentromeres was observed for K27M but without significance (**Figure 45c right**), probably due to the high clonal variability.

In addition to the minor satellites overexpression, a genomic amplification has been observed on the chromosome 14 of G34R mESC (**Figure 45d**). To note, the number of chromosomes was verified for all mESC clones by karyotyping and this amplification could only be seen when inputs from the ChIP-seq were sequenced.

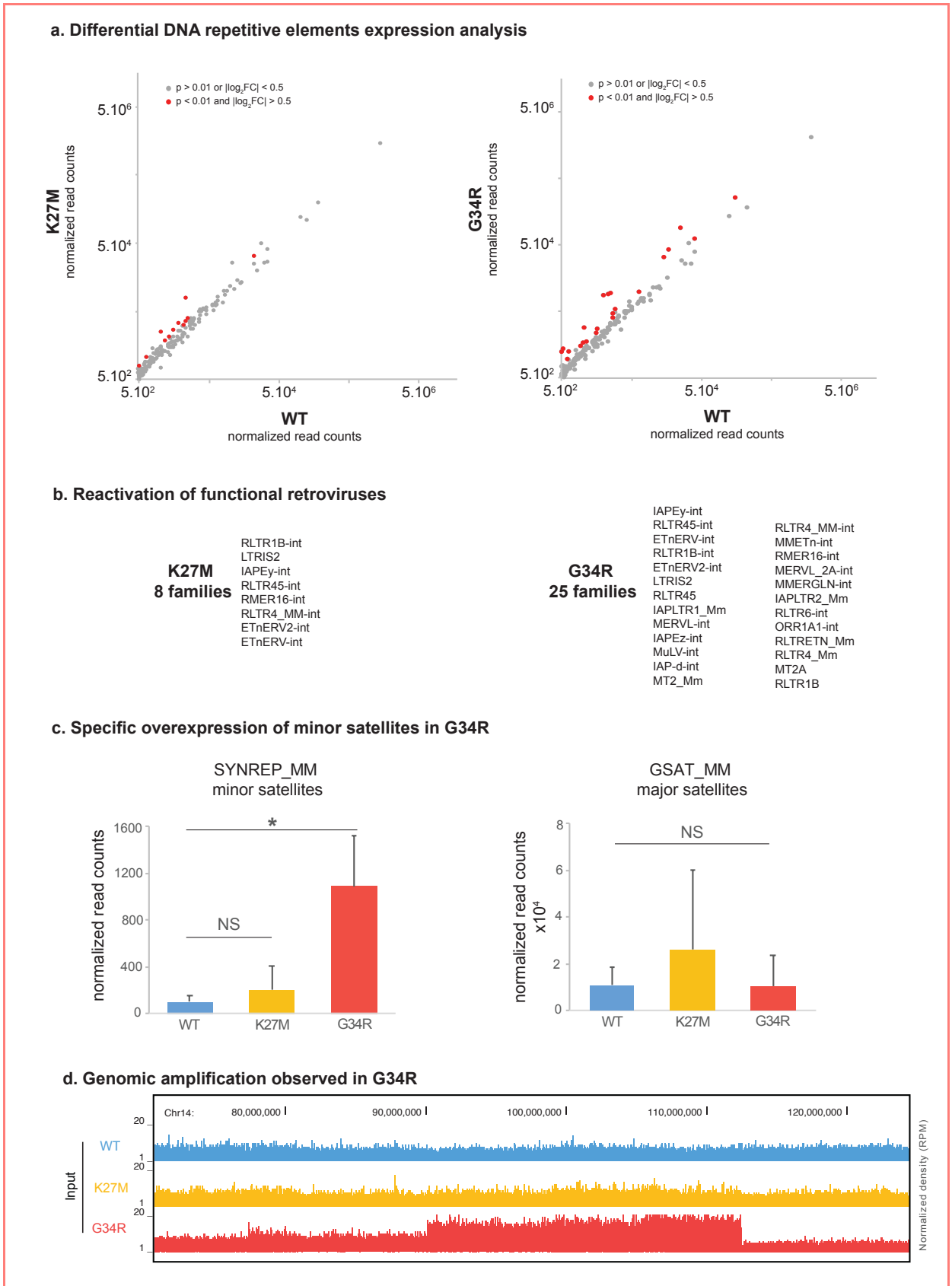


Figure 45: H3.3 mutant expression lead to ERVs overexpression.

(Legend continues on next page)

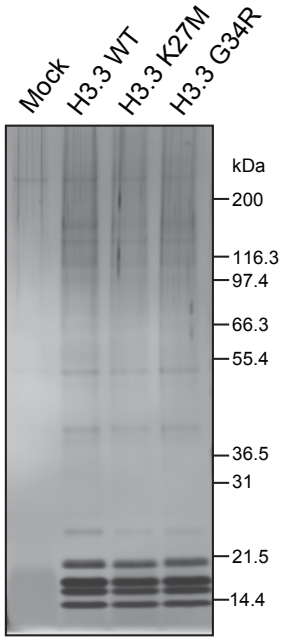
a. DNA repetitive elements families' differential expression analysis for H3.3K27M (n=6) vs. H3.3WT (n=12) (left) and H3.3G34R (n=8) vs. H3.3WT (n=12) (right). Differentially expressed DNA repetitive element families are shown in red (adjusted p-value < 0.01 and |log2FC|>0.5). 13 and 26 DNA repetitive element families are deregulated in H3.3K27M and H3.3G34R respectively. **b.** List of the endogenous retroviruses overexpressed in H3.3 mutants (K27M left and G34R right), sorted in line with the level of deregulation. **c.** Histogram of minor satellites (left) and major satellites (right) expression comparison between wildtype and mutants. * adjusted p-value < 0.01, NS: not significant. **d.** Genome browser view showing the distribution of native ChIP inputs (normalized density, RPM) for H3.3 WT and mutant on chromosome 14, showing a specific chromosomal aberration (amplification and potential deletion) for G34R.

H3.3 WT and mutants have been found enriched at recently integrated ERVs and the latter show overexpression under H3.3 mutant expression. H3.3 has been previously proposed to participate in ERVs silencing in mESC as loss of H3.3 led to ERVs overexpression (Elsässer et al., 2015). Our results thus show that H3.3 mutants lead to ERVs overexpression. The ERVs overexpressed are mainly from ERVK and ERV1 subfamilies and correspond to active members of the Class II ERVs (e.g. IAP, ETn, RLTR, **I.3.3.2.1**). Several mechanisms of repression have been proposed for ERVK and ERV1 subfamilies, namely KRAB-ZFP/KAP1 system, DNA methylation (5-mC) and H3K9me3 which is proposed to be the main mechanism of ERVs repression in mESC (**I.3.2.2**). H3.3 has been linked to these repression machineries as the recruitment of DAXX, H3.3 and KAP1 to ERVs was shown to be co-dependent. H3.3 mutant and wildtype are enriched at similar level at recently integrated and potentially active ERVs, thus H3.3 mutant might alter the recruitment of the ERVs repression machineries. In order to test this hypothesis, the chromatin-associated H3.3 WT and mutant complexes were purified and analyzed by mass spectrometry.

2.3. Step in the mechanism of ERV overexpression

To identify proteins recruited by WT or mutant H3.3 nucleosomes, tandem affinity purification FLAG-HA was performed on crosslinked mESC chromatin. The pulled-down H3.3 associated protein complexes were analyzed by mass spectrometry (**Figure 46**). As expected, the top proteins found in the complex were the histones. Kap1 was also an abundant associated protein. The full KAP1 repressor complex could be identified with the three HP1 (α, β, γ), DNMTs (1, 3a and 3l), SETDB1 as well as the NuRD/HDAC complex (**Figure 46b**). Variation in the abundance of some proteins could be observed in H3.3 mutants. For instance, HP1α has been detected only in H3.3 mutant-associated complexes. On the other hand, a decrease in DNMT1, DNMT3a and MDB3 abundance has been observed for both K27M and G34R, and loss of SETDB1 and decrease of CHD4 was specifically observed for K27M.

a.



b.

			Mock		WT		K27M		G34R	
	Gene Symbol	MWT (kDa)	U	T	U	T	U	T	U	T
Nucleosome	Hist1h4a	11,36	3	3	16	161	13	176	13	160
	Hist1h2bk	13,91	2	2	14	143	12	171	13	169
	Hist1h2ad	14,13	2	3	6	64	7	76	7	53
	Hist1h3a	15,39	0	0	8	23	8	26	7	24
KAP1 Repressive complex	Kap1	88,79	0	0	16	26	14	19	15	20
	HP1β	21,4	0	0	5	7	4	5	5	6
	HP1γ	20,84	0	0	5	6	5	5	3	3
	HP1α	22,17	0	0	0	0	**2**	2	**3**	3
	Dnmt1	183,07	0	0	11	11	**7**	7	**7**	7
	Dnmt3l	47,96	0	0	5	5	4	4	4	4
	Dnmt3a	101,61	0	0	3	3	**1**	1	**1**	1
	Setdb1	144,46	0	0	2	2	**0**	0	3	3
Chaperonnes	Daxx	81,44	0	0	2	3	3	3	4	5
	Atrx	278,41	0	0	5	6	2	2	3	3
NuRD/HDAC	Chd4	217,61	0	0	11	12	**5**	6	8	9
	Mbd3	32,15	0	0	1	2	**0**	0	**0**	0
	Rbbp7	47,76	0	0	5	5	5	5	5	5
	Rbbp4	47,63	0	0	2	3	3	4	3	5
	Hdac2	55,27	0	0	1	2	1	2	1	2
	Hdac1	55,04	0	0	4	4	3	3	4	4
	Mta1	80,75	0	0	3	3	2	2	1	1
	Mta2	74,98	0	0	3	3	2	2	1	1
	Mta3	67,03	0	0	1	1	1	1	1	1
	Gatad2a	67,29	0	0	1	1	0	0	0	0
	Gatad2b	65,37	0	0	2	2	2	2	1	1

Figure 46: The KAP1 and its associated repressive protein complexes bind to H3.3

a. Silver staining of proteins associated with tagged H3.3 WT, K27M or G34R after FLAG-HA tandem affinity purification of crosslinked mESC chromatin. 'Mock' indicated purification from a non-tagged cell line. **b.** Subset of the protein complexes shown in (a.) analyzed by mass spectrometry. MWT: Molecular Weight in; U: number of Unique peptides; T: Total number of peptides.

To note, low amounts of peptides were identified for the majority of those components, so no quantitative interpretation can properly be made on this single result. This pull-down will need to be repeated with several replicates to get a proper quantification. In addition, making differences in protein abundance can be challenging taking into account the potential presence of several neighboring non-mutated H3.3/H3.1 nucleosomes. I hereby propose that H3.3 K27M and G34R are leading to a differential recruitment of the repression machineries at ERVs, for example by decrease of the binding of one or several of the actors or by decreasing their residency time. HP1 proteins have been previously shown to be dispensable for ERV silencing (Maksakova et al., 2011). All of the results presented here are obtained in primed mESC (grown in serum+LIF), thus suggesting that the ERVs repression in performed by both SETDB1-dependent H3K9me3 deposition and DNA methylation.

The ERV families which are enriched in H3.3 in mESC are also marked by DNA methylation. Indeed, all the ERV families enriched in H3.3 are either enriched in 5mC or in 5hmC or both, further supporting a repression at least partially performed by DNA methylation (**Figure 47**). As suggested by the protein composition of the different complexes presented in **Figure 46**, DNMT1 and DNMT3a seems to be less abundant in H3.3 mutant complexes. A decrease in the DNMTs could lead to ERVs hypo-methylation and participate in their overexpression. To confirm this hypothesis, DNA methylation immunoprecipitation should be performed on H3.3 wildtype and mutants mESC.

Further molecular studies will be needed to understand the underlying mechanism of differential recruitment at ERVs of the repressor machineries in order to identify the major cause of the overexpression of the specific set of ERVs in our mESC model.

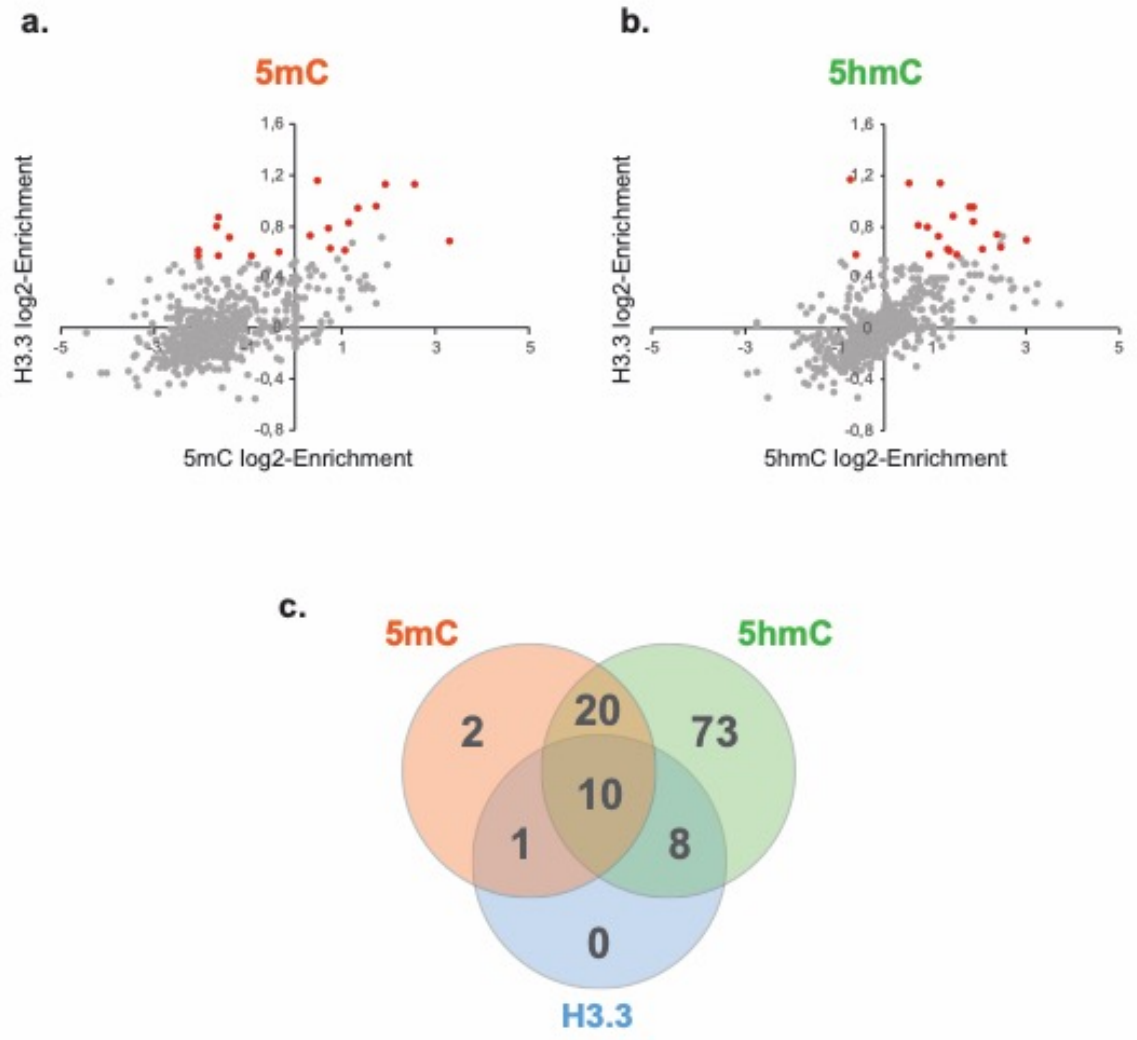


Figure 47: ERVs enriched in H3.3 are enriched in DNA methylation in mESC

a-b. Scatter plots representing H3.3 WT enrichment (ChIP-seq FLAG-HA enrichment for H3.3 WT, IP vs. Input, n=2) vs. 5mC or 5hmC enrichment (Me-Dip 5mC and 5hmC IP vs Input). The ERV families significantly enriched in H3.3 are shown in red (adjusted p-value < 0.01 and log2FC>0.5). **c.** Venn diagram representing ERV families' enrichments in 5mC, 5hmC and H3.3. ERV families enriched in H3.3 in blue (19, adjusted p-value < 0.01 and log2FC>0.5), ERV families enriched in 5mC and 5hmC in orange and green respectively (33 and 111, log2FC>0.4).

Another described way of repressing ERVs is performed by the SETDB1-dependent H3K9me3 deposition. SETDB1 is recruited at ERVs through KAP1 which is itself recruited by a KRAB-ZFP. KRAB-ZFPs have been proposed to specifically recognize and bind subsets of ERVs and to participate in the recruitment of the repressive machinery. For instance, ZFP932 and its paralog Gm15446 have been

shown to regulated overlapping yet distinct sets of ERVKs in mice (Ecco et al., 2016). Both of these KRAB-ZFPs are for instance enriched at IAP-d-int and RLTR44-int, while Gm15446 is more found at IAPEy-int, MMERVK10C-int and IAPEY3-int. Both ZFP932 and Gm15446 are found downregulated in H3.3 K27M and G34R compared to wildtype with a stronger downregulation in G34R (**Figure 48**).

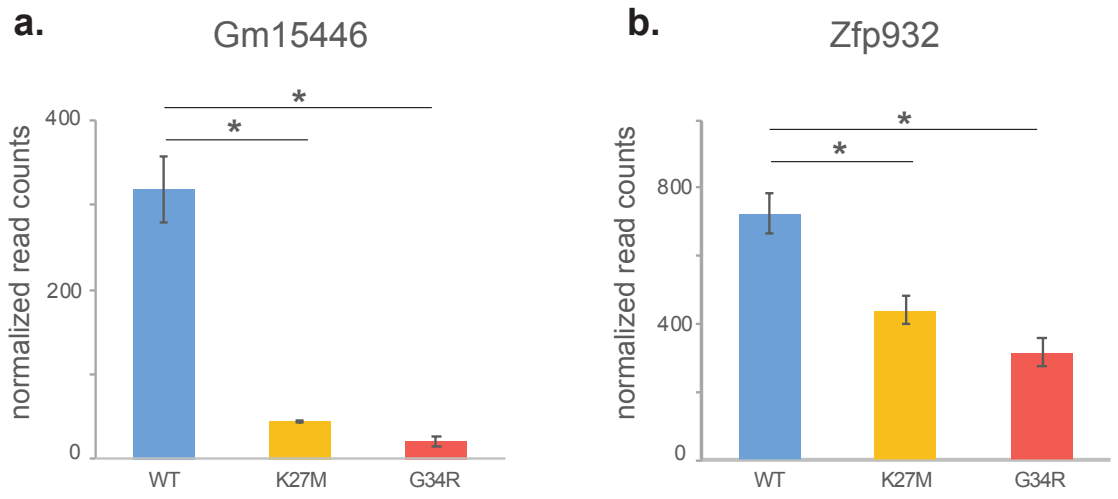


Figure 48: KRAB-ZNP binding and repressing subsets of ERVK are downregulated in H3.3 K27M and G34R.

Histograms of **a.** Gm15446 and **b.** Zfp932 expression comparison between wildtype (blue) and K27M (yellow) or G34R (red). * adjusted p-value < 4.10^{-7}

The transcription of the two KRAB-ZFPs Gm15446 and Zfp932 is highly decreased in the presence of H3.3 mutant, though the protein level should be verified to understand whether the lack of those two KRAB-ZFPs could play a role in the overexpression of the subset of ERVK they bind. The predicted KRAB-ZFP Gm17353 as well as Gm6020, Gm9805, Gm14393, Zfp418 have also been found to be downregulated in H3.3 mutants, and this downregulation was always higher for G34R than K27M and proportional to the overexpression of ERVs. Each KRAB-ZFP is thought to recognize, bind and potentially regulate different set of DNA repetitive elements, so the downregulated KRAB-ZFPs could play a role in the ERV overexpression under H3.3 mutant expression.

In summary, H3.3 wildtype and mutants are located and enriched at the same level in mESC, both at active (promoter, enhancer) and repressive (ERVs) chromatin regions.

In the presence of H3.3 mutations, a set of recently integrated and potentially functional retroviruses are overexpressed. ERV elements have been proposed to influence the expression of neighboring genes by acting as alternative promoters either during oncogenesis (Jang et al., 2019) or development (Karimi et al., 2011; Macfarlan et al., 2012) but their molecular mechanism of action remains poorly understood. It appears that at least two repressive pathways are involved in the repression of transposable elements, the SETDB1-H3K9me3 pathway and the DNMT1 pathway. Whether the two pathways are antagonistic or cooperate to repress the same set of TEs remain to be determined. The SETDB1 pathway is able to maintain silencing of IAPs even in the absence of DNMT1 (Sharif et al., 2016).

As the set of genes deregulated upon mutant expression was not related to a specific biological function, we hypothesized that ERVs overexpression could lead to deregulation of neighboring genes.

2.4. Link between repetitive elements and genes transcriptional deregulation

In order to investigate the link between ERV overexpression and genes deregulation, the nearest gene has been associated to each LTR full length (> 4 kb). Gene deregulation (-log10(adjusted p-value)) has been plotted according to the distance between the gene and the LTR (proximal/other) and according to the deregulation of the LTR (**Figure 49**). The genes proximal to a deregulated LTR are more de-regulated than the other. This result is significant for G34R and constitute only a tendency for K27M, probably due to the small number of genes in the K27M LTR deregulated-associated genes group (n=72). We can thus propose that overexpression of ERVs leads to deregulation of neighboring genes.

Nevertheless, our method of selection for the ‘neighboring’ is not optimal for several reasons. First, only one gene (the closest) was associated for each LTR. The deregulation of one LTR could deregulate several genes in the neighborhood which might still be close (< 10 kb) but which are lost in this analysis because they are not the nearest. In addition, we are lacking the information about the topology of the chromatin as we are considering ‘close’ two elements which are on the same

chromosome in a distance under 5 kb. A better way to analyze the potential link between ERVs and genes would be to integrate the three-dimensional architecture of the genome by using Hi-C data.

ERV overexpression is hereby proposed to be the cause of the gene deregulation upon H3.3 mutant expression.

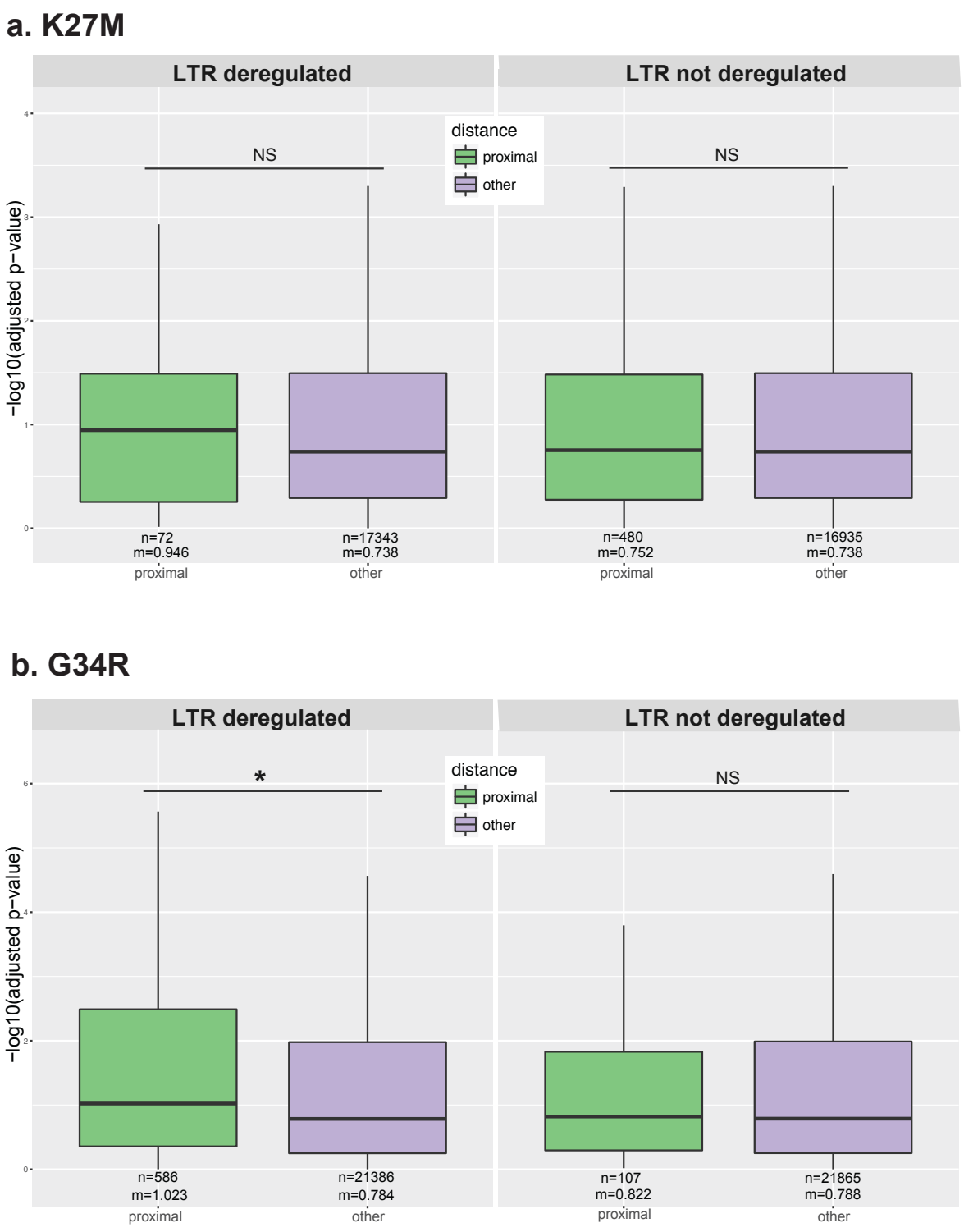


Figure 49: ERV overexpression leads to deregulation of neighboring genes.

Boxplots showing the gene deregulation (-log10(adjusted p-value)) according to the distance LTR-closest gene (proximal <10 kb in green; other >10 kb in purple) when the LTR is overexpressed upon H3.3 K27M (**a.**) or G34R (**b.**) mutant expression (LTR deregulated, left) or not (LTR not deregulated, right). * adjusted p-value < 0.001; NS: not significant. n = number of genes in the group; m = median.

In our mESC model, H3.3 mutant expression has been shown to lead to a specific set of ERV overexpression which might further deregulate neighboring genes, leading to the observed transcriptional deregulation. In order to validate this model in pHGG, several strategies were used.

2.5. Clinical validation

First, we reanalyzed total RNA-seq dataset published by Krug et al. (2019) of 32 pHGG total RNA-seq from patient harboring or not K27M mutation. Two LTR families were significantly up-regulated in K27M harboring tumors compared to H3.3 WT tumors: HERVKC4-int (ERVK) and ERVL47-int (ERVL) (**Figure 50**).

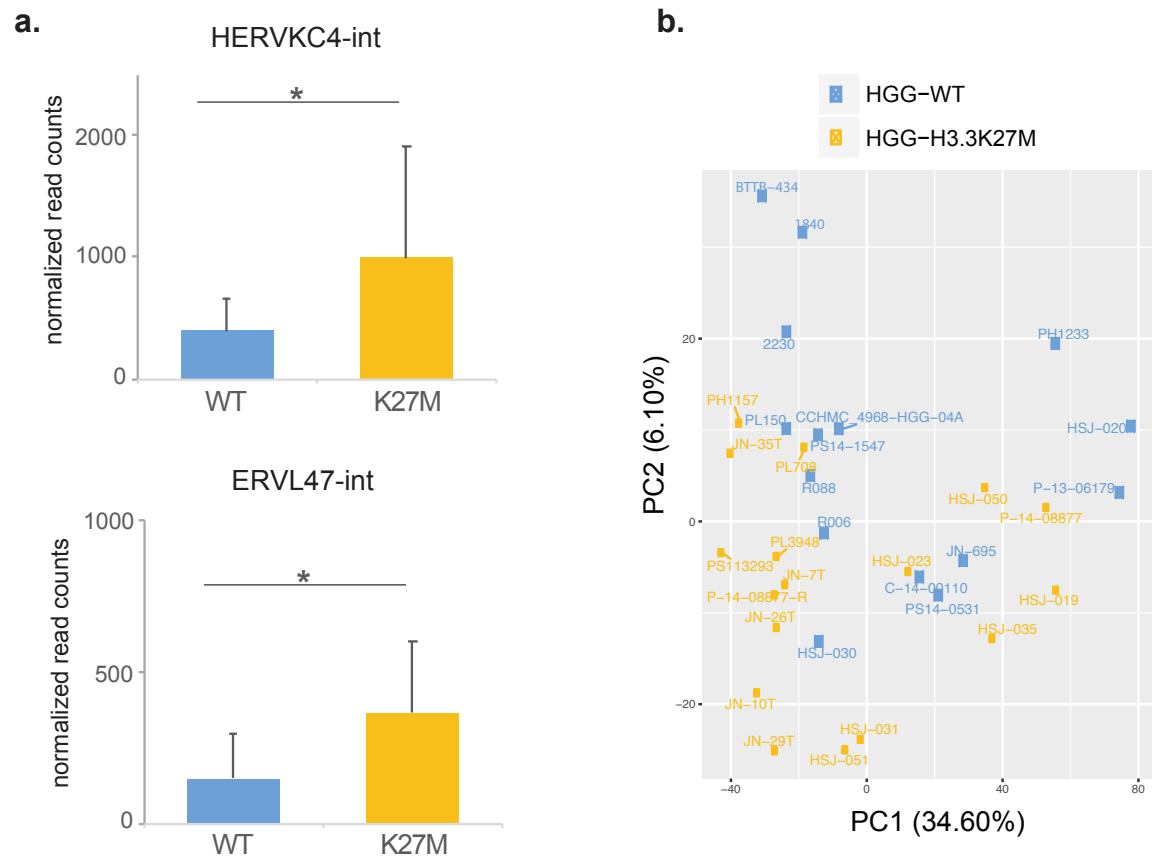


Figure 50: HERVKC4-int and ERVL47-int are specifically upregulated in the K27M tumors

a. Histograms of HERVKC4-int (ERVK, top) and ERVL47-int (ERVL) expression comparison between H3.3 K27M (n=17) and WT tumors (n=15). * adjusted p-value < 0.05. **b.** Principal Component Analysis of the samples reanalyzed. WT tumors in blue and K27M tumors in yellow.

Thanks to a collaboration with Pr. Chris Jones at ICR London, total RNA-seq was also performed on pHGG patient-derived cell lines harboring either H3.3 WT or H3.3 G34R/V. Two LTR families were significantly up-regulated in G34R/V harboring patient-derived cell lines compared to H3.3 WT patient-derived cell lines: HERVH-int and HERV9-int (ERV1) (**Figure 51**).

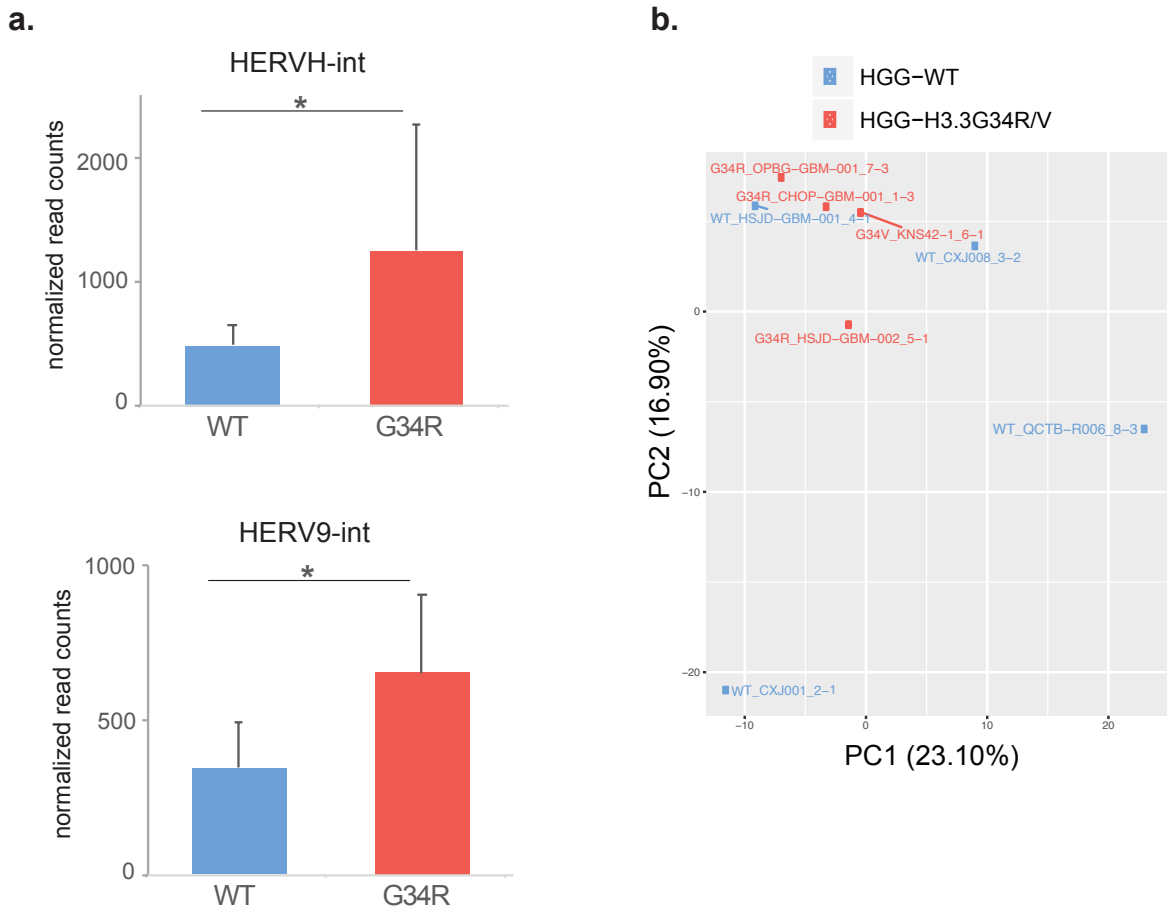


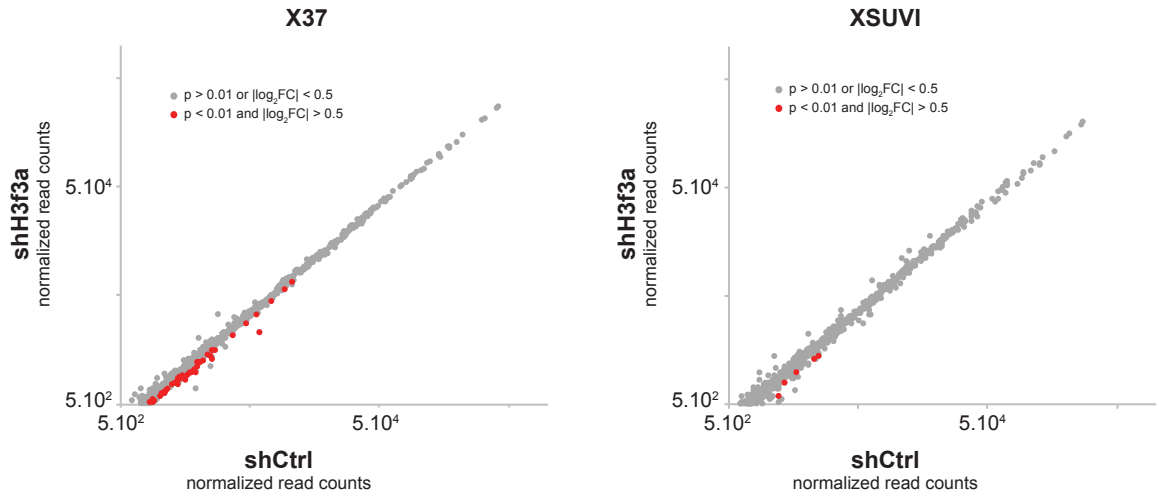
Figure 51: HERVH-int and HERV9-int are specifically upregulated in the G34R/V patient derived cell lines.

a. Histograms of HERVH-int (ERV1, top) and HERV9-int (ERV1) expression comparison between H3.3 G34R/V (n=3/1) and WT patient-derived cell lines (n=4). * adjusted p-value < 0.05. **b.** Principal Component Analysis of the samples. WT patient derived cell lines in blue and G34R/V patient derived cell lines in red.

The comparison of H3.3 K27M and WT pHGG as well as H3.3 G34R/V and WT pHGG-derived cell lines highlighted ERVs families significantly upregulated in H3.3 mutant. Only two ERV families for each mutant are found overexpressed, but this could be explained by the high heterogeneity between the samples of the same group (**Figure 50b** and **Figure 51b**). Indeed, according to the principal component analysis, the variation between the samples is more important than the variation caused by the mutation. When comparing tumors, it is thus difficult to extract the phenotype due to the single H3.3 mutation in a background which has accumulated several mutations and is known to have high clonal heterogeneity. In addition, ERVs are known to be overexpressed in several types of cancer, independently of H3.3 mutation (Hancks and Kazazian, 2012; Criscione et al., 2014; Bannert et al., 2018), so comparing H3.3 mutant tumors to H3.3 WT tumors might not be adequate. Taking into account the tumor and patient (age, sexe) heterogeneity as well as their complexity, analyzing ERVs differential expression in this context turns out not the best strategy. The right way would be to compare each tumor to its adjacent healthy tissue from the same patient. However, for ethical reasons, this is not possible in the case of pediatric pHGG.

Thus, analyzing ERVs expression in the same tumor background under knockdown of H3.3 mutant would be a better strategy. Silveira et al. (2019) have performed knockdown of H3.3 K27M or H3.3 WT in patient-derived xenograft (PDX) harboring K27M. They showed that shH3f3a(K27M) but not shH3f3b(WT) is sufficient to slow down tumor progression. We reanalyzed the total RNA-seq performed under shH3f3a against the shCtrl and focused on differentially expressed repetitive elements. We found that 66 and 6 ERV families are downregulated under shH3f3a for X37 and XSUVI PDX respectively (**Figure 52**). In summary, shH3f3a (but not shH3f3b) treatment in PDX harboring K27M is sufficient to slow tumor progression and leads to repression of ERVs. So K27M is proposed to be necessary for tumor maintenance and its downregulation is associated with ERVs repression.

a. Differential DNA repetitive elements expression analysis



b. Repression of functional retroviruses under H3.3K27M knockdown

X37 66 families

ERV24B_Prim-int	LTR16C	LTR42	LTR88b	MER51C	MLT1A1-int
ERVL-B4-int	LTR16D	LTR49	LTR90A	MER54A	MLT1B-int
HERV16-int	LTR19A	LTR54B	MamGypLTR2b	MER57A-int	MLT1H1-int
HERV4_I-int	LTR1E	LTR56	MamGypLTR3a	MER57B1	MLT2B2
HERVK11-int	LTR2752	LTR62	MER110	MER57C2	MLT2E
HERVK13-int	LTR28B	LTR64	MER34B-int	MER57D	MSTC-int
LTR101_Mam	LTR2C	LTR6A	MER41-int	MER61D	
LTR103_Mam	LTR37-int	LTR76	MER41A	MER65C	
LTR103b_Mam	LTR38	LTR81B	MER41B	MER67B	
LTR10A	LTR40b	LTR82B	MER4A1_	MER70A	
LTR13	LTR40c	LTR85c	MER4CL34	MER76	
LTR13A	LTR41C	LTR86A2	MER51A	MER9a3	

XSUVI 6 families

HERV4_I-int
HERVE_a-int
HERVE-int
HERVK13-int
LTR13
MSTD-int

Figure 52: K27M knockdown slows down tumor progression and leads to ERVs repression in patient-derived xenografts.

a. DNA repetitive elements families' differential expression analysis for X37 PDX shH3f3a (n=4) vs. shCtrl (n=5) (left) and for XSUVI PDX shH3f3a (n=3) vs. shCtrl (n=5) (right). right). Differentially expressed DNA repetitive element families are shown in red (adjusted p-value < 0.01 and |log2FC|>0.5). 66 and 6 DNA repetitive element families are deregulated under shH3f3a in X37 and XSUVI PDX respectively. **b.** List of the endogenous retroviruses repressed under K27M knockdown (X37 left and XSUVI right), sorted by alphabetical order. ERV families in common between the two PDX are shown in red.

2.6. Discussion

Using a mESC model, I have shown that H3.3 K27M and G34R mutations lead to a mild gene transcriptional deregulation but without any link with H3.3 enrichment at promoters. H3.3 enrichment is known to follow transcription at TSS, but it might rather constitute a marker than an actor of transcriptional activity in non-dividing cells. However, the presence of H3.3 mutants at recently integrated ERVs albeit at the same level as the WT is enough to cause their reactivation and overexpression which further lead to de-regulation of neighboring genes. This long-range control of gene expression is not surprising given that several LTRs were previously shown to act at distance to regulate developmental genes (Karimi et al., 2011) or oncogenes (Jang et al., 2019). Deregulation of repetitive and mobile elements could be considered as an emerging mechanism disrupted in disease.

H3.3 is also found at recently integrated ERVs in mESC. H3.3 mutants are also enriched at the same level at ERVs but cause their overexpression which further lead to deregulation of neighboring genes. In primed mESC, recently integrated ERVs repression is mediated through a tuned balance of SETDB1-dependent H3K9me3 deposition and DNA methylation (Deniz et al., 2018). The awakening of transposable elements in mESC has recently been proposed to occur both through erasing of repressing mark, or gain/retain of active mark (He et al., 2019). H3.3 is interacting with the KAP1 repressor complex as well as with the NuRD/HDAC and recruitment of H3.3 and KAP1 to ERVs has been reported to be co-dependent and occurring upstream of SETDB1 recruitment (Elsässer et al., 2015). Our study thus suggests that H3.3 mutations interfere with the proper recruitment of machineries in charge of ERVs regulation and lead to overexpression of the families where mutant H3.3 is located. Most of the works support a unique repression of ERVs by H3K9me3 deposition in mESC (Matsui et al., 2010, Rowe et al., 2013b). But the ERVs overexpressed under H3.3 mutants are targeted by an active methylation/de-methylation process in mESC (Papin et al., 2017), thus we rather suggest a regulation through a balance between H3K9me3 and DNA methylation. Protein interaction analysis will help deciphering the impact of H3.3 mutations on the KAP1 repressor machinery. H3K9me3 enrichment analysis as well as methylation enrichment analysis will give better insight in the mechanism behind ERV overexpression.

The choice of the study model for pHGG is of main importance. As shown above, the high background heterogeneity can make any results interpretation challenging when analyzing directly patient tumors or derived cell lines. Indeed, the main issue when studying pHGG is the lack of proper biological controls. Here, the best solution found to corroborate our mESC model results was to use already published datasets of patient-derived xenografts which have been challenged by shRNA treatments. By analyzing the effect of the treatment by using the same background as control (PDX with control treatment), the direct effect of H3.3 K27M in tumor maintenance and ERV regulation is easier to understand. Patient biopsies and derived cell lines are of main importance to understand this deadly cancer, but these models should rather be challenged (e.g. shRNA treatment) than compared with other patients and with non-adequate controls (H3.3 WT tumors). The direct link between H3.3 K27M and ERVs overexpression has been validated by reanalyzing RNA-seq datasets from Silveira et al. (2019). Indeed, under downregulation of K27M, tumor growth has been shown to be decreased and slowed down and was associated with a repression of ERVs. The results obtained in the mESC model are supporting the direct role of H3.3 on ERV repression and their overexpression under H3.3 mutant expression. Complementing this observation, the results following PDX treatment by shK27M are validating the ongoing role of H3.3 mutation in tumor maintenance but also on ERV regulation.

Hereby, H3.3 mutations are proposed to be the leading cause of overexpression of recently integrated ERVs further leading to deregulation of neighboring genes. Two KRAB-ZFPs known to repress specific sets of ERVK have been found to be strongly downregulated under H3.3 mutant expression, which could decrease the recruitment of the repressor complexes at these ERVs and thus lead to their overexpression. The repression of the latter is thought to be performed by a tuned balance between SETDB1-dependent H3K9me3 deposition and DNA methylation. Comparison of H3K9me3 and 5mC/5hmC between H3.3 WT and mutants mESC will help deciphering the mechanism leading to ERVs overexpression.

Studying the dynamics of ERVs expression under neural differentiation and in an *in vivo* model would be of high significance. While some K27M-DIPG mouse models start to arise, no current model exists for G34R/V-pHGG. In order to link our results in mESC with the one obtained in patients, I investigated the potential of mESC harboring H3.3 WT or mutants for neural differentiation toward Neural Stem Cells (NSC).

Chapter 3: H3.3 plays a major role in neural differentiation through ERV regulation

3.1. mESC-to-NSC differentiation model

In order to investigate the impact of H3.3 mutations and ERV deregulation on neural differentiation, I adapted the mESC-to-NSC protocol from Colombo et al. (2006) (**Figure 53a**). H3.3 wildtype mESC were successfully differentiated in NSC and the latter were validated by RT-PCR and immunofluorescence (**Figure 53b top, c and d**). H3.3 K27M mESC seemed to undergo proper NSC differentiation, but H3.3 G34R mESC failed to differentiate (**Figure 53b middle and bottom, c and d**). Indeed, already at day 5 of differentiation, G34R cells failed to properly adhere on matrigel and had a semi-adherent spherical shape all the way toward the end of the differentiation (**Figure 53b bottom**). After the first dissociation, they kept the spherical shape and died in the following days, thus no mESC-derived-NSC could be obtained for G34R.

To better understand the problem of differentiation for G34R and to decipher the dynamics of ERVs during differentiation, total RNA-seq was performed at day 0 (mESC), day 4 (Embryonic bodies) and day 9 (NSC) for WT, K27M and G34R.

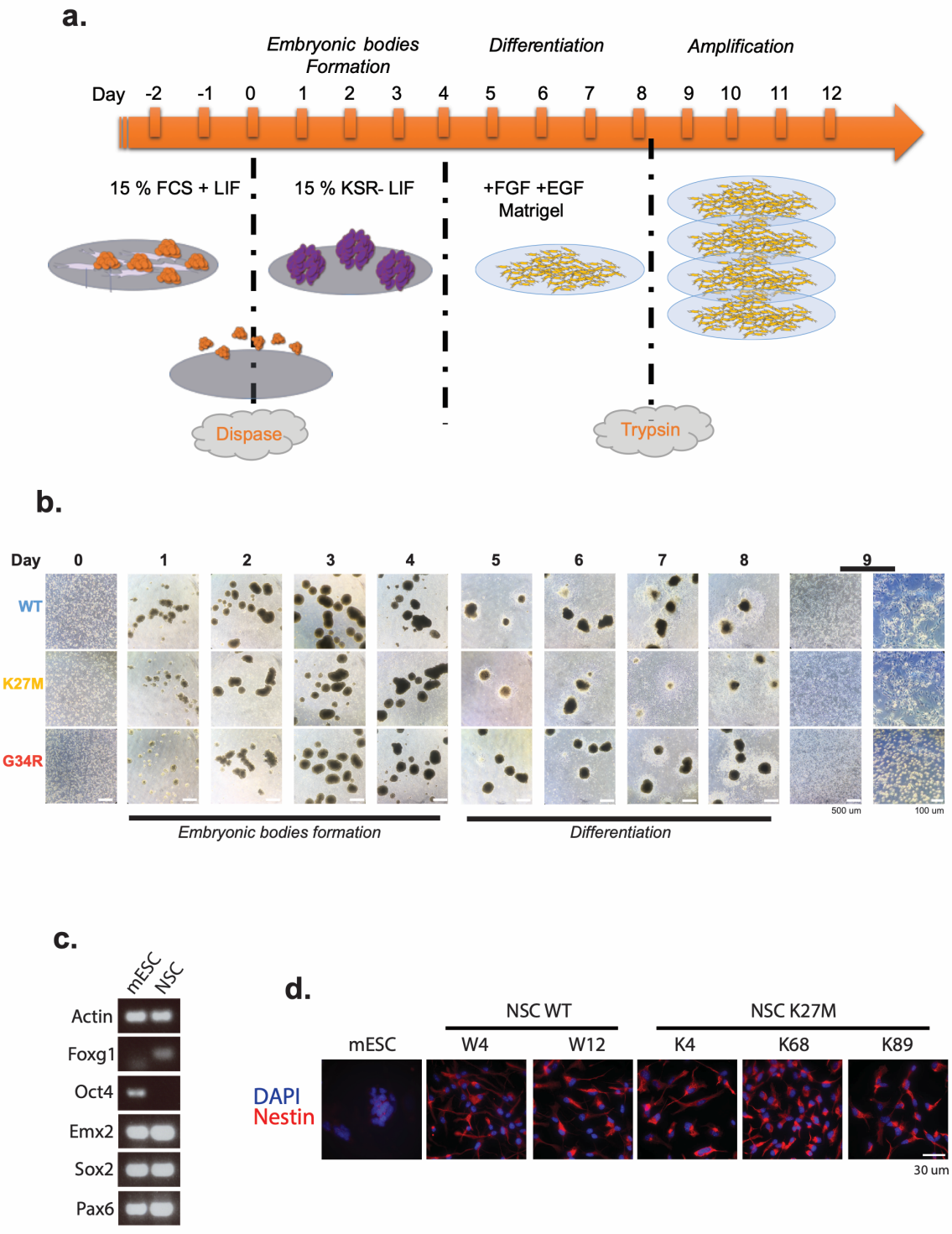


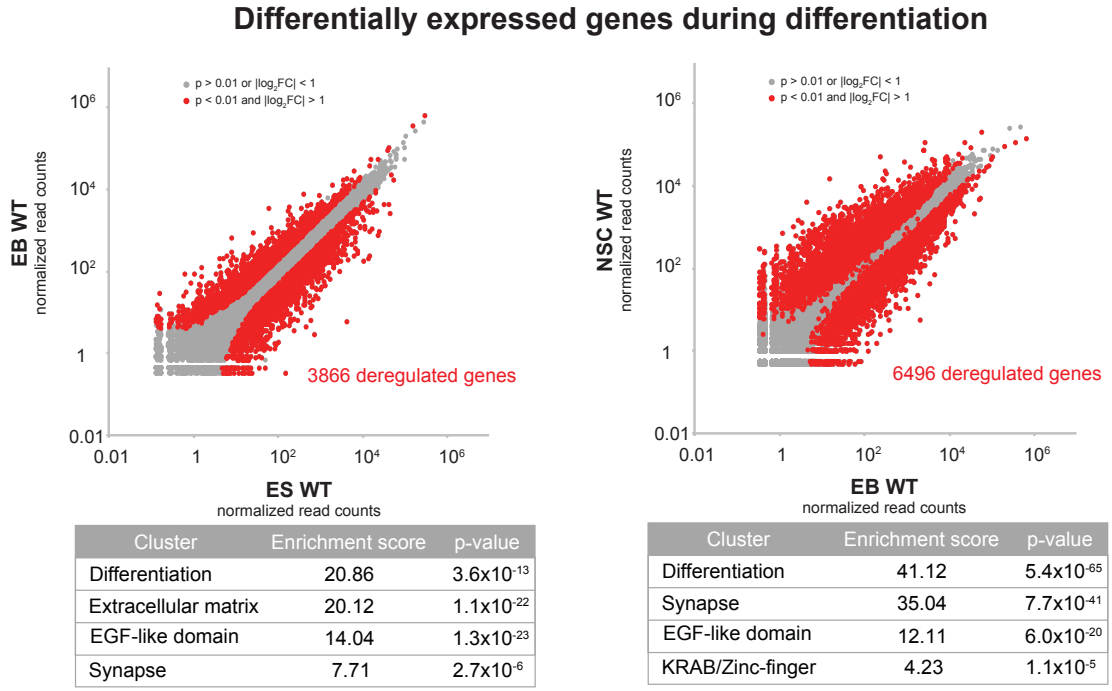
Figure 53: Differentiation of mESC in Neural Stem Cells

a. Schematic view of the protocol of differentiation. The mESC are separated from the feeder layer by a dispase treatment and go toward embryonic bodies formations followed by a differentiation step. FCS: Fetal Calf Serum, LIF: Leukemia Inhibitory Factor, KSR: Knockout-Serum Replacement, FGF: Fibroblast Growth Factor, EGF: Epidermal Growth Factor. (Legend continues on next page)

b. Photos of mESC-to-NSC differentiation steps for mESC harboring H3.3 WT (n=3), K27M (n=3) or G34R (n=3). Day 1-4: Embryonic bodies formation; Day 5-8: Differentiation toward NSC; Day 9 onward: NSC amplification. **c.** Proper differentiation in NSC was validated for all the mESC-derived NSC by RT-PCR for absence of Oct4 and presence of Foxg1, Emx2, Sox2 and Pax6 according to Colombo et al. (2006). **d.** Nestin-immunofluorescence in mESC (negative control) or mESC-derived-NSC validated in RT-PCR and harboring H3.3 WT or K27M. Scale bar 30 μm.

3.2. H3.3 mutations lead to failure of differentiation in NSC

First, differential gene expression analysis was performed for the wildtype mESC at each step of differentiation: EB vs. ES and NSC vs. EB (**Figure 54**).



Cluster	Enrichment score	p-value
Differentiation	20.86	3.6×10^{-13}
Extracellular matrix	20.12	1.1×10^{-22}
EGF-like domain	14.04	1.3×10^{-23}
Synapse	7.71	2.7×10^{-6}

Cluster	Enrichment score	p-value
Differentiation	41.12	5.4×10^{-65}
Synapse	35.04	7.7×10^{-41}
EGF-like domain	12.11	6.0×10^{-20}
KRAB/Zinc-finger	4.23	1.1×10^{-5}

Figure 54: mESC differentiation in NSC is coupled with main transcriptional changes. Gene differential expression analysis for EB H3.3WT (n=3) vs. ES H3.3WT (n=6) (left) and NSC H3.3WT (n=2) vs. EB H3.3WT (n=3) (right). Differentially expressed genes are shown in red (adjusted p-value < 0.01 and |log2FC|>1). 3866 and 6496 genes are deregulated in the first and second differentiation step respectively. Functional annotation clustering tables of the differentially expressed genes are shown under each graph (performed with DAVID 2.0).

mESC differentiation toward NSC has been accompanied by a huge transcriptional change. The differentially expressed genes were associated with functional clusters like differentiation, extracellular matrix, EGF-like domain or synapse (**Figure 54 bottom**). KRAB-ZFPs, known to recruit repressor machineries to repetitive element families, are downregulated at the end of the differentiation process.

In addition, for each differentiation step (EB and NSC), gene differential expression analysis was performed between mutant and WT (**Figure 55**). At day 4 of differentiation, EB arising from mutant or WT H3.3 mESC had a similar shape, while a defect in differentiation could already be observed at the transcriptional level (**Figure 55b**). At day 9 of differentiation, 'NSC' arising from K27M and WT H3.3 had a similar shape, but G34R showed a different phenotype from day 5 onward. Even though K27M-derived NSC were validated by RT-PCR, Nestin-immunofluorescence and could be maintained in culture, the gene differential expression analysis showed a clear defect of differentiation (cluster significantly enriched at day 9: differentiation and neurogenesis). As expected with the morphological observations, G34R-derived 'NSC' presented a huge transcriptional deregulation compared to WT NSC which confirmed an important defect of differentiation (clusters significantly enriched: synapse, differentiation, neurogenesis) (**Figure 55c**).

Expression of H3.3 K27M and G34R mutation in mESC leads to differentiation failure toward NSC. The phenotype was stronger for G34R than K27M with a total blockade during the last differentiation step along with cell death. The genes and ERVs deregulation was also stronger for G34R than K27M in mESC. Considering the impact of H3.3 mutation on ERVs expression in mESC, I investigated the effect of H3.3 mutations on the dynamics of repetitive elements expression during differentiation.

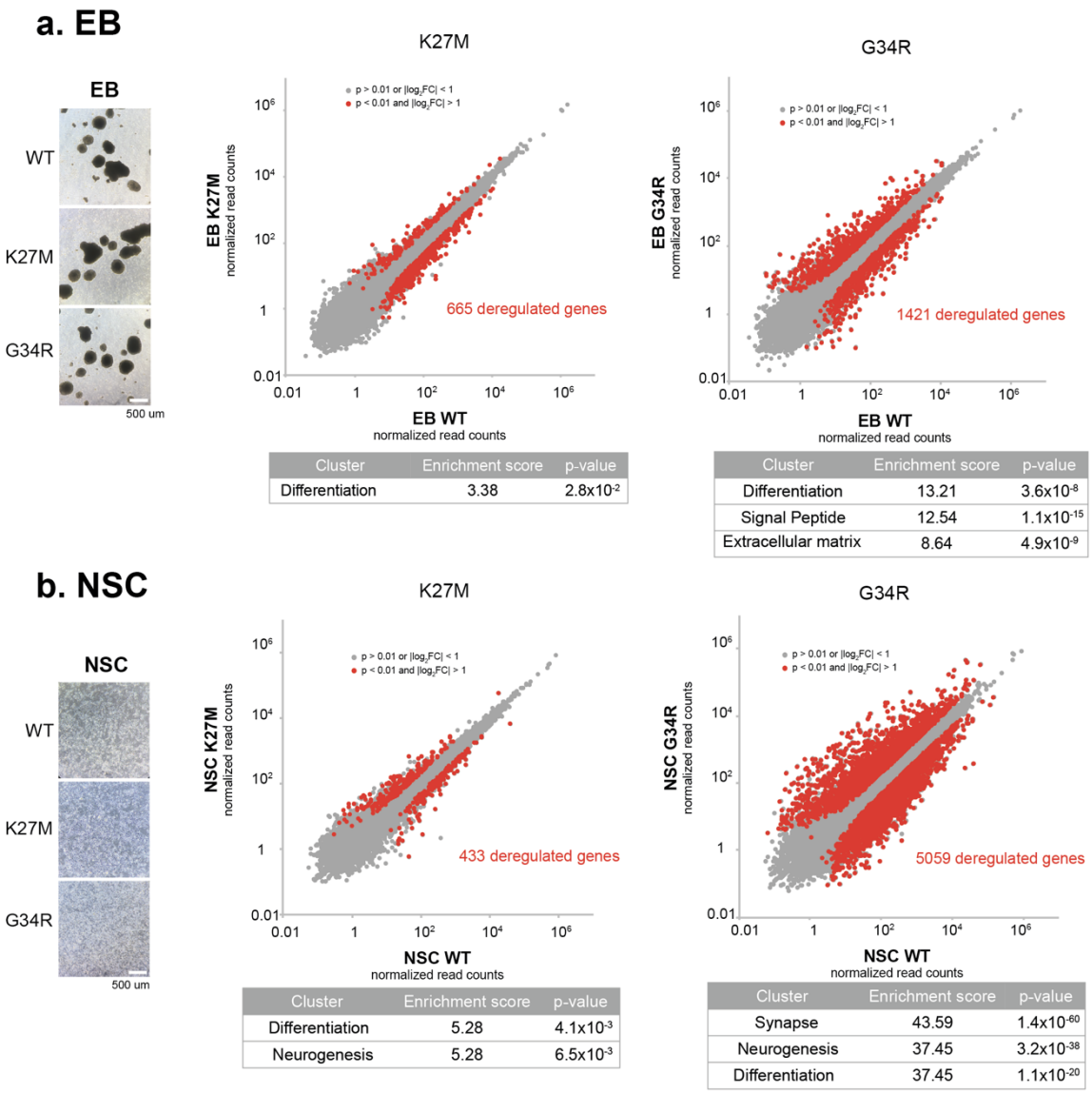


Figure 55: H3.3 mutations lead to transcriptional deregulation and mESC-to-NSC differentiation failure

Gene differential expression analysis for **a.** Day 4 embryonic bodies - EB K27M (n=3, left) of EB G34R (n=3, right) vs. EB H3.3WT (n=3) and for **b.** Day 9 neural stem cells - NSC K27M (n=2, left) or NSC G34R (n=2, right) vs. NSC H3.3WT (n=2). Differentially expressed genes are shown in red (adjusted p-value < 0.01 and |log2FC|>1). Functional annotation clustering tables of the differentially expressed genes are shown under each graph (performed with DAVID 2.0). Photos of each step extracted from **Figure 53** are shown on the left.

3.3. DNA repetitive elements dynamics during mESC-to-NSC differentiation

The dynamics of DNA repetitive elements during mESC to NSC differentiation were first investigated for wildtype mESC.

3.3.1. DNA repetitive elements are repressed during wildtype mESC-to-NSC differentiation

During mESC to NSC differentiation, the vast majority of the DNA repetitive elements were repressed (**Figure 56**). 1010 out of the 1209 families of DNA repetitive elements were repressed during the ES to EB transition, while a single family (IAPEz-int) was overexpressed. During the EB to NSC transition, an additional set of families were repressed with the exception of 3 families which were upregulated (IAPEY3-int, MMERVK10C, RLTR4_Mm-int). The repressed families were 94 families continuously repressed from the ES to EB transition, and 8 families newly repressed (SYNREP_MM, (GGGAA)n, IAP-d-int, L1M3f, LTRIS2, Lx2A, Lx2A1, RLTR1B-int).

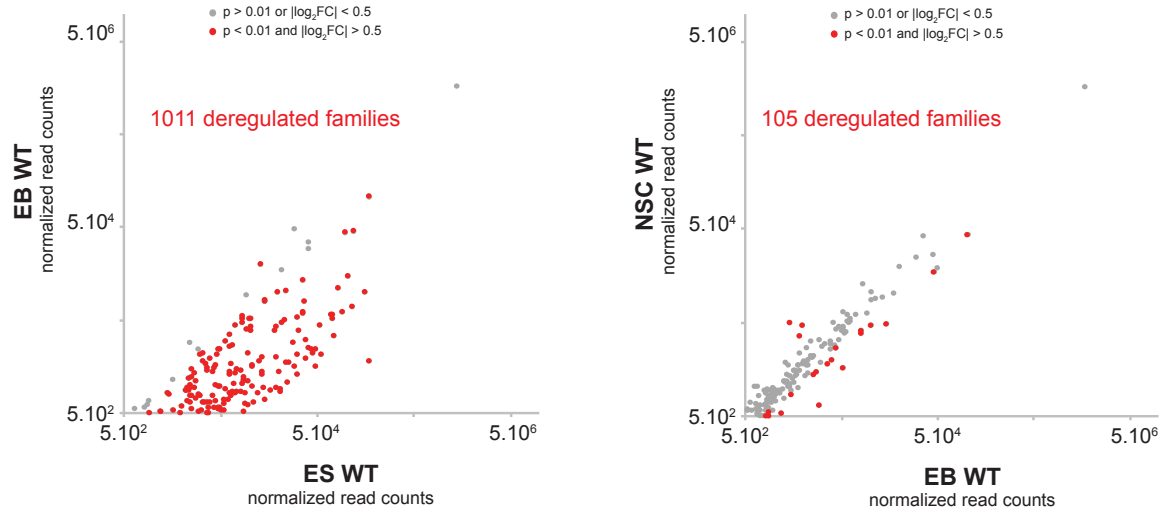


Figure 56: DNA repetitive elements are globally repressed during differentiation

DNA repetitive elements families' differential expression analysis during wildtype differentiation, for EB WT (n=3) vs. ES WT (n=6) (left) and NSC WT (n=2) vs. EB WT (n=3) (right). Differentially expressed DNA repetitive element families are shown in red (adjusted p-value < 0.01 and |log2FC|>0.5). 1011 and 105 DNA repetitive element families are deregulated (mainly repressed) during differentiation, respectively during ES to EB and EB to NSC transition. To note, only the families represented by more than 500 normalized read counts for each condition were represented on the scatter plots.

A global repression of DNA repetitive elements is occurring during the differentiation of wildtype mESC toward NSC. The impact of DNA repetitive elements dynamic under H3.3 mutant expression was then investigated.

3.3.2. DNA repetitive elements are deregulated in H3.3 mutant context upon differentiation

At the EB step, no DNA repetitive elements families were deregulated for K27M compared to the WT. This result was surprising considering the 600 genes deregulated (**Figure 57a left**). Nevertheless, taking into account the overexpression of ERVs at the mESC stage, the dynamic of inter-regulation between ERVs/genes is potentially shifted with a repression of those families taking more time and leading to gene deregulation. A defect in the timing of DNA repetitive elements regulation might have a direct impact on gene regulation. On G34R side, 12 families were significantly upregulated at the EB step compared to WT (**Figure 57a right**). Out of them, 11 were among the most overexpressed ERVs at the mESC stage (IAPEy-int, ETnERV-int, ETnERV2-int, RLTR45-int, RLTR1B-int, IAPLTR1_Mm, LTRIS2, MMETn-int, RLTRETN_Mm, MERVL-int and IAPEz-int). In addition to those ERVs, a LINE family was also upregulated in G34R EB compared to WT (L1Md_A). The remainder 11 upregulated ERVs families is suggesting a delay in repeats repression under G34R mutant.

At the NSC step, 18 and 28 families of DNA repetitive elements were found deregulated in K27M and G34R respectively. On K27M side, 3 ERVs were downregulated (ETnERV3-int, IAPEY3-int, MMERGLN-int) and the 15 upregulated ones were composed of ERVs (e.g. RLTR1B-int, MMERVK10C, IAP-d-int, IAPEy-int) and LINEs (L1_Mur and Lx subfamilies) (**Figure 57b left**). For G34R, 4 families out of the 28 deregulated in NSC were downregulated (MMSAT4, MMETn-int, ETnERV3-int, IAPEY3-int). The 24 others were overexpressed and 6 of them were already overexpressed at the EB stage (IAPEy-int, RLTR45-int, RLTR1B-int, IAPLTR1_Mm, LTRIS2, IAPEz-int). The 18 other overexpressed families were composed of ERVs (e.g. MMERVK10C-int, IAP-d-int) but also several LINEs (e.g. L1_Mur and Lx subfamilies) (**Figure 57b right**). Thus, under H3.3 mutation expression, a subset of ERVs as well as some LINEs are deregulated during mESC to NSC differentiation, suggesting an additional role of H3.3 in LINEs regulation in the last step of

differentiation. Interestingly, like ERVs, the targeted LINEs are also CG-rich, recently integrated, and potentially functional (Castro-Diaz et al., 2015). In addition, they have been reported to be specifically regulated through DNA methylation in differentiated cells in line with their later emergence during differentiation (Papin et al., 2017).

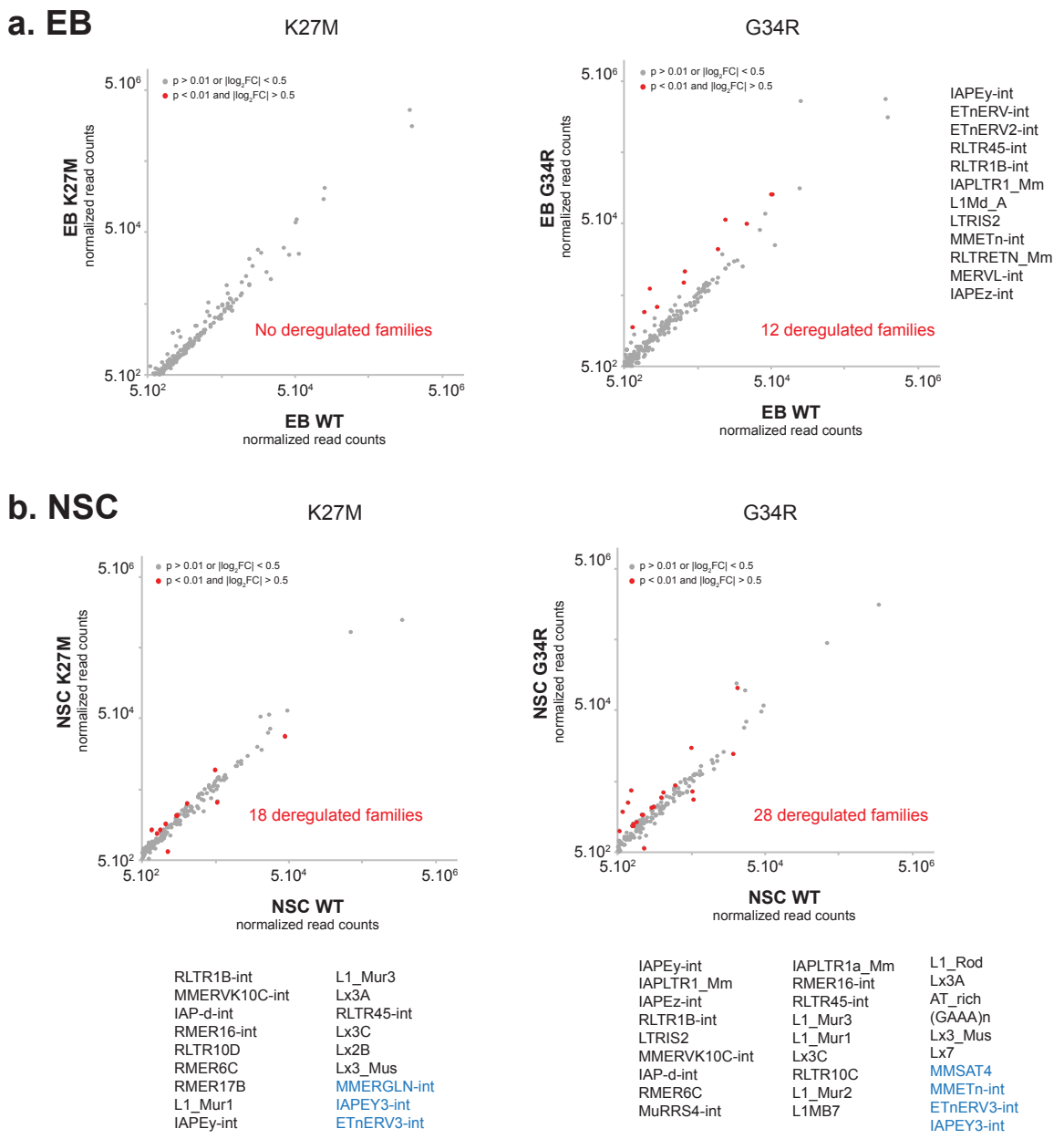


Figure 57: H3.3 mutation lead to deregulation of DNA repetitive elements during differentiation

(Legend continues on next page)

DNA repetitive elements families' differential expression analysis at each differentiation step between mutant and wildtype. **a.** EB step: EB K27M (n=3) vs. EB WT (n=3) (left) and EB G34R (n=3) vs. EB WT (n=3) (right). List of the families overexpressed in H3.3 G34R EB is shown on the right sorted in line with the level of deregulation. **b.** NSC step: NSC K27M (n=2) vs. NSC WT (n=2) (left) and NSC G34R (n=2) vs. NSC WT (n=2) (right). List of the families de-regulated are shown below the graphs, sorted in line with the level of deregulation. Families which are up-regulated are shown in black, families which are down-regulated are shown in blue. Differentially expressed DNA repetitive element families are shown in red (adjusted p-value < 0.01 and |log2FC|>0.5).

3.4. H3.3 plays a role in ERV regulation during differentiation

To have a better view of repetitive elements dynamics over neural differentiation, we analyzed the expression patterns of 'repeatome' across the different stages of differentiation in the wildtype context. DNA repeat families whose expression varied during differentiation were clustered according to their variation direction (see Material and Methods).

3.4.1. H3.3 marks ERVs that are expressed during differentiation

The majority of the DNA repetitive elements are repressed in early differentiation, from the transition to ES toward EB (clusters 1-2-3, **Figure 58a** and **Figure 59a**). A small subset is continuously repressed (cluster 4), variably expressed (cluster 5) or stably expressed (cluster 6) toward the full differentiation (**Figure 58a** and **Figure 59a**). The latter are differentially expressed over differentiation so they are very likely to play a role in the process of differentiation. Interestingly, the majority of those families were enriched in H3.3 and/or deregulated under H3.3 mutant enrichment (e.g. IAPEY3-int, IAPEz-int, LTRIS2, RLTR1B, RLTR4_Mm, RLTR4_MM-int, RLTR9E, RLTRETN_Mm, **Figure 60**). Thus, the differential expression of the DNA repetitive elements was further investigated in the context of H3.3 mutant over differentiation.

3.4.2. ERVs marked by H3.3 are de-regulated upon H3.3 mutant expression

Clusters of differentially expressed DNA repetitive element families determined for H3.3 WT differentiation were applied to H3.3 K27M and G34R. While early repressed repeats were globally repressed in the same manner than H3.3 wildtype (Clusters 1-2-3), several families from the clusters 4-5-6 were deregulated in the context of H3.3 mutant (**Figure 58b-c, Figure 59b-c**). The set of families which are potentially important for mESC-to-NSC differentiation are deregulated under H3.3 mutant expression. Of note, ERV families which are enriched in H3.3 in mESC and overexpressed under mutant expression are also deregulated during differentiation. In addition to recently integrated families of ERVs on which H3.3 has been found in mESC, a set of recently integrated LINEs are also part of the differentially expressed clusters and some of them are deregulated in H3.3 mutants (**Figure 60**).

The deregulation of DNA repetitive elements under H3.3 mutants is variable. Indeed, some families such as ERVs enriched in H3.3 are already overexpressed at the mESC stage and fail to be repressed on time or at the same level than H3.3 wildtype (e.g. RLTR1B-int, LTRIS2, **Figure 60** and **Figure 61**). Other families are being even more overexpressed instead of being repressed during differentiation (e.g. IAPEz-int for G34R, **Figure 60** and **Figure 61**).

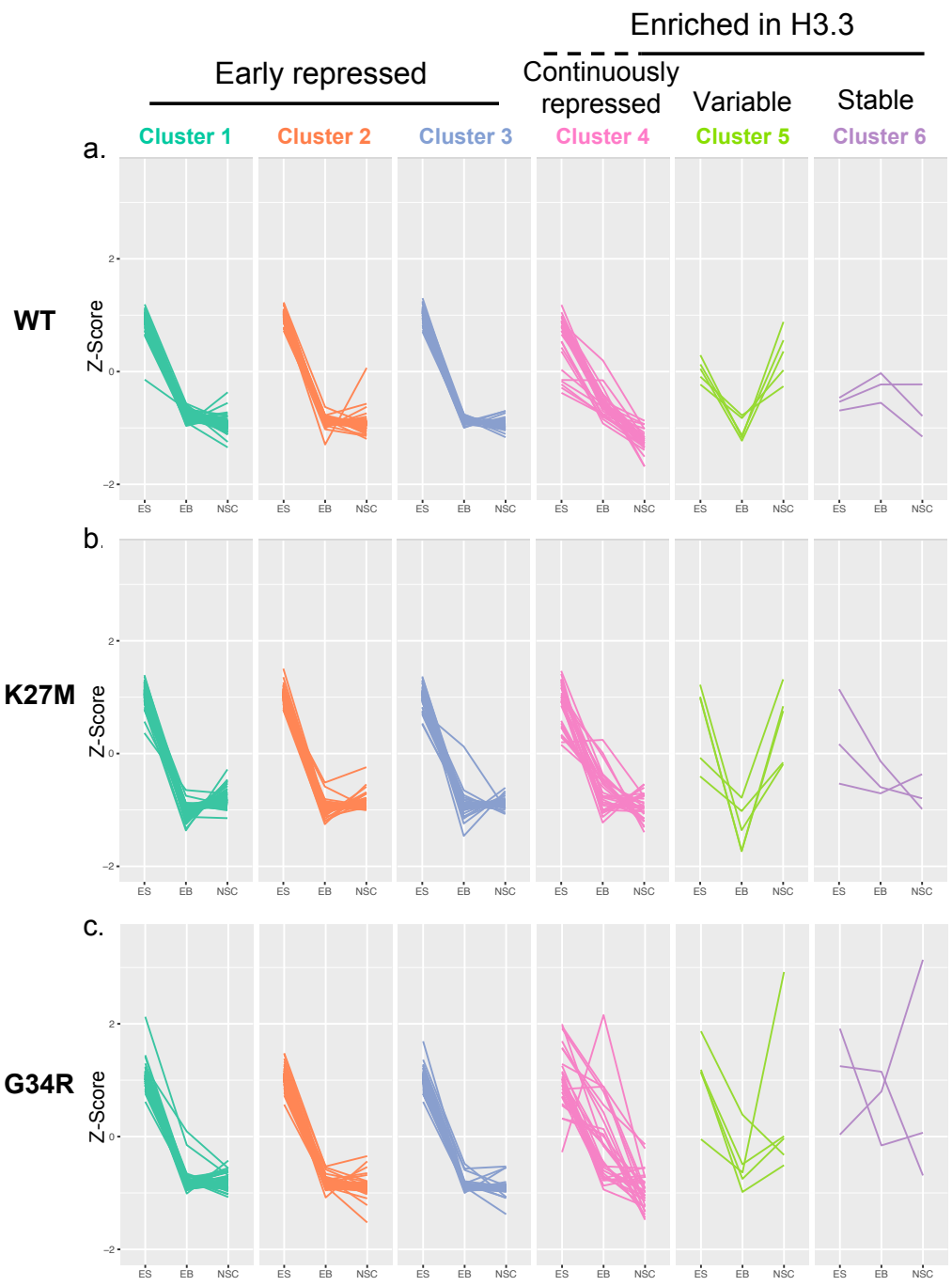


Figure 58: ERVs marked by H3.3 are de-regulated upon H3.3 mutant expression

a. Clustering of the DNA repetitive elements which are differentially expressed over the differentiation of WT mESC toward NSC. Each line represents a DNA repetitive element family. Z-score = (repeat value – mean)/standard deviation. Clusters 1-2-3: early repressed (n=419), Cluster 4: continuously repressed (n=30), Cluster 5: variable (n=5), Cluster 6: stable (n=3). Clusters 5 and 6 and part of cluster 4 families are enriched in H3.3 in mESC (recently integrated ERVs). Same clusters have been applied for **b.** K27M and **c.** G34R mutants. Clusters 5 and 6 and part of cluster 4 families are enriched in H3.3 in mESC (recently integrated ERVs) and are deregulated under mutant expression.

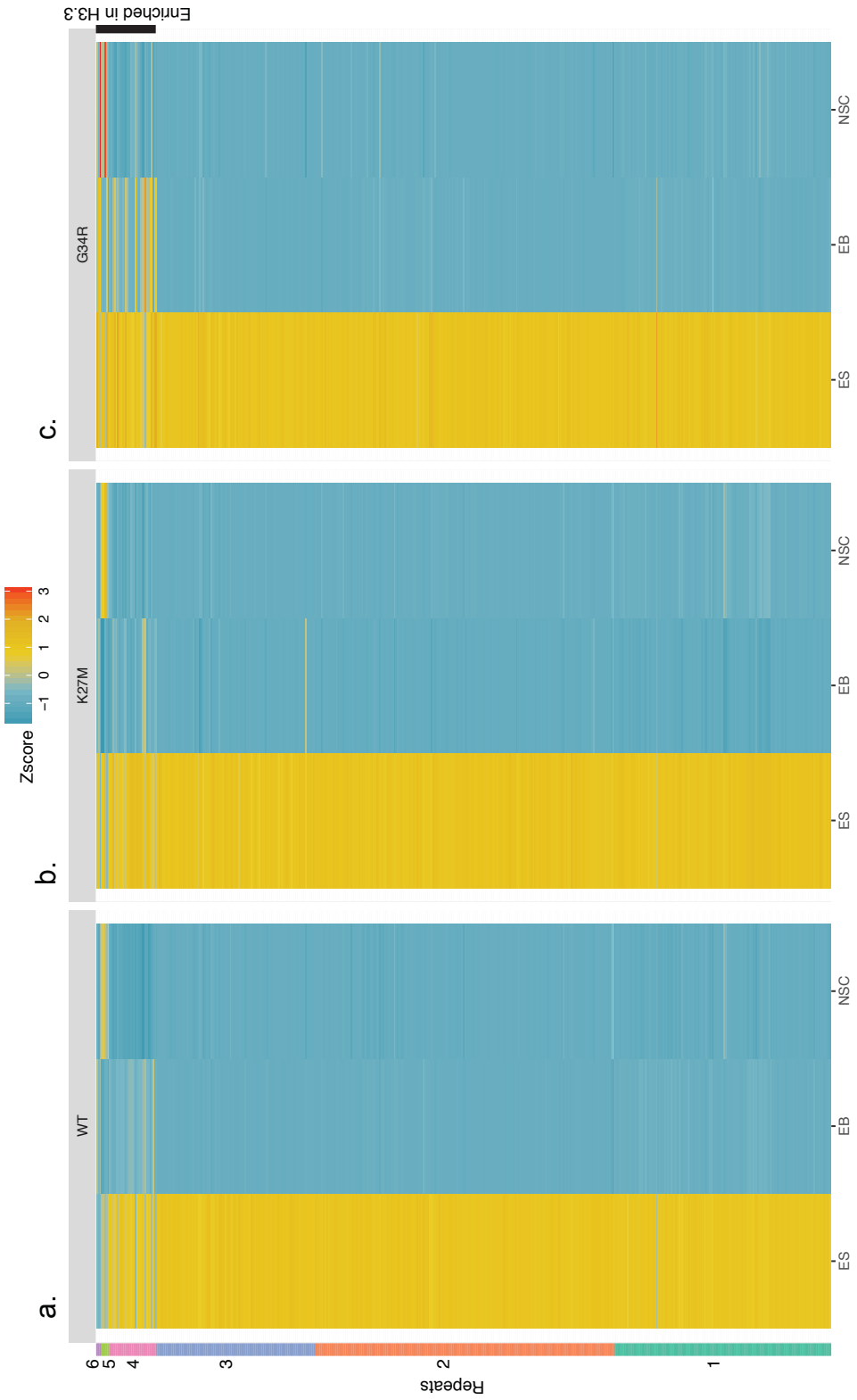


Figure 59: DNA repetitive elements potentially important for differentiation are deregulated under H3.3 mutant expression

Heatmaps of the repetitive elements expression from the 6 clusters shown in Figure 58 for the 3 differentiation steps (ES>EB>NSC), for H3.3 WT, K27M and G34R. Clusters are indicated on the left.

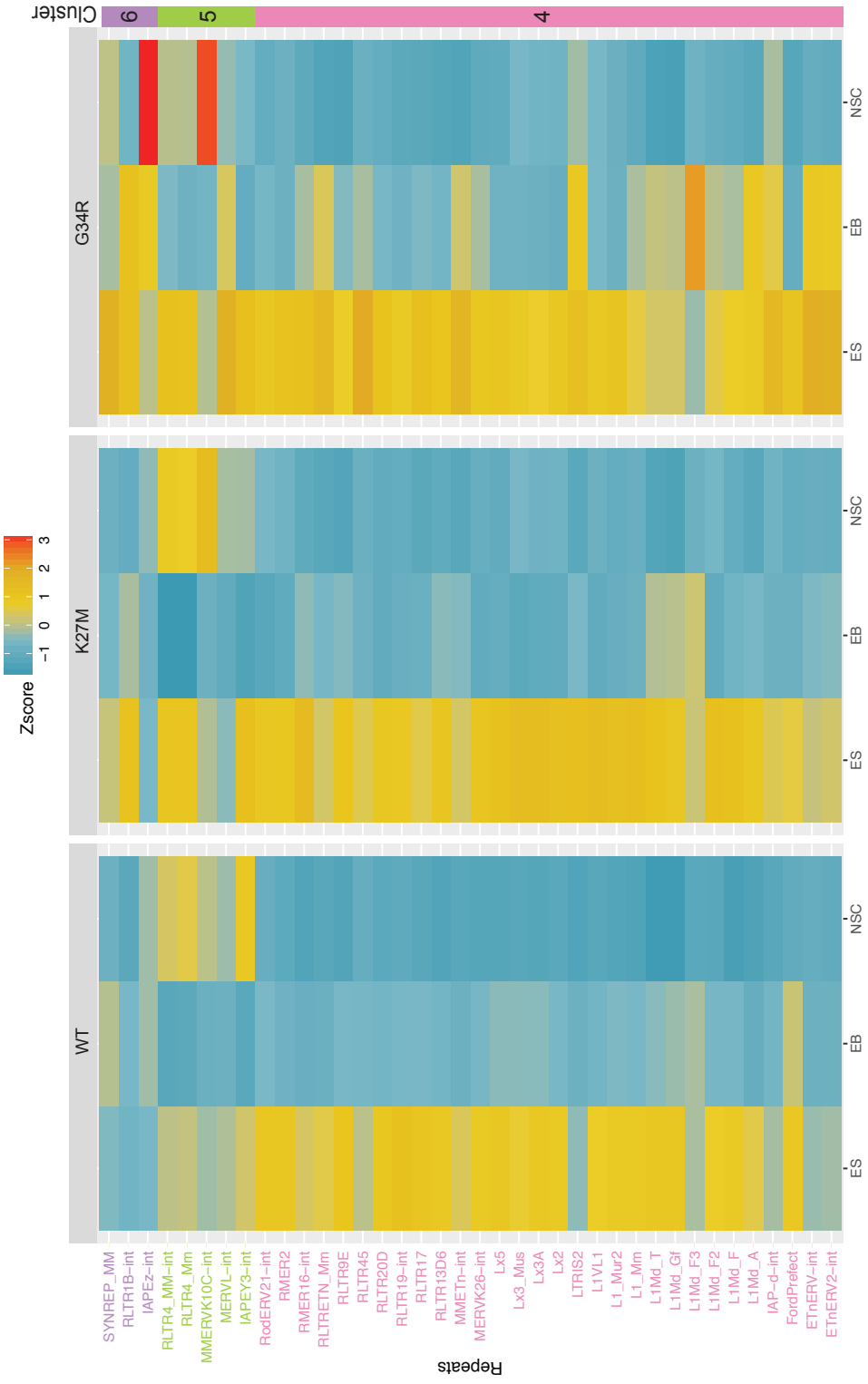


Figure 60: Zoom on Cluster 4-5-6 of **Figure 59**

Heatmaps of the repetitive elements expression zoomed on clusters 4-5-6 shown in **Figure 59** for the 3 differentiation steps (ES>EB>NSC), for H3.3 WT, K27M and G34R. Clusters are indicated on the right, DNA repetitive element families on the left.

As described in **Figure 56**, a wide repression of DNA repetitive elements is occurring over differentiation in the wildtype context. However, 3 recently integrated ERV families have been shown to be reactivated at the NSC stage: MMERVK10C, IAPEY3-int and RLTR4_Mm-int. The late overexpression of those families is also impaired under H3.3 mutant expression (**Figure 61**). For instance, MMERVK10C was over activated in mutant compared to wildtype from EB to NSC transition. IAPEY3-int was expressed at a higher level in mutant compared to wildtype at ES stage, got repressed at the EB stage, but its reactivation toward NSC in mutants failed to reach the same level than wildtype. RLTR4_MM-int was deregulated at each step of differentiation in mutants compared to wildtype, highlighting the complexity of the deregulation led by the presence of H3.3 mutant K27M or G34R.

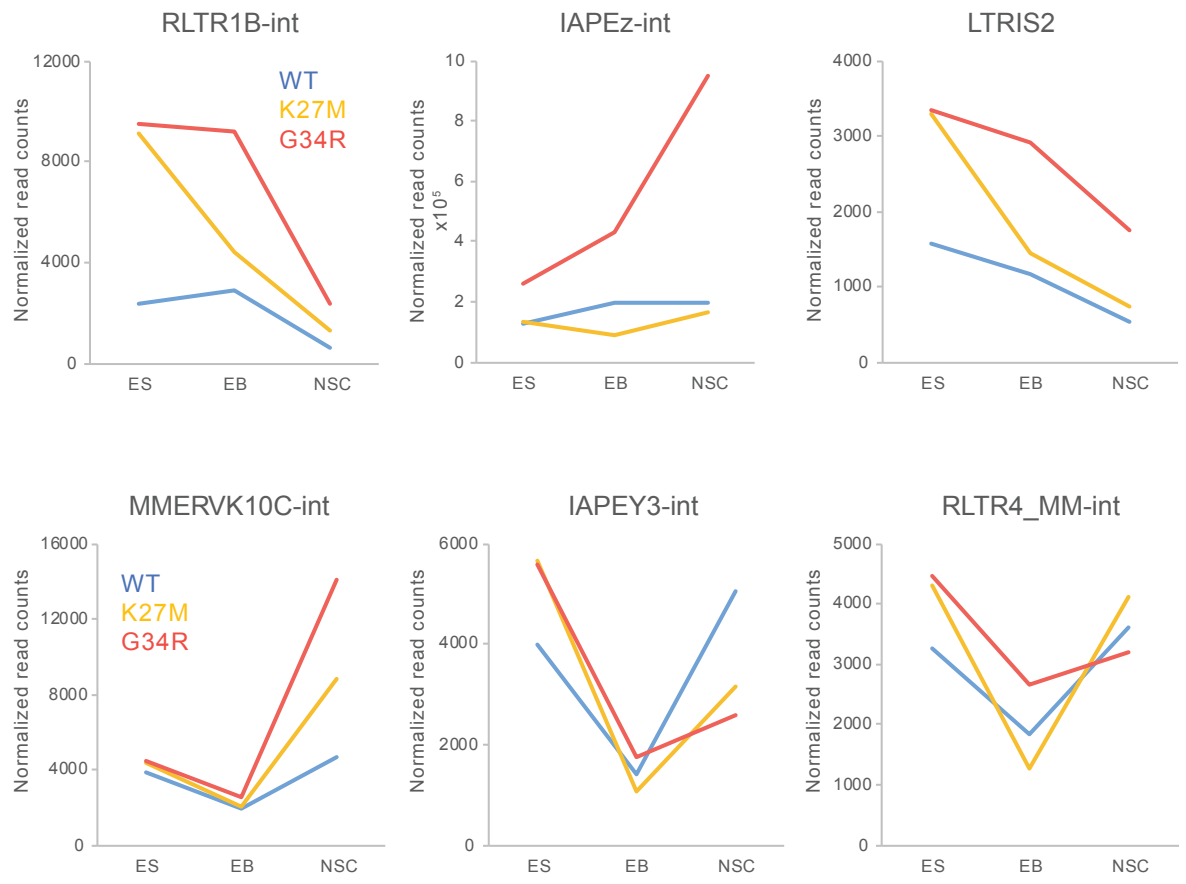


Figure 61: Global deregulation of ERV enriched with H3.3 and/or activated during differentiation under H3.3 mutant expression during mESC-to-NSC differentiation

Expression variation of ERV families shown in Figure 59 for the 3 differentiation steps (ES>EB>NSC), for H3.3 WT (blue), K27M (yellow) and G34R (red).

3.5. Discussion

Directing the differentiation of mESCs to specific cell fates in vitro is a powerful tool to recapitulate developmental processes and permits understanding of how repetitive elements regulate these processes and how H3.3 mutations impact the differentiation process. To have a better view of repetitive elements dynamics during neural differentiation, we analyzed their expression pattern across the different stages of differentiation.

Our data suggest an important role of H3.3 in differentiation through regulation of recently integrated CG-rich and potentially functional retroelements (ERVs and LINEs). Indeed, a general trend of repression of DNA repetitive elements over mESC-to-NSC differentiation has been observed with the exception of a small subset of families which were differentially expressed over time. The majority of the latter were found enriched for H3.3 and strongly impacted by its mutations. In addition, mESC expressing H3.3 K27M or G34R failed to differentiate into NSC and exhibited a strong deregulation of H3.3-enriched ERVs. A subset of recently integrated CG-rich LINEs was also de-regulated during differentiation. H3.3 might also become enriched at these LINE families during differentiation and participate in their regulation. The deregulation of those ERV/LINE families is most likely deregulating the nearby genes driving differentiation. In order to validate H3.3 localization at newly deregulated retrotransposons families, H3.3 enrichment should be performed by repeating native FLAG-HA ChIP-seq at EB and NSC stage in the wildtype and mutants (e.g. MMERVKC10-int or subset of LINEs). In addition, as performed for repeats, a clustering of gene expression over differentiation should be done. We could then investigate if genes from similar clusters with the same expression variation over time, are associated/close to specific variating retroelements families. To understand the failure of differentiation in H3.3 mutants, the distance between the deregulated genes and the deregulated retroelements should be thoroughly investigated. We could test all the genes in the vicinity of the deregulated retroelements (<10 kb) and verify if they are involved in differentiation. Another way to confirm the impact of retroelements on gene regulation would be to visualize if the genes whose expression vary over differentiation are grouped by clusters on the genome and if those regions are enriched in recently integrated CG-rich retroelements.

Interestingly, downregulated genes from EB to NSC wildtype differentiation are enriched for KRAB-ZFPs. KRAB-ZFPs are known for their repressor role on DNA repetitive elements (Ecco et al., 2017; Yang et al., 2017) and their downregulation (EB>NSC) is coming right after the global repression of DNA repetitive elements (ES>EB). H3.3 WT and mutants have been shown to bind the KAP1 repressive complex in mESC, which is known to be recruited by KRAP-ZFPs (Friedman et al., 1996; Ecco et al., 2017). The KRAB/KAP1 system has been reported to be the key element in the control of ERV/LINEs (L1) in mESC (Castro-Diaz et al., 2015). The overexpression of H3.3 mutant-bound ERVs as well as the global deregulation of recently integrated ERV/LINEs during differentiation is likely due to the decreased capacity of H3.3 mutant to recruit or activate the repression machineries such as KAP1-dependent recruitment of DNMTs and SETDB1.

I hereby propose that H3.3 has a major role in neural differentiation through the regulation of recently integrated and potentially functional retroelements such as subsets of ERVs and LINEs.

IV. Conclusion and Perspectives

While DNA repetitive elements constitute more than half of the human genome, the majority of the current genome-wide studies do not take them into account because of their low sequencing coverage. Since their identification in 2012, histone variant H3.3 K27M and G34R/V mutations have been intensively studied in the field of pediatric high-grade glioma, although little is known about the role of H3.3 mutations in the development of this deadly cancer.

In this study, we have analyzed the genomic distribution of the H3.3 K27M and G34R mutants in mESC. Our data show that there is no difference in the H3.3 distribution between WT and mutants at promoters and enhancers suggesting that K27M or G34R mutations have no effect on the mechanism of H3.3 deposition. Mild gene transcription deregulation has been observed under mutant expression but could not be linked to the level of H3.3 enrichment at transcriptional start sites. Nevertheless, H3.3 has been shown to be located at a subset of recently integrated and potentially functional ERVs which were overexpressed in the presence of the mutant. ERVs overexpression has been further proposed to be responsible of the deregulation of neighboring genes.

Through differentiation of mouse embryonic stem cells in neural stem cells, transposable elements have been shown to become widely repressed during differentiation. A subset of transposons composed mainly of recently integrated CG-rich ERVs and LINEs remains differentially expressed over the differentiation and is thus suggested to play a key role in the regulation of differentiation. The majority of those expressed ERVs is enriched in H3.3 in mESC and de-regulated under mutant H3.3. In addition, H3.3 mutations have led to a failure of differentiation in NSC together with a massive retrotransposons deregulation. This work places H3.3 as a critical player in neural differentiation through the regulation of recently integrated retrotransposons (ERVs, LINEs). While ERV regulation seems to be important in the early differentiation stage, families of recently integrated LINEs are entering the game later. H3.3 might sequentially travel from retrotransposon families to others and play a role in the regulation of the differentiation steps (e.g. from subsets of ERVs to other ERVs or LINEs). Whether H3.3 is only vital in neural differentiation or plays a general role in differentiation through retrotransposons regulation remains to be elucidated. Other differentiation systems will be used to decipher whether the role of H3.3 in differentiation is specific or global.

The next step will be to understand the mechanism behind H3.3 regulation of recently integrated retroelements, with the latter proposed to be mostly repressed by SETDB1-dependent H3K9me3 in mESC. However, the recently integrated CG-rich ERVs and LINEs are also specifically targeted by a dynamic methylation/demethylation process in mESC, suggesting a role of DNA methylation in their regulation. We propose that the deregulation of recently integrated H3.3-enriched retroelements is likely due to the decreased capacity of the H3.3 mutant to recruit or activate the repression machineries such as KAP1-dependent recruitment of DNMTs and SETDB1. We will perform biochemical interaction studies, as well as enrichment analysis of KAP1, H3K9me3 and DNA methylation in the wildtype and mutant context to decipher which of these factors' binding/activity is affected by H3.3 mutations.

Unlike studies comparing patient tumors, the current study's use of an mESC model has a great advantage as the genetic background is fully controlled. The only variation in the model is the presence or not of H3.3 mutation in the endogenous model (K27M or G34R). Thus, the mutation of a single copy of the *H3f3a* gene is sufficient for H3.3-enriched ERV overexpression and for failure of differentiation in neural stem cells.

A stronger phenotype could be observed for G34R compared to K27M, while G34R tumors have a better prognosis and are diagnosed later than K27M tumors. Of note, H3.3 G34R led to a specific overexpression of minor satellites in mESC, which could cause genomic instability. If applied to the tumor development, a too strong phenotype could be detrimental and possibly lead to cell death, and thus to a slower tumor formation. On the other hand, Chiappinelli et al. (2015) have reported that strong upregulation of endogenous retroviruses in tumor cells induces a growth-inhibiting immune response. So, the stronger ERV overexpression caused by G34R could have an anti-tumor-like effect. Whether overexpression of retroelements constitute a threat or a benefit to the cell remains to be determined. Our current view highlights a threat for a mild retrotransposons' overexpression to cause gene deregulation and potentially genomic instability which could further constitute a favorable environment for tumorigenesis. On the other hand, a more important retrotransposons overexpression could constitute a benefit for the organism by inducing a growth-inhibiting immune response. Repetitive elements have played an undeniable positive role in evolution,

but increasing evidence highlights their implication in several cancers, suggesting that their domestication is subjected to a finely-tuned regulation balance.

The somatic onset of H3.3 K27M or G34R/V is thus very likely to de-regulate the non-coding genome expression further contributing to a favorable environment for tumor development. In addition, Krug et al. (2019) highlighted the importance of H3.3 mutation in tumor maintenance. We showed that knockdown of H3.3 K27M in patient-derived xenografts led to the repression of recently integrated ERVs together with an anti-tumor effect, validating the direct role of H3.3 mutation on ERV deregulation and on tumor maintenance.

Understanding the molecular mechanism of recently integrated retroelements' regulation by H3.3 and deciphering the thresholds of benefit/threat in the expression of retroelements will pave the way toward new pHGG treatment strategies.

Future research should also address the capacity to retrotranspose of the de-regulated recently integrated retroelements. The present study focused on their transcription but did not investigate whether or not they are translated and have the potential to retrotranspose. Indeed, high retrotransposition rate of retroelements is very likely to promote genomic instability and novel mutations, further constituting a conducive environment for tumor development.

Recently integrated repetitive elements are proposed to take part in the genome regulation dynamics and to represent a fast and global way to modify the genome expression on a large scale. In a nutshell, H3.3 is proposed to regulate neural differentiation through the regulation of recently integrated CG-rich retrotransposons. H3.3 mutations lead to deregulation of those retrotransposons and impaired differentiation. Deregulation of recently integrated and potentially functional retroelements under H3.3 K27M and G34R mutations represent a novel mechanism potentially implicated in pediatric high-grade glioma development and maintenance.

V. References

Adam, S., Polo, S. E., and Almouzni, G. (2013). Transcription recovery after DNA damage requires chromatin priming by the H3.3 histone chaperone HIRA. Cell *155*, 94-106.

Ahmad, K., and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol Cell *9*, 1191-1200.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

Apel, A., Zentgraf, H., Buchler, M. W., and Herr, I. (2009). Autophagy-A double-edged sword in oncology. Int J Cancer *125*, 991-995.

Artandi, S. E., and DePinho, R. A. (2010). Telomeres and telomerase in cancer. Carcinogenesis *31*, 9-18.

Avery, O. T., Macleod, C. M., and McCarty, M. (1944). Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. J Exp Med *79*, 137-158.

Bachu, M., Tamura, T., Chen, C., Narain, A., Nehru, V., Sarai, N., Ghosh, S. B., Ghosh, A., Kavarthapu, R., Dufau, M. L., and Ozato, K. (2019). A versatile mouse model of epitope-tagged histone H3.3 to study epigenome dynamics. J Biol Chem *294*, 1904-1914.

Baeriswyl, V., and Christofori, G. (2009). The angiogenic switch in carcinogenesis. Semin Cancer Biol *19*, 329-337.

Bailey, C. P., Figueroa, M., Mohiuddin, S., Zaky, W., and Chandra, J. (2018). Cutting Edge Therapeutic Insights Derived from Molecular Biology of Pediatric High-Grade Glioma and Diffuse Intrinsic Pontine Glioma (DIPG). Bioengineering (Basel) *5*.

Banaszynski, L. A., Wen, D., Dewell, S., Whitcomb, S. J., Lin, M., Diaz, N., Elsasser, S. J., Chapgier, A., Goldberg, A. D., Canaani, E.*, et al.* (2013). Hira-dependent histone H3.3 deposition facilitates PRC2 recruitment at developmental loci in ES cells. Cell *155*, 107-120.

Bannert, N., Hofmann, H., Block, A., and Hohn, O. (2018). HERVs New Role in Cancer: From Accused Perpetrators to Cheerful Protectors. Front Microbiol *9*, 178.

Bannert, N., and Kurth, R. (2004). Retroelements and the human genome: new perspectives on an old relation. Proc Natl Acad Sci U S A *101 Suppl 2*, 14572-14579.

Bax, D. A., Mackay, A., Little, S. E., Carvalho, D., Viana-Pereira, M., Tamber, N., Grigoriadis, A. E., Ashworth, A., Reis, R. M., Ellison, D. W.*, et al.* (2010). A distinct spectrum of copy number aberrations in pediatric high-grade gliomas. Clin Cancer Res *16*, 3368-3377.

Beck, C. R., Garcia-Perez, J. L., Badge, R. M., and Moran, J. V. (2011). LINE-1 elements in structural variation and disease. Annu Rev Genomics Hum Genet *12*, 187-215.

Bednar, J., Garcia-Saez, I., Boopathi, R., Cutter, A. R., Papai, G., Reymer, A., Syed, S. H., Lone, I. N., Tonchev, O., Crucifix, C.*, et al.* (2017). Structure and Dynamics of a 197 bp Nucleosome in Complex with Linker Histone H1. Mol Cell *66*, 384-397 e388.

Behjati, S., Tarpey, P. S., Presneau, N., Scheipl, S., Pillay, N., Van Loo, P., Wedge, D. C., Cooke, S. L., Gundem, G., Davies, H.*, et al.* (2013). Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. Nat Genet *45*, 1479-1482.

Belancio, V. P., Hedges, D. J., and Deininger, P. (2008). Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res *18*, 343-358.

Bender, S., Tang, Y., Lindroth, A. M., Hovestadt, V., Jones, D. T., Kool, M., Zapatka, M., Northcott, P. A., Sturm, D., Wang, W.*, et al.* (2013). Reduced H3K27me3 and DNA hypomethylation are major drivers of gene expression in K27M mutant pediatric high-grade gliomas. Cancer Cell *24*, 660-672.

Bernstein, E., Duncan, E. M., Masui, O., Gil, J., Heard, E., and Allis, C. D. (2006). Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol Cell Biol *26*, 2560-2569.

Berx, G., and van Roy, F. (2009). Involvement of members of the cadherin superfamily in cancer. Cold Spring Harb Perspect Biol *1*, a003129.

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209-213.

Biswas, S., and Rao, C. M. (2018). Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. Eur J Pharmacol *837*, 8-24.

Bjerke, L., Mackay, A., Nandhabalan, M., Burford, A., Jury, A., Popov, S., Bax, D. A., Carvalho, D., Taylor, K. R., Vinci, M.*, et al.* (2013). Histone H3.3. mutations drive pediatric glioblastoma through upregulation of MYCN. Cancer Discov *3*, 512-519.

Blomberg, J., Benachenhou, F., Blikstad, V., Sperber, G., and Mayer, J. (2009). Classification and nomenclature of endogenous retroviral sequences (ERVs): problems and recommendations. Gene *448*, 115-123.

Boeke, J. D., and Stoye, J. P. (1997). Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements. In Retroviruses, J.M. Coffin, S.H. Hughes, and H.E. Varmus, eds. (Cold Spring Harbor (NY)).

Braunstein, S., Raleigh, D., Bindra, R., Mueller, S., and Haas-Kogan, D. (2017). Pediatric high-grade glioma: current molecular landscape and therapeutic approaches. J Neurooncol *134*, 541-549.

Brennan, C. W., Verhaak, R. G., McKenna, A., Campos, B., Noushmehr, H., Salama, S. R., Zheng, S., Chakravarty, D., Sanborn, J. Z., Berman, S. H.*, et al.* (2013). The somatic genomic landscape of glioblastoma. Cell *155*, 462-477.

Brown, Z. Z., Muller, M. M., Jain, S. U., Allis, C. D., Lewis, P. W., and Muir, T. W. (2014). Strategy for "detoxification" of a cancer-derived histone mutant based on mapping its interaction with the methyltransferase PRC2. J Am Chem Soc *136*, 13498-13501.

Buczkowicz, P., Hoeman, C., Rakopoulos, P., Pajovic, S., Letourneau, L., Dzamba, M., Morrison, A., Lewis, P., Bouffet, E., Bartels, U.*, et al.* (2014). Genomic analysis of diffuse intrinsic pontine gliomas identifies three molecular subgroups and recurrent activating ACVR1 mutations. Nat Genet *46*, 451-456.

Buschbeck, M., and Hake, S. B. (2017). Variants of core histones and their roles in cell fate decisions, development and cancer. Nat Rev Mol Cell Biol *18*, 299-314.

Bush, K. M., Yuen, B. T., Barrilleaux, B. L., Riggs, J. W., O'Geen, H., Cotterman, R. F., and Knoepfler, P. S. (2013). Endogenous mammalian histone H3.3 exhibits chromatin-related functions during development. Epigenetics Chromatin *6*, 7.

Campos, E. I., and Reinberg, D. (2009). Histones: annotating chromatin. Annu Rev Genet *43*, 559-599.

Carrozza, M. J., Li, B., Florens, L., Suganuma, T., Swanson, S. K., Lee, K. K., Shia, W. J., Anderson, S., Yates, J., Washburn, M. P., and Workman, J. L. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581-592.

Castel, D., Philippe, C., Kergrohen, T., Sill, M., Merlevede, J., Barret, E., Puget, S., Sainte-Rose, C., Kramm, C. M., Jones, C.*, et al.* (2018). Transcriptomic and epigenetic profiling of 'diffuse midline gliomas, H3 K27M-mutant' discriminate two subgroups based on the type of histone H3 mutated and not supratentorial or infratentorial location. Acta Neuropathol Commun *6*, 117.

Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., Duc, J., Jang, S. M., Turelli, P., and Trono, D. (2014). Evolutionally dynamic L1 regulation in embryonic stem cells. Genes Dev *28*, 1397-1409.

Castro-Diaz, N., Friedli, M., and Trono, D. (2015). Drawing a fine line on endogenous retroelement activity. Mob Genet Elements *5*, 1-6.

Chan, K. M., Fang, D., Gan, H., Hashizume, R., Yu, C., Schroeder, M., Gupta, N., Mueller, S., James, C. D., Jenkins, R.*, et al.* (2013a). The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. Genes Dev *27*, 985-990.

Chan, K. M., Han, J., Fang, D., Gan, H., and Zhang, Z. (2013b). A lesson learned from the H3.3K27M mutation found in pediatric glioma: a new approach to the study of the function of histone modifications in vivo? Cell Cycle *12*, 2546-2552.

Chen, P., Zhao, J., Wang, Y., Wang, M., Long, H., Liang, D., Huang, L., Wen, Z., Li, W., Li, X.*, et al.* (2013). H3.3 actively marks enhancers and primes gene transcription via opening higher-ordered chromatin. Genes Dev *27*, 2109-2124.

Chiappinelli, K. B., Strissel, P. L., Desrichard, A., Li, H., Henke, C., Akman, B., Hein, A., Rote, N. S., Cope, L. M., Snyder, A.*, et al.* (2015). Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. Cell *162*, 974-986.

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell *168*, 442-459 e420.

Clapier, C. R., Iwasa, J., Cairns, B. R., and Peterson, C. L. (2017). Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nat Rev Mol Cell Biol *18*, 407-422.

Clement, C., Orsi, G. A., Gatto, A., Boyarchuk, E., Forest, A., Hajj, B., Mine-Hattab, J., Garnier, M., Gurard-Levin, Z. A., Quivy, J. P., and Almouzni, G. (2018). High-resolution visualization of H3 variants during replication reveals their controlled recycling. Nat Commun *9*, 3181.

Codner, G. F., Lindner, L., Caulder, A., Wattenhofer-Donze, M., Radage, A., Mertz, A., Eisenmann, B., Mianne, J., Evans, E. P., Beechey, C. V.*, et al.* (2016). Aneuploidy screening of embryonic stem cell clones by metaphase karyotyping and droplet digital polymerase chain reaction. BMC Cell Biol *17*, 30.

Collins, P. L., Kyle, K. E., Egawa, T., Shinkai, Y., and Oltz, E. M. (2015). The histone methyltransferase SETDB1 represses endogenous and exogenous retroviruses in B lymphocytes. Proc Natl Acad Sci U S A *112*, 8367-8372.

Colombo, E., Giannelli, S. G., Galli, R., Tagliafico, E., Foroni, C., Tenedini, E., Ferrari, S., Ferrari, S., Corte, G., Vescovi, A.*, et al.* (2006). Embryonic stem-derived versus somatic neural stem cells: a comparative analysis of their developmental potential and molecular phenotype. Stem Cells *24*, 825-834.

Coluccio, A., Ecco, G., Duc, J., Offner, S., Turelli, P., and Trono, D. (2018). Individual retrotransposon integrants are differentially controlled by KZFP/KAP1-dependent histone methylation, DNA methylation and TET-mediated hydroxymethylation in naive embryonic stem cells. Epigenetics Chromatin *11*, 7.

Consortium, M. G. S., Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M.*, et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Cordero, F. J., Huang, Z., Grenier, C., He, X., Hu, G., McLendon, R. E., Murphy, S. K., Hashizume, R., and Becher, O. J. (2017). Histone H3.3K27M Represses p16 to Accelerate Gliomagenesis in a Murine Model of DIPG. Mol Cancer Res *15*, 1243-1254.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D.*, et al.* (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell *146*, 67-79.

Couldrey, C., Carlton, M. B., Nolan, P. M., Colledge, W. H., and Evans, M. J. (1999). A retroviral gene trap insertion into the histone 3.3A gene causes partial neonatal lethality, stunted growth, neuromuscular deficits and male sub-fertility in transgenic mice. Hum Mol Genet *8*, 2489-2495.

Couture, J. F., Collazo, E., Ortiz-Tello, P. A., Brunzelle, J. S., and Trievel, R. C. (2007). Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. Nat Struct Mol Biol *14*, 689-695.

Cox, S. G., Kim, H., Garnett, A. T., Medeiros, D. M., An, W., and Crump, J. G. (2012). An essential role of variant histone H3.3 for ectomesenchyme potential of the cranial neural crest. PLoS Genet *8*, e1002938.

Crichton, J. H., Dunican, D. S., Maclennan, M., Meehan, R. R., and Adams, I. R. (2014). Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline. Cell Mol Life Sci *71*, 1581-1605.

Criscione, S. W., Zhang, Y., Thompson, W., Sedivy, J. M., and Neretti, N. (2014). Transcriptional landscape of repetitive elements in normal and cancer human cells. BMC Genomics *15*, 583.

Curto, M., Cole, B. K., Lallemand, D., Liu, C. H., and McClatchey, A. I. (2007). Contact-dependent inhibition of EGFR signaling by Nf2/Merlin. J Cell Biol *177*, 893-903.

Das, C., Tyler, J. K., and Churchill, M. E. (2010). The histone shuffle: histone chaperones in an energetic dance. Trends Biochem Sci *35*, 476-489.

Davies, M. A., and Samuels, Y. (2010). Analysis of the genome to personalize therapy for melanoma. Oncogene *29*, 5545-5555.

Delbarre, E., Ivanauskiene, K., Spirkoski, J., Shah, A., Vekterud, K., Moskaug, J. O., Boe, S. O., Wong, L. H., Kuntziger, T., and Collas, P. (2017). PML protein organizes heterochromatin domains where it regulates histone H3.3 deposition by ATRX/DAXX. Genome Res *27*, 913-921.

Deng, H., Zeng, J., Zhang, T., Gong, L., Zhang, H., Cheung, E., Jones, C., and Li, G. (2018). Histone H3.3K27M Mobilizes Multiple Cancer/Testis (CT) Antigens in Pediatric Glioma. Mol Cancer Res *16*, 623-633.

Deniz, O., de la Rica, L., Cheng, K. C. L., Spensberger, D., and Branco, M. R. (2018). SETDB1 prevents TET2-dependent activation of IAP retroelements in naive embryonic stem cells. Genome Biol *19*, 6.

Derynck, R., and Weinberg, R. A. (2019). EMT and Cancer: More Than Meets the Eye. Dev Cell *49*, 313-316.

Diaz, A. K., and Baker, S. J. (2014). The genetic signatures of pediatric high-grade glioma: no longer a one-act play. Semin Radiat Oncol *24*, 240-247.

Dignam, J. D. (1990). Preparation of extracts from higher eukaryotes. Methods Enzymol *182*, 194-203.

Drane, P., Ouararhni, K., Depaux, A., Shuaib, M., and Hamiche, A. (2010). The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. Genes Dev *24*, 1253-1265.

Duarte, L. F., Young, A. R., Wang, Z., Wu, H. A., Panda, T., Kou, Y., Kapoor, A., Hasson, D., Mills, N. R., Ma'ayan, A.*, et al.* (2014). Histone H3.3 and its proteolytically processed form drive a cellular senescence programme. Nat Commun *5*, 5210.

Duhl, D. M., Vrieling, H., Miller, K. A., Wolff, G. L., and Barsh, G. S. (1994). Neomorphic agouti mutations in obese yellow mice. Nat Genet *8*, 59-65.

Dupressoir, A., Lavialle, C., and Heidmann, T. (2012). From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation. Placenta *33*, 663-671.

Ecco, G., Cassano, M., Kauzlaric, A., Duc, J., Coluccio, A., Offner, S., Imbeault, M., Rowe, H. M., Turelli, P., and Trono, D. (2016). Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues. Dev Cell *36*, 611-623.

Ecco, G., Imbeault, M., and Trono, D. (2017). KRAB zinc finger proteins. Development *144*, 2719-2729.

Elsasser, S. J., Noh, K. M., Diaz, N., Allis, C. D., and Banaszynski, L. A. (2015). Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. Nature *522*, 240-244.

Elsasser, S. J., Noh, K. M., Diaz, N., Allis, C. D., and Banaszynski, L. A. (2017). Elsasser et al. reply. Nature *548*, E7-E9.

Erwin, J. A., Marchetto, M. C., and Gage, F. H. (2014). Mobile DNA elements in the generation of diversity and complexity in the brain. Nat Rev Neurosci *15*, 497-506.

Fang, D., Gan, H., Cheng, L., Lee, J. H., Zhou, H., Sarkaria, J. N., Daniels, D. J., and Zhang, Z. (2018a). H3.3K27M mutant proteins reprogram epigenome by sequestering the PRC2 complex to poised enhancers. Elife *7*.

Fang, H. T., El Farran, C. A., Xing, Q. R., Zhang, L. F., Li, H., Lim, B., and Loh, Y. H. (2018b). Global H3.3 dynamic deposition defines its bimodal role in cell fate transition. Nat Commun *9*, 1537.

Fang, J., Huang, Y., Mao, G., Yang, S., Rennert, G., Gu, L., Li, H., and Li, G. M. (2018c). Cancer-driving H3G34V/R/D mutations block H3K36 methylation and H3K36me3-MutSalpha interaction. Proc Natl Acad Sci U S A *115*, 9598-9603.

Fasching, L., Kapopoulou, A., Sachdeva, R., Petri, R., Jonsson, M. E., Manne, C., Turelli, P., Jern, P., Cammas, F., Trono, D., and Jakobsson, J. (2015). TRIM28 represses transcription of endogenous retroviruses in neural progenitor cells. Cell Rep *10*, 20-28.

Ferrara, N. (2009). Vascular endothelial growth factor. Arterioscler Thromb Vasc Biol *29*, 789-791.

Fontebasso, A. M., Papillon-Cavanagh, S., Schwartzentruber, J., Nikbakht, H., Gerges, N., Fiset, P. O., Bechet, D., Faury, D., De Jay, N., Ramkissoon, L. A.*, et al.* (2014). Recurrent somatic mutations in ACVR1 in pediatric midline high-grade astrocytoma. Nat Genet *46*, 462-466.

Frank, D., Doenecke, D., and Albig, W. (2003). Differential expression of human replacement and cell cycle dependent H3 histone genes. Gene *312*, 135-143.

Frey, A., Listovsky, T., Guilbaud, G., Sarkies, P., and Sale, J. E. (2014). Histone H3.3 is required to maintain replication fork progression after UV damage. Curr Biol *24*, 2195-2201.

Friedli, M., and Trono, D. (2015). The developmental control of transposable elements and the evolution of higher species. Annu Rev Cell Dev Biol *31*, 429-451.

Friedman, J. R., Fredericks, W. J., Jensen, D. E., Speicher, D. W., Huang, X. P., Neilson, E. G., and Rauscher, F. J., 3rd (1996). KAP-1, a novel corepressor for the highly conserved KRAB repression domain. Genes Dev *10*, 2067-2078.

Funato, K., Major, T., Lewis, P. W., Allis, C. D., and Tabar, V. (2014). Use of human embryonic stem cells to model pediatric gliomas with H3.3K27M histone mutation. Science *346*, 1529-1533.

Futschik, M. E., and Carlisle, B. (2005). Noise-robust soft clustering of gene expression time-course data. J Bioinform Comput Biol *3*, 965-988.

Goke, J., Lu, X., Chan, Y. S., Ng, H. H., Ly, L. H., Sachs, F., and Szczerbinska, I. (2015). Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. Cell Stem Cell *16*, 135-141.

Goldberg, A. D., Banaszynski, L. A., Noh, K. M., Lewis, P. W., Elsaesser, S. J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X.*, et al.* (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell *140*, 678-691.

Goodier, J. L., and Kazazian, H. H., Jr. (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell *135*, 23-35.

Grivennikov, S. I., Greten, F. R., and Karin, M. (2010). Immunity, inflammation, and cancer. Cell *140*, 883-899.

Guo, G., von Meyenn, F., Rostovskaya, M., Clarke, J., Dietmann, S., Baker, D., Sahakyan, A., Myers, S., Bertone, P., Reik, W.*, et al.* (2017). Epigenetic resetting of human pluripotency. Development *144*, 2748-2763.

Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., Yan, J., Ren, X., Lin, S., Li, J.*, et al.* (2014a). The DNA methylation landscape of human early embryos. Nature *511*, 606-610.

Guo, R., Zheng, L., Park, J. W., Lv, R., Chen, H., Jiao, F., Xu, W., Mu, S., Wen, H., Qiu, J.*, et al.* (2014b). BS69/ZMYND11 reads and connects histone H3.3 lysine 36 trimethylation-decorated chromatin to regulated pre-mRNA processing. Mol Cell *56*, 298-310.

Gurard-Levin, Z. A., Quivy, J. P., and Almouzni, G. (2014). Histone chaperones: assisting histone traffic and nucleosome dynamics. Annu Rev Biochem *83*, 487-517.

Hake, S. B., Garcia, B. A., Duncan, E. M., Kauer, M., Dellaire, G., Shabanowitz, J., Bazett-Jones, D. P., Allis, C. D., and Hunt, D. F. (2006). Expression patterns and post-translational modifications associated with mammalian histone H3 variants. J Biol Chem *281*, 559-568.

Hake, S. B., Garcia, B. A., Kauer, M., Baker, S. P., Shabanowitz, J., Hunt, D. F., and Allis, C. D. (2005). Serine 31 phosphorylation of histone variant H3.3 is specific to regions bordering centromeres in metaphase chromosomes. Proc Natl Acad Sci U S A *102*, 6344-6349.

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

Hancks, D. C., and Kazazian, H. H., Jr. (2012). Active human retrotransposons: variation and disease. Curr Opin Genet Dev *22*, 191-203.

Harutyunyan, A. S., Krug, B., Chen, H., Papillon-Cavanagh, S., Zeinieh, M., De Jay, N., Deshmukh, S., Chen, C. C. L., Belle, J., Mikael, L. G.*, et al.* (2019). H3K27M induces defective chromatin spread of PRC2-mediated repressive H3K27me2/me3 and is essential for glioma tumorigenesis. Nat Commun *10*, 1262.

He, J., Fu, X., Zhang, M., He, F., Li, W., Abdul, M. M., Zhou, J., Sun, L., Chang, C., Li, Y.*, et al.* (2019). Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. Nat Commun *10*, 34.

He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L.*, et al.* (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303-1307.

Hegi, M. E., Diserens, A. C., Gorlia, T., Hamou, M. F., de Tribolet, N., Weller, M., Kros, J. M., Hainfellner, J. A., Mason, W., Mariani, L.*, et al.* (2005). MGMT gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med *352*, 997-1003.

Hetey, S., Boros-Olah, B., Kuik-Rozsa, T., Li, Q., Karanyi, Z., Szabo, Z., Roszik, J., Szaloki, N., Vamosi, G., Toth, K., and Szekvolgyi, L. (2017). Biophysical characterization of histone H3.3 K27M point mutation. Biochem Biophys Res Commun *490*, 868-875.

Hochberg, Y. B. a. Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple

Testing. Journal of the Royal Statistical Society *Vol. 57*, pp. 289-300.

Hodl, M., and Basler, K. (2009). Transcription in the absence of histone H3.3. Curr Biol *19*, 1221-1226.

Hodl, M., and Basler, K. (2012). Transcription in the absence of histone H3.2 and H3K4 methylation. Curr Biol *22*, 2253-2257.

Hoffman, L. M., Veldhuijzen van Zanten, S. E. M., Colditz, N., Baugh, J., Chaney, B., Hoffmann, M., Lane, A., Fuller, C., Miles, L., Hawkins, C.*, et al.* (2018). Clinical, Radiologic, Pathologic, and Molecular Characteristics of Long-Term Survivors of Diffuse Intrinsic Pontine Glioma (DIPG): A Collaborative Report From the International and European Society for Pediatric Oncology DIPG Registries. J Clin Oncol *36*, 1963-1972.

Hu, Y., Yu, X., Xu, G., and Liu, S. (2017). Metastasis: an early event in cancer progression. J Cancer Res Clin Oncol *143*, 745-757.

Huang, H., Sabari, B. R., Garcia, B. A., Allis, C. D., and Zhao, Y. (2014). SnapShot: histone modifications. Cell *159*, 458-458 e451.

Huang, H. P., Chen, P. H., Yu, C. Y., Chuang, C. Y., Stone, L., Hsiao, W. C., Li, C. L., Tsai, S. C., Chen, K. Y., Chen, H. F.*, et al.* (2011). Epithelial cell adhesion molecule (EpCAM) complex proteins promote transcription factor-mediated pluripotency reprogramming. J Biol Chem *286*, 33520-33532.

Huntley, S., Baggott, D. M., Hamilton, A. T., Tran-Gyamfi, M., Yang, S., Kim, J., Gordon, L., Branscomb, E., and Stubbs, L. (2006). A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors. Genome Res *16*, 669-677.

Hutnick, L. K., Huang, X., Loo, T. C., Ma, Z., and Fan, G. (2010). Repression of retrotransposal elements in mouse embryonic stem cells is primarily mediated by a DNA methylation-independent mechanism. J Biol Chem *285*, 21082-21091.

Huttner, A. (2017). Molecular Neuropathology and the Ontogeny of Malignant Gliomas. 15-29.

Imbeault, M., Helleboid, P. Y., and Trono, D. (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature *543*, 550-554.

Inoue, A., and Zhang, Y. (2014). Nucleosome assembly is required for nuclear pore complex assembly in mouse zygotes. Nat Struct Mol Biol *21*, 609-616.

Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129-1133.

Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300-1303.

Ivanauskiene, K., Delbarre, E., McGhie, J. D., Kuntziger, T., Wong, L. H., and Collas, P. (2014). The PML-associated protein DEK regulates the balance of H3.3 loading on chromatin and is important for telomere integrity. Genome Res *24*, 1584-1594.

Jang, C. W., Shibata, Y., Starmer, J., Yee, D., and Magnuson, T. (2015). Histone H3.3 maintains genome integrity during mammalian development. Genes Dev *29*, 1377-1392.

Jang, H. S., Shah, N. M., Du, A. Y., Dailey, Z. Z., Pehrsson, E. C., Godoy, P. M., Zhang, D., Li, D., Xing, X., Kim, S.*, et al.* (2019). Transposable elements drive widespread expression of oncogenes in human cancers. Nat Genet *51*, 611-617.

Jang, H. S., Shin, W. J., Lee, J. E., and Do, J. T. (2017). CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. Genes (Basel) *8*.

Jang, S. M., Kauzlaric, A., Quivy, J. P., Pontis, J., Rauwel, B., Coluccio, A., Offner, S., Duc, J., Turelli, P., Almouzni, G., and Trono, D. (2018). KAP1 facilitates reinstatement of heterochromatin after DNA replication. Nucleic Acids Res *46*, 8788-8802.

Jha, D. K., and Strahl, B. D. (2014). An RNA polymerase II-coupled function for histone H3K36 methylation in checkpoint activation and DSB repair. Nat Commun *5*, 3965.

Jiang, B. H., and Liu, L. Z. (2009). PI3K/PTEN signaling in angiogenesis and tumorigenesis. Adv Cancer Res *102*, 19-65.

Jiao, L., and Liu, X. (2015). Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. Science *350*, aac4383.

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev *21*, 1519-1529.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K., and Felsenfeld, G. (2009). H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. Nat Genet *41*, 941-945.

Jones, C., Perryman, L., and Hargrave, D. (2012). Paediatric and adult malignant glioma: close relatives or distant cousins? Nat Rev Clin Oncol *9*, 400-413.

Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet *13*, 484-492.

Jones, R. G., and Thompson, C. B. (2009). Tumor suppressors and cell metabolism: a recipe for cancer growth. Genes Dev *23*, 537-548.

Jullien, J., Astrand, C., Szenker, E., Garrett, N., Almouzni, G., and Gurdon, J. B. (2012). HIRA dependent H3.3 deposition is required for transcriptional reprogramming following nuclear transfer to Xenopus oocytes. Epigenetics Chromatin *5*, 17.

Juratli, T. A., Qin, N., Cahill, D. P., and Filbin, M. G. (2018). Molecular pathogenesis and therapeutic implications in pediatric high-grade gliomas. Pharmacol Ther *182*, 70-79.

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res *110*, 462-467.

Justin, N., Zhang, Y., Tarricone, C., Martin, S. R., Chen, S., Underwood, E., De Marco, V., Haire, L. F., Walker, P. A., Reinberg, D.*, et al.* (2016). Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. Nat Commun *7*, 11316.

Kaatsch, P. (2010). Epidemiology of childhood cancer. Cancer Treat Rev *36*, 277-285.

Kallappagoudar, S., Yadav, R. K., Lowe, B. R., and Partridge, J. F. (2015). Histone H3 mutations--a special role for H3.3 in tumorigenesis? Chromosoma *124*, 177-189.

Karimi, M. M., Goyal, P., Maksakova, I. A., Bilenky, M., Leung, D., Tang, J. X., Shinkai, Y., Mager, D. L., Jones, S., Hirst, M., and Lorincz, M. C. (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. Cell Stem Cell *8*, 676-687.

Kato, M., Takemoto, K., and Shinkai, Y. (2018). A somatic role for the histone methyltransferase Setdb1 in endogenous retrovirus silencing. Nat Commun *9*, 1683.

Khorasanizadeh, S. (2004). The nucleosome: from genomic organization to genomic regulation. Cell *116*, 259-272.

Kinzler, K. W., and Vogelstein, B. (1997). Cancer-susceptibility genes. Gatekeepers and caretakers. Nature *386*, 761, 763.

Kornberg, R. D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868-871.

Korshunov, A., Ryzhova, M., Hovestadt, V., Bender, S., Sturm, D., Capper, D., Meyer, J., Schrimpf, D., Kool, M., Northcott, P. A.*, et al.* (2015). Integrated analysis of pediatric glioblastoma reveals a subset of biologically favorable tumors with associated molecular prognostic markers. Acta Neuropathol *129*, 669-678.

Krimer, D. B., Cheng, G., and Skoultchi, A. I. (1993). Induction of H3.3 replacement histone mRNAs during the precommitment period of murine erythroleukemia cell differentiation. Nucleic Acids Res *21*, 2873-2879.

Krug, B., De Jay, N., Harutyunyan, A. S., Deshmukh, S., Marchione, D. M., Guilhamon, P., Bertrand, K. C., Mikael, L. G., McConechy, M. K., Chen, C. C. L.*, et al.* (2019). Pervasive H3K27 Acetylation Leads to ERV Expression and a Therapeutic Vulnerability in H3K27M Gliomas. Cancer Cell *35*, 782-797 e788.

Kurihara, Y., Kawamura, Y., Uchijima, Y., Amamo, T., Kobayashi, H., Asano, T., and Kurihara, H. (2008). Maintenance of genomic methylation patterns during preimplantation development requires the somatic form of DNA methyltransferase 1. Dev Biol *313*, 335-346.

Lapin, D. H., Tsoli, M., and Ziegler, D. S. (2017). Genomic Insights into Diffuse Intrinsic Pontine Glioma. Front Oncol *7*, 57.

Larson, J. D., Kasper, L. H., Paugh, B. S., Jin, H., Wu, G., Kwon, C. H., Fan, Y., Shaw, T. I., Silveira, A. B., Qu, C.*, et al.* (2019). Histone H3.3 K27M Accelerates Spontaneous Brainstem Glioma and Drives Restricted Changes in Bivalent Gene Expression. Cancer Cell *35*, 140-155 e147.

Lee, C. H., Yu, J. R., Granat, J., Saldana-Meyer, R., Andrade, J., LeRoy, G., Jin, Y., Lund, P., Stafford, J. M., Garcia, B. A.*, et al.* (2019). Automethylation of PRC2 promotes H3K27 methylation and is impaired in H3K27M pediatric glioma. Genes Dev.

Leeb, M., Pasini, D., Novatchkova, M., Jaritz, M., Helin, K., and Wutz, A. (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. Genes Dev *24*, 265-276.

Lehnertz, B., Zhang, Y. W., Boivin, I., Mayotte, N., Tomellini, E., Chagraoui, J., Lavallee, V. P., Hebert, J., and Sauvageau, G. (2017). H3(K27M/I) mutations promote context-dependent transformation in acute myeloid leukemia with RUNX1 alterations. Blood *130*, 2204-2214.

Lewis, P. W., and Allis, C. D. (2013). Poisoning the "histone code" in pediatric gliomagenesis. Cell Cycle *12*, 3241-3242.

Lewis, P. W., Elsaesser, S. J., Noh, K. M., Stadler, S. C., and Allis, C. D. (2010). Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. Proc Natl Acad Sci U S A *107*, 14075-14080.

Lewis, P. W., Muller, M. M., Koletsky, M. S., Cordero, F., Lin, S., Banaszynski, L. A., Garcia, B. A., Muir, T. W., Becher, O. J., and Allis, C. D. (2013). Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science *340*, 857-861.

Li, F., Mao, G., Tong, D., Huang, J., Gu, L., Yang, W., and Li, G. M. (2013). The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSalpha. Cell *153*, 590-600.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Liao, J., Karnik, R., Gu, H., Ziller, M. J., Clement, K., Tsankov, A. M., Akopian, V., Gifford, C. A., Donaghey, J., Galonska, C.*, et al.* (2015). Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. Nat Genet *47*, 469-478.

Lin, C. J., Conti, M., and Ramalho-Santos, M. (2013). Histone variant H3.3 maintains a decondensed chromatin state essential for mouse preimplantation development. Development *140*, 3624-3634.

Lin, C. J., Koh, F. M., Wong, P., Conti, M., and Ramalho-Santos, M. (2014). Hira-mediated H3.3 incorporation is required for DNA replication and ribosomal RNA transcription in the mouse zygote. Dev Cell *30*, 268-279.

Loppin, B., Bonnefoy, E., Anselme, C., Laurencon, A., Karr, T. L., and Couble, P. (2005). The histone H3.3 chaperone HIRA is essential for chromatin assembly in the male pronucleus. Nature *437*, 1386-1390.

Louis, D. N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W. K., Ohgaki, H., Wiestler, O. D., Kleihues, P., and Ellison, D. W. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol *131*, 803-820.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Lowe, B. R., Maxham, L. A., Hamey, J. J., Wilkins, M. R., and Partridge, J. F. (2019). Histone H3 Mutations: An Updated View of Their Role in Chromatin Deregulation and Cancer. Cancers (Basel) *11*.

Loyola, A., and Almouzni, G. (2007). Marking histone H3 variants: how, when and why? Trends Biochem Sci *32*, 425-433.

Lu, C., Jain, S. U., Hoelper, D., Bechet, D., Molden, R. C., Ran, L., Murphy, D., Venneti, S., Hameed, M., Pawel, B. R.*, et al.* (2016). Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. Science *352*, 844-849.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S. L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature *487*, 57-63.

Mackay, A., Burford, A., Carvalho, D., Izquierdo, E., Fazal-Salom, J., Taylor, K. R., Bjerke, L., Clarke, M., Vinci, M., Nandhabalan, M.*, et al.* (2017). Integrated Molecular Meta-Analysis of 1,000 Pediatric High-Grade and Diffuse Intrinsic Pontine Glioma. Cancer Cell *32*, 520-537 e525.

Maiti, A., and Drohat, A. C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J Biol Chem *286*, 35334-35338.

Maksakova, I. A., Goyal, P., Bullwinkel, J., Brown, J. P., Bilenky, M., Mager, D. L., Singh, P. B., and Lorincz, M. C. (2011). H3K9me3-binding proteins are dispensable for SETDB1/H3K9me3-dependent retroviral silencing. Epigenetics Chromatin *4*, 12.

Margueron, R., Justin, N., Ohno, K., Sharpe, M. L., Son, J., Drury, W. J., 3rd, Voigt, P., Martin, S. R., Taylor, W. R., De Marco, V.*, et al.* (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762-767.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343-349.

Matsui, T., Leung, D., Miyashita, H., Maksakova, I. A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M. C., and Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. Nature *464*, 927-931.

Maze, I., Wenderski, W., Noh, K. M., Bagot, R. C., Tzavaras, N., Purushothaman, I., Elsasser, S. J., Guo, Y., Ionete, C., Hurd, Y. L.*, et al.* (2015). Critical Role of Histone Turnover in Neuronal Transcription and Plasticity. Neuron *87*, 77-94.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A *36*, 344-355.

McKittrick, E., Gafken, P. R., Ahmad, K., and Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. Proc Natl Acad Sci U S A *101*, 1525-1530.

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T. K., Koche, R. P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Mohammad, F., Weissmann, S., Leblanc, B., Pandey, D. P., Hojfeldt, J. W., Comet, I., Zheng, C., Johansen, J. V., Rapin, N., Porse, B. T.*, et al.* (2017). EZH2 is a potential therapeutic target for H3K27M-mutant pediatric gliomas. Nat Med *23*, 483-492.

Morozov, V. M., Gavrilova, E. V., Ogryzko, V. V., and Ishov, A. M. (2012). Dualistic function of Daxx at centromeric and pericentromeric heterochromatin in normal and stress conditions. Nucleus *3*, 276-285.

Muotri, A. R., Chu, V. T., Marchetto, M. C., Deng, W., Moran, J. V., and Gage, F. H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature *435*, 903-910.

Nacev, B. A., Feng, L., Bagert, J. D., Lemiesz, A. E., Gao, J., Soshnev, A. A., Kundra, R., Schultz, N., Muir, T. W., and Allis, C. D. (2019). The expanding landscape of 'oncohistone' mutations in human cancers. Nature *567*, 473-478.

Nagaraja, S., Quezada, M. A., Gillespie, S. M., Arzt, M., Lennon, J. J., Woo, P. J., Hovestadt, V., Kambhampati, M., Filbin, M. G., Suva, M. L.*, et al.* (2019). Histone Variant and Cell Context Determine H3K27M Reprogramming of the Enhancer Landscape and Oncogenic State. Mol Cell.

Nazarian, J., Mason, G. E., Ho, C. Y., Panditharatna, E., Kambhampati, M., Vezina, L. G., Packer, R. J., and Hwang, E. I. (2016). Histological and molecular analysis of a progressive diffuse intrinsic pontine glioma and synchronous metastatic lesions: a case report. Oncotarget *7*, 42837-42842.

Nikbakht, H., Panditharatna, E., Mikael, L. G., Li, R., Gayden, T., Osmond, M., Ho, C. Y., Kambhampati, M., Hwang, E. I., Faury, D.*, et al.* (2016). Spatial and temporal homogeneity of driver mutations in diffuse intrinsic pontine glioma. Nat Commun *7*, 11185.

Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Olins, D. E., and Olins, A. L. (2003). Chromatin history: our view from the bridge. Nat Rev Mol Cell Biol *4*, 809-814.

Orgel, L. E., and Crick, F. H. (1980). Selfish DNA: the ultimate parasite. Nature *284*, 604-607.

Ors, A., Papin, C., Favier, B., Roulland, Y., Dalkara, D., Ozturk, M., Hamiche, A., Dimitrov, S., and Padmanabhan, K. (2017). Histone H3.3 regulates mitotic progression in mouse embryonic fibroblasts. Biochem Cell Biol *95*, 491-499.

Ostrom, Q. T., Gittleman, H., Truitt, G., Boscia, A., Kruchko, C., and Barnholtz-Sloan, J. S. (2018). CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2011-2015. Neuro Oncol *20*, iv1-iv86.

Ou, H. D., Phan, S., Deerinck, T. J., Thor, A., Ellisman, M. H., and O'Shea, C. C. (2017). ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science *357*.

Oudet, P., Gross-Bellard, M., and Chambon, P. (1975). Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. Cell *4*, 281-300.

Pai, C. C., Deegan, R. S., Subramanian, L., Gal, C., Sarkar, S., Blaikley, E. J., Walker, C., Hulme, L., Bernhard, E., Codlin, S.*, et al.* (2014). A histone H3K36 chromatin switch coordinates DNA double-strand break repair pathway choice. Nat Commun *5*, 4091.

Papillon-Cavanagh, S., Lu, C., Gayden, T., Mikael, L. G., Bechet, D., Karamboulas, C., Ailles, L., Karamchandani, J., Marchione, D. M., Garcia, B. A.*, et al.* (2017). Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas. Nat Genet *49*, 180-185.

Papin, C., Ibrahim, A., Gras, S. L., Velt, A., Stoll, I., Jost, B., Menoni, H., Bronner, C., Dimitrov, S., and Hamiche, A. (2017). Combinatorial DNA methylation codes at repetitive elements. Genome Res *27*, 934-946.

Pathania, M., De Jay, N., Maestro, N., Harutyunyan, A. S., Nitarska, J., Pahlavan, P., Henderson, S., Mikael, L. G., Richard-Londt, A., Zhang, Y.*, et al.* (2017). H3.3(K27M) Cooperates with Trp53 Loss and PDGFRA Gain in Mouse Embryonic Neural Progenitor Cells to Induce Invasive High-Grade Gliomas. Cancer Cell *32*, 684-700 e689.

Paugh, B. S., Qu, C., Jones, C., Liu, Z., Adamowicz-Brice, M., Zhang, J., Bax, D. A., Coyle, B., Barrow, J., Hargrave, D.*, et al.* (2010). Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease. J Clin Oncol *28*, 3061-3068.

Pfister, S. X., Ahrabi, S., Zalmas, L. P., Sarkar, S., Aymard, F., Bachrati, C. Z., Helleday, T., Legube, G., La Thangue, N. B., Porter, A. C., and Humphrey, T. C. (2014). SETD2-dependent histone H3K36 trimethylation is required for homologous recombination repair and genome stability. Cell Rep *7*, 2006-2018.

Piazzesi, A., Papic, D., Bertan, F., Salomoni, P., Nicotera, P., and Bano, D. (2016). Replication-Independent Histone Variant H3.3 Controls Animal Lifespan through the Regulation of Pro-longevity Transcriptional Programs. Cell Rep *17*, 987-996.

Pina, B., and Suau, P. (1987). Changes in histones H2A and H3 variant composition in differentiating and mature rat brain cortical neurons. Dev Biol *123*, 51-58.

Piunti, A., Hashizume, R., Morgan, M. A., Bartom, E. T., Horbinski, C. M., Marshall, S. A., Rendleman, E. J., Ma, Q., Takahashi, Y. H., Woodfin, A. R.*, et al.* (2017). Therapeutic targeting of polycomb and BET bromodomain proteins in diffuse intrinsic pontine gliomas. Nat Med *23*, 493-500.

Pollack, I. F., Hamilton, R. L., Burger, P. C., Brat, D. J., Rosenblum, M. K., Murdoch, G. H., Nikiforova, M. N., Holmes, E. J., Zhou, T., Cohen, K. J.*, et al.* (2010). Akt activation is a common event in pediatric malignant gliomas and a potential adverse prognostic marker: a report from the Children's Oncology Group. J Neurooncol *99*, 155-163.

Pontis, J., Planet, E., Offner, S., Turelli, P., Duc, J., Coudray, A., Theunissen, T. W., Jaenisch, R., and Trono, D. (2019). Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. Cell Stem Cell *24*, 724-735 e725.

Qian, B. Z., and Pollard, J. W. (2010). Macrophage diversity enhances tumor progression and metastasis. Cell *141*, 39-51.

Quenneville, S., Turelli, P., Bojkowska, K., Raclot, C., Offner, S., Kapopoulou, A., and Trono, D. (2012). The KRAB-ZFP/KAP1 system contributes to the early embryonic establishment of site-specific DNA methylation patterns maintained during development. Cell Rep *2*, 766-773.

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P. V., Grimaldi, G., Riccio, A., and Trono, D. (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. Mol Cell *44*, 361-372.

Raica, M., Cimpean, A. M., and Ribatti, D. (2009). Angiogenesis in pre-malignant conditions. Eur J Cancer *45*, 1924-1934.

Ray-Gallet, D., Quivy, J. P., Scamps, C., Martini, E. M., Lipinski, M., and Almouzni, G. (2002). HIRA is critical for a nucleosome assembly pathway independent of DNA synthesis. Mol Cell *9*, 1091-1100.

Ray-Gallet, D., Ricketts, M. D., Sato, Y., Gupta, K., Boyarchuk, E., Senda, T., Marmorstein, R., and Almouzni, G. (2018). Functional activity of the H3.3 histone chaperone complex HIRA requires trimerization of the HIRA subunit. Nat Commun *9*, 3103.

Ren, M., and van Nocker, S. (2016). In silico analysis of histone H3 gene expression during human brain development. Int J Dev Biol *60*, 167-173.

Ricketts, M. D., and Marmorstein, R. (2017). A Molecular Prospective for HIRA Complex Assembly and H3.3-Specific Histone Chaperone Function. J Mol Biol *429*, 1924-1933.

Rothbart, S. B., and Strahl, B. D. (2014). Interpreting the language of histone and DNA modifications. Biochim Biophys Acta *1839*, 627-643.

Rowe, H. M., Friedli, M., Offner, S., Verp, S., Mesnard, D., Marquis, J., Aktas, T., and Trono, D. (2013a). De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. Development *140*, 519-529.

Rowe, H. M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P. V., Layard-Liesching, H., Verp, S., Marquis, J.*, et al.* (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. Nature *463*, 237-240.

Rowe, H. M., Kapopoulou, A., Corsinotti, A., Fasching, L., Macfarlan, T. S., Tarabay, Y., Viville, S., Jakobsson, J., Pfaff, S. L., and Trono, D. (2013b). TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. Genome Res *23*, 452-461.

Rowe, H. M., and Trono, D. (2011). Dynamic control of endogenous retroviruses during development. Virology *411*, 273-287.

Sakai, A., Schwartz, B. E., Goldstein, S., and Ahmad, K. (2009). Transcriptional and developmental functions of the H3.3 histone variant in Drosophila. Curr Biol *19*, 1816-1820.

Saksouk, N., Simboeck, E., and Dejardin, J. (2015). Constitutive heterochromatin formation and transcription in mammals. Epigenetics Chromatin *8*, 3.

Saletta, F., Seng, M. S., and Lau, L. M. (2014). Advances in paediatric cancer treatment. Transl Pediatr *3*, 156-182.

Schwartz, B. E., and Ahmad, K. (2005). Transcriptional activation triggers deposition and removal of the histone variant H3.3. Genes Dev *19*, 804-814.

Schwartzentruber, J., Korshunov, A., Liu, X. Y., Jones, D. T., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A. M., Quang, D. A., Tonjes, M.*, et al.* (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature *482*, 226-231.

Sharif, J., Endo, T. A., Nakayama, M., Karimi, M. M., Shimada, M., Katsuyama, K., Goyal, P., Brind'Amour, J., Sun, M. A., Sun, Z.*, et al.* (2016). Activation of Endogenous Retroviruses in Dnmt1(-/-) ESCs Involves Disruption of SETDB1-Mediated Repression by NP95 Binding to Hemimethylated DNA. Cell Stem Cell *19*, 81-94.

Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A. C., Fung, H. L., Zhang, K., and Zhang, Y. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. Cell *153*, 692-706.

Shi, L., Shi, J., Shi, X., Li, W., and Wen, H. (2018). Histone H3.3 G34 Mutations Alter Histone H3K36 and H3K27 Methylation In Cis. J Mol Biol *430*, 1562-1565.

Silveira, A. B., Kasper, L. H., Fan, Y., Jin, H., Wu, G., Shaw, T. I., Zhu, X., Larson, J. D., Easton, J., Shao, Y.*, et al.* (2019). H3.3 K27M depletion increases differentiation and extends latency of diffuse intrinsic pontine glioma growth in vivo. Acta Neuropathol *137*, 637-655.

Sitbon, D., Podsypanina, K., Yadav, T., and Almouzni, G. (2017). Shaping Chromatin in the Nucleus: The Bricks and the Architects. Cold Spring Harb Symp Quant Biol *82*, 1-14.

Smith, Z. D., Chan, M. M., Humm, K. C., Karnik, R., Mekhoubad, S., Regev, A., Eggan, K., and Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. Nature *511*, 611-615.

Stafford, J. M., Lee, C. H., Voigt, P., Descostes, N., Saldana-Meyer, R., Yu, J. R., Leroy, G., Oksuz, O., Chapman, J. R., Suarez, F.*, et al.* (2018). Multiple modes of PRC2 inhibition elicit global chromatin alterations in H3K27M pediatric glioma. Sci Adv *4*, eaau5935.

Steliarova-Foucher, E., Colombet, M., Ries, L. A. G., Moreno, F., Dolya, A., Bray, F., Hesseling, P., Shin, H. Y., Stiller, C. A., and contributors, I.-. (2017). International incidence of childhood cancer, 2001-10: a population-based registry study. Lancet Oncol *18*, 719-731.

Stephens, A. S., Stephens, S. R., Hobbs, C., Hutmacher, D. W., Bacic-Welsh, D., Woodruff, M. A., and Morrison, N. A. (2011). Myocyte enhancer factor 2c, an osteoblast transcription factor identified by dimethyl sulfoxide (DMSO)-enhanced mineralization. J Biol Chem *286*, 30071-30086.

Stocking, C., and Kozak, C. A. (2008). Murine endogenous retroviruses. Cell Mol Life Sci *65*, 3383-3398.

Sturm, D., Witt, H., Hovestadt, V., Khuong-Quang, D. A., Jones, D. T., Konermann, C., Pfaff, E., Tonjes, M., Sill, M., Bender, S.*, et al.* (2012). Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. Cancer Cell *22*, 425-437.

Suzuki, M. M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet *9*, 465-476.

Szenker, E., Lacoste, N., and Almouzni, G. (2012). A developmental requirement for HIRA-dependent H3.3 deposition revealed at gastrulation in Xenopus. Cell Rep *1*, 730-740.

Szenker, E., Ray-Gallet, D., and Almouzni, G. (2011). The double face of the histone variant H3.3. Cell Res *21*, 421-434.

Tachiwana, H., Osakabe, A., Shiga, T., Miya, Y., Kimura, H., Kagawa, W., and Kurumizaka, H. (2011). Structures of human nucleosomes containing major histone H3 variants. Acta Crystallogr D Biol Crystallogr *67*, 578-583.

Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell *116*, 51-61.

Talbert, P. B., Ahmad, K., Almouzni, G., Ausio, J., Berger, F., Bhalla, P. L., Bonner, W. M., Cande, W. Z., Chadwick, B. P., Chan, S. W.*, et al.* (2012). A unified phylogeny-based nomenclature for histone variants. Epigenetics Chromatin *5*, 7.

Tang, M. C., Jacobs, S. A., Mattiske, D. M., Soh, Y. M., Graham, A. N., Tran, A., Lim, S. L., Hudson, D. F., Kalitsis, P., O'Bryan, M. K.*, et al.* (2015). Contribution of the two genes encoding histone variant h3.3 to viability and fertility in mice. PLoS Genet *11*, e1004964.

Tang, M. C., Jacobs, S. A., Wong, L. H., and Mann, J. R. (2013). Conditional allelic replacement applied to genes encoding the histone variant H3.3 in the mouse. Genesis *51*, 142-146.

Tatavosian, R., Duc, H. N., Huynh, T. N., Fang, D., Schmitt, B., Shi, X., Deng, Y., Phiel, C., Yao, T., Zhang, Z.*, et al.* (2018). Live-cell single-molecule dynamics of PcG proteins imposed by the DIPG H3.3K27M mutation. Nat Commun *9*, 2080.

Taylor, K. R., Mackay, A., Truffaux, N., Butterfield, Y., Morozova, O., Philippe, C., Castel, D., Grasso, C. S., Vinci, M., Carvalho, D.*, et al.* (2014). Recurrent activating ACVR1 mutations in diffuse intrinsic pontine glioma. Nat Genet *46*, 457-461.

Theunissen, T. W., Friedli, M., He, Y., Planet, E., O'Neil, R. C., Markoulaki, S., Pontis, J., Wang, H., Iouranova, A., Imbeault, M.*, et al.* (2016). Molecular Criteria for Defining the Naive Human Pluripotent State. Cell Stem Cell *19*, 502-515.

Torres-Padilla, M. E., Bannister, A. J., Hurd, P. J., Kouzarides, T., and Zernicka-Goetz, M. (2006). Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. Int J Dev Biol *50*, 455-461.

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H. R.*, et al.* (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells *11*, 805-814.

Turelli, P., Castro-Diaz, N., Marzetta, F., Kapopoulou, A., Raclot, C., Duc, J., Tieng, V., Quenneville, S., and Trono, D. (2014). Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. Genome Res *24*, 1260-1270.

Udugama, M., FT, M. C., Chan, F. L., Tang, M. C., Pickett, H. A., JD, R. M., Mayne, L., Collas, P., Mann, J. R., and Wong, L. H. (2015). Histone variant H3.3 provides the heterochromatic H3 lysine 9 tri-methylation mark at telomeres. Nucleic Acids Res *43*, 10227-10237.

Vakoc, C. R., Mandat, S. A., Olenchock, B. A., and Blobel, G. A. (2005). Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Mol Cell *19*, 381-391.

van der Heijden, G. W., Derijck, A. A., Posfai, E., Giele, M., Pelczar, P., Ramos, L., Wansink, D. G., van der Vlag, J., Peters, A. H., and de Boer, P. (2007). Chromosome-wide nucleosome replacement and H3.3 incorporation during mammalian meiotic sex chromosome inactivation. Nat Genet *39*, 251-258.

Venneti, S., Garimella, M. T., Sullivan, L. M., Martinez, D., Huse, J. T., Heguy, A., Santi, M., Thompson, C. B., and Judkins, A. R. (2013). Evaluation of histone 3 lysine 27 trimethylation (H3K27me3) and enhancer of Zest 2 (EZH2) in pediatric glial and glioneuronal tumors shows decreased H3K27me3 in H3F3A K27M mutant glioblastomas. Brain Pathol *23*, 558-564.

Vinci, M., Burford, A., Molinari, V., Kessler, K., Popov, S., Clarke, M., Taylor, K. R., Pemberton, H. N., Lord, C. J., Gutteridge, A.*, et al.* (2018). Functional diversity and cooperativity between subclonal populations of pediatric glioblastoma and diffuse intrinsic pontine glioma cells. Nat Med *24*, 1204-1215.

Voon, H. P. J., Udugama, M., Lin, W., Hii, L., Law, R. H. P., Steer, D. L., Das, P. P., Mann, J. R., and Wong, L. H. (2018). Inhibition of a K9/K36 demethylase by an H3.3 point mutation found in paediatric glioblastoma. Nat Commun *9*, 3142.

Walsh, C. P., Chaillet, J. R., and Bestor, T. H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet *20*, 116-117.

Wang, L. H., Wu, C. F., Rajasekaran, N., and Shin, Y. K. (2018). Loss of Tumor Suppressor Gene Function in Human Cancer: An Overview. Cell Physiol Biochem *51*, 2647-2693.

Wang, X., Long, Y., Paucek, R. D., Gooding, A. R., Lee, T., Burdorf, R. M., and Cech, T. R. (2019). Regulation of histone methylation by automethylation of PRC2. Genes Dev.

Wang, X., Paucek, R. D., Gooding, A. R., Brown, Z. Z., Ge, E. J., Muir, T. W., and Cech, T. R. (2017). Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA. Nat Struct Mol Biol *24*, 1028-1038.

Watanabe, M., Fung, E. S., Chan, F. B., Wong, J. S., Coutts, M., and Monuki, E. S. (2016). BMP4 acts as a dorsal telencephalic morphogen in a mouse embryonic stem cell culture system. Biol Open *5*, 1834-1843.

Watson, J. D., and Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*, 737-738.

Wen, D., Banaszynski, L. A., Liu, Y., Geng, F., Noh, K. M., Xiang, J., Elemento, O., Rosenwaks, Z., Allis, C. D., and Rafii, S. (2014a). Histone variant H3.3 is an essential maternal factor for oocyte reprogramming. Proc Natl Acad Sci U S A *111*, 7325-7330.

Wen, D., Banaszynski, L. A., Rosenwaks, Z., Allis, C. D., and Rafii, S. (2014b). H3.3 replacement facilitates epigenetic reprogramming of donor nuclei in somatic cell nuclear transfer embryos. Nucleus *5*, 369-375.

Wen, H., Li, Y., Xi, Y., Jiang, S., Stratton, S., Peng, D., Tanaka, K., Ren, Y., Xia, Z., Wu, J.*, et al.* (2014c). ZMYND11 links histone H3.3K36me3 to transcription elongation and tumour suppression. Nature *508*, 263-268.

White, E., and DiPaola, R. S. (2009). The double-edged sword of autophagy modulation in cancer. Clin Cancer Res *15*, 5308-5316.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O.*, et al.* (2007). A unified classification system for eukaryotic transposable elements. Nat Rev Genet *8*, 973-982.

Wolf, D., and Goff, S. P. (2009). Embryonic stem cells use ZFP809 to silence retroviral DNAs. Nature *458*, 1201-1204.

Wolf, G., Rebollo, R., Karimi, M. M., Ewing, A. D., Kamada, R., Wu, W., Wu, B., Bachu, M., Ozato, K., Faulkner, G. J.*, et al.* (2017). On the role of H3.3 in retroviral silencing. Nature *548*, E1-E3.

Wolf, G., Yang, P., Fuchtbauer, A. C., Fuchtbauer, E. M., Silva, A. M., Park, C., Wu, W., Nielsen, A. L., Pedersen, F. S., and Macfarlan, T. S. (2015). The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. Genes Dev *29*, 538-554.

Wong, L. H., McGhie, J. D., Sim, M., Anderson, M. A., Ahn, S., Hannan, R. D., George, A. J., Morgan, K. A., Mann, J. R., and Choo, K. H. (2010). ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. Genome Res *20*, 351-360.

Wong, L. H., Ren, H., Williams, E., McGhie, J., Ahn, S., Sim, M., Tam, A., Earle, E., Anderson, M. A., Mann, J., and Choo, K. H. (2009). Histone H3.3 incorporation provides a unique and functionally essential telomeric chromatin in embryonic stem cells. Genome Res *19*, 404-414.

Wu, G., Broniscer, A., McEachron, T. A., Lu, C., Paugh, B. S., Becksfort, J., Qu, C., Ding, L., Huether, R., Parker, M.*, et al.* (2012). Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. Nat Genet *44*, 251-253.

Wu, G., Diaz, A. K., Paugh, B. S., Rankin, S. L., Ju, B., Li, Y., Zhu, X., Qu, C., Chen, X., Zhang, J.*, et al.* (2014). The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma. Nat Genet *46*, 444-450.

Wu, H., and Zhang, Y. (2014). Reversing DNA methylation: mechanisms, genomics, and biological functions. Cell *156*, 45-68.

Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. Nat Rev Genet *18*, 517-534.

Xia, W., and Jiao, J. (2017). Histone variant H3.3 orchestrates neural stem cell differentiation in the developing brain. Cell Death Differ *24*, 1548-1563.

Xiong, C., Wen, Z., Yu, J., Chen, J., Liu, C. P., Zhang, X., Chen, P., Xu, R. M., and Li, G. (2018). UBN1/2 of HIRA complex is responsible for recognition and deposition of H3.3 at cis-regulatory elements of genes in mouse ES cells. BMC Biol *16*, 110.

Xu, C., Bian, C., Yang, W., Galka, M., Ouyang, H., Chen, C., Qiu, W., Liu, H., Jones, A. E., MacKenzie, F.*, et al.* (2010). Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). Proc Natl Acad Sci U S A *107*, 19266-19271.

Yadav, R. K., Jablonowski, C. M., Fernandez, A. G., Lowe, B. R., Henry, R. A., Finkelstein, D., Barnum, K. J., Pidoux, A. L., Kuo, Y. M., Huang, J.*, et al.* (2017). Histone H3G34R mutation causes replication stress, homologous recombination defects and genomic instability in S. pombe. Elife *6*.

Yang, L., Pang, Y., and Moses, H. L. (2010). TGF-beta and immune cells: an important regulatory axis in the tumor microenvironment and progression. Trends Immunol *31*, 220-227.

Yang, P., Wang, Y., and Macfarlan, T. S. (2017). The Role of KRAB-ZFPs in Transposable Element Repression and Mammalian Evolution. Trends Genet *33*, 871-881.

Ye, T., Krebs, A. R., Choukrallah, M. A., Keime, C., Plewniak, F., Davidson, I., and Tora, L. (2011). seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res *39*, e35.

Yuan, T. L., and Cantley, L. C. (2008). PI3K pathway alterations in cancer: variations on a theme. Oncogene *27*, 5497-5510.

Yuen, B. T., Bush, K. M., Barrilleaux, B. L., Cotterman, R., and Knoepfler, P. S. (2014). Histone H3.3 regulates dynamic chromatin states during spermatogenesis. Development *141*, 3483-3494.

Zhang, H., Gan, H., Wang, Z., Lee, J. H., Zhou, H., Ordog, T., Wold, M. S., Ljungman, M., and Zhang, Z. (2017). RPA Interacts with HIRA and Regulates H3.3 Deposition at Gene Regulatory Elements in Mammalian Cells. Mol Cell *65*, 272-284.

Zhao, Y., and Garcia, B. A. (2015). Comprehensive Catalog of Currently Documented Histone Modifications. Cold Spring Harb Perspect Biol *7*, a025064.

Zhou, L., Baibakov, B., Canagarajah, B., Xiong, B., and Dean, J. (2017). Genetic mosaics and time-lapse imaging identify functions of histone H3.3 residues in mouse oocytes and embryos. Development *144*, 519-528.

Université de Strasbourg

Alexia VIDAL

inoviem scientific

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

Rôle des mutations de l'histone variante H3.3 dans le développement des gliomes pédiatriques de haut grade

Résumé

Les tumeurs cérébrales constituent le deuxième cancer pédiatrique le plus fréquent après les leucémies et présentent un mauvais pronostic dû à l'absence de traitement efficace. Les mutations de l'histone variante H3.3 K27M ou G34R ont été décrites comme moteur du développement des gliomes pédiatriques de haut-grade mais le mécanisme sous-jacent reste à établir. Le but de ma thèse a été de comprendre l'impact des mutations de H3.3 sur sa distribution chromatinienne et sur l'expression du génome, grâce au développement d'un nouveau modèle de cellule souche embryonnaire contenant une forme étiquetée de H3.3 sauvage ou mutée. Mon travail montre que H3.3 marque les rétrovirus endogènes (ERVs) récemment intégrés, et que ces rétroéléments sont spécifiquement impliqués dans la différentiation neurale. Les mutations de H3.3 provoquent une dérégulation globale des rétrotransposons, qui aboutit à un défaut de différentiation. Le lien direct entre H3.3 et les ERVs a été confirmé cliniquement, ce qui suggère que ce mécanisme est impliqué dans la pathogénicité des gliomes pédiatriques de haut-grade.

Mots clés : H3.3, mutations K27M/G34R, gliome pédiatrique de haut-grade, rétrovirus endogènes, rétrotransposons, différentiation

Summary

Brain tumors are the second most frequent pediatric cancer after leukemia and the absence of efficient treatment leads to a poor prognosis with a survival rate of less than two years. Histone variant H3.3 mutations have been described as drivers of pediatric high-grade glioma, but the underlying mechanisms remain unknown. The main goal of this thesis is to dissect the mechanistic aspects of H3.3 mutation functions and the molecular mechanisms through which these mutations contribute to oncogenesis. Using a new constitutively tagged H3.3 (wildtype, K27M or G34R) mouse embryonic stem cell model, we showed that recently integrated endogenous retroviruses (ERV) are enriched in H3.3. ERV become overexpressed under mutant H3.3 and lead to deregulation of neighboring genes. The direct link between H3.3 and ERV regulation could be clinically confirmed on patient samples. In addition, a novel role for H3.3 in neural differentiation has been highlighted through the regulation of recently integrated retrotransposons, with H3.3 mutants leading to their deregulation and failure of differentiation. Deregulation of recently integrated retroelements under H3.3 K27M and G34R represents a novel mechanism potentially implicated in pediatric high-grade glioma development and progression.

Key-words: H3.3, K27M/G34R mutations, pediatric high-grade glioma, endogenous retroviruses, retrotransposons, differentiation