

Master 2 GID
Génétique, Immunité et Développement
Faculté des Sciences de l'Université de Nice Sophia-Antipolis

Recherche des partenaires moléculaires d'un long ARN non codant, régulé par l'hypoxie, dans les adénocarcinomes pulmonaires

Marine GAUTIER

Stage réalisé sous la direction du **Dr Roger REZZONICO**

*IPMC – CNRS UMR 7275
Equipe Pascal Barbry
660 Route des Lucioles
SOPHIA ANTIPOLIS -06650 Valbonne*

A stylized, handwritten signature in blue ink that reads "ipmc". The signature is written in a cursive, lowercase font and is underlined with a thick blue stroke.

2016

TABLE DES MATIERES

RESUME	3
ABREVIATIONS	4
INTRODUCTION	5
1. Les adénocarcinomes pulmonaires, un problème clinique majeur.....	5
2. Le contexte tumoral hypoxique.....	5
3. Les Longs ARN non codants, d'importants régulateurs de l'expression génique.....	6
4. LinCH3, un LncRNA induit par l'hypoxie dans les ADC.....	6
OBJECTIFS	9
MATERIELS ET METHODES	10
RESULTATS	16
1. Mise au point de la technique de ChIRP.....	16
a) Vérification de la réponse hypoxique.....	16
b) Fragmentation.....	17
2. Validation de la précipitation des LncRNA par la technique de ChIRP.....	19
3. Identification de séquences d'ADN génomique associées à LinCH3.....	21
a) Validation de la précipitation d'ADN génomique associées à LinCH3.....	21
b) Résultats des données de ChIRP-Seq.....	22
4. Impact de l'inactivation de LinCH3 sur le transcriptome de cellules d'ADC au cours de la réponse hypoxique.....	25
5. Comparaison des résultats de transcriptomique et de ChIRP-Seq.....	27
6. Etude préliminaire de recherche des partenaires protéiques des LncRNA.....	29
DISCUSSION	31
CONCLUSIONS ET PERSPECTIVES	34
1. Conclusion.....	34
2. Perspectives.....	34
BIBLIOGRAPHIE	36

RESUME

Mots-clés :

Adénocarcinomes pulmonaires, longs ARN non codants, Hypoxie, ChIRP
(Chromatin Isolation by RNA Purification), LincH3.

Lung cancer is the leading cause of cancer death worldwide, with poor prognosis and a high rate of recurrence despite early surgical removal. It is therefore essential to identify new prognostic markers and new therapeutic targets. My team is interested in gene regulation related to hypoxia, a factor associated with relapse of lung adenocarcinomas (LUAD), the most represented form of lung cancer. While long non-coding RNAs (lncRNA) are increasingly recognized as major gene expression regulators through various mechanisms involving nucleic acid and protein partners, their roles in cancer development and hypoxic response are still largely unexplored. We previously identified a "hypoxaLinc", LincH3 that is significantly up-regulated both in LUAD tumor tissues and cell lines in response to hypoxia. This 10 kb nuclear transcript is regulated by NF κ B and may participate in the regulation of hypoxic or inflammatory responses.

My project focused on the characterization of LincH3 mode of action through the identification of its DNA-binding sites. My goal was to set up a novel technique, the ChIRP (Chromatin Isolation by RNA Purification), based on affinity capture of lncRNA: chromatin complexes by tiling antisense-oligos and mapping of its genomic binding sites at high resolution. Using WT and CRISPR/Cas9 LincH3-deleted LUAD cells, I performed ChIRP-Seq to identify 254 LincH3 binding sites on chromatin. Through transcriptome analysis I showed dysregulation of TGF β , NRF2 and NF κ B signaling pathways in LincH3-deficient LUAD cells. Finally, the comparative analysis of these data identified 11 potential target genes of LincH3 including *CTGF*, a central regulator of TGF β and angiogenic pathways.

ABREVIATIONS

lncRNA: Long ARN non codant

ChIRP: Chromatin Isolation by RNA purification

NEAT1: Nuclear Enriched Abondant Transcript

LincH3: Long ARN non codant intergénique H3

MALAT: Metastasis Associated Lung Adenocarcinoma Transcript 1

HOTAIR: HOX Antisens Intergenic RNA

HIF: Hypoxia-inducible factors

HRE: HIF-Responsive Element

ARN: acide ribonucléique

ADN: acide désoxyribonucléique

ATP: Adenosine Tri-Phosphate

WT: Wild type – phénotype sauvage

DEL: Clones de cellules A549 invalidés pour l'expression de *LincH3*

PCR: polymerase chain reaction

RTqPCR: Reverse Transcription quantitative polymerase chain reaction

CTGF: Connective Tissue Growth Factor

MREG: Melanoregulin

IGFBP3: Insulin-like Growth Factor-Binding Protein 3

LTBP1: Latent-Transforming growth factor Beta-binding Protein 1

MTSS1: Metastasis Suppressor 1

PTPRJ: Protein Tyrosine Phosphatase Receptor type J

RPL23AP32: Ribosomal protein L23a Pseudogene 32

FBN2: Fibrillin 2

ZSWIM6: Zinc Finger SWIM-type containing 6

BCL2L1: B-cell CCL/Lymphoma 2 Like 1

PSPC1: Paraspeckles Protein Component 1

CHART : Capture Hybridization Analysis RNA Targets

RIP : RNA Immunoprecipitation

TEM : Transition épithélio-mésenchymateuse

INTRODUCTION

1. Les adénocarcinomes pulmonaires, un problème clinique majeur

Les cancers broncho-pulmonaires représentent la première cause de décès par cancer dans les pays développés (30%, 32000 décès/an en France). D'un point de vue histologique, on distingue les carcinomes à petites cellules (CPC) et les carcinomes "non-à-petites-cellules" (CNPC) qui regroupe trois sous-types principaux : les carcinomes épidermoïdes, les carcinomes à grandes cellules et les adénocarcinomes (ADC) (Figure 1A) ^[19].

Les CNPC sont des cancers liés au tabagisme avec une incidence croissante et un mauvais pronostic lié à une résistance aux traitements conventionnels (chimio et radiothérapies) ainsi qu'aux nouvelles approches de thérapies ciblées ou d'immunothérapies. Ainsi, bien que la détection précoce et l'exérèse curative aient significativement amélioré la survie des patients, 50% des patients de stade précoce récidivent encore dans les 5 ans suivant le diagnostic, puis échappent le plus souvent aux traitements. L'enjeu actuel est donc d'identifier de nouveaux marqueurs pronostiques et de nouvelles cibles thérapeutiques.

Mon projet porte exclusivement sur les ADC qui sont la forme la plus fréquente de CNPC, et plus particulièrement sur l'influence de l'environnement tumoral et notamment de l'hypoxie sur l'agressivité et la résistance aux traitements de ces tumeurs.

2. Le contexte tumoral hypoxique

Au cours du développement tumoral, la diffusion des nutriments et de l'oxygène est limitée au sein d'une tumeur solide, induisant un environnement hypoxique. Afin d'échapper à la mort cellulaire programmée, les cellules tumorales vont s'adapter à ces conditions en effectuant un changement métabolique, appelé effet de Warburg, qui consiste en un basculement de la respiration oxydative mitochondriale vers la glycolyse anaérobie. Les cellules tumorales hypoxiques induisent de l'angiogénèse et vont devenir plus agressives, plus résistantes aux traitements et acquérir des propriétés invasives ^[9, 20].

Ces effets sont médiés par la voie de signalisation impliquant les facteurs de transcription HIF (Hypoxia Inducible Factor) (Figure 1B) qui agissent sur des séquences régulatrices HRE (HIF-Response Element) afin d'induire la transcription de gènes codants pour des protéines en lien avec l'angiogénèse (ex : *VEGF*) ou encore avec le changement métabolique (ex : *PGK1*, *CA9*) ^[18].

D'autres études, réalisées notamment dans notre équipe, ont mis en évidence l'induction d'ARN non codants par les facteurs de transcription HIF [11]. C'est le cas du microARN miR-210 qui participe à la mise en place de l'effet Warburg [2, 7, 14].

Plus récemment d'autres ARN non codants appartenant à la famille émergente des longs ARN non codants (LncRNA) ont été identifiés comme étant régulés par l'hypoxie dans différents types de cancers [8, 11, 22, 27].

3. Les Longs ARN non codants, d'importants régulateurs de l'expression génique

Les LncRNA sont définis comme des transcrits ayant des tailles supérieures à 200 pb, et ne codant pour aucune protéine fonctionnelle connue. Les données actuelles suggèrent l'existence de plusieurs dizaines de milliers de LncRNA chez l'homme incluant des pseudogènes, des antisens ou des LincRNA (long intergenic ncRNA). Cependant, très peu de ces transcrits (~1%) ont été caractérisés aujourd'hui [12].

En permettant le recrutement de ribonucléoprotéines régulatrices, les LncRNA contrôlent l'expression génique par divers mécanismes incluant notamment des modifications épigénétiques de l'organisation de la chromatine, l'activation ou la répression de la transcription en *cis* ou en *trans*, le contrôle de la stabilité des ARNm ou dans le cas de NEAT1 qui est induit par l'hypoxie, le contrôle de l'épissage nucléaire (Figure 1C) [3, 13, 21, 27].

De plus en plus de données suggèrent l'implication des LncRNA dans la physiopathologie du cancer [17, 28]. Cependant, leur rôle dans la réponse à l'hypoxie n'a jamais été étudié de manière globale et notamment dans les ADC. Dans ce contexte, mon équipe d'accueil a identifié plusieurs LncRNAs régulés par l'hypoxie dans ces tumeurs.

4. LincH3, un LncRNA induit par l'hypoxie dans les ADC

L'analyse comparative des profils de LncRNAs modulés par l'hypoxie, dans des biopsies d'ADC de stade précoce et dans différentes lignées d'ADC (A549, H1975, H2228) soumises à des conditions hypoxiques, a permis d'identifier une dizaine de LncRNA modulés par l'hypoxie *in vivo* et *in vitro*. Les chercheurs de l'équipe se sont focalisés sur LincH3 car ce transcrit est induit par l'hypoxie et n'est pas caractérisé dans la littérature (Figure 2A).

La caractérisation structurale de LincH3, par des techniques de RT-PCR, RNA-Seq, Northern blot et d'hybridation *in situ* ARN (FISH RNA), a montré qu'il s'agit d'un transcrit nucléaire de 10kb composé de 6 exons (Figure 2B). L'étude fonctionnelle a ensuite été abordée par une

approche d'inactivation en utilisant le système CRISPR/Cas9. La région génomique comprenant les exons 1 à 3 et 200 pb en amont du site d'initiation de la transcription a été déléetée afin d'inhiber l'expression du transcrit (Figure 2C). L'étude de l'impact de cette inactivation est en cours dans le laboratoire.

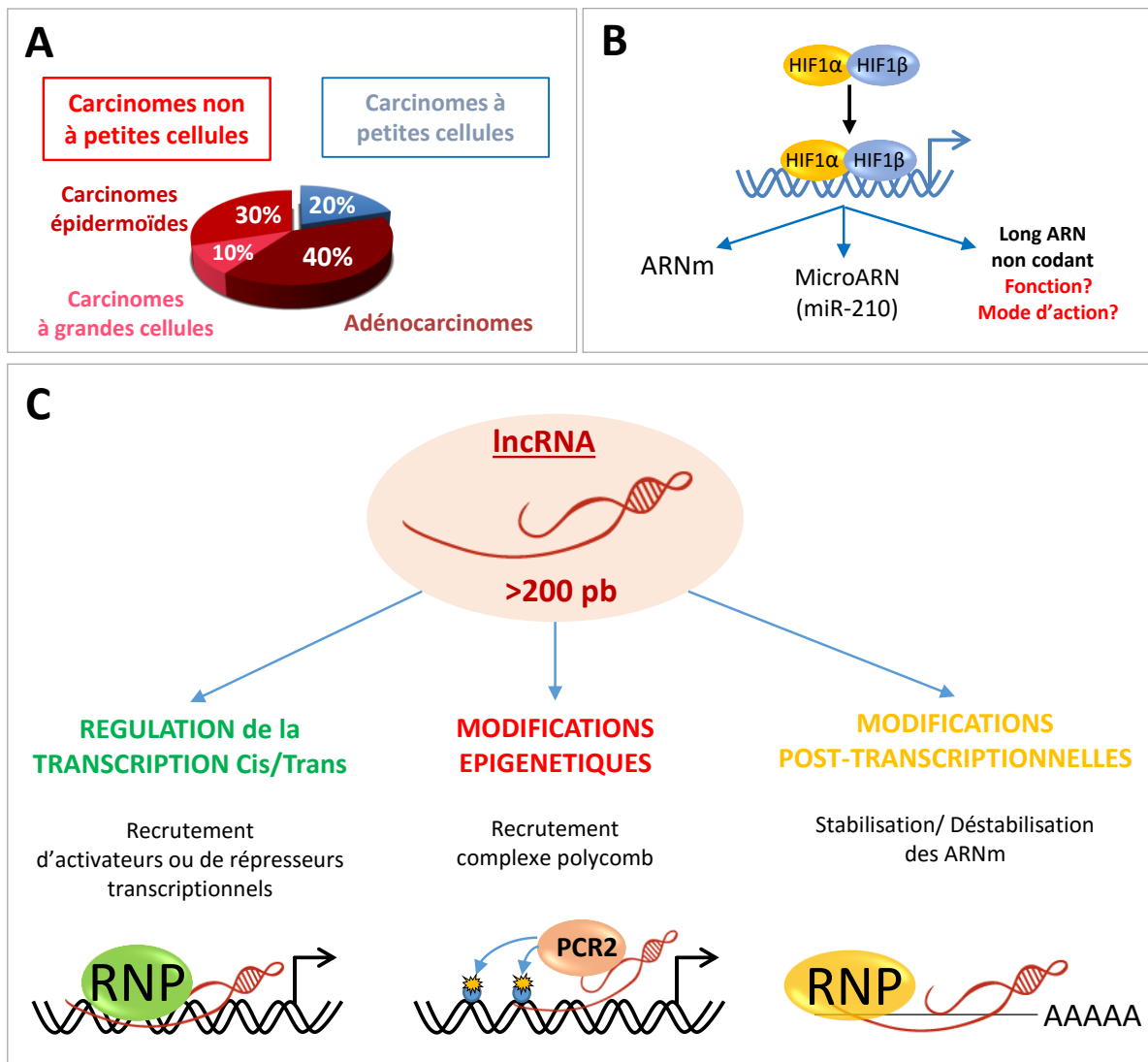


Figure 1 : Généralités

A. Diagramme de répartition histologique des cancers broncho-pulmonaires.

B. Schéma de régulation de l'expression génique par les facteurs de transcription HIFs.

C. Schéma représentant différents modes d'actions des longs ARN non codants intergéniques.

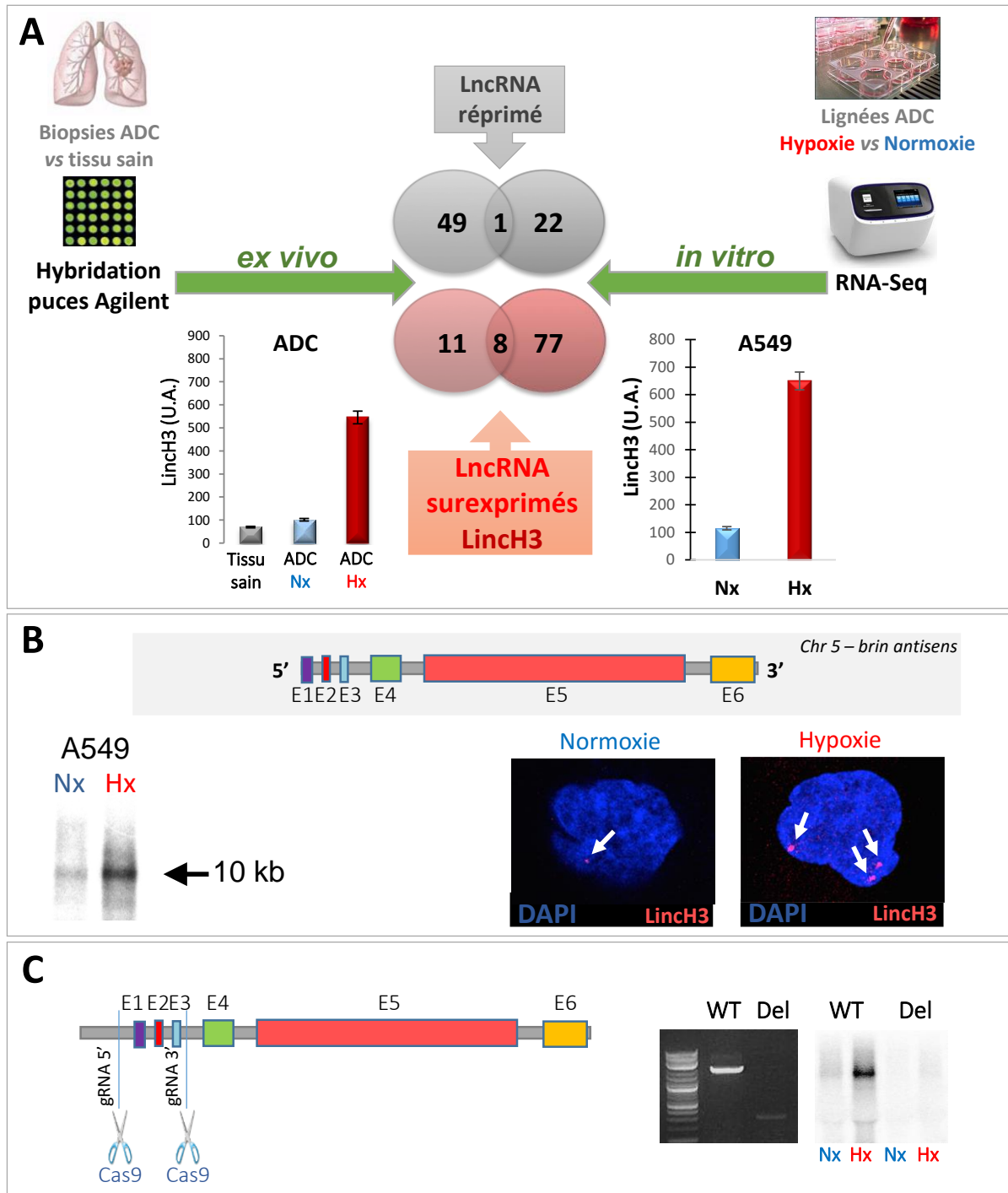


Figure 2 : Résultats préliminaires concernant le long ARN non codant : LincH3

A. Représentation du double crible transcriptomique réalisé sur les biopsies d'ADC et les lignées cultivées en hypoxie. Les histogrammes illustrent l'induction de LincH3 par l'hypoxie in vivo et in vitro.

B. En haut, schéma de la structure du transcrit Linch3. En bas, à gauche, Northern Blot du transcrit Linch3. En bas, à droite, RNA FISH du transcrit Linch3.

C. Schéma représentant la région délétée CRISPR/Cas9 sur le gène *Linch3*. A droite, analyse RT-PCR de la région délétée au niveau génomique et Northern Blot du transcrit Linch3 dans les clones WT et délétés.

OBJECTIFS

Dans le cadre de mon stage, je me suis intéressée à une problématique centrale dans le domaine des lncRNA, la caractérisation du mode d'action de Linch3 par l'identification de ses partenaires protéiques et surtout de ses gènes cibles (ADN ou ARN) dans les ADC pulmonaires.

Le but de mon projet a plus précisément consisté à lever un verrou technologique par la mise en place au sein du laboratoire d'une technique récente et encore peu utilisée : le ChIRP ou isolement de chromatine par purification d'ARN. Cette technique combine des approches biochimiques visant à précipiter les complexes macromoléculaires associés au transcrit Linch3 et des analyses à haut-débit de séquençage et/ou de protéomique par spectrométrie de masse.

Un second aspect de mon stage a été de recouper les informations issues des expériences de ChIRP et les données de transcriptomique obtenues après invalidation de Linch3 dans des cellules d'ADC en hypoxie, afin de relier le mode d'action de Linch3 à sa fonction physiologique.

MATERIELS ET METHODES

Culture cellulaire

Les cellules de la lignée d'ADC A549 proviennent de l'ATCC et sont cultivées en milieu DMEM (Dulbecco's Modified Eagle Medium, Life Technologies) supplémenté avec 10% de sérum de veau fœtal (SVF) et 2mM de L-glutamine (Life Technologies) dans un incubateur à 37°C et 5% de CO₂. Les traitements hypoxiques sont réalisés sur des cellules cultivées dans un incubateur à 1,2% d'oxygène. Les clones qui ont subi une délétion dans le gène *Linch3* par le système CRISPR/Cas9 sont cultivés dans les mêmes conditions.

ChIRP

Préparation de noyaux isolés : Les cellules (100×10^6) sont trypsinées, culotées par centrifugation à 4°C, lavées en PBS puis incubées 10min dans du PBS contenant 1% de formaldéhyde afin de créer des liaisons covalentes entre les différents partenaires moléculaires. La fixation est stoppée par incubation 5 min avec de la glycine 0,125M. Les cellules sont reprises dans un tampon hypotonique (KOAc 10 mM, MgOAc 10mM, NP40 1%, Tris/HCl 100mM pH 7.0, PMSF 1mM, SUPERaseIN 1X, Protease Inhibitor (PI) 1X) permettant leur lyse mécanique à l'aide d'un piston en verre de type "dounce", tout en préservant l'intégrité des noyaux. Les noyaux sont centrifugés à 2000xg, à 4°C durant 10 min, lavés deux fois dans 50mL de PBS puis fixés dans du PBS contenant 3% de formaldéhyde durant 30 min. L'action de formaldéhyde est stoppée par incubation 5 min avec de la glycine 0,125M.

Fragmentation de la chromatine : Les noyaux sont lysés dans un tampon Tris/Hcl 50mM pH 7.0 contenant SDS 1%, EDTA 10mM, PMSF 1mM, PI 1X, et SUPERaseIN 1X. La chromatine et les transcrits sont alors fragmentés (150-800 pb) à l'aide d'un sonicateur Covaris® E2 durant 9 minutes (Intensité 4, Duty cycle 5%, cycle/ burst 200) à 4°C. La chromatine est alors centrifugée 10min à 16000xg à 4°C afin d'éliminer les débris nucléaires.

Sondes ChIRP LincH3 : 44 sondes oligonucléotidiques antisens de 20 nt non recouvrantes ont été déterminées sur toute la longueur du transcrit LincH3 à l'aide d'un algorithme (Stellaris ChIRP Probe Designer, LGC Biosearch Technologies) prenant en compte la structure secondaire, le % en GC et la température d'hybridation (Figure 3A). Elles ont été séparées en deux lots afin de comparer les résultats obtenus avec chacun des lots et d'éliminer les interactions non spécifiques. Dans un premier temps, chaque lot a été marqué avec un fluorophore différent puis testé en hybridation in situ (FISH RNA), le co-marquage nucléaire

obtenu indiquant la spécificité d'interaction avec Linch3 (Figure 3B). Pour le ChIRP, les sondes ont été synthétisées sous la forme d'oligos ADN comprenant des modifications TEG-biotine en 3'. Des lots de sondes oligonucléotidiques biotinylées commerciales NEAT1 et LacZ (Millipore) ont été utilisés pour la mise au point du protocole.

Hybridation : 1mL de chromatine fragmentée est hybridée avec 30 nM de chaque lot de sondes biotinylées durant 4h à 37°C dans un tampon d'hybridation qui va dénaturer les transcrits et permettre la fixation des sondes (Tris/HCl 35 mM, pH 7.0, Formamide 10%, NaCl 500 mM, SDS 0,75%, EDTA 3 mM, PMSF 1mM, PI 1X, et SUPERaseIN 1X). On récupère ensuite les sondes associées à Linch3 par incubation 30 min à 37°C avec 100 µl de billes magnétiques couplées à de la streptavidine. Cela permet de récupérer également les partenaires moléculaires associés à Linch3. Après 5 lavages dans un tampon 2X SSC, SDS 0,5%, PMSF 1mM, les échantillons sont aliquotés pour extraction des ARN, de l'ADN et/ou des protéines.

Extraction des ARN : Les échantillons associés aux billes sont dé-protéinés dans un tampon Protéinase K (Tris/HCl 10mM, pH 7.0, NaCl 100mM, EDTA 1mM, 0,5% SDS, protéinase K 1mg/mL) durant 1h à 55°C puis les liaisons covalentes sont dénaturées en incubant 30 min à 65°C. La protéinase K est inactivée 10 min à 95°C. Ensuite on traite les échantillons avec de la DNase 1 pendant 30 min à 37°C et on inactive l'enzyme par de l'EDTA 0,5 mM pendant 10 min à 65°C. Pour l'extraction, les échantillons (billes, chromatine ou cellules) sont lysés dans 1 mL de Qiazol (Qiagen) et les ARNs sont extraits par ajout de 0,2 ml de chloroforme et centrifugation à 12 000 rpm pendant 15 min à 4°C. Les ARN contenus dans la phase aqueuse sont précipités par ajout d'un carrier (Glycoblue 50µg/mL) et d'isopropanol puis incubation 8h à -20°C. Après 1h de centrifugation à 12 000 rpm à 4°C, les culots sont rincés avec de l'éthanol à 70%, puis séchés et repris dans de l'eau "RNase free". La concentration des différents échantillons est déterminée par lecture au spectrophotomètre Nanodrop à 260 nm. La qualité des ARN est ensuite déterminée par électrophorèse capillaire microfluidique sur un Bioanalyzer (Agilent) à l'aide du kit RNA 6000 Pico Assay. Les échantillons ARN vont être analysés par RTqPCR.

Extraction de l'ADN associé à Linch3 : Le matériel associé aux billes est incubé 30 min à 65°C dans un tampon NaHCO₃ 50 mM, SDS 1% contenant de la RNase A 100 µg/mL afin de libérer les séquences d'ADN génomique co-précipitées avec le transcrit d'intérêt. Puis, les échantillons sont incubés 45 min à 55°C en présence de protéinase K afin d'éliminer les protéines liés à la chromatine. Pour l'extraction, les éluats sont lysés dans 300µl de phénol-

chloroforme-isoamyl puis centrifugés 15min à 16,100xg, à 4°C. La phase aqueuse est récupérée puis l'ADN est précipité à l'aide d'un carrier (acrylamide linéaire 0,5mg/mL), NaOAc 0,3M et d'éthanol 100% durant 8h à -20°C. Après 1h de centrifugation à 16,100xg, les culots sont lavés à l'éthanol 70% puis resuspendu dans de l'eau « RNase free ». Les échantillons ADN sont analysés en qPCR et en séquençage à haut débit.

Extraction des protéines partenaires : Pour l'analyse par Western blot des protéines associées à NEAT1, les culots de billes magnétiques sont repris dans du tampon SDS 1%, TAE 1X pH : 8,3 contenant de la RNase A 100µg/mL puis ils sont incubés 30min à 37°C afin de décrocher les protéines des transcrits. Puis, on ajoute du tampon Laemmli (Tris/HCl 62,5mM pH 6,8, glycérol 25%, 3% SDS) contenant du βmercaptoéthanol (350mM) que l'on incube à 95°C durant 1h afin de dénaturer et réduire les protéines. Ultérieurement, les protéines co-précipitées avec LinCH3 seront traitées afin de permettre leur identification par spectrométrie de masse.

Transcription réverse et PCR quantitative

Les amorces utilisées pour les réactions de PCR et de qPCR ont été élaborées à l'aide de l'outil Primer3plus (bioinformatics) (Figure 3C) et commandées chez IDT (Integrated DNA Technologies). Les ARN totaux sont rétro-transcrits à l'aide du kit High-Capacity cDNA Reverse Transcription (Applied Biosystems). Pour cela, on ajoute à 1 µg d'ARN totaux repris dans 10 µL d'eau, 2 µl de tampon de RT 10X, 2 µl de RT Random primers 10X, 1 µl de MultiScribe™ Reverse Transcriptase, 0,8 µl de dNTP 100 mM, 1 µl de RNase Inhibitor et 3.2 µl d'eau RNase free. La transcription réverse est ensuite réalisée dans un thermocycler (10 min d'annealing à 25°C, 2 h d'élongation à 37°C et 5 min de dénaturation à 85°C). L'analyse des niveaux d'expression des transcrits de *LinCH3*, *NEAT1*, *MALAT1*, *CA9*, *PGK1* et *RPLP0* par PCR quantitative a été réalisée à l'aide du réactif LightCycler® 480 DNA SYBR Green I Master (Roche). Brièvement, 10 ng de cDNA sont repris dans 5 µl d'eau. Puis on ajoute 12,5 µl de SYBR Green I Master Mix (2x), 300 nM d'amorce sens, 300 nM d'amorce antisens et 6 µl d'eau RNase free. La qPCR est réalisée à l'aide du système de détection LightCycler 480 (Roche). Chaque réaction est réalisée en triplicat et les niveaux d'expression calculés par rapport au gène de ménage *RPLP0* en utilisant la méthode de comparaison de C_T ($2^{-\Delta\Delta C_T}$). L'efficacité de précipitation des ARN et de l'ADN par ChIRP est calculée par l'équation: $2^{(Ct(matériel\ précipité) - \mu(Ct(ARN\ chromatine\ fragmentée)) - \log(dilution\ réelle))}$

La significativité des résultats a été établie par un test *t de Student* pour 2 conditions comparées et par un test d'ANOVA lorsque $n > 2$ et est représentée par * : $p < 0,05$; ** : $p < 0,01$.

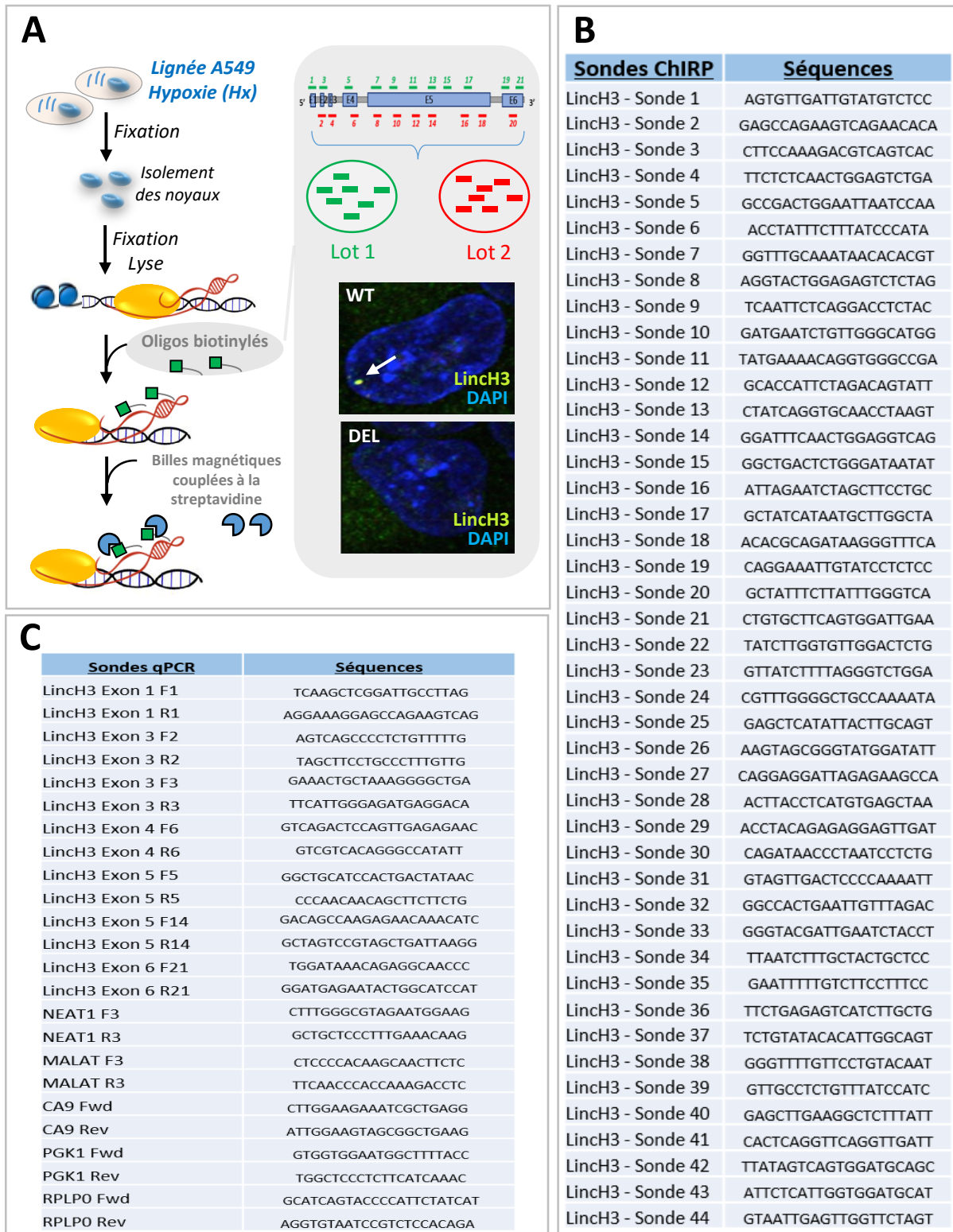


Figure 3 : Tableau récapitulatif des sondes de ChIRP et de qPCR

A. Schéma du protocole de ChIRP. **B.** Liste des sondes utilisées pour le ChIRP LincH3. **C.** Liste des amorces utilisées en qPCR.

Séquençage de l'ADN génomique associé aux LncRNA (DNA-Seq)

Création de banques d'ADN : Nous avons utilisé le kit diagenode® Microplex Library preparation V2 pour amplifier par PCR les séquences génomiques de l'échantillon et intégrer l'adaptateur (5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3') aux séquences puis les index (séquences connues de 8 pb) qui vont se fixer sur l'adaptateur et nous permettre de différencier les échantillons. Nous amplifions nos échantillons en réalisant 22 cycles (20s à 98°C, 50s à 72°C). Les banques obtenues sont analysées au bioanalyseur puis mixées (elles ont des index différents) et confiées à la plateforme génomique de l'IPMC pour être séquencées sur Illumina® NextSeq500. Le séquençage s'effectue avec 10µl de banque d'ADN à 2nM. Le séquenceur effectue 130 millions de lectures de 75 pb afin d'identifier les potentielles régions génomiques interagissant avec le transcrit LinCH3.

Analyse de séquences d'ADN : Les séquences obtenues sont « trimmées » ce qui signifie qu'on élimine de l'analyse les séquences correspondant à l'adaptateur et aux dimères d'index. Ensuite, nous alignons les séquences sur le génome humain de référence GRCh37/hg19 grâce au logiciel STAR. Enfin nous avons utilisé un outil de comparaison dCLIP afin d'observer les séquences enrichies dans nos échantillons WT par rapport aux échantillons DEL. Nous avons visualisé les locus d'intérêt grâce au logiciel IGV (Integrative Genomics Viewer) qui permet d'observer la superposition des séquences obtenues sur le génome de référence. Enfin, les séquences ont été soumises à l'outil BLAT© (UCSC) afin d'observer si elles sont répétées ailleurs dans le génome.

Analyse du transcriptôme par Hybridation de biopuces à ADN

Brièvement, 200 ng d'ARN totaux sont dénaturés et rétrotranscrits durant 2h à 40°C en présence d'amorce T7-Oligo(dT), de tampon 5X, de dNTP (10 mM), de DTT (0,1 M), de RNase inhibitor et de Reverse Transcriptase. La synthèse du second brin d'ADN s'effectue à 16°C pendant 2h par ajout de : tampon 10X, de dNTP, de RNase H et de d'ADN polymérase.

L'ajout de ce promoteur T7 va permettre le marquage des ARN par transcription *in vitro* 14 h à 37°C en présence d'ARN polymérase T7 et de CTP marqué avec un fluorophore, la cyanine 3 (Cy3). Les ARN marqués sont purifiés sur des colonnes (Kit RNeasy Mini, Qiagen) et quantifiés au spectrophotomètre. 750 ng d'ARN marqués au Cy3, sont mélangés avec 10µl de tampon de fragmentation 25X (Agilent) dans un volume final de 250 µL et incubés pendant 30 min à 60°C. 250 µl de tampon d'hybridation 2X (Agilent) sont ensuite ajoutés et ce mélange est déposé sur une lame de biopuce Agilent SurePrint G3 Human Gene Expression 8x60K v2

Microarray dédiée à l'analyse du transcriptome. Chaque lame est disposée dans une chambre d'hybridation et incubée durant 18 heures à 62°C. Les lames sont ensuite lavées, séchées et scannées avec une résolution de 3 µm à l'aide du « scanner C Agilent ». Le logiciel « Agilent scan control » permet l'acquisition des images et leur analyse par un processus en trois étapes : Les spots sont localisés par application d'une grille sur l'image, les données sont extraites puis segmentées et enfin, les images en format .tif ainsi que les fichiers de quantifications sont produits grâce au logiciel « Features extraction » afin de réaliser l'analyse statistique et la visualisation des résultats de l'expérience.

L'analyse statistique des résultats est réalisée à l'aide du package d'analyse limma disponible sous le logiciel "R" (<http://cran.rproject.org/>). Les données sont normalisées afin de rééquilibrer l'intensité du fluorochrome. Une normalisation inter-lame est effectuée par la méthode quantile afin d'homogénéiser la distribution des intensités entre les lames. L'analyse statistique des données est ensuite effectuée et permet d'obtenir une liste de gènes différentiellement exprimés à l'aide d'un score statistique dérivé d'un modèle linéaire utilisant une approche empirique bayésienne. Une P-Value ajustée (correction de Benjamini et Hochberg) < 0.05 a été utilisée comme seuil de significativité.

Western Blot

Afin d'observer la protéine PSPC1 partenaire de NEAT1 dans des échantillons nucléaires fragmentés ainsi que dans des échantillons provenant de la technique du ChIRP, nous avons séparé les échantillons par SDS-PAGE sur un gel d'acrylamide 10% sur la nuit à 12mA. Les protéines sont ensuite transférées sur une membrane de PVDF durant 3h à 400 mA. Après 3 lavages dans du TBST (Tris/HCl 20 mM pH 7.5, NaCl 150 mM, Tween 20 1%), la membrane est incubée sur la nuit avec l'anticorps anti-PSPC1 (Santa-Cruz) dilué à 1/100 dans du TBST contenant 5% de lait. Après lavages, puis incubation 1h avec un anticorps secondaire anti-lapin couplé à la peroxydase (1/5000), la membrane est à nouveau lavée 3 fois en TBST et enfin révélée par chimioluminescence puis détectée sur un appareil fusion-FX imager (Vilber Lourmat).

RESULTATS

1. Mise au point de la technique de ChIRP

Le premier aspect de mon projet a été la mise au point de la technique de ChIRP-Seq développée par le groupe de Howard Chang à l'Université de Stanford qui permet de cartographier les sites de liaison d'un LncRNA nucléaire à l'échelle du génome [3]. Cette approche complexe consiste dans un premier temps à "cross-linker" par fixation au formaldéhyde les complexes macromoléculaires, constitués du LncRNA, de ses ribonucléoprotéines partenaires et des régions d'ADN génomique auxquelles il se fixe (Figure 3A). Puis, après lyse des noyaux et fragmentation de la chromatine, les complexes sont hybridés avec deux lots de sondes oligonucléotidiques antisens biotinylées non recouvrantes et réparties le long du transcrit LinCH3. Après précipitation à l'aide de billes couplées à de la streptavidine, les séquences génomiques associées à LinCH3 sont extraites et identifiées par DNA-Seq. Pour plus de spécificité, nous avons fait une analyse différentielle des séquences précipitées à partir de cellules d'ADC sauvages ou invalidées pour LinCH3 par le système CRISPR/Cas9.

a) Vérification de la réponse hypoxique

La première étape du ChIRP est donc une étape de fixation des cellules au formaldéhyde. Elle permet i) de protéger les transcrits de la dégradation (LinCH3 est un transcrit induit par l'hypoxie qui a une demi-vie d'environ 30 min) et ii) de faire des liaisons covalentes entre les différents partenaires potentiels du LinCH3 constituant un complexe macromoléculaire. A titre d'exemple, le LncRNA HOTAIR s'associe au complexe multiprotéique polycomb PRC2 et à des séquences promotrices [13]. Cependant la fixation perturbe l'extraction des ARN et l'analyse par RT-PCR [4]. Afin de vérifier l'amplitude de la réponse hypoxique, nous avons donc prélevé des aliquots des différents types cellulaires utilisés pour les expériences de ChIRP avant fixation au formaldéhyde afin d'extraire les ARN totaux et mesurer l'expression génique par RT-qPCR (Figure 4A). Les résultats ont été normalisés par rapport à l'expression du gène *RPLP0* qui code pour une protéine ribosomale. On observe une induction de 3 à 5 fois de l'expression de LinCH3 par l'hypoxie dans la lignée A549 et dans le clone sauvage WT1, alors qu'il n'est pas détectable dans les clones Del1 et Del2, invalidés par CRISPR/Cas9. En revanche,

l'induction des gènes *CA9*, *PGK1* et *NEAT1* par l'hypoxie est comparable dans les différents types cellulaires et n'est pas affectée par l'inactivation de *LinCH3*.

b) Fragmentation

C'est la 2^{ème} étape du protocole. En effet, pour obtenir des conditions de précipitation optimales, la chromatine doit être cassée en fragments de 200 à 800 pb. Cette étape s'effectue par sonication et a nécessité un travail de mise au point car elle est variable selon les types cellulaires. Nous avons donc soumis nos échantillons à différents temps de sonication puis analysé la taille des fragments par électrophorèse microfluidique (Figure 4B). Les résultats montrent que 5 min de traitement aux ultrasons fragmente la chromatine en segments supérieurs à 1000 pb alors qu'une sonication de 15 min génère des fragments inférieurs à 200 pb. La fragmentation optimale est obtenue pour un sonication de 10 min.

Afin d'uniformiser les différentes conditions expérimentales nous avons ensuite quantifié l'ADN génomique extrait à partir de la chromatine fragmentée provenant des cellules d'ADC WT ou invalidées pour l'expression de *LinCH3* (Figure 4C). Dans ce but, nous avons amplifié par qPCR les gènes *LinCH3* et *NEAT1*, un autre lncRNA nucléaire modulé par l'hypoxie que nous avons utilisé pour la mise au point du protocole de ChIRP. Les résultats montrent peu de variations du matériel génomique entre les différentes préparations de chromatine utilisées pour réaliser les expériences de ChIRP.

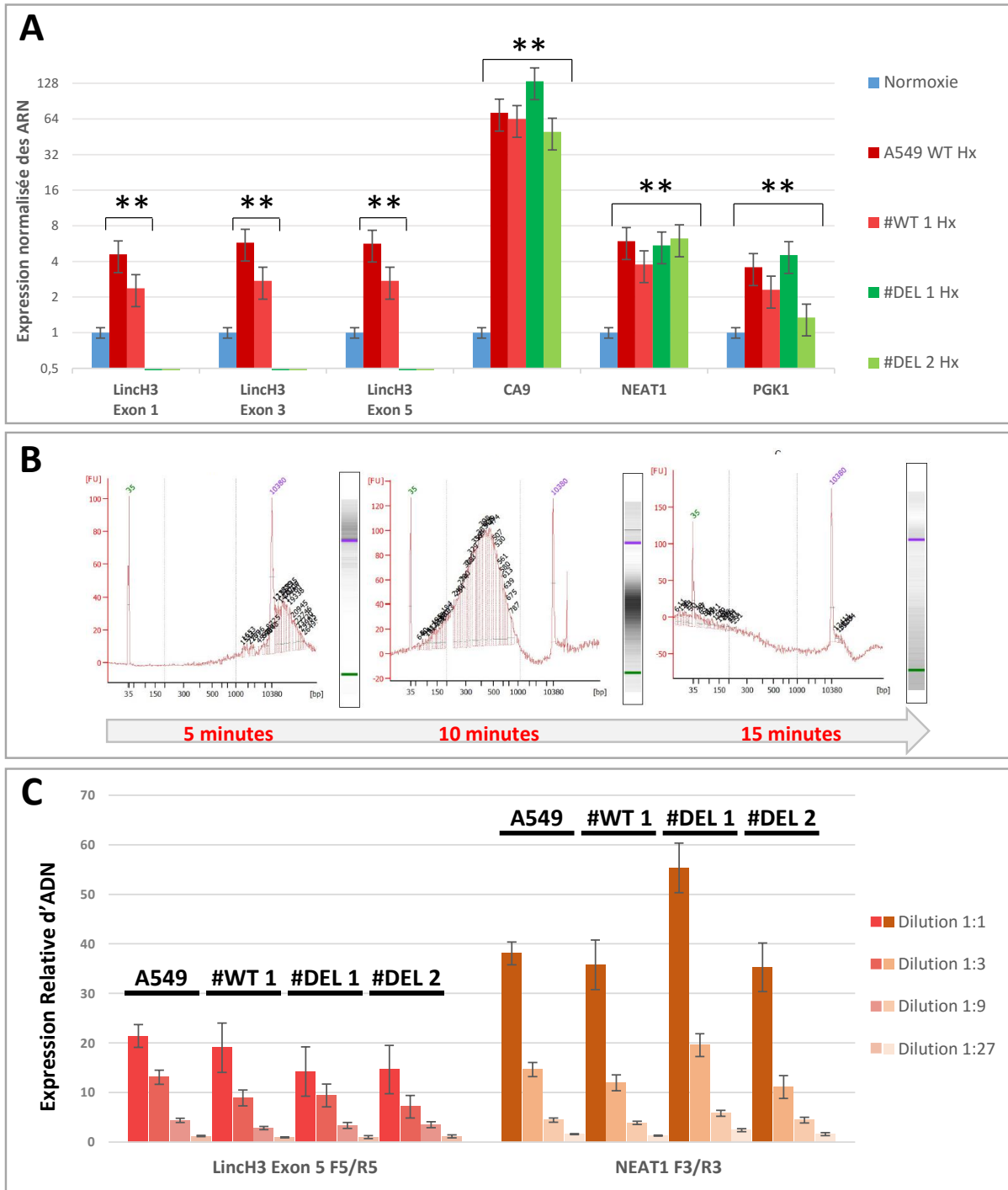


Figure 4 : Mise au point de la technique de ChIRP

- A.** Analyse de l'expression de LincH3, d'un autre LncRNA NEAT1 et de gènes régulés par l'hypoxie (*CA9*, *PGK1*) dans la lignée parentale A549 (n=3), un clone sauvage (WT1) (n=2) et deux clones CRISPR/Cas9 invalidés pour LincH3 (DEL1, DEL2) (n=2) au cours de la réponse hypoxique. Les données représentent les moyennes \pm SD de 3 expériences (* : $p < 0,05$; ** : $p < 0,01$).
- B.** Analyse de la fragmentation de la chromatine par sonication en fonction du temps.
- C.** Analyse génomique des loci LincH3 et NEAT1 extraits à partir de la chromatine fragmentée des différents types cellulaires utilisés pour la technique de ChIRP.

2. Validation de la précipitation des LncRNA par la technique de ChIRP

Une des étapes clé du protocole de ChIRP concerne la précipitation du transcrite cible par les sondes oligonucléotidiques biotinylées. Il est donc nécessaire de déterminer l'efficacité de cette capture. Nous avons souhaité dans un premier temps mettre au point cette technique, complexe en utilisant deux lots de sondes commerciales ciblant le LncRNA NEAT1. Nous avons également utilisé un lot de sondes dirigées contre le transcrite LacZ comme contrôle négatif. LacZ est un transcrite présent uniquement chez les procaryotes et donc absent de nos cellules eucaryotes. L'efficacité de la précipitation est déterminée en mesurant par RTqPCR les niveaux de LncRNA récupérés en fin de protocole (Figure 5A). Dans les échantillons de chromatine on mesure une forte expression des transcrits NEAT1 et MALAT, un autre LncRNA nucléaire utilisé ici comme contrôle négatif de précipitation. L'analyse des échantillons "ChIRP" NEAT1 montre que nous avons précipité spécifiquement ce transcrite, puisqu'aucune expression de MALAT n'est détectée. De même, les sondes biotinylées dirigées contre LacZ ne précipitent aucun des deux LncRNA testés. Nous avons ensuite quantifié l'efficacité de précipitation du transcrite NEAT1 par rapport à la quantité de matériel présente initialement dans la chromatine fragmentée. Les résultats montrent que le second lot de sondes précipite plus efficacement le transcrite NEAT1 (5,4% vs 2,4% du transcrite NEAT1 total) (Figure 5A, droite).

Nous avons ensuite testé la capture de notre transcrite d'intérêt LinCH3 par deux lots de sondes biotinylées que nous avons sélectionnées sur leur capacité à reconnaître spécifiquement LinCH3 dans des expériences de FISH (Figures 3A et 5B). Afin d'analyser l'efficacité de cette capture par qPCR, nous avons utilisé différents couples d'amorces qPCR dirigées contre les différents exons de LinCH3 (Figure 3C). La fixation au formaldéhyde peut fortement affecter l'extraction des ARN et leur analyse par RT-PCR ^[4]. Cependant, dans les conditions de "décross-link" que nous avons utilisées (1h de rupture des liaisons covalentes à 55°C, voir Matériels et méthodes), on mesure un niveau conséquent d'expression de LinCH3 dans les ARN totaux extraits à partir de la chromatine fragmentée (Figure 5B haut). Nous avons détecté une expression significative de LinCH3 dans les ChIRP réalisés sur la lignée A549 parentale incubée dans des conditions d'hypoxie mais bien évidemment aucune expression dans les échantillons provenant des cellules dans lesquelles LinCH3 a été invalidé.

La quantification de l'efficacité de capture de LinCH3 montre que selon le lot de sonde et/ou les essais qPCR utilisés, seulement 0,5 à 2% du transcrite est récupéré à l'issue du protocole ChIRP. De plus, l'analyse de l'expression de NEAT1 confirme la spécificité de la précipitation.

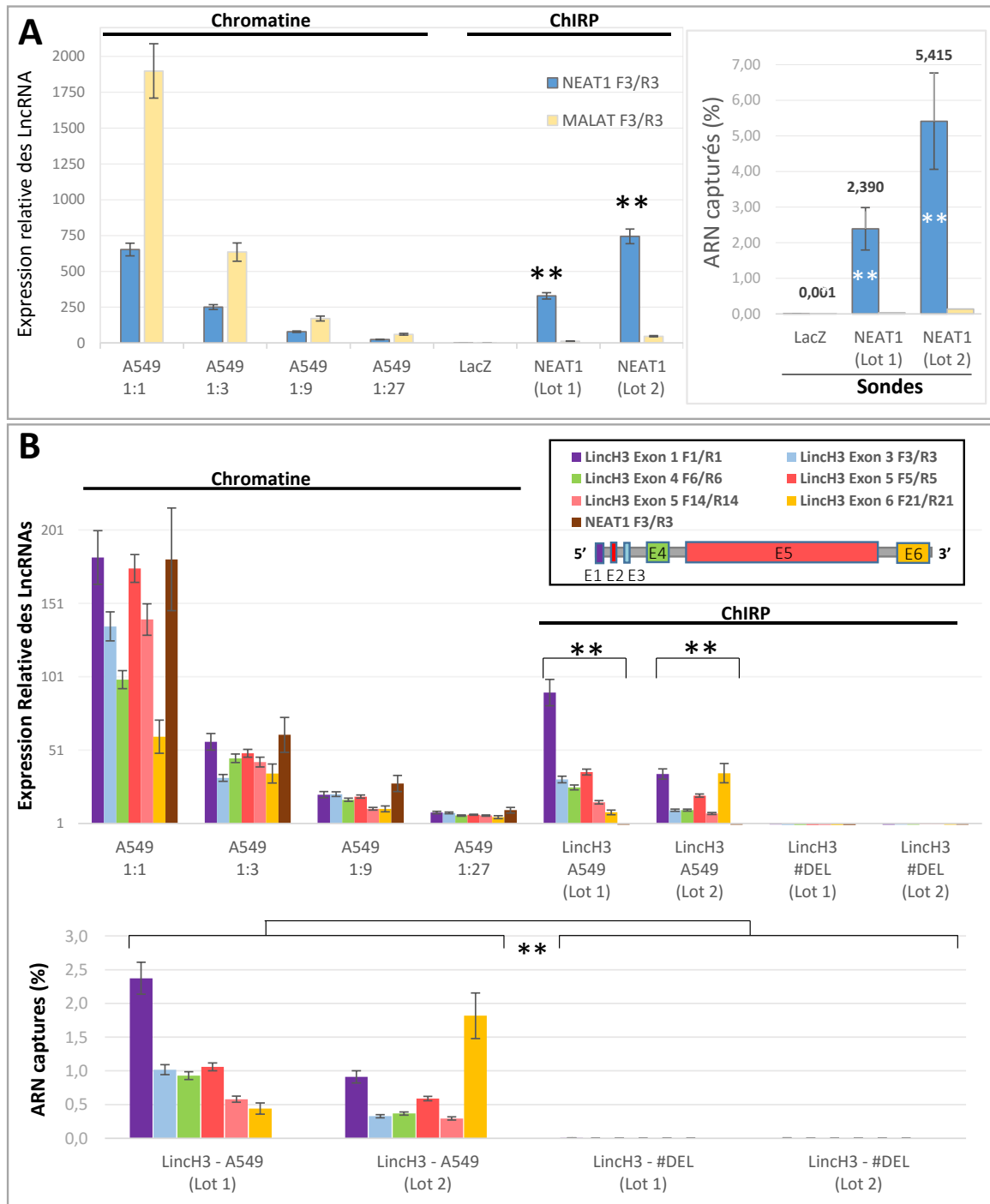


Figure 5 : Efficacité de capture des transcrits NEAT1 et LincH3 par la technique de ChIRP

A. Expérience de ChIRP NEAT1 sur cellules A549. Gauche: Amplification PCR de NEAT1 et MALAT à partir i) de différentes dilutions d'ARN extraits de la chromatine de cellules A549 et ii) des ARN issus d'une expérience de ChIRP NEAT1. Droite: Efficacité de capture des transcrits par ChIRP exprimé en % du matériel de départ présent dans la chromatine (* : $p < 0,05$; **: $p < 0,01$).

B. Expérience de ChIRP LincH3 sur cellules A549 WT et délétées cultivées en hypoxie. Haut: Niveaux de LincH3 présent dans i) différentes dilutions d'ARN extraits de la chromatine de cellules A549 et ii) dans les échantillons issus du ChIRP LincH3. Bas: Efficacité de capture de LincH3 et NEAT1 par ChIRP LincH3 exprimé en % du matériel de départ présent dans la chromatine. Des essais qPCR ciblant différents exons de LincH3 ont été utilisés (* : $p < 0,05$; **: $p < 0,01$).

3. Identification de séquences d'ADN génomique associées à LinCH3

a) Validation de la précipitation d'ADN génomique associées à LinCH3

Suite à ces résultats et étant donné la localisation nucléaire de NEAT1 et LinCH3 nous nous sommes prioritairement intéressés à l'identification des séquences d'ADN génomique avec lesquelles ces deux lncRNA pourraient interagir.

De récents travaux indiquent que le transcrit NEAT1 est capable de se fixer sur une région de son propre locus [4, 10]. Nous avons réalisé des PCR, sur l'ADN génomique extrait à partir des échantillons de ChIRP NEAT1 et LacZ, à l'aide d'amorces ciblant les gènes *NEAT1* et *MALAT* (Figure 6A). Les résultats montrent une amplification significative du gène *NEAT1* (mais pas de *MALAT*) dans le matériel précipité avec les deux lots de sondes NEAT1 correspondant respectivement à 0,57% et 0,64% du matériel de départ présent dans la chromatine. En revanche aucun signal n'est détecté dans les échantillons de ChIRP LacZ, montrant la spécificité de la précipitation. Ces résultats indiquent que les conditions expérimentales que nous avons utilisées dans les étapes de "cross-link", de fragmentation, d'hybridation, de lavages, de "décross-link" et d'extraction permettent de visualiser des associations des lncRNA avec des séquences d'ADN génomique.

Nous nous sommes ensuite intéressés aux séquences d'ADN génomiques associées à LinCH3. Aucune séquence d'ADN ciblée par LinCH3 n'étant déjà identifiée, nous avons dans un premier temps procédé comme pour NEAT1. Le matériel précipité par ChIRP, à partir de cellules sauvages ou invalidées cultivées en hypoxie, a été amplifié à l'aide d'amorces ciblant l'exon 5 de LinCH3, situé en dehors de la région délétée par CRISPR/Cas9 (Figure 6B). Dans les cellules A549 parentales (n=3) et dans les clones sauvages (n=2), la capture de séquences génomiques correspondant à une partie de l'exon 5 de LinCH3 représente de 0,25% à 0,38% du matériel de départ. Cet amplicon n'est pas détecté dans les échantillons de ChIRP LinCH3 réalisés sur les clones invalidés, démontrant que cette capture nécessite la présence du transcrit.

Nous avons ensuite effectué une recherche sans à priori des séquences pouvant interagir avec NEAT1 et LinCH3. Les fractions ADN extraites des différentes expériences de ChIRP (NEAT1 et LinCH3 sur cellules WT et DEL) ont été amplifiées pour constituer des banques génomiques qui ont été analysées par séquençage (Figure 6C). On observe que les fragments amplifiés correspondent à la gamme de taille de fragments obtenus par sonication de la chromatine. Le séquençage a été réalisé au sein de la plateforme génomique de l'IPMC.

b) Résultats des données de ChIRP-Seq

Nous avons débuté l'analyse des résultats de séquençage par la banque génomique NEAT1 (Figure 7A). Nous avons utilisé le navigateur de génome Integrative Genomics Viewer (<https://www.broadinstitute.org/igv/>) afin d'aligner les séquences obtenues sur le génome humain GRCh37/hg19. Nous avons focalisé sur le locus de *NEAT1* afin de vérifier les données obtenues par ChIRP-qPCR montrant l'association de NEAT1 avec son gène. On observe que des séquences génomiques précipitées avec le lncRNA NEAT1 correspondent au gène NEAT1 (Figure 7B). Inversement, comme en qPCR, aucun fragment séquencé ne correspond au gène *MALAT1*.

Les séquences obtenues à partir des banques génomiques des expériences de ChIRP LincH3 ont également été alignées sur le génome de référence (Figure 7B). Nous avons éliminé de l'analyse toutes les séquences répétées s'apparentant à du signal non-spécifique. Puis, à l'aide d'un programme de comparaison de séquences (dCLIP), nous avons exclu les séquences présentes dans les échantillons correspondant aux cellules délétées. Nous avons ainsi obtenu 254 séquences proches de différents gènes codants et non codants. Parmi ces séquences, 43% correspondent à des séquences introniques, 42% à des séquences intergéniques et 7% à des séquences promotrices (Figure 7C).

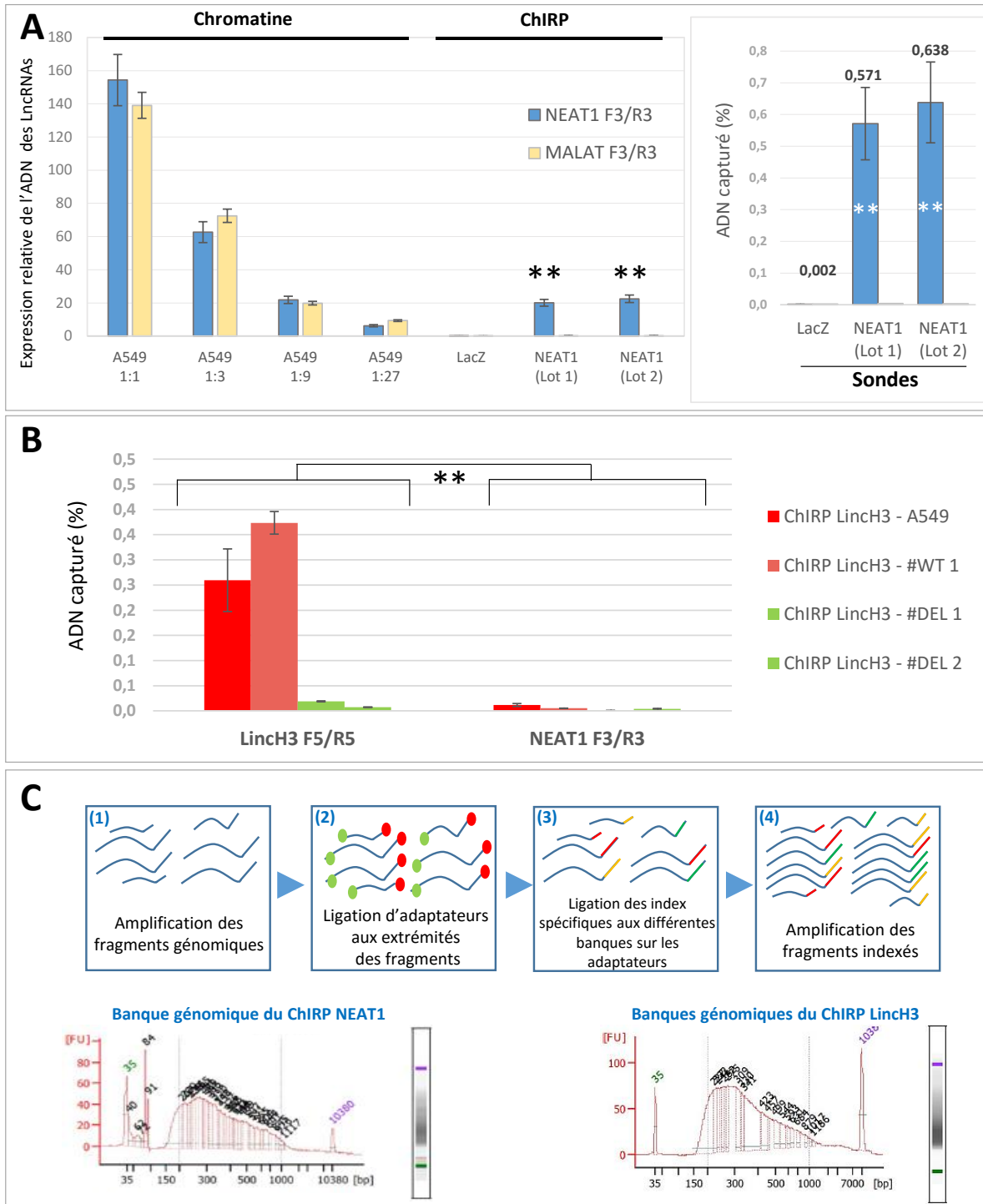


Figure 6 : Analyse de l'ADN génomique associé à NEAT1 et LincH3

A. Expérience de ChIRP NEAT1 sur cellules A549. Gauche: Analyse par qPCR de la quantité d'ADN des gènes *NEAT1* et *MALAT* à partir i) de différentes dilutions d'ADN génomique extrait de la chromatine de cellules A549 et ii) d'ADN génomique capturé par ChIRP NEAT1. Droite: Efficacité de capture de l'ADN du gène *NEAT1* par ChIRP exprimé en % du matériel de départ (* : $p < 0,05$; ** : $p < 0,01$).

B. Expérience de ChIRP LincH3 sur cellules A549 WT et délétées cultivées en hypoxie. Efficacité de capture de l'ADN des gènes *LincH3* (exon 5) et *NEAT1* par ChIRP LincH3 exprimé en % du matériel de départ (* : $p < 0,05$; ** : $p < 0,001$).

C. Haut: Schéma des étapes de création des banques génomiques. **Bas:** Profils électrophorétiques des banques génomiques issues des ChIRP NEAT1 et LincH3.

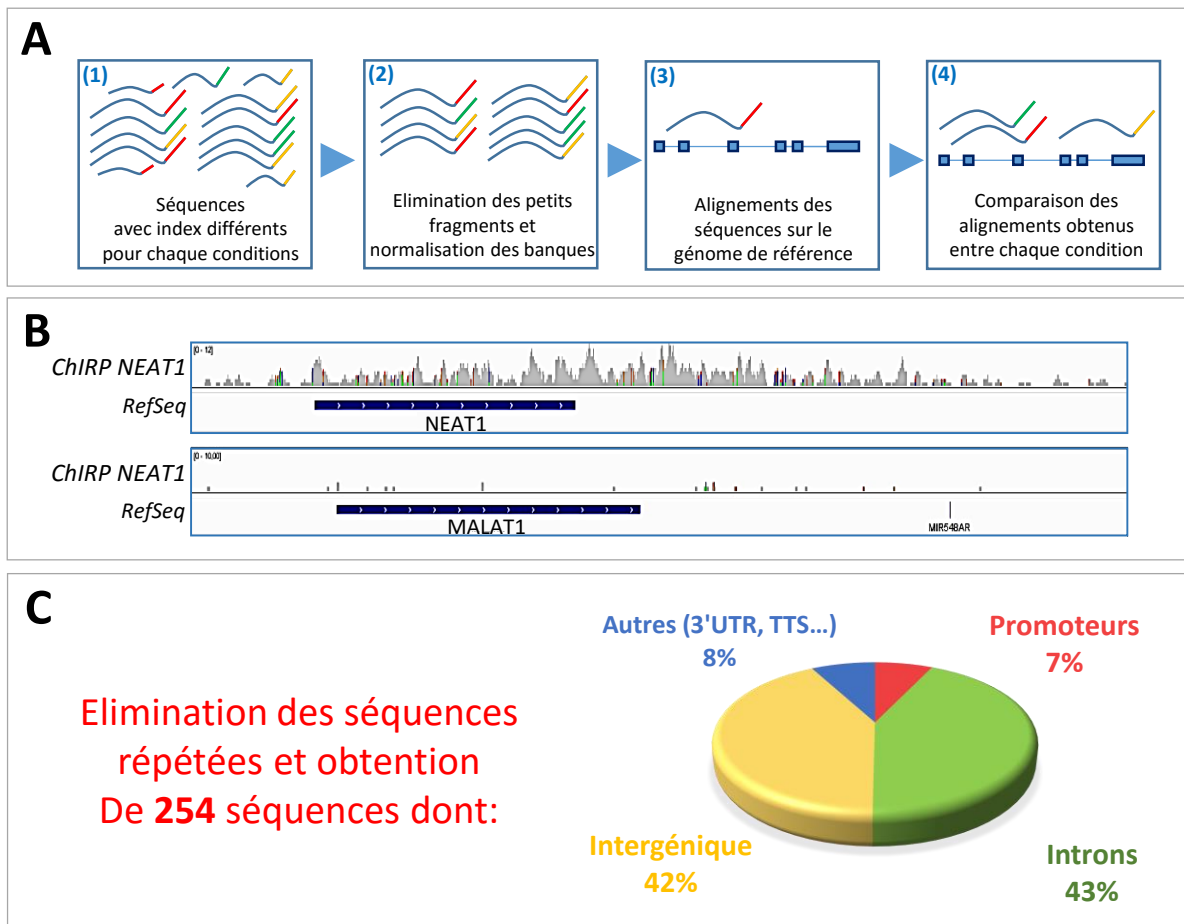


Figure 7 : Identification par DNA-Seq des séquences d'ADN capturées par ChIRP.

A. Schéma du protocole d'analyse des données de ChIRP-Seq. **B.** Résultats du séquençage de la banque génomique obtenu par ChIRP NEAT1 et observation de l'alignement des séquences obtenues sur les régions génomiques NEAT1 et MALAT. **C.** Résultats de séquençage des banques génomiques obtenues par ChIRP LincH3 sur les cellules A549 WT et invalidées en conditions hypoxiques. Le diagramme montre la répartition des séquences obtenues en fonction de leur position par rapport aux gènes.

4. Impact de l'inactivation de Linch3 sur le transcriptome de cellules d'ADC au cours de la réponse hypoxique

Afin de préciser la fonction de Linch3 dans les cellules d'ADC nous avons réalisé une analyse comparative de la signature d'expression génique de clones de cellules A549 WT ou invalidés pour l'expression de Linch3 par le système CRISPR/Cas9. Les cellules ont été cultivées en conditions normoxiques ou hypoxiques pendant 24h et les ARN hybridés sur des puces Agilent comme détaillé dans la section "matériels et méthodes".

La Figure 8A représente la classification hiérarchique des 449 gènes les plus affectés par l'absence d'expression de Linch3 ($\log_2(\text{intensité}) > 5$; p. value ajustée $< 0,05$; $|\log_2(\text{DEL}/\text{WT})| > 0,7$). On observe une bonne dichotomie entre les échantillons WT et DEL, indiquant que l'absence d'expression de Linch3 a des conséquences significatives sur le profil d'expression génique des cellules d'ADC. Le diagramme de Venn montre qu'une partie de la réponse hypoxique est altérée dans les cellules invalidées pour Linch3. De plus, l'annotation fonctionnelle des 449 gènes, par le logiciel IPA (Ingenuity Pathway Analysis), a révélé une exacerbation des voies TGF β et NF κ B dans les cellules déficientes pour Linch3 par rapport aux cellules sauvages (Figure 8B). A l'inverse, la voie anti-oxydante NRF2 semble réprimée dans les clones invalidés. Ces données suggèrent que le Linch3 pourrait être un important régulateur de l'expression génique par un mécanisme d'action impliquant d'une part la répression des voies de signalisation contrôlées par le TGF β et NF κ B, et d'autre part l'activation de voies régulées par NRF2.

5. Comparaison des résultats de transcriptomique et de ChIRP-Seq

Nous avons ensuite comparé ces résultats de transcriptome avec nos résultats de ChIRP-Seq (Figure 8C). Parmi les 254 séquences identifiées par ChIRP-Seq, 11 sont localisées dans des régions proches de gènes modulés par l'inactivation stable de Linch3, 8 étant induits (**CTGF** (Connective Tissue Growth Factor), **IGFBP3** (Insulin-like Growth Factor-Binding Protein 3), **MTSS1** (Metastasis Suppressor 1), **LTBP1** (Latent-Transforming growth factor Beta-binding Protein 1), **PTPRJ** (Protein Tyrosine Phosphatase Receptor type J), **RPL23AP32** (Ribosomal protein L23a Pseudogene 32), **FBN2** (Fibrillin 2), **ZSWIM6** (Zinc Finger SWIM-type containing 6)) et 3 étant réprimés (**MREG** (Melanoregulin), **BCL2L1** (B-cell CCL/Lymphoma 2 Like 1) et **LOC100422737** (ncRNA non caractérisé)). Le positionnement de ces séquences associées à Linch3 montre qu'elles correspondent au promoteur de *CTGF*, à des séquences introniques des gènes *BCL2L1*, *LTBP1*, *PTPRJ*, *RPL23AP32*, *FBN2*, *ZSWIM6* et *LOC100422737*, et à des séquences intergéniques proche des gènes *IGFBP3*, *MTSS1* et *MREG*.

Afin d'approfondir les résultats de cette première expérience de ChIRP-Seq, nous avons regardé la position exacte des séquences génomiques des gènes *CTGF*, *IGFBP3* et *BCL2L1* associés à Linch3 (Figure 9A). L'alignement de ces séquences sur le génome, montre qu'elles se trouvent respectivement à 200 pb du site d'initiation de la transcription de *CTGF*, dans l'intron 2 de *BCL2L1* et à 60 kb d'*IGFBP3*. Nous avons recherché si ces séquences étaient répétées ailleurs dans le génome, et donc non spécifiques de ces gènes. L'utilisation de l'outil BLAT© dans le navigateur de génome UCSC a confirmé qu'il s'agit bien de trois séquences uniques et spécifiques. L'analyse des niveaux d'expression des ARNm de ces gènes dans des cellules sauvages ou déficientes pour Linch3, en normoxie et en hypoxie, a mis en évidence une induction de *CTGF* et *IGFBP3* en absence de Linch3 (Figure 9B). Au contraire, *BCL2L1* est réprimé dans les cellules invalidées. L'analyse de la modulation des 8 autres gènes a été effectuée (résultats non présentés) et montre qu'à l'exception d'*IGFBP3*, et en accord avec la Figure 8A, l'expression de ces gènes est peu affectée par l'hypoxie.

Parmi ces différentes séquences régulatrices, à valider dans des expériences ultérieures, la piste CTGF apparaît particulièrement intéressante, ce facteur extracellulaire interagissant avec de nombreuses cytokines, facteurs de croissance, protéines de la matrice et récepteurs et modulant un grand nombre de voies de signalisation incluant la voie du TGF- β et l'angiogenèse [5, 24, 25].

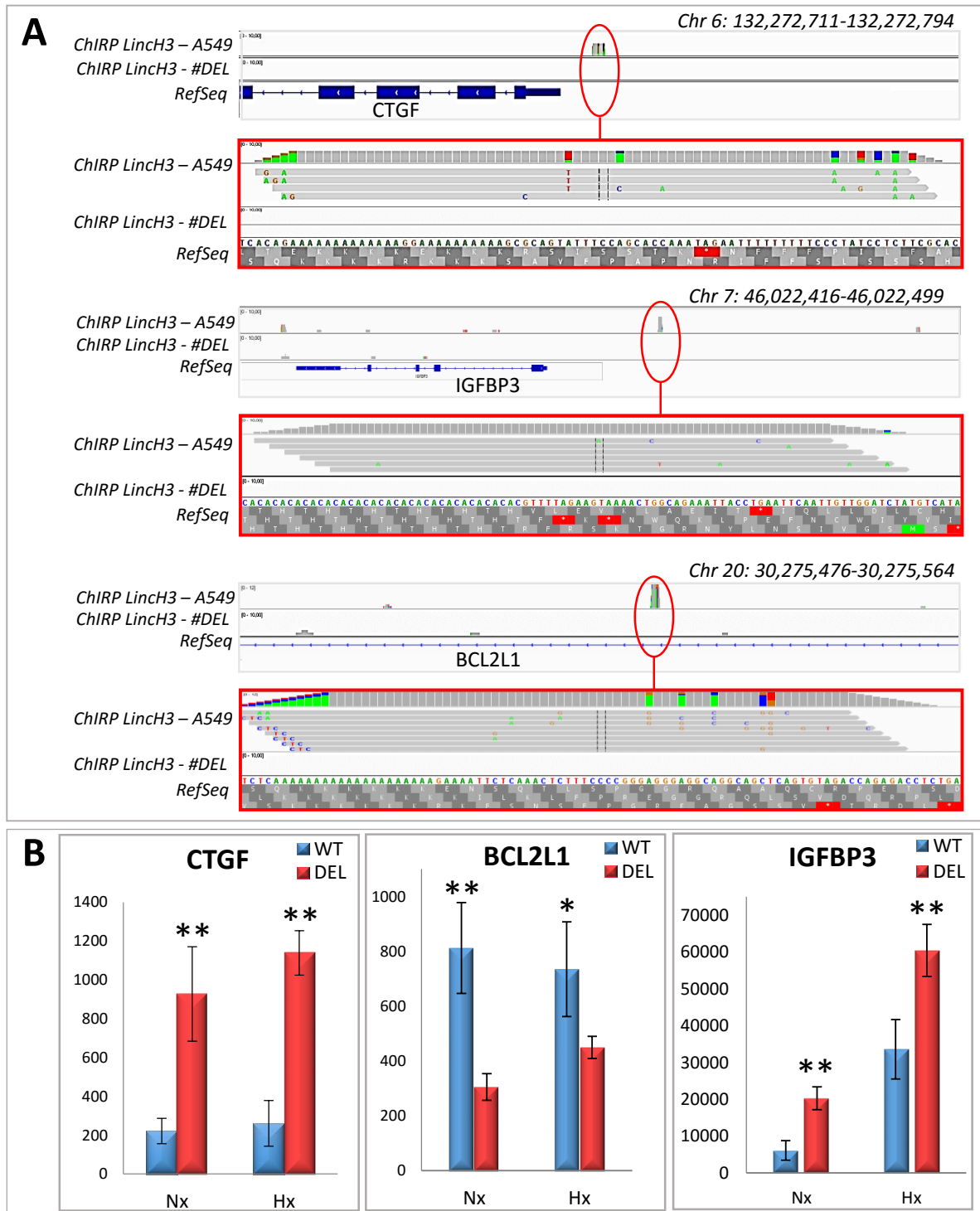


Figure 9 : CTGF, IGFBP3 et BCL2L1, trois gènes cibles de LincH3 ?

A. Alignement des séquences génomiques obtenues par ChIRP-Seq LincH3 sur les régions correspondant au promoteur du *CTGF*, à la région intergénique proche d'*IGFBP3* et à l'intron 2 de *BCL2L1* dans les cellules A549 sauvages.

B. Niveau d'expression des transcrits *CTGF*, *IGFBP3* et *BCL2L1* obtenus par hybridation de puces à ADN. Les histogrammes représentent la moyenne \pm SD des niveaux d'expression de 6 échantillons de cellules WT et DEL (* : $p < 0,05$; ** : $p < 0,01$).

6. Etude préliminaire de recherche des partenaires protéiques des LncRNA

Parallèlement aux analyses de ChIRP-Seq, nous avons débuté quelques expériences préliminaires dans le but d'aborder ultérieurement une étape clé de l'étude du mode d'action des LncRNA, l'identification des ribonucléoprotéines avec lesquelles ils s'associent et qui portent l'activité enzymatique des complexes.

Pour cela, nous avons réalisé des expériences de ChIRP à l'aide des sondes commerciales dirigées contre NEAT1 puis nous avons analysé par Western Blot l'expression d'une protéine précédemment identifiée pour s'associer à NEAT1, la protéine PSPC1 [\[1, 10\]](#).

A l'aide d'un anticorps spécifique, nous avons observé la présence de cette protéine dans la chromatine fragmentée et dans les échantillons de ChIRP NEAT1 (Figure 10). Cependant, l'analyse de l'expression de la protéine dans le surnageant du ChIRP montre que dans nos conditions, seule une petite fraction de PSPC1 est associée à NEAT1.

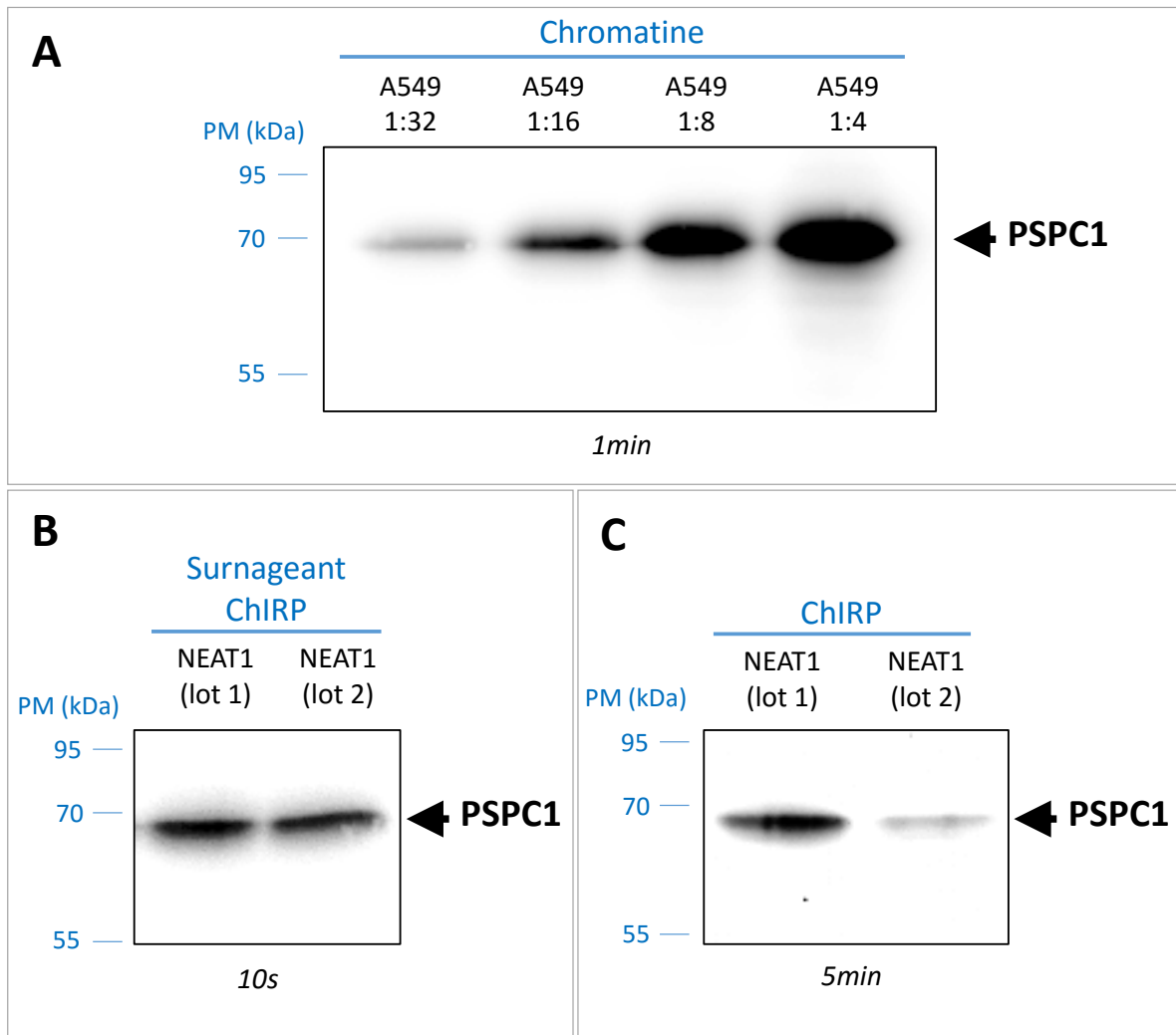


Figure 10 : Précipitation de la protéine PSpC1 par ChIRP-NEAT1

Analyse par Western blot de l'expression de la protéine PSpC1 dans:

A. 50 μ l de différentes dilutions de chromatine fragmentée de cellules A549.

B. 50 μ l de surnageant de ChIRP NEAT1.

C. des échantillons de ChIRP NEAT1.

DISCUSSION

La première étape de mon projet a été de lever un verrou technologique en mettant en place la technique de ChIRP afin de cartographier les sites de liaison de Linch3 sur le génome.

Après diverses mises au point techniques, nous avons précipité les complexes macromoléculaires associés à NEAT1 et Linch3 à l'aide de différents lots de sondes biotinylées puis nous avons réalisé des expériences de RTqPCR afin de quantifier la capture des transcrits d'intérêts. Les sondes ont spécifiquement précipité les LncRNA qu'elles ciblaient mais la quantité de transcrits capturés est faible que ce soit pour NEAT1 ou Linch3 (Figure 5). En ce qui concerne NEAT1, la quantité de transcrit capturée est toutefois comparable à celle obtenue par une autre équipe avec un protocole de capture appelé CHART (Capture Hybridization Analysis of RNA Targets) utilisant des oligos plus longs ^[10]. Nous avons émis deux hypothèses pouvant expliquer ces résultats: i) soit une partie des sondes utilisées n'accède pas au transcrit ciblé par gêne stérique (protéine associée au LncRNA, structure secondaire des LncRNA...), ii) soit la dissociation des complexes (décross-link) et l'extraction des acides nucléiques n'est pas assez efficace. Actuellement, nous mettons au point de nouveaux protocoles visant à dénaturer les liaisons covalentes créées par le formaldéhyde afin de récupérer plus de matériel.

Par la suite, nous avons analysé la capture de séquences génomiques par ChIRP. Comme précédemment décrit dans la littérature, nous avons mis en évidence la capture du gène *NEAT1* par le ChIRP NEAT1 (Figure 6A) ^[1, 10]. De façon similaire on observe la capture de *Linch3* par ChIRP Linch3 (Figure 6B). En essayant de comprendre les raisons de ces captures, nous nous sommes demandé s'il ne s'agissait pas d'un artéfact de la technique car les sondes antisens biotinylées pourraient directement précipiter des fragments génomiques complémentaires. Cependant, nous n'avons observé aucune capture de l'exon 5 de *Linch3* dans les expériences de ChIRP Linch3 sur cellules invalidées (DEL) (Figure 6B). Ces cellules ont une délétion des exons 1,2 et 3, mais elles ont conservé la séquence génomique des exons 4, 5 et 6 de *Linch3*. La capture du gène dépend donc de la présence du transcrit. Deux hypothèses subsistent: i) Linch3 et NEAT1 se fixent sur leurs propres séquences génomiques afin de s'autoréguler (transcription ou épissage), ii) La fixation au formaldéhyde crée des

liaisons covalentes entre les différents acteurs de la machinerie transcriptionnelle (gène, LncRNA néo-synthétisé et protéines).

Après avoir visualisé les fragments d'ADN dans nos échantillons ChIRP, nous avons construit des banques génomiques (Figure 6C). La quantité d'ADN génomique capturée étant faible, la complexité des banques amplifiées, c'est à dire la variété des séquences présentes dans les banques, est également faible. Là encore, l'amélioration des conditions de "décross-link" et du rendement d'extraction des acides nucléiques devrait permettre d'obtenir des banques de plus grande complexité afin d'identifier plus de sites de liaisons de Linch3 sur le génome. Dans ces conditions, l'alignement des séquences issues du ChIRP-Seq NEAT1 sur le génome a confirmé les résultats de qPCR et ceux obtenus par d'autres équipes ayant utilisé l'approche CHART montrant une association avec le gène *NEAT1* (Figure 7B) [1, 10].

L'analyse des séquences génomiques capturées par ChIRP Linch3 dans les cellules WT et absentes dans les cellules invalidées a permis d'identifier 254 sites potentiels de liaison de ce lncRNA sur le génome (Figure 7C). Certaines séquences sont localisées dans des régions promotrices (7%) ou intergéniques (47%) suggérant que Linch3 pourrait réguler la transcription de ces gènes au niveau de promoteurs ou de séquences régulatrices éloignées (ex : enhancer, silencer). D'autres séquences sont localisées au niveau d'introns (43%) et pourraient participer au contrôle de l'épissage de ces transcrits.

Parallèlement à l'identification des sites de liaison sur le génome, j'ai essayé de préciser la fonction de Linch3 dans les cellules d'ADC. Dans ce but, j'ai comparé les profils d'expression génique des cellules WT et des cellules invalidées pour Linch3 en normoxie et en hypoxie par l'analyse du transcriptome (Figure 8A/8B). Les résultats ont mis en évidence une exacerbation des voies TGF β et NF κ B et inversement une diminution de la voie anti-oxydante NRF2 dans les cellules déficientes pour Linch3. Ces données sont en accord avec d'autres études fonctionnelles réalisées au sein de l'équipe qui montrent que les cellules délétées ont: i) une augmentation de l'expression de cytokines pro-inflammatoires (IL1A, IL1B, IL6) régulées par NF κ B, et ii) une sensibilité accrue à l'apoptose induite par le cisplatine, une drogue chimiothérapeutique qui induit un stress oxydatif. Il est également intéressant de noter que NF κ B et NRF2 régulent l'expression de Linch3 [8]. Nos résultats suggèrent que Linch3 pourrait exercer un rétrocontrôle sur les voies contrôlées par ces facteurs de transcription. De plus, en accord avec la signature génique montrant une augmentation de la voie TGF β , les cellules

invalidées pour Linch3 ont un phénotype mésenchymateux alors que les cellules WT sont plus épithéliales suggérant que l'absence de Linch3 favorise une transition épithélio-mésenchymateuse (TEM).

La comparaison de la listes des gènes altérés par l'absence de Linch3 avec celle des gènes capturés par ChIRP Linch3 a permis d'identifier 11 gènes cibles potentiels de ce LncRNA (Figure 8C). Les gènes *CTGF*, *IGFBP3*, *MTSS1*, *LTBP1*, *PTPRJ*, *RPL23AP32*, *FBN2*, *ZSWIM6* sont induits tandis que les gènes *MREG*, *BCL2L1* et *LOC10042273* sont réprimés en l'absence de Linch3. Parmi ces 11 gènes, 6 sont liés à la voie TGF β , 1 à la voie NRF2 et les 4 restants n'ont pas de fonction bien caractérisée. Ainsi, CTGF est induit par la voie TGF β 1 et participe à la formation de métastases dans les cancers hépatiques et à la TEM dans les carcinomes pulmonaires [5, 24, 25]. De même, IGFBP3 semble avoir un lien avec la formation de métastases dans les cancers œsophagiens [23]. Enfin, des études indiquent que BCL2L1 serait induit par NRF2 et aurait des fonctions anti-apoptotique et dans la résistance aux traitements anti-tumoraux [26, 29].

Cependant, la démonstration d'un lien de causalité entre la liaison de Linch3 sur des séquences génomiques situées à l'intérieur (intronique) ou à proximité (promotrice, intergénique) de ces gènes et la dérégulation de leur expression reste toutefois à établir.

CONCLUSIONS ET PERSPECTIVES

1. Conclusion

Au cours de mon stage, nous avons pu mettre en place le ChIRP, une technique récente permettant de cartographier les sites de liaison des lncRNA sur le génome. Nous avons également caractérisé l'impact de l'inactivation de LinCH3 par CRISPR/Cas9 sur le transcriptome de cellules d'ADC au cours de la réponse hypoxique. Ces résultats nous ont permis d'appréhender la fonction de LinCH3 en mettant en évidence des modulations des voies de signalisation TGF β , NF κ B et NRF2 en absence de LinCH3. Ces données, étayées par d'autres résultats, suggèrent que LinCH3 pourrait réprimer la TEM, l'inflammation et être une intéressante cible thérapeutique.

L'analyse croisée de ces deux approches de criblage à grande échelle a permis l'identification de 11 gènes cibles potentiels de LinCH3, dont six sont liés à la voie de signalisation du TGF β et un à la voie NRF2. Parmi ces gènes cibles potentiels déjà identifiés, le gène *CTGF* apparaît comme une cible particulièrement intéressante. En effet, LinCH3 se fixe au niveau de son promoteur et ce gène est fortement lié à la voie de signalisation du TGF β .

Nous sommes toutefois conscients que des améliorations du protocole ChIRP-Seq seront encore nécessaires afin d'identifier d'autres cibles et mieux appréhender le mode d'action et la fonction de LinCH3. Dans cette optique, l'identification des protéines s'associant à LinCH3 sera également une étape déterminante dans la compréhension du mécanisme d'action de ce lncRNA.

2. Perspectives

a) Recherche des partenaires protéiques de LinCH3

L'identification des ribonucléoprotéines associées à LinCH3 sera abordée par deux types d'approches :

La première consistera à faire des recherches ciblées en utilisant notamment le protocole de ChIRP-Western Blot mis au point durant mon stage (Figure 10). Dans le cas de LinCH3, nous nous focaliserons sur les acteurs majeurs des voies NF κ B, TGF β et NRF2. Nous utiliserons donc des anticorps ciblant les protéines NF κ B (RelA...) ou leurs inhibiteurs I κ B, les

facteurs de transcription SMAD régulés par le TGF β ou le facteur de transcription NRF2. Le cas échéant, dans un second temps, nous validerons les résultats obtenus en utilisant la technique de RIP (RNA-binding protein Immunoprecipitation) qui consiste à récupérer les complexes macromoléculaires à l'aide d'anticorps ciblant les partenaires protéiques puis à vérifier la présence de LinCH3 par RT-qPCR [16, 21].

L'autre approche consistera à faire une recherche sans a priori en utilisant du ChIRP couplé à de l'analyse protéomique par spectrophotométrie de masse [30]. Une alternative de type "pull-down" consistera à incuber des fragments de LinCH3 biotinylés ou tagués MS2 avec des extraits nucléaires puis à identifier les protéines associées par spectrométrie de masse [6].

b) Etude fonctionnelle des protéines associées à LinCH3

Nous utiliserons des approches classiques de gain (expression ectopique) et de perte de fonction (siARNs, inhibiteurs pharmacologiques) ciblant les protéines partenaires identifiées afin de reproduire ou de corriger, dans des cellules WT ou déficientes, les conséquences de l'invalidation de LinCH3 sur le transcriptome et sur différents processus cellulaires régulés par ce transcrit. La fonction des gènes cibles potentiels de LinCH3 pourra également être étudiée par les mêmes approches. Il sera notamment intéressant de voir si des siRNA dirigés contre le CTGF sont capables de corriger, au moins partiellement, les dérégulations observées dans les cellules invalidées pour LinCH3.

Par la suite, nous pourrons déterminer la relevance physiopathologique de ces interactions en co-marquant, le transcrit LinCH3 par hybridation in situ et les protéines partenaires par immuno-marquage, dans des coupes de tumeurs pulmonaires.

c) Recherche des sites d'interaction de LinCH3 avec ses partenaires

Afin de préciser les régions de LinCH3 qui interagissent avec les séquences génomiques et les protéines partenaires précédemment identifiées nous utiliserons une variante du ChIRP, le domain-ChIRP [15]. Les lots de sondes antisens seront redistribués afin de cibler différentes régions ou domaines du LncRNA et la capture des partenaires sera analysée par DNA-Seq ou Western-blot.

Des expériences de "pull-down" à l'aide de fragments de LinCH3 biotinylés ou tagués (synthétisés par transcription *in vitro*) permettront de valider ces associations et seront complétées par la mutagénèse des sites identifiés.

BIBLIOGRAPHIE

1. Simon, M.D., Wang, C.I., Kharchenko, P.V., West, J.A., Chapman, B.A., Alekseyenko, A.A., Borowsky, M.L., Kuroda, M.I., and Kingston, R.E. (2011). **The genomic binding sites of a noncoding RNA**. *Proc. Natl. Acad. Sci. U.S.A.* *108*, 20497–20502.
2. Puisségur, M.-P., Mazure, N.M., Bertero, T., Pradelli, L., Grosso, S., Robbe-Sermesant, K., Maurin, T., Lebrigand, K., Cardinaud, B., Hofman, V., et al. (2011). **miR-210 is overexpressed in late stages of lung cancer and mediates mitochondrial alterations associated with modulation of HIF-1 activity**. *Cell Death Differ.* *18*, 465–478.
3. Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). **Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions**. *Mol. Cell* *44*, 667–678.
4. Evers, D.L., Fowler, C.B., Cunningham, B.R., Mason, J.T., and O’Leary, T.J. (2011). **The effect of formaldehyde fixation on RNA: optimization of formaldehyde adduct removal**. *J Mol Diagn* *13*, 282–288.
5. Lipson, K.E., Wong, C., Teng, Y., and Spong, S. (2012). **CTGF is a central mediator of tissue remodeling and fibrosis and its inhibition can reverse the process of fibrosis**. *Fibrogenesis Tissue Repair* *5*, S24.
6. Yoon, J.-H., Srikantan, S., and Gorospe, M. (2012). **MS2-TRAP (MS2-tagged RNA affinity purification): tagging RNA to identify associated miRNAs**. *Methods* *58*, 81–87.
7. Grosso, S., Doyen, J., Parks, S.K., Bertero, T., Paye, A., Cardinaud, B., Gounon, P., Lacas-Gervais, S., Noël, A., Pouysségur, J., et al. (2013). **MiR-210 promotes a hypoxic phenotype and increases radioresistance in human lung cancer cell lines**. *Cell Death Dis* *4*, e544.
8. Thai, P., Statt, S., Chen, C.H., Liang, E., Campbell, C., and Wu, R. (2013). **Characterization of a novel long noncoding RNA, SCAL1, induced by cigarette smoke and elevated in lung cancer cell lines**. *Am. J. Respir. Cell Mol. Biol.* *49*, 204–211.
9. Razungles, J., Cavallès, V., Jalaguier, S., and Teyssier, C. (2013). **[The Warburg effect: from theory to therapeutic applications in cancer]**. *Med Sci (Paris)* *29*, 1026–1033.
10. West, J.A., Davis, C.P., Sunwoo, H., Simon, M.D., Sadreyev, R.I., Wang, P.I., Tolstorukov, M.Y., and Kingston, R.E. (2014). **The Long Noncoding RNAs NEAT1 and MALAT1 Bind Active Chromatin Sites**. *Molecular Cell* *55*, 791–802.
11. Ricciuti, B., Mecca, C., Crinò, L., Baglivo, S., Cenci, M., and Metro, G. (2014). **Non-coding RNAs in lung cancer**. *Oncoscience* *1*, 674–705.
12. White, N.M., Cabanski, C.R., Silva-Fisher, J.M., Dang, H.X., Govindan, R., and Maher, C.A. (2014). **Transcriptome sequencing reveals altered long intergenic non-coding RNAs in lung cancer**. *Genome Biol.* *15*, 429.
13. Kornfeld, J.-W., and Brüning, J.C. (2014). **Regulation of metabolism by long, non-coding RNAs**. *Front Genet* *5*, 57.
14. Yang, F., Zhang, H., Mei, Y., and Wu, M. (2014). **Reciprocal regulation of HIF-1 α and lincRNA-p21 modulates the Warburg effect**. *Mol. Cell* *53*, 88–100.
15. Quinn, J.J., Ilik, I.A., Qu, K., Georgiev, P., Chu, C., Akhtar, A., and Chang, H.Y. (2014). **Revealing long noncoding RNA architecture and functions using domain-specific chromatin isolation by RNA purification**. *Nat. Biotechnol.* *32*, 933–940.
16. Yang, Y., Wen, L., and Zhu, H. (2015). **Unveiling the hidden function of long non-coding RNA by identifying its major partner-protein**. *Cell Biosci* *5*, 59.

17. Khandelwal, A., Bacolla, A., Vasquez, K.M., and Jain, A. (2015). **Long non-coding RNA: A new paradigm for lung cancer.** *Mol. Carcinog.* *54*, 1235–1251.
18. Masoud, G.N., and Li, W. (2015). **HIF-1 α pathway: role, regulation and intervention for cancer therapy.** *Acta Pharm Sin B* *5*, 378–389.
19. Yu, H., Xu, Q., Liu, F., Ye, X., Wang, J., and Meng, X. (2015). **Identification and validation of long noncoding RNA biomarkers in human non-small-cell lung carcinomas.** *J Thorac Oncol* *10*, 645–654.
20. Uzunhan, Y., Bernard, O., Marchant, D., Dard, N., Vanneaux, V., Larghero, J., Gille, T., Clerici, C., Valeyre, D., Nunes, H., et al. (2015). **Mesenchymal stem cells protect from hypoxia-induced alveolar epithelial-mesenchymal transition.** *Am. J. Physiol. Lung Cell Mol. Physiol.* ajplung.00117.2015.
21. Carlson, H.L., Quinn, J.J., Yang, Y.W., Thornburg, C.K., Chang, H.Y., and Stadler, H.S. (2015). **LncRNA-HIT Functions as an Epigenetic Regulator of Chondrogenesis through Its Recruitment of p100/CBP Complexes.** *PLoS Genet* *11*.
22. Choudhry, H., and Mole, D.R. (2015). **Hypoxic regulation of the noncoding genome and NEAT1.** *Brief Funct Genomics.*
23. Yen, Y.-C., Hsiao, J.-R., Jiang, S.S., Chang, J.S., Wang, S.-H., Shen, Y.-Y., Chen, C.-H., Chang, I.-S., Chang, J.-Y., and Chen, Y.-W. (2015). **Insulin-like growth factor-independent insulin-like growth factor binding protein 3 promotes cell migration and lymph node metastasis of oral squamous cell carcinoma cells by requirement of integrin β 1.** *Oncotarget* *6*, 41837–41855.
24. Uzunhan, Y., Bernard, O., Marchant, D., Dard, N., Vanneaux, V., Larghero, J., Gille, T., Clerici, C., Valeyre, D., Nunes, H., et al. (2015). **Mesenchymal stem cells protect from hypoxia-induced alveolar epithelial-mesenchymal transition.** *Am. J. Physiol. Lung Cell Mol. Physiol.* ajplung.00117.2015.
25. Liu, H., Wang, S., Ma, W., and Lu, Y. (2015). **Transforming Growth Factor β 1 Promotes Migration and Invasion of Human Hepatocellular Carcinoma Cells Via Up-Regulation of Connective Tissue Growth Factor.** *Cell Biochem. Biophys.* *73*, 775–781.
26. Zhang, M., Zhang, C., Zhang, L., Yang, Q., Zhou, S., Wen, Q., and Wang, J. (2015). **Nrf2 is a potential prognostic marker and promotes proliferation and invasion in human hepatocellular carcinoma.** *BMC Cancer* *15*, 531.
27. Chang, Y.-N., Zhang, K., Hu, Z.-M., Qi, H.-X., Shi, Z.-M., Han, X.-H., Han, Y.-W., and Hong, W. (2016). **Hypoxia-regulated lncRNAs in cancer.** *Gene* *575*, 1–8.
28. Tao, H., Yang, J.-J., Zhou, X., Deng, Z.-Y., Shi, K.-H., and Li, J. (2016). **Emerging role of long noncoding RNAs in lung cancer: Current status and future prospects.** *Respir Med* *110*, 12–19.
29. Choi, S., Chen, Z., Tang, L.H., Fang, Y., Shin, S.J., Panarelli, N.C., Chen, Y.-T., Li, Y., Jiang, X., and Du, Y.-C.N. (2016). **Bcl-xL promotes metastasis independent of its anti-apoptotic activity.** *Nat Commun* *7*, 10384.
30. Xing, Z., Lin, C., and Yang, L. (2016). **LncRNA Pulldown Combined with Mass Spectrometry to Identify the Novel LncRNA-Associated Proteins.** *Methods Mol. Biol.* *1402*, 1–9.