



HAL
open science

Développement d'une technique d'identification des protéines associées à un locus spécifique

Sandrine Sauzet

► **To cite this version:**

Sandrine Sauzet. Développement d'une technique d'identification des protéines associées à un locus spécifique. Génomique, Transcriptomique et Protéomique [q-bio.GN]. 2020. hal-03159205

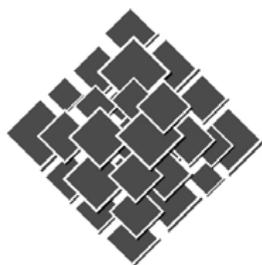
HAL Id: hal-03159205

<https://ephe.hal.science/hal-03159205>

Submitted on 4 Mar 2021

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

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



École Pratique
des Hautes Études



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

ÉCOLE PRATIQUE DES HAUTES ÉTUDES
Sciences de la Vie et de la Terre

MÉMOIRE

Présenté par
SAUZET SANDRINE

pour l'obtention du Diplôme de l'École Pratique des Hautes Études

TITRE : Développement d'une technique d'identification des protéines associées à un locus spécifique

soutenu le 17 décembre 2020 devant le jury suivant :

DEMIGNOT SYLVIE – Président

DEJARDIN JEROME – Tuteur scientifique

NEILDEZ THI-MY-ANH – Tuteur pédagogique

REINA SAN MARTIN BERNARDO – Rapporteur

SEVENO MARTIAL – Examineur

Mémoire préparé sous la direction de :

Jérôme DEJARDIN

Intitulé de la structure d'accueil : UMR 9002 - Institut de Génétique Humaine - Montpellier

Directeur : M. BENKIRANE

et de

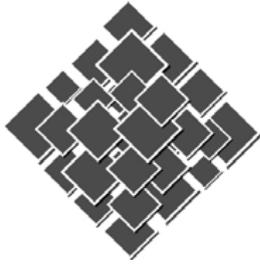
Thi My Anh NEILDEZ

**Intitulé de la structure d'accueil EPHE : UMR_S 938 Centre de Recherche Saint-Antoine (CRSA)
- Paris**

Directeur : B. FEVE

EPHE (Sciences de la Vie et de la Terre)

**Groupe de Recherche et d'Enseignement Thématiques de l'EPHE : Modélisation,
Épigénétique et Développement (MÉDé)**



École Pratique
des Hautes Études



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

ÉCOLE PRATIQUE DES HAUTES ÉTUDES
Sciences de la Vie et de la Terre

MÉMOIRE

Présenté par
SAUZET SANDRINE

pour l'obtention du Diplôme de l'École Pratique des Hautes Études

TITRE : Développement d'une technique d'identification des protéines associées à un locus spécifique

soutenu le 17 décembre 2020 devant le jury suivant :

DEMIGNOT SYLVIE – Président

DEJARDIN JEROME – Tuteur scientifique

NEILDEZ THI-MY-ANH – Tuteur pédagogique

REINA SAN MARTIN BERNARDO – Rapporteur

SEVENO MARTIAL – Examineur

Mémoire préparé sous la direction de :

Jérôme DEJARDIN

Intitulé de la structure d'accueil : UMR 9002 - Institut de Génétique Humaine - Montpellier

Directeur : M. BENKIRANE

et de

Thi My Anh NEILDEZ

**Intitulé de la structure d'accueil EPHE : UMR_S 938 Centre de Recherche Saint-Antoine (CRSA)
- Paris**

Directeur : B. FEVE

EPHE (Sciences de la Vie et de la Terre)

**Groupe de Recherche et d'Enseignement Thématiques de l'EPHE : Modélisation,
Épigénétique et Développement (MÉDé)**

REMERCIEMENTS

Je tiens tout d'abord à remercier les membres de mon jury d'avoir accepté cette tâche : Dr Bernardo REINA SAN MARTIN, rapporteur, Dr Martial SEVENO, examinateur et Dr Sylvie DEMIGNOT, président de jury.

Je remercie également l'Institut de Génétique Humaine et notamment Monsef BENKIRANE d'avoir permis que ce projet soit possible en acceptant que je passe ce diplôme ainsi que pour le financement de mes déplacements au cours de ce diplôme. Merci également à Catherine LAROSE pour son aide dans le montage de mon dossier de financement auprès du CNRS. J'en profite également pour remercier la DR13 du CNRS pour son soutien financier dans ce projet.

Merci également à Thi-My-Anh NEILDEZ, qui a accepté d'être ma tutrice pédagogique, pour son encadrement EPHE durant tout mon cursus mais également pour tous ses conseils.

Un autre grand merci plus particulier à Jérôme DEJARDIN qui a été tout au long de ses 2 années mon tuteur scientifique. Merci pour ta grande disponibilité, tes conseils, ton aide mais également pour les corrections que tu as apporté à mon mémoire de fin d'étude. Mais avant tout, merci d'avoir cru en ce projet.

Mes sincères remerciements à Amandine BARRAL et Mathieu TARDAT qui m'ont encouragé au quotidien. Merci pour nos discussions scientifiques ou non scientifiques et pour nos pauses toujours aussi agréables.

Merci à Lucas DUCROT pour sa bonne humeur et ses conseils en mathématiques !

Je remercie également Sihem, Varinia et Bariaa, pour tous ces bons moments au laboratoire.

Merci également à Laetitia GUINTINI pour m'avoir encouragé à me lancer dans cette aventure et à Lauriane FRITSH pour m'avoir permis de persévérer.

C'est enfin à ma petite famille, que j'adresserai mes plus chers remerciements. Rémy, Zoé et Yanis, tout au long de ce projet, vous m'avez incroyablement soutenu, encouragé et supporté...

TABLE DES MATIERES

INTRODUCTION GENERALE	13
I. La chromatine et les éléments de régulation des gènes : contexte scientifique.....	13
I.1. La chromatine.....	13
I.1.1. Modifications d’histones.....	14
I.1.2. L’hétérochromatine.....	14
I.1.2.1. L’hétérochromatine constitutive	15
I.1.2.2. L’hétérochromatine facultative	16
I.1.3. L’euchromatine	16
I.2. Les éléments de régulation des gènes	17
I.2.1. Les promoteurs	18
I.2.2. Les enhanceurs.....	18
I.2.3. Les silencers.....	19
I.2.4. Les insulateurs.....	19
II. Purification des protéines associées au niveau d’un locus spécifique	20
II.1. Enrichissement nécessaire à la purification d’un locus unique	20
II.1.1. Matériel de départ	20
II.1.2. Sensibilité et profondeur d’analyse	21
II.1.3. Taux d’enrichissement	22
II.2. Techniques de purification des protéines associées à un locus spécifique	24
II.2.1. Techniques de purification avec des sondes spécifiques.....	25
II.2.2. Techniques de purification par CRISPR/Cas9	25
II.2.2.1. Purification des protéines avec un système contenant la biotine ligase.....	26
II.2.2.2. Purification des protéines avec un système contenant l’APEX2.....	26
II.2.2.3. Capture par une dCas9 immobilisée	27
II.2.2.4. Capture grâce à un tag	27
OBJECTIF DU PROJET	29

METHODOLOGIE.....	30
I. Culture cellulaire	30
II. Immunofluorescence (Duolink Proximity Ligation Assay –PLA).....	31
III. Western-blot	32
IV. Mesure de la taille des télomères	33
V. PICh : Proteomics of isolated chromatin segments	35
V.1. Crosslink et lyse des cellules	36
V.2. Préparation de la chromatine	36
V.3. Hybridation et capture de la chromatine	37
V.4. Précipitation des protéines et reverse crosslink.....	38
VI. Gel d'argent.....	39
VII. Spectrométrie de masse.....	39
VII.1. Préparation des échantillons	40
VII.2. Spectrométrie de masse et analyse	40
CHAPITRE I : Détermination de l'impact de la protéine dCas9 sur la composition locale de la chromatine	42
I. Problématique de recherche.....	42
II. Méthodologie.....	44
III. Résultats	45
III.1. Caractérisation des populations sélectionnées	45
III.2. Purification des protéines associées aux télomères	47
III.3. Analyse des protéines associées aux télomères	48
IV. Discussion et perspectives	50
CHAPITRE II : Utilisation de la technique du PICh modifiée pour la purification d'un locus unique.....	53
I. Problématique de recherche.....	53
II. Méthodologie.....	54
III. Résultats	55
III.1. Modification des sondes de capture	56

III.2. Augmentation de la sensibilité de l'analyse par spectrométrie de masse	58
III.2.1. Modifications de la préparation des échantillons.....	58
III.2.2. Modification de la technique de spectrométrie de masse	59
III.3. Ajout d'une seconde étape de purification.....	61
III.4. Elution de la seconde étape de purification	64
IV. Discussion.....	67
V. Perspectives.....	70
 CONCLUSION GENERALE	 79
 BIBLIOGRAPHIE.....	 80
 Annexe 1 : Carte des plasmides	 85
Annexe 2 : Tableau des anticorps	87
Annexe 3 : Liste des sondes de PICH et PICH en tandem	88
Annexe 4 : Protocole du PICH en tandem	89

LISTE DES FIGURES

Figure 1 : Représentation schématique de l'euchromatine et de l'hétérochromatine sous forme de fibres de nucléosomes accessibles ou condensées

Figure 2 : Représentation schématique des différents éléments de régulation des gènes

Figure 3 : Principe de la technique Tandem Mass Tag

Figure 4 : L'information obtenue a un locus donné dépend de l'enrichissement de l'échantillon cible.

Figure 5 : Principe de la technique Duolink Proximity Ligation Assay

Figure 6 : Montage de transfert des ADN sur une membrane nylon

Figure 7 : Protocole de la technique PICh

Figure 8 : Protocole de la technique S-TRAP

Figure 9 : Représentation schématique des limitations de la technologie CRISPR/Cas9

Figure 10 : Représentation schématique de la méthodologie employée afin de comprendre l'impact possible de la protéine dCas9 sur la chromatine au niveau local.

Figure 11 : Vérification des populations de cellules HEK293T contenant la dCas9 ou la dCas9 et le guide télomérique

Figure 12 : Vérification de la purification des protéines télomériques par PICh.

Figure 13 : Résultats de spectrométrie de masse sur des cellules HEK293T contenant la dCas9 non guidée ou guidée aux télomères.

Figure 14 : Représentation schématique de la méthodologie pour la compréhension de la spécificité de l'impact de la protéine dCas9 sur la chromatine au niveau local.

Figure 15 : Représentation schématique de la méthodologie employée pour optimiser la technique du PICh.

Figure 16 : Test de différentes sondes de capture.

Figure 17 : Amélioration de la préparation des échantillons de spectrométrie de masse (sensibilité).

Figure 18 : Amélioration de la profondeur d'analyse de spectrométrie de masse (sensibilité).

Figure 19 : Vérification des étapes de purification en tandem sur un locus spécifique, le télomère.

Figure 20 : Vérification de l'élution aux UV de la purification en tandem sur un locus spécifique, le télomère.

Figure 21 : Comparaison du PICh et du PICh en tandem sur un locus spécifique, le télomère.

Figure 22 : Structure et fréquence de mutation du promoteur du gène *TERT*.

Figure 23 : Profils de survie en fonction du statut génétique de glioblastomes.

Figure 24 : Niveaux d'expression du gène *TERT* dans les différents types de glioblastomes.

Figure 25 : Réarrangements situés en amont de *TERT* analysés par séquençage du génome dans des glioblastomes ALT négatifs.

LISTE DES TABLEAUX

Tableau 1 : Analyse des groupes de protéines obtenus par spectrométrie de masse sur des cellules HEK293T contenant la dCas9 non guidée ou guidée aux télomères.

ABBREVIATIONS

32P- γ ATP : Adenosine triphosphate marquée au gamma phosphate 32

ADN : Acide désoxyribonucléique

APEX2 : Peroxydase d'ascorbate modifiée 2

ARN : Acide ribonucléique

BirA : Biotine ligase A

BSA : Bovine serum albumin

Cellules ES : Cellules souches embryonnaires

CHIP : Chromatine immunoprécipitation

COMPASS : Complex proteins associated with Set1p

DAPI : 4',6-diamidino-2-phénylindole

DIG : Digoxigénine

DMEM : Dulbecco's modified eagle's medium

DTB : Desthiobiotine

ECL : Électrochimiluminescence

EDTA : Éthylènediaminetétraacétate

EGTA : Ethylene glycol-bis(2-aminoethylether)-N,N,N',N'-tetraacetate

ETS : E26 transformation-specific

GABPA : GA-binding protein alpha chain

GBM : Glioblastome

HAT : Histone acétyltransférase

HDAC : Histone déacétylase

HEK293T : Human embryonic kidney 293 cells transformed with T antigen

HEPES : 4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid

HPLC : Chromatographie en phase liquide à haute performance (high performance liquid chromatography)

HRP : Peroxydase de raifort (horseradish peroxidase)

IDH : Isocitrate dehydrogenase
LIF : Leukemia inhibitory factor
MEM : Minimal essential medium
MS : mass spectrometry
NaCl : Chlorure de sodium
NGS : Next generation sequencing
PBS : Phosphate-buffered saline
PCR : réaction de polymérisation en chaîne
PLA : Proximity Ligation Assay
POT1 : Protection of telomeres protein 1
PRC2 : Complexe de répression polycomb 2
PVDF : Polyfluorure de vinylidène
RAP1 : Repressor/Activator Protein 1 homolog
RPM : Rotation par minute
SDS : Sodium dodecyl sulfate
SETD2 : Set domain containing protein 2
SILAC : Stable isotope labeling by amino acids in cell culture
SSC : Saline-sodium citrate
SVF : Sérum de veau fœtal
T4 PNK : T4 Polynucleotide Kinase
TALE : Transcription activator-like effector
TBS : Tris-buffered saline
TINF2 : TERF1-interacting nuclear factor 2
TMT : Tandem mass tag
TPP1 : Adrenocortical dysplasia protein homolog
TRF1 : Telomeric repeat-binding factor 1
TFIID : Facteur de transcription II D
TSS : Transcription start site

INTRODUCTION GENERALE

I. La chromatine et les éléments de régulation des gènes : contexte scientifique

I.1. La chromatine

Le génome est constitué d'ADN mais également de protéines permettant à celui-ci de se structurer et de réguler l'accès à la séquence (expression des gènes, réplication et réparation). Cet ensemble est nommé la chromatine. Deux types de chromatine peuvent être distingués : l'euchromatine, la partie « active » du génome, de structure globalement décondensée permettant l'expression génique et l'hétérochromatine, de structure condensée, composée en grande partie de régions répétées.

Les différents états de la chromatine peuvent également être identifiés en fonction des modifications post-traductionnelles présentes sur la queue des histones. Ces modifications jouent le rôle de « marques » épigénétiques et entraînent le recrutement de différentes protéines (Figure 1), favorisant ainsi spécifiquement différentes transactions génomiques.

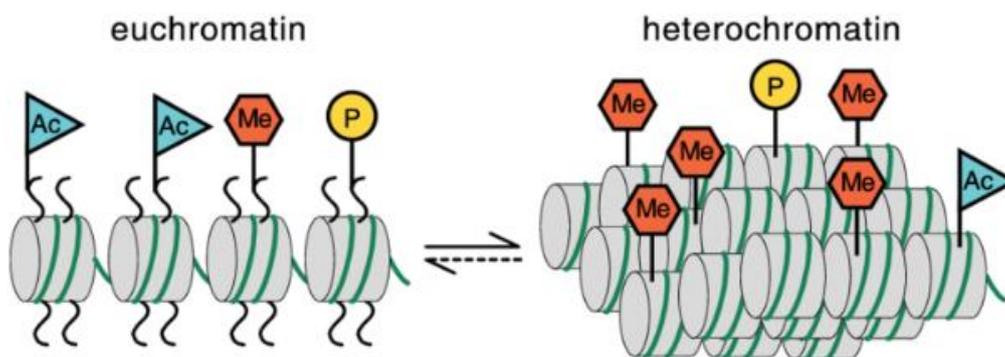


Figure 1 : Représentation schématique de l'euchromatine et de l'hétérochromatine sous forme de fibres de nucléosomes accessibles ou condensées contenant des histones acétylés (Ac), phosphorylés (P) et méthylés (Me). D'après T. and Jenuwein et C. David Allis. Science 293(5532):1074-80 · September 2001

I.1.1. Modifications d'histones

De nombreux facteurs et complexes associés à la chromatine interagissent de façon spécifique avec les histones modifiées.

Les histones peuvent être phosphorylées. Cette modification est très dynamique. La phosphorylation est réalisée sur les sérines, les thréonines et les tyrosines, généralement sur la partie N-terminale des queues d'histones (1).

Tout comme la phosphorylation, l'acétylation des histones est très dynamique. Les lysines des queues d'histones peuvent être acétylées. Cette modification est effectuée par des histones acétyltransférases (HAT) (1). L'acétylation des histones entraîne une diminution des interactions avec l'ADN et les nucléosomes voisins et donc le relâchement de la structure chromatinienne. L'acétylation est par conséquent associée à une activation de la transcription et est facilement réversible par l'action des histones désacétylases (HDAC).

La méthylation des histones est réalisée sur des lysines ou des arginines et peut consister en l'ajout d'un à trois groupements méthyles. Contrairement à l'acétylation ou la phosphorylation, la méthylation ne modifie pas la charge des protéines d'histones (1). Selon les résidus méthylés et le nombre de groupements ajoutés, la méthylation des histones peut être une marque d'activation ou de répression de la transcription. La méthylation des histones est également un processus dynamique et réversible, puisqu'il existe des déméthylases d'histones.

I.1.2. L'hétérochromatine

Grâce à des observations histologiques, Emil Heitz (1928) a identifié l'hétérochromatine comme étant des régions de chromosome qui apparaissent très condensés et très colorés dans le noyau interphasique. Ces régions correspondent à des séquences répétées telles que les péricentromères et les télomères. Ces 20 dernières années, la définition cytologique de

l'hétérochromatine a été étendue à d'autres régions chromosomiques qui ont, comme les péricentromères ou les télomères, une action de répression des gènes.

L'hétérochromatine est définie comme composée principalement de séquences répétées, pauvres en gènes, de structure dense et permettant la répression de l'expression des gènes. L'hétérochromatine est caractérisée par une hypoacétylation des histones et est associée à une méthylation de l'histone 3 au niveau de la Lysine 9 (H3K9me) et à la protéine de l'hétérochromatine (HP1) ou à une méthylation de la lysine 27 de l'histone 3 (H3K27me) (2).

1.1.2.1. L'hétérochromatine constitutive

La majorité des génomes de mammifères est constituée de séquences répétées (télomères, péricentromères, transposons...). Ces séquences posent un problème pour l'intégrité du génome du fait de leur homologie et donc d'un potentiel de recombinaison ou de duplication de ces séquences. Ces séquences sont généralement transcriptionnellement inactivées grâce à une compaction de la chromatine qui limite l'accès pour les facteurs de transcription. Ces régions répétées forment l'hétérochromatine constitutive et sont retrouvées dans l'ensemble des lignées cellulaires (3).

La triméthylation de la lysine 9 de l'histone 3 (H3K9me₃) et HP1 sont abondants et conservés au sein des domaines d'hétérochromatine constitutive. Ces éléments sont nécessaires à la structure et à la fonction de l'hétérochromatine constitutive. La lysine 9 de l'histone 3 est généralement méthylée par l'enzyme SU(VAR)3-9 (Supressor of variegation 3-9) notamment dans les régions péricentromériques, qui représentent la quasi-totalité de cette hétérochromatine (2). Une autre famille de SET methyl-transférases qui utilisent H3K9 comme substrat a été caractérisée au niveau des rétrovirus endogènes mais aussi au niveau des télomères (4). Il s'agit de l'histone méthyl-transférase SETDB1 (SET Domain Bifurcated Histone Lysine Methyltransferase 1).

I.1.2.2. L'hétérochromatine facultative

L'hétérochromatine facultative est localisée dans des régions dynamiques, au niveau de gènes et des régions activatrices des gènes, de façon spécifique à un type cellulaire (5).

La répression de l'expression génique dépendante du type cellulaire de beaucoup de gènes est sous contrôle de la triméthylation de la lysine 27 de l'histone 3 (H3K27me3). Cette marque d'histone est catalysée par le complexe de répression polycomb 2 (PRC2). Ce type de répression de la transcription est retrouvé au niveau de nombreux gènes de facteurs de transcription spécifiant la lignée tels que la famille des gènes homéotiques (HOX) (6).

I.1.3. L'euchromatine

L'euchromatine est de structure globalement ouverte, riche en gènes et permet l'activation de l'expression des gènes.

L'euchromatine est riche en histones H3 et H4 acétylés ainsi qu'en H3K4 méthylé (H3K4me). Le complexe Set1 (complexe COMPASS : complex proteins associated with Set1p) catalyse la méthylation de H3K4. La triméthylation de H3K4 est une marque d'histone associée à un gène actif. H3K4me3 est enrichi autour du site d'initiation de la transcription (7). L'initiation de la transcription correspond à la synthèse de la première liaison phosphodiester réalisée par la sous-unité catalytique de l'ARN-polymérase.

H3K36 triméthylé est également retrouvée au niveau des gènes activement transcrits. Cette marque est catalysée par l'histone méthyltransférase SETD2. SETD2 se lie à l'ARN polymérase II phosphorylée lors de l'élongation transcriptionnelle. La marque H3K36me3 se trouve donc dans le corps du gène et est une marque de l'élongation (7). L'élongation correspond au déplacement de la transcription le long du gène.

1.2. Les éléments de régulation des gènes

L'expression génique est liée à l'état de la chromatine mais est aussi sous le contrôle des éléments de régulation des gènes tels que les promoteurs, les enhanceurs, les silencers et les insulateurs (8).

La transcription des gènes est un processus très complexe et hautement régulé. L'expression de chaque gène est sous contrôle d'éléments localisés en cis des gènes (promoteur) ou à une plus grande distance (enhancer, silencer, insulateur) (Figure 2). Le développement de nouvelles technologies a permis l'identification et la caractérisation des différents éléments de régulation des gènes permettant ainsi une meilleure compréhension de leur lien sur l'expression génique dans un contexte biologique normal ou dans des conditions pathologiques (9).

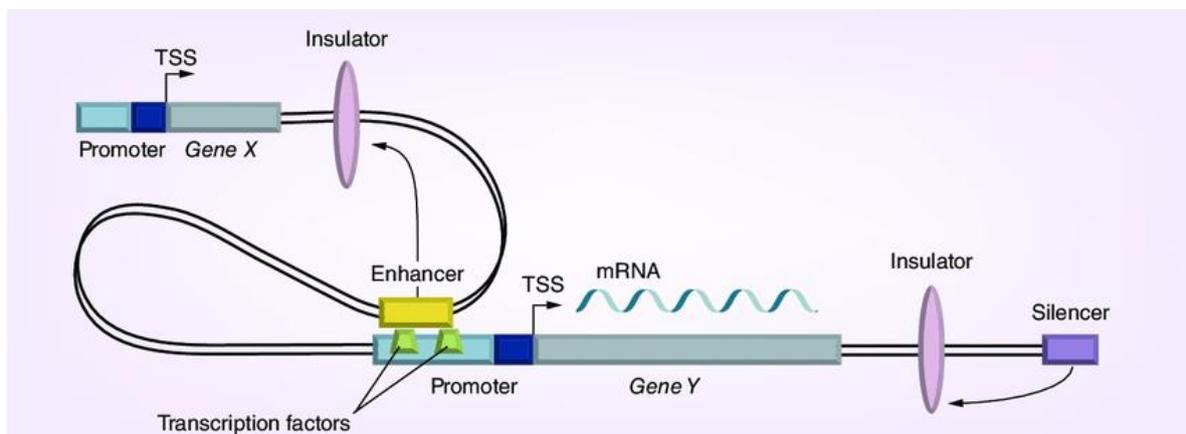


Figure 2 : Représentation schématique des différents éléments de régulation des gènes. Le promoteur est localisé proche du TSS (Transcription start site) des gènes. Les promoteurs ont des sites de fixation des facteurs de transcription qui servent également d'ancrage aux enhanceurs. La formation de la boucle promoteur-enhancer active la transcription du gène cible. Les silencers ont un effet opposé aux enhanceurs (inactivation de l'expression des gènes). Les insulateurs agissent comme une barrière pour empêcher les enhanceurs et les silencers de réguler les gènes cibles. D'après Marcelo R Luizon and Nadav Ahituv. *Pharmacogenomics*. 2015 Nov; 16(16): 1829–1841.

1.2.1. Les promoteurs

Les promoteurs des gènes sont des éléments qui se trouvent en amont du TSS des gènes et permettent l'initiation de la transcription. Le corps du promoteur est défini comme la séquence minimale qui permet l'initiation de la transcription. Le rôle du promoteur est de se lier au complexe d'initiation de la transcription dont l'activité catalytique principale est l'ARN polymérase (10).

Différents types de promoteurs ont été montrés chez les mammifères. Les promoteurs principalement rencontrés au niveau des gènes constitutifs contiennent fréquemment des îlots CpG. La plupart des promoteurs avec un faible taux de CpG sont localisés au niveau d'un TSS précis et possèdent une boîte TATA (séquence d'ADN présente au niveau de la séquence promotrice et qui sert de lieu de reconnaissance de l'ARN polymérase). Les promoteurs ayant une boîte TATA sont associés à des gènes ayant une transcription tissu-spécifique (10,11).

La triméthylation de la lysine 4 de l'histone 3 est généralement présente au niveau du TSS. L'ARN polymérase II est présente sur les gènes actifs (12).

1.2.2. Les enhancers

Les gènes peuvent aussi être fortement influencés par la présence d'enhancers qui sont des séquences pouvant se situer à une distance considérable du promoteur (parfois plus de 100kb) et contenant des sites de fixation des facteurs de transcription qui vont ensuite interagir avec le complexe Mediator ou le facteur de transcription II D (TRIID) et aider au recrutement de l'ARN polymérase 2 au niveau du promoteur des gènes (13).

Les enhancers peuvent activer la transcription indépendamment de leur localisation, de leur distance ou de leur orientation. En effet, les facteurs de transcription liés à l'enhancer forment une boucle qui va permettre la mise en contact de l'enhancer avec le promoteur (14). Des études ont montré l'implication de la cohésine pour connecter les enhancers aux promoteurs de façon physique et fonctionnelle (15).

L'activation des gènes nécessite également la décompaction de la chromatine. Les facteurs de transcription liés aux enhancers peuvent ainsi recruter des enzymes de modification des histones ou des complexes permettant de remodeler la chromatine afin d'augmenter l'accessibilité de l'ADN à d'autres protéines. Les enhancers actifs ou inactifs peuvent être respectivement distingués par la présence d'H3K27ac ou d'H3K27me3 (13).

1.2.3. Les silencers

L'hétérochromatine entraîne une répression de la transcription. On peut la trouver au niveau de séquences répétées ou au niveau de séquences spécifiques appelées des silencers. Les silencers sont des éléments qui établissent un état répressif qui peut être stable. Les silencers répriment les promoteurs à distance via des complexes tels que les protéines polycomb afin d'initier la formation d'hétérochromatine (14), c'est notamment le cas chez la Drosophile. Le recrutement du complexe polycomb entraîne ainsi la triméthylation de la lysine 27 de l'histone 3 sur une large région et corrèle avec une inhibition de la transcription au niveau du gène cible (15).

1.2.4. Les insulateurs

Les insulateurs permettent aussi la régulation des enhancers et des silencers. Ils peuvent avoir 2 types d'activité : 'bloquante' ou 'barrière'. Les insulateurs bloquants peuvent interrompre la communication entre un enhancer et son promoteur lorsqu'ils sont positionnés entre les 2. Les insulateurs barrières sont localisés entre un silencer et un promoteur et protègent le promoteur de la répression (15).

La plupart des insulateurs peuvent utiliser des facteurs qui ont la propriété d'interagir à une longue distance avec d'autres éléments de régulation des gènes (14).

II. Purification des protéines associées au niveau d'un locus spécifique

L'établissement de la composition de loci spécifiques permet l'identification des facteurs de régulation spécifiques de manière non biaisée. Les protéines associées au niveau d'un locus et l'état de la chromatine donnent des indications essentielles pour la compréhension de l'activation ou de la répression spécifique au locus. L'identification des protéines associées au niveau d'un locus donné est donc cruciale pour identifier les protéines qui interagissent au locus dans des conditions normales ou lors du développement ou de pathologies.

Actuellement, la biologie de la chromatine (régulation du génome par les protéines) est essentiellement étudiée par la technique du CHIP (immunoprécipitation de la chromatine). Cette technique peut être couplée à la PCR quantitative ou au NGS. Toutefois, cette technologie nécessite des connaissances préalables à propos de la protéine candidate à étudier. De plus, cette technique nécessite l'obtention d'un anticorps approprié pour la protéine à étudier.

Différents verrous technologiques doivent être levés pour pouvoir purifier une très petite quantité d'ADN et ses protéines associées, par rapport au reste du génome (16).

II.1. Enrichissement nécessaire à la purification d'un locus unique

II.1.1. Matériel de départ

Actuellement, les quantités de peptides analysables par spectrométrie de masse sont de l'ordre de la femtomole. Les peptides, contrairement aux ADN, ne peuvent pas être amplifiés. Il est donc nécessaire de partir d'une quantité de matériel de départ importante qui permettra d'obtenir suffisamment de chaque peptide pour qu'il soit identifiable. Par conséquent, il est plus facile d'identifier des protéines abondantes au locus cible d'autant plus si elles sont en faible quantité sur le reste du génome (17). En effet, dans ce type de technique, l'ensemble du reste du génome représente le bruit de fond. Par conséquent, si une protéine est présente

essentiellement au locus cible, alors son identification sera plus facile puisqu'elle sera en beaucoup plus petite quantité dans le bruit de fond (i.e. le reste du génome).

Quelques femtomoles de peptides représentent 600 millions de molécules. Donc en partant de l'hypothèse que 100% de la cible est isolée, il faut partir de 300 millions de cellules diploïdes pour obtenir 600 millions de molécules afin de purifier des protéines abondantes à un locus unique. La quantité de matériel devra donc être encore augmentée pour purifier les protéines moins abondantes. De plus, la quantité de chaque protéine a un locus donné ou à un moment donné peut varier (16).

II.1.2. Sensibilité et profondeur d'analyse

La composition de la chromatine au niveau local est très complexe d'une part par la très grande variété de protéines pouvant être associées et d'autre part, par les très grandes différences de niveaux d'abondance de chacune de ces protéines à un temps donné. Par exemple, une histone (composant structural ubiquitaire de la chromatine) sera toujours retrouvée en abondance alors qu'une enzyme de modification ou de remodelage sera souvent 100 fois, voire 1000 fois moins abondante. Une analyse par spectrométrie de masse avec une bonne profondeur est donc nécessaire pour identifier de tels partenaires. Le mélange peptidique peut être rendu moins complexe par un fractionnement des peptides sur un gradient de méthanol ou d'acétonitrile. De cette manière, chaque fraction sera moins complexe et des protéines moins abondantes ont plus de chances d'être identifiées.

De plus, afin de comparer les résultats obtenus d'un type cellulaire à l'autre pour un même locus, une analyse quantitative est nécessaire. Récemment, différentes stratégies de protéomique quantitative ont été développées. Ces techniques sont basées sur l'incorporation d'isotopes en culture cellulaire ou dans un organisme ou via un marquage isotopique au niveau des protéines ou des peptides (18). Le SILAC (Stable Isotope Labelling by amino acids in cell culture) est la technique la plus connue. Des isotopes d'acides aminés naturels (Lysine et Arginine) dits « légers » (condition standard) ou bien leurs équivalents « lourds » sont ajoutés dans le milieu de culture des cellules. Les acides aminés lourds contiennent des isotopes non radioactifs tels que C13 et N15 (19). L'incorporation

métabolique peut être très efficace et peut donner de hauts niveaux de marquage. Toutefois, cette technique est très coûteuse. Le marquage isotopique à posteriori des peptides est une autre alternative. Différentes entités stables peuvent être incorporées. Le marquage isobarique (Tandem Mass Tag : TMT) est une de ces techniques (Figure 3). La quantification est réalisée au niveau MS/MS. Actuellement, différents kits commerciaux permettent de multiplexer les échantillons. Le fait de multiplexer les échantillons permet de diminuer les variations entre les échantillons car leur analyse dans le spectromètre de masse sera réalisée en une seule fois (pas de variations liées à l'analyse). Par conséquent, le multiplexage des échantillons augmente l'efficacité de quantification et de comparaison des échantillons. La comparaison des résultats obtenus par ces différentes techniques démontre que la technique TMT est la plus précise (18).

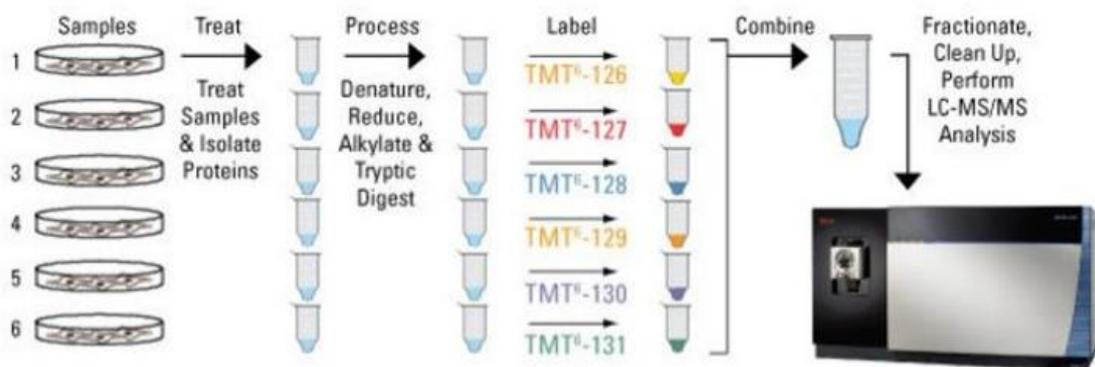


Figure 3 : Principe de la technique Tandem Mass Tag d'après la compagnie Life Technologie

II.1.3. Taux d'enrichissement

De nombreuses protéines de la chromatine sont dynamiques et donc s'associent à un locus donné de façon transitoire. Des protéines peuvent donc échapper à l'analyse si la quantité de matériel de départ est trop faible ou si le taux d'enrichissement n'est pas suffisant (16). Un enrichissement important a pour but de discriminer le signal du bruit de fond. Le taux d'enrichissement pour un locus cible doit être calculé en fonction de la taille de ce locus

rapportée à la taille du génome car dans cette approche, c'est l'ensemble du reste du génome qui constitue le bruit de fond (Figure 4). Ce point est crucial (17).

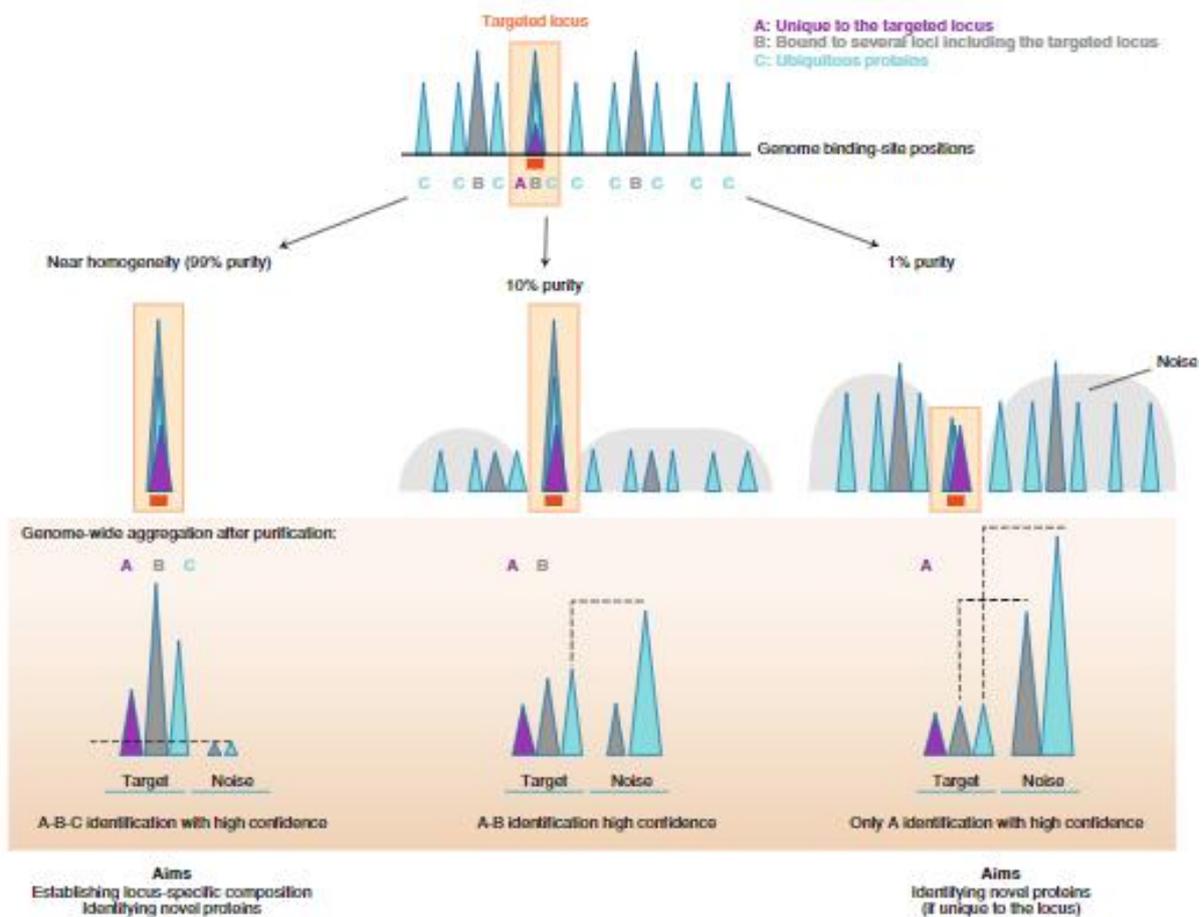


Figure 4 : L'information obtenue a un locus donné dépend de l'enrichissement de l'échantillon cible. D'après Gauchier M. et al. Nat Methods. avr 2020;17(4):380-9.

Par exemple, la région péricentromérique représente environ 3% du génome murin. Pour enrichir 3% du génome jusqu'à homogénéité (99% du matériel purifié est d'origine péricentromérique), il faut un taux d'enrichissement de 33 fois. Selon la même logique, pour enrichir un locus unique de 3kpb (~0.0001% du génome humain), le taux d'enrichissement pour obtenir un mélange quasi-homogène doit être de 1 million de fois (17).

Une seule étape de purification ne peut pas être suffisante pour obtenir un taux d'enrichissement aussi élevé. Il est donc nécessaire de réaliser des étapes successives de

purification pour caractériser un segment unique du génome et ainsi retrouver les éléments caractéristiques de ce locus (par exemple, les facteurs de transcription et l'ARN polymérase II au promoteur). Si ces protéines caractéristiques ne sont pas retrouvées lors de la purification, les résultats devront être interprétés avec précaution (16).

II.2. Techniques de purification des protéines associées à un locus spécifique

Ces 10 dernières années, différentes techniques de purification de locus spécifique ont été proposées. Le locus est principalement ciblé grâce à des sondes spécifiques (20,21) ou grâce aux technologies CRISPR/Cas9 (22–29).

D'autres stratégies ont été développées. Notamment une utilisant les domaines TALE (Transcription activator-like effector) fusionnées à un domaine capturable par anticorps. Ces protéines permettent de créer des domaines de liaison à l'ADN spécifiques d'une séquence d'intérêt, dans le but de lier le système de capture de façon spécifique à une séquence cible. Les protéines TALE sont taguées afin d'être immunoprécipitées pour enrichir les protéines de la chromatine de ce segment d'ADN (30). Dans d'autres stratégies, le locus d'intérêt est modifié par une séquence d'ADN spécifique qui peut être capturée par des protéines adaptatrices (par exemple Tet ou LexA). Le locus fixe est enrichi par immunoprécipitation grâce à ces protéines adaptatrices (31–33). Enfin, une technique appelée Epidecoder, ne nécessitant pas d'analyse protéomique a été développée chez la levure (34). Cette technique se base sur l'étiquetage systématique des protéines par des codes-barres et une analyse quantitative par CHIP. Cette technique très prometteuse a permis l'identification de facteurs au niveau d'un locus unique chez la levure.

II.2.1. Techniques de purification avec des sondes spécifiques

Actuellement 2 techniques ont été utilisées pour purifier un locus spécifique grâce à des sondes : le PICh (20) et l'HYCCAPP (21). Ces 2 techniques utilisent la spécificité de l'ADN cible pour purifier la chromatine. Ces techniques ne nécessitent aucune modification du génome et sont basées sur une capture par hybridation spécifique (17).

Le PICh (proteomics of isolated chromatin segment) est la première méthode publiée qui combine la purification d'un locus et la spectrométrie de masse. Le PICh est basé sur l'hybridation de sondes synthétiques sur un locus spécifique de façon stable. Afin d'augmenter la stabilité de l'interaction de la sonde à l'ADN, les sondes utilisées sont de type LNA (locked nucleic acid). Les sondes contiennent également un analogue de la biotine, la desthiobiotine qui permet une capture réversible par des billes de streptavidine (20).

L'HYCCAP (capture of chromatin-associated proteins for proteomics) est une technologie similaire employée chez la levure et qui utilise également des sondes pour cibler le locus d'intérêt. Les sondes utilisées sont des sondes ADN qui de ce fait ont une plus faible stabilité. Les taux d'enrichissement retrouvés dans cette technique sont par conséquent plus faibles que ceux obtenus avec le PICh (17). Les sondes contiennent une biotine et la capture, irréversible, du locus cible est réalisée par des billes de streptavidine (21).

II.2.2. Techniques de purification par CRISPR/Cas9

Les systèmes CRISPR/Cas9 peuvent se lier de façon précise à une région d'intérêt du génome grâce à un ARN guide d'environ 20 nucléotides. Des techniques de purification à un locus spécifique utilisent cette spécificité pour l'adressage du système de capture des protéines grâce à la combinaison d'une Cas9 inactive (dCas9) et d'un ou plusieurs ARN guide(s). Différents systèmes de capture des protéines sont utilisés pour réaliser la purification par affinité.

II.2.2.1. Purification des protéines avec un système contenant la biotine ligase

Différentes techniques de purification des protéines de locus spécifiques utilisent un système contenant la biotine ligase (BirA). La BirA biotinyler de façon spécifique les lysines d'une séquence peptidique donnée. Après mutation de la BirA (BirA*), l'activité de la biotine ligase est modifiée. La BirA* peut biotinyler toutes les lysines des protéines environnantes. Cette caractéristique a permis le développement d'une technique appelée BioID. Cette technique permet l'identification de protéines qui se trouvent à proximité d'un facteur protéique donné et leur analyse par spectrométrie de masse (35).

Deux techniques utilisant la biotine ligase ont été développées : CAPTURE (CRISPR affinity purification *in situ* of regulatory elements) et CasID.

La technique CAPTURE utilise la biotinylation par une BirA, d'une Cas9 inactive (dCas9) guidée à une séquence spécifique par un ARN guide. La dCas9 taguée avec un site accepteur de biotine et son ARN guide servent au bon adressage du système au niveau du locus. Après biotinylation *in vivo* par la BirA, les protéines sont capturées par des billes de streptavidine (22).

La technique CasID utilise également une dCas9 et un ARN guide pour le bon adressage du système. Par contre, la biotine ligase utilisée est une BirA* qui biotinyler l'ensemble des protéines environnantes. Les protéines étiquetées par la BirA* sont purifiées grâce à des billes de streptavidine. Du fait de la biotinylation *in vivo*, les protéines n'ont pas besoin d'être crosslinkées (23)

II.2.2.2. Purification des protéines avec un système contenant l'APEX2

D'autres techniques basées sur l'étiquetage des protéines environnantes ont été développées plus récemment. Ces méthodes sont basées sur l'utilisation d'une dCas9 en fusion avec une peroxydase d'ascorbate modifiée (APEX2). L'APEX2 génère des radicaux biotine-phénoxy très réactifs à partir de biotine-phénol et de peroxyde d'hydrogène. La biotinylation par l'APEX2

est très rapide et efficace et le temps d'incubation est très court. Les radicaux formés ont également une durée de vie très courte et peuvent interagir avec la tyrosine des protéines à une distance maximale de 20 nm. La résolution spatiale est donc meilleure qu'avec les techniques basées sur la BirA car les protéines identifiées seront beaucoup plus proches de la cible.

Les 3 technologies développées grâce à l'APEX2 sont GloPro (Genomic locus proteomics) (24), C-Berst (dCas9–APEX2 biotinylation at genomic elements by restricted spatial tagging) (25) et CAPLOCUS (Combining CRISPR and peroxidase APEX2 system to identify local chromatin interactions) (26). Ces techniques ont été utilisées pour purifier des loci uniques grâce à une purification des protéines biotinylées par des billes de streptavidine et par une analyse par spectrométrie de masse.

II.2.2.3. Capture par une dCas9 immobilisée

Une technologie originale a été développée récemment, il s'agit de CLASP (Cas9 locus-associated proteome). Dans cette technique, la chromatine est ciblée *in vitro* par une dCas9 immobilisée. Cette technique a été utilisée chez *Drosophila melanogaster* (27).

Dans la technique CLASP, la capture est réalisée *in vitro* en passant une chromatine crosslinkée sur une dCas9 immobilisée qui se lie de façon spécifique à des guides ARN. Cette technique ne nécessite pas d'opération préalable de génie génétique.

II.2.2.4. Capture grâce à un tag

Enfin, deux technologies basées sur une immunoprécipitation d'une Cas9 inactive ont été développées : enCHIP (engineered DNA-binding molecule-mediated chromatin immunoprecipitation) (28) et CRISPR CHAP-MS (Chromatin Affinity Purification with Mass Spectrometry) (29).

Dans ces techniques, la dCas9 est guidée au locus spécifique grâce à un ARN guide. Dans la technologie enCHIP, la purification de la chromatine crosslinkée est réalisée grâce à un anticorps dirigé contre un tag, fusionné à la dCas9. Dans la technologie CHAP-MS, la purification est réalisée grâce à une protéine A, fusionnée à la dCas9. La protéine A permet une purification de la chromatine crosslinkée par des billes d'IgG.

OBJECTIF DU PROJET

Un des axes majeurs de l'équipe « Biologie des séquences répétées » de l'Institut de Génétique Humaine à Montpellier, est de caractériser la composition de loci spécifiques et d'identifier des facteurs spécifiques à leur régulation. Des études ont été initiées en utilisant une version quantitative de la technologie PICh (proteomics of isolated chromatin segments) (4,20,36). Cette approche permet la caractérisation non biaisée des protéines liées à un locus spécifique et ainsi d'identifier des éléments déterminants dans leur régulation. Toutefois à ce jour, cette technique est applicable uniquement sur des régions relativement abondantes, donc répétées (télomères, centromères...). En effet cette méthode ne reposant que sur une seule étape d'affinité, les taux d'enrichissements ne dépassent pas les 10,000.

Différentes techniques de purification d'un locus unique ont été publiées ces dernières années (21–34). Ces techniques reposent principalement sur un système CRISPR/Cas9. Les taux d'enrichissement obtenus ne sont à l'heure actuelle pas suffisants pour que le signal obtenu permette une identification des éléments constitutifs des locus purifiés au niveau d'un locus unique de mammifère. De plus, la taille de la protéine dCas9 (> 150kDa) peut influencer la structure de la chromatine ou concurrencer la liaison d'interacteurs biologiques.

Dans une première partie, l'impact de la fixation de la protéine dCas9 sur la composition de la chromatine au niveau local est étudiée afin de déterminer si sa fixation entraîne des modifications. Dans une seconde partie, une nouvelle technique de purification des protéines associées au niveau d'un locus unique est proposée. Cette technique est basée sur une capture par des sondes spécifiques avec une purification par affinité en tandem.

METHODOLOGIE

I. Culture cellulaire

Les cellules HEK293T sont maintenues en culture dans un milieu DMEM supplémenté avec 10% de SVF (Sérum de veau Fœtal), 2 mM de glutamine, 100U/mL de pénicilline et 100U/mL de streptomycine.

Les cellules HEK293T sont transfectées par le plasmide pdCas9-HA (annexe 1) grâce à la lipofectamine 3000 (Life Technologie) selon le protocole donné par le fournisseur. 24 heures après la transfection, les cellules sont sélectionnées par la blasticidine à 5 µg/mL. Après 15 jours de sélection, la population est récupérée puis remise en culture. La population est testée afin de vérifier la présence de la protéine dCas9 puis une partie de la population est transfectée par le plasmide contenant le guide et sélectionnée 2 semaines par la puromycine à 2µg/mL. La deuxième population (dCas9 guidée aux télomères) est récupérée puis testée.

Les cellules souches embryonnaires (ES) de souris sont maintenues en culture dans des boîtes de culture préalablement coatées avec 0.18% de gélatine. Le milieu de culture des cellules ES est composé de DMEM supplémenté avec 15% de SVF, 2mM de glutamine, 1X de MEM (Minimal essential medium), 0.2mM de 2-mercaptoethanol, 2mL de LIF (Leukemia inhibitory factor), 100U/mL de pénicilline et 100U/mL de streptomycine.

Toutes les cellules en culture sont maintenues à 37°C avec 5% de CO₂ dans des incubateurs.

II. Immunofluorescence (Duolink Proximity Ligation Assay –PLA)

La technique *d'in situ* Proximity Ligation Assay (PLA) permet une détection directe des interactions entre 2 protéines. En effet, deux anticorps primaires (de type A) spécifiques des protéines à caractériser sont utilisés. Deux autres anticorps spécifiques aux anticorps A et possédant un brin d'ADN vont se lier aux anticorps A. Lorsque ces deux anticorps sont proches, les brins d'ADN peuvent interagir et deux amplifications circulaires de l'ADN sont réalisées (Figure 5).

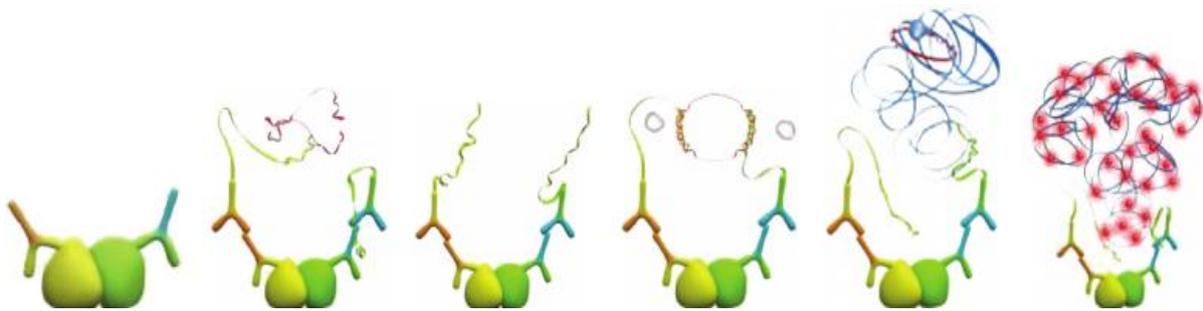


Figure 5 : Principe de la technique Duolink Proximity Ligation Assay d'après la compagnie SIGMA-ALDRICH.

Les cellules sont cultivées dans des plaques 6 puits contenant des lamelles. Lorsque les cellules ont une bonne confluence sur les lamelles, elles sont fixées avec 3.7% de formaldéhyde et perméabilisées avec 0.3% de triton pendant 5 min à RT, sous agitation. Les lamelles sont ensuite lavées 3 fois 5 min avec du PBS avant d'être recouverte de la solution bloquante et incubées 60 minutes à 37°C. Les anticorps primaires sont ensuite dilués dans la solution diluante. Les lamelles sont incubées 1 heure à 37°C avec les anticorps primaires. Les lamelles sont lavées 2 fois 5 min dans le tampon de lavage A à température ambiante. Les sondes MINUS et PLUS PLA sont diluées dans la solution diluante et ajoutées sur les lamelles puis incubées 1 heure à 37°C. Les lamelles sont à nouveau lavées 2 fois 5 min dans le tampon de lavage A puis incubées 30 min à 37°C avec la solution de ligation. Après 2 lavages de 5 min avec le tampon de lavage A, les lamelles sont incubées 100 min à 37°C avec la solution d'amplification. Enfin, les lamelles sont lavées 2 fois 10 min avec le tampon de lavage B puis 1 min avec le tampon de lavage B dilué à 0.01X. Les lamelles sont montées sur des lames en

utilisant un milieu de montage contenant du DAPI avant d'être visualisées au microscope à épifluorescence à champ plein (Zeiss Axioimager avec un Apotome).

L'ensemble des anticorps utilisés est répertorié dans l'annexe 2.

III. Western-blot

Les échantillons utilisés pour le western-blot sont soit des échantillons issus de la technique du PICh ou du PICh en tandem soit des extraits totaux de protéines préparés par lyse dans un tampon contenant 50mM de Tris pH 7.5, 600mM de NaCl, 1% d'IGEPAL, 0.1% de SDS et 0.5% de sodium déoxycholate. La lyse est réalisée sur la glace, dans les boîtes de culture après lavage des cellules au PBS. Les cellules lysées sont récupérées et placées dans un microtube. 0.5µL de benzonase sont ajoutés par 200µL d'échantillon. Les échantillons sont incubés 1H sur la glace. Enfin, la quantité de protéines est mesurée par dosage colorimétrique avec le kit Pierce BSA (Invitrogen). Les protéines sont conservées à -20°C.

Pour les extraits totaux de protéines, un tampon de charge et 10% de β-mercaptoéthanol sont ajoutés aux échantillons qui sont ensuite placés 5 min à 95°C avant d'être déposés sur le gel d'acrylamide.

Les échantillons de PICh ou de PICh en tandem ne doivent pas être à nouveau bouillis avant d'être déposés dans le gel d'acrylamide pour éviter leur dégradation.

Les protéines sont déposées sur un gel de polyacrylamide (Bio-Rad). La migration est réalisée durant environ 1h à 110V dans un tampon contenant 25mM de Tris, 192 mM de glycine et 0.1% de SDS. Après leur migration, les protéines sont transférées sur une membrane PVDF préalablement activée dans du méthanol. Le transfert s'effectue dans un tampon Tris-glycine (25 mM de Tris et 192 mM de Glycine) avec 20% de méthanol pendant 1h20 à 110V.

Après transfert, la membrane est incubée dans un tampon de blocage (TBS 1X contenant 0.1% de Tween et 10% de lait) pendant 30 min sous agitation. L'incubation de l'anticorps primaire

s'effectue dans la solution de blocage, la nuit, à 4°C, sous agitation. Après 2 lavages dans un tampon TBS avec 0.1% de Tween à température ambiante, la membrane est incubée dans un anticorps secondaire-HRP (spécifique de l'espèce de l'anticorps primaire) pendant 45 min sous agitation. Enfin, 2 lavages sont effectués au TBS avec 0.1% de Tween et la membrane est révélée à l'ECL.

L'ensemble des anticorps utilisés est répertorié dans l'annexe 2.

IV. Mesure de la taille des télomères

La taille des télomères est mesurée par la technique de « telomere restriction fragment analysis », qui consiste à digérer tout l'ADN génomique par restriction à l'aide d'enzymes qui coupent fréquemment le génome sauf les télomères (qui ne contiennent pas de tels sites) et de mesurer la taille de ces télomères par séparation sur gel d'agarose, transfert de l'ADN par Southern et révélation du locus télomérique par hybridation à l'aide d'une sonde radiomarquée spécifique.

Pour cela, les cellules sont lavées au PBS puis l'ADN est extrait avec 200µL de tampon de lyse (20mM de NaCl, 20mM de Tris pH8, 4mM d'EDTA et 1% de SDS) par boîte de culture de diamètre 10cm. 200µL de RNase (25µg/mL) sont ajoutés et les échantillons sont incubés 1H à 37°C. Après cette incubation, 5µL de protéinase K sont additionnés (20mg/mL) et les tubes sont placés la nuit à 55°C. L'ADN est ensuite extrait avec 400µL de phénol-chloroforme. Après 5 minutes de centrifugation, la phase aqueuse est récupérée et 400µL de chloroforme sont ajoutés. Les échantillons sont à nouveau centrifugés. La phase contenant l'ADN est récupérée et 10mL d'éthanol sont ajoutés avec 1/10^e d'acétate de sodium à 3M. L'ADN génomique forme une méduse qui est récupérée et lavée à l'éthanol 70% puis séchée et resuspendue dans de l'eau.

Les ADN sont quantifiés au spectrophotomètre (Nano-Drop). 2µg d'ADN pour chaque échantillon est digéré, 2H à 37°C par un mélange d'enzymes de restriction (Hinfl et RsaI).

Les ADN digérés sont déposés sur un gel d'agarose à 0.8%. La migration est réalisée à 5V/cm pendant 4H. Après la migration, le gel est dépuriné à l'HCl 0.25N puis dénaturé (solution

dénaturante : 0.5M NaOH, 1.5M NaCl) et enfin neutralisé (solution neutralisante : 0.5M Tris-HCl pH 7.5, 3M NaOH). Les ADN sont ensuite transférés la nuit, sur une membrane nylon (Figure 6).

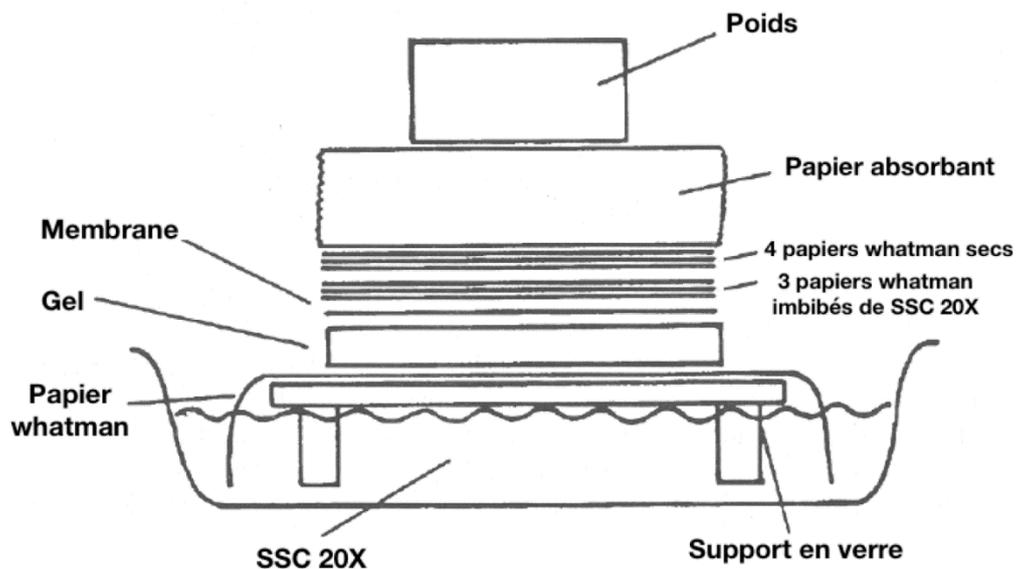


Figure 6 : Montage de transfert des ADN sur une membrane nylon

Après le transfert, la membrane est crosslinkée aux UV puis placée dans le tampon de préhybridation (PerfectHyb plus – Sigma) pendant 30 min à 42°C.

Les ADN sont hybridés avec une sonde ((CCCTAA)₆) radiomarquée. La sonde est préparée en mélangeant 50pmol de sonde avec 5µL de tampon de T4PNK 10X, 2µL de T4PNK, 5µL de ³²P-gATP et 37.5µL d'eau. Ce mélange est ensuite incubé 30 minutes à 37°C puis inactivé 2 minutes à 95°C. La radioactivité est ensuite filtrée sur une colonne G7.

La sonde est dénaturée 5 min à 95°C avant d'être ajoutée au tampon d'hybridation (PerfectHyb plus – Sigma) et la membrane est incubée 3 H à 42°C dans cette solution.

Après l'hybridation, la membrane est lavée 2 fois à température ambiante avec un tampon contenant du 2X SSC et 0.1% de SDS.

La lecture est réalisée sur un Typhoon (Amersham) après exposition de la membrane sur un écran, dans une cassette, pendant la nuit.

V. PICh : Proteomics of isolated chromatin segments

La technique du PICh permet la purification des protéines associées au niveau d'un locus spécifique. Actuellement, cette technique est utilisable uniquement sur des séquences répétées.

Cette technique est basée sur l'utilisation de sondes spécifiques qui permettent la capture d'un locus répété, spécifique et des protéines qui y sont associées après une fixation de ces protéines par crosslink avec du formaldéhyde. La purification est ensuite réalisée grâce à des billes de streptavidine qui vont se fixer à la sonde spécifique qui contient une desthiobiotine en 5'. La desthiobiotine est un analogue de la biotine qui peut se détacher de la streptavidine par compétition avec la biotine. L'élution est donc réalisée par ajout de biotine. Après avoir réversé les crosslinks, les protéines peuvent être analysées par spectrométrie de masse et western-blot (Figure 7).

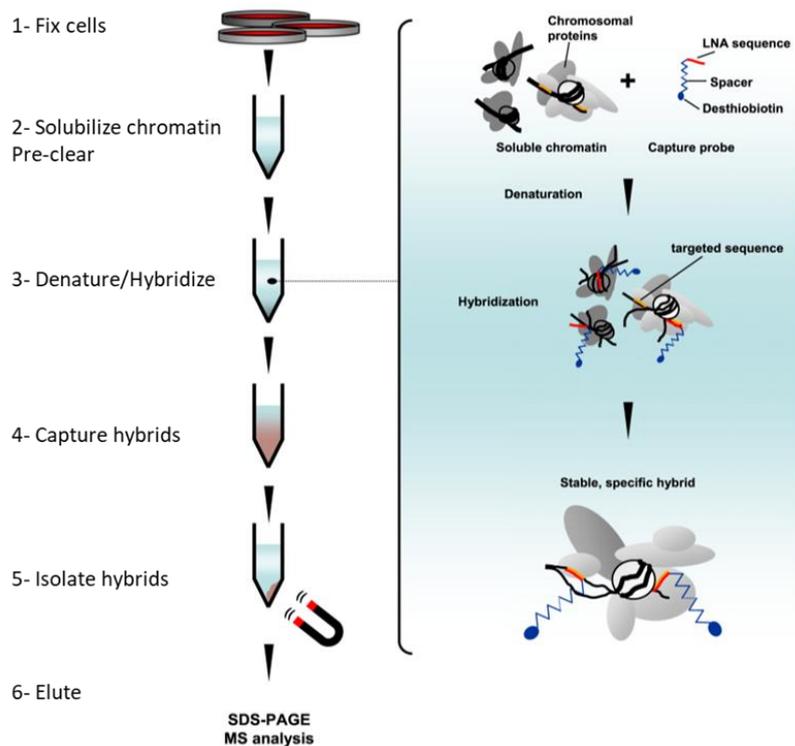


Figure 7 : Protocole de la technique PICh d'après J. Dejardin, RE. Kingston - Cell. 2009 Jan 9; 136(1): 175–186.

V.1. Crosslink et lyse des cellules

Il s'agit de la première étape du PICh (voir figure 7). Les cellules sont cultivées dans des boîtes de culture de diamètre 15cm. Elles sont fixées, par ajout de formaldéhyde au 1/10 dans le milieu de culture et incubées 45 min sous agitation à température ambiante. Les cellules sont lavées 2 fois avec du PBS afin d'éliminer le formaldéhyde. Après ajout d'1mL de solution de scrapping (PBS contenant 0.05% de Tween), les cellules sont grattées et collectées dans des tubes de 50mL maintenus dans la glace. Après 10 minutes de centrifugation à 3600g à 4°C, les cellules sont poolées dans un seul tube 50mL par condition. Elles sont à nouveau lavées 4 fois avec 50mL de PBS. Entre chaque lavage, les cellules sont centrifugées 10 minutes à 3600 g à 4°C. Elles sont ensuite lavées dans du tampon sucrose (0.3M sucrose, 10mM HEPES pH 7.9, 1% Triton, 2mM MgOAc) puis centrifugées 10 minutes à 3600g à 4°C. Après ce dernier lavage, les cellules sont resuspendues dans 15mL de tampon sucrose et transférées dans un homogénéisateur de type douce. Les cellules sont 'doucées' 15 fois avec le pestle 'tight'. La solution est ensuite transférée dans un tube 50mL et centrifugée 10 min à 3600g à 4°C. Le surnageant est jeté et le culot est resuspendu dans 20mL de PBS 0.5% Triton. Après ajout de 50µL de RNase A à 100mg/mL et 200µL de PMSF, les tubes sont incubés la nuit à 4°C, sous rotation.

V.2. Préparation de la chromatine

La deuxième étape du PICh permet d'obtenir la chromatine soluble et pré-clearée (voir figure 7).

Après avoir ajouté 30mL de PBS au tube contenant les 20mL de cellules lysées, le tube est centrifugé 10 minutes à 3600g à 4°C. Le culot est lavé avec 50mL de PBS et centrifugé 10 min à 3600g à 4°C. Puis, le culot est lavé 2 fois avec 50mL de LB4 (50mM Tris pH 8, 200mM NaCl, 20mM EDTA, 1% SDS). Entre chaque lavage, le tube est centrifugé 10 min à 3600g à 4°C. Après le dernier lavage, le surnageant est jeté et le culot est resuspendu avec 2,5 volumes de LB4. La chromatine est aliquotée dans des tubes 15mL (maximum 4mL par tube) et la sonication

est réalisée, afin de solubiliser la chromatine, après ajout de 50µL de PMSF par tube. Les paramètres de sonication sont : amplitude 70%, 15 sec de pulse ON, 45 sec de pulse OFF, 5 min totales de sonication. Après la sonication, la chromatine est incubée 5 min à 58°C. La chromatine est à nouveau aliquotée dans des tubes de 1,5mL et centrifugée 10 min à 16 000 g à RT. Le surnageant (la chromatine solubilisée) est récupéré et poolé dans un tube de 15mL.

Après avoir équilibré 1mL de billes Ultralink Streptavidin dans 10mL de LB4 (centrifuger 5 min à 300g et éliminer le surnageant), la chromatine est ajoutée aux billes afin d'éliminer les protéines endogènes biotinylées (essentiellement des carboxylases). Une incubation de 3H à RT est réalisée puis le tube est stocké la nuit à 4°C. Une colonne avec 1 volume de Sephacryl S-400-HR (1 volume de slurry = 1 volume de chromatine) est préparée. Après une centrifugation de 5 min à 750g afin d'éliminer le tampon de stockage de la Sephacryl, la chromatine est ajoutée sur la colonne. La colonne est centrifugée 5 min à 750g pour faire passer la chromatine et la collecter dans un tube de 50mL. Cette chromatine est ensuite dosée dans un spectrophotomètre à 260 nm. La concentration obtenue est de 1.5 à 2.5 mg/mL avec une DO 260/260 comprise entre 1.3 et 1.45.

V.3. Hybridation et capture de la chromatine

Il s'agit des étapes 3 et 4 de la technique du PICCh (voir figure 7). Après avoir conservé 10µL de chromatine (input), la chromatine est divisée en autant de tubes que d'échantillons (scramble, télomère ...). Afin de réaliser un PICCh sur les télomères, il faut 15mg de chromatine pour des cellules humaines ou 5mg de chromatine pour des cellules ES.

15µL d'une sonde spécifique (annexe 3) à 100µM sont ajoutés dans chaque tube puis vortexés. La chromatine mélangée avec la sonde est distribuée dans des plaques 96 puits (150µL par puit) puis placée dans un thermocycleur. Le programme d'hybridation est le suivant : 3 min à 25°C, 5 min à 82°C, 60 min à 37°C, 3 min à 60°C, 30 min à 37°C, 3 min à 60°C, 30 min à 37°C puis une température finale de 25°C. A l'issue de l'étape d'hybridation, la chromatine est récupérée dans des tubes 1,5mL puis centrifugée 15 min à 18 000g. Le surnageant est ensuite dilué au ½ dans de l'eau puis placé avec 900µL de billes magnétiques MyOne C1 (Life

Technologie) préalablement lavées dans 10mL de tampon LB3 (10mM HEPES pH 7.9, 100mM NaCl, 2mM EDTA pH 8, 1mM EGTA pH 8, 0.2% SDS, 0.1% sodium sarkosyl). Après 90 min sous rotation, les billes sont immobilisées sur un portoir magnétique et le surnageant est récupéré (FT). Il s'agit de l'isolation des hybrides (étape 5 du PICh – voir figure 7). Les billes sont ensuite lavées 5 fois avec 10mL de LB3 (les tubes sont placés sur le portoir magnétique entre chaque lavage pour éliminer le tampon). Les billes sont transférées avec 1mL de tampon LB3 dans des tubes 1.5mL puis lavées avec 1mL de tampon LB3. Un dernier lavage est effectué avec 1mL de tampon LB3S (10mM HEPES pH 7.9, 30mM NaCl, 2mM EDTA pH 8, 1mM EGTA pH 8, 0.2% SDS, 0.1% sodium sarkosyl) à 42°C, sous agitation pendant 5 min.

Après avoir éliminé le surnageant, l'étape d'élution est réalisée (étape 6 – voir figure 7). Pour cela, 900µL de tampon d'élution ($\frac{3}{4}$ LB3 et $\frac{1}{4}$ biotine à 50mM) sont ajoutés aux billes. Puis, les billes sont incubées 30 min à 37°C sous agitation puis 10 min à 65°C avant d'être immobilisées sur le portoir magnétique. Le surnageant est récupéré dans un nouveau tube 1.5mL.

V.4. Précipitation des protéines et reverse crosslink

170µL de TCA sont ajoutés aux éluats. Les tubes sont ensuite incubés 10 min (max 20 minutes) dans de la glace puis centrifugés 15 min à 16 000g à 4°C. Après avoir retiré le surnageant, 1mL d'acétone froide (-20C) est ajoutée. Les tubes sont vortexés puis centrifugés 7 min à 16 000g à 4°C. Le surnageant est retiré et le culot est à nouveau lavé avec 1mL d'acétone froide puis centrifugé 7 min à 16 000g. Après avoir retiré le surnageant, les culots sont séchés quelques secondes à 99°C puis resuspendus avec 100µL de tampon de reverse crosslink (250mM de Tris pH 8.8, 2% SDS, tampon de charge 1X, 1M de β-mercaptoéthanol). Les échantillons sont ensuite placés 5 min à 99°C. Après cette incubation, 80µL de chaque échantillon est prélevé et permettra l'analyse par spectrométrie de masse. Les 20µL restants sont placés à nouveau 20 min à 99°C. Les 20µL d'échantillons incubés 25 min au total à 99°C permettent de réaliser un gel d'argent afin de vérifier le profil de capture mais également des western-blot afin de valider la capture et de confirmer les résultats qui seront obtenus par spectrométrie de masse.

VI. Gel d'argent

Le gel d'argent est réalisé avec le kit SilverQuest Staining (Life Technologie) et permet de visualiser les profils de purification des échantillons.

Après avoir fait migrer les échantillons sur un gel d'acrylamide, le gel est placé dans une boîte de pétri propre et rincé rapidement à l'eau. Le gel est ensuite placé dans 100mL de solution de fixation (10mL d'acide acétique, 40mL d'éthanol, 50mL d'eau ultrapure) et mis 30 sec au micro-onde à 700 watts. Après avoir retiré le gel du micro-onde, il est incubé 5 min sous agitation. La solution de fixation est retirée et le gel est lavé avec 100mL d'une solution à 30% d'éthanol. Puis, le gel est placé 30 secondes au micro-onde puis 5 min sous agitation. L'éthanol est remplacé par 100mL de solution de sensibilisation (10mL de solution de sensibilisant, 30mL d'éthanol qsp 100mL eau ultrapure) et le gel est remplacé 30 sec au micro-onde puis 2 min sous agitation. La solution de sensibilisation est retirée. Le gel est lavé 2 fois dans 100mL d'eau ultrapure (30 sec au micro-onde puis 2 min sous agitation). Le gel est placé dans 100mL de solution de coloration (1mL de solution de coloration qsp 100mL eau ultrapure) puis mis 30 sec au micro-onde et 5 min sous agitation. Le gel est rapidement rincé dans 100mL d'eau ultrapure puis immergé dans 100mL de solution de développement (10mL de solution de développement, 1 goutte d'activateur de développement qsp 100mL eau ultrapure). Le gel est agité jusqu'à obtention de l'intensité de coloration désirée. Lorsque le gel est suffisamment développé, le développement est stoppé avec 10mL de solution de stop et agité 10 min. Le gel est lavé avec 100mL d'eau ultrapure pendant 10 min sous agitation, puis scanné.

VII. Spectrométrie de masse

Les échantillons obtenus par les technologies PICh sont analysés par spectrométrie de masse quantitative. La préparation des échantillons, leur analyse par spectrométrie de masse ainsi que les analyses réalisées sous Max Quant et Perseus sont effectuées par la plateforme de protéomique FPP (Plateforme de Protéomique Fonctionnelle) de Montpellier. Les analyses sont réalisées en collaboration avec cette plateforme.

VII.1. Préparation des échantillons

Les échantillons de PICH ou PICH en tandem contiennent de fortes concentrations en SDS. Le SDS n'étant pas compatible avec les étapes de digestions à la trypsine, les échantillons sont passés sur des colonnes de type S-TRAP (Protifi) selon le protocole du fabricant afin de capturer et nettoyer les protéines avant de les digérer (Figure 8).

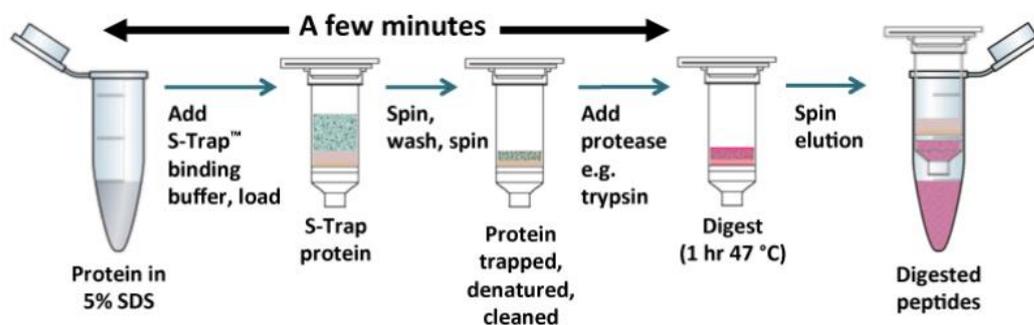


Figure 8 : Protocole de la technique S-TRAP d'après la compagnie Protifi

Après avoir été digérés à la trypsine, les échantillons sont marqués grâce à un marquage isobarique (TMT : tandem Mass Tag) et multiplexés en suivant les instructions du kit TMTduplex ou sixplex™ Isobaric Label Reagent Set (Life technologies). Enfin, ils sont fractionnés en 8 fractions sur un gradient d'acétonitrile afin de séparer les protéines en fonction de leur facteur de rétention (lié à leur polarité).

VII.2. Spectrométrie de masse et analyse

Les échantillons peptidiques sont repris dans 10 µl de tampon A (0,1% acide formique). Les échantillons sont injectés après dilution pour une analyse en ligne en utilisant une HPLC nano débit (RSLC U3000, Thermo Fisher Scientific) couplée à un spectromètre de masse équipé d'une source nanoelectrospray (Q Exactive HF, Thermo Fisher Scientific).

Les peptides sont séparés sur une colonne capillaire (0.075 mm × 150 mm, Acclaim Pepmap 100, phase inverse C18, NanoViper, Thermo Fisher Scientific) suivant un gradient de 2-40% de tampon B en 120 min (A = 0,1% acide formique B = 0.1 % acide formique, 80% acetonitrile) à un débit de 300 nl/min.

Les spectres sont enregistrés via le logiciel Xcalibur 4.1 (Thermo Fisher Scientific) avec la méthode 128LowQtt.meth.

Les données spectrales sont analysées via les logiciels MaxQuant v1.6.10.43, avec utilisation du script leading FPP v3.2. Les bases de données utilisées sont : RefProteome_MOUSE-cano_2020_02_UP000000589 ou RefProteome_HUMAN-cano_2019_10_UP0000005640 avec la modification fixe suivante : Carbamidomethylation (C) et les modifications variables suivantes : Oxidation (M) ; Acetyl (Protein N-term). La validation des données a été réalisée avec les filtres suivants : FDR 1% au niveau peptidique puis protéique.

Le retraitement des données, la comparaison des échantillons et la réalisation des graphiques sont réalisés sur le logiciel Perseus 1.6.10.43.

CHAPITRE I : Détermination de l'impact de la protéine dCas9 sur la composition locale de la chromatine

Le système CRISPR/Cas9 est un système de réponse immunitaire bactérien qui a été adapté comme outil d'édition du génome. Grâce à son utilisation simple et sa grande efficacité, de nombreuses applications et technologies ont été développées (37).

La version catalytiquement inactive de la Cas9 (dCas9) est utilisée dans de nombreuses applications (38–41), l'étude de l'activation ou de la répression des gènes à l'échelle du génome (42), l'imagerie de la chromatine (43) ou la purification de la chromatine (22–29).

I. Problématique de recherche

Ces 10 dernières années, différentes techniques de purification d'un locus spécifique ont été initiées dans différents laboratoires (20–34). Une grande majorité de ces techniques vise à purifier un locus unique mais les enrichissements obtenus actuellement ne sont pas suffisants pour permettre d'affirmer que les protéines identifiées sont effectivement liées au locus unique cible chez les mammifères. En effet, les protéines caractéristiques des loci purifiés n'ont pas été identifiées lors des purifications sur un locus unique de mammifère (16,17).

Plus de la moitié de ces technologies ont été conduites grâce à la protéine dCas9 (22–29). Lors des purifications au niveau d'un locus unique, les auteurs ont noté l'absence de protéines attendues telles que les facteurs de transcription ou l'ARN polymérase II au promoteur mais également l'identification de protéines inattendues. Ceci est particulièrement problématique. Lorsqu'un locus unique est ciblé, une seule molécule dCas9 par génome haploïde est spécifiquement liée à un moment donné. Or dans ces cellules, il existe des milliers (voire plus) de molécules dCas9 non liées. Une grande majorité des protéines identifiées sont donc non spécifiques du locus cible et contribuent par conséquent à augmenter le bruit de fond.

De plus, la protéine dCas9 est une grosse protéine (>150kDa, à peu près la taille d'un nucléosome). Son efficacité de liaison pourrait par conséquent être fortement influencée par

la structure de la chromatine (Figure 9A). De même, la liaison de la protéine dCas9 pourrait concurrencer la liaison d'interacteurs biologiques (Figure 9B). Enfin, la protéine dCas9 peut se fixer de façon non spécifique (off-targets) avec une probabilité significative (Figure 9C) (17).

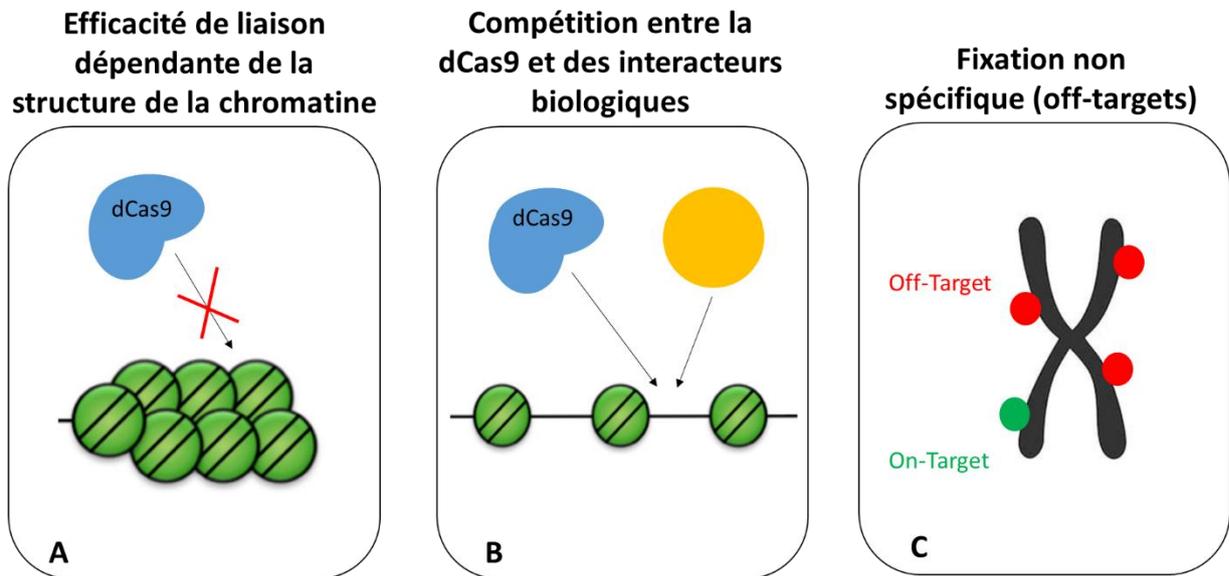


Figure 9 : Représentation schématique des limitations de la technologie CRISPR/Cas9

Par conséquent, lors d'une purification d'un segment de chromatine à l'aide d'un système CRISPR/Cas9, la composition locale de la chromatine pourrait être affectée directement par la présence de la dCas9 au locus.

Une première partie de mon projet de recherche consiste à comprendre l'impact de la protéine dCas9 sur la composition locale de la chromatine lorsqu'elle est liée à un locus donné. Pour cela, je réalise la purification des protéines associées au niveau des télomères dans une lignée cellulaire sauvage, dans une lignée cellulaire contenant la protéine dCas9 non guidée et dans une lignée cellulaire contenant la dCas9 guidée aux télomères.

II. Méthodologie

Une population de cellules HEK293T contenant la protéine dCas9 et une population de cellules HEK293T contenant la dCas9 et un guide télomérique sont établis et validés. Les 2 populations contenant la protéine dCas9 (guidée aux télomères ou non guidée) ainsi que la population de cellules HEK293T sauvages sont amplifiées afin de réaliser une purification des protéines présentes aux télomères par la technique du PICh. Ces purifications sont analysées par western-blot ainsi que par spectrométrie de masse quantitative (Figure 10).

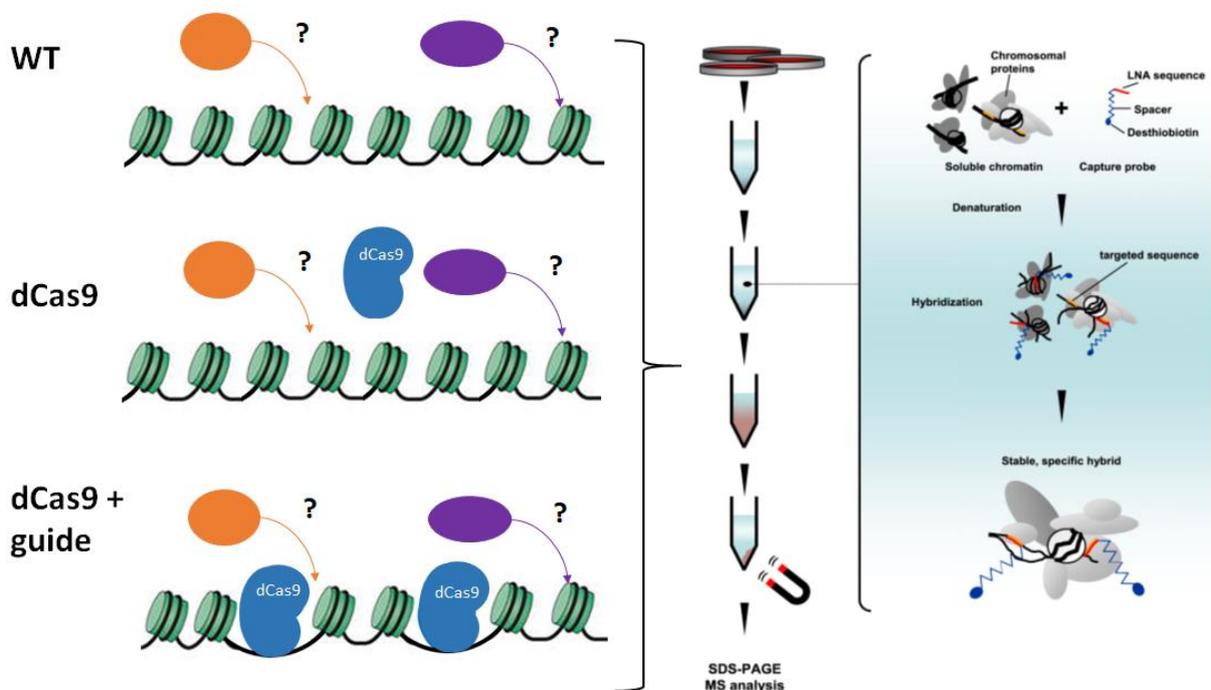


Figure 10 : Représentation schématique de la méthodologie employée afin de comprendre l'impact possible de la protéine dCas9 (en bleu) sur la chromatine au niveau local.

III. Résultats

III.1. Caractérisation des populations sélectionnées

Deux populations de cellules sont préparées :

- Une population de cellules HEK293T contenant la protéine dCas9 (taguée HA)
- Une population de cellules HEK293T contenant la protéine dCas9 (taguée HA) et guidée aux télomères

Pour cela, le plasmide pdCas9-HA (annexe 1) est transfecté dans des cellules HEK293T. Les cellules sont sélectionnées par un antibiotique (la blasticidine) puis validées par westernblot. Un second plasmide contenant le guide télomérique est transfecté dans la population sélectionnée de cellules HEK293T contenant déjà la dCas9. Les cellules exprimant le guide télomérique sont sélectionnées par un second antibiotique (la puromycine) puis validées par westernblot et *in situ* PLA (Proximity Ligation Assay).

La présence de la protéine dCas9 et son bon adressage est vérifié par western-blot et immunofluorescence (*in situ* PLA). Dans la population de cellules contenant la dCas9, sans ou avec guide, le tag HA (dCas9-HA) est détecté (Figure 11A). Les protéines TRF2 (marqueur des télomères) et HA (dCas9-HA) sont co-localisées par *in situ* PLA uniquement dans la population de cellules contenant la dCas9 guidée aux télomères (Figure 11B).

Enfin, les 2 populations (cellules contenant la dCas9 et cellules contenant la dCas9 et le guide) ainsi que les cellules contrôles (WT) ont une taille de télomères comparable. La quantité de protéines associées aux télomères sera donc purifiée dans des proportions équivalentes et pourra donc être comparée entre les 3 populations (Figure 11C).

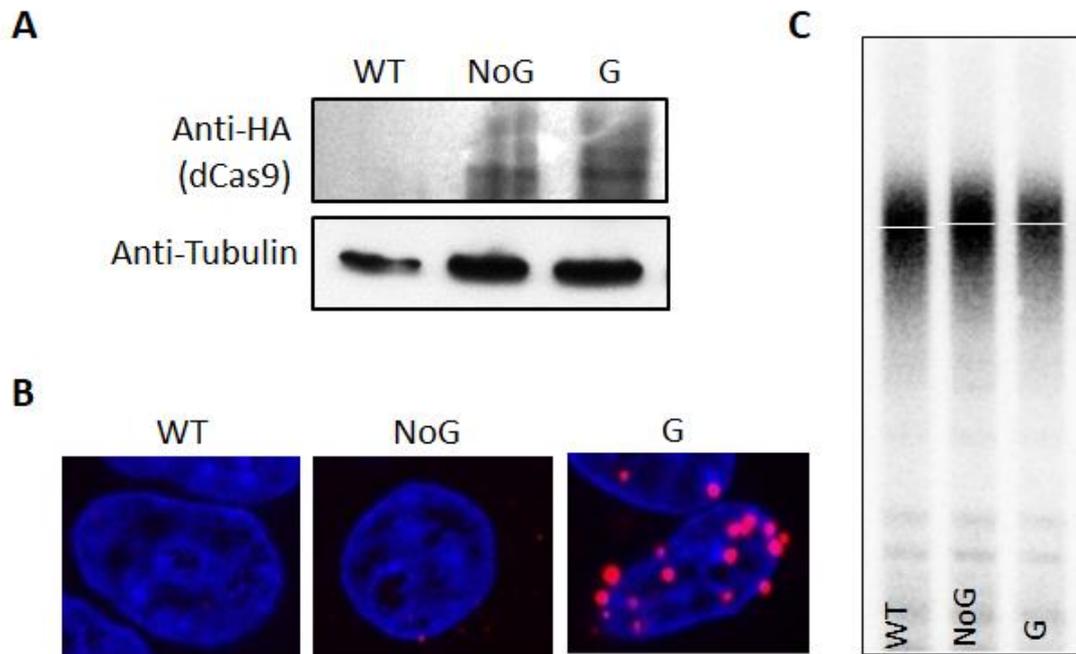


Figure 11 : Vérification des populations de cellules HEK293T contenant la dCas9 ou la dCas9 et le guide télomérique. (A) western-blot avec l’anticorps anti-HA sur protéines totales des 3 populations (HEK293T : WT, HEK293T avec dCas9-HA : NoG et HEK293T avec dCas9-HA et guide télomère : G). (B) co-localisation des protéines TRF2 et dCas9-HA par *in situ* proximity liagtion assay avec les anticorps anti-HA et anti-TRF2. (C) Southern-blot réalisé avec une sonde télomère avec un marquage radioactif sur les 3 populations de cellules (HEK293T : WT, HEK293T avec dCas9-HA : NoG et HEK293T avec dCas9-HA et guide télomère : G) pour la mesure de la taille des télomères.

Les 3 populations de cellules peuvent donc être comparées entre elles. Les cellules HEK293T WT et la population de cellules HEK293T contenant uniquement la dCas9 seront utilisés comme contrôle. La population de cellules contenant la dCas9 et le guide télomérique permettra de déterminer l’impact de la fixation de la protéine dCas9 au niveau d’un locus spécifique (i.e. télomères).

III.2. Purification des protéines associées aux télomères

Les protéines associées aux télomères des 3 populations sont purifiées grâce à la technique du PICh avec des sondes spécifiques du télomères (sondes Fluoro-ARN – annexe 3). Cette expérimentation est réalisée en triplicat.

Le profil des protéines purifiées est vérifié sur gel d'argent ainsi que par western-blot avec les anticorps anti TRF2 (vérification de la capture) et anti-HA (présence de la dCas9-HA).

Le profil des protéines sur le gel d'argent montre la présence de protéines dans l'élution. Le profil des éluions est différent de celui du matériel de départ (input) (Figure 12A). Le western-blot réalisé sur les échantillons de PICh avec l'anticorps anti TRF2 valide l'enrichissement des télomères dans l'élution (Figure 12B). De plus, la protéine dCas9 est retrouvée uniquement dans les éluions de PICh de la population de cellules qui contiennent la protéine dCas9 guidée aux télomères. Ceci démontre la présence de la dCas9 et son bon adressage aux télomères (Figure 12B).

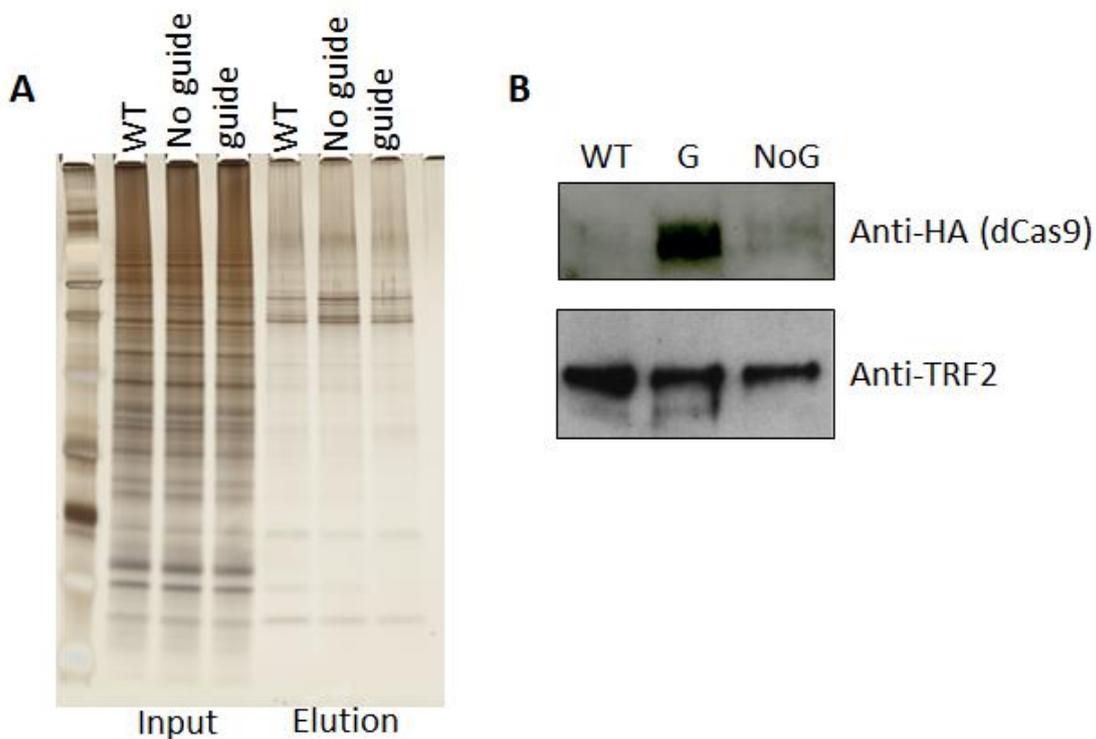


Figure 12 : Vérification de la purification des protéines télomériques par PICh. (A) Gel d'argent du matériel de départ (input) et des éluions du PICh pour les cellules HEK293T

sauvages (WT), la population de cellules contenant la dCas9 (No guide) et la population de cellules contenant la dCas9 et le guide télomérique (guide). (B) westernblot réalisés sur les éluions des PICh des cellules HEK293T sauvages (WT), de la population de cellules contenant la dCas9 et le guide télomérique (G) et de la population de cellules contenant la dCas9 (NoG) avec les anticorps anti-HA (dCas9-HA) et anti TRF2.

III.3. Analyse des protéines associées aux télomères

Les différents PICh sont analysés par spectrométrie de masse quantitative (TMT). L'analyse des données obtenues est réalisée avec le logiciel Perseus.

Les protéines identifiées dans les cellules HEK293T WT sont comparées à celles obtenues dans la population de cellules contenant la dCas9 non guidée. Pour cela, un volcano-plot est réalisé. Le volcano-plot montre une distribution en nuage ce qui démontre qu'il n'y a pas de différence de protéome entre ces 2 conditions (Figure 13A).

Puis, les protéines identifiées dans les cellules HEK293T WT sont comparées aux protéines identifiées dans la population de cellules contenant la dCas9 guidée aux télomères. Le volcano-plot montre une distribution en V et non en nuage ce qui signifie qu'il y a une différence entre les 2 conditions. Les protéines les moins fixées sont principalement des shelterins (complexe de protéines spécifiques du télomère) et des histones, mais ces pertes sont non significatives (Figure 13B).

Enfin, les protéines identifiées dans les populations de cellules contenant la dCas9 guidée ou non guidée sont comparées. L'ensemble des protéines télomériques (en bleu) et une grande majorité des histones (en rouge) sont des protéines dont la quantité est diminuée dans la population de cellules contenant la dCas9 guidée par rapport à la population de cellules contenant la dCas9 non guidée. Toutefois, ces pertes restent non significatives (Figure 13C).

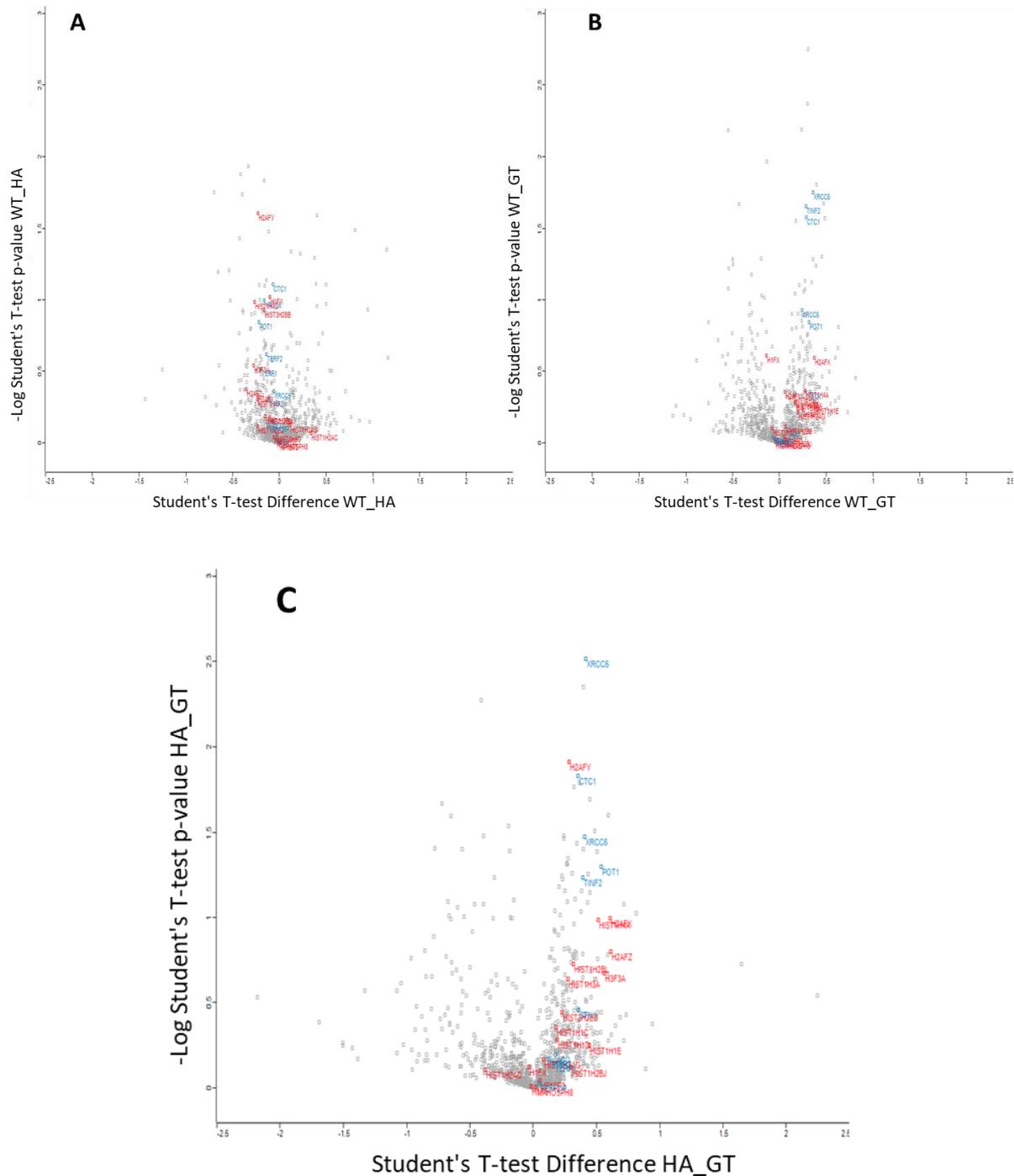


Figure 13 : Résultats de spectrométrie de masse sur des cellules HEK293T contenant la dCas9 non guidée ou guidée aux télomères. (A) Volcano-plot des résultats obtenus par spectrométrie de masse quantitative (TMT) entre la lignée sauvage (WT) et la lignée contenant la dCas9 (HA). (B) Volcano-plot des résultats obtenus entre la lignée sauvage (WT) et la lignée contenant la dCas9 guidée aux télomères (GT). (C) Volcano-plot des résultats obtenus entre la lignée contenant la dCas9 (HA) et la lignée contenant la dCas9 guidée aux télomères (GT). Les protéines d'histones sont marquées en rouge. Les protéines télomériques sont en bleu.

Par la suite, une analyse des groupes de protéines est effectuée par la méthode du “1D enrichment” grâce au logiciel Perseus. Cette analyse regroupe les protéines selon leur rôle biologique. L’algorithme utilise l’abondance absolue des protéines. Ainsi, une comparaison de la distribution de l’intensité d’un groupe de protéines d’un échantillon est réalisée afin de savoir si ce groupe de protéines est significativement enrichi à des fortes ou des faibles valeurs par rapport à l’ensemble des protéines de la catégorie des différents échantillons (44).

Cette analyse a été effectuée entre les protéines purifiées dans les populations de cellules contenant la dCas9 guidée ou non guidée. Différents groupes de protéines de la population de cellules contenant la dCas9 guidée aux télomères sont déplétées de façon significative. Les catégories les plus affectées par la présence de la dCas9 guidée au locus (p-value la plus basse) sont des protéines de l’organisation du télomère (telomere organization) et de la “coiffe télomérique” (telomere cap complex) ainsi que les nucléosomes (Tableau 1).

C: Name	N: Size	N: Score	N: P value	N: Benj. Hoch. FDR	N: Mean	N: Median
DNA geometric change	27	0,393263	0,00048016	0,00605922	0,143445	0,205377
DNA repair	108	0,218241	0,00020111	0,00296078	0,113195	0,121138
DNA replication	64	0,287794	0,00011262	0,00186533	0,142295	0,169726
nuclear chromosome	20	0,444966	0,00064524	0,00531081	0,26645	0,232096
nucleosome	18	0,44607	0,00116231	0,0088834	0,235615	0,252006
telomere cap complex	10	0,636399	0,00052574	0,00468783	0,298281	0,35749
telomere organization	45	0,446938	3,83E-07	8,46E-06	0,207854	0,218118

Tableau 1 : Analyse des groupes de protéines obtenus par spectrométrie de masse sur des cellules HEK293T contenant la dCas9 non guidée ou guidée aux télomères. Analyse des groupes de protéines (1D enrichment) obtenus par spectrométrie de masse quantitative entre la lignée contenant la dCas9 non guidée et la lignée contenant la dCas9 guidée aux télomères.

IV. Discussion et perspectives

Cette première partie de mon projet de recherche consistait à comprendre l’impact de la protéine dCas9 sur la composition locale de la chromatine lorsqu’elle est liée à un locus spécifique. Les données obtenues suite à une purification des protéines du télomère par PICh en présence de la protéine dCas9 faiblement associée au télomère, ont démontré une

diminution de la quantité de certains groupes de protéines caractéristiques du locus par rapport à la lignée cellulaire contenant la dCas9 non guidée. En effet, lors de l'analyse des groupes de protéines, certains groupes caractéristiques du télomère sont statistiquement diminués lorsque la dCas9 est liée à ce locus. De plus, le volcano-plot comparant les protéines obtenues après purification en présence de la dCas9 guidée ou non guidée aux télomères, montre une distribution en V et non en nuage ce qui signifie qu'il y a une différence entre les 2 conditions. Les protéines les moins fixées sont principalement les protéines de l'organisation du télomère et des histones. Toutefois, ces pertes restent non significatives et n'ont donc probablement pas d'effet biologique, expliquant aussi l'absence d'effets sur la taille des télomères.

La protéine dCas9 est une très grosse protéine (>150kDa). Il est donc possible que la liaison de la dCas9 exclue certaines protéines spécifiques du locus. La dCas9 pourrait donc concurrencer la liaison d'interacteurs biologiques endogènes. La diminution des nucléosomes, des protéines de l'organisation du télomère et de la "coiffe télomérique" pourraient donc être liés à une compétition entre ces protéines et dCas9.

Différentes technologies utilisent la dCas9 pour la purification d'un locus unique. Une partie de ces techniques ont été utilisées pour la purification des télomères comme preuve de principe. Ainsi, CAPTURE (22) a permis d'identifier différentes protéines spécifiques du télomère mais 3 shelterins n'ont pu être identifiées (TRF1, TINF2 et TPP1). CLASP (27) a également permis la purification de différentes shelterins mais TRF1, TINF2 et POT1 étaient absents. De même, TRF1, RAP1 et POT1 étaient absents de la purification par la technique CasID (23) alors que TRF2, TINF2 et TPP1 n'étaient pas identifiés par la technique CAPLOCUS (26). L'absence de ces protéines caractéristiques du locus pourraient être dû à un taux d'enrichissement trop faible ou à une modification de la composition locale de la chromatine par compétition de la dCas9 avec ces interacteurs biologiques. Il semble donc que la quantité de dCas9 associée au locus soit critique dans le changement de composition de ce locus. Ce qui pose potentiellement un problème pour l'utilisation de dCas9 : une fixation efficace est nécessaire à une purification efficace par dCas9, mais cette fixation entraîne un changement de composition d'autant plus profond que cette fixation est efficace.

Afin de déterminer si ces modifications sont spécifiques de la présence de la dCas9 et si elles pourraient avoir un effet biologique, des expérimentations avec un contrôle de la quantité de la dCas9 guidée doivent être menées. De plus, une modulation de la quantité de dCas9 permettrait de déterminer si les modifications observées sont liées de façon directe à la présence de la dCas9.

Pour cela, une caractérisation de la chromatine par CHIP (chromatine immunoprécipitation) ou par la technique du PICh pourrait être réalisée au niveau du télomère, en présence de la protéine dCas9 guidée ou non guidée par un ARN guide avec de plus forts niveaux d'expression de la dCas9. De plus, afin de vérifier la spécificité de la réponse en fonction de la quantité de dCas9 présente au locus spécifique, il serait intéressant d'utiliser d'une part, une faible quantité de dCas9 liée au locus et d'autre part, une forte quantité de dCas9 au même locus. Ainsi, un plasmide dont le promoteur permet une modulation de la quantité de dCas9 grâce à l'ajout de doxycycline (promoteur TET ON) pourrait être utilisé (Figure 14).

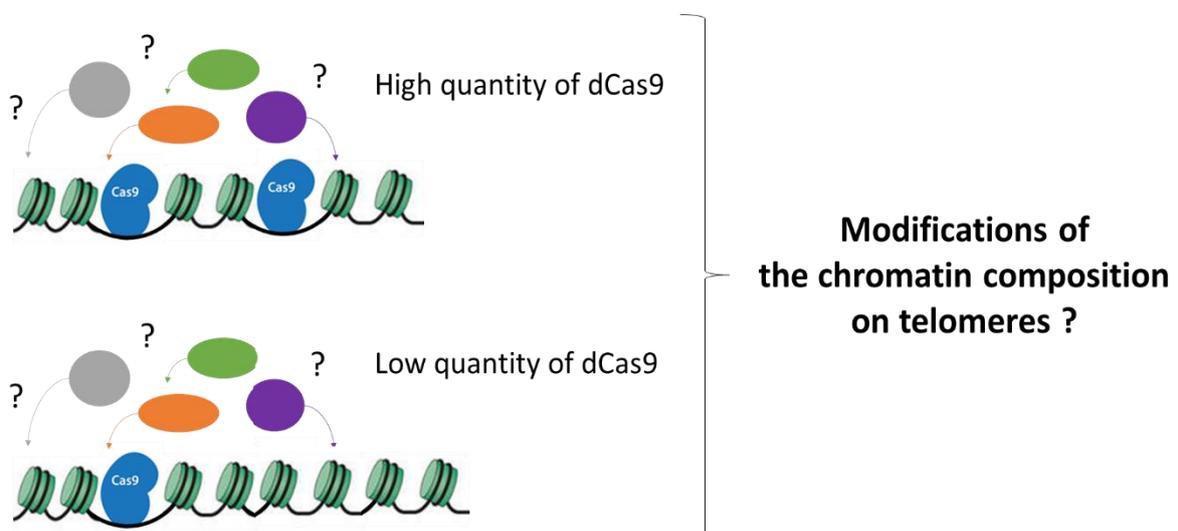


Figure 14 : Représentation schématique de la méthodologie pour la compréhension de la spécificité de l'impact de la protéine dCas9 sur la chromatine au niveau local.

CHAPITRE II : Utilisation de la technique du PICh modifiée pour la purification d'un locus unique

I. Problématique de recherche

La technologie PICh repose sur une seule étape d'affinité entre la desthiobiotine et des billes de streptavidine. Les taux d'enrichissement ne dépassent donc pas les 10 000, ce qui est déjà remarquable pour une unique affinité.

La caractérisation d'un locus unique nécessite un enrichissement bien supérieur à 10 000. Ainsi, pour enrichir un locus unique de 3kpb soit 0.0001% du génome humain, le taux d'enrichissement doit être de 1 million de fois pour obtenir un mélange quasi-homogène (17). Un taux d'enrichissement aussi élevé ne peut pas être obtenu à l'issue d'une seule étape d'affinité, c'est pourquoi deux étapes successives de purification par affinité sont nécessaires.

De plus, les limites de détection des instruments de spectrométrie de masse se situent au niveau de la femtomole. Les peptides, contrairement aux acides nucléiques ne peuvent pas être amplifiés. Le matériel de départ devra donc être produit en quantités importantes pour pouvoir obtenir un signal spécifique suffisamment élevé par rapport au bruit de fond. Enfin, afin de comparer les échantillons entre eux, une analyse par spectrométrie de masse quantitative doit être réalisée.

Les techniques initiées dans différents laboratoires (21–34) ont principalement utilisé un système de purification contenant la dCas9 (22–29). Toutefois, les techniques utilisant la dCas9 ont démontré des limitations liées au système CRISPR/Cas9. Ainsi, la dCas9 peut se lier de façon non significative à d'autres loci (off-targets), la conformation de la chromatine pourrait également moduler l'efficacité de fixation de la dCas9 au locus cible (17) et enfin, la dCas9 semble modifier la composition locale de la chromatine en se fixant au locus. En effet, nous avons démontré que des groupes de protéines spécifiques du locus sont significativement diminués lorsque la dCas9 est liée au locus cible. Par conséquent, l'utilisation d'un système contenant la dCas9 pour la purification d'un locus unique entraîne une faible

identification du signal du fait d'un bruit de fond élevé (off-targets) et affecterait la composition locale de la chromatine.

La deuxième partie de mon projet de recherche consiste au développement d'une stratégie permettant d'identifier les protéines associées à un locus unique. Cette caractérisation non biaisée permettrait une meilleure compréhension de la régulation des gènes. En effet, l'activation ou la répression de l'expression des gènes est contrôlée par les protéines liées aux éléments de régulation, présents en copies uniques.

Pour cela, j'adapte la technique du PICh (capture par sondes spécifiques) en y ajoutant une seconde étape de purification par affinité afin d'obtenir un taux d'enrichissement suffisant (PICh en tandem). De plus, le type de sonde est optimisé ainsi que l'analyse par spectrométrie de masse quantitative afin d'obtenir un signal plus important tout en diminuant les coûts de l'expérimentation.

La mise au point de la technique du PICh en tandem est réalisée sur des locus répétés (péricentromère et télomère). Ces locus ont déjà été caractérisés par la technique du PICh et leur composition est donc connue, ce qui permet d'identifier facilement les protéines spécifiques du locus et donc le bon déroulement du protocole. De plus, la quantité de matériel de départ nécessaire pour la purification d'un locus répété est beaucoup plus faible que pour un locus unique ce qui facilite la mise en œuvre des expériences. Enfin, la comparaison de la technique du PICh avec le PICh en tandem est possible et permet de visualiser l'efficacité de purification et de calculer le taux d'enrichissement de la technique du PICh en tandem.

II. Méthodologie

Des modifications concernant différentes étapes de la technique du PICh sont réalisées. Ainsi, le type de sonde est optimisé afin de faciliter leur synthèse et diminuer leur coût. La spectrométrie de masse quantitative est également modifiée afin d'améliorer la sensibilité de l'analyse et permettre une bonne quantification des protéines identifiées. Ce travail est réalisé en collaboration avec la plateforme de protéomique de Montpellier (Plateforme de

Protéomique Fonctionnelle - FPP). Enfin, une seconde étape de purification par affinité est ajoutée à la technique du PICh afin d'augmenter le taux d'enrichissement (Figure 15).

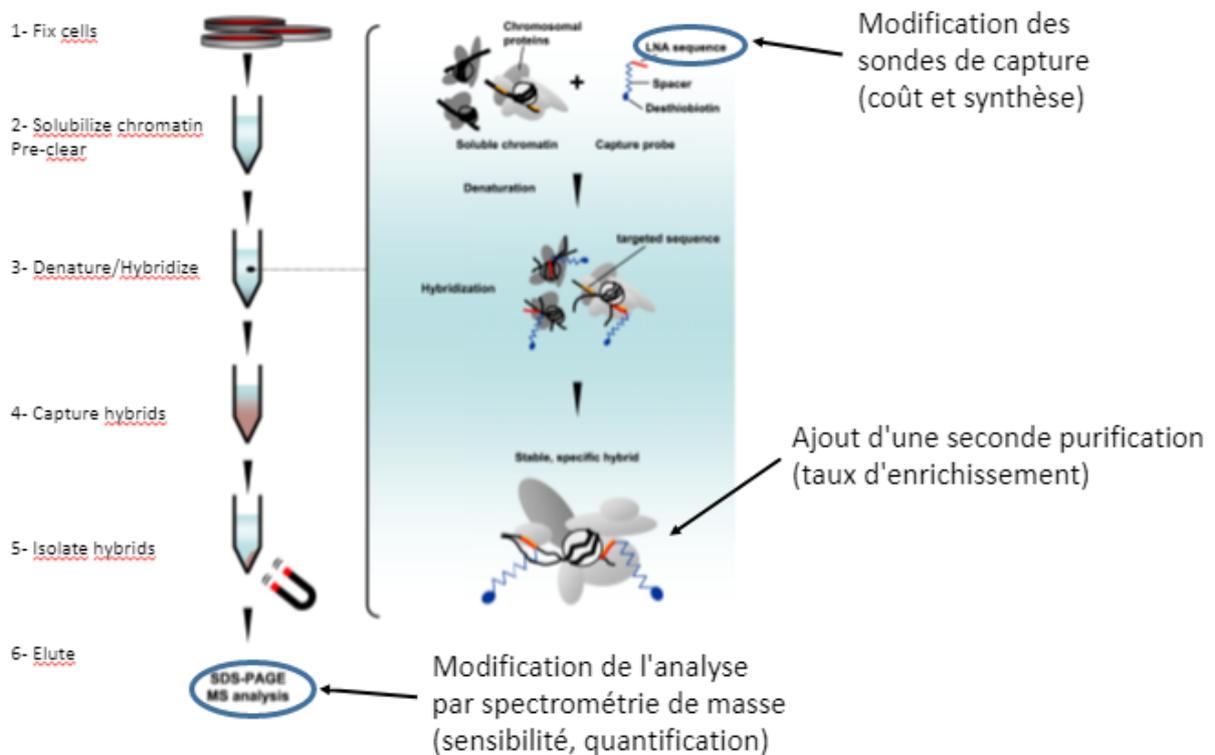


Figure 15 : Représentation schématique de la méthodologie employée pour optimiser la technique du PICh.

III. Résultats

Dans le but d'adapter la technique, utilisée pour la purification de loci répétés, des modifications concernant les sondes, la purification par affinité et l'analyse par spectrométrie de masse ont été réalisées afin d'augmenter la sensibilité et le taux d'enrichissement de la technique qui se situait à environ 10 000 fois. Les modifications de cette technique ont pour but l'enrichissement d'un locus unique soit environ 3kpb (0.0001% du génome humain). Pour cela, le taux d'enrichissement pour obtenir un mélange quasi-homogène doit être de 1 million de fois.

III.1. Modification des sondes de capture

Les sondes qui sont actuellement utilisées pour la technologie PICCh sont des sondes 2'Fluoro-ARN. Ces sondes sont résistantes aux nucléases et ont un coût d'environ 500 euros pour 200 nmol de sonde (sonde de 65nt). Deux autres types de sonde ont été testées. Il s'agit de sondes ADN et de sondes 2'O-Methyl ARN. Les sondes 2'O-Methyl ARN sont stables aux nucléases mais augmentent le Tm des duplex de 1 à 4°C. Leur coût est de 500 euros pour 200 nmol (sonde de 65nt). Les sondes ADN sont stables et ont un coût beaucoup plus faible (environ 80 euros pour une sonde de 65nt).

Ces 3 types de sonde ont été testés sur 2 loci différents, les télomères et les satellites majeurs (les régions péri-centromériques murines). Dans le cas des télomères, une seule sonde de 65 nt est nécessaire à la purification du locus. Pour les satellites majeurs, un mélange de 6 sondes est utilisé pour purifier ce locus (annexe 3). Les sondes sont testées sur de la chromatine de cellules ES de souris, à une concentration de 100µM. Pour la purification des télomères, 4 mg de chromatine ont été utilisés pour chaque test. Pour les satellites majeurs 1 mg de chromatine (environ une boîte de pétri) est utilisé pour chaque purification.

Après avoir réalisé les purifications, les échantillons sont visualisés sur un gel d'argent. Les profils de purification sont comparés à l'échantillon de départ (input) ainsi qu'au matériel non capturé (FT). Les profils d'élution (EL) obtenus sont différents du matériel de départ ainsi que de la fraction non capturée. Les quantités de protéines obtenues à chaque locus pour les différentes sondes semblent assez semblables et les profils identiques. Des western-blot ont été réalisés afin de mieux quantifier les éléments caractéristiques de chaque locus et donc l'efficacité de chaque capture. La shelterin TRF1 a été utilisée pour les télomères et H3K9me3 pour les satellites majeurs. Au niveau des télomères, les sondes ADN ont une efficacité de capture qui semble similaire aux sondes 2'Fluoro ARN. Les sondes 2'O-Methyl ARN semblent avoir une meilleure efficacité. Au niveau des satellites majeurs, les sondes 2'Fluoro ARN et 2'O-Methyl ARN ont une capture équivalente. La sonde ADN a une capture plus faible que les 2 autres à ce locus (Figure 16).

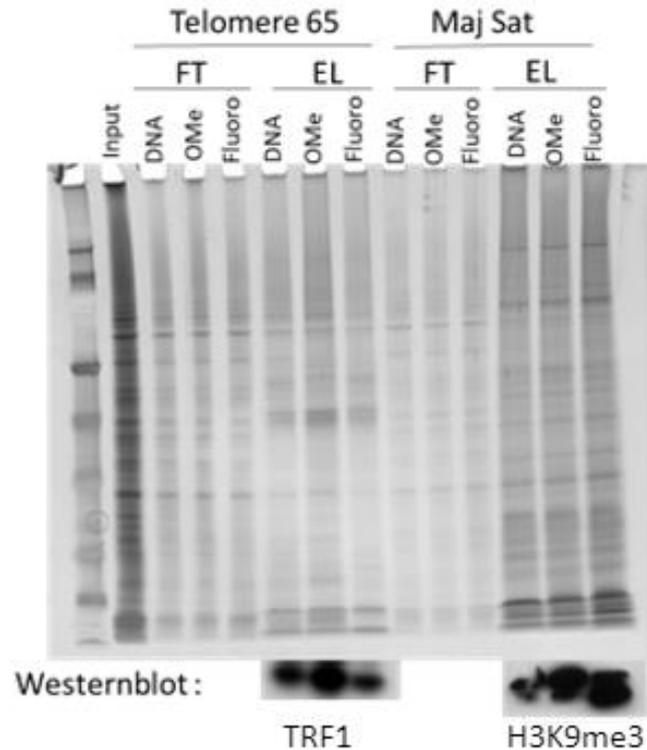


Figure 16 : Test de différentes sondes de capture. Gel d'argent des protéines de départ (input), du matériel non capturé (FT) et des éluions (EL) pour les 2 loci spécifiques (télomère et satellites majeurs (Maj Sat)) pour les 3 types de sondes (ADN, 2'O-Méthyl ARN (OMe) et 2'Fluoro ARN (Fluoro)) et westernblot sur les éluions de PICh pour les 3 sondes avec l'anticorps anti-TRF1 sur les échantillons purifiés aux télomères et avec l'anticorps anti H3K9me3 pour les purifications du satellite majeur.

L'efficacité de capture dépend des sondes mais également du locus à étudier. Les différences observées sont faibles. Comme le coût est beaucoup moins important pour les sondes ADN, nous choisissons d'utiliser ce type de sonde pour la purification d'un locus unique. En effet, afin de couvrir 3kpb, il faudra un grand nombre de sondes de 65nt (environ 45 sondes). Le coût est donc trop important avec les 2 autres types de sonde (environ 22 500 euros pour 45 sondes).

III.2. Augmentation de la sensibilité de l'analyse par spectrométrie de masse

Afin d'étudier un locus unique, l'analyse par spectrométrie de masse doit être très sensible afin d'identifier des protéines en très petite quantité. De plus, afin de comparer des conditions d'expérimentation différentes, l'analyse doit être quantitative.

Pour cela, en collaboration avec la plateforme de protéomique FPP, je réalise différentes améliorations de la technique PICh au niveau de l'analyse par spectrométrie de masse.

III.2.1. Modifications de la préparation des échantillons

Tout d'abord, je teste les conditions de réversion des crosslinks induits par le formaldéhyde. En effet, les échantillons se dégradent durant cette étape ce qui entraîne une diminution du nombre de peptides identifiés. Je réalise ces tests sur des PICh obtenus avec 1 mg de chromatine de cellules ES de souris avec des sondes spécifiques du satellite majeur (Figure 17A). Le nombre de peptides identifiés et comptés plusieurs fois lors de l'analyse par spectrométrie de masse est inférieur lorsque l'étape de réversion est réalisée avec une faible concentration de β -mercaptoethanol (0.1M), par rapport à une forte concentration de β -mercaptoethanol (1M). De plus, à forte concentration en β -mercaptoethanol, le nombre d'identifications et de comptage des peptides diminuent lorsque le temps d'incubation à 99°C augmente. Les conditions permettant un nombre d'identification et de comptage des peptides maximal correspond à une incubation de 5 minutes à 99°C en présence de 1M de β -mercaptoethanol (Figures 17B-C). Il ne s'agit pourtant pas de conditions idéales pour une analyse sur gel (moins de signal, voir figure 17A)

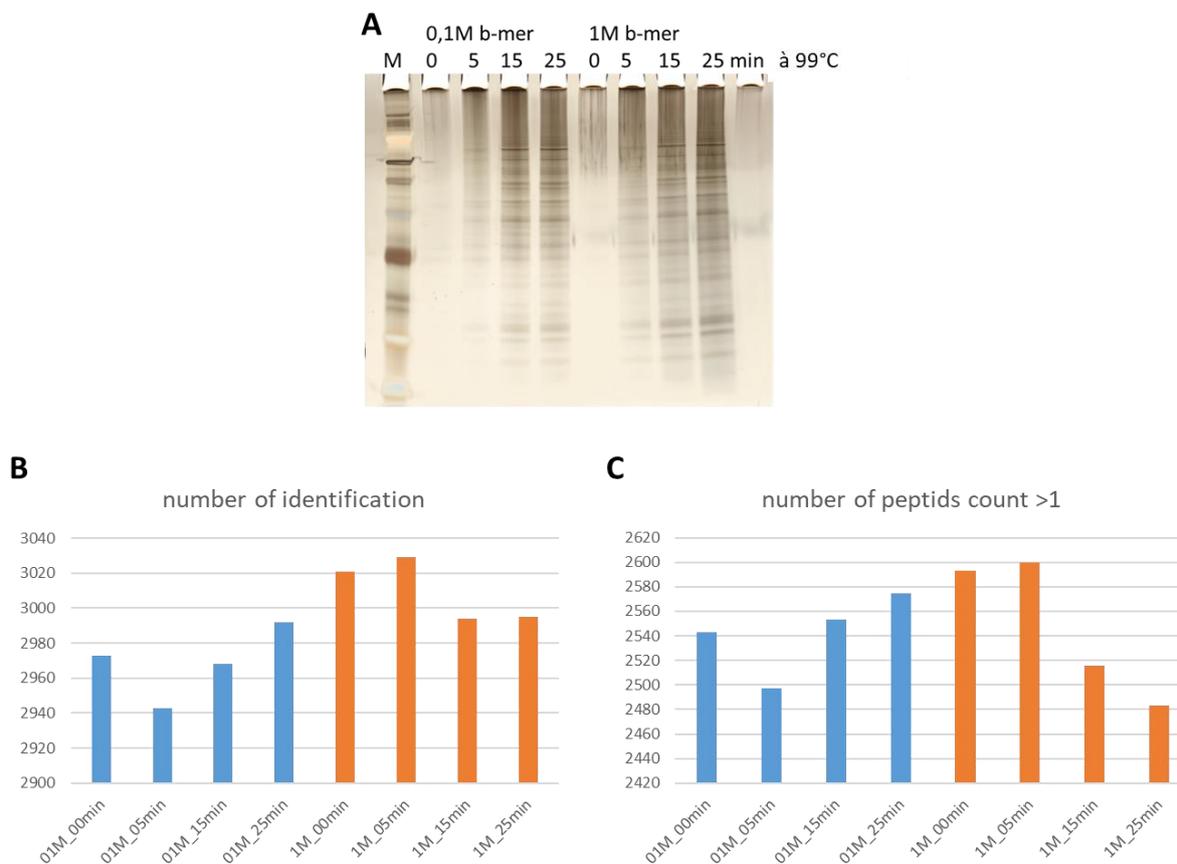


Figure 17 : Amélioration de la préparation des échantillons de spectrométrie de masse (sensibilité). (A) Gel d'argent du matériel de départ (input), des protéines non capturées (FT) et de l'éluion (EL) du PICh des satellites majeurs de cellules ES de souris. M : marqueur. (B) graphique représentant le nombre de protéines identifiées par spectrométrie de masse classique (MSMS) à différentes concentrations de β -mercaptoethanol (0.1M et 1M) et à différents temps d'incubation à 99°C (0, 5, 15 et 25 minutes). (C) graphique représentant le nombre de peptides comptés plus d'une fois lors de l'analyse par spectrométrie de masse classique (MSMS) à différentes concentrations de β -mercaptoethanol (0.1M et 1M) et à différents temps d'incubation à 99°C (0, 5, 15 et 25 minutes).

III.2.2. Modification de la technique de spectrométrie de masse

Une technique de spectrométrie de masse quantitative utilisant un marquage des peptides purifiés est testée. Il s'agit d'un marquage de type isobarique (TMT). Ce marquage est réalisé

après la lyse des protéines par la trypsine, juste avant l'injection du matériel dans le spectromètre. Afin de vérifier la qualité de ce type de marquage, la technique TMT a été comparée à une analyse par spectrométrie de masse classique (MSMS) sur des PICh de satellites majeurs réalisés avec 1 mg de chromatine de cellules ES de souris (Figure 18A). Le nombre total de protéines identifiées par TMT est légèrement supérieur au nombre de protéines totales identifiées par MSMS. De plus, la qualité d'identification par TMT est très supérieure à l'analyse par MSMS. En effet, par MSMS, les protéines sont identifiées à 40% par correspondance avec la base de données du protéome de souris (by matching) alors que 90% des protéines analysées par TMT sont identifiées de façon directe (by MSMS). L'identification des protéines est donc meilleure de façon quantitative mais surtout qualitative par TMT (Figure 18B).

Le mélange obtenu à l'issue d'une purification par PICh est très complexe. Lorsque le mélange est trop complexe, une partie des peptides peut échapper à la détection si plusieurs peptides se trouvent dans la même fenêtre de taille en même temps lors de l'analyse (fenêtre de 1kDa). En diminuant la complexité du mélange par fractionnement, un plus grand nombre de peptides pourrait être identifié. Pour cela, une comparaison d'un PICh TMT avec un PICh TMT fractionné a été réalisée sur des échantillons de PICh de satellite majeur d'ES de souris. Le fractionnement est réalisé après le marquage TMT, sur un gradient d'acétonitrile. Huit fractions sont récoltées. Toutes les fractions sont analysées par spectrométrie de masse quantitative. Le nombre de protéines identifiées est très largement supérieur après fractionnement et 100% des protéines sont identifiées par MSMS (Figure 18B).

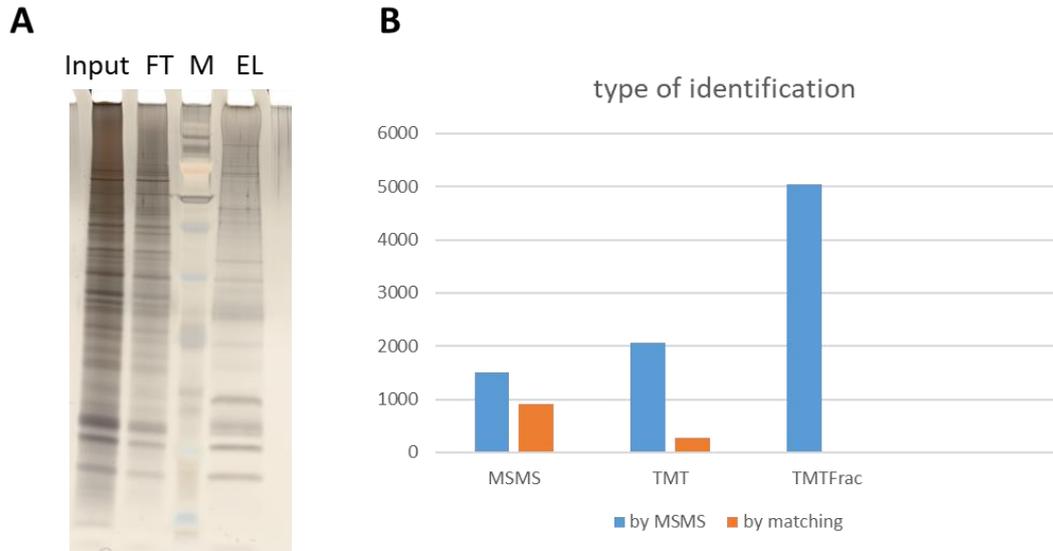


Figure 18 : Amélioration de la profondeur d'analyse de spectrométrie de masse (sensibilité). (A) Gel d'argent du matériel de départ (input), des protéines non capturées (FT) et de l'élution (EL) du PICh des satellites majeurs de cellules ES de souris. M : marqueur. (B) graphique représentant le nombre de protéines identifiées de façon directe (by MSMS) ou par comparaison avec la base de données du protéome de souris (by matching) par spectrométrie de masse classique (MSMS), par spectrométrie de masse quantitative (TMT) ou par spectrométrie de masse quantitative avec fractionnement sur un gradient d'acétonitrile (TMT Frac).

L'analyse de spectrométrie de masse quantitative la plus sensible est donc une analyse après marquage TMT et fractionnement des protéines sur un gradient d'acétonitrile.

III.3. Ajout d'une seconde étape de purification

Afin d'obtenir des taux d'enrichissement de l'ordre du million de fois, une seconde étape de purification par affinité est nécessaire après la première élution. La purification par la digoxygénine (DIG) est choisie comme deuxième étape car l'affinité de la DIG avec son anticorps est très élevée ($K_d \sim 10^{-9}M$) et permet donc d'obtenir un fort taux d'enrichissement. La DIG est ajoutée aux sondes ADN en 5'.

Pour la réalisation des tests, une sonde ADN spécifique des télomères de 65 nt contenant la DTB est synthétisée avec un espaceur amino-modifié C12 en 5'. Le espaceur amino modifié C12 contient un groupement réactif de type amine primaire -NH₂. Ce groupement peut réagir avec un substrat NHS-ester. Les sondes sont préparées au laboratoire par ajout de la DIG-NHS-ester sur le C12 (annexes 3 et 4).

Dans un premier temps, je réalise une purification des télomères de cellules ES de souris avec la sonde avant l'ajout de la DIG afin de vérifier que le C12 ne masque pas la DTB et que la première étape de purification sur billes de streptavidine n'est pas modifiée par le design de la sonde. Cette purification est comparée à une purification par la DTB sur billes de streptavidine avec une sonde ADN ne contenant pas le C12. Afin de visualiser l'efficacité de capture, des western-blots sont réalisés avec la shelterin TRF1 sur les produits de purification (sans élution). La quantité de TRF1 enrichie est équivalente avec ces 2 sondes (Figure 19A). Le C12 ne perturbe donc pas la capture de la DTB par les billes de streptavidine.

Dans un second temps, je réalise des captures par la DIG avec un anticorps anti DIG et des billes magnétiques de protéine A. Cette capture est comparée à une purification à la DTB sur billes de streptavidine. Le but de cette expérimentation est de vérifier l'efficacité de purification par la DIG. Les échantillons sont visualisés par western-blot avec un anticorps anti TRF1. Une quantité équivalente de shelterin TRF1 est retrouvée dans les 2 purifications (Figure 19A). L'efficacité de capture par la DIG est donc comparable à celle obtenue par la DTB.

Enfin, un essai de double purification (PICH en tandem : DTB - DIG) est réalisé et comparé à une capture uniquement par la DTB. La quantité de TRF1 visualisée à l'issue d'un western blot sur les produits de purification est comparable entre les 2 types de purification. La perte de protéines est donc très faible entre une purification DTB et une purification en tandem DTB et DIG (Figure 19B). Les profils de ces 2 purifications sont visualisés sur un gel d'argent (Figure 19C). On observe un plus grand nombre de bandes dans la purification DTB par rapport à la purification DTB et DIG. Ceci démontre une diminution du bruit de fond et donc une augmentation de l'enrichissement puisque la quantité de protéines spécifiques du loci (i.e. TRF1) ne sont pas diminuées dans le western-blot. Toutefois, de fortes bandes à 150, 50 et 25kDa sont observées. Ces bandes semblent correspondre à l'anticorps anti-DIG (2 chaînes lourdes de 50kDa et 2 chaînes légères de 25kDa soit 150kDa au total). En effet, les billes ont

été incubées à 99°C dans la solution de reverse crosslink avant d’être déposées sur gel ce qui a entraîné le détachement de l’anticorps en même temps que les protéines capturées.

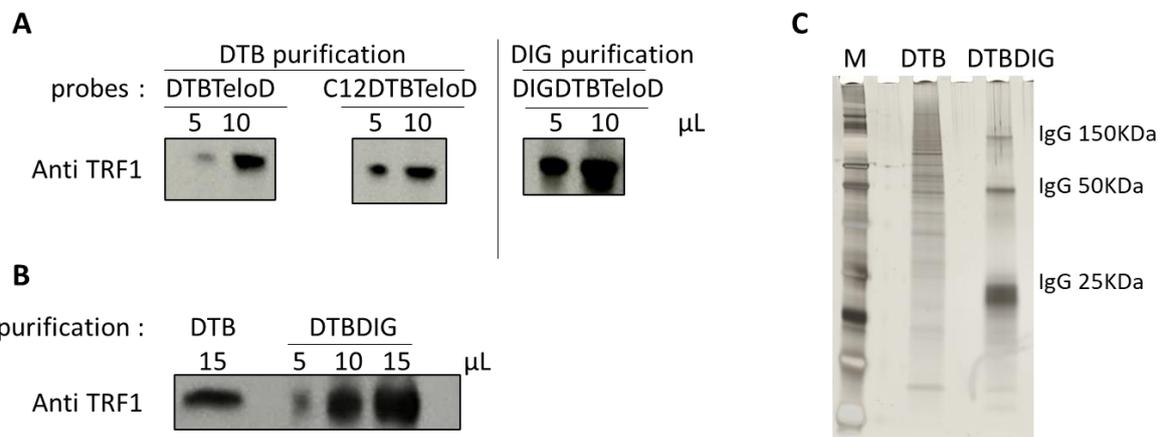


Figure 19 : Vérification des étapes de purification en tandem sur un locus spécifique, le télomère. (A) western-blots réalisés sur les protéines purifiées par la technique du PICh (purification par la desthiobiotine - DTB purification) ou par purification avec la digoxygénine (DIG purification) avec l’anticorps anti-TRF1 (i.e. télomères). Purifications réalisées avec une sonde ADN contenant uniquement la desthiobiotine (DTBTeloD), contenant la desthiobiotine et un espaceur amino C12 (C12DTBTeloD) ou contenant la digoxygénine et la desthiobiotine (DIGDTBTeloD). A l’issue de la capture et des lavages, les billes contenant les protéines capturées sont placées dans le tampon de reverse crosslink à 99°C pendant 25 minutes avant d’être déposés sur gel. (B) westernblot avec l’anticorps anti-TRF1 sur les échantillons purifiés avec une sonde ADN soit par le PICh simple (DTB) soit par le PICh en tandem (DTBDIG). A l’issue de la capture et des lavages, les échantillons sont placés dans le tampon de reverse crosslink à 99°C pendant 25 minutes avant d’être déposés sur gel. (C) gel d’argent pour la visualisation des profils de purification d’un PICh simple (DTB) et d’un PICh en tandem (DTBDIG) avec des sondes ADN. M : marqueur.

L’addition d’une étape de purification par la DIG à l’issue d’une purification par la DTB est donc possible. La perte de protéines spécifique (TRF1) au long de ces 2 étapes successives de

purification est très faible. Le taux d'enrichissement pourrait donc être très significativement augmenté.

III.4. Elution de la seconde étape de purification

L'affinité de la DIG avec l'anticorps anti-DIG est très élevée mais cette dernière ne peut pas être détachée de son anticorps par des tampons d'élution classiques.

Un système permettant une élution "douce" doit donc être mis en place afin d'obtenir une élution sans contaminants issus d'une interaction avec les billes ou l'anticorps sur les billes afin de diminuer le bruit de fond lors de l'analyse par spectrométrie de masse. En effet, si les conditions d'élution de la DIG sont trop drastiques, l'anticorps anti-DIG pourrait se retrouver dans l'éluat ainsi que des protéines capturées de façon non spécifique par les billes.

Afin de réaliser cette élution, nous choisissons donc d'utiliser un espaceur photo-clivable situé entre la DTB et la DIG afin de libérer la sonde et la chromatine des billes. Cet espaceur se clive lorsqu'il est exposé aux UV à une longueur d'onde de 360nm (il s'agit d'un rayonnement UV de faible énergie ne modifiant pas les acides nucléiques ou les protéines).

Comme les billes magnétiques sont opaques, elles ne laissent pas passer les UV facilement, le temps d'élution est donc largement rallongé par rapport aux recommandations d'utilisation d'un espaceur photoclivable. L'élution est réalisée sous agitation afin que le maximum de faces des billes soit exposé aux UV pour maximiser l'élution. L'élution de la DIG sera réalisée pendant 2 H sous rotation sous UV (365nm).

Afin de mesurer le taux d'enrichissement et le bruit de fond de la technique du PICh en tandem (DTBDIG) avec une élution aux UV, un PICh classique (DTB) avec une sonde ADN est réalisé sur les télomères de cellules ES de souris et comparé à un PICh en tandem avec une sonde ADN sur la même chromatine. Les profils de purification sont visualisés sur gel d'argent (Figure 20). Le profil obtenu avec le PICh en tandem contient beaucoup moins de bandes ce qui suggère que le bruit de fond est fortement diminué avec cette technique par rapport au PICh classique.

De plus, les bandes caractéristiques de l'anticorps utilisé pour la capture ne se retrouvent plus dans l'élution ce qui démontre le bon clivage de l'espaceur. La capacité d'élution est évaluée par un western-blot avec un anticorps anti TRF1 sur les 2 purifications (Figure 20). On observe une diminution des protéines spécifiques du locus (i.e. TRF1) ce qui indique que l'élution n'est pas complète. Néanmoins, les échantillons sont envoyés pour une analyse par spectrométrie de masse (MSMS) afin d'identifier les protéines purifiées. Les protéines identifiées lors du PICh sont les protéines caractéristiques du télomère. Les 6 shelterins sont parmi les 10 protéines qui ont le plus de comptages MS. Concernant le PICh en tandem, aucune shelterin n'a pu être identifiée. La quantité de protéines éluées est donc probablement trop faible pour permettre une identification des peptides (sous les limites de sensibilité) ou les protéines sont modifiées chimiquement après la première élution, ce qui les rendrait incompatibles avec une analyse par spectrométrie de masse.

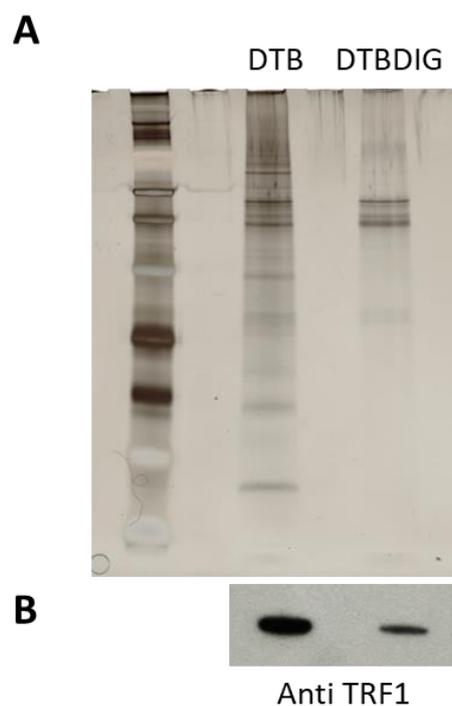


Figure 20 : Vérification de l'élution aux UV de la purification en tandem sur un locus spécifique, le télomère. (A) gel d'argent pour la visualisation des profils de purification d'un PICh simple (DTB) et d'un PICh en tandem élué aux UV (DTBDIG) avec des sondes ADN. M : marqueur. (B) westernblot réalisé sur les protéines purifiées par la technique du PICh simple (DTB) et d'un PICh en tandem élué aux UV (DTBDIG) avec des sondes ADN avec l'anticorps anti-TRF1.

Afin de tester l'efficacité de la purification en tandem et vérifier si l'absence de protéines dans l'analyse par spectrométrie de masse est due à une faible élution par les UV, un PICh en tandem sur les télomères est réalisé en faisant bouillir les billes dans le tampon de reverse crosslink afin de détacher l'ensemble des protéines présentes sur les billes. Ce PICh en tandem est comparé à un PICh simple. La sonde utilisée est identique (sonde ADN de 65nt contenant la DTB et la DIG) ainsi que la quantité de chromatine (2 mg de chromatine de cellules ES de souris). Les profils observés sur gel d'argent montrent des profils différents entre les 2 purifications. Le nombre de bandes est diminué dans la purification en tandem mais des bandes fortes sont observées. Celles-ci correspondent probablement aux anticorps qui ont été détachés des billes par la haute température de l'élution (Figure 21A). Des western-blots sont réalisés avec les anticorps anti TRF1 et RAP1 afin de visualiser l'efficacité des 2 purifications. Les intensités sont identiques pour TRF1 entre les 2 types de PICh ce qui démontre que le PICh en tandem n'engendre pas de perte de matériel. Toutefois, un changement de taille est observé pour les protéines TRF1 et RAP1 (Figure 21B). Les échantillons sont envoyés pour une analyse par spectrométrie de masse (MS/MS).

Les protéines identifiées lors du PICh simple sont les protéines caractéristiques du télomère. Les 6 shelterins sont parmi les 10 protéines qui ont le plus de spectres MS. Concernant le PICh en tandem, seuls TRF1 et TINF2 (TRF1-interacting nuclear factor 2) ont été identifiées mais avec un nombre de comptage MS extrêmement faible et avec une identification indirecte par matching. De façon surprenante, aucune immunoglobuline n'est identifiée. L'analyse par spectrométrie de masse n'est donc pas en adéquation avec les résultats obtenus par gel d'argent ou par western-blot. Il est donc probable que les protéines aient subi une modification chimique très profonde lors d'une étape de la purification en tandem qui n'est pas compatible avec leur identification par spectrométrie de masse. Cette hypothèse est renforcée par la présence d'un switch de taille observé dans les western-blots (Figure 21B).

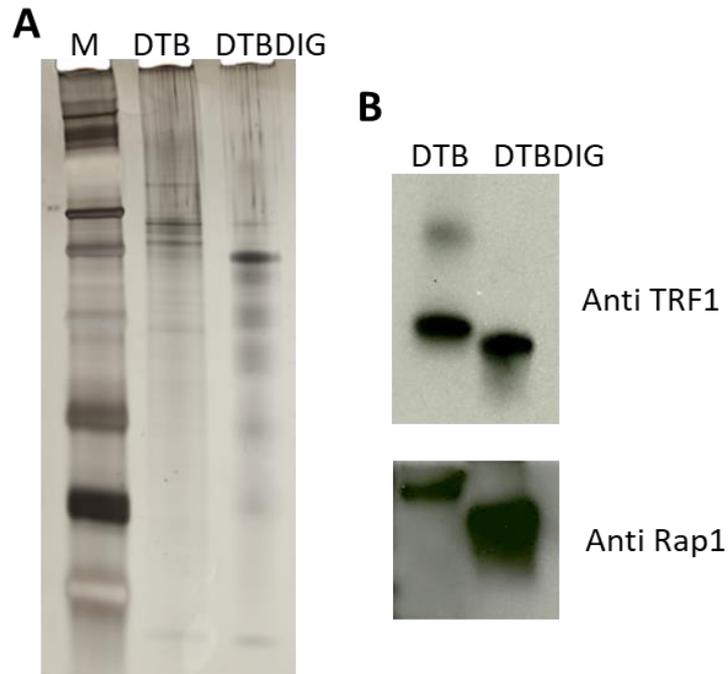


Figure 21 : Comparaison du PICh et du PICh en tandem sur un locus spécifique, le télomère. (A) gel d'argent pour la visualisation des profils de purification d'un PICh simple (DTB) et d'un PICh en tandem (DTBDIG) avec des sondes ADN. M : marqueur. (B) westernblots réalisés sur les protéines purifiées par la technique du PICh simple (DTB) et d'un PICh en tandem (DTBDIG) avec des sondes ADN avec les anticorps anti TRF1 et Rap1.

IV. Discussion

Cette seconde partie de mon projet de recherche consistait au développement d'une technique permettant la caractérisation des protéines associées à un locus spécifique et pouvant être utilisée pour la purification des protéines associées au niveau d'un locus unique. Pour cela, j'ai adapté la technique du PICh (capture par sondes spécifiques) en y ajoutant une seconde étape de purification par affinité afin d'obtenir un taux d'enrichissement supérieur au PICh simple (PICh en tandem). De plus, le type de sonde ainsi que l'analyse par spectrométrie de masse quantitative ont été modifiés afin d'augmenter la sensibilité de l'analyse tout en diminuant les coûts de l'expérimentation.

Trois types de sondes de capture ont été testés sur deux locus (péricentromère et télomère). Les 3 types de sonde (DNA, O-methyl RNA et 2' Fluoro RNA) ont permis la capture des protéines associées au locus spécifique. Néanmoins, des petites variations ont été observées en fonction de la sonde mais aussi du locus à purifier. Toutefois, le coût est beaucoup moins important pour les sondes ADN. En effet, une sonde ADN de 65nt a un coût de 80 euros alors que les 2 autres types de sonde ont un coût de 500 euros pour la même taille de sonde. Pour cette raison, les sondes de types ADN seront préférées.

La préparation des échantillons a également été optimisée afin de maximiser le nombre de protéines identifiées par spectrométrie de masse. Pour cela, différents temps d'incubation à 99°C et différentes solutions de reverse crosslink ont été testés. Lorsque les protéines sont incubées un temps court à 99°C (maximum 5 minutes), dans la solution de reverse crosslink, le nombre de protéines identifiées est augmenté. De plus, la concentration en β -mercaptoéthanol influence le nombre d'identifications. Ainsi, lorsque la solution de reverse crosslink contient 1M de β -mercaptoéthanol, le nombre d'identification est supérieur par rapport à une solution de reverse crosslink contenant 0.1M de β -mercaptoéthanol. Ces modifications ont donc permis de diminuer le nombre de protéines non identifiées lors de l'analyse par spectrométrie de masse en limitant la dégradation des protéines lors de l'étape de reverse crosslink.

Enfin, différents types de spectrométrie de masse ont été testés. Une purification des protéines associées au niveau des péricentromères a été réalisée. Cette purification a été analysée par spectrométrie MS/MS, par spectrométrie MS/MS après marquage au TMT et par spectrométrie MS/MS après marquage au TMT et fractionnement des peptides sur un gradient d'acétonitrile. Le nombre de peptides identifiés ainsi que la qualité de leur identification est supérieure lorsque les peptides sont marqués au TMT. De plus, le nombre d'identification est fortement augmenté (2 à 3 fois supérieur) après fractionnement de ces peptides sur un gradient d'acétonitrile. Leur identification est également de meilleure qualité (uniquement par MS/MS). Le marquage des peptides par TMT a également l'avantage de permettre une quantification statistique des peptides lorsque l'expérimentation est réalisée en triplicat. Enfin, les échantillons marqués au TMT peuvent être multiplexés pour pouvoir comparer de multiples échantillons entre eux. Le multiplexage a pour intérêt de diminuer les variations expérimentales puisque tous les échantillons sont analysés en même temps dans le

spectromètre de masse. De ce fait, le coût de l'analyse est fortement diminué puisqu'une seule injection des échantillons dans le spectromètre de masse est nécessaire.

Ainsi, ces modifications ont permis de diminuer le coût de l'expérimentation et d'augmenter sa sensibilité. Néanmoins, le taux d'enrichissement n'est pas ou peu augmenté. Pour modifier ce taux d'enrichissement, une seconde étape de purification a été ajoutée.

La digoxigénine (DIG) a été choisie pour la seconde étape de purification. La capacité de capture de la DIG a été testée. Elle se révèle équivalente à une capture par la DTB avec des billes de streptavidine. L'efficacité de capture d'une double purification (DTB et DIG) a également été mesurée et comparée à une capture uniquement de la DTB par la streptavidine. De nouveau, la capture est équivalente. La DIG permet donc la capture des protéines associées à un locus spécifique sans entraîner de perte significative.

Néanmoins, la DIG ne peut être éluée de son anticorps avec des tampons classiques d'éluion. Un espaceur photo-clivable est donc utilisé. Lors d'une exposition des échantillons aux UV à 365nm, l'espaceur se clive et libère la sonde et la chromatine des billes alors que la DIG, l'anticorps anti-DIG et toutes protéines non spécifiquement liées aux billes restent attachés. La chromatine spécifique atteint donc un niveau de pureté largement supérieur à ce l'on obtient après une simple purification. Il me faudra encore mesurer cet enrichissement pour vérifier s'il atteint le million de fois, requis pour procéder à une purification de locus unique.

Le nombre de pics observés en spectrométrie de masse dans le PICH en tandem est extrêmement faible. Il semblerait donc que les protéines ne sont plus présentes dans l'échantillons au moment de l'analyse. Les protéines sont donc modifiées de façon drastique avant la spectrométrie de masse ce qui rend les échantillons incompatibles avec ce type d'analyse. Une modification chimique est donc supposée. Cette modification devrait être liée à la seconde étape du PICH en tandem puisque les protéines du PICH simple sont identifiables. La réaction chimique entraînant la modification des protéines semble donc liée à la digoxigénine.

En effet, la DIG est ajoutée sur la sonde par réaction d'un NH₂ libre avec une DIG-NHS ester. La présence, même résiduelle de cet ester, utilisé en très vaste excès lors de la réaction de

couplage, pourrait réagir lors de l'étape de reverse crosslink. En effet, la réaction de carboimidation est favorisée par un pH basique et la température. Comme le reverse crosslink se déroule à 99°C à pH 8.8, il se pourrait que le DIG-NHS ester se conjugue aux amines primaires des protéines de l'échantillon : donc en N-terminal et sur toutes les lysines. Afin de réaliser une analyse par spectrométrie de masse, les protéines doivent être lysées en peptides par la trypsine. Comme la trypsine clive les polypeptides après les résidus lysine et arginine, une modification par NHS-ester rendrait cette digestion impossible. De ce fait, un nombre très limité de protéines sont analysables par spectrométrie de masse. Cette hypothèse est renforcée par l'observation d'un changement des propriétés de migration des protéines dans les gels acrylamide dénaturants. La migration plus rapide des protéines pourrait s'expliquer par la modification de la charge des protéines engendrée par cette réaction. De plus, la quantité d'anticorps et de billes utilisées pour une capture complète est en vaste excès pour favoriser une capture totale. Ceci pourrait expliquer la présence de la DIG-NHS ester toujours présente lors des captures. La DIG-NHS ester associée à l'anticorps pourrait donc réagir avec les protéines capturées mais également avoir un impact sur l'efficacité d'élution aux UV. En effet, le temps de réaction nécessaire à une élution totale est probablement allongé car le volume de billes est très élevé.

Afin de pallier ce problème une purification de la sonde doit être réalisée après le couplage de la DIG. Cette purification permettrait d'éliminer la DIG-NHS ester en excès ce qui diminuerait la quantité d'esters présents lors des étapes de purification de la chromatine. De plus, le temps d'exposition aux UV devra être optimisé afin d'augmenter significativement l'efficacité d'élution.

V. Perspectives

Après la mise au point de cette technique, une preuve de principe devra être réalisée sur un locus spécifique afin de confirmer que les taux d'enrichissement obtenus sont suffisants pour cibler un locus unique par la technique du PICh en tandem. Pour cela, un locus unique sera mimé en réalisant une purification des télomères avec une quantité de matériel de départ

diminuée d'un facteur 100, mélangé à un vaste excès de chromatine extraite de cellules de drosophile (qui n'ont pas de séquences TTAGGG aux télomères). En effet, les télomères représentent 0.01% du génome alors qu'un locus unique représente 0.0001% du génome. Dans cette expérimentation, l'abondance relative de la cible sera donc équivalente à celle d'un locus unique d'environ 3kpb dans un génome mammifère. De plus, la technique du PICh permet la purification de l'ensemble des éléments constitutifs des télomères (e.g. shelterins) pour une quantité de matériel de départ donnée qui est fonction de la taille des télomères à caractériser et de l'espèce à étudier. En diminuant d'un facteur 100 le matériel de départ nécessaire à cette purification, le signal obtenu par la technique du PICh sera au niveau du bruit de fond. En revanche, la technique du PICh en tandem devrait permettre d'identifier les éléments constitutifs de ce locus à partir d'une très faible quantité de matériel de départ. Cette expérimentation aura pour but la validation d'une purification et d'une identification de protéines à partir d'une très petite quantité de matériel de départ grâce à l'augmentation du taux d'enrichissement et à la diminution du bruit de fond. Cette expérimentation permettra également de valider les limites de détection liées aux instruments de spectrométrie de masse.

Pour cela, un PICh sera réalisé avec une sonde ADN de 65nt spécifique des télomères à partir de 50µg de chromatine de cellules ES de souris mélangés à 5 mg de chromatine de drosophile ainsi qu'un PICh en tandem avec également une sonde ADN de 65nt spécifique des télomères et 50µg de chromatine de cellules ES de souris mélangés à 5 mg de chromatine de drosophile. L'expérimentation sera réalisée en triplicat afin d'être quantifiée.

Par la suite, une preuve de concept doit être réalisée pour un locus unique. Nous choisissons de la réaliser sur le promoteur du gène de la télomérase (TERT). En effet, TERT est très fréquemment réactivé dans le cancer. Comprendre les mécanismes de réactivation de ce gène a donc un fort intérêt. De plus, les protéines identifiées au niveau du promoteur du gène réactivé pourraient être des cibles potentielles en thérapie.

TERT a un promoteur unique entouré d'îlots CpG (-1800 à +2300 par rapport à l'ATG). Les îlots CpG sont des régions à forte densité en Cytosine et Guanine qui peuvent être méthylés sur le C. Généralement, leur méthylation entraîne une inactivation de l'expression génique.

Toutefois, le promoteur de TERT n'est pas méthylé dans les cellules humaines normales. Différents résultats suggèrent qu'une forme non méthylée du promoteur TERT favoriserait la fixation de facteurs répresseurs. Le corps du promoteur de TERT se situe de -330 à +37 par rapport à l'ATG et contient des sites de fixation de facteurs de transcription mais ne contient pas de boîtes TATA ou CAAT. Ce promoteur contient également des sites de fixation d'éléments répresseurs de la transcription (45,46).

Dans les cellules cancéreuses, la réactivation du gène de la télomérase est fréquemment liée à une mutation au niveau du promoteur du gène. Deux mutations ponctuelles au niveau du promoteur de TERT sont retrouvées dans la plupart des cas. Ces mutations C > T sont observées en position -124 et -146 pb en amont du codon ATG de TERT et sont généralement notées C228T et C250T respectivement. L'observation de la fréquence de ces mutations est très élevée et pourrait donc suggérer qu'elles jouent un rôle important dans différentes étapes de la tumorigénèse (46). Ces mutations sont associées à une augmentation de la protéine TERT et une augmentation de l'activité de la télomérase.

Généralement un seul allèle du promoteur de TERT est muté. L'allèle muté présente ainsi une di- ou une tri-méthylation de la lysine 4 de l'histone 3 (chromatine active) alors que l'allèle non muté présente une tri-méthylation de la lysine 27 de l'histone 3 (chromatine répressive). Seul l'allèle muté est actif au niveau de la transcription de TERT. Un seul nucléotide muté entraîne donc un changement épigénétique important et une expression mono-allélique (47).

De plus, ces 2 mutations créent un motif de fixation des facteurs de transcription ce qui permet la transcription du gène *TERT* (Figure 22A). Une moyenne de 27% des tumeurs présentent une mutation au niveau du promoteur de TERT ce qui en fait la mutation la plus fréquente observée dans une région non codante dans les cancers humains. Le glioblastome, le mélanome et le cancer urothélial présentent une fréquence de mutation très élevée (Figure 22B) (45).

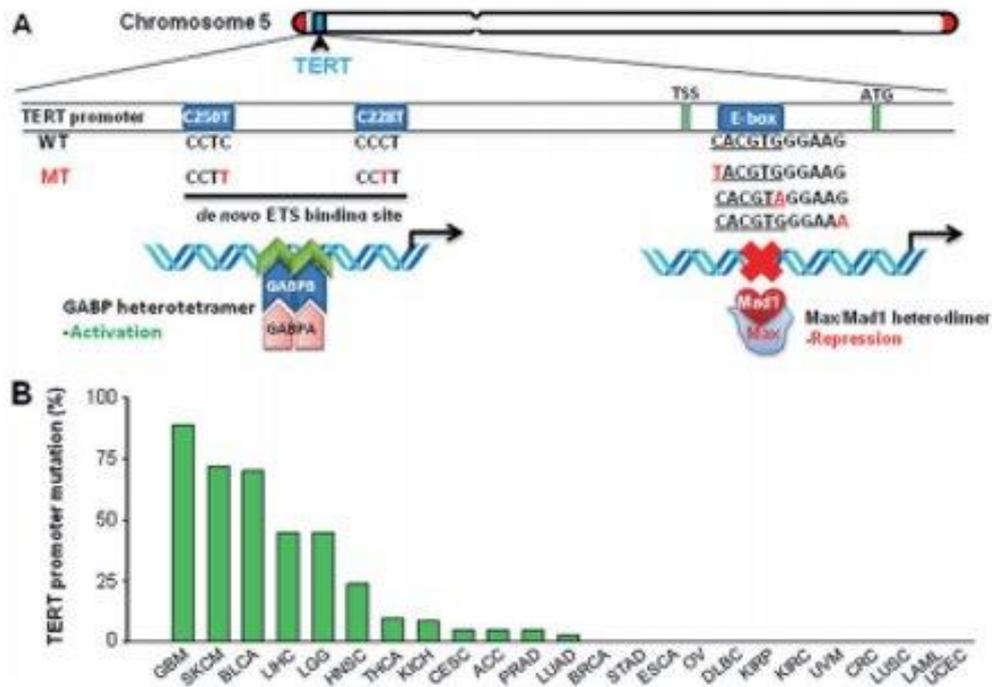


Figure 22 : Structure et fréquence de mutation du promoteur du gène *TERT*. A. Schematic presentation of *TERT* promoter mutations and relevant transcription factors. The *TERT* gene at chromosome 5p and its promoter. C > T mutation occurs at one of both positions of the *TERT* proximal promoter (C228T and C250T) in malignant cells, which create de novo ETS binding motifs. The ETS family members GABPA and GABPB form heterotetramers that bind to the de novo ETS site and activate *TERT* transcription. The E-box (CACGTG) sequence mutation may lead to the dissociation of the repressor MAX/Mad1 complex from E-box, thereby de-repressing the *TERT* gene. B. The frequency of *TERT* promoter mutations in a panel cancer types from the TCGA dataset analyses. D'après Yuan X et al. *Oncogene* (2019) 38(34);6172–6183

C'est pour cela que nous avons choisi de réaliser la preuve de concept sur une lignée de cellules de glioblastome. Le glioblastome (GBM) est la tumeur cérébrale primaire la plus courante et la plus mortelle avec une survie globale médiane à moins de 15 mois malgré un traitement agressif. Il est donc important de bien connaître les marqueurs moléculaires du GBM afin d'améliorer le diagnostic et développer de nouvelles thérapies pour ce cancer (48).

La preuve de concept sera réalisée par la technique du PIC_h en tandem sur le promoteur de la télomérase dans une lignée de cellules de glioblastome ayant les 2 allèles mutés au niveau du promoteur TERT. Le deuxième allèle sera muté (par CRISPR/Cas9) afin d'obtenir une expression bi-allélique de TERT. Cette caractérisation sera comparée à la même lignée de glioblastome ne contenant pas de mutation au niveau du promoteur de TERT (modification par CRISPR/Cas9) et immortalisée par un plasmide contenant le gène de la télomérase. Enfin une troisième lignée sera créée. Dans cette lignée contrôle, le promoteur de la télomérase sera totalement enlevé par CRISPR/Cas9 et un plasmide contenant la télomérase sera également ajouté afin de permettre la culture de cette lignée (immortalisation). Les protéines purifiées seront analysées par spectrométrie de masse quantitative.

Par conséquent, le fait d'identifier les ETS (facteurs de transcription) sur la lignée présentant une mutation au niveau du promoteur de TERT permettra de valider la technique de purification à un locus unique. L'analyse des autres protéines identifiées permettra également de mieux comprendre la biologie de la réactivation de ce gène dans ces lignées cancéreuses. La lignée contrôle sans mutation permettra de discriminer les protéines spécifiques d'un promoteur actif ou inactif. Enfin, la troisième lignée permettra d'établir le bruit de fond expérimental. Il s'agira de la première expérimentation permettant de déterminer le bruit de fond dans ce type de technologie et ainsi d'éliminer les protéines non spécifiques du locus ciblé. En effet, ce type d'expérimentation ne peut pas être réalisée sur les télomères ou les péricentromères car il n'est pas possible de les enlever du génome.

Dans un second temps, cette technique pourra être utilisée dans le but de mieux comprendre les mécanismes mis en jeu dans différents types de tumeur et ainsi permettre d'identifier de nouvelles cibles thérapeutiques. Les protéines associées au niveau du promoteur de la télomérase seront purifiées dans plusieurs types de glioblastome. En effet, il existe différents profils génétiques de glioblastome pouvant conduire à des temps de survie différents (Figure 23).

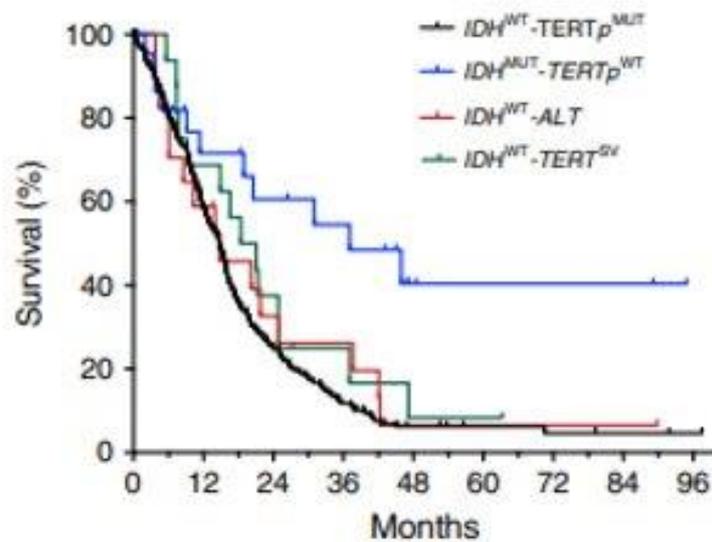


Figure 23 : Profils de survie en fonction du statut génétique de glioblastomes. D’après Diplas BH et al. Nat Commun. 2018 May 25;9(1):2087.

De plus, ces différents profils génétiques entraînent une activation ou une répression de l’expression de la télomérase. Ainsi, il existe 2 formes de glioblastome ayant une activité télomérase (activation de l’expression de TERT) et 2 formes de type ALT (répression de l’expression de TERT). Enfin, une autre catégorie de glioblastome (IDH WT-other) présente un mélange de formes permettant une activation (via la fixation de c-MYC) ou une répression du gène de la télomérase (Figure 24).

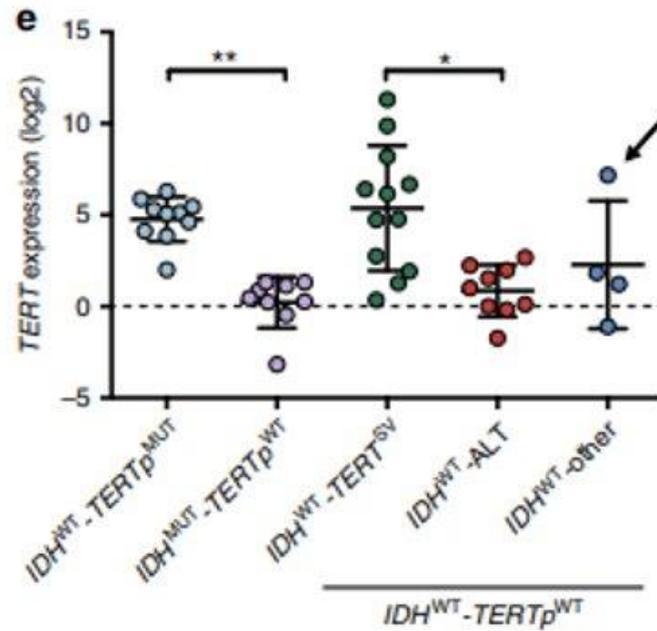


Figure 24 : Niveaux d'expression du gène *TERT* dans les différents types de glioblastomes.
D'après Diplas BH et al. Nat Commun. 2018 May 25;9(1):2087.

Ces différents profils génétiques sont liés au promoteur de la télomérase qui peut être WT ou muté ou présenter un variant structural mais sont également liés à la mutation du gène de l'isocitrate déshydrogénase (IDH).

La mutation du gène de l'IDH est rencontrée dans les glioblastomes de type ALT. Cette mutation joue un rôle dans l'hyperméthylation, notamment dans la méthylation de l'H3K9. L'IDH entraîne donc des changements épigénétiques dans les cellules de glioblastome et les rend vulnérables à l'acquisition de nouveaux facteurs oncogènes tels que TP53 (tumor suppressor protein 53) (49). De plus, le promoteur de la télomérase est entouré d'îlots CpG. Ces îlots sont non méthylés dans les cellules normales et méthylés dans les cellules malignes (43). Enfin, les cellules de glioblastome de type ALT présentent une perte de ATRX ou de SMARCA1. Ces 2 protéines appartiennent à la même famille de remodelleur de la chromatine et possèdent à la fois des domaines ATP-binding et C-terminal helicase (46). Il serait donc

intéressant de comprendre le mécanisme de répression du gène de la télomérase en fonction du statut génétique de l'IDH afin de mieux appréhender sa fonction dans ce type de cancer.

D'autre part, il existe 2 types de statuts génétiques des glioblastomes entraînant l'activation de l'expression de la télomérase, la mutation du promoteur de la télomérase ou un variant structural de ce promoteur. Ces 2 modifications génétiques sont retrouvées dans des glioblastomes en absence de mutation du gène de l'IDH. La mutation du promoteur de la télomérase entraîne la fixation de facteurs de transcriptions tels que GABPA et donc l'expression du gène *TERT*. Lorsque le promoteur de *TERT* présente un variant structural, il n'y a pas de mutations sur le promoteur mais un point de cassure situé 50kpb en amont du promoteur qui entraîne un réarrangement permettant l'activation du gène *TERT* via un super enhancer (Figure 25). L'expression du gène *TERT* peut donc se produire sans mutations au niveau du corps du promoteur.

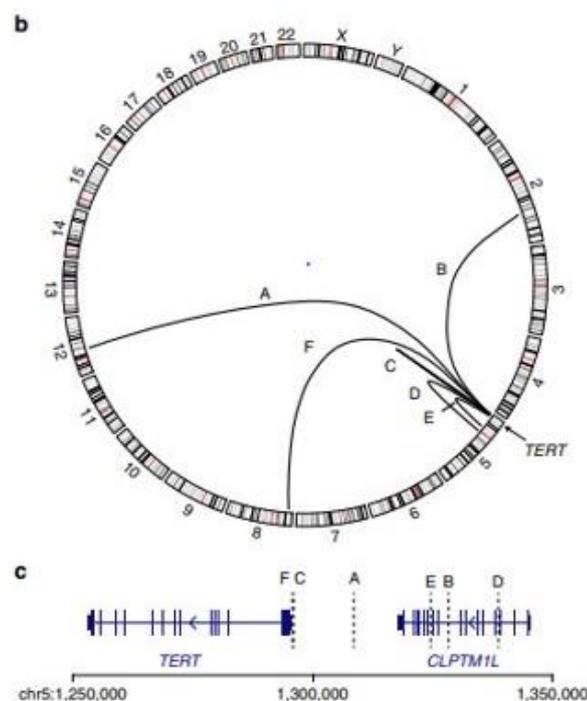


Figure 25 : Réarrangements situés en amont de *TERT* analysés par séquençage du génome dans des glioblastomes ALT négatifs. Circos plot of rearrangements identified upstream of *TERT* by whole genome sequencing of ALT-negative GBMs ($N = 8$). Several cases were

interchromosomal translocations (A, B, F), while the remaining cases were intrachromosomal (C, D, E). D'après Diplas BH et al. Nat Commun. 2018 May 25;9(1):2087.

La purification des protéines associées au niveau du promoteur du gène *TERT* dans des cellules de glioblastomes présentant un variant structural de ce promoteur permettrait de mieux comprendre les mécanismes qui permettent la réactivation de l'expression de *TERT* (sans mutations du corps du promoteur) dans ce type de glioblastome.

La purification non biaisée des protéines associées au niveau du promoteur de *TERT* dans ces différents cas de figure (répression ou activation de *TERT*) permettrait d'identifier les facteurs mis en jeu dans différents cancers et ainsi de mieux appréhender leurs mécanismes. De plus, les protéines identifiées pourraient être utilisées pour le diagnostic des différents types de GBM ou comme cibles thérapeutiques afin d'améliorer le traitement de ce type de cancer.

CONCLUSION GENERALE

Dans le génome, chaque locus est défini par une combinaison unique de protéines qui déterminent les propriétés et la fonction de ce locus. Etablir la composition locale de la chromatine à un locus spécifique est donc une étape essentielle à la compréhension de la régulation de la chromatine et à sa fonction biologique.

Ces 10 dernières années, différents laboratoires ont développé des techniques de purification d'un locus spécifique et notamment d'un locus unique. Toutefois, les analyses protéomiques d'un locus unique de mammifère retrouvent des facteurs protéiques inattendus alors que les protéines caractéristiques du locus sont absentes. Ceci démontre donc les limitations des méthodes actuelles. En effet, dans ce type d'approche, le rapport entre le signal et le bruit de fond doit être parfaitement appréhendé pour permettre l'identification des protéines spécifiques au niveau d'un locus unique de mammifère.

De plus, contrairement aux acides nucléiques, les peptides ne peuvent pas être amplifiés actuellement. Ceci implique donc de travailler avec de très grandes quantités de matériel de départ car les spectromètres de masse ont une limite de détection de l'ordre de la femtomole. L'identification d'un peptide dépend donc de son abondance mais aussi de sa séquence. En effet, certains peptides sont moins ionisés et donc séparés. Par exemple, les peptides "flyables" sont identifiés plus facilement que les "non flyables". Toutefois, les avancées dans le marquage des protéines (notamment le TMT) permettent d'obtenir une meilleure fragmentation des peptides et donc un meilleur signal à condition que les marqueurs ne soient pas en excès. Enfin, les spectromètres de masse ont une capacité limitée pour le traitement des ions en parallèle ce qui entraîne une accumulation des ions dans le spectromètre et donc la quantification d'un nombre limité de peptides par unité de temps. Pour pallier cela, une analyse avec une acquisition lente doit être utilisée (50,51).

Toutefois, l'émergence de la protéomique sur cellule unique permet de nouvelles avancées dans le domaine de la spectrométrie de masse et de l'analyse des données de protéomiques à partir d'échantillons en très faible abondance. La combinaison de ces dernières permettra une augmentation des enrichissements et de la sensibilité des analyses (51).

BIBLIOGRAPHIE

1. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res.* 2011;21(3):381-95.
2. Tamaru H. Confining euchromatin/heterochromatin territory: jumonji crosses the line. *Genes Dev.* 2010;24(14):1465-78.
3. Saksouk N, Simboeck E, Déjardin J. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin.* 2015;8(1):3.
4. Gauchier M, Kan S, Barral A, Sauzet S, Agirre E, Bonnell E, et al. SETDB1-dependent heterochromatin stimulates alternative lengthening of telomeres. *Sci Adv.* 2019;5(5):eaav3673.
5. Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun Z-W, Schmid M, et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature.* 2000;406(6796):593-9.
6. Becker JS, Nicetto D, Zaret KS. H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes. *Trends Genet.* 2016;32(1):29-41.
7. Kimura H. Histone modifications for human epigenome analysis. *J Hum Genet.* 2013;58(7):439-45.
8. Riethoven J-JM. Regulatory Regions in DNA: Promoters, Enhancers, Silencers, and Insulators. *Methods Mol Biol.* 2010;674:33-42.
9. Elkon R, Agami R. Characterization of noncoding regulatory DNA in the human genome. *Nat Biotechnol.* 2017;35(8):732-46.
10. Lenhard B, Sandelin A, Carninci P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet.* 2012;13(4):233-45.
11. Danino YM, Even D, Ideses D, Juven-Gershon T. The core promoter: At the heart of gene expression. *Biochim Biophys Acta.* 2015;1849(8):1116-31.
12. Ernst J, Kellis M. Discovery and Characterization of Chromatin States for Systematic Annotation of the Human Genome. *Nat Biotechnol.* 2010;28(8):817-825.
13. Ong C-T, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet.* 2011;12(4):283-93.
14. Raab JR, Kamakaka RT. Insulators and promoters: closer than we think. *Nat Rev Genet.* 2010;11(6):439-46.
15. Dean A. In the loop: long range chromatin interactions and gene regulation. *Brief Funct Genomics.* 2011;10(1):3-10.

16. Gauchier M, van Mierlo G, Vermeulen M, Déjardin J. Purification and enrichment of specific chromatin loci. *Nat Methods*. 2020;17(4):380-9.
17. Vermeulen M, Déjardin J. Locus-specific chromatin isolation. *Nat Rev Mol Cell Biol*. 2020;21(5):249-50.
18. Altelaar AFM, Frese CK, Preisinger C, Hennrich ML, Schram AW, Timmers HThM, et al. Benchmarking stable isotope labeling based quantitative proteomics. *J Proteomics*. 2013;88:14-26.
19. Vermeulen M, Eberl HC, Matarese F, Marks H, Denissov S, Butter F, et al. Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell*. 2010;142(6):967-80.
20. Déjardin J, Kingston RE. Purification of Proteins Associated with Specific Genomic Loci. *Cell*. 2009;136(1):175-86.
21. Kennedy-Darling J, Guillen-Ahlers H, Shortreed MR, Scalf M, Frey BL, Kendzioriski C, et al. Discovery of Chromatin-Associated Proteins via Sequence-Specific Capture and Mass Spectrometric Protein Identification in *Saccharomyces cerevisiae*. *J Proteome Res*. 2014;13(8):3810-25.
22. Liu X, Zhang Y, Chen Y, Li M, Zhou F, Li K, et al. In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell*. 2017;170(5):1028-1043.e19.
23. Schmidtman E, Anton T, Rombaut P, Herzog F, Leonhardt H. Determination of local chromatin composition by CasID. *Nucleus*. 2016;7(5):476-84.
24. Myers SA, Wright J, Peckner R, Kalish BT, Zhang F, Carr SA. Discovery of proteins associated with a predefined genomic locus in living cells via dCAS9-APEX-mediated proximity labeling. *Nat Methods*. 2018;15(6):437-9.
25. Gao XD, Tu L-C, Mir A, Rodriguez T, Ding Y, Leszyk J, et al. C-BERST: Defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. *Nat Methods*. 2018;15(6):433-6.
26. Qiu W, Xu Z, Zhang M, Zhang D, Fan H, Li T, et al. Determination of local chromatin interactions using a combined CRISPR and peroxidase APEX2 system. *Nucleic Acids Res*. 2019;47(9):e52.
27. Tsui C, Inouye C, Levy M, Lu A, Florens L, Washburn MP, et al. dCas9-targeted locus-specific protein isolation method identifies histone gene regulators. *Proc Natl Acad Sci USA*. 2018;115(12):E2734-41.
28. Fujita T, Fujii H. Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochem Biophys Res Commun*. 2013;439(1):132-6.

29. Waldrip ZJ, Byrum SD, Storey AJ, Gao J, Byrd AK, Mackintosh SG, et al. A CRISPR-based approach for proteomic analysis of a single genomic locus. *Epigenetics*. 2014;9(9):1207-11.
30. Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol*. 2013;14(1):49-55.
31. Fujita T, Fujii H. Direct identification of insulator components by insertional chromatin immunoprecipitation. *PLoS One*. 2011;6(10):e26109.
32. Hamperl S, Brown CR, Garea AV, Perez-Fernandez J, Bruckmann A, Huber K, et al. Compositional and structural analysis of selected chromosomal domains from *Saccharomyces cerevisiae*. *Nucleic Acids Res*. 2014;42(1):e2.
33. Byrum SD, Raman A, Taverna SD, Tackett AJ. ChAP-MS: A Method for Identification of Proteins and Histone Posttranslational Modifications at a Single Genomic Locus. *Cell Rep*. 2012;2(1):198-205.
34. Korthout T, Poramba-Liyanage DW, van Kruijsbergen I, Verzijlbergen KF, van Gemert FPA, van Welsem T, et al. Decoding the chromatin proteome of a single genomic locus by DNA sequencing. *PLoS Biol*. 2018;16: e2005542.
35. Roux KJ, Kim DI, Raida M, Burke B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol*. 2012;196(6):801-10.
36. Kan SL, Saksouk N, Déjardin J. Proteome Characterization of a Chromatin Locus Using the Proteomics of Isolated Chromatin Segments Approach. *Methods Mol Biol*. 2017;1550:19-33.
37. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012;337(6096):816-21.
38. Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, et al. Phage Response to CRISPR-Encoded Resistance in *Streptococcus thermophilus*. *J Bacteriol*. 2008;190(4):1390-400.
39. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, et al. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell*. 2013;152(5):1173-83.
40. Cheng AW, Wang H, Yang H, Shi L, Katz Y, Theunissen TW, et al. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res*. 2013;23(10):1163-71.
41. Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods*. 2013;10(10):973-6.

42. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*. 2014;159(3):647-61.
43. Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li G-W, et al. Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell*. 2013;155(7):1479-91.
44. Cox J, Mann M. 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics*. 2012;13(S16):S12.
45. Yuan X, Larsson C, Xu D. Mechanisms underlying the activation of TERT transcription and telomerase activity in human cancer: old actors and new players. *Oncogene*. 2019;38(34):6172-83.
46. Ramlee MK, Wang J, Toh WX, Li S. Transcription Regulation of the Human Telomerase Reverse Transcriptase (hTERT) Gene. *Genes*. 2016;7(8):50.
47. Stern JL, Theodorescu D, Vogelstein B, Papadopoulos N, Cech TR. Mutation of the *TERT* promoter, switch to active chromatin, and monoallelic *TERT* expression in multiple cancers. *Genes Dev*. 2015;29(21):2219-24.
48. Diplas BH, He X, Brosnan-Cashman JA, Liu H, Chen LH, Wang Z, et al. The genomic landscape of TERT promoter wildtype-IDH wildtype glioblastoma. *Nat Commun*. 2018;9(1):2087.
49. Miller JJ, Shih HA, Andronesi OC, Cahill DP. Isocitrate dehydrogenase-mutant glioma: Evolving clinical and therapeutic implications. *Cancer*. 2017;123(23):4535-46.
50. Budnik B, Levy E, Harmange G, Slavov N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol*. 2018;19(1):161.
51. Marx V. A dream of single-cell proteomics. *Nat Methods*. 2019;16(9):809-12.

LISTE DES ANNEXES

Annexe 1 : Carte des plasmides

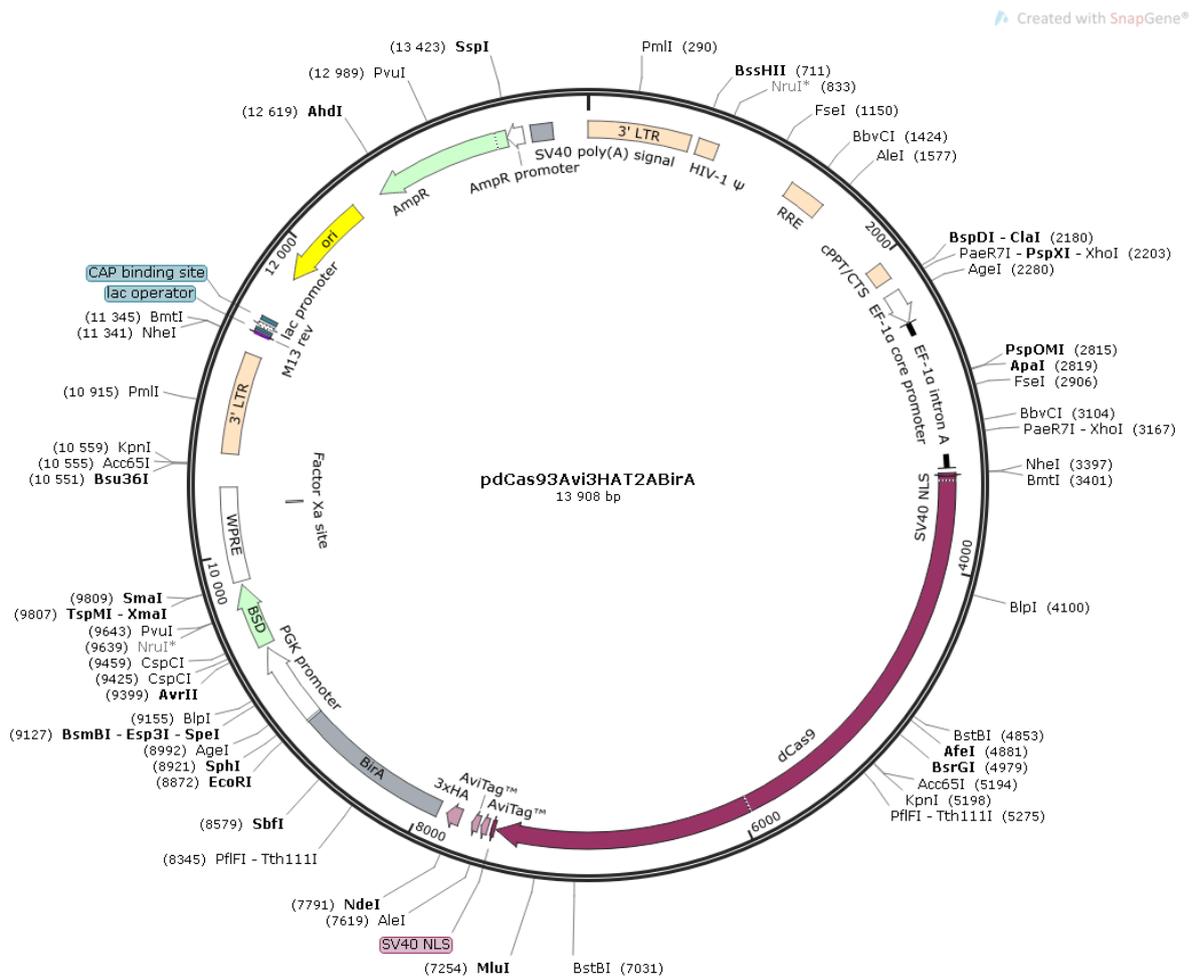
Annexe 2 : Tableau des anticorps

Annexe 3 : Liste des sondes de PICH et PICH en tandem

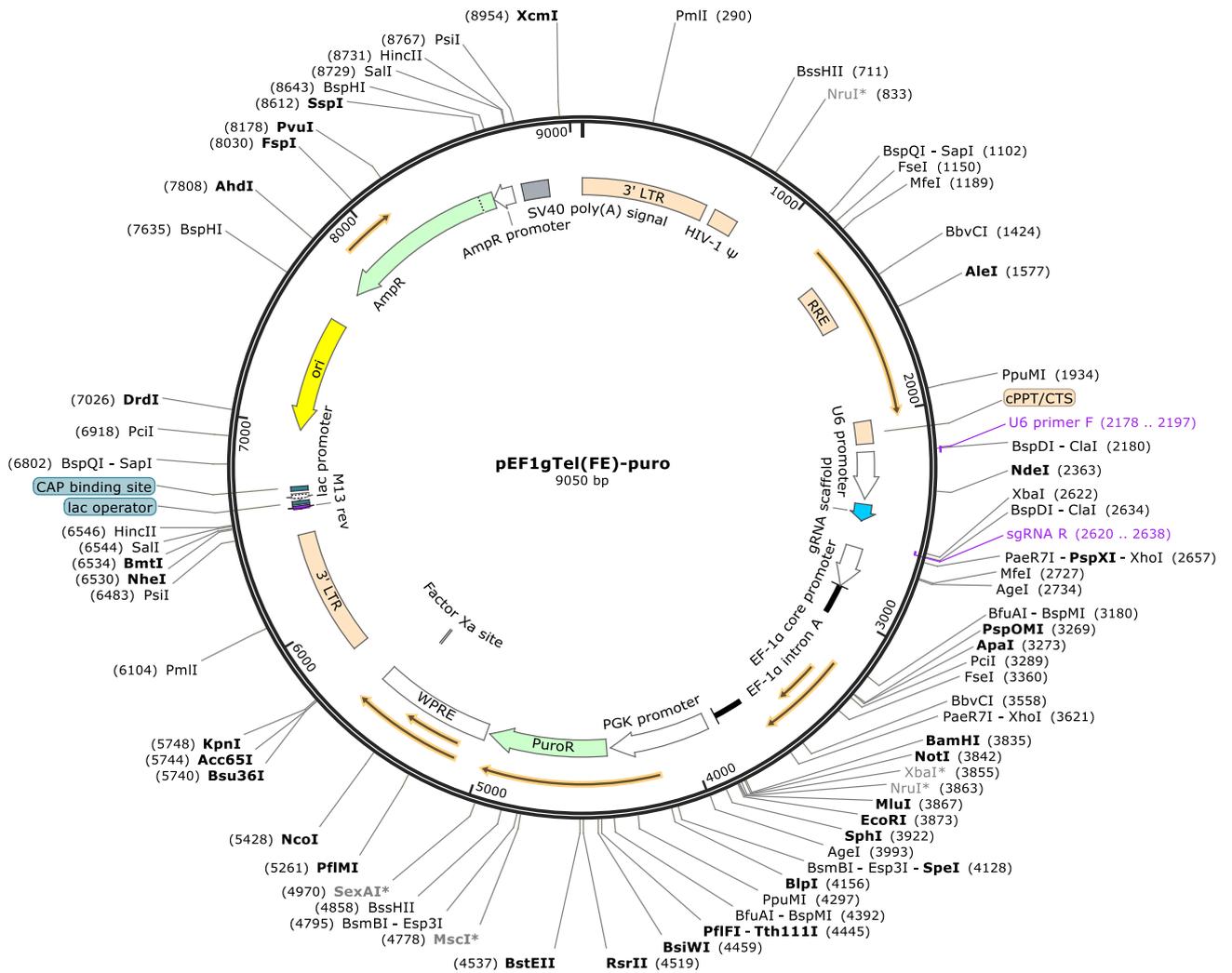
Annexe 4 : Protocole du PICH en tandem

Annexe 1 : Carte des plasmides

pdCas9-HA - blasticidine



pEF1 - guideTélomérique - puromycine



Annexe 2 : Tableau des anticorps

Nom	Type	Compagnie	Application
HA.11	Mouse Monoclonal	BIOLEGEND	WB/IF
Tubuline Alpha	Mouse Monoclonal	SIGMA T9026	WB
TRF2	Mouse Monoclonal	Abcam ab 13579	WB/IF
DIG	Mouse Monoclonal	Abcam ab 420	
TRF1	Rabbit Polyclonal	Biologics Corp	WB/IF
H3K9me3	Rabbit Polyclonal	Abcam Ab 8898	WB
RAP1	Rabbit Polyclonal	Abcam ab 4181	WB

Annexe 4 : Protocole du PICh en tandem

La technique de PICh en tandem a été mise au point dans le but de purifier des protéines associées au niveau d'un locus unique.

Cette technique est basée sur l'utilisation de sondes spécifiques qui permettent la capture d'un locus spécifique et des protéines qui y sont associées après une fixation de ces protéines par crosslink avec du formaldéhyde. La première purification est ensuite réalisée grâce à des billes de streptavidine qui vont se fixer à la sonde spécifique qui contient une desthioiotine du côté 5'. L'élution est réalisée par ajout de biotine. Une seconde purification est réalisée grâce à des billes (protéines A) couplées à un anticorps anti-DIG. L'anticorps se lie à la digoxigénine placée à l'extrémité 5' de la sonde. La deuxième élution est réalisée par photo-clivage de la sonde par les UV grâce à un espaceur photoclivable placé entre la digoxigénine et la desthioiotine. Après avoir la réversion des crosslinks, les protéines peuvent être analysées par spectrométrie de masse et western-blot (Figure 1).

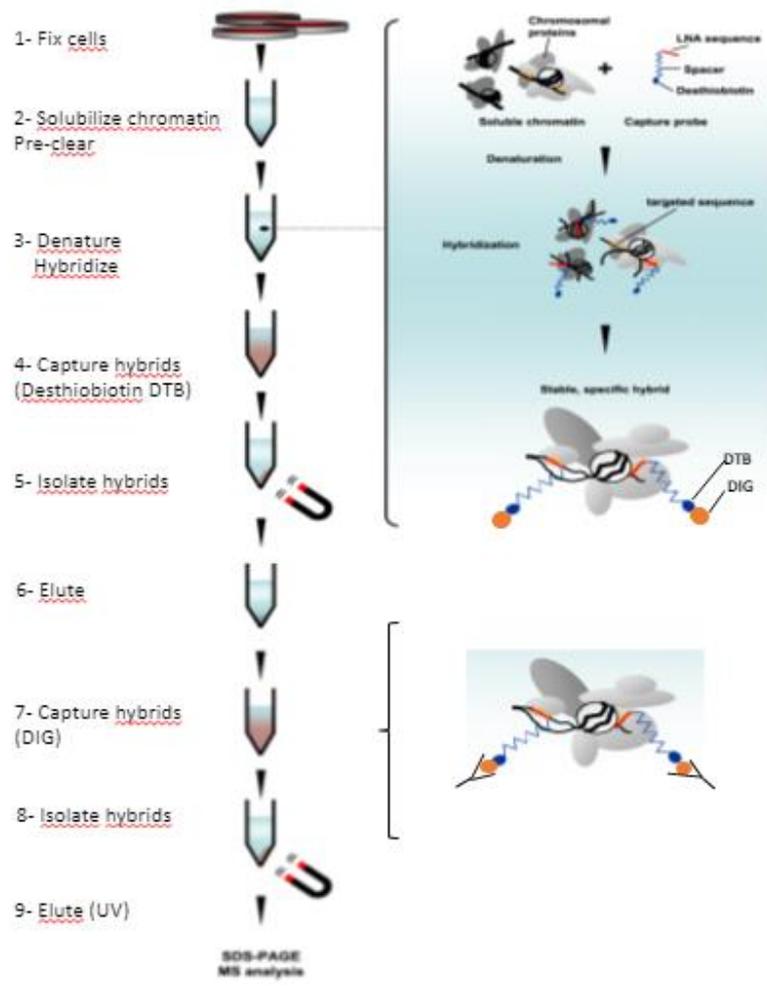


Figure 1 : Protocole de la technique du PICCh en tandem adapté d’après J. Dejardin, RE. Kingston - Cell. 2009 Jan 9; 136(1): 175–186.

Préparation des sondes du PICCh en tandem

Les sondes sont synthétisées par le laboratoire “Keck” de l’Université de Yale (Oligo Synthesis Resource). Il s’agit de sondes ADN de 65 nt. Dans le sens 5’ vers 3’, les sondes contiennent un espaceur amino-modifié C12, un espaceur photo-clivable, une desthiobiotine, 2 espaceurs puis les 65 nucléotides spécifiques de la séquence cible (annexe 3).

Le espaceur amino-modifié C12 contient un NH₂ libre. Ce radical est utilisé pour rajouter la digoxigénine. Les oligonucléotides arrivent purifiés, séchés et dessalés à une quantité de 200 nmol. Chaque sonde est tout d’abord reprise dans 500µL de tampon carbonate 0.1M pH 9. La

poudre de DIG-NHS ester est reprise dans du DMSO anhydre à une concentration de 0.2M. 6µL de DIG-NHS ester à 0.2M est ajouté à chaque sonde. Les sondes sont ensuite vortexées et incubées 1 à 2 heures. Après ajout de 500µL de tampon carbonate, les sondes sont dessalées sur des colonnes Gel Pak 1.0. Pour cela, après avoir équilibré la colonne avec 15mL de tampon carbonate, les sondes sont passées sur les colonnes. L'élution est réalisée avec 1mL d'eau. Enfin, les sondes sont précipitées à l'éthanol 100% puis centrifugées et lavées à l'éthanol 75% avant d'être séchées. Les sondes sont reprises dans du tampon TE 0.1X à une concentration de 100µM, stockées à -80°C et emballées dans du papier d'aluminium afin d'éviter le photo-clivage de la sonde.

Une étape de purification des sondes sur gel d'acrylamide ou par HPLC devra être ajoutée afin de purifier la sonde (élimination de la DIG-NHS ester libre).

Crosslink, lyse des cellules et préparation de la chromatine

Les étapes de fixation et de lyse des cellules pour la technique de PICh en tandem sont identiques à celles utilisées dans la technique de PICh. De même, la préparation de la chromatine reste inchangée.

Hybridation et captures de la chromatine

Toutes les étapes d'hybridation et de capture doivent être effectuées sous du papier d'aluminium afin de ne pas entraîner un photo-clivage de la sonde.

La chromatine est divisée en autant de tube que d'échantillons. 15µL d'une sonde spécifique à 100µM sont ajoutés dans chaque tube puis vortexés. La chromatine mélangée avec la sonde est distribuée dans des plaques 96 puits (150µL par puit) puis placée dans un thermocycleur. Le programme d'hybridation est le suivant : 3 min à 25°C, 5 min à 82°C, 60 min à 37°C, 3 min

à 60°C, 30 min à 37°C, 3 min à 60°C, 30 min à 37°C puis une température finale de 25°C. A l'issue de l'étape d'hybridation, la chromatine est récupérée dans des tubes 1,5mL puis centrifugée 15 min à 18 000g.

Le surnageant est ensuite dilué au ½ dans de l'eau puis placé avec 900µL de billes magnétiques MyOne C1 (Life Technologie) préalablement lavées dans 10mL de tampon LB3-0.1% SDS (10mM HEPES pH 7.9, 100mM NaCl, 2mM EDTA pH 8, 1mM EGTA pH 8, 0.1% SDS, 0.1% sodium sarkosyl). Après 90 min sous rotation, les billes sont immobilisées sur un portoir magnétique et le surnageant est récupéré (FT). Les billes sont ensuite lavées 3 fois avec 10mL de LB3-0.1% SDS (les tubes sont placés sur le portoir magnétique entre chaque lavage pour éliminer le tampon). Les billes sont transférées avec 1mL de tampon LB3-0.1% SDS dans des tubes 1.5mL puis lavées avec 1mL de tampon LB3-0.1% SDS. Un dernier lavage est effectué avec 1mL de tampon LB3-0.1% SDS à 42°C, sous agitation pendant 5 min.

Après avoir éliminé le surnageant, 1mL de tampon d'élution (50mM Tris pH8, 200mM NaCl, 20mM EDTA pH 8, 0.1% SDS, 12.5mM biotine) est ajouté aux billes. Puis, les billes sont incubées 30 min à 37°C sous agitation puis 15 min à 50°C avant d'être immobilisées sur le portoir magnétique. Le surnageant est récupéré dans un nouveau tube 1.5mL.

Pour chaque échantillon, 200µL de billes magnétiques protéines A pré-équilibrées dans le tampon LB3-0.1% SDS sont incubées avec 8µL d'anticorps anti DIG pendant 2H. L'éluat de la première capture est ensuite ajouté aux billes protéine A - Anticorps anti-DIG et incubé 2 H sous rotation à RT. Les billes sont ensuite lavées 5 fois avec 1mL de LB3-0.1% SDS puis transférées dans des petits sachets plastiques. Les sachets sont ensuite placés sur un bloc froid (à 4°C), sur un agitateur 3D et incubés 2H sous une lampe à UV (UV à 360nm). Le surnageant est ensuite récupéré.

Les quantités de billes protéine A et d'anticorps anti-DIG devront être optimisés ainsi que l'étape d'élution de la DIG.

Précipitation des protéines et reverse crosslink

Les étapes de précipitation des protéines et reverse crosslink sont identiques entre le PICH et le PICH en tandem.

ÉCOLE PRATIQUE DES HAUTES ÉTUDES SCIENCES DE LA VIE ET DE LA TERRE

Développement d'une technique d'identification des protéines associées à un locus spécifique

SAUZET SANDRINE

17 décembre 2020

RESUME

L'état de la chromatine et les éléments de régulation des gènes tels que les promoteurs permettent la régulation de l'expression génique. Actuellement, des techniques ciblées tels que l'immunoprécipitation de la chromatine (CHIP) sont principalement utilisées pour l'étude de la biologie de la chromatine. La purification des protéines associées au niveau des éléments de régulation des gènes de façon non biaisée permettrait de mieux comprendre leur activation ou leur répression au cours du développement ou lors de maladies. De plus, les protéines identifiées seraient susceptibles d'apporter de nouvelles connaissances pour le diagnostic ou la thérapie de ces maladies.

Ces 10 dernières années différentes techniques de purification d'un locus unique ont été développées. En effet, pour enrichir un locus unique de 3kbp soit 0.0001% du génome humain, le taux d'enrichissement doit être de 1 million de fois pour obtenir un mélange quasi-homogène. Les techniques actuellement publiées utilisent majoritairement des systèmes CRISPR/deadCas9. La protéine dCas9 est une grosse protéine (>150kDa). Son efficacité de liaison pourrait par conséquent être fortement influencée par la structure de la chromatine. En outre, la liaison de la protéine dCas9 pourrait concurrencer la liaison d'interacteurs biologiques endogènes. La compréhension de l'impact de ce type de système sur la composition locale de chromatine est un point clé pour évaluer l'efficacité de ces techniques. D'autres systèmes basés sur l'utilisation de sondes de captures ont été développés. Ces techniques, comme celles utilisant dCas9 ne comportent qu'une seule étape de purification par affinité ce qui ne permet pas d'obtenir un taux d'enrichissement suffisant pour la caractérisation d'un locus unique.

Dans une première partie, l'impact de la protéine dCas9 sur la composition locale de la chromatine est étudiée au niveau d'un locus spécifique. Dans une seconde partie, une nouvelle technique de purification d'un locus unique basée sur l'utilisation de sondes de capture est développée. Il s'agit d'une adaptation de la technologie PICH (Proteomics of isolated chromatin segments) rendue très sensible et quantitative, et avec un taux d'enrichissement augmenté grâce à l'ajout d'une seconde étape de purification par affinité.

MOTS-CLÉS : purification d'un locus spécifique, développement méthodologique, chromatine, spectrométrie de masse quantitative, purification d'un locus unique, télomère, CRISPR/Cas9.